Security levels do not provide insight into the extraction or add to context, unless a security level and categorization mapping is somehow made available. Object-level, field, or pointer encryption renders the recovery map difficult to use. Since the recovery map contains limited contextual information, and much of the information is not even ordered in any way, forced decryption is not necessarily of any value. Encrypted pointers can indicate an arbitrary number of other encrypted fields or pointers. The extraction can be dispersed to various remote storage locations, can be stored in multiple locations, be validated, can be encrypted through 3rd party add-ins or use any of the DigitalDoors encryption and multicryption methods.

[1617] The recovery map points to dispersed recovery files. Access to recovery files is predicated on knowledge of the remote dispersed locations (network mappings), access passwords to those storage locations, and any DigitalDoors password restrictions to internal storage locations or granular storage at the extraction-level. Normally, this detail is cached for each user as part of the desktop configuration, cached by DigitalDoors for each user as different relative paths map to absolute storage locations, and contained within the recovery map and recovery files. Caching is protected by encryption. Although it could seem that the performance overhead for these layers is high, the CPU processing power available to the average desktop user is many powers of ten faster than even the fastest network connection. Network latency is the primary performance concern.

[1618] This single file or data unit, or dispersed data unit contains the extractions for the source, several extractions, one extraction, or part of a single extraction. This information is delivered sequentially to the user desktop, decrypted, validated, and recovered for the reconstitution. Disallowed extractions are not delivered to the user desktop, even in encrypted form except in the one case when the DigitalDoors security system is specifically configured for single object (one file option) that

IPR2023-00973
International Business Machines Corporation v. Digital Doors, Inc.
IBM's Exhibit 1002, Vol 2 of 9
Ex. 1002, Page 592

contains user-level access control, declassified source, recovery map, recovery data, and encryption.

[1619] In addition, certain failsafe options are designed to maintain security and protect reconstitution. The most obvious method to unravel encryption is to break the administrator-defined encryption key. However, this does provides limited access to administrator information only as other databases, maps, recovery files, and support files are widely dispersed.

[1620] Damage to the storage linkages and particular recovery files represents another approach to defeat what cannot be forced. Because DigitalDoors functions within standard file systems, backups provide the best integrity. Should that be destroyed, an automated recovery based on global network search would be necessary. This could not be performed without administrator access and oversight. The level of CPU overhead and network access would be very obvious. A scattershot approach below this threshold would not break the NP-complete problem of single-file recovery. This is a needle in many haystacks.

[1621] Damage to recovery files is assumed in the DigitalDoors security system. At worst, a user, even a high-level user, will see the reconstitution for the most part. At best, secondary storage will provide per extract recovery. Broken links, missing files, and other network or file failures do not preclude reconstitution. With standard backup and recovery process, recovery of damaged or missing files will allow full reconstitution.

AA.7.2.0 Malware attack the tentacles of Insecurity

[1622] Following is a description of how we understand a malware attack vector and how to mount appropriate responses.

[1623] It demonstrates why perimeters, application, workflow, and data are now longer protected by the methods derived from traditional security paradigms.

[1624] The number of open doors -- whether they are called edges, touchpoints, browsers, SaaS processes, remote users -- that provide a channel for attack are already overwhelming. They also increasing as hackers aggregate opportunities, make effective inferences about the value and vulnerability of targets, and interact with tools, applications, and systems accessible through e-mail, firewalls, or published services. Any open avenue, even legitimated ones, provides a doorway for breaches. Open doors create open sets that undermine perimeter-based security systems or stretch the perimeter out of shape and expose presumed closed sets to the ravages of Internet risk and errors and mistakes by insiders. Additionally, the efforts made to protect against malware before or after the fact require excessive expensively resources, are not always effective, and frequently add more difficulty that just the damage from the malware itself.

[1625] DigitalDoors laboratory's backup storage machine experienced a malware attack early in January 2008. This created a significant compromise and eventually a full meltdown with temporary loss of that resource. Restoration required a full system wipe, reinstallation from known source media, and reconfiguration of the backup environment. Although we burn backups by choice to read-only media like DVD/R and CD/R with single sessions that preclude additional sessions and prevent erasure or modification of backup sources, documents, demos, and support infrastructure, it is not sufficient. Metadata, SharePoint workflow, and complex Internet-enabled processes open doors for exploits to bypass process-oriented security. This incident revealed the failure of site, perimeter-based protection, and workflow or application security traditionally available today, and the corresponding need to rethink security hygiene. Security hygiene is a strategic business obligation from the top as a strategic continuity and competitive requirement but filters down and throughout the organization.

[1626] According to DigitalDoors any new doctrine must reflect total loss, immediate tear down

of assets, and relocation of both process and data to immediately-available alternatives. Assets should be cheap, easily replaced, and generally as low in value to the attacker. That is an expression of formlessness. Recovery or marshalling of these assets needs to be simple, rapid, and easily positioned to respond to business requirements or situational threats. Computing has to enter the age of rapid deployment as a reactionary force. This certainly puts a new spin on the easily-interchangeable value of commercial-off-the-shelf (COTS) over specialized or proprietary solutions at every level of the infrastructure.

AA.7.3.0 Security is broken

[1627] Fear, uncertainly and doubt (or doom), better known as FUD, is leveraged to maintain the security status quo. It is used to further entrench unsustainable methods. When the best and brightest experts propose variations of methods that are indefensible, or at best just delaying actions, the reaction is instinctual rather than intelligent and reactive. We are trapped in a cycle of resource depletion applied to the wrong war.

[1628] The current technology of security is ineffective. Efforts to support it, the variations f it, or the incremental improvements to it, are misplaced. Security is broken and no amount of bailing wire and toothpicks or duct tape will stem the leaks and thefts. The best efforts to date include burocratic creation of new titular posts for security as part of some blue ribbon commission. The commission rehashes old ideas with snappy new bullet-points, recommends new regulations, and suggests more interdepartmental cooperation.

[1629] In the end, if the blue ribbon commission is not soon disbanded, if the recommendations do not threaten the status quo, the commission creates a subsidiary panel to explore how the problems should be plugged. If it is government, the executive branch passes this hot potato to a legislative body.

If it is the senate, they hand the reports and problems to the representatives; the representatives trade their notes with senators. They all create new committees that pair off in topical investigative teams to detail the latest security threat.

[1630] Doubtless, the security breaches and leaks still occur, perhaps with greater frequency and severity. Some unlikely leader or even the panel or blue ribbon commission get saddled with the blame and are shown the door. But more likely, these movers and shakers move on to better titles, get reelected, and even campaign for higher offices based on this very visible report with press coverage. This process is constrained by too many chiefs and not enough indians, and a narrow concept that limits security change to piecemeal incremental change lacking cross-cultural innovation with leaps of technical advancement. While security is collapsing and undermining the competitive strategic advantages of the economic ecosystem, the blue ribbon panel is merely enumerating the reasons for the coming precipitous collapse. This is rearrangement of desk chairs on the Titanic. We are not implementing the necessary fundamental shifts that can forestall this catastrophe. This fear, uncertainty, and doubt is not converted into promises that might take away that fear. Fear is pointless unless converted into action with change.

[1631] We are defending ourselves with each insignificant pyrrhic victory against an unsustainable war of attrition. Attempts to augment barrier security by analyzing incoming and outgoing flows do little to caulk the pinpoint leaks at these unsustainable perimeters. These leaks are singly insignificant but together aggregate into a profound threat. Point solutions represent an endless arms race against asymmetric threats brought by unseen foes. Attackers bear resources to target evolving weaknesses of concern only in aggregate. It is not the one or two, or a even few attacks that threaten our sustainability, but the myriad assaults, each alone of limited concern, but the aggregate that presage doom.

[1632] As we try to maintain our organizational business, we are like a herd of migrating caribou under the suffocating onslaught of a cloud of mosquitoes. Caribou are ill-equipped to fend off such nuisance-level attacks that nonetheless decimate the herd. The ecosystem has changed, and without adaptation, the system will collapse completely. Instead, we run faster, climb higher, expend more resources against a threat that has fundamentally different rather than addressing the necessary goal in different ways. We are just not equipped for this situation. It will be a shame if we succumb to a suffocating onslaught without seeking other ways to address fundamental ecosystem shifts.

AA.7.4.0 Meta data dangerous double hedge sword

[1633] Data lacks purpose, structure and exacting patterns, context, and relevance without external references. Metadata provides those references as internal aspects of the data construct. As such, data has been downgraded as an important facility in favor of the more precise, extensible, and useful metadata. Long live metadata.

[1634] The information processing and library science worlds define the concept of data. Whereas information processing perceived data as a structured source for applications, library sciences sees data as the answers to questions and problems. Although the Internet, Gopher, Archie, WAAS, the world wide web and browser search technology began as an information processing concept, these now the complement to library sciences. As these search technologies and taxonomies for setting purpose, context, and references have grown to organize data, all data has become contingent on meta-tagging or labeling.

[1635] Applications, data, configuration baggage, scripts, complex formatting, documents with other types of documents inserted or referenced and external links are all forms of metadata. Many documents contain scripts, applets, functional code that might be static but more often is dynamic.

Viruses and worms are examples of unwanted dynamic modifications.

[1636] Applications often contain tables of data or reference externally linked modules with tables of data. This data can change, and with it the linked modules or even the application itself. Applications frequency update configurations and other variable data streams internal to itself. Some applications modify themselves and rewrite its internal code. Viruses and worms are examples of unwanted dynamic modifications. Applications often support persistence of data; this is the situation when user data, configurations, or metadata is stored within the storage space of the saved application. Data persistence provides several avenues for security lapses, modification of the data and the usage of the data storage space as a platform for a virus or worm. In addition, metadata provides all sorts of opportunities to hide polymorphic traps that can bypass signature scanners and even illegal process monitors. It is a rich game where the very extensible nature of metadata makes for a very extensible solution set but also a very perverse battleground. Every tool, even metadata, becomes a weapon.

AA.7.5.0 The transition content security via Data Leakage Prevention

[1637] Security has evolved from physical access control, computer rooms, and perimeter restrictions on hardware to include logical controls of networks and resources. Now security is evolving from perimeters protection for infrastructure, processes, and websites that process data to include the protection of data itself as well. The relative value of infrastructure and workflow supersedes the value of hardware, and now the value of information and the costs of reparations after information loss or ID theft supersedes the value of the infrastructure and workflow. The pressure to protect data has created new efforts and classes of products. Access control, encryption, multi-factor authentication, packet routing controls, in/outbound firewalls, deep packet inspection, and data loss prevention represent the bleeding edge of information security.

[1638] While new efforts address specific shortfalls in data security, each one of them represents a specific point solution, the repackaging of existing substandard solutions, and/or the aggregation of many other products into an integrated suite. These efforts parallel mathematically-proven but functionally-futile multiple level security (MLS) now being reinvented for commercial applications. MLS is expensive to implement, and after huge investments of, precludes knowledge discovery or information sharing, and disrupts operations to a point where functionality is preferred over any security.

[1639] For example, the newest security effort is data loss prevention (DLP), which provides perimeter control of data. It is not so different from any physical barrier, process-control firewall, or other perimeter control system. The difference is that the emphasis has shifted from hardware infrastructure to workflow and now to the data itself within the workflow. Perimeter security control for data will be as effective as any other perimeter has been. All the limitations and inherent flaws apply to this new focus.

[1640] Data loss prevention is hobbled by a number of functional impediments. Wholesale protection -- the construction of data silos -- have proven to recreate the functionality versus security conflict. Data must be classified to ascertain what subsets should be specially protected. Organizations implementating DLP have discovered that massive amounts of existing data and the magnitude of data storage growth precludes real-time categorization. Categorization is impeded by classification bias limitations. Furthermore, the creation of categorized information silos only shifts the flaws of perimeterization into a larger number of microperimeterizations, where resulting security is no better than with a single perimeterization.

[1641] DLP additionally presumes that categorization is universal. An entire document is all [TS] or all [U] with a binary restriction. It can be released or it cannot. A binary sharing decision is usually

too restrictive in practice. However, restrictions are functionally granular, because users, situations, roles, and other complex conditions modify the release. Categorization is also granular because some parts of document might be [TS] where others are [S] but the overall utilitarian view of the document is that it is [U]. MLS and MLS-like systems upcode that tagging to the highest level and preclude release. This creates a black hole accumulating most of the data ultimately lost in this silo. This recreates the conflicts inherent between functionality and security, where functionality typically wins.

[1642] Categorization for DLP suffers the typical tagging problems elsewhere described as reflexivity, classification bias, and the lack of a multitier hierarchical overlapping tag structure that establishes tagging for the presumptive process requirements but also other conceivable needs and likely later data repurposing. DigitalDoors enhances DLP and overcomes many of its inherent limitations through granular content control. Automated and assisted manual classification resolves the primary functional complaint of DLP implementation by tagging everything and coping with the data storage explosion. Bias and reflexivity issues are established by the supported multitier hierarchical overlapping tag structure. These compound and combination tags (out-of-channel) enable tuned release of sensitive materials based on the balance of facilitating production functionality while yet maintaining any arbitrary level of security at the same time. While DLP is not the solution for security, microperimeterizing with granular content control goes a long way to overcome the clear and present limitations is this technology

AA.7.6.0 The failure of best breed security

[1643] The "Best Breed" in security does not deliver the needed solutions. The best of breed chips away at the background noise but does little to protect against evolving attacks and intelligent adversaries. It is a rationalized effort based on inertia, customary roles, continuing practices, systemic reinforcement by internal supervisors and external vendors that assert it makes sense and that it has

always been done that way.

[1644] The best we have stops the loss of the lowest hanging fruit. Best of breed does nothing to divert the motivated, incessant, and evolving threats. If an attack fails but the prize is worth winning, the attackers return with revised efforts. There is a known the maxim, "if at first you don't succeed, try, try again." (Teacher's Manual (1840), Palmer, Thomas H. "'There is a lesson you should heed, Try, try again. If at first you don't succeed, Try, try again."). There is no reward for security attack failure, but no punishment either. There is so much noise that even law enforcement is interested in their lowest hanging fruit. Most failed attempts go unnoticed. Even the successful breaches of security rarely garner reproach, arrest, or repercussion. Botnet attacks hide the source and shield the attackers from any and all repercussions, and perhaps even generate opportunities to make new enemies where none existed, diverting suspicion to innocent parties.

[1645] The traditional best of breed for security is an aggregation of perimeter-based security methods and point solutions designed against specific attack types. Perimeter security is sufficient when the perimeter is actually sacrosanct and not distended to incorporate workflow integration and the ambiguous pel-mel of incremental process evolution. Calling a weakness in the perimeter an 'endpoint' is merely another means to hide the insecurity of touch point(s). Point solutions are sufficient when the security mechanism matches the risk profile one-for-one and covers multiphasic and evolving risk. Unfortunately, both perimeter and point solution sufficiency is unlikely. \

[1646] Perimeters get distended, and point risks tend to have multiple attack vectors but generally only one designated security fix, not the many discovered over the long tails. The application of some or even all of these approaches does repel some attacks and maybe even the majority of attacks, but not all of the attacks and allows more of the PII and trade secrets to leak out. Attacks that get through

are increasingly damaging. The lowest hanging fruit has either been protected or picked clean. The more valuable and more difficult fruit is under assault. Data fields and structured tables will give way to unstructured intangible property because the risk is dispersed and the monetary value is so much greater.

[1647] The best of efforts is forming panels to study security problems and pick the best of breed to avoid the mistake of doing something different. The best of breed is at best a faux security. Both effort and breed provide some rationalization that security does something; the incident logs prove that. It provides rationalization that incremental improvements will address new venues of attacks and patch existing gaps, so perhaps the environment will improve. Ultimately that is self-deception, a rationalization that somehow, something will change. But it has not and is unlikely to change within the traditional security venue. The best of breed is ultimately based on a broken security metaphor than cannot be patched, cannot be augmented, and can only be replaced by a new paradigm. Meanwhile, the best of breed is faux security.\

[1648] Best of breed falls into two categories. The first is the application of tools that are acknowledged to be provide the best techniques understood and available providing really nothing but the protection of lowest hanging fruit. That is faux security. It just feels good. The second involves a dramatic shift to a new paradigm filled with changes, new methods, and changes to the ways business function and the shift to a production model based on the information revolution. Such change is difficult, requiring agents of change, and an openness to concepts of asymmetry, knowledge-based production, and formlessness that describe the future of both security and wealth creation.

AA.7.7.0 Why Security of information

[1649] The reason for security is so we can anticipate the future. We want to known that what we have today we still have tomorrow. We want to know that the work we do generates benefits that

accrue to us, and are not stolen by someone else. It is a basic requirement for society and civilization. It is a particularly important requirement in complex societies where property is increasingly less tangible and can be stolen in many new ways. Security of property is relevant to the protection of intellectual property, the ideas and collections of information that define the information revolution, and the development of more complex civilizations.

[1650] Security enables us to plan, forecast, implement, and anticipate. Without the physical and social tools, we are unwilling to invest in a future. Without security or with inferior security, we do not have a future. We become unwilling to invest in it. We lack trust of individuals, groups, and any systems that might enable us and our activities. We have chaos, an inability to plan for and prevent technical, man-made, and natural disasters, or to preserve the various survival and social systems that enable civilization. Without security, at best we have a social status quo but more likely a descent into barbarism.

[1651] Although there seem to many different types of security predicated on the things each type protects, ultimately there is an overarching one. It is the security of information that frames data, ideas, process expressions, all tangible products and services, and all intangible ideas. Physical security and logical security are not different, but just different process expressions and protections at different points in the wealth production process.

AA.7.8.0 Security as an open set

[1652] The overwhelming limitation to achieving security is that a security space is an open set. Think of a set as a space that defines the physical and functional space bounded by the security mechanism, which is generally a perimeter. It could also be a wall, a moat, a fortress exterior, a building, a vault, an area surrounded by a firewall. The interior -- all of the interior -- is at risk when the perimeter

is breached. A closed set might indicate an air gap, a sandbox, a single device, or a single process. If that closed set can be appropriately defined and maintained as a closed set without a single path of entry, that closed set can be secured. The functional problem is that this security space can rarely be defined as a closed set. However, few sets can maintain that rigor and provide value. Most have a door for access. Even when that single door is controlled, that set is not longer closed. It is a limited closed set at best, and thus an open set.

[1653] This security problem explodes with process, functionality, access, and networking. Whether physical or virtual, most assets exist in this limited closed set. The limits are provided by access controls, routers, firewalls, guards, locked doors, user passwords, roles, and anti-malware. However, despite these limits, the limited closed set is exposed to unlimited access combinations. This is NP! in fact. The growth of risk is exponential. The more venues that are available, the more the limited closed set is opened, and the closed set really should be viewed instead as an open set. See also Security based on Closed Sets and Open Sets.

[1654] The issue of whether a limited closed set is an open set or not is pure semantics. Reality has demonstrated that lockdowns and access controls reflect anticipated security requirements. Reality falsely presumes that the paths of entry remain under control. It is also false to presume that any open set can be subdivided into closed sets with security. Closed sets preclude any access, thus functionality. Whereas, closed sets with limited access depend on the integrity of the access method, its resistance to direct attacks, indirect attacks, and social engineering. Closed sets with limited access might exist in the world of provably secured operating systems, but the reality of these systems is that at some point, assets get distributed to enable functionality (as with information sharing) and undermine the initial integrity of these closed sets.

[1655] Security based on the premise of an open set is more honest and has a better chance at productive success. Security based on the open set must presume formlessness. Formlessness masks the assets in terms of their values, utilities, and purposes. This formlessness creates a venue of assets without known value and risk assessment lacking basis, therefore, it yields a statistical environment far different from physical and virtual security systems.

AA.7.9.0 Dangers in information interactions, the need for granular content isolation

[1656] Purpose and interaction are as important -- or even more so -- to the virtual world than in the real world. The difference is important, although the results seem much the same. Purpose and interaction become very important within the security ecosystem but critical to the data ecosystem.

[1657] There are lessons that need to be learned from the physical world and implements into the digital world to avoid damaged caused by interaction.

[1658] In September 2006, spinach across America was tainted with a potentially lethal strain of Escherichia coli. Two hundred and sixty five people were sickened enough to be admitted to hospitals, some requiring organ transplants, and one woman died. Raw spinach was recalled. The reputations of growers, processors, shippers, and supermarkets were stained. Customers suffered. The assumption was, and the legal hope was, that growers or processors somehow precipitated a new endogenous factor that spread lethal bacteria. Two months elapsed before the vector was determined to be feral pigs roaming through California's San Fernando Valley farmland, an exogenous factor. Growers and processors had done nothing wrong.

[1659] Purpose and interaction clashed. The fields were operated as a controlled environment for a food source. Unfortunately, the feral pigs used the field also as a food source, a living space, and a natural bathroom. The pigs literally "interacted" and "aggregated" with the spinach to cause the

pandemic outbreak. The use of the fields for cultivation is sanctio. The freedom of feral pigs was sanctioned by hunting restrictions. Vociferous environmental groups exert political pressure to save the pigs. Surely, feral pigs will no longer be sanctioned given the conflict between health and environment. It is unlikely that farmers and politicians will consider effective preventive measures against the possibility of flying pigs. It is unclear how they will deal with wind-borne pollution or migrating international birds on a practical basis and also a legal one. The primary solutions are not viable.

[1660] This discussion is not belittling the problem, the cause of the problem, or the scope of the problem. In fact, if anything, this discussion is trying to show that the problem is really severe and the risk for similar scenarios will increase with the complexity and integration of the food distribution system. More complexity and the increased integration will lead to aggregation, inference, and interaction of minor problems into a pandemic outbreak.

[1661] Furthermore, this physical problem becomes even more perverse within a virtual cyber-environment. While the real world minimizes aggregation, inference, and interaction, this is represents opportunity for growth and increases in efficiency in the virtual world. The aggregation, inference, and interaction of data to data can replicate, clone, and transmorph risk factors far more readily than in the physical world because the time scales are counted in nanoseconds and process cycle speeds in Mhertz. However, herein the solutions are indeed viable.

[1662] The pigs interacted within a growing ecosystem. They might have been noticed. They might not. However, the risk vector was not considered until the damage was done. Next time, feral pigs will be considered a threat. It became a known cause and effect, thus foreseeable and preventable risk. If the vectors can be defined, what really can be done to isolate the vectors? You keep them physically apart in the real world with a fence because spinach and pigs reside in the same two-dimensional space.

Flying pigs, pollution, or migrating birds represent a different problem because of the difficulty in controllable the another dimension.

[1663] In the virtual world of information, the time scales, the cycle speeds, and the ability to create separated unlimited dimensions enable granular content isolation. You can apply this technology to tag information into taxonomies for three-fold results; the search for new information, the measurement and assessment of risk with a concurrent security control, and the classification and delivery of versioned information to enable secured multilateral information sharing.

AA.7.10.0 Prevention of leakage through context and range control

[1664] Access control and tagging (labeling and taxonomy) with subsequent redaction is the technology of choice to prevent information leakage by restricting access to information silos or the damage of information distributed as part of normal process, interagency sharing, or business partnerships. Access control and tagging with redaction creates an either/or scenario where security restriction gives way to getting work done. Granular content control enables more security and less leakage with fewer work restrictions. It is all about assessing context of content and its applicable range.

[1665] Context, the extra meaning that colors the inference, utility, purpose, and risk, extends beyond the apparent content in the data stream. DigitalDoors recognizes the relationship between context and risk to blend the measurement of content with the measurement of context and concept. The differentiation is the aggregation, inference, and interaction that extends the data stream through process, system, and infrastructure. Context extends beyond the evident, intended, and expected because content can be repurposed. The social security number was established for social insurance, but this identifier is unfortunately used as a unique customer key. You have open-ended risk without clear understanding of possible, likely, and probable extended meanings and usages. Knowledge discovery, risk

measurement, security, and information sharing is constrained within a traditional notion of the choice of security versus operations. DigitalDoors is measuring beyond the evident, to expose what others might see and certainly what a determined attacker might exploit.

[1666] Tagging and redaction is critically dependent on the meaning of words, phrases, sentences, paragraphs. It is also dependent on the hidden interstitial connections that extend across the words, phrases, sentences, and paragraphs. Although it is easy and common to redact a document in full and prevent its release, such complete lockdowns cripple information sharing and workflow. This reinforces the traditional notion of the choice of security versus operations. DigitalDoors and its paradigm shift of granular content control masters this limitation. However, in doing so, DigitalDoors changes the basic view of a document from simply content to a more complex view of the document as content, context, and concept. This is reflected in the structures, encoding, format, and metadata that part of the document itself, or wrapped external around the document through style sheets, process controls, application tools, or more complex workflow.

[1667] Filtering words, black list, white list, and gray list of potential, the extra-lingual connections of words that transcend content alone and are tied together by context and concept. Partly is human selection, partly is replicating this human selection against future process in the form of automation, partly applying constructs of context abstraction on top of the linear content across a range of content.

[1668] Content can be appreciated for its obvious meaning and how it is used. Content alone does not uniquely define meaning, and thus tagging. Note that content is modified heavily by context and concepts that might contained with the same data stream as represented by encoding, formats, presentation styles, linkages, and metadata. Also note that content is modified heavily by context and

concepts that might loosely coupled or externally referenced to encoding, formats, styles, purposes, workflow, other data streams, and well-understand situations. These internal and external effects in effect catalyze the meaning in ways not apparent within the content data stream.

[1669] For example, five years ago how many people would question a free promotional gift for signing up for a credit card inside an airline terminal? How many people today would question the very veracity of booth, the delivery of sensitive information, and privacy with which that information is protected on their behalf? The situation has been so altered that most people now weigh the risk of information disclosure against any freebies.

[1670] The combinations of words, phrases, words described by a specific context, the application, derivation, routing, purpose, ownership, imbedded or externally-referenced metadata, links, references, implicit or explicit aggregations of other data streams, inferences, data-to-data interactions represent a catalyst to the content and the meaning of the information as a whole.

AA.7.11.0The vulnerability of perimeter security and the need for granular data control

[1671] A fire all is a solid construction intended to separate one area from another area with a high risk of heat, fire, or explosion. A firewall might preclude smoke infiltration, but usually other orifices allow for smoke penetration leading to site evacuation, smoke inhalation, minor injuries, and smoke damage. A firewall is outright protection against gross events but not leakage. The firewall limits the extent of the profound destruction but not secondary effects. A firewall is not a perimeter but only a point solution of a perimeter.

[1672] A firewall has been extended to mean any separation between areas requiring protection from those with potential or active attack vectors. The problem with firewall security is that the extension of meaning is an attractive metaphor but not a realistic implementation. Security by firewall

provides protection from preconceived attack vectors but not attacks from other risk factors. A firewall protects one part of a building from intrusion of a fire from parked gas-filled vehicles; it does not protect from cat burglars, floods, storm-tossed debris, or an airplane falling from the sky. Within this tangible world, a firewall can protect against threats for which it was specifically designed, and maybe some additional ones, but not usually against unforeseen or specifically-engineered attacks.

[1673] Within an environment of information storage and processing a firewall can protect against preconceived, known, and anticipated threats but not against uncertain, novel, or engineered polymorphic ones. A firewall is outright protection against gross events but not data leakage and toxicity, or the intrusion, infiltration, and tainting of internally-stored data.

[1674] As the information revolution pushed the disproportionate value of intangibles over tangibles ever higher, secondary effects become more important. Downtime and smoke damage could undermine the credibility of a business, or prevent continuity and even survivability. An informational firewall might not be sufficient protection. Within an environment of information storage and processing a firewall can protect against preconceived, known, and anticipated threats but not against uncertain, novel, or engineered polymorphic ones. A firewall is outright protection against gross events but not data leakage.

[1675] Another reaction to effective but limited protection, like that of the firewall, is that failures that occur through neglect, other doorways, or because of the limitations are often misapplied to failures of that protection method and an institutional disdain for its worth. Whereas the firewall does work, reliance on it to perform a job beyond its design parameters often leads to rejecting it in its entirety. When the perimeter firewall is extended to a cloud, a grid, or other amorphous topologies, any visualization of a boundary is fiction, and protection is at best haphazard or incomplete. Virtualization

presents similar problems because the portable package is both application and data, neither of which is vetted for integrity or authenticity. Also, a virtualized chain-of-custody is extremely fragile because all subsequent steps depend on prior ones.

[1676] Because of this, a firewall is a poor metaphor for total security and even a poor expression for physical or logical security. A firewall is a plan against typical and topical risks but not likely and possible risks. When the risk profile changes, changes to an informational or logical firewall do allow for incremental adjustments to the known threat profile but not against uncertain, novel, or engineered polymorphic ones. A firewall is only a synonym for limited perimeter security, both physically and logically, and a perimeter is no match for conceivable threats of today's and tomorrow's environment.

[1677] The only total security solution known to DigitalDoors addressing the unknown risks, the polymorphic advance of threats, that overcomes the traditional tensions between security and functionality is granular content control. A firewall is not a solution for peripheral security or even perimeter security, only a functional shield against a predefined risk or a simple static threat that is not polymorphic.

AA.7.12.0 Leakage from Data combination

[1678] An emergent property is any newly realized attribute that arises from a combination of low-level entities, a change in ecosystem, a combination of previously unforeseen factors, or the increase of knowledge and wisdom that changes the perspective or repurposing of information. The primary concern for a emergent property is that the classification for information changes in any number of ways previously unforeseen or the perspective is radically altered leading to new security risks or actual leakage exposures. Although this is not per se a classification bias, the ramifications are the same.

AA.7.13.0 The Risks from XML

[1679] XML provides an unifying architecture that serves many masters, that of people, of applications, of process, of flexibility, of integration, of future-proofing workflows. The pertinent characteristics of this XML is that it is both human-readable but structured enough for computer operations. XML is a combined data presentation, format, and structure. Data from other types of presentations, formats, and structures are migrated or transformed to create the new XML package along with additional information needed to create that new package. In other words, XML is complex but supposedly transparent.

[1680] This transparency transformation comes with a price, a very serious price, in terms of the information revolution. It unintentionally leaks, delivers, and exports trade secrets and intangible property. It undermines security both actively by passing sources and methods to people and processes without the need to know, and it provides a covert channel for trojan horse attacks as a carrier for any and every type of data signal format. Its very flexibility is two-edged providing capabilities as intended and unintended.

[1681] What makes XML so flexible? XML rolls the data and formats and the structures into a single monolithic hierarchical package. Data now includes tags. Tags explain purpose, context, utility, sources, method, intent, and many other attributes at face content value. Tags link to other resources establishing complex and concise purpose, context, utility, sources, method, intent, and many other attributes at face content value. Tags link to other resources establishing more complex and more concise purpose, context, utility, sources, method, intent, and many other attributes at context value. Together, content and context defines concept. Format and structures add punctuation, arrangement, coordination, and meaning. Unauthorized people and processes can understand the XML package, even if you wouldn't want them to. That is too transparent. Code or process can be imbedded within for a self-

replicating, self-forming, and self-processing entity. What content and context is not obviously a security compromise can be combined to activate a functioning attack that passed content and context risk filtration. That is not transparent enough.

[1682] This flexibility enables an unlimited ranges of process migrations and data transformations that undermine security. XSL style sheets and XML transformations can redact the XML package, but generally this is a one way trip. Changes to the redaction are difficult to reintegrate back into the original package. As result, the orginal monolithic hierarchical package is preserved for all intended and extended functions. XML carries too much baggage. XML can also carry hidden baggage with dead, unknown, or blank tags and references. This is a way to covertly insert any signal inside XML. This establishes both untenable active and passive risks for XML. Transparency is not the solution for security; control is.

[1683] Traditional perimeter security methods fail, not as inbound filters but also as outbound filters. Additionally, the increased complexity and interconnectivity of workflow pierces the perimeter in so many ways that checks at each potential door become astronomical. Note templates and style sheets impart metadata. Finally, combinations of content and context, even if inert alone or even together, combined through this perimeter with the detonator to activate an actualized attack. The content and context does become concept until all the data, formats, and structure are together in a single place, like the XML package.

[1684] The only known method to us to protect data, in general, and specifically XML within the public information sharing environment is through granular content control. This method would apply XSL style sheets and XML transformations to redact XML information in versions and sensitivity levels with white, black, and gray lists to minimal formats and content suitable for each task, breaking

possible combinatorial concept trojans and preventing aggregation, inference, and interaction. However, reverse integration is absolutely required to synchronize changes to distributed data back into the original monolithic hierarchical package and maintain the functional integration of complex and interrelated workflows. This method both continues the flexibility inherent in XML while layering on a method to enable concurrent data discovery, information sharing, and sharing.

AA.7.14.0 The risks of new information infrastructures

[1685] The GIG (Global Information Grid) is a military-sponsored replacement for the successful but security-deficient Internet. Access to the original ARPAnet was restricted and it was secure within that very small and trusted community. ARPAnet only became useful, effective, and wildly successful as the familiar and unrestricted Internet. The Internet succeeded because it provided a general-purpose data information distribution channel with a high level of operational security in terms of continuity and integrity. However, because that small and trusted community has grown to encompass nearly the population of the planet, the Internet fails to provide data security in terms of confidentiality and privacy. This type of data security is fundamentally absent from the Internet. If encryption, firewalls, guards, intrusion prevention, antiviral solutions, and other revisional or traditional technology were sufficient to reverse the absent Internet data security, there would be no security plague, and thus no rationale for a separate but parallel GIG.

[1686] So, the rationale for the GIG is the creation of a secured Internet-like infrastructure. The proposed GIG security mechanism is rooted in limited-access, special-purpose parallel infrastructure "extending perimeter security to-the-edge." However, four cultural disparities play against the success of this security model for the GIG. First, separate but parallel is not sustainable and controllable even were viable federated identity controls available because the granularity of access controls transcends

traditional rank, role, and location to include aggregating and inferencing knowledge management risks; aggregating and inferencing knowledge management risks exists because contexts and concepts give new unforeseen value to disparate collections of data. Second, containment of data within a separate but parallel infrastructure defies the premise for the very universality mechanisms that made the ARPAnet into what we now know as the Internet. Third, security based on the premise of access restriction does not consider that a trusted GIG community is unlikely to be small. Fourth, data containment explicitly within the trusted GIG will be procedurally impossible because many other existing embedded applications and processes exist within the untrusted Internet requiring data transport between the two creating new avenues for at-the-edge perimeter breaches; a demarcation between GIG and Internet is a fiction and the introduced cancer that will eat away at GIG security. The separate but parallel GIG is neither a sustainable vision nor a security solution. The data security failure must be addressed specifically on its own terms and not through the false parochial perimeterization of a separate but parallel infrastructure.

[1687] As such, the assumption that the currently-envisioned security architecture for the GIG will resolve these flaws is woefully short-sighted. Even the Internet, the prototype for the GIG, is crying out for something specifically beyond traditional or revisional technology to provide the missing confidentiality and privacy. Yet market pressures have provided no effective solution. Perimeter security, bulk encryption, and controlled access fails. This does not address the decay of confidentiality and privacy because data replication, mining, aggregation and inference render control over data ineffective. Therefore, without the introduction of truly novel security paradigms -- not just revisions or more of the same ideas repackaged to look novel -- neither the prototype Internet nor the derivative GIG will achieve the necessary confidentiality and privacy to provide the missing systemic security.

[1688] Given that existing continuity and integrity security techniques have proven effective but not extensible to address security deficiencies in confidentiality and privacy, we need new paradigms. To merely assert that "privacy is dead and learn to live with it (Scott McNealy)" fails to address the forthcoming disruptions to infrastructure services, loss of productivity, and material intrusions to national security and interests. This will have profound economic consequences and military ramifications, if the minor samples already experienced are any indication. Instead, realize that data security is different and requires new paradigms to restore the security to confidentiality and privacy

AA.7.15.0 Granular content dispersion as security

[1689] When an information data stream (that is, a document, file, message, audio, video, or compound metadata structure) is parsed into its separate components the context is broken. The value of information is contingent on structure, format, encoding, linguistic reference, and other factors. As a result, granularization and dispersion enables security by similar mechanism to the one-time encryption pad. Although the one-time pad presumes a random seed, DigitalDoors approaches a level of randomness for each granular content dispersion. Although there are some indications of context (location, language, storage facility, date, etc.), padding occurrence counts and chaff can generate any arbitrary level of statistical randomness. Basically, all interpretations of the granular content are equally likely, thereby providing no special or usual meaning.

AA.7.16.0 Security by the One-Time Pad

[1690] In cryptography, the one-time pad (OTP) is an encryption algorithm where the plaintext is combined with a random key that is as long as the plaintext and used only once. The key is called a "pad" because the encryption key was often preprinted as a page in a bound pad of paper. A modular addition through XOR is used to combine the plaintext with the pad. It was invented in 1917 and

patented a couple of years later. If the key is truly random, never reused, and kept secret, the one-time pad provides perfect secrecy.

[1691] It has been proven that any cipher with perfect secrecy must use keys with the same requirements as OTP keys. The key normally consists of a random stream of numbers, each of which indicates the number of places in the alphabet (or number stream, if the plaintext message is in numerical form) which the corresponding letter or number in the plaintext message should be shifted. Messages in the Latin alphabet, for example, apply a key consisting of a random string of numbers between 0 and 25, the 26 capitalization-independent letters of the standard alphabet. The key for binary messages consists of a string of 0s and 1s.

[1692] The top sheet of the "pad" could be easily torn off and destroyed after use. For easy concealment, the pad was sometimes reduced to such a small size seen only by a powerful magnifying glass. Captured KGB pads fit in the palm of the hand, or even within a walnut shell. To increase security, one-time-pads were printed onto flammable nitrocellulose (as in old-style photographic paper or film).

[1693] The one-time pad is derived from the Vernam cipher, named after Gilbert Vernam, one of its inventors. Vernam's system was a cipher that combined a message with a key read from a paper tape loop. In its original form, Vernam's system was not unbreakable because the key could be reused. One-time use came a little later when Joseph Mauborgne recognized that if the key tape was totally random, cryptanalytic difficulty would be increased.

[1694] There is some ambiguity due to the fact that some authors use the term "Vernam cipher" synonymously for the "one-time-pad," while others refer to any additive stream cipher as a "Vernam cipher," including those based on a cryptographically secure pseudo-random number generator (CSPRNG). This latter encryption workflow is not as secure as the one-time pad or true randomness

because pseudo-random numbers can be recreated from the initial seed state and key reuse compromises security.

AA.7.16.1 Perfect Secrecy

[1695] The Vernam-Mauborgne one-time pad was recognized early on as difficult to break, but its special status was only established by Claude Shannon in 1947. (A unifying theory known as information theory was developed and became the subject of intensive research. Claude E. Shannon, whose initial ideas appeared in the article "The Mathematical Theory of Communication" in the Bell System Technical Journal (1948) defined information to include the messages occurring in any of the standard communications media, such as telegraphy, radio, or television, and the signals involved in electronic computers, servomechanism systems, and other data-processing devices. The theory is even applied to the signals appearing in the nerve networks of humans and other animals. The signals or messages do not have to be meaningful in any ordinary sense.)

[1696] He proved, using information theory considerations, that the one-time pad has a property he termed perfect secrecy.

[1697] The ciphertext C gives absolutely no additional information about the plaintext. Thus, the á priori probability of a plaintext message M is the same as the a posteriori probability of a plaintext message M given the corresponding ciphertext. Mathematically, this is expressed as $H(M) = H(M \mid C)$, where H(M) is the entropy of the plaintext and $H(M \mid C)$ is the conditional entropy of the plaintext given the ciphertext C. Perfect secrecy is a strong notion of cryptanalytic difficulty.

[1698] Despite Shannon's proof of its security, the one-time pad has serious drawbacks in practice. It requires perfectly random one-time pads, the secure generation and exchange of the one-time pad material, which must be at least as long as the message, the careful treatment to make sure that it

continues to remain secret from any adversary, and is disposed of correctly preventing any reuse in whole or part -- hence "one time." Furthermore, delivery of the keys to receiving parties and protection of the message after decrytpion and secondary process or redistribution breaks the security workflow. Distribution of wholesale messages between a large number of people or automated processes today stresses the ability and practically of one-time keys and hybrid encryption with PKI systems when unique keys are generated. There are also problems of message size and efficiencies even when hybrid encryption is applied such that the message is encrypted with random symmetrical key and that key is protected with a public/private key pair.

[1699] Because the pad must be passed and kept secure, and the pad has to be at least as long as the message, there is often no point in using one-time padding, as you can simply send the plain text instead of the pad (as both are the same size and have to be sent securely). However, once a very long pad has been securely sent, it can be used for numerous future messages, until the sum of their sizes equals the size of the pad. Implementation difficulties have led to one-time pad systems being broken, and are so serious that they have prevented the one-time pad from being adopted as a widespread tool in information security.

[1700] In particular, one-time use is absolutely necessary. If a one-time pad is used just twice for two plaintext messages, simple mathematical operations can reduce it to a running key cipher. When both plaintexts are in a natural language (e.g., English, Russian, or Chinese), even though both are secret, each stands a very high chance of being recovered by heuristic cryptanalysis, with few possibly ambiguities. Of course, the longer message can only be broken for the portion that overlaps the shorter message, plus, perhaps, a little more by completing a word or phrase. The most famous exploit of this vulnerability is the VENONA project. This is particular a concern in modern communications where

information process, structure, source and destination, prior message contexts, aggregation of messages over time, and inference can be used to exploit patterns or individual words in the message.

[1701] The one-time pad does not provide a mechanism to ensure message integrity and, in theory, a man-in-the-middle attacker who knows the exact message being sent can straightforwardly replace all or part of that message with text of their choosing which is the same length. Standard techniques to prevent this, such as the use of a message authentication code, can be used along with a one-time pad system, but they lack the perfect security the OTP itself has. The one-time pad is effective for one-time security. Because the forced decryption creates equally likely outcomes, the one-time pad is a perfect example of implementing formlessness. Once the document is decrypted and revealed, security for that document is exposed and questionable. Decryption initiates a rapid security decay. Reliance even on the integrity of the still-encrypted form for later use is inadvisable because the key and the content has already been exposed creating a process vulnerability and a contextual weakness by means of the release of source, method, and content itself.

AA.7.17.0 Risk in repurposing of information

[1702] Information in storage, information in motion, and information in process is at risk for diversion and repurposing. Often primary protections of encryption, firewalls, and isolation are bypassed because of the ongoing workflow and the information is exposed directly to misuse or reuse in distribution. More frequently, the pressure to perform and generate results puts the security at risk as well as the data itself.

[1703] In the article, Virtuality becomes Reality, DigitalDoors demonstrates how virtuality becomes reality when the context of reality can be manipulated specifically to enable virtual crime. Although virtual crime and kidnappings has become reality in Mexico, Uraguay, and Brazil, the extent

of information manipulation is far more extensive. As transportation, military events, commercial distribution, banking, and insurance rely increasingly on automation, all of these activities are at risk for diversion, fraud, and repurposing. While authorization, encryption, and validation are a basic necessity to assuring integrity in communications, the chain-of-custody for such communication channels is simply insufficient.

[1704] Control sensors can be manipulated to feed a deliberate data stream, capture control and reposition, or gather your own data stream. Deliver a false data stream from a prior session or an alternate realtime session. Misrepresent the location of the sensor or the device posing as a sensor. Denial of service attacks, static communication, or false feeds are all valid threats that can tip the competitive balance. It is also useful to statistically evaluate potential damage assessments and divert operations to less valuable targets or run a shell game ahead of operations to reposition valuable assets before these attacks. An example of such information repurposing is the actual disembarcation timing and location of the U.S. Pacific carrier fleet on December 7, 1942. Its position, or the reason for its position outside of Pearl Harbor, has never been confirmed. The U.S. Navy states the Saratoga and Lexington were at Wake and Midway Islands, and the Enterprise was in San Diego for retrofit. Conspiracy theorists suggest these three ships and their escorts were actively and secretly dispatched to intercept the Japanese armada just after their bombers were launched but either the timing was wrong, or they could not locate them in the vast ocean. Alternatively, the U.S. carrier fleet was simply respositioned for passive protection with political gain.

[1705] As sensors and assets become virtualized, the risks for manipulation are real and increase. Take control of a Predator drone armed with a hellfire missile and redirect it against friendly forces, an alternate target, or return it to base and the control center to destroy that control site. The core

DigitalDoors philosopy presumes dispersion of key resources. The return of the armed drone to the command center provides navigational and location information to a key facility, a venue for disruption or destruction of that critical control site, and potential to leverage destruction to other drones and support facilities at that site. The potential force multiplier is higher than direct enemy attack. Attacks through proxies muddy the origination of the attacks and retaliation is most likely to escalate the initial damage.Virtualized attacks allow for plausible deniability, distance from control or responsibility, and the inclusion of conspiracy theories as explanations.

[1706] The bombing of the Chinese Embassy in Seriabo remains a public mystery. It is always possible that intended target was indeed hit with the expected collateral damages and political consequences. However, it is not definite whether obsolete or out-of-date information lead to the mistaking targeting of the wrong building, whether the pilots were diverted by in-flight information, or whether an external party -- either a known enemy or even a competitive party with secondary gains -- manipulated the bombing run and redirected the target to a friendly site. We can speculative and invent all sorts of conspiracy theories. That does mean that conspiracy theories are not true, do not actually happen, or cannot be actualized in the future. Also, a true explanation of the events as they really happened and the intent behind these events reveals sources and methods are best kept secure. If history is any lesson, even the revelation of the Zimmermann Telegram 80 years after the fact exposes political secrets between friendly governments and still-valuable spy secrets.

AA.7.18.0 The ineffectiveness of Perimeter security

[1707] The effectiveness of perimeter security is a myth. Perimeter security has inherent weaknesses and flaws that preclude information sharing and distribution. The literature for defeating perimeter security includes flying under the radar, blockages, sieges, stealth mode, insiders, spies, and

so forth. It has always been flawed because every perimeter must have its doorway, if even just one designed into the architecture. Containers with any openings have no viability. Perimeters often exhibit other access points in dimensions not considered or assumed low risk, such as befriending the guards, kidnapping relatives and blackmailing them, or crashing through a non-load-bearing wall or even a reinforced firewall.

[1708] A perimeter is the line (or complex and convoluted shape) that provides the demarcation between outside and inside, the same definition for the open and closed set. The geometric difficulty with circumscribing the multi-dimensional perimeter becomes more apparent when the process and data storage is extended to grids, clouds, or virtualization. Too many processes and distributed data storage schemes defy even arbitrary, fiat, or judgmental determinations of the boundary. The overlaps between yours and mine and theirs is so uncertain. What at one instant might be correct changes to something else; these battle lines are just not stable.

[1709] Perimeter security is trustworthy only so long as the defenses equal or outmatch the offenses of the attacker. Unfortunately, this does not work and really has never worked. As the perimeter becomes a virtualized concept, its very definition becomes harder to ascertain with methods less certain for any physical and logical defense.

[1710] Perimeter security is simple, however economically viable, but not absolute, and therefore subject to defeat. There are four ways to defeat the perimeter:

Overwhelming force

Flying under the radar

Stealth mode

Corruptive insider

[1711] Sieges throughout history defeated perimeters through overwhelming force or the action of wearing down the defenders and the industrial support for the war machine. Every immovable object can be confronted with an irresistible force. The actual results, while unclear, are certain, the failure of the perimeter. An overwhelming point of effort is applied at some point in a perimeter -- generally where least expected -- and no amount of response or reinforcement is effective or timely. Even castle gates, however narrow and byzantine, provide a point of access that must exist and can be overwhelmed. Often such doorways seem economically impregnable, so new doors are punched through the castle walls at a presumed weak point. Although the virtuality of information technology is not subject to attacks by force, denial of service and repurposing of resources are effective attack vectors on the weakest part of the perimeter. Furthermore, the physical infrastructure of the virtual world can be physically attacked and disabled. It is sometimes effective to attack the industrial capacity supporting the war machine or the physical computing infrastructure, including power, communications, cooling, or the people maintaining it.

[1712] "Flying under the radar" means quite literally bypassing the perimeter security system by exploiting weaknesses in the monitoring system. The fact that such a term exists at all speaks to the known failures in perimeter security systems and the inability to plug them. Less literally, "Flying under the radar" means forging an effective new technology that is not perceived as a competitive risk until barriers to entry are established and a sustainable competitive advantage is assured. It is the arms race. Aerial or subterranean efforts represent historical methods to fly under or over the radar. Ladders and siege machines, tunnels, ramparts and towers overcame the architectural integrity of the perimeter security. This is no difference in the virtual world today, despite efforts at software quality assurance or secure coding techniques.

[1713] In stealth mode, aircraft and ships pass through the perimeter radar by directly defeating the ability of the system to see them or by appearing deceptively as something less threatening. Examples include warriors hidden inside a wooden horse (Trojan War) or in delivery wagons under the animal stock (Robin Hood). This is a problem of classification bias. Sometimes, appearances are masked so that dangers are cloaked in false purposes, pretenses, structures, normal traffic, convoluted and dispersed into pieces without perceived risk, or hidden so deep as sleepers as not to be recognized as a systemic risk. Stealth mode is literally the process of forging and deploying a radical new technology in secret, the craft expressed today as malware.

[1714] A corruptive insider represents a breach of trust. Insiders are spies, turncoats, or the repurposing of resources and processes against the security infrastructure and the industrial war machine. The insider is allowed through the access controls, passed through the single castle gate, or is a process with outside, inside, and perimeter components. An insider belongs and breaches his trust or exceeds his authority for his hidden agenda. Air gaps and guards preclude functionality, minimize bandwidth, disallow flexibility, but still provide for a crafty attack on all systems and methods. Even guarded access can be overcome with earned trust that is not warranted, however narrow or long required to establish that trust. Compromise of passwords, defeat of multifactor authentication, expansion of authorization beyond assignment is a form of insider corruption. Once within the perimeter of the castle, everything, every secret, every risk, every potential is within this soft underbelly. Perimeters within perimeters, also known as microperimeters or microperimeterization, merely represent defensive positions to regroup and do not really enhance security; they merely raise the cost to defeat the perimeters. It is merely a matter of time until this corrosion is exploited in its allowed format or expanded and extended for bigger and faster exploits.

[1715] Microperimeterization is also a method to imply the creation of information scarcity or dispersion, but it is really only the repackaging of one perimeter into many. In some cases, more perimeters does not mean more perimeters as difficult to breach as the original one; they could require less effort, less time, and even provide overall effective security. Microperimeterization does not create scarcity, just the implicit distribution of information that might seem like scarcity. Distributing something in a thin layer is not scarcity because reconsolidation dispels any reality of scarcity when it is all there again. Likewise, it is just distribution but not dispersion unless more than content is dispersed. Effective security dispersion requires that context and concept also be dispersed and stored independently from its content.

[1716] Other attacks are composite and polymorphic, a blend of these four basis perimeter attack styles. Layered security is undermined by the effort of layered attacks, each layer designed to penetrate or otherwise bypass each complication in turn. Polymorphic blends start as one type of attack, cloaked in a stealth mode or appearing as something else, then transforms progressively to bypass each layer of security. As perimeter security becomes more layered and more complicated, or microperimeterized, attacks will match these techniques step for step to unravel them. Taller siege weapons forced taller walls, gunpowder and cannons forced thicker walls and reactive armor, aerial warfare precipitated three-dimensional defenses, and ongoing technological developments promote this endless arms race.

[1717] Because the economic value of perimeter security is clear as a viable first line of defense, it persists as a security method, but it alone is insufficient due to its inherent weaknesses and is logically flawed in the face of security open sets. Mobility, asymmetric warfare, n^{th} generation methods, and formlessness counteract this attack escalation by masking, dispersing, and even intermingling the target resources and providing a different venue for security. Intermingled dispersion raises the ante in that

information from 3rd parties, perhaps even the attacker, is at risk too. Security by this type of formlessness is about sacrifice, deflection, and transformation. It is subtle this formlessness with granularity that is the only viable response to the security arms race and the perfect adjunct to the weaknesses of perimeter security.

AA.7.19.0 Deficiencies in perimeter security and encryption

[1718] Information security is onerous, perhaps even impossible because controlled closed sets are for all intents and purposes just exposed open sets with controlled touch points. Although touch points might seem under control, security reality is usually very different and these tend to be at risk. This situation is made more complicated because the information content of products and services increases daily. Intentionally hostile conditions expose opportunities for reuse of information to embezzle resources linked to or referenced by that information. Many legacy systems and applications were never designed for security in the first place.

[1719] Even evaluated systems (per Common Criteria assurance) are only defined to be protected from all but determined and well-funded attackers. Common Criteria was intended to describe a government effort, but the context has changed to include criminal organizations and even lone individuals. Most financially-motivated attacks are now well-funded, so even the most security-conscious and tested systems are not assured to resist what are now the typical threat profiles. Most financially-motivated attacks are funded better than the collective efforts for security against those attacks. The bar for collective security assurance is set at the same height for at least the last 20 years but the common attack profile is far above that level. This defines the current state for information security deficiencies.

[1720] Security used to be a mature but niche product until the paradigm of the information

revolution significance lowered the relative importance of hardware and process infrastructure. Networking and Internet access created this new 'wild wild west.' The pain today is due to this disconnect between this changed security significant and a market which has not reflected these changes or adequately responded to the new security requirements. While analysts propose a comprehensive multifaceted layered approach to security, where are all the tools? The paradigms for providing security are inconsistent with the changed threat landscape.

[1721] Few organizations, even those with huge budgets, are achieving information security, privacy, and confidentiality. Most organizations are merely trying to maintain ongoing services by deprecating the full potential of the security threat. As a result, identity theft breaches are incurring huge costs and becoming an extensive legal obligation. Agency and military programs are being suspended due to the conflicts over benefits exceed the risks from fundamental first amendment rights and the unsubstantiated potential for abuse. Recent publicly-acknowledged security project suspensions include:

Total Information Awareness (TIA)
Talon
Sentinel
MUSIC
Truth Project
National Security Decision Directive No. 77
Rail Head

[1722] Conflicting goals undermined these long-term agency and Pentagon military projects. It is consistent with the complexities of security versus functionality, a controversy still unresolved to date, and the root of security inadequacies. No one else, other than DigitalDoors, has found balance between operational security and preservation of operations, or the means to adapt between these two contentious obligations.

[1723] The increasing information content in virtually all products and services increases the

significance and the risk of this information. The global economy and the multinational sources of both products and services increase the complexity of production and distribution workflow, and correspondingly the reliance of information communications and content. It has raised the predicted risk and the actuality of these security lapses. Agencies and other major organizations have tried to address the risk for decades. The success of such efforts remains ambiguous, or for others very clear and very poor as number and success of attacks demonstrates. The complexity and overhead obligatory to confirm to some minimally-sanctioned basis of protection means that multiple solutions overlap and compete for resources with line-of-business operations. Some of these solutions might make these problems even worse.

[1724] This might explain why security journals are floating a trial balloon against any need for AV and perimeter protection. How could no security be better than at least some? Could the state of security be so bad that people and organizations would rather do nothing than something? Is the technology so poor that risk management, self-insurance, and remediation after-the-fact with its attendant legal and criminal consequence is more palatable than securing the crown jewels. It will get far worse before it gets any better.

[1725] Efforts to modify or enhance existing security tools or integrate point solutions together generate inadequate results. Point solutions defend against specific threats, and the bundling of point solutions suggests a broader protection net. However, point solutions only protect against specific threats and the bundling still leaves holes in the net and opens channels through which malware and other attacks slither through. Some of these retrograde integration efforts are opening new holes and avenues of attack, including the bundled tools themselves, and doing more harm than good. The reactionary cobbling together of existing marginal solutions to address a presumed market need does not in fact

address the legitimate compliance and technical threats. More is not necessarily better. Furthermore, small pieces do not always add up to a whole requirement. Most of these tools are sited at what are perceived as organizational perimeters and rely on data stream filtration using black lists and white lists. Even though advanced artificial intelligence techniques or statistical tools like Bayesian filters augment the lists, it is an arms race being won by the attackers at these perimeters.

[1726] Cobbling together point solutions also retains the value of existing tools, knowledge, experience, and workflow. It retains its familiarity. It retains the value of the prior investments. It sometimes doesn't matter whether these tools work sufficiently or address the next generation of attack vectors so long as they provide some benefit. Change is disruptive, expensive, and provides no guarantee that it will be better. Change is a big hurdle to overcome in any system. It is an even bigger hurdle in the security sector where the future is a cipher and the where the future is prepared under a cloud of fear, uncertainty, and doubt.

[1727] Information security efforts to date are based on perimeter protections (inbound and increasingly outbound too), protection of proxy devices and processes rather than the specific protection of the information itself. Efforts include encryption of actual information and displacement and reduction in the size of perimeters. None of these efforts directly address the risk inherent in information itself or provide more than surrogate protection. In fact, that is exactly what proxy means, and the ramification of proxy protection methods.

[1728] Specifically, protection of the perimeter, the servers, the storage, the applications, the workflow, the filtering software itself in no way directly addresses the content of the information or addresses an explicit security need. Because these technologies protect the container and not the content, the content remains at risk. This is proxy protection. It is a surrogate -- not the real thing. There are

legitimate reasons for protecting the proxies, the infrastructure, and the ecosystem, but this is proxy infrastructure protection not explicit information content protection.

[1729] Although denial of service attacks against the proxies, disruption of operations, and defacement of web sites are costly attacks, data breaches are not specifically protected by the existing security technology. The risks exposed today, the actual breaches making headlines, the billion-dollar legal torts, and disclosure of functional military response lapses, specifically result from information theft. Information is slipping through this leaking perimeter and proxy-bound security technology.

[1730] A perimeter is a fuzzy concept at best, all the more porous when pierced by local networks, wide-area networks, storage networks, and the Internet. Perimeters can contain internal structures, a perimeter within a perimeter, a gated community within a policing zone. Security tools rarely explore the containers with perimeters within those perimeters. Data files transport other data files, metadata, tags, inclusions, attachments, XML, DICOM, binary objects, geospatial files, and all types of databases. Most of these themselves can contain all those file types and more inside themselves. Therefore, a good perimeter check does not really stop at the perimeter but explores other internal structures. Because the hierarchy of these layers within layers, like Russian Matrioska dolls, is endless, realistic perimeter protection is insufficient. This is true for inbound and outbound protection, and since context, and encryption being a possible context, usually stops exploration at a functional perimeter and does not explore the granular details inside.

[1731] Perimeters cease to have legitimate meaning with grids, clouds (SOA and SaaS), and virtualization. The interchange of application packages with data sets midstream from one system to another represents a security crisis. Certainly, VMware and other product lines are enhancing management tools with security, but how can we expect these to work when security for simpler flatter

architectures has yet to be resolved? Virtualization is beginning to support the redelivery of applications, themselves supporting virtualization like SQL Server, creating another level of Matrioska dolls.

[1732] Encryption is the only widespread solution to protect mobile devices, Internet transactions, and backup tapes. While encryption, alone of all the legacy technologies discussed in this paper, does explicitly protect the content of information, it is undermined with major functional, application, and systemic flaws. Insufficient implementation and complex workflow also undermine the effectiveness of encryption.

[1733] Encryption performance is not a convincing flaw. While the complexity and overhead of encryption is a concern, it is not really a viable compliant since public key delivery, symmetric methods, and coprocessors can reduce overheads to acceptable ranges. It is not so much that encryption takes time and space, but that it gums up the existing operational workflow and tangles that workflow too.

[1734] Other encryption workflow problems are severe. Encryption imposes a security versus functionality penalty. Encryption key management and distribution might be too complex for widespread utility and the implementations to date collapse as the scope and number of included applications increase. Encryption key management and partial (as in granular) control creates NP!-complete management problems. Perhaps this complexity imposes a larger security and control problem than the initial problem itself.

[1735] Encryption is not a free ride. It is a weapon you might think you control, but you can lose it or find it used against you. It is a weapon that complicates the utility of the rest of your security arsenal. Encryption of metadata, covert channels, malware, and other risk factors impair perimeter protection and point solution technology. Consider how encryption can be used internally or externally

against your own resources, as in encryption ransom, outbound leakage of IP within encrypted streams, and the application of encryption to mask covert channels in plain view. Because the data flow is encrypted it will seem safe, and it will be very difficult to expose the content and check it for its potential for self-damage. When encryption tools are available they are used as initially intended but eventually for nefarious uses.

[1736] Additionally, key escrow and inspection of encrypted content at perimeters creates a new perimeter exploit opportunity, since encrypted content needs decryption at this perimeter for deep inspection and imposition of the same or the new encryption to the next endpoint in the chain of distribution. Renewed or new encryption represents an attack point and violation of source-to-destination integrity promised with digital signatures. While encryption and digital signatures provide content security and authentication, they both create a security lapse at perimeters.

[1737] This is a simple risk that a simple example explains. The physical security countermeasures at airports and buildings force you to open packages and demonstrate that the contents are not contraband and if innocuous really work as expected. A complication of open packages includes those lost at these checkpoints. Another complication is that even if a demonstration with a piece of electronics works does not preclude other covert functions too for that same package. A working radio can also be a bomb; a working data set can also be a carrier for any malware, botnet, or trojan. Even encryption unwrapped does not preclude layers of covert attack channels with perimeters of encryption or containers with seemingly innocuous content.

[1738] Air gaps provide security through restriction of access and the total disruption of external network functionality and operational workflow. The use of air gaps undermines workflow, reinforces the inherent segregation of information silo storage, and increases paranoia without addressing internal

security flaws. This is the ultimate clash of security versus functionality. The idea is well-meaning since it provides complete security against external risks. It also halts all external or inter-domain functionality. However, air gaps contribute nothing against internal threats.

[1739] The creation of sandboxes within perimeters or the reduction in the size of perimeters while correspondingly increasing the number of perimeters and thereby spreading risks in no way addresses the basic flaw of perimeter security. Although it seems like a rational solution, it is not a logical or provable solution. This will not withstand rigorous design and analysis. Perimeter replication and displacement just spreads the footprint of the risks without reducing it. In fact, the surface area of the replicated and displaced risks often exceeds that of an initial consolidated perimeter. Replication and displacement moves the perimeters somewhere else and creates more of them with different flaws. Replication and displacement creates flaws that can be compromised one-by-one rather than at some integrated aggregate perimeter. This de-aggregates integrated solutions into point solutions, which contribute to another path of attack as described previously. Displacement of perimeters, in addition to these other flaws, adds to the physical and logical complexity of management.

[1740] Secure coding methods, code analysis, and other techniques reducing the number of bugs and design flaws represent a respectable avenue to minimize security flaws. However, this technique is a variation on perimeter security. Basically, security is a test of the ingress and egress through the perimeter, where specifically the perimeter is an application workflow. This application workflow is both at the mercy of other security methods in place and part of the breach that exposes the perimeter to risk. Integration, applications referenced by the code, patches, counterfeit patches, access to the information through alternate insecure channels or other applications, code from prior versions, can compromise the target code itself and the information workflow. It is only as effective as the

thoroughness of the analysis, and only as complete as the protection of the code deliverable. Code deliverables can be compromised by replacement of operating system elements, dynamic libraries, external libraries, failure in core components, or by diddling the deliverable package. Code deliverables are also easily repurposed in way unintended. Code deliverables with digital signatures suggest security through formal authentication, this security is but a small part of a larger exposed ecosystem. Hackers are attacking code structure vulnerabilities despite secure coding methods, encrypted application deliverables, and vetting with digital signatures by altering the executable in memory during runtime. At that point, the application is presumed to be secure and already adequately validated, so exploitation starts by locating it in the task list, mapping its memory footprint, and then altering key functional return routines and those values directly in memory. Storm, for example, has been observed to defeat well-known innoculating antivirus software in just this manner. This same parasitic technology can be used to compromise the BIOS, a trusted control module, an MILS micro-kernel, an MLS proven structure, an operating system, applications, and any configuration files, databases, data files, and inbound or outbound data streams.

[1741] Access control is seen as a security panacea. However, it is a proxy for functional access. Actually, access control is a perimeter restriction policy. It funnels access through the perimeter using a control point within the supposedly watchful control of administrators. The failure of access control is one of granularity, although vendors are trying to address this issue with role, timeframe, scope, and permission limitations. However, access can be compromised with stolen or counterfeit credentials, misused credentials, or migration to other areas or functions lacking such rigorous restrictions. Insiders can open a side door to bypass a front door, or an outsider once inside can create a trap door for future entrance, or a weak alternate entrance can be compromised. One way exits are often compromised as

covert entrances.

[1742] Mandatory access control (MAC) and discretionary access control (DAC) is an enhancement over generic functional access control when applied to file systems. In this case, such security is indeed applied directly to information rather than the proxy infrastructure. In fact, most operation systems support one, the other, or both methods, such as Trusted Solaris. However, MAC and DAC recreate the conflict between security and functionality, with the predisposed result of either a severe functional lockdown or information leakage. So while this enables information security directly, it often imposes complications in provisioning access control to applications and workflow and the associated matrix of data streams.

[1743] MAC and DAC are great ideas and work... after a fashion, but organizations chafe under the complexity and limitations. Vendors have created tools to ease the burden and reduce the security lockdown to regain functionality. This still represents a direct tradeoff between security and functionality. The problem is that the processes and information under the controls of MAC or DAC is not granular enough. For example, even when application code is created to distinguish between a payroll clerk and a human resource manager to protect employee salary data, other applications often are able to bypass the controlled field. When database security is enforced by row- and/or column-level locks, user with database administrator access or programming and design privileges can apply brute or subtle methods to select, extract, dump, reformat, or link to expose employee salary data. Database security has been enhanced by limitations on what the administrator, root, super user can do, but there are interesting ways to bypass these new restrictions.

[1744] In many case, the means to compromise the salary application is to access the database with the core database application and use the native query tools. If that does not work and security

precludes row or column access even after table export or dump, opening a Trusted Oracle database with Microsoft SQL, for example, could fully expose the complete table structure and row and column values. Consider opening a newer database file with older release or a non-trusted installation. Even if Oracle security is locked down with user IDs, passwords, and table restrictions, these are often bypassed through an alternative tool that predates the new security restrictions or fetches data sets with drivers bypassing the locks. Some vendors simplify conversions, upgrading, and transitions, especially with universal BI tools, and these yield a treasure trove of skeleton keys that make light of new database security restrictions.

[1745] Even with Trusted Oracle, multilevel security, and transparent encryption, exposing privileged database data often is as simple as rethinking the how you bypass the front door. Try a side door or a partially open window that is generally extant through classification bias. SELECT employee FROM table.employee WHERE rate > 120000 and rate < 121000 will expose a highly paid employee because the SQL request is not returning a locked field value but a result contingent on that protected field value. That is all part of normal operations.

[1746] Security restrictions are often badly implemented and do not deny functional access but prevent display access. When employee rows are locked by explicit rate constraints, an inner join on the employee list with rate ranges will conspicuously infer the desired answer by the absence of an employee response to a query. Even if an employee record is not even accessible because of record-level controls, inference with a relatively small number of queries will still expose the desired data. Lockouts and failures can be probed for answers just as well as the actual information itself. This is very traditional but scientific black box exploration. When the actual problem or data set cannot be observed directly, analysis by proxy, perturbations, or shadows infers results.

[1747] Security based on redaction for the creation of versions provides niche-specific security. This works so long as the context is consistent with the initial state and when results are returned within the boundaries of the design criteria. However, change the requirements, assumptions, design limitations, become creative with public data aggregation and secured internal data in order to penetrate security. Restore, re-identify, or expose the context of the supposedly secured information. Results might initially show an unusual level of ambiguity, but inference, aggregation, and interaction eventually will compromise distributed versions.

[1748] Versions increase flexibility and streamline management. However, the duplication -- even with a mediated risk -- still propagates the security risk with more avenues of attack. The risk increases because of the additional instances and potentially wider distribution. This effort is contrary to efforts at centralization described previously, however decentralization with information item-level control does change the risk profile when key items are asymmetrically extracted, as with DigitalDoors.

[1749] Centralized methods and resources falsely promise ease of implementation under the watchful control of a highly-skilled and efficient administration. However, protection of centralized methods and resources recreate the previously detailed perimeter security problem. Breach the perimeter and the risk is not just one information stream but all information streams. Break into the castle and pillage the soft underbelly of the site. Centralization increases the risk of cascading exposure. Application integration, workflow integration, source code errors, access control errors, and complexity remain an insoluble avenue for information security breaches. Disaggregation of the centralized resources does not resolve the inherent problems in displacing and dispersing perimeters, however, it does statistically spread the risk from actual information breach when information content is dispersed in a partial or granular form.

[1750] In this manner, and this manner only, decentralization and dispersion increases security. Dispersion of information into multiple replicated locations raises the point risk because there are more instances of the information to protect. Instead of making the risk smaller, this makes it bigger. However, the solution to this is to disperse some content of the information to replicated locations, breaking the context of the data stream and lowering the risk of information items. This more than compensates for the security replication risk. It is a significant paradigm shift.

[1751] Existing point solutions address risk and security problems strictly within the context for which they were designed; they are rarely if ever extensible. They fail to address the extension of risk permutations, perimeter intrusions, process and information repurposing. Reinvention and revisions of existing point solutions for newly perceived risk and security problems work as point solutions but also are not extensible. The aggregation of existing point solutions into a unified security marketable product addresses only the point risks in aggregate. It does not do more than that, and often creates conflicts between the integrated point solutions themselves. In addition, many of these integration results create adverse side effects equivalent to or exceeding the costs from the observable attacks themselves. That is bad medicine that is getting worse even as the disease itself gets worse. Aggregating point solutions just creates a new perimeter defense.

[1752] Notice the two traditional approaches to security today. Disaggregate the perimeter defenses which have been ineffective into point solutions that can be refitted as new point solutions, or aggregate the point solutions which have been ineffective into an aggregate solution that can be marketed as an new all-in-one integrated solution. Logically, each option is the inverse of the other. This converts a solution that is known not to work well into another solution that is known not to work well, and at the same time, converting the other solution known not to work into the former. It is churn. This is a

shell game where every player loses.

[1753] Further, the incomplete concepts of point solutions, such as routers, NAC, firewalls, IDS, IDS, virus scanners, and so forth are being reintroduced as new point solutions to limit cross-site scripting, SQL injection, and site-wide database attacks. These new products are point solutions for newly identified channels of attack, extensions of modest technology into new reincarnations of modest technology that do not do any better than the original modest solutions. They address some security flaws by proxy protection rather than specific and directed information protection. Clearly, nothing has changed.

[1754] So why did it take DigitalDoors so long? The problem itself is complex and the existing technology is insufficient, the solutions are not evident in existing literature, academic research, government efforts, or commercial products. DigitalDoors is swimming upstream against the contradictory efforts either to disaggregate perimeter solutions or integrate point solutions into new bundled tools. Consider a layered and geographical-driven approach to information security directly than indirect infrastructure protection.

[1755] That perception represents a paradigm shift not easily acknowledged, a new technology not yet accredited or widely tested, conflicts with years of security inculcation and propagandizing, undermines the experience and even the successful efforts of known subject matter experts, and contradicts the efforts of huge and influential industry. Recognition of limitations and the need for change does not come easily but usually as the result of some massive and consequential defeat. Maybe it is time to throw in the towel, realize the security deficiencies plaguing us today, admit defeat, and explore alternative paradigms.

AA.7.20.0 The breakdown of perimeter security and the need for Dispersion

[1756] "Crunchy Outside, Chewy Inside "is a common security metaphor based on outdated concept of security as a physical barrier and only as a perimeter. The metaphor describes containment of valuable assets within a tough external barrier -- the crunchy outside -- that when breached allows total exploitation of the desirable and unprotected interior -- the chewy inside. The chewy inside is also known as the soft underbelly. The outside is tough, but the inside is fully vulnerable when any part of the exterior catastrophically fails or even slowly leaks. One way leaks are only perceived as one-way; they are logically bidirectional and 'slow' is merely a velocity measure that is a function of bandwidth, time, and context.

[1757] The traditional military concept is shifting in response to asymmetric warfare, randomness, global interlocking relationships, and formlessness. There is a lack of clear demarcation between inside and outside, my assets and your assets, and unilateral damage when the infrastructure is most likely interrelated and collateralized between defender and attacker. My cloud is your cloud; it is an entangled resource.

[1758] The addition of single point solutions, aggregations of point solutions, and microperimeterization to partition the chewy inside makes practical sense but the reality provides little additional logical security. Stretched and ambiguous perimeters do little to help establish an integrity to a perimeter or define exactly where it exists. These techniques do not address difficult logical definitions of what is inside and what is outside, where outside and inside overlap, and where custodianship and interrelationships are entangled. Inside and outsite are also blurred with integration, communications networking, and the increased valuation of intangibles over tangible assets. Asymmetric warfare, randomness, and global interlocking relationships enable formlessness as the obvious offensive model but also with a corresponding defensive metaphor reallignment. Formlessness is the new security

paradigm.

[1759] Intangibles are increasingly represented as data that blurs any distinction between asset, copy of asset, loss of asset, partial destruction of asset, and the diddling violation of the asset. Furthermore, this duplication, replication, and accessibility to intangible assets blurs any fixed distinction between asset valuation to the defending steward, partners, and an attacking aggressor. Multiple values depend on context and repurpose concept. Point solutions, aggregations of point solutions, and micro-perimeterization adds complexity and compartmentalization with a corresponding loss in functionality and still without addressing the core security requirements. Security mega-suites do not address the changing environment adequately.

[1760] Security breaches, leaks, and collapses are not only the results of direct assault, siege, infiltration, but functional avenues as well. The types of attacks are no longer so overt, so physical, so tangible. Security failure is more complex with assets frequently not so clearly owned. Assets frequently overlap in custodianship, sharing, and applicability, and the information revolution is increasing the intangibility of these assets. The PCI/DSS documentation for security defines a touchpoint as any access point, system, conduit, or workflow that exposes an information item to exposure. It is an edge or an endpoint. A touch point is breach of perimeter security. Traditional perimeter security methods do not incorporate the risks of the touch point or the likelihood of many of them, or show how to wrap the security around these transient doorways. Consider the security failures when the touch point is perceived as transient, but to all intents and purposes, breaches the perimeter on a permanent basis.

[1761] The only method to address the chewy inside is to partition it and disperse it beyond functional value and reach. This formlessness of content, context, and concept when reversible, provides reconstituted value in intangible assets when and as needed, decreasing catastrophic failures and

exploitation of even slow leaks. Partitioning (microperimeterizing) is not sufficient with dispersal to preclude the reuse and repurpose of these assets. Reversibility through reconstitution with granular content control maintains the requisite functional accessibility with out the traditional conflict between functionality and security.

AA.7.21.0 Asymmetric Security

[1762] Cyber-attacks are often initiated by a small organization with distributed stolen resources against a facility that is large, immovable, and target-rich. The attacker drives a wedge at a deficient or the weakest component thereby applying virtually unlimited and irresistible force. When components are interwoven, failure of the weakest component provides both systemic failure and entrée into the soft underbelly of the facility. Countermeasures are few. Retreat or redeployment is impossible. Retaliation is impractical because subterfuge by the attacker is likely. This represents the unequal application of power through covert, concealed, ghostly, and distributed subterfuge, the obvious evidence for (indication of) asymmetric warfare.

[1763] If the initial attack against the facility fails, subsequent ones can be tried with parameter changes, alternate attack vectors, or the application of additional stolen resources. If the initial attack fails, it can be tried against another similar target and still be reapplied to the initial target at a later time. Every cyber battlefield is not unique but a variation on a very common theme. This situation is most advantageous for the attacker using asymmetric warfare. Should the initial attack is detected and repulsed, it can be tried again. Although known attacks have economic impetus, some agency might be amassing the unknown ones for a first strike. Perhaps what is most scary is that the initial attack (or at least many of the multitude that take place) probably succeeded and remains undetected because the attacker refrains from profiting then and there; instead the attack vector is mothballed for a later-stage

planned military conquest. Successful attacks might include more than denials of service and theft of data and services but also seemingly benign probes that pass for noise or unknown anomalies.

[1764] How to thrive in such an environment? no symmetric countermeasures but asymmetric ones... attacks, counterattacks, and certainly facilities that are not a large, target-rich, immovable sitting duck.

[1765] Facility can not be monolithic... failure of whole is failure without recourse

[1766] Facility must be dispersed, moveable, mobile, flexible

[1767] Facility must have backup and alternatives

[1768] Components must be independent so that failure of one does not lead to failure of another

[1769] Components must be independent so that failure of one does create unlimited internal access

[1770] Components as in data must be dispersed so that that they cannot be turned to the benefit of the attacker or other third parties, or used to the detriment of the defender or other third parties...

[1771] Dispersed components should be inert, inactive, or neutral, or useless

AA.7.21.0 Suspicious packages as covert channels

[1772] Suspicious packages come in innocuous forms that ipso facto are not intrinsically suspicious. They are generally innocent, adjusted to look like formats from familiar and even trusted parties, or plainly Spartan. This is the bootstrap vector for computer attacks, data leakage, security lapses, and privacy compromises. Security requires that we exercise caution with all data streams. Botnet trojans, viruses, malware, and bootstrap slurpers that steal confidential data come in innocuous packages but these insidious traps are designed to get past perimeter firewalls and circumvent any natural suspicion of risk. Malware never says "hey, try me out, I'm malware"? Nevertheless, such ploys defy

both rational discretionary and mandatory protection mechanisms.

[1773] Confidential data, intellectual property, and proprietary bits and bytes go out freely or skip past outbound leak detectors in innocuous packages too. Inside risks are hidden under cover of friendships, trust built up over years or decades, man-in-the-middle attacks, or a lack of time and attentiveness to root out these uncertainties. A lot of risks fall below a threshold of recognition or are overlooked because the problem is not squeeky enough to get the necessary attention. Sub-clinical attacks do not initiate a defensive posture and can linger unnoticed indefinitely, and still be ultimately incurable.

[1774] Covert channels are concealed in innocuous files, attachments, inclusions, whole file encryption, encrypted insertions, metadata, tags, formatting, strange errors in data streams, hiding under, on, even inside every proverbial rock. Management agents and monitors are designed to work for your best interests, but hackers readily repurpose these tools against you. Protocols, network packet structures, agents working on your behalf and also the delivery wrappers for these vetted tools can be infused with covert channels. There are millions of creative people out there with a motive to pick every lucrative pocket. Encryption might be a solution to one problem but it is also a wonderful tool to create new uncertainties and even hide them in plain site. Encryption can mask the malware under a wrapper of security and cloaked in the complexity and authentication by digital signatures. Every package, and even just its wrapper, is suspicious.

[1775] Digital signatures and other attribution tools can vet the source and destination of such packages until such time as we become inattentive about that too, but never the sanitation of those packages. Every data stream and every file is a package without attribution. Every tool we use to protect those packages and the sanitation of the content can be repurposed to create covert channels. So every

data stream and every file package is inherently suspicious. The chain of custody, while vetted throughout its length, still has a starting point that is ultimately based on some unsubstantiated level of trust.

[1776] Be suspicious of innocuous packages, small files, data sets, common everyday messages that do not fulfill the mandates of processing, require operations, or belong. Why, because we can create programs to read a lunch menu, the Oxford OED, or... a trade secret formula. Just one bit at time, mind you, and tweak just one bit with that extra bit in some gray-flannel suit files within an arraigned or calculated position. We only have to pass just one bit per file. It is a leak, a leak so small and so slow nobody will mind it. The sub-clinical theft is only a matter of bandwidth, and bandwidth is almost unmanaged today. We can smuggle out that trade secret in a stream of traffic, over one hour, one day, one month, even over one year -- past any guard. We need only extract those bits to reassembly them into an exact copy of your trade secret. If we think that the leak has any risk of detection, we can even spread chaff into the traffic and recycle a CD of old viruses as a ruse to busy detectors and keep the watchers self-satisfied that they are accomplishing a job while we are surreptitiously doing the job on them.

[1777] Information is a complex assemblage that includes raw data, structured data, other forms, metadata, intellectual capital and property, processes, communication infrastructures, and forms far more intangible. Issues of context, purpose, intent, and how the information can be aggregated for new purposes also become critical security consideration. Information includes structured, unstructured, semi-structured, Intellectual Capital and Intellectual Property. The definition of information is subject to much academic debate and competing theory. However, DigitalDoors whitepaper, What is Information?, provides overview on this topic in a review of the types of information, the format of information, the purpose of information, and the risk factors inherent in the aggregation, inference, and interaction of

information.

[1778] We need to protect information by creating information security that is security that specifically protects the information rather than the infrastructure, the perimeters, the processes, and the workflow, as we do now. Information is not about protecting the proxy but protecting the information itself. We go naked. Not naked like in no firewalls, but naked as in reducing the complexity of information formats to the minimum required for the job. No metadata, no compound files, and complete understanding of the intent and context of the information. For additional information about how to implement a Security Hygiene as part of the DigitalDoors Doctrine, see these whitepapers for additional depth in these subject matters.

[1779] Then, how do we protect the information? We know encryption usually works, except when encryption is used against us, so that is problematic. Instead, we need to disperse the risk, but hide our secrets in the clear. DigitalDoors protects information by extracting the crown jewels and creating a single generic and universal redaction. Multiple redactions, that is versions of redactions, are useful too, except when the versions get distributed out of control. The DigitalDoors universal redaction is controlled. The DigitalDoors universal redaction can be reversed with none, some, or all the extractions, with fine-grained reconstitution under a granular content control. The information itself is secured with control over contents, purposes, potential for reuse and abuse.

AA.7.22.0 Dispersal security versus perimeter security

[1780] The subdivision of a perimeter, whether by moat, castle wall, firewall, silos, containment, or physical protections, into microperimeters disperses the risk of complete security failure but does not disperse the content at risk or provide security through dispersion. Microperimeterization creates subsets of perimeters that might be better, equal, or no better than separate protections. Microperimeterization

is compartmentalization or even subcompartmentation. Think of ice cube trays that minimize slosh but do not really provide isolation between adjacent compartments. It provides internal delays and battlements when an outer perimeter is breached but more confusion than security within the greater perimeter. However, when the microperimeterization uses the same basis as perimeterization, the level of security is the same although the number of instances is increased. Often microperimeterization provides delays and complications without any real increase in the magnitude of the security.

[1781] Dispersion is the structural disassembly of the content at risk, often into units with minimal value, and its distribution such that compromise of any single unit is equal to or less than the substantative value of the dispersion itself. Partial and incomplete compromise might preclude partial or full value of the content at risk. The process of granular information dispersion breaks the content, context, and concept of the information to preclude aggregation, inference, or interaction. The distribution of the fragments precludes intelligent aggregation or partial inference with certainty.

AA.7.23.0 Metadata risks, the need for a lockdown

[1782] The following shows how metadata extends the value of data, introduces obvious and hidden risks, and how to lock down data security while enabling concurrent means to share data. Although many published articles outline the risks of hidden metadata in document distribution, these articles do not explain how metadata introduces these risks. It is of utmost importance to realize that semiotics, through the search for meaning in the greater environment, extends the risk for data starting from the definitive to the unforeseen and merely hypothetical. If metadata is after all just data, not so different from raw data streams, how does it create such risks? This paper explains why granular content is a solution for the data ecosystem.

[1783] The simple answer is that metadata is not inherently riskier than any other data. Metadata

is just data that is allotted so that it is not always visible nor accessible. Furthermore, most data includes implicit or explicit metadata. The difference is format and whether styles and structural formats are internalized or externalized. XML tags, database table structures, and other internal demarcations are obviously metadata. A separate style sheet or transformational application that formats the fetched data for display is also two different forms of metadata. An externalized implicit format is still metadata even if it is not written anywhere; it might be hard-coded into an application that segments and parses the fields hidden in packed data or obvious to a human reader who makes assessments of the syntactical or grammatical structure. In fact, application code, macros, control functions, and most other aspects of information technology are reducible to just data, and should be construed as a data file. Compiled application code, while clearly a process workflow, is also a transformed data set. Also see Punctuation--Context alters Concept

[1784] If you harbor any illusions that a raw data file is not metadata and no matter how simple and minimal does not include reference to metadata somewhere, reconsider the meaning of its existence, name, extent, properties, and location in any filing system. For example, because the raw data file exists, it implies work accomplished. The file name provides some designation of utility or purpose. Extent often is linked to the application or process that opens the file. Properties define ownership and user access controls, date of creation and last use, size, numbers of access; and in trusted systems logs of who, when, how often, and with what application accessed the file. The location of the file within a hierarchical file system also ascribes the functional purpose of the file. Consider that the file path components define ownership or stewardship and designations of utility when the path contains a directory with a file type name or purpose. Image-based file systems with a hierarchical tree view of files and thumbnails of content or unique icons also impart meaning.

[1785] Although it is possible to create a file system hierarchy with random paths and names, semiotic analysis of the metadata defining the environment surrounding that file still provide rich meaning to that file independent of its content. When data file includes metadata, the semiotic implications of the file increase. The risk of data containing metadata is not hypothetical; it is a statistical certainty.

[1786] The meaning, value, risk, security, and sharing of data is far more complex than white lists, black lists, and gray lists or firewalls, encryption, access controls, and perimeter protection. The semiotics (meaning and interpretive understanding) of data is complex but represents a first step in recognizing the value and consequential risk of data and how to secure it and share it.

[1787] DigitalDoors represents data with a triumvirate of content, context, and concept as a simple mnemonic to illustrate the complexity in meaning. Content, context, and concept is not the full spectrum of meaning; it is just a shorthand. The data stream, its encoding, format, structure, references, labels and tags, hidden text, comments, macro code, and internal structures such as tables of contents, indexes, and footnotes establish the content, context and concept. It also sets intent, purpose, utility, the decay rate and half-life for the data, and who or what should use it. These categorizations overlap in that content also sets context and concept. Likewise context and concept establish content too. Order, placement, tags and labels, a defined metadata hierarchy, and other data structures establish meaning.

AA.7.24.0 Competitive Analysis of Security Methods

AA.7.24.1 security Technology

[1788] Digita Doors security technology is represented by several techniques. This includes everything from:

Table:DigitalDoors techniques

Making the process complicated in some way

Encryption

Ciphering (non-character substitution)

Data compression and other encoding

Data hiding (or data masking)

Steganography (the process of hiding a messaging within the content of a more obvious one)

Winnowing (validating integrity and non-repudiation)

Access control

Biometric access control

Situational access control

Data Hiding

Data Sharing

Remote monitoring

User identification

Remote visual identification

Identification of friend or foe

User authorization

Cascade of authentication

Chaffing (injection of rubbish)

Winnowing (checksum compression)

Secure maintenance

Firewalls (or moats)

Data dispersion (storage and delivery)

Binary (ternary and et cetera) information delivery

[1789] Digital Doors applies all of these techniques in one form or another, but primarily focuses on the two core techniques of data hiding, and binary storage with delivery. Note that process times are at least comparable to encryption—in fact, several magnitudes faster—plus provide some other functional and workflow benefits. Specifically, while security might be the primary focus for these technologies, DigitalDoors are enables survivability (risk mitigation), privacy, confidentiality, and anonymity under the same umbrella.

Table of other security techniques,

Virtual private networks (VPNs)

Intrusion Detection

Firewalls

Proxy servers

Encapsulation

Convolution encoding (taken to an extreme by Rinjdael)

Obfuscation

Statistical encoding

Forensic control

Strikeback

[1790] DigitalDoors does not require as large an initial commitment to creating new infrastructure and it works with existing infrastructure. It is not dependent on a bet of cipher technology or the distribution of it.

AA.7.24.2 Encryption and Security History

[1791] The earliest recorded history of message encryption and message hiding (steganography) goes back at least to the failure of Persians to conquer Greece. The Persians had expected to surprise the Greeks, but the Greeks turned the tide and trapped the Persian fleet in a confined bay. Advanced warning delivered as a message on a wooden board obscured by a layer of wax from a Greek expatriate spurred the Greeks to build up their munitions and surprise the Persians instead.

[1792] So-called "unbreakable" encryption unraveled by code breakers lead to the loss of her head by Mary Queen of Scots, the German U-Boat fleet, the Japanese Navy Admiral Yamamoto, and many others. No encryption has proven secure in actual use. No method has been mathematically proven beyond doubt to be undecipherable. In fact, even the best of the modern algorithms rely on the difficulty of locating prime numbers, which are now under pressure from faster CPUs and new technology based on quantum mechanics. Some encryption methods require a trusted third party, which in the past has shown itself to be untrustworthy, susceptible to bribery, and even been the puppet (CryptoAG) for an unfriendly spy agency. The germinating philosophy is to address the damage and risk through mitigation rather than barrier-based security.

AA.7.24.3 Cipher Science

[1793] Lengthening the encryption key is a false effort to improve security because the distribution of prime numbers is asymptotically logarithmically decreasing rather than exponentially increasing. A 56-bit key creates a number that is 22 digits in length with 612489549322870 prime numbers. A 112 bit key is 46 digits in length with 4309606392583910000000000 prime numbers. In other words, while the key space is 72057594037927900 times larger, the prime number space is only 70362121237035500 times larger, by direct prime number generation and counted enumeration.

Approximation methods other than enumeration (practical as the key gets larger) to count primes include the Gauss's empirical postulate, von Mangoldt value, Chebyshev's ϕ function, Perron's formula, Dirichlet series, and Reimann's Zeta function.

[1794] There are 2.4% fewer prime numbers between the 56-bit and 112-bit space, representing an 11% asymptotically logarithmic distribution of prime numbers. This slight reduction might seem inconsequential, but as the key size is doubled and doubled again, the loss is non-linear (2.7 and 3.4%). This undermines the potential for prime numbers as encryption keys and increases the effectiveness of brute force prime number factoring in decryption. Although the key sizes and the sizes of prime number sets are enormous (103 digits at 256 bits and 206 digits at 512 bits), Moore's Law wins on computational deciphering. A Swedish team cracked 512-bit DES in 14 hours in 1999. Besides, unwanted prime-number encryption has a non-linear 2 to 7% helper in the form of the asymptotic distribution. Note that the AES replacement for DES and Triple-DES is Rinjdael. Because it is a block dislocation cipher, it is at first glance immune to brute force factoring. However, Rinjdael (and any other encryption method) will succumb to modified cracking efforts. The government analysts for the AES group in charge of selecting from the five encryption process finalists in fact cite the following:

[1795] "The standard techniques of differential and linear cryptanalysis can be adapted to be used against Rijndael. Because of the way matrix multiplication works, and because in GF(2^8), all the coefficients of the Mix Column matrix (as indeed all numbers from 1 to 255) have reciprocals, a specific attack, originally developed for use against its predecessor Square, called the "Square attack", can be used as well... This interesting property, although not trivial to exploit, can be used to impose certain conditions on the key when one additional round, before or after the two rounds involved, is present. The possibility of this attack was first noted by the developers of Square and Rijndael themselves, and was

noted in the paper that initially described Square."

[1796] Encryption only provides protection point-to-point and so a long as all information is contained within the encrypted container. If the security needs require only point-to-point, virtual private networks are faster, more efficient, and less prone to complications. This conflicts with normal workflow by impeding the benefits of data sharing, networking, and Internet distribution. Encryption is also order-sensitive, such that any encryption applied to a previously encrypted container must be decrypted in the reverse order. DigitalDoors is not order-dependent, unlike encryption and other security algorithms. The many features and concepts of DigitalDoors bypasses the conflicts between information delivery and the business workflow needs of openness and information dispersion. Furthermore, it can always be applied, not only to protect against external attacks but also internal subversion. Encryption can be applied as delivery encapsulation, if so desired, but for practical reasons must be the last security step.

AA.7.24.4 integrating diciplines

[1797] DigitalDoors applied methods and technologies borrowed from other endeavors and solutions in information technology. The implementation is revolutionary, but the core is derived from the prior knowledge of other disciplines. As one example, the disassembly of information into small pieces and restoration of that information is based on theories of data packet communication. Ethernet chops data into 1512 byte segments, wraps this data into a packet, and sends these packets in arbitrary order, via paths that are arbitrarily defined by routers, into a temporary buffer where packets are received in arbitrary order, sometimes a few missing packets are requested to be redelivered, in order to recreate the original data at a new location. In a corresponding way, DigitalDoors creates controlled chaos within rigorously defined parameters, data redundancy, and a carefully orchestrated process to restore the original data in part or completely. DigitalDoors chops the data in more meaningful and contextual

segments—by meaning and for many different application purposes—but nonetheless the technology is evolutionary.

AA.24.5.Unique Architecture

[1798] The current security paradigms include access control and encryption. Access control implemented through passwords, authentication, biometric authentication, and physical certification is recognizable as username and password pairs and firewalls. Usernames and passwords are breached through a number of workflow complications, active demon dialing (brute force), adventitious mistakes, failure to enforce basic procedures, and social engineering. Firewalls are breached by brute force, implementation errors, software bugs, design flaws, new combinations of multi-pronged attacks, as well as adventitious mistakes, failure to enforce basic procedures, and social engineering. It is also important to recognize that firewalls do not protect against internal attacks or external attacks aided by internal help.

[1799] Access control, while widely implemented, is seen as a necessary evil that complicates workflow. Few applications enforce or include random or spot security checks. The number of physical or logical bypasses limits the effectiveness of access control. In addition, the massive workflow shifts attributable to E-mail, shared application access, application sharing, data sharing, networks of business functionality shared with partners and even competitors, exchanges, data distribution, web hosting, and wireless enablement severely compromises the effectiveness of current access control methodologies. Specifically, access is no longer physically controlled or actual location(s) even known. It is often mobile too. Distributed information extends the time frame for its accessibility to an unlimited interval, through screen shots, ongoing access, or downloads.

[1800] Encryption, variously implemented through obfuscation, ciphering, table lookup

substitutions, or mathematical processing, suffer from a number of flaws. The most obvious is its lack of mathematical integrity. No encryption method has been proven. All historical obfuscation, ciphering, and substitutions have been broken. The strength and protection from mathematical encryption, based on symmetric, public key, asymmetric, shared key, and others are subject to faster computers, parallel processing, new algorithms, and even new technology such as quantum mechanics. Despite the well-publicized flaws, failures, and breaches of encryption, the workflow and implementation problems with encryption are actually superincumbent.

[1801] The most undermining actuality of encryption is that the seeds of its own decryption are embedded within the data and encryption encapsulation. Statistical analysis is often applied to rapidly decipher the data contents and brute force is also a solution with sufficient time or improve algorithms. Brute force is not necessary exponential either. Content analysis of the decryption of partial results streamlines rejection of unlikely keys. In addition, since most modern encryption is based on paired prime numbers, prime number factoring also streamlines deciphering. Prime number lookup tables and prime number factor tables reduces the set space for decryption keys. Note that the distribution of prime numbers decreases by Reimann's Zeta function as the size of the key increases. When the key size is doubled (100%), the number of prime numbers only increases by about 2%. In other words, as keys get longer as they inevitably must, given Moore's Law and development of new technology and better algorithms, the keys become easier to guess... let alone factor.

[1802] Although the new US encryption standard will eventually use Rinjdael, RSA applies prime number generation. Consider also prime number searching using harmonic analysis described in the September 1988 German issue of the Scientific American, Spektrum der Wissenschaft. The article on number theory briefly mentions Stanislaw Ulam's Spiral, which is a way of mapping the set N of

natural numbers onto the grid N x N. Marking all prime numbers black, Ulam saw patterns emerge providing useful technology for mapping and locating prime numbers based on Hough's Transform.

[1803] Specifically, encryption is slow, even with supplemental, external, or coprocessor technologies. The increases in key bit size from 56 bits to 256 bits imposed a 64-fold increase in processing overhead. As it has become clear that even 512-bit and 1024-bit keys can be defeated within hours by corporate hackers and minutes by government agencies, the orders of magnitude required to apply more complex and longer encryption will slow encryption to a crawl. Although computing capacity has doubled every 18 months (in harmony with the principles of Moore's Law), the pressure on key length has outstripped that pace. Furthermore, computing capacity has doubled on the desktop and server end; supercomputer capacity has not kept pace, even with the parallel and scalar designs.

[1804] Encryption sluggishness is a two-way street. Encryption time is symmetric with decryption time. This is fundamentally true with both asymmetric and symmetric methods, private and public key methods, and most escrow and third party methods. As data access shifts from screen display to shared processing metaphors, the process latency times and cumulative queuing delays will represent times in excess of routing and network delays, undermining effective real-time operations.

[1805] Of even more concern are the complications encryption imposed on workflow. Although encryption can be implemented at both ends of a process (or multiple steps in a process) technically as an inline piped procedure, there are fundamental security and distribution control problems with keys. Specifically, the two parties or multiple parties must agree on the symmetric or asymmetric encryption keys. The key, whether symmetric or two part must somehow be delivered to the recipient. Keys can be intercepted, copied, and redistributed. Distributions of a single source to multiple recipients increases the workflow complexity in that either multiple unique keys must be distributed or partially

compromised by the single key. SSL and similar technology ships the key in the same time window as the delivery of the data thereby creating a very viable interception possibility. Interception has been demonstrated in the press.

[1806] Although the interception of encrypted information is supposedly protected by the encryption wrapper, encryption can fail with some social engineering. Specifically, a hacker or unintended recipient can participate in the distribution of keys by posing on the intended recipient, and thereby defeat all security whatsoever. Although encryption is in place, the hacker or unintended recipient has the key anyway.

[1807] Encryption keys have other social engineering problems. Alternatively, keys can be issued by key escrow agents or third party trusted vendors. As has been evidenced by many such similar episodes as the CryptoAG debacle, there is no guarantee that the encryption technology, the key issue, key escrow, or even implementation is free from trap doors, Trojan horses, and secret back entries. In addition, in spite of any trust assured or guaranteed through third party agents, there are no guarantees that keys will not be released by insiders, moles, paid informants, or under duress of court orders.

[1808] Encryption is also order-sensitive. If a data object is encrypted by one method and set of keys, then encrypted again by other methods or keys, recovery can only be effected by the reverse application of method and keys. As encryption proliferates inside of workflows, maintaining this order is not always feasible or possible, and decryption failures are becoming more frequent. When processing is combined with encrypting as part of workflow, the stepwise failure to decrypt and then process renders not only the encrypted message irrecoverable, but is also likely to abend the data processing operations with flawed "data." Note that knowing whether to use DES, Triple-DES, or how to handle multipart keys as part of an automated workflow is also important because the use of the wrong method will merely

pipe gibberish into the step in the process.

[1809] Failure to recovery encrypted data often entails complete data set loss or a complex process beyond the capabilities of most organizations to recover or reconstruct the sources. Lost encryption keys, damaged encryption containers, and process order failures render encrypted data irrecoverable. This could require a process restart at best or require data regeneration from the start. DigitalDoors is designed to address and overcome the flaws in existing data security technologies. It resolves a number of problems in security, survivability, privacy, anonymity, and confidentiality. The DigitalDoors paradigm is based on compartmentalization, dispersion, and chaos theory. It describes a condition of security based on the absence or critical and important information, the diffusion of critical information and components, and the use of chaos theory to distort context and reference with information scattering and increase complexity or entropy within the data information system (as defined by Jeremy Campbell in the Grammatical Man, Simon & Schuster, 1982). This technology has been replicated from agency, military, historical, and theoretical work by such as Sun Tzu, Ron Rivest, George Patton, etc....

AA.7.24.6 DigitalDoors versus Other Technologies

[1810] Please see Analysis of Competing Securities for greater depth on this subject. Security, privacy, anonymity, confidentiality, and survivability are today's prerequisites to functional information technology. Existing methods fail to achieve these goals in toto, in part, or in aggregate. As a result, a new approach is necessary to address each of these goals separately and all of these goals together.

[1811] Existing security technologies are incomplete and fragile. The security technologies are designed for deadbolt-like exclusion rather than internal security, external security against internally-mounted attacks, defense against intrusion, and custody for outbound distributions. Furthermore, existing

security technologies are not designed to provide privacy, anonymity, confidentiality, and survivability. Privacy is a problem addressed with ad hoc methods by merchants, customers, at both server- and client-side, with unimpressive results. For the most part, security is currently implemented with encryption and firewalls to thwart surveillance. These limit access to camouflaged assets. However, everyone of agency or hacking significance knows exactly where the assets are located, whatever the perimeter protection, and as such, can focus unlimited resources to defeat defenses and capture of those assets. This is an outdated mentality.

[1812] Anonymity and confidentiality are mandated by EU data directives, the Safe Harbors agreement, by the U.S. Medical Privacy Act of 1996 (to be effective by 2004), by other foreign and U.S. privacy directives that are in force now, will apply after statute dates, or are likely to be enacted in the future given the political climate. Frankly, implementation is manual and frighteningly archaic. Products, services, and computer technology seem aimed more at circumventing police surveillance than in proactive consumer and organizational protection.

[1813] Technologies for survivability, disaster recovery planning, and business continuity have been developing for quite a while. Backup, hot sites, and management planning is advanced and effective when properly implemented and maintained. Indeed, this industry is showing growth rates of 290% annually and representing about 40% of IT budgets. The Internet and remote backup technology in conjunction with high-speed links has made live, offsite backup and restoration feasible. This nexus of new technology makes DigitalDoors the next step in data continuity by parameterizing data, lowering storage costs, and creating new doorways for access authentication.

[1814] The short answer to "Why DigitalDoors?" is that this paradigm, technology, implementation, and application suite is designed to provide security, privacy, anonymity,

confidentiality, and survivability as a complete, interrelated, and integrated solution. DigitalDoors represents a paradigm for safety based on data selection, data extraction, data dispersion, data hiding, and data delivery with layered-access reconstitution not previously discussed in the literature or implemented as host, server, appliance, or end-user applications. We have implemented this technology and it is available now.

[1815] The more technical answer for "Why DigitalDoors?" can be given as a comparison with competing technologies, such as encryption, steganography, and firewalls. For example, DigitalDoors does not need to establish key exchange with a 3rd party who is using a different computer hardware or software; 3rd parties do not need identical encryption and decryption programs, they only need a password or a biometric verifier. Strong encryption is strong relative only to the availability of computing power to break the keys. Moore's Law, or rather his prognostication, suggests that because the power of a microprocessor doubles every 18 months, new technology will have the power to sequentially break encryption keys. No encryption is mathematically proven unbreakable. Social engineering attacks are common to guess keys. Brute force attacks are common. DES encryption system, the existing standard for law enforcement, was cracked by a $200,000 computer of the Frontier Foundation and has been cracked within a period of days with keys as large as 512 bits. Third party key escrow is no guarantee of protection. Key recovery by law enforcement agencies creates a questionable back door. Should encryption regulations become prohibitive again, this system does not rely on encryption for its core functionality; therefore, DigitalDoors can avoid any current or future restrictions on encryption. By definition, breakable security measures include:

Table: breakable security measures

Encryption

Firewalls (or moats)

Ciphering (non-character substitution)

Steganography (the process of hiding a messaging within another obvious one)

Access control

Biometric access control

Situational access control

Remote visual identification

Identification of friend or foe

User authorization

Data dispersion (storage and delivery)

Binary (ternary and et cetera) information delivery

[1816] Even chaffing and winnowing have serious functional limitations, as explained since Ron Rivest first suggested his alternative to strong encryption export limitations. Specifically, chaffing and winnowing have not been implemented to date. Although they demonstrate some potential (and indeed DigitalDoors applies some of the concepts), chaffing and winnowing have not been implemented alone or endorsed as a viable security software. Some research scientists have formerly criticized it as complicated, as using extraordinary amounts of bandwidth, and its mathematics as insecure. Steganography, a method of communicating secrets by hiding secret information within another piece of information in such a way that the adversary can not discern the presence or contents of the hidden message, also purports to be an open method to protect data while publicly sharing that data. Examples include:

Microdots

Watermarks

Data embedding

Secret sharing (dual key access)

Data hiding

Digital signatures

Covert channels

Spread spectrum communications

[1817] However, steganography has serious limitations. Primary, the assumption is that the defender is smarter than the attacker, such that the attacker will not figure out the methods used to hide useful data within the structure of chaff. Ultimately, multiple contexts of chaff and the subverted information become apparent. Tools get developed to find edges where edges do not belong, where information somehow represents disjoint or clashing sets, or semantic confusion is at odds with the functional purpose of the information. Ultimately, the key data stands out from the chaff in how it functions, looks, reads, or formats.

[1818] Serious limitation includes that microdots are not practical in a digital environment and generally can be located very easily. Watermarks require bandwidth, can be removed, manipulated, or decrypted from the source. Watermarks are often obvious and do not protect information. Secret sharing is difficult to implement and usually requires the co-location of sharers. Data embedding and hiding suffer the same limitations of chaffing and winnowing, in terms if overhead and lack of mathematical certainty. Covert channels are effective only so long as they remain unknown. Spread spectrum is effective only so long as analog and digital spectrum analysis is not used to define the channels and the hops, unlikely even now with current technology. Spread spectrum efforts to protect cellular telephone

NAM and PINs are broken within hours of each new technology release.

AA.7.25.0 Security by dispersion to the cloud

[1819] The DigitalDoors selection, classification, and dispersion technology was designed primarily to address a distributed universe of processing and data storage. The cloud is a composite of data, process, and workflow where boundaries represent a distinction no longer viable. The cloud is a single integrated ecosystem that ignores firewalls, perimeters, guards, and security point solutions.

[1820] The first DigitalDoors illustrations defined a network within the traditional framework of a cloud, even multiple clouds reflecting multiple layers of security both isolated from each other and interacting together through Digitaldoors technology. Implementation was applied to enable and support traditional data processing operations within a structured, field-oriented, free-form, unstructured, and database-oriented architecture. Data structure (the lack thereof or complexity) was not perceived as a design constraint or limiting factor. Processing technology for this new paradigm, the generations of information processing equipment, and networking interconnections, were not constrained by available concepts. DigitalDoors anticipated information as the primary resource and the expression of both process and distribution within its own framework as an extension of an evolutionary information and production model.

[1821] Furthermore, the basic dispersion technologies were invented to address the security requirements of a distributed agency and military global information grid (GIG) supporting centralized and distributed processing resources, multihomed delivery channels, uncertain security at the edge, edges under constant attack, siloes and stovepipes, uneven methods to interconnect high and low risk resources (MLS interconnectivity), and the realization that warfare had evolved to a point where attacks are amorphous, asymmetric, easily involving both friend and foe alike at the same time.

[1822] Of course, the concept of the 'cloud' that we know today is not the same cloud described by the mainframe and terminal architecture, the network of peer servers, clients and services, remote procedure calls, or a distributed LAN and WAN network architecture that comprises the Internet. It is possible to think of these subsets within the cloud as spaces delineated by perimeters, with the cloud comprising the superset of all possible sets and subsets. The cloud we recognize today is a ecosystem cloud comprising all possibilities, with seamless interconnection and integration, not the specialized VPN generally presumed. Although the cloud of today actually incorporates VPNs and specialized links for silos and stovepipes, any presumption of security microperimeterization is based on archaic views of trustworthiness.

[1823] The cloud today attempts to describe an uniform environment of process, storage, and utilization not defined by any where or any time, and not static in any terms familiar to us now. This definition is still evolving as information and process are mashed together in interrelated but dependent forms. This is the concept foreseen by the GIG and what DigitalDoors anticipated would provide security through dispersion as a fundamental paradigm shift. The cloud transcends perimeters and point solutions and any expectations of when and what order, so solutions that presume a defined location or definitive risk are inadequate for this cloud paradigm shift. The cloud becomes our functional universe where perimeters and microperimeters are archaic delineations lacking force of security or control.

[1824] DigitalDoors sees information as the primal resource, reflected in data, application definitions, workflows, and expressions of design blueprints that become programs, process, products, services, intangibles, and intellectual property. DigitalDoors sees most resources as a reflection of information, and the enablement of it cannot be based in place or in time. The DigitalDoors technology addresses this cloud as it is exists today, as it develops in new ways, as the distribution of information

evolves into new forms, formats, structures, composite processes, and purposes where the primary production model inputs and outputs are data and the production model is an omnipresent cloud.

AA.7.26.0 Defenses in Depth: layered security

[1825] Layered security or defense-in-depth is an historical and biological approach. Perimeter security is usually layered, in that the area surrounding a fortress is denuded of vegetation and hiding places. Concentric walls are often built in progressive heights with inside moats and battlements. Walls and moats define the final perimeter of the castle. Walls have evolved in terms of height, depth under the ground, angles, and thicknesses as siege weapons evolved. Entry into the fortress is usually tortuous so as to prevent the defenders from being suddenly overrun or defenders from facing a friendly party turned aggressive.

[1826] Biological security is typically layered. Individuals forms herds or pods to provide gross control against natural events, predators, and competitors. Some animals include aggressive and defensive fortifications, such as turtles, porcupines, snakes. Bone, nails, scales, fur or hair, provide site protection. Skin provides protection against some physical, microscopic intrusions, and leakage. The immune system deals with gross damage but also bacterial or viral attacks. Organisms replicate, evolve, and disperse to further the species, but the more important social and cultural information is taught intergenerationally.

[1827] This layered security metaphor extends to information technology. Most infrastructures have outer perimeter defenses that are physical and limit access. Firewalls and routers (plus all the evolving IDS, IPS, and other technologies) provide a perimeter defense. Biometric or password access control limits functionality and data access. Antiviral products provide some low-level protections. Distribution of assets and encryption of data complete this layered approach. Hot sites and backups

provide integrity and survivability.

AA.7.26.1 Why layered security works

[1828] This methodology works because the defense mechanisms overlap most of the attack vectors. It is simple. People understand it. It supports evolutionary changes that fulfill a view of incremental improvements. New point solutions can be created to parry new threats and attacks methods as they become known. Layered security is flexible and resilient, and its also supports polymorphic, evolving, and layered or multipronged attacks through defenses-in-depth.

AA.2.27.0 Perimeter security, the failure of the Sandbox

[1829] A sandbox is any security space compartmentalized within a larger environment. The metaphor is based on the concept of retaining the messy, abrasive, supposedly clean sand within a restricted children's play area. This is the application of perimeter security inverted so as to create a perimeter to keep valuable assets within and prevent cross-contamination rather than keep dangers out.

[1830] A sandbox is a technique of security isolation and functionality restriction within a designated internally controlled perimeter rather than an external separation beyond the perimeter. Simply put, a sandbox protects what is inside from damaging things on the outside. In contrast, a perimeter protects what is inside from being damaged by things on the outside. A perimeter also keeps the two apart.

[1831] The sandbox is the predominant security method applied for remote procedures, ill-behaving applications, cloud computing, and virtualization security. A sandbox is a perimeter protection method, keepings things separated rather than keeping things out.

[1832] A sandbox is a variation of a perimeter, just geometrically different. A sandbox is the simple difference between a belly button "iny" versus "outy." The geometric configuration is basically

turned inside-out, so both are really the same. The context differs only by whether the observer is inside or outside the shape and whether the shape is used to keep something inside or keep something out. Inside and outside are relative contexts functionally equivalent in geometry. Imagine a balloon turned inside out, doing the same functional job, but contextually inside is now outside and outside is now inside, balloon surface printing notwithstanding. Therefore, while a sandbox and perimeter might seem different, logically the sandbox is just a geometric inversion of a perimeter.

[1833] The sandbox demarcation is still there, serves the same function -- really nothing -- and is no more or no less effective than any other perimeter-based security variation. The terminology is used to impart a difference between it and other failing perimeter security methods like firewalls, guards, intrusion detection, extrusion protection, deep packet inspection, and the like. It seems cleaner and more concrete than say a perimeter because a sandbox is defined in terms of space, resources, area, placement, containment, and access points. However, the logical structure is still a defined space that requires interactive doorways into the other spaces. The security is no different from any other perimeter-based method and suffers the same definitional problems of defining what is inside from what is outside.

[1834] The apparent difference between a sandbox and the perimeter is the false simplification that the sandbox is defined as the secured space protecting everything else rather than an amorphous space that must be protected from everything else. Stretched and ambiguous perimeters do little to help establish an integrity to a perimeter or define exactly where it exists.

[1835] When a sandbox is created within a workflow or process space, this an example of microperimeterization. It is also possible to create two, or an infinite number of sandboxes, which by virtue of the quantity is somehow more attack-resistant and secure than just one. It is also possible to create a temporary sandbox. Although this apparently minimizes the time window for an attack, a

sandbox still needs the same doorways to interconnect inherently interactive communications processes. The timing of temporary sandboxes is not random since it corresponds to a process that can be predicted. Access during these termporary windows through the ever-present doorways represents the same channel for attack for the one, two, or infinite sandboxes. Timing is obscurity that an intelligent attacker will nevertheless learn to exploit. Process concurrency, virtualization, and any shared service or resource also enables security bleed-through.

[1836] Often the sandbox is 'enhanced' with additional security methods. This includes mandatory access control, roles, and expectation of the stability and integrity of the infrastructure. However, mandatory and role control presumes one-way directional and protected control over communications flow. This is not a proven methodology, and in fact piggybacking communications opens a two-way venue for malware and process reengineering. Hardening the infrastructure certainly improves its stability and integrity but does not address limitless avenues for complex breaches, polymorphic leaks, and attacks from 'trusted' parties.

[1837] Instead of creating an additional defensible citadel within the perimeter through microperimeterization, in effect the sandbox is creating a controlled internalized space with the logical functional position externalized beyond the perimeter. This is the illusion of containment and managerial control without any functional differentiation. This illusion also relocates the risk itself nearer to any potential targets rather than isolating the risk from the targets.

[1838] The additional risk includes that of an expanding sandbox blowing up within your perimeter. This creates additional adverse effects, such as buffer overruns, denial of services, and resource hogging. There really is no difference even for this internalized risk except for resource allocation issues. A sandbox is a perimeter guard rather than a perimeter firewall. The result is just so

much verbal hair splitting without any substentative difference or any substantative security improvement.

[1839] Hybrids of various point security solutions and perimeters make the rounds as new opportunities. Unfortunately, they are not perceived as variations on the same incomplete schemes. Hybrids are variations of techniques that do not work separately, rarely work better together, and generally create more venues for exploits when combined. If we repeat the definition for sandbox and perimeter, perhaps there is a hybrid sandbox/perimeter waiting in the security future; it would seem to make sense, after all creating a double wall suggests a protection from the external risk separated from the internal risks. Recall that the sandbox protects what is inside from damaging things on the outside, whereas a perimeter protects what is inside from being damaged by things on the outside. Combined, wouldn't that be best of all security techniques? This is the hybrid of the firewall and guard together.

[1840] However, the logical failure of this hybrid stems from the opening of any doorway. This undermines valved one-way communications. One-way communications is not functionable. All applications and workflow effectively requires two-way communications in order to integrate business process silos. One-way communications would be secure, but the lack of confirmation creates the proverbial 'on a wing and a prayer.' Two-way communications exposes the doorway to a piggybacked exploit compromising both inside and outside even with such a double perimeter. See the white paper In & Out for explanation why two-way contamination is inevitable. As hybrid megasuite security products incorporate additional point solutions for each new exploit and composite solutions are engineered to identify polymorphic threats, the result is an escalating arms race without any asymptotic end of the rainbow.

[1841] The 'solutions' become more complex and unwieldy, less efficient, more difficult. They

eventually undermine the core sustainable strategic advantage with adverse economic impact. This arms races revs the gridlock between functionality and security where any compromise always defers to functionality. It is a fast nickel versus the slow dime or versus even nothing at all. Functionality is measurable now but risk is the prosnostication of some future that might not even happen. Fast nickel usually wins.

[1842] A sandbox works as well as any perimeter. A sandbox suffers the same functional, economic, and practical benefits and security limitations as any perimeter. A sandbox is just a different name for the same geometric construct, albeit turned inside out. A sandbox functions without measurable effectiveness. A sandbox is a perimeter, and sandbox security is just as dead too.

[1843] The sandbox fails as a security technique.

AA.2.28.0 The critical danger of Commercial off the shelf software

[1844] Although commercial off-the-shelf (COTS) represents a cost-effective and functional alternative to specialized, certified, and formally secured technology, it also portends a coming catastrophe because it is an open set layered over another flawed security open set infrastructure. Workflow can be reengineered to conform within the limitation of COTS and augmented to achieve the existing or desired level of operative functionality. However, COTS represents a general-purpose solution this is both a subset and superset of the workflow requirements. The subset represents missing functionality and the superset represents functionality beyond the desired functionality. Although many specification and purchasing managers are likely to bulldoze through such COTS projects with allusion to all the extra functions and features beyond the core requirements that are "free," such extras come with real risks and consequential costs. The lack of subset functionality and overage of superset functionality includes critical avenues to attack security, privacy, confidentiality, survivability,

continuity, integrity and information superiority. The deficiencies in existing perimeter-based security infrastructures, the lack of an obtainable security paradigm change, and this trend toward COTS forecast both the increase in the number and the severity of security catastrophes through subset deficiencies and superset abundance.

AA.2.29.0 COTS Implementation

[1845] COTS solutions represent software, hardware, operating system, network, and integration efforts of a known quality. Costs are defined, implementation is based on historical experience, functionality has a chronicle with public track record, and many cheerleaders press for its functionality versus risk benefits. COTS implementation is desirable because final outcomes tend to overtly match estimates, proposals, and approved requisitions. Projects get completed within budget, time, and resources with a utility corresponding to the initial functional objective. Furthermore, with many COTS solutions migrating toward published interfaces or even open source standards, essential functionality that is missing can be integrated either off-the-shelf, supplemented with third-party tools, or through customization. Because information technology and security costs are rising ahead of inflation, many corporations, government agencies, and the military are migrating to COTS to rebalance spending against other objectives.

[1846] However, frequently there is a disconnect between functional objectives and security for several reasons. First, traditionally, functionality is in conflict with security. Pick functions and features or security, but not both at the same time. Second, functionality is vetted against a perimeter infrastructure that is already entirely breached by networks, the Internet, cooperation, information sharing initiatives, permitted partners, insiders, and normal workflow. All aspects of information superiority, the global information grid, data mining and search, attempts to connect-the-dots, and

legislated interagency information sharing trash any nominal or even exceptional existing perimeter-based security technology. Third, security is impossible to measure within the current infrastructure. It is generally assumed to be adequate until the repercussions of security lapses are observed and/or published; even at that point, security is layered on the perimeter as a bandaid with limited reference to security measurements of any kind. Fourth, security at the COTS level or within the infrastructure is overlooked as a functional objective with clear and concise definitions. Functions have definitions, while security is existential. Fifth, COTS often represents a lowest common denominator that does not allow for competitive advantages or assumption of security issues subordinated to functional outcomes. So while COTS is a managerially-desirable implementation, it is integrated on a flawed operational and security infrastructure.

AA.7.29.2 Unfolding Catastrophe

[1847] The primary coming failure with COTS software is underscored by the reliance on perimeter security, systemic infrastructure designs, and chains of chains-of-custody. Metadata, extensible data sets, code triggers or tables of critical data that drives the software, or frequently configurable and represent a security openness. Security is operational-centric rather than data-centric. Rather, security requires information lockdown with granular content control.

[1848] In conclusion, COTS will increase in the number and the severity of security catastrophes until security becomes a functional requirement rooted within a universal measurement system

AA.7.30.0 DigitalDoors Access and Rights

[1849] The technology secures documents and enforces privacy through extraction of critical and important data from within documents. These extracts are transferred to remote storage locations for granular security. The resulting declassified documents are reconstituted subject to presentation and

validation of the user identity.

AA.7.30.1 Authorization

[1850] The DigitalDoors proprietary authorization module is supplemented by a second channel, which utilizes best of breed security technology. The system uses the second channel for authorization of the different parties when sending and receiving E-mail. The second channel is a secured server, which can be an RSA ACE/Server using a two-factor authentication like the RSA secure ID which uses a PIN number with a number from a physical token, or NSA-approved biometric and other token-based systems. After access to the system, the user information is processed through one or many databases, which reside within the secured server or in other locations within or outside firewalls.

AA.7.30.2 Authentication Database

[1851] When a local user wants to access sensitive data, that user must provide identity information with a two-factor authentication. The user must provide a password and corresponding token. A preferred add-on system is the RSA Secure ID product integrated with the RSA ACE/Server. The local computer routes the user login data to the secured server. The server validates the two-factor authentication and enables the access to the databases to establish the user rights and access to extracted data.

AA.7.30.3 Registration/ Rights database

[1852] The Registration/Rights database defines access rights to which each user has to the extracted data. A ser with a high level clearance have access to all extracts enabling the view of the original document in full. A user with a lower clearance level will have access to fewer extracts enabling the view of a restricted but workable document with sensitive items secured

AA.7.30.4 Location Map Database(s)

[1853] Each authenticated user will transparently receive a map (or abstracted map) providing the address of the storage locations, as allowed by clearance level, containing the missing extracts. The map data might include URLs, pointers, encrypted links, or abstracted references to other databases. Access to maps, links to subordinate maps, access to remote storage, or subordinate storage locations are allowed by clearance level. Insufficient clearance levels preclude access to maps, storage locations, maps to storage locations, and operational access to or reconstitution with extracts.

AA.7.30.5 Architecture Overview

[1854] The core idea of the DigitalDoors architecture is securing critical data through a process that secures the important and critical data elements. It then disperses these important and critical data elements to different storage locations for secured storage. The architecture enables reconstitution of documents subject to validation of user identification. The declassified document can be reconstituted in different levels according to the different clearance levels of the users.

AA.7.30.6 Primary Benefits

[1855] The primary benefits are as follows:

[1856] The architecture enables timely leveled-sharing of information across an organization or organizations in a secured way. This enables many different users with varying clearance levels to view and collaborate on the same documents.

[1857] The architecture enables protection against insider attacks by creation of "digital doors" throughout the documents. Those security checkpoints enable release of critical content into a document subject to presentation of validated identifications.

[1858] The architecture enables instant searching for key words without compromising the security of the secret documents. The searching is done at the extract storage level and not at the

document level, thereby avoiding potential security compromises.

[1859] The architecture enables automatic declassification and reclassification of documents saving substantial money and labor.

AA.7.31.0 Securing the Internet

[1860] The Internet succeeded because it provided a general-purpose data information distribution channel with a high level of operational security in terms of continuity and integrity. However, because the community has grown to encompass nearly the population of the planet, the Internet fails to provide data security in terms of confidentiality and privacy. This type of data security is fundamentally absent. If encryption, firewalls, guards, intrusion prevention, antiviral solutions, and other revisional or traditional technology were sufficient to reverse the absent Internet data security, there would be no security plague.

[1861] An Internet security mechanism rooted in limited-access "extending perimeter security to-the-edge" with increased use of bulk encryption and multiple factor authentication methods will not address data pollution from information propagation, aggregation, and inference. Specifically, three cultural disparities play against the success of this security model. First, containment of data defies the premise for the very universality mechanisms that made the Internet as successful as it is now. Second, security based on the premise of access restriction does not consider the lack of trust and control in a worldwide community with competing and conflicting agenda. Third, data containment will be procedurally impossible because many other existing embedded applications and processes exist that require data transport creating new avenues for at-the-edge perimeter breaches. The data security failure must be addressed specifically on its own terms and not through the false parochial perimeterization. The risks are with the data itself not within the transport infrastructure or creative mash of processes that

increasingly add value to the economy.

[1862] The Internet is crying out for something specifically beyond traditional or revisional technology to provide this missing confidentiality and privacy. Yet market pressures have provided no effective solution. Perimeter security, bulk encryption, and controlled access fails. This does not address the decay of confidentiality and privacy because data replication, mining, aggregation and inference render control over data ineffective. Therefore, without the introduction of truly novel security paradigms—not just revisions or more of the same ideas repackaged to look novel—the Internet will not achieve the necessary confidentiality and privacy to provide the missing systemic security.

[1863] Given that existing continuity and integrity security techniques have proven effective but not extensible to address security deficiencies in confidentiality and privacy, we need new paradigms. To merely assert that "privacy is dead and learn to live with it" as stated by Scott McNealy fails to address the forthcoming disruptions to infrastructure services, loss of productivity, and material intrusions to national security and interests. This will have profound economic consequences and military ramifications, if the minor samples already experienced are any indication. Instead, realize that data security is different and requires new paradigms to restore the security to confidentiality and privacy. Such security paradigms must address the protection of the data itself in its most fundamental form rather than the Internet transport and processing infrastructure.

AA.7.32.0 Spread of viruses within meta data

[1864] The speed at which disease spreads, how quickly it manifests symptoms, and how virulent it is determines the speed of response to quarantine the disease. This is true for both biological infestation as well as computer system disease. If a disease has a long dormancy, defies treatment, hides in places where it cannot be recognized, or appears to do little or no damage, the disease is largely

ignored. A beneficial parasitic often enhances the well-being of its host and often masks a longer term nefarious objective. Just as many human diseases fall within this threshold, so to do information infections if they exhibit the long latency of information storage, the wide distribution of information, and the ease with which information is copied. In such circumstances, the theft of information often goes unrecognized because the long latency hides the discovery of the damage. Usually discovery is incidental to problems of PII exposure or competitors' exploitation of trade secrets with an unknown etiology. With longer latency, there is also the issue of chain-of-custody or the infection patterns that frequently include a carrier that is symptom-free.

[1865] The risk of computer disease that lacks obvious and immediate damage grows as positive functions and features masks a trickle of information leakage, although many of these ploys such as cookies, DoubleClick, Microsoft software registration with system configuration capture have been exposed for their deception. Increasingly, there is growing host of computer diseases, with no known primary enticement, no obvious contagion vector and no immediate potential to damage, primary signature markers that rapidly morph, and a long period of dormancy that defies recognition, such as Storm.

[1866] (Storm (a.k.a. the Peacomm Botnet) applies many of the techniques described by DigitalDoors for the selection, extraction, dispersion, and reconstruction of information to evade detection.)The continued existence and lack of Storm inoculation tools demonstrates the failure of tradition methods for virus detection, perimeter security, and protection by functional proxy and the need to syntactically analyze data streams for content, context, and concept in conformance to acceptable and expected utility.[2] Risk grows in direct correlation to the complexity of the application, platform, network,

---

2

and infrastructure. Risk grows when multiple inert parts (that is, exons) combine to create a hydra. Risk grows when combinations of innocuous information undermine security, privacy, confidentiality, and trust. For additional reference to this topic, please see the DigitalDoors white papers The Structure of Information and Cookie Crumbs, Trails, Tracks, and Footsteps, or the article, Implementing MS Office Security, published in the CISSP Journal of September 2002 by the ICS² security certification organization.

[1867] This risk latency grows as computer disease becomes completely hosted within information or metadata rather than transported in the matrix of applications, platforms, networks, or infrastructure. The risk of computer disease that lacks obvious and immediate damage grows when only information is copied without immediate functional consequences or the recognition of this leakage. This risk becomes preeminent as information becomes the driving resource in human endeavor. The information economy is and has been dependent on the ongoing processing of information. Reassignment of process (theft by diversion) or denial of service (redirection of services or shutdown of a competitor) have been the most obvious infection vectors to date. However, information itself is becoming the preferred target. Theft of raw information, misappropriation of information, and the aggregation and reuse of published information represents a new threat to continuity and welfare of the information economy.

[1868] PII, privacy, confidential, IP, process sequences and controls, the chain of command and the informal managerial information channels, the relationships between vendor and client delineate the new risk venues. The theft of such information is most often not noticed and will become more difficult to trace back to its source if even noticed. Given the lengthy incubation period between infection and recognition of damage or leaks from the infection, the chain-of-custody for any etiology has often

decayed beyond recovery. Masking techniques and dispersion of information theft to multiple independent parties can effectively disguise the source or delay any efforts to trace the breach and the exploit back to the source and perpetrators. These will be seen as leaks with no obvious sources. The resulting crisis will be demonstrated by the internal loss of trust, widespread wariness between partners, rampant apprehension between adversaries, breakdowns of political partnerships, and the erosion of primary intellectual property that enable capital, governmental, and military systems to function. This risk is not merely an academic exercise, but a historical fact with a distress signal to the future for potential information theft. Sometimes the fear of the infection and attempts to isolate its vector become self-destructive internal battles that pale in comparison to the real external risks, as exemplified by the McCarthy Red Scare and congressional hearings of the 1950s that ignored more sedate but legitimate threats from the USSR and China.

[1869] Fear actually undermined an U.S. agency that realized it had a mole after overseas field agents began disappearing. For a number of years internal agents were mistrusted, reassigned, removed from high-level assignments, or retired. This hidden disease from information theft creates panic and erodes trust between and within organizations. In fact, one famous agent cashiered from service joined another agency and used the new position to politically attack the integrity and effectiveness of his old agency. It took almost ten years to discover the real leak caused by Aldrich Ames. But in the meantime, fear, mistrust, distrust, suspicion, uncertainty, witch hunts, loss of effectiveness, and outright havoc plagued the agency, leading to various intelligence failures and breakdowns. The value of managerial organization is well-documented in literature, as is its decay and impairment too. Internal damage from the mistrust and the agency infighting might well have done more damage to reputation and outcomes than the actual Soviet mole himself.

[1870] Theft of information, damage of information, or disruption of process represent different types of theft. The theft of most things leaves an entry trail or an apparent absence, whereas the theft of information is usually silent until damage derived from the utility of the information theft is asserted and linked if at all possible to its original source. Because information can be copied with stealth and secrecy -- with no person or system realizing that theft -- the infection latency can be unlimited. Loss of information represents preeminent infection without time limits, physical boundaries, and any quarantine containment when loosed. It undermines trust in effective systems, such as perimeters and firewalls, but does not recognize the limitations of those security methods and instead ascribes the causes of the damage wrongly to those very organizational factors and systems that probably are working best of all.

[1871] As information becomes a greater component of the every human endeavor, the need for security of the information itself grows larger, more important, and harder to control. Several factors play into this exponential risk explosion. First, the nature of information storage and distribution makes the recall and deniability of information nearly impossible. Second, the veracity of information is almost impossible to confirm so that information and chaff become indistinguishable. Third, the proliferation of information sources and the quantity of information sources make it practicable to assemble similar units of information from many different sources so that control over the end information product becomes impossible. Fourth, since information is derivable from multiple sources, leak recognition and containment is complicated. Fifth, information is unitary but a complex conglomeration that can be assembled through aggregation, inference, and interaction providing a simultaneous existence for what otherwise would be considered private, confidential, trade secret, or proprietary information. Sixth, with unlimited infection latency and unconstrained recombination of information, information damage can be endless and ongoing.

[1872] Detection does not initiate sufficient treatment as there might be none of consequence. Even if the damage were noticed, the damage can be endless and ongoing, not only to the source organization but primary and subsequent parties to transactions and dependencies where information value is undermined. The stakes are very large because we lack solutions after the fact. A security breach even when repaired does address the ramifications for the loss of privacy. See the white paper, Context: Difference between Security and Privacy.

[1873] The ideal logical protection for information is control and containment before the fact. This speaks to the need to preempt the release and leakage of information. However, since information is the driving resource in this economy, its protection is critical to the continuity of the economy. The need to create trust is critical between and within cooperating organizations. Trust is the underpinning of human activities. It is not just the raw inputs to PII, privacy, confidential, IP, process sequences and controls, the chain of command and the informal managerial information channels, the relationships between vendor and client. The resulting crisis will be demonstrated by the internal loss of trust, widespread wariness between trading partners, disruption of joint military operations, rampant apprehension between adversaries, an erosion of primary intellectual property that enable capital, governmental, and military systems to function.

[1874] The only logical protection for information is demonstrated by the technologies of DigitalDoors, the preemptive measurement of informational value, the extraction and isolation of information from all data streams within a flexible and robust hierarchy of categorization, the delivery of data streams retaining operational functionality despite gaps in the information context, and the granular content control through reconstitution on an as needed basis to extend and support operational functionality with controllable and recallable information elements.

AA.7.33.0 Touchpoints and gateways as back doors

[1875] The enigmatic security touchpoint is a operational doorway piercing a perimeter to distribute PII and other security-sensitive information. Its enables transaction processing and sharing. It is a necessity to the functionality and efficiency of a process-interconnected world. A touchpoint opens a security weakness that transits across otherwise presumably secured perimeters. In effect, a touchpoint is a designation of a potential, established, or recognized barrier transit condition. In effect, a touchpoint is a rationalization of a necessary risk generally acceptable to enable functional workflow processing. Somehow, a touchpoint sounds so much better with its potential for control than a designated security weakness.

[1876] Two logical problems are evident with a touchpoint. First, aggregations of perimeter and point solutions do not secure the infrastructure with its integrated processes and cross-domain data flows. Second, where there is one touchpoint there are likely to be more touchpoints if only because most workflows are not designed but evolve other time. Touchpoints are often seen for it single dimension or purpose, but not recognized for the many other possible simultaneous communication channels, alternate protocol delivery, or its flexibility to piggyback just about any process or information flow. Furthermore, when such gateways are consolidated, the absolute volume of traffic is not altered but focused within singular points of failure and services likely to overload and leak risks because the singular point is no longer convenient and optional but now process-critical. As the overall number of touchpoints increases, the security risk increases non-linearly; A reduction in touchpoints is not likely to reverse this effect. Fewer touchpoints as a solution to leakage is a reorganization too little too late.

[1877] Touchpoints incrementally address business complexity and new business requirements but stretch the perimeter beyond its effective security zone for new unnoticed touchpoints. Additionally,

the downstream waterfall of integrated workflow and information distrubution is likely to mask many other touchpoints where the distended (that is, 'rubber-banded') perimeter is again breached in many places to promote specific PII and security-sensitive intangible property repurposed for different contexts. Stretched and ambiguous perimeters do little to help establish an integrity to a perimeter or define what it protects.

[1878] In effect, a touchpoint is a well-known and inventoried potential for a security breach. Calling a weakness in the perimeter an 'endpoint' is merely another means to hide the insecurity of touchpoint(s). Designating a weakness with other verbiage does not raise awareness and does do anything to address the specific security weakness. A touchpoint is discussed no longer as a "hole in the perimeter" but some specific process improvement and security red flag. Unless there is a process for noticing red flags and responding to them, it is empty symbolism that intends well but does poorly.

[1879] Traditional security tools do not sufficiently address point weaknesses or their evolving multiphasic risks, aggregated weaknesses collectively, or even maintain an arms-race parity with polymorphic ones. See the white paper Security is Rationalized for an assessment of traditional security weaknesses. A touchpoint is at best an admission of weakness lacking a solution but establishing a (typically short-lived when watched since nothing goes wrong) focus to watch this point of weakness. Watchfulness decays as the organization reasserts primary production goals and also as people lose interest in all but active and disruptive security breaches. It is hard to maintain accountability with conflicting goals of functionality versus security.

[1880] In recognition of this bombastic rubber-banding of the perimeter with unfettered insertion of insecure touchpoints, DHS is reducing its edge exposure from more than 5000 agency sites through the OBM Trusted Internet Connection Program (TICP). The intent to reduce the number of gateways to

less than 50 and eliminate all access through anonymous proxies is a reorganization too little too late. While a reduction in the number of perimeter gateways by 99% is commendable, the remaining 1% still represent a two-way piggybacked exposure through an existing and well-trodden barrier. The numerical size of the problem has been reduced but the traffic load is forced through increasingly overloaded and circuitous doorways, focusing the extent of the problem but in no way providing solution. It might inadvertently be the wrong solution because it creates chokepoints subject to denials of service attacks and empty guarantees for service level agreements often required for mission-critical functions that are just not viable.

[1881] Most security experts and organizations incorrectly presume that a touchpoint can be specifically protected with one or more point solutions or an aggregation of technology specifically addressing this transit. Many touchpoints are exposed as the sensors, probes, IDSs, IPSs, appliances routers, firewalls, analysis software, and other monitoring tools that are added to the infrastructure with the expectation of making it more secure. Instead, each added point becomes a touchpoint which itself adds a weakness to the overall infrastructure, or even many of new weaknesses masked by new functionality. So-called endpoint security tools are just another name for point solutions to cloak a gap in the security perimeter with unwarranted respectibility. Security appliances address perimeter or point solutions. Note that their very complexity has the additional potential to destabilize security. They are, after all, just more touchpoints.

[1882] Specific protection of information through granular content control represents a paradigm that goes with the data wherever it flows and despite the number of admissions of weakness and unknown ones. This represents a contexual content solution rather than a process workflow or perimeter barrier.

AA.8.1.0 Multi level Security Information Sharing:

AA.8.1.1 Information sharing overview

[1883] DigitalDoors technology is important to provide security against insiders and augment traditional perimeter-based methods. It represents a major step toward granular data security and leveled information sharing necessary for unified national security and democratically-acceptable defense against terrorism. Security is fully sustainable with protection for civil liberties and a more rigorous implementation of privacy. It introduces granular content security with support for taxonomies with security and privacy protection, while enabling data sharing without exposures of source, methods, and activities. Information is categorized at a content and contextual basis for both security and privacy issues. Data sharing can be safety encouraged with financial incentives.

[1884] DigitalDoors technology operates within the data to implement security, privacy, and information taxonomy. Because it is inherently platform-independent and process-independent, it is a single enterprise-wide solution for all types of platforms, network infrastructures, cross-platform distribution, extended information workflows, and traditional data processing operations. The DigitalDoors technology is fully flexible and compliant with existing and developing standards for databases, XML, and hierarchical metadata.

[1885] Given the practical implementation of data security at a granular data level and the modular, multi-platform initiative, the core taxonomic technology enables anonymous information sharing so that agencies can share information to connect the dots in a practical, sustainable, and effective process.

AA.8.2.0 Multi Level Security: new concepts for information sharing

[1886] Current technology for MLS security presupposes multiple network domains and high

equipment requirements. Encryption of deliverables precludes granular access to data required for cross-domain collaboration, white-boarding, and document or workflow routing, which are typical with modern composite reporting requirements. Guarded data downgrades, perimeter retention, and the human overhead for enabling collaboration sharing is untenable given the current work flows and technological implementation. The core problem is neither the tools, networks, infrastructure, nor the work flows but rather the inherent security risks of compound metadata documents.

[1887] DigitalDoors addresses this impasse with several novel technologies. First, all data is stripped into content components and metadata. Second, all such components are analyzed for taxonomic, sensitivity level, and content that could undermine source, method, and ongoing action security. Third, components are downgraded to a lowest generic acceptable security level (e.g. Unclassified or Sensitive But Unclassified) through the extraction of element, phrase or other differentiable unit, linear, sentence, paragraph, or sections that undermine security. Four, multiple (potentially unlimited) components are created for distribution at all required security levels (e.g. Top Secret, Secret, Confidential, TS SI TK, SBU, U, and/or TS NATO, etc.) Fifth, reconstitution of the original content components is available subject to user identity, security controls, network reach, and other ad hoc or controlled access limitations. Reconstitution is fully granular, supporting any level from partial to complete recovery of the source content.

[1888] Fundamental design, implementation, and work flows of existing collaboration tools and COTS products prevent any implementation of secure sharing and collaborative editing of sensitive documents. DigitalDoors' work with Microsoft, Macromedia, Adobe, and other like products contain inherent architectural limitations based on a unified hierarchical metadata file and in-memory control data. Trusted workstations, segregated security networks, and hosted terminal server sessions provide

compartmentalized security of data and work flow; this will not secure collaborative work flows predicated on the flow of linked composite documents, partial access to sections or material accessible at or below user sensitivity level. The inherent limitation of composite documents bound by metadata in the form of EDI, HTML, XML, DoD tagging, proprietary vendor structures, imbedded macros, user history, redlining, process control, routing, indexes, footnotes, links, references, and other higher-order designs across different versions create a nightmare of unknown triggering and activating events that undermine any EAL certification. In simple words, every composite document represents a security risk; the COTS tools do not represent the risk, they create the risk in the form of the hierarchical metadata file or in-memory control data.

[1889] DigitalDoors circumvents these limitations to achieve what could be certified as a high-assurance collaboration scenario, tagged data control, assured data pedigrees, and appropriate delivery to peer or lower security domains. DigitalDoors bypasses these inherent architectural limitations by decomposing composite documents into introns and exons that can be individually validated or extracted. This is part of core proprietary technology from DigitalDoors. The decompositions can be taxonomically encoded, downgraded, shared, and then recombined into the original composite document or other security-responsible production formats.

AA.8.3.0 Information sharing through granular content control

[1890] Security with stovepipes and silos requires granular content security and access control with reconstitution to protect sources and methods and enable information sharing. Stovepipes and silos generally enjoy a level of security through oddity, obscurity, general isolation, all of which is unraveled when these are interconnected by networking or integrated into organizational workflows. When mainframes, hosts, servers, and data warehouse silos are connected into internal networks, the Internet,

and the GIG, these resources need security too. Because the integration is both manual and automated, security must be applied to the data within the workflow.

AA.8.3.1 Gaps in Controls

[1891] Cyberspace, The Global Grid, (GIG), and Internet security is focused on data in transit and data at rest along with the aggregation, inferencing, and interaction of search results. However, more than ninety percent of information technology (IT) today involves the technical relics of stovepipes and silos. Accounting, inventory, payroll, planning, production, and logistics or transportation constitutes the bulk of IT. Faxing, messaging, and other communications represent about seven percent. Although Internet search, Web 2.0, and virtualized services provide necessary and cutting edge results, they represent a novel component. Although many security failures and privacy crises seem related to the Internet and might well be perpetrated through the Internet, most in fact relate to the insufficient security and integration of stovepipes and silos. The lost 25 million veterans records derived from a database on a host mainframe. The stolen credit card account information was accessed in silos of vast data warehouses. The multiple breaches of personal accounts from information brokers were also stored in silos, although the distribution mechanism was in all cases facilitated by Internet access and mobile or distributed computing.

[1892] The proportion of commercial, military, and government workflow hosted by stovepipes and silos is very high. Dollars applied for security within an Internet-like venue represents less than three percent of the risk. While traditional security for stovepipes and silos is quite elaborate and robust -- think data centers -- the transition to complex workflow and Internet-enabled distribution creates the bulk of the new security risk. Security must be included within standard business operations.

[1893] Consider, for example, this typical IT request. A member of military general staff is

planning for the obsolescence in 18 years and overall future lifecycle costs for the Trident missile program. This requires a complex inventory, manufacturing, maintenance, transportation, testing, and operations reporting system. This is not a cutting-edge search or Internet project. This is paper, data, and report aggregation project with host-database integration overseen by an IT manager, 8 programmers, 5 coders, 2 testers, an accountant, and two secretaries. Sites include 49 different IT locations and an unknown number of storage warehouses and deployment locations, such as missile cruisers and submarines. Security is not just a computer or Internet issue but a workflow problem.

[1894] The project requirements include estimates of available stocks, stocks that can be produced over a defined timeframe and future costs, maintenance parts inventoried, available and forecast numbers of trained personnel, logistical plans, actual field placements, future deployment plans, and hierarchy of all prime contractors and alternates. An eighteen-year plan will also entail vetting the integrity and ongoing availability of the contractors; after all, if GM owns Hughes, and Hughes owns the subsidiary that manufactures the missile guidance system, and GM sells Hughes and Hughes divests itself of the subsidiary, is the contractor still qualified? Is there a long-term contract in place and who in legal reviews those contracts yearly? Additionally, that missile guidance system is a 1960s relic, built from transistors on old-style fiberglass circuit boards with lead/tin solder. Will production facilities be subsidized and maintained? What would be miniaturization and certification costs for electronic upgrades? How will the legislative transitional requirements to use only tin solder cope with tin whisker reliability issues? There are many other questions and answers that cross over the security perimeters of many agencies. This project requires the automation and the allocation of human resources to compiled data from many different mainframes, databases, and paper files. Some paper files might be converted to images or keyed into new documents and databases.

[1895] The example IT operation is designed to interconnect and integrate data from a large number of different sources to facilitate intermediate or final reports. These reports are printed on paper or perhaps to Adobe Acrobat for real-time delivery over the Internet or GIG. These reports are security-sensitive. They contain routing information and metadata with significant data and workflow risk. Access to the many stovepipes and silos are security-sensitive. Intermediate record sets, intermediate reports, and other collateral are security-sensitive. Management reports, designs, maps, lists of data silos, programming plans, and flight arrangements to vet sites, inventory locations, and contractors are also likely to be security-sensitive. This project becomes a huge operation generating a lot of paperwork, both printed and stored. Distribution of printed formats is generally protected by traditional and well-understood security process. However, reliance on the stored data is new and represents leakage. Until all such intermediate and final reports are fed through DigitalDoors (subject to Internet style sheets, push, pull, RSS syndication, and blogging) to assess the sensitivity levels of items, sentences, paragraphs, metadata, tagging, inclusions, and references, security will not be granular and the workflow will be disrupted by the traditional lack of information sharing. Extraction of granular items and reconstitution with graduated versions, as defined by the 2004 Markle Report on Information Sharing, provides security with information sharing.

AA.8.3.2 Risks from Distributed Stovepipes and Silos

[1896] The dispersion of information, information subsets, the separation of content from context, and the otherwise partial replication of information create the apparent integrity of the information without reality. Sources and methods are often removed from versions of information and in particular removed from the views created by DigitalDoors. Although DigitalDoors views are centralized and recovery maps ultimately maintain integrity, distributions in silos are not by definition.

As a result, the reliance on partial version information could undermine operations and analysis. It is important to retain correspondence between versions and initial sources and remember that DigitalDoors views are subsets of the initial sources.

[1897] There is another issue that warrants consideration. Applicability, functionality, and utility of IT stovepipes and silos decay over time and the data within becomes harder to retrieve, integrate, maintain, or migrate. Data is not independent from the applications that process or report it. Often migrations retain the data without the physical or logical means to access it within the available infrastructure as of hardware obsolescence and retirement. However, despite such decay in these legacy relics, security is still a problem for two reasons. First, other infrastructures usually exist that can retrieve that data. Second, the disruption of the infrastructure also erodes the security procedures so that future access to the data trumps security protections and concerns. However, the intermediate data within a workflow must nonetheless be protected through access control and data granularity.

[1898] Security with stovepipes and silos requires granular content security and access control with reconstitution to protect sources and methods and enable information sharing within the compliant security ecosystem. When mainframes, hosts, servers, and data warehouse silos are connected to internal networks, the Internet, and the GIG, these resources need security too. Security must be applied to the data within the workflow.

AA.8.4.0 The Failure of Tag Reflexivity for Information Sharing

[1899] Information tagging is deficient as the sufficient and necessary mechanism for information sharing. It misses the true requirement to establish data ownership, stewardship, and control over subsequent downstream serialized and hierarchical information sharing. Serialized information sharing is based on a continuous and linear chain-of-custody between the original information source

and all subsequent uses. Hierarchical information sharing is based on information reuse, repurposing, transformation, aggregation, inferencing, and data-to-data interaction. Serialized and hierarchical information sharing are different, and this discrepancy is the fundamental explanation for information sharing denials and leakage.

[1900] Tagging is not sufficient and necessary to maintain security with serialized and hierarchical information sharing for three reasons. There is no strong method to bind the tag to the tagged information and have that security metadata persist throughout the life and downstream usage of the tagged information. Even inline in-channel tagging does not assure the integrity and resolve intention of hard-coded tags. There is a fundamental inability to institute tag reflexivity and multi-jurisdictional interagency reciprocity with a universal tag dictionary. There is no way to assert any persistent universal meaning to tagging. There is no way to resolve the lack of objective classification bias so dependent on intent and personal understanding of frameworks to tag information. Efforts to consolidate, rationalize, or federate tags are misplaced; the real issue is data stewardship and the specific control over subsequent information sharing.

AA.8.4.1 Tag Reflexivity

[1901] Reflexivity indicates a stable and persistent overlap of semiotic significance and utility. A label, markup, or tag is a linguistic designation for the sensitivity level of information with context perspective only to me; it has no perspective value to you except within the original context as it might relate to me. I care only about what information you get and what you do with it after you have it. You only care about whether the information you get has value to you and what you can do with it after you have it. Within the study of semiotics, this relates to the interpretation of meaning, whereas in other disciplines, this relates to the "precisiation" of meaning driven by objective. As a result, you and I have

no reflexivity in terms of the tag significance or utility, and in fact we assert no tag reflexivity (that is, a meaning overlap) whatsoever.

[1902] Any tag reflexivity overlap is by chance, short-lived, and incidental. It is not possible to truthfully assert any more than that. Furthermore, that tag is liable to alteration during any subsequent information sharing. A tag is a non-normalized secondary-order metadata element that purports the value of the primary information but does not assert or guarantee the value or integrity of that information. Subsequently, the tag asserts no real sharing control and no persistent universal meaning whatsoever in that information sharing workflow. It is not possible to normalize a tag so that it has a universal, static, or definitive meaning. Tags are slippery, transitional, amorphous, and culturally dependent. Even when generic tags are applied such the meaning of the tag is externally defined, that external definition is still slippery, transitional, amorphous, and culturally dependent.

[1903] This is an intractable metacontent problem.

[1904] Efforts to create information sharing within a community of interest (COI) using a predefined dictionary of tags merely create a uniform vocabulary. That has tremendous value at least by creating a common basis but does not address the actual sharing issue. Any downstream information sharing presumes that the tagged content is uniform in terms of its potential leakage risk, relevancy, and inference. However, that is not the general case. The workflow needed for information sharing will require a guard between each pairing of interests in the community to filter out the risks before distribution. These pairings are NP! and therefore computationally intensive, and therefore intractable as stated previously. The solution is rationalization through granularization, tagging with multiple dimensions at an item level so that redaction or versions can support granular content control.

[1905] As a result, the tag is a metadata element lacking absolute basis in significance and utility.

Therefore, the meaning of the sensitivity-level and information sharing tag, or any tag for that matter, depends on the eye of the beholder. Despite the desire for a universal tag dictionary, the application of a tag is limited within a single objective, perspective, purpose, or point of view. Application of a tag beyond interagency or inter-domain, in other words for the COI, lacks sufficient reflexivity.

[1906] For example, what is sensitive but unclassified [SBU] for internal operations to me could well reveal security risks, sources, and methods when released to you. I view the information as [SBU] for my internal use but would release it to you only with a top secret [TS] sensitivity. [SBU] does not map to [SBU] for you; it might map to [TS] even though you have other information that is [SBU]. [SBU] to me looks like the [SBU] you use, but the context is different both to me and to you. That [SBU] information might be unclassified [U] to me, [TS] to me, or even [SBU] to me too. The meaning of the [SBU] is sufficient and necessary within my context but is not universal to your context, or any other context for that matter. The establishment of a common and sweeping definition of [SBU] is misinformed because every agency participating directly or indirectly in information sharing will have an aggressive, competing, and potentially antagonistic context.

[1907] The failure is that [SBU] can be a syntactical certainly but not the contextual certainty. This confusion stems from the widespread use of [SBU] and all other tags. Even though these tags are the same at face value, the different assignment of these tags lack context in common. They have just a syntactical usage in common. But, the meaning is not the same. Tag usage is not reflexive between you and I. This creates a tower of linguistic Babel where the syntax and the vocabulary is identical but the meaning is reflexively diverse. Attempts to consolidate, rationalize, or federate this vocabulary are pointless because even if everyone were to understand the meaning of [SBU] -- I contend that everyone already understands the meaning of [SBU] and other tags -- consolidation fails to address the

fundamental data stewardship problem that rapidly becomes aggressive, conflicting, and potentially antagonistic.

[1908] When I send you a redacted part of my [SBU] document and retag it as [TS] specifically for you, you might cite several pieces of my information (and its source to me). Because of your data transformation of my source information, you no longer see the subsequent report as anything but [C] and only need to distribute it as [C] for both legal and operational requirements. However, that wider release could undermine my [TS] security, as well as my [TS] and [SBU] sources and methods. Even when tidbits of information like that are aggregated or inferenced, others apart from you might infer content of the redacted [TS] or even the original [SBU]. Specifically, Judge T.S. Ellis III in the 2006 AIPAC case ruling asserts that "classification may not be sufficient to render information protected." The reason is that the tag fails to protect the content or the context of the information; this failure is really the byproduct from the wrong data sharing architecture.

[1909] Tags usually reflect a singular understanding of the data, either [TS], [S], [C], [U], or [SBU]. It is a consolidation of the most viable designation or the most restrictive requirement. Unfortunately, a single designation is not sufficient to represent complications of timing, roles, degradation, the change in cultural contexts, or the multiple applicability or aggregation of different types of purposes of information in the unit of information described by the tag. A binary sharing decision is usually too restrictive in practice. Compounded or combination tags with flexible tagging overcomes a bias classification of the data. Note, however, that many processes expect a single in-channel tag and lack facilities for resolving multiple and conflicting designations. Even XML or other metadata formats confuse the issue because few processes understand how to resolve competing values. When a multitier hierarchical overlapping tag structure is supported out-of-channel, a version of the data

can be transformed specifically to match the required in-channel tag formats of each subsequent process and redacted to overcome the conflicts of functionality versus security enabling secured production or information sharing.

[1910] The issue is defined in terms of data stewardship -- my context of my data stewardship -- your context of your data stewardship. I collect the data, therefore I own the data, so that I alone determine how I use, decide how I distribute it, and ultimately judge how and what can be shared. I distribute it to you, I still want to control your use of that information, because I own it and it is mine. However, your likely attitude is that you received it, you now how downstream control of that information, that there is no rational and realistic copyright enforcement issue, and therefore that information is now yours and under your revised ownership with explicit rights for stewardship.

[1911] The issue of stewardship reduces to establishing reciprocity over security controls, an operational problem of reflexivity that fails in the same way or rights or ownership. Conflicting legal rules require agency information sharing but also reinforce the necessity for protection of state secrets. This is true for non-military, non-governmental, and non-agency workflow too in commercial and private business. It is not possible to legislate or negotiate away this conflict with a dictionary of predefined tags because of the inherent lack of universal reflexivity and the difference in opinion over the data stewardship and ownership issue. I cannot make the decision for you, and you cannot make the decision for me. I cannot say or even conceive of how you will reuse my information and when that information will become stale or obtain an even higher risk to me. It is a very complex issue that lacks a contextual and controllable temporal basis.

[1912] Consider this scenario. It is my data. I do not want you to have it for a number of reasons that include its potential to undermine internal security or jeopardize my sources and methods. Clearly,

you want the data so that you do not have to duplicate the effort to collect the same data and you want to fill the gaps in your collection efforts. In order for me to redact the data for information sharing, I have to reassess the [SBU] label and the content in order to reveal what I choose to you. I also need to understand how you are going to use it for your own requirements and the potentials uses that are not in my interests. I also have to guess what potentials exist for aggregation, inference, and interaction with other data sources both known and unknown to me, and future data sources that do not know but will compromise my security interests. I also need to assess your access to private and public sources of data, prior relevant data releases, and contextual repurposing that might undermine my contextual security requirements. The data stewardship issue is that it is my contextual security requirements, but yours only indirectly.

[1913] I can tag that redaction with any tag I want, and you can reassign any new tag you choose since your think you now "own" it. The traditional information sharing environment provides no controls over your ability to effectively own it and reuse the information as your choose and assert what is effectively metadata to that information. The ability to append metadata and have flexibility to rework or repurpose information is powerful attribute of the information revolution. You do not have the same information in the same manner that I did. My new tag provides only value to me, but none to you; likewise your new tag provides only value to you and is frequently pointless to me. Nothing says these changes need to be collaborative. However, my data stewardship might wish to retain control over your subsequent serial or hierarchical redistribution of my redaction. Subsequent redaction and assigned of new tags is still pointless to me. Realistically, such control is lost with tagging controls unless control is retained over granular content.

[1914] Efforts to consolidate, rationalize, or federate tags are misplaced; the real issue is the data

stewardship and the specific control over subsequent information sharing. I want to control the subsequent downstream redistribution of my granular information within whatever sensitivity context that I designate for subsequent usage. I might not assert [TS] but rather impose restrictions that the information is unavailable under all conditions. Solution is through the DigitalDoors architecture of data stewardship and digital rights management where information is dispersed through granular content control.

[1915] In conclusion information tagging is not the sufficient and necessary mechanism for information sharing. It misses the true requirement to establish data stewardship and control over subsequent downstream serialized and hierarchical information sharing. There is no strong method to bind the tag to the tagged information. There is fundamental lack of tag reflexivity between agencies and securable method for reciprocity. There is no way to assert any persistent universal meaning to that tag.

[1916] However, the DigitalDoors paradigm and granular content control makes it possible to assign multiple hierarchical contextual tags -- and then map granular contexts to granular content items ,which enables security for leak-proof and inference-limited downstream information sharing.

AA.8.5.0 Multi Level Security needs to adopt granular data control to achieve its objective

[1917] MLS or Multi Level Security has been used to try to share information between different levels of security within defined location and defined application spaces. Day-to-day tradeoffs included a mentality of security versus functionality. However, as these spaces multiplied and applications became more complex, the solution implementation became too expensive, too unwieldy, too complex, and frankly non-operational. People worked around it. This is prove of a breakdown.

[1918] The existing technologies, classified, private, or commercial, and available or foreseeable variations on them, simply cannot deliver secured information sharing between different security levels.

[1919] The solution space of location and application no longer fits the operating environment for security. There is an evolution in the ecosystem that must be reflected by a change in how we view the environment itself. If you cannot admit that there are widespread security problems that are not addressed by the currently available solutions, you have a larger security problem than you can even imagine. If you cannot admit that the scope of the current security problems are escalating and the frequency and complexity of the security problems are themselves escalating, you are ignoring the inherent environmental danger and forthcoming extinction of those that will fail to change and adapt. You will fight a pitched battle protection your location and application, while leaking information and preventing the necessary distribution needed to even win that battle. It is a fundamental strategic failure that adversaries continue to exploit. If you cannot admit to this environmental change, you are attacking an archaic security problem with obsolete tools. You might delay the inevitable extinction for some period of time, but it is still inevitable.

[1920] The security problem is no longer about location and application but rather information and distribution. Actually the security ecosystem must also include knowledge discovery, integrity, security, and sharing, not some single-dimensional notion of security. If you can accept some fundamental ecosystem shift is possible, conceivable, or literally upon us now, you have take the first step to perceiving a new paradigm to addressing the security ecosystem. This new paradigm repositions intractable security problems with very simplified views of information defined by a semantic structure supporting configurable distributions. This security is asserted through granular content control. Location and application remains intact and is really no longer part of the security problem. Even when location and application becomes more complicated as need be through such issues as SaaS, SOA, Web 2.0, distributed processing, and dynamic with continuous cyclical support operations, security is applied

specifically to information and distribution. This bypasses the complications and complexities inherent in the obsolete tools. That is the fundamental archetypal shift and operative means to reassert security within this profound evolution. When the battle is fought specifically over information and distribution, old friends like MLS can be redeployed based on implementation of granular content control. It is proven and extensible into the new security ecosystem. However, you regain control over security by shifting the security problem ecosystem from location and application to the ecosystem of information and distribution.

AA.8.6.0 Accommodating the Bell LaPadula model

[1921] This document describes how to configure Bell-LaPadula model (BLM) *-property compliance and options. Information sharing controls and BLM compliance are "baked into" DigitalDoors. The BLM and information sharing features emphasize adaptive risk and privilege management in an environment where there is the "need to share world". In simple words, DigitalDoors compartmentalizes the risk taken by one so that it is not assumed by all. DigitalDoors includes major three options for information security and compartmentalization in conjunction with collaborative information sharing. Information sharing is controlled on a highly granular basis not only for traditional documents, database queries, or data streams but also for metadata, directories, search results, and catalogs. These options balance strict BLM compliance, strict compliance to information sharing laws, or risk and privilege management with smart push/pull while facilitating a black core based on a platform- and process-independent operational stack.

[1922] DigitalDoors enables redaction of documents and the delivery of versions for each sensitivity level to different sites or guards at these different sensitivity levels to fulfill operational needs and enable information sharing. Delivery can either (Option 1) adhere strictly to the Bell-LaPadula *-

property model, (Option 2) enable complete information sharing as obligated by the 2004 Information Sharing Act, or (Option 3) balance the BLM model with specifically-defined granular information collaboration pass-through. The object of functional range is provide tools to balance strict information compartmentalization and with information distribution and sharing. Specifically, these options avoid the usual overclassification of intelligence sources and to increase the distribution of functional information in order to connect the dots while rigorously protecting sources and methods.

[1923] The Bell-LaPadula model defines a formal state transition model of computer security policy based on a set of access control rules. In this model, the entities in a computer system are divided into abstract sets of subjects and objects. The notion of a secure state is defined and it is proven that each state transition preserves security by moving from secure state to secure state; thus, inductively proving that the system is secure. A system state is defined to be "secure" if and only if the permitted access modes of subjects to objects are in accordance with a specific security policy. The clearance of a subject is compared to the classification of the object and a determination is made as to whether the subject is authorized for the specific access mode in order to determine whether or not a specific access mode is allowed. For the purposes of this documentation, hierarchical sensitivity levels are described as the generic [U], [C], [S], and [TS] with a nod to TSOL [Admin High] and [Admin Low]. Sensitivity levels, however, are easily configured for other specialized entries like [SBU], [FOUO], [TS/NSA], [Rel/NATO], or [StoneGhost] for any other labeling requirement. Compartments (multiple assignment and non-hierarchical) are not described herein but correspond to TSOL compartment implementations and labeling.

[1924] Strict BLM compliance (and the inverse downward compliance) is enabled only by the administrator through the checkbox on the configuration tab of the Administration Configuration

application. This is indicated by the next screen image.

[1925] When this checkbox is checked without any subsidiary options, strict BLM compliance is enabled (Option 1). Notice that the users are listed for inclusion in this strict enforcement. All options for information sharing are disabled (shown by the boxes in red); while BLM conformance is always enforced for all users, the subsidiary options for information sharing between sensitivity levels are not enabled.

[1926] Cross-domain distribution on-the-fly is enabled because multitier tagging is reassessed prior to upgrading, downgrading, or multiple sensitivity level distributions. The clearance/classification scheme is expressed in terms of a lattice with emphasis that information with a higher sensitivity level is not released to a lower sensitivity level. However, in addition to the normal implementation, DigitalDoors has expanded the BLM functionality to also restrict inverse information flow—from a lower sensitivity level to a higher sensitivity level—so as to preclude upward insertion of virus, spyware, malware, attachments, applets, and data dilution and the creep toward overclassification. Note that this upward containment is configurable separately from the classic BLM implementation for backward model compatibility.

[1927] When the checkbox is not checked, full compliance to the 2004 Information Sharing Act is enabled (Option 2); compliance with BLM is not enabled. Information sharing within strict BLM star security compliance or with extended DigitalDoors overrides is also not enabled. Notice that the users are not shown because the options for granular information sharing are already broadly enabled.

[1928] A balance between BLM compliance and information sharing is configurable to control the level of information passed through the star security model, from each user's defined sensitivity level to any and all sensitivity levels that are up or down from that level. Option 3 is enabled when the

checkbox is checked and subsidiary options are also engaged. This section explains how.

[1929] These subsidiary information-sharing options must be enabled step-by-step. Each user has three selections. (1) To enable information sharing at that user's sensitivity level (only). (2) To enable sharing down to a lower sensitivity level (as per the classic BLM implementation). (3) To enable sharing up to a higher sensitivity level; this is additionally implemented by DigitalDoors to minimize the necessary intranet/network complications and the risk taken by one so that it is not assumed by all. Note that these settings are only accessible to the administrator for security reasons.

Information sharing with BLM compliance requires that DigitalDoors reconstitution, versions, dispersion, and replication output vary substantiality. Granular selection of content is paired with granular distribution to users, and also the granular packaging for wide scale networking. The limitations of selection and extraction by compartmentalization and the delivery of base sanitized documents with the recovery map and recovery data must reflect the sensitivity level at the source and also at each target level. In simple words, the output from the DigitalDoors process for a [S] document processed delivered to [TS] level must be self-contained and different from the results delivered to [C] or [U] level. Each such self-contained unit would—as that unit—comply with BLM. Separate units must be delivered to each sensitivity level without overlap to prevent security leakage and support the need for wide scale networking with propagating the risk assumed by one to the entire networked environment.

[1930] BLM compliance with optional overrides for information sharing. Sharing is enabled across all comparable sensitivity levels subject to the network infrastructure and arrangements for shared data access. In simple words, if the network is in place with MAC and DAC access controls, sharing between [U] and [U] will available or between [TS] and [TS]. Sharing up from [U] to [TS] or down from [TS] to [U] also is subject to the network infrastructure and arrangements for shared data access. Sharing

with BLM compliance still requires a manual two-man guard, but DigitalDoors automates the creation of document versions applicable for up or down sharing without leakage of sensitive information. In other words, DigitalDoors automates the production and population of the networks with documents sanitized at each or all sensitivity levels so users can connect the dots; guarded access is not affected and remains the under the existing organizations and their policies. DigitalDoors does not violate BLM, information compartmentalization, need-to-know rules; it automatically produces the granularly sanitized documents in order to populate the GIG and all the agency and parallel networks that exist at each sensitivity level classification.

[1931] DigitalDoors enables granular distribution based on sensitivity level, compartments, category, and other supported taxonomies and ontologies. None of these features are lost with BLM compliance. However, strict BLM compliance precludes information sharing. Therefore, optionally, granular leakage can be controlled on a need-to-know basis in order to expedite interagency and information sharing operations so that the dots can be connected. Specifically, the subsidiary settings enable distribution of content at the user's sensitivity level for sharing upward. Specifically, the subsidiary settings enable distribution of content at either/or/both the user's or the recipient's sensitivity level for sharing upward. Specifically, the subsidiary settings enable distribution of content at the user's sensitivity level for sharing downward. Specifically, the subsidiary settings enable distribution of content at either/or/both the user's or the recipient's sensitivity level for sharing downward.

[1932] DigitalDoors can create a distribution targeted for a single sensitivity level, such as [S] from a [TS] base source. The DigitalDoors dispersion delivers the data streams either to a single structure on a local data storage device, or disperses the major pieces to remote storage devices. Furthermore, these data streams can be further granularized to disrupt any context and protect against

insiders and sophisticated data aggregation and inferencing. One source sensitivity level to another sensitivity level is the simplest case. Versions represent this same process repeated once for each enabled sensitivity level. However, the selection and extraction process is optimized with a DigitalDoors metadata collection and generated once and only once; all tagging, taxonomy, and ontology are performed upfront. This depth of selection and extraction process is configurable for more or less aggressive categorization and performance improvements.

AA.8.6.1.0 Maps and Recovery Information

[1933] Another aspect of BLM that affects the DigitalDoors implementation is the format for the maps and recovery files. While this is addressed elsewhere in the document, it is also pertinent to recall that some formats are specifically non-compliant with BLM and others are implemented specifically to enabled information sharing capabilities with full BLM compliance. Maps and recovery information can be produced in four categories of formats. These are:

Table: Map categories

Consolidated

Inclusive

Exclusive

Granular itemization

[1934] The consolidated format is an aggregated rollup of the maps and all extraction information. While it is usually encrypted, this format is non-compliant for BLM. Inclusive, exclusive, and granular itemization are BLM-compliant. Each format serves a different function. Inclusive provides a stack or rolled up map and recovery information for each sensitivity level. The exclusive format strictly supports need-to-know and provides map and recovery information specific to a defined sensitivity level.

The granular itemization format disperses the maps and each extraction into separate contains so that reconstitution can be controlled at any combination or selection of sensitivity level, compartment, roles, missions, users, groups, and other categorical hierarchies. In addition, this granular lockout enables controlled indexing, searching, and mining opportunities without releasing the details of sensitive information

[1935] Although the selection of maps and recovery information format is codependent on BLM compliance, the selection of the format is independent from the implementation of BLM compliance because these two functions have separate utilities. Neither compliance sets formats nor do formats set compliance, but some lockout "cams" ensure security, privacy, and full functionality with flexible administrator configuration.

AA.8.7.0 Challenges of Multi level security information sharing

[1936] Current technology for Multi Level security information sharing - MLS security presupposes multiple network domains and high equipment requirements. Encryption of deliverables precludes granular access to data required for cross-domain collaboration, white-boarding, and document or workflow routing, which are typical with modern composite reporting requirements. Guarded data downgrades, perimeter retention, and the human overhead for enabling collaboration sharing is untenable given the current work flows and technological implementation. The core problem is neither the tools, networks, infrastructure, nor the work flows but rather the inherent security risks of compound metadata documents.

[1937] DigitalDoors addresses this impasse with several novel technologies. First, all data is stripped into content components and metadata. Second, all such components are analyzed for taxonomic, sensitivity level, and content that could undermine source, method, and ongoing action

security. Third, components are downgraded to a lowest generic acceptable security level (e.g. Unclassified or Sensitive But Unclassified) through the extraction of element, phrase or other differentiable unit, linear, sentence, paragraph, or sections that undermine security. Four, multiple (potentially unlimited) components are created for distribution at all required security levels (e.g. Top Secret, Secret, Confidential, TS SI TK, SBU, U, and/or TS NATO, etc.) Fifth, reconstitution of the original content components is available subject to user identity, security controls, network reach, and other ad hoc or controlled access limitations. Reconstitution is fully granular, supporting any level from partial to complete recovery of the source content.

[1938] Fundamental design, implementation, and work flows of existing collaboration tools and COTS products prevent any implementation of secure sharing and collaborative editing of sensitive documents. DigitalDoors' work with Microsoft, Macromedia, Adobe, and other like products contain inherent architectural limitations based on a unified hierarchical metadata file and in-memory control data. Trusted workstations, segregated security networks, and hosted terminal server sessions provide compartmentalized security of data and work flow; this will not secure collaborative work flows predicated on the flow of linked composite documents, partial access to sections or material accessible at or below user sensitivity level. The inherent limitation of composite documents bound by metadata in the form of EDI, HTML, XML, DoD tagging, proprietary vendor structures, imbedded macros, user history, redlining, process control, routing, indexes, footnotes, links, references, and other higher-order designs across different versions create a nightmare of unknown triggering and activating events that undermine any EAL certification. In simple words, every composite document represents a security risk; the COTS tools do not represent the risk, they create the risk in the form of the hierarchical metadata file or in-memory control data.

[1939] DigitalDoors circumvents these limitations to achieve what could be certified as a high-assurance collaboration scenario, tagged data control, assured data pedigrees, and appropriate delivery to peer or lower security domains. DigitalDoors bypasses these inherent architectural limitations by decomposing composite documents into introns and exons that can be individually validated or extracted. This is part of core proprietary technology from DigitalDoors. The decompositions can be taxonomically encoded, downgraded, shared, and then recombined into the original composite document or other security-responsible production formats.

AA.8.8.0 Implementing Cross-Domain MLS Information Flow

[1940] Cross-domain information flow is undermined for three reasons. Primarily, organizations are culturally unable and pathologically unwilling to share information. Information represents a point of power and control, security risk, and ongoing legal strife. Agency or agent will not share with any other; any act of sharing gives up power and reward, opens the door to the disastrous exposure of sources and methods, while national security laws preclude sharing with terrorist-combating-related laws mandating data sharing. Sharing data, when data is mostly noise, is not the same as sharing signal or sharing content, which are relevant subsets. In other words, data sharing could be mandated and actually implemented but the noise factor undermines functional result. Besides, human nature being what it is, real sharing requires a tit-for-tat accounting of chips at the very personal level, like "I gave you something useful, and you owe me something comparable in return." Both parties to an information transfer must agree to the value, extent of the information, and consent to bilateral transfers. Ultimately, morsels of information are owned by someone, and that someone will extract the most value and power from them.

[1941] Secondarily, existing computer and networking technology overvalues the workflow

rather than information. The existing paradigm where workflows are fed with data must advance into an inverted recognition where information is the primary asset and is transformed by the workflows. This disconnect can be defined as the inability of work flows to technically define the content and value for its potential as information flows. Considerable effort is applied to the process of information transformation because we easily comprehend the order of events rather than the complex characteristics, comprehension, metadata assessment, and recognition of value within the information.

[1942] Tertiarily, multi-level security (MLS) and multiple independent levels of security (MILS) presuppose a compartmentalized processing network with a complete secured channel for each domain. Unfortunately, a disconnect in implementing protected but interconnected MLS and MILS domains have undermined this technology. Cross-domain guards have focused on partitioning process and data rather than on content, context, and concept—the fundamentals that enable granular content management. Therefore, cross-domain information flow will not transpire until we can overcome these cultural and technical barriers, align these to stated and unstated organizational missions, and implement granular content management environments. In summary, any cross-domain information sharing will remain ineffective and be actively sabotaged unless, the following conditions are met:

1)content value can be measured both objectively

2)content release can be limited and controlled granularly in a cross-domain matrix

3)realistic rewards can be assessed and awarded for successful information sharing

[1943] Content is a complex evaluation not traditionally supported except for structured data such as transactions, databases, and other records defined in terms of purpose. This excludes non-structured data such as documents, XML, style sheet transformations, containers, zones, collections, messages, and most type of metadata. Content is retained by a stream, message, or file packaging. Such

containers, XML tagging, and formats modify the content in context and concept in workflow but do effect this content in substance itself. XML, metadata tagging, and other taxonomies modify and become part of the information sharing environment as an infinite cycle. Effectively, all hierarchical data must necessarily have both internal and external metadata designations about utility, purpose, functional transformation (application code), inference, and aggregation. This is the point for objective measurement and consensual determination for content release. Here content release becomes multidimensional, both in terms of the strict content, the inferred context, and functional concepts. Because content is measured at creation (up front) rather than ad hoc or at a guarded translation, value is measured inherent to the content allowing real-time decision-making. The release rules must consider the source side and the target side security levels and the ramifications of the content with the metadata and subsequent measurements, which represent a new layer of metadata. Every sensitive network and each mission represents additional N-dimensional risks. As a result, simplistic matrix of source rows and target columns under-represent the other N-dimensional metadata risks.

[1944] DigitalDoors aligns with this cultural and technical barriers by providing:

1)means to and objectively measure content in terms of content, context, and concept

2)method to N-dimensionally tag content and create predefined or ad-hoc releases to each of source side, target side, and metadata dimensions for controlled information sharing

3)accounting system to measure information value, accumulate inter-aggregate transfers, and report net surplus or deficits for information sharing

[1945] The DigitalDoors differentiation creates a core for granular information management for security, privacy, and connecting-the-dots independent of platform, environment, and networking. Any translation of the content retains the original content to continue the integrity of any applied value

measure, a limitation of structured data, semi-structured data, even XML, MS Word documents, and other modern hierarchical data formats. Although DigitalDoors cannot create the actual organizational, managerial, or contributive incentives and rewards for effective information sharing, the measurement, method to release, and the accounting system represents a direct input to such a mandated legislation.

AA.8.9.0 Multi level Security for the Internet

[1946] The Internet succeeds because it provides a general-purpose data information distribution channel with a high level of operational security in terms of continuity and integrity. However, because the community has grown to encompass nearly the population of the planet, the Internet fails to provide data security in terms of confidentiality and privacy. This type of data security is fundamentally absent. If encryption, firewalls, guards, intrusion prevention, antiviral solutions, and other revisionist or traditional technology were sufficient to maintain Internet data security, there would be no security plague. Yet the obvious loss of confidentiality and privacy shows the growing fundamental lapses in Internet security and other interconnected and distributed network infrastructures like Constellation and the GIG.

[1947] An Internet security mechanism rooted in limited-access "extending perimeter security to-the-edge" with increased use of bulk encryption and multiple factor authentication methods does not address data pollution from information propagation, explosive availability of answers through mining and search, and the threats from aggregation, inference, and data-to-data interaction. Specifically, three cultural disparities play against the success of a perimeter security access model. First, containment of data defies the premise for the very universality mechanisms that made the Internet as successful as it is now. Second, security based on the premise of access restriction does not consider the lack of trust and control in a worldwide community with competing and conflicting agenda. Third, data containment

will be procedurally impossible because many other existing embedded applications and processes require data transport and create new avenues for at-the-edge perimeter breaches. The data security failure must be addressed specifically on its own terms and not through the false parochial perimeterization. The risk derives from propagation of the data itself and not from the transport infrastructure or creative mash of processes that increasingly add value to the economy.

[1948] The Internet is crying out for something specifically beyond traditional or revisionist technology to provide this missing confidentiality and privacy. Market pressures have provided no effective solutions, while group think has encouraged many vendors and users to acquire modernized versions of nonetheless obsolete technology. Perimeter security, bulk encryption, and controlled access fail. Perimeterization does not address the decay of confidentiality and privacy because data replication, mining, aggregation, inference, data-to-data interaction functionally render control over data ineffective. Therefore, without the introduction of truly novel security paradigms -- not just illusionary revisions or more of the same ideas repackaged to look novel -- the Internet and the GIG will not achieve the necessary confidentiality and privacy to provide the missing systemic security.

[1949] Given that existing continuity and integrity security techniques have proven effective but not extensible to address security deficiencies in confidentiality and privacy, we need new paradigms. To merely assert that "privacy is dead and learn to live with it" as stated by Scott McNealy, fails to address the forthcoming disruptions to infrastructure services, loss of productivity, and material intrusions to national security interests. This will have profound economic consequences and military ramifications, if the minor samples already experienced are any indication. Instead, realize that data security is different and requires new paradigms to restore the security to confidentiality and privacy. Such security paradigms must address the protection of the data itself in its most fundamental form

rather than the Internet transport and processing infrastructure.

[1950] Recognize the discrepancy between governance and compliance with actual security. Governance and compliance are band-aids not cures. Conformance with rules and regulations is bound to be a responsible but prolonged reaction to risk. Although such conformance represents a baseline approach to security and risk management, the baseline defines only a starting point predicated by a trivial measurement system. Most organizations will be required to exceed baselines to maintain a legal, moral, or operational integrity, survivability, and continuity. There will always be the responsibility for compliance and, more importantly, proof of that compliance, but most organizations will ultimately be held to a higher standard that "you should have known that the state of the art has migrated to ...." In such a situation, think security cure, not just band-aid.

AA .8.9.1 Cultural and Technical Constraints

[1951] For over forty years, security was predicated on the Multi-Level Security (MLS) model based on information compartmentalization. Although this model withstands the test of time, validation for integrity, and thusly remains the foundation for security modeling, unfortunately the cultural and technical constraints have undermined implementations. The MLS model is not invalid. The MLS implementation is not defective. The applicability was just too restrictive and it's underlying data model flawed. Cross-platform complexities, the lack of a simple model for semiotic meaning, and the lack of consistent executions have undermined effective, efficient, and widespread adoption. However, MLS has never been invalidated; it just never found its critical mass or its comprehensive applicability. MLS compartmentalization had been tied to security at the perimeter, and now the concept of defining and maintaining a perimeter is invalidated. Multiple Independent Levels of Security (MILS) is an illusionary attempt to implement MLS by creating parallel communication infrastructures with guarded interfaces;

however, MILS does not address the fundamental security issues of the information structure which need to be addressed instead. As a result, MLS is being reinvented as a mainstream security technology as the Trusted Computing Module (TCM). TCM is consistent with trusted computing environments, like Sun Microsystems Trusted Solaris. TCM is neither new, nor different, just relabeled as a retailored marketing strategy for Unix, Linux, and WinTel platforms. MLS is alive and well as TCM.

[1952] Another transition undermines MLS and data security. Specifically, the migration of data formats from field-oriented, record-oriented, or database storage into so-called unstructured formats typically prevents clear and concise assignment of sensitivity levels to the data. Additionally, when the data is consolidated with mixed data types, wrapped with metadata, and infused with macros and application code, risk assessment errs heavily on the side of caution. As a result, data is typically over-classified and fundamentally restricted for reuse and distribution. The explosion of unstructured data and complex streams from signal sources will only increase the pressure to bypass security perimeterization.

[1953] The current fears over escaped data in the form of CDs that walk, USB finger storage devices or iPODs with the downloaded databases, or stolen laptops are not new problems. They are recurrences of past events or variations on old themes. However, it is now clear that we correctly value data over the value of the media or the pilfered computers. There are several current legislative attempts to limit data collection, alter data ownership, and force new restrictive retention rules. While the lofty goal is to minimize ID theft and newsworthy operational security lapses, such efforts will backfire and undermine data sharing initiatives by reverting to the cultural and technical limitations undermining traditional MLS. The concept should not be security versus information sharing, but rather both, in harmony. It need not be either/or when it can be both. Information is not some monolithic entity but rather an assemblage of granular elements with unlimited potential for meaning in a "disaggregated"

format. Even unstructured data, compound signals, and complex metadata can be deconstructed into elements that can be tagged at appropriate sensitivity levels. This subtle but simple new model for information creates the paradigm shift that enables a different view of security compatible with all existing operational requirements for security and information sharing on current and foreseeable computing infrastructures and command, control, and communication platforms.

[1954] The era of glass-walled stovepipe computing has evolved into the Internet model for information distribution and virtualization. Security too has to reflect that evolution too. So, as computing becomes distributed and data is collected, aggregated, replicated, and dispersed, security must migrate with the data. It is not that the concepts for security based on compartmentalization and perimeters are flawed, it is simply that the demarcation is blurred and the ability to lock down the composite data entity is undermined. This risk becomes increasingly large as the data entity grows in complexity with tags and labels, hierarchical structures, metadata, and the imbedded engines of knowledge management and processing. This risk is also analogous to the priority given unstructured data as structured formats become more hierarchical, tagged, and marked with metadata. Therefore, the perimeter can no longer be asserted as a containment around the entire data entity since that precludes operational utility and dispersion to connect the dots. Rather, we want to perimeterize the dots and disaggregate the content, context, and concept.

AA.8.10.0 A Paradigm Shift

[1955] The technology in DigitalDoors makes MLS and TCM possible for widespread cross-platform adoption. It overcomes the technical and cultural hurdles of the past by viewing information as a stream that can be composed from data, metadata, hierarchical relationships, and compound data flows at any arbitrary complexity. The operative principle is that information can be separated into

elements of content, context, and concept. These elements can be tagged with any level of sensitivity and compartmentalization or unlimited multiple competing levels of tag (label) assignment, and removed from the data stream to defeat aggregation, inference, and interaction. Therefore, DigitalDoors implements a granular MLS consistent with the new TCM hardware model that circumscribes a perimeter on granular elements with access controlled by sensitivity level, compartmentalization, and other assignments. Multiple tags for the same element drive computing with sensitivity-level-awareness. This technology piggybacks on MLS and TCM and can function as sophisticated filters as part of inbound and outbound guards. Sharing of the parts as parts or assembled in subsets as allowed lets users connect the dots and yet constrain information release to preclude leakage or process failures consistent with the older generation of MLS technology.

AA.8.11.0 Implementing Cross-Domain Information Flow - Multi level security

[1956] Cross-domain information flow is undermined for three reasons. Primarily, organizations are culturally unable and pathologically unwilling to share information. Information represents a point of power and control, security risk, and ongoing legal strife. Agency or agent will not share with any other; any act of sharing gives up power and reward, opens the door to the disastrous exposure of sources and methods, while national security laws preclude sharing with terrorist-combating-related laws mandating data sharing. Sharing data, when data is mostly noise, is not the same as sharing signal or sharing content, which are relevant subsets. In other words, data sharing could be mandated and actually implemented but the noise factor undermines functional result. Besides, human nature being what it is, real sharing requires a tit-for-tat accounting of chips at the very personal level, like "I gave you something useful, and you owe me something comparable in return." Both parties to an information transfer must agree to the value, extent of the information, and consent to bilateral transfers. Ultimately,

morsels of information are owned by someone, and that someone will extract the most value and power from them.

[1957] Secondarily, existing computer and networking technology overvalues the workflow rather than information. The existing paradigm where workflows are fed with data must advance into an inverted recognition where information is the primary asset and is transformed by the workflows. This disconnect can be defined as the inability of work flows to technically define the content and value for its potential as information flows. Considerable effort is applied to the process of information transformation because we easily comprehend the order of events rather than the complex characteristics, comprehension, metadata assessment, and recognition of value within the information.

[1958] Tertiarily, multi-level security (MLS) and multiple independent levels of security (MILS) presuppose a compartmentalized processing network with a complete secured channel for each domain. Unfortunately, a disconnect in implementing protected but interconnected MLS and MILS domains have undermined this technology. Cross-domain guards have focused on partitioning process and data rather than on content, context, and concept—the fundamentals that enable granular content management. Therefore, cross-domain information flow will not transpire until we can overcome these cultural and technical barriers, align these to stated and unstated organizational missions, and implement granular content management environments. In summary, any cross-domain information sharing will remain ineffective and be actively sabotaged unless, the following conditions are met:

1)content value can be measured both objectively

2)content release can be limited and controlled granularly in a cross-domain matrix

3)realistic rewards can be assessed and awarded for successful information sharing

[1959] Content is a complex evaluation not traditionally supported except for structured data

such as transactions, databases, and other records defined in terms of purpose. This excludes non-structured data such as documents, XML, style sheet transformations, containers, zones, collections, messages, and most type of metadata. Content is retained by a stream, message, or file packaging. Such containers, XML tagging, and formats modify the content in context and concept in workflow but do effect this content in substance itself. XML, metadata tagging, and other taxonomies modify and become part of the information sharing environment as an infinite cycle. Effectively, all hierarchical data must necessarily have both internal and external metadata designations about utility, purpose, functional transformation (application code), inference, and aggregation. This is the point for objective measurement and consensual determination for content release. Here content release becomes multidimensional, both in terms of the strict content, the inferred context, and functional concepts. Because content is measured at creation (up front) rather than ad hoc or at a guarded translation, value is measured inherent to the content allowing real-time decision-making. The release rules must consider the source side and the target side security levels and the ramifications of the content with the metadata and subsequent measurements, which represent a new layer of metadata. Every sensitive network and each mission represents additional N-dimensional risks. As a result, simplistic matrix of source rows and target columns under-represent the other N-dimensional metadata risks.

[1960] DigitalDoors aligns with this cultural and technical barriers by providing:

1)means to and objectively measure content in terms of content, context, and concept

2)method to N-dimensionally tag content and create predefined or ad-hoc releases to each of source side, target side, and metadata dimensions for controlled information sharing

3)accounting system to measure information value, accumulate inter-aggregate transfers, and report net surplus or deficits for information sharing

[1961] The DigitalDoors differentiation creates a core for granular information management for security, privacy, and connecting-the-dots independent of platform, environment, and networking. Any translation of the content retains the original content to continue the integrity of any applied value measure, a limitation of structured data, semi-structured data, even XML, MS Word documents, and other modern hierarchical data formats. Although DigitalDoors cannot create the actual organizational, managerial, or contributive incentives and rewards for effective information sharing, the measurement, method to release, and the accounting system represents a direct input to such a mandated legislation.

AA.8.12.0 Negotiated multi level information sharing

[1962] DigitalDoors introduces a method and system for maximizing the sharing of information intelligence among different parties. The system delivers a process for securing and secured sharing of information through extraction of critical data from a document and a mechanism for reconstitution. The reconstitution is done through a controlled-release process affecting partial reconstitution from none to full based on sophisticated access controls. The process also enables reconstitution of extracted data for a "consideration" or if needed a "negotiated consideration". The consideration may be a monitory consideration, a service or release of requested information.

[1963] Sharing a document whose critical data was extracted i.e. a declassified document creates an interest and an appetite for seeing the missing critical extracts. This creates the willingness to give something in consideration in order to be able to access the missing critical elements of the document. DigitalDoors technology, in essence, enables various sharing transactions. For example, it enables an e- commerce negotiated transactions, intelligence sharing processes, etc,

[1964] The critical vulnerability before and since after September 11, 2001 attack is the sharing of intelligence. There is a lot of lip service to sharing intelligence after September 11, 2001 but nothing

of substance is filtering down to the local [], firefighters, police people and health care responders, the first responders who see themselves as canaries in the coal mine.

[1965] The process of controlled release of critical extracts into a document, a declassified document/file or data stream for a "consideration" enables a successful sharing of intelligence and information. DigitalDoors enables overcoming the obstacles sharing of sensitive data and granular data between parties with a controlled release mechanism for release of data in conformance to benchmarks that can include the giving of consideration, submitting of other information, et cetera. A major benefit of the DigitalDoors is the ability of parties to exchange their information for other party's information, in a controlled method that enables implementation of security protection on the sensitive data.

[1966] DigitalDoors solves major cultural and bureaucratic barriers, which impede sharing of information between government officials and state and local officials. For example local responders may decide to release to another organization (local or federal) information they have gathered on an anthrax attack only if they get in exchange information they need.

[1967] In order to view the critical extracts, the receiving party will need to reciprocate and present a consideration which may be releasing some of its organization's secret information.(for example: a local respondent will release information on specific types of symptoms if the other party will release information and studies on those symptoms and fatalities following such symptoms in other localities).

[1968] Although intelligence sharing is mandated by a number of laws new since September 11, 2001, the reality of intelligence sharing is cloudy for four major reasons.. First, other laws mandate the absolute protection of information. Second, ingrained methods for doing business will preclude effective liberal information sharing. Third, most national, state, and local agencies are unwilling to compromise

sources and methods, which is understandable given the potential for the careless release of information and the potential for the release of information to undermine the very intent of its collection and protection. Fourth, the process for sharing of information has no prior history and no defined functional workflows.

[1969] DigitalDoors technology resolves these issues in controllable, automated, and effective methods. In addition, it creates the workflow infrastructure and reason d etat to encourage information sharing.

[1970] First, DigitalDoors incorporates the legal and business rules for protecting information by predefined categories, word and phrase dictionaries, language translations, and special overrides.

[1971] Second, automation removes the risk for errant distribution of classified information and simplifies the process and workflow.

[1972] Third, the DigitalDoors technology for targeting, categorization, extraction, encryption, and dispersion makes is possible and simple to remove all reference to sources and methods which could undermine the data gathering organization and its very intent.

[1973] Fourth, the automation of targeting, categorization, extraction, encryption, and dispersion makes it simple control, distribute, monitor, and fine-tune leveled-access. In addition, control of extractions, indexes, links, and innate categorization of security, privacy, and risk enables internal or external control over a table of contents or the control over the hierarchical response to requests for information specific to needs or problems. Furthermore, this creates an infrastructure that engenders the sharing of information since DigitalDoors can establish leveled-access based on any number of access controls including username, actual location, time of day, security threat mode, permission, or level of participation in providing information for sharing in the mutual pool.

AA.8.13.0 Security through information Distribution

[1974] Multi Level Security - MLS is the provable solution for security within defined location and defined application spaces. It worked. Day-to-day tradeoffs included a mentality of security versus functionality. However, as these spaces multiplied and applications became more complex, the MLS solution implementation became too expensive, too unwieldy, too complex, and non-operational. People worked around it.

[1975] However, as a result of MLS complications and workarounds, new views of CTW, EAL, and TCW were implemented to enable operational functionality. Do such implementation meet the requirements for security? Does it allow functional information sharing? Is this a functional solution with guaranteed security? Whether or not CTW, EAL, and TCW are provable to provide security is no longer a useful question. Consider whether it is deployed widely and if it is even preventing the current security attacks with the massive information leakage and denial of services. The concept of the GIG is creating an infinity of at-the-edge locations and an explosion of sensitivity level tagging dictionaries with too many entries to allow rationalization. CTW, EAL, and TCW do not seem poised to present a coherent and effective security foundation for the GIG, transformation, and the complex new world of distributed applications.

[1976] The security problem is no longer about location and application but information and distribution. Actually the evolving security ecosystem must include knowledge discovery, integrity, security, and sharing, rather than some single-dimensional notion of security. If you can accept that a fundamental security ecosystem shift is possible, conceivable, or literally upon us now, you have take the first step to perceiving a new paradigm. This new paradigm repositions intractable security problems with very simplified views of information defined by a semantic structure supporting configurable

distributions. This implementation is asserted through granular content control (cross-platform, platform-agnostic, interagency, code and language independent). Location and application remain intact but is no longer part of the new security problem. Even when location and application becomes more complicated through such issues as complex computer applications, SAAS, SOA, Web 2.0, distributed processing, and dynamic with continuous cyclical support operations, security is applied specifically to information and distribution. Such is the fundamental archetypal shift and operative means to reassert security within this profound evolution. When the battle is fought specifically over information and distribution, old friends like MLS can be recycled because it is proven and extensible into the new security ecosystem. However, you regain control over security by shifting the security problem ecosystem from location and application to the ecosystem of information and distribution in order to win this war.

AA.8.14.0 Implementation of information sharing with Rigorous Protections

[1977] Information sharing occurs on a daily basis across all organizations -- both with and without permission or control. These examples constitute both the planned sharing and the inadvertent distribution of information, like:

Table:example information sharing methods

The distribution of a report to a defined list or readers
the push or pull of information is based on user authentication
the removal of a desktop computer for upgrade and recycling
the theft of a laptop; the loss of USB thumb drives
backup media lost out the tail gate of a courier
output sent to a group printer or FedEx Kinkos
messages pushed and pulled through a GIG infrastructure

messages expanded with additional human insight and intelligence

reports distributed, cited, copied in whole or in part, and referenced in new reports

insiders copying database tables for new projects bypassing original access rights

[1978] The issue is not so much that information sharing occurs or that new laws aim to encourage additional inter-agency information sharing, but rather that information sharing is ad hoc, inadvertent, and undermined by advances in data mining, aggregation, inference, and data-to-data interaction. When information sharing lacks formal usage controls, it undermines BLM security, compartmentalization, and privacy and confidentiality, and rules for data stewardship fail.

AA.8.14.1 General Concept

[1979] You can protect resources, like information, in just two ways; with perimeters; or with dispersion. For the first option, a wall around resources in the whole defends that multi-dimensional perimeter. For the second option, dispersed resources are no longer useful in parts, replicate the parts for survivability and mobile delivery, and enable reassembly of parts into the original functional whole or reconstruction with repurposing for other results.

[1980] The first option is the long-standing traditional way, effective and efficient but not completely protective. It works so long as the perimeter is intact, so long as the resources are not compromised from the inside, and so long as these resources are not repurposed and redistributed. Perimeters are created with firewalls, guards, access controls, encryption, and need-to-know hierarchies. This is effective for materials but not information since information can be replicated, reformatted, distributed and redistributed by secondary recipients, dispersed, inferenced, aggregated, and repurposed in new forms -- all the while the original information seems intact, seems never to have been compromised, and even seems never to have been touched. Each digital copy is exactly bit-for-bit the

original but does not increment access counters or indicate attribution or distribution. There is no control over any full or partial copies and copies of copies.

[1981] The second option is extraordinarily effective for information. You can pull the structure and the content from the immediate storage or from any available replicated remote storage at wire speed and respond instantly with reassembly at wire speed. In contrast, dispersion is not effective for materials. For example, if you disassemble your weapons with parts sorted into bins, secure the ammunition in lockers, and put trigger guards on the guns, you would be slow to respond in an emergency with weapon ready and loaded. In contrast, reassembled information can be pushed and pulled to thousands of users simultaneously anywhere at wire speed; not so with bins of gun parts in the armory and users at the front lines or in impromptu defensive positions. Wire speed data replication and the semblance of parts interchangeability plays into the strengths of dispersion for security because reassembly with almost the correct items is indistinguishable from the exact items. When perimeter security is combined with DigitalDoors dispersion and reassembly methods you can protect granular items at will, anywhere and anytime, at rest and in transit, from outside attackers, "trusted" insiders, and from data inferencing, aggregation, data-to-data interaction, and repurposing. Backups on tape, replicated computers, or at remote sites remain safe because of the dispersed granularity.

AA.8.18.2 Implementation map:

[1982] The DigitalDoors' paradigm illuminates information security in new light. This paradigm shift is not just an academic theory; it is a matter of fact and available demonstration. The software actually runs on servers, desktops, laptops, tablets, diskless trusted clients, NAS devices, web servers, local area networks, wide area networks, and email infrastructures. The paradigm shift as implemented is a series of procedures. Specifically, the procedures includes steps defining:

Table: DigitalDoors Procedures

How to identify information on a minimum granular basis

How to granulate the whole to any arbitrary level of parts (i.e. "items" or "granules")

How to assign items to a multi-tiered taxonomy

How to split items from metadata, cultural, semantic, or semiotic context

How to separate content from context and from concept

How to disperse items to different (remote and replicated) storage

How to assign item-level security based on taxonomy

How to create a redacted whole

How to create an architectural map for mining, inference, aggregation, connecting the dots, and reconstitution and protect against ad hoc data-to-data interaction

How to reconstitute the items into the whole

How to partially reconstitute the parts subject to security dominance (or any other taxonomy)

[1983] The multi-tiered taxonomy is a critical aspect to implement effective and efficient inter-agency information control and sharing. Items, tear lines, references, and reports fulfill multiple purposes within a single agency or organization and potentially many more for multiple agencies. As a result, tagging information with a single SBU or TS tag constrains information sharing with additional levels of perimeter guard revaluation. The introduction of metadata tagging for purpose, content control, and additional security assessment only increases the complexity of perimeter guard revaluation and the potential for inadvertent out-of-channel security lapses. This is one of the complexities of defined cross-domain tags that frequently preclude freely flowing information sharing. This is also one of the complexities of metadata. Instead, out-of-band multi-tier tagging enables assessment of information

sharing based on the potential for information leakage and loss of privacy through aggregation, inference, and data-to-data interaction.

[1984] An "item" is any part of the whole from as little as... a bit, a pixel, punctuation, a bullet point or footnote, a field marker, a character, a word, a phrase, a line of a document, a sentence, a paragraph, a sound bite, an arbitrary area of an image, a link, an encapsulated object, a metadata definition, or the metadata for the item, to as much as... a document inclusion or the entire whole data stream. The "granule" is an alternate designation for "item" as described in the academic literature of "Granular Computing".

AA.8.18.3 Technical Implementation

[1985] DigitalDoors information selection, extraction, dispersion, and reassembly technology worked immediately from the initial experimental program. The concept and the basic implementation has been stable, reliable, flexible, and fault tolerant. Information can derive from almost any format; a database, a structured field-oriented flow, freeform text, to arbitrary data streams of text, signal, audio, video, or hierarchical metadata. DigitalDoors can interface with your data and your stovepipe infrastructures, SOA web integration, on-demand applications, or software-as-a-service, and COTS workflow. Certain complexities inherent in purpose, dispersion security, and achieving multi-process integration have dictated optional dispersion formats so that many goals can be achieved at the same time, specifically security and sharing, privacy and confidentiality, continuity and survivability, as well as information utility.

[1986] The DigitalDoors machinery is the black box, like a turbine engine. You can use it almost anywhere. However, this facility that must be engineered to the platform, size, location, system requirements, user needs, and the existing cultural processes that must be supported. DigitalDoors has

created shrink-wrapped applications for Windows, MS Office (Word, Outlook, PowerPoint), and interconnected with trusted workstations, SMB and Samba, FTP, web servers, virtualized server environments, and Sun Trusted Solaris. This demonstrates our ability to integrate with other complex platforms, external devices and processes, and coexist with COTS software, service-oriented architectures (SOA), and proprietary components. DigitalDoors supports granular content extraction with placeholder substitution with 30 or more options, including misinformation. DigitalDoors includes an infrastructure with supporting tools and toolkits for access control, encryption, hashing, and random number generator, strong password creation, complex security and sharing policy administration, selected disclosure, anti-inferencing, information taxonomy, data mining, and sophisticated content, pattern, and context recognition, all running easily on COTS laptops or terminal-services or web-enabled infrastructures.

[1987] The demonstration implementations are simple in order to show ease of use and applicability for an inexperienced workforce. Other implementations and demonstrations are complex because of the veritable buffet of options readied to support one user or thousands of users. Our goal is an invisible user interface so that Word, Outlook, and MANMAN with DigitalDoors in play look no different from these COTS products. We can integrate with CAC access cards and the operating system for seamless access control and user sensitivity level and security compartmentalization.

[1988] We strive to simplify the user experience, and therefore many functions and features are hidden to provide a one-button user interface or less through automatic integration. We cannot show you all the bells and whistles in one sitting. In fact, you might not even want these bells and whistles because every added option increases training and administration overhead. However, we present this complex infrastructure as a gauge of the existing available architecture, depth, flexibility, extensibility, and

capacity for integration.

AA.8.18.4 A Tool in a Toolbox

[1989] DigitalDoors is just a tool—just like MS Word, MS PowerPoint, PhotoShop, a hammer, a wrench, a turbine engine, or a carbine. DigitalDoors is not a solution any more than a hammer is a solution. Just as a hammer can pound a nail, pull out a bent one, or open a glass soda bottle, DigitalDoors is a solution for granular security within a network of open or secured networks over a range of configurations and applicability. We do not insist or emphasize the best way to use this tool; we only provide it you as part of a larger toolbox and work with your system engineering teams to effect integration, workflow, or create new solution sets. We can help you with issues of applicability or define the problem set in order to craft solutions, no button or one button functionality, and seamless integration into operational workflow and stovepipe applications. We strive, as stated in the prior section, for simplicity with obvious utility.

AA.8.18.5 Maturity of the Paradigm Shift

[1990] Dispersion of information—to enable security, privacy, survivability within the complex and counterproductive environment needed to support knowledge management, data stewardship, security, confidentiality, continuity, survivability information sharing, and dispersion—is a paradigm shift. It is diametrically transposed to the traditional policies of security isolations, compartmentalization, air gaps, and need to know hierarchies. Perimeter defenses thwarts effective information sharing, networking, and push/pull data analysis. If you want to connect the dots, you need to disperse the information but without revealing sources and methods. Too much information or unlimited access allows opponents the opportunity to infer and repurpose that information. Control of granular items on a contextual basis maintains an operational balance between security and usage while retaining

information control.

[1991] DigitalDoors is a turbine engine with functions and features ranging from manual to automatic content selection to complex support for network dispersion and replication of information. There are implementations of declassification redaction so that output is automatically directed to unlimited sensitivity levels and compartments. Information can be versioned and delivered to enable users in different groups, organizations, and external partnerships to connect the dots without violating sources and methods. An administrative facility enables taxonomy rules to be added, deleted, or changed to support individuals, groups, or entire organizations. Security can be suggested, overridden, or fiercely enforced through OPSEC and COMSEC oversight. Taxonomy tools classify granular information units subject to sensitivity levels, multiple compartments, groups, categories, and targeted analysis tracks. Cross-domain distribution on-the-fly is enabled because multitier tagging is reassessed prior to upgrading, downgrading, or multiple sensitivity level distributions. Automatic conversion of documents from the raw to the partially coded can be turned into tear lines. Reporting, defenses in depth, and numerous other features extend the basic to the extreme.

[1992] Just as many other security technologies were introduced as applets and add-ins but later became so indispensably as to be incorporated directly into mainstream application and operating system kernels, many DigitalDoors features could track the same path to maturity. The sample applications, integration with an SOA web site, and Office and Outlook add-ins are just beginnings for the implementation of our paradigm shift. Define your requirements and engineer this DigitalDoors turbine to your needs and desired platform.

AA.9.0.0 Intangibles:

AA.9.1.0 Creation and production of intangibles

[1993] DigitalDoors makes the assertion that everything tangible and intangible can be represented as information and in its reduction is information. Plants and animals are based on a DNA blueprint. Rocks and air are created from an atomic structure represented by information. Buildings and cars are built per design from other tangible inputs (themselves the manifestations of information). Everything else, that is data, is information. Energy delivery to enable the production process is directed by information through relays, communication channels, and meters. The fundamental difference between the intangible information and its physical counterpart is the manufacturing process that turns the information into a physical representation. The physical reality of tangible information is the conversion of matter and energy into other matter and energy. Information is represented as arrangements of matter and energy, whether as letters in a book, magnetic fields on a disk drive, or power differentials within storage vessels. These are physical (and tangible) expressions of logical (and intangible) information.

[1994] An expression is a mode, means, or use of significant representation or symbolism. It is the manifestation of the intangible design. RNA, the functional building materials of life, is the expression (or one of many possible expressions) of the DNA blueprint. Industrial manufacturing is the expression of inputs and process to create an output product. These are things we do not normally think of as physical manifestations of information. Generally, we cannot convert electrons, protons, and neutrons into atoms or atoms into molecules. However, our technology is getting to the stage where we can replicate some of the fundamental physics to create these microscopic manifestations through information-driven factories. Even light and energy is something we manipulate and convert from other physical forms and sources to carry information and enable the manufacturing engines of the industrial age. This same technology has transformed into the engines of the information revolution where the

inputs and outputs are sometimes completely intangible.

[1995] The physical manufacturing process is not that much different from a computerized application. Directions realign information into new forms, and these become products and services. The manufacturing process is represented by a cookbook, a blueprint, or a design method expressed as an informational description. A computer application, a workflow, a macro, and computer code is merely a series of directions patterned as information and expressed to create the logical or physical output. Computer chips and circuits are physical manifestations of computer code and instructions. These are all reducible in their lowest forms to information. Some outputs are pure information, some are manifestations in physical form of information and the information that drives the manufacturing engines. However, in reduction, everything is information or the expression of information.

[1996] It is also essential to realize that process is a modifier of data in ways unanticipated. As such, process is an expression of data or the blueprint that is always based on a raw data format. Consider this example of the MS Word document routed through an interoffice document management system. Each user in turn reviews the document presumed to be a work-in-progress. However, as the document is distributed downstream, individual user-level macro settings for the spelling and grammar check could either highlight suspected problems or automatically address them to actually change the document content. Words change if unknowns and convoluted syntax is chopped up into shorter phrases. These alterations often substitute the wrong common words and break the meaning. [Need a visual sample here].

[1997] Because individual users can change/add/delete dictionary and syntax entries, the ultimate distribution of the original document could bear little resemblance to its initial concept and frequently makes little sense. This is not an exercise in possibility, but a reality of MS Word with common

SharePoint or Exchange workflow systems. Although this might seem a minimal risk, this type of macro process as is poses a covert communications channel. Enhanced, it poses even more of a sophisticated attack. Substitution of altered dictionaries and grammar rules easily pervert a low value plain-text into a sophisticated substitution cipher at the contextual and conceptual levels. This should be viewed as a risk where raw content is perverted by expressions of metadata and contextual changes not inherent within the raw content alone.

[1998] In some cases, we have data describing information as tags, classification, attributes, and properties. Some of this data describes how the data should be applied or formatted, or it describes the processes for manufacturing something else from it. This so-called metadata is just the expression of the data and the methods to turn data alone into building blocks. It is still information, albeit more complex information, and is still the expression of information.

[1999] If there is something that cannot be represented as information, DigitalDoors would like to know. There are many things that can be represented as information but cannot yet be created de novo from information alone. That is a limitation of our available production methods, not the theory that everything can be represented by information, created from information, and springs from information and its expression into another form. However, we can stick our heads in the sand rejecting the potential for a security risk because it seems too far out in some future fiction.

[2000] What once was unknown became fiction and then became reality. The information revolution has transformed our world exposing heretofore unconsidered risks. Information is content, context, and concept. These are all forms of the expression of information, and all are at risk. Forms of security that block some forms of this information, for example, perimeters, firewalls, guards, deep packet inspections, safe coding methods, are not sufficient security methods. They are efficient security

methods for the current expression of that information but not sufficient to reflect possible transformations of the expression. Expression must be recognized for its asset value and protected.

AA.9.2.0 Intangible data assetization

[2001] The assetization of both intangible and tangible assets requires a level of granularity in order to reflect the complex nuances of finance, legal, and operational workflow. This creates the algebra for measuring raw, aggregated, and processed information for both risk exposure and asset value. This includes initial collection through customer distribution and inclusion of information and other intangibles as a production input to a finished product.

AA.9.2.1 Assetization for Asset Monetization

[2002] In the case of intangibles, specifically information, it can be transformed by content, format, aggregation, inference, interaction, structure, presentation, and other attributes that alter its valuation. Granularity is represented in terms of describing the raw material or finished goods with variations on the same, and in terms of redaction, distribution, and utilization. Granularlity is the semantic, contentual, cultural, or semiotic representation of elements within a data stream. For this reason explicitly, any intangible assetization process should reflect the granularity of information and generate monetary valuation reflecting specific attribute expression and its value (decay or growth) over time. Monetary valuation need not be fixed or static based on internal content or attributes, external cultural or contextual frameworks, or the passage of time.

[2003] For example, consider an initial MS Word document. It contains both obvious content, context, and concept in plain sight. It also contains a tremendous amount of contextual and covert channel information. This is normal hidden metadata or even that which is expressively secreted into metadata and only accessible through secondary encryption keys and other add-in applications within

MS Word or a workalike application. It is easy to hide metadata inserted with earlier versions of Word from later releases because the document structure has changed between versions; similarly, MS Word metadata can be hidden from the web-based Google desktop tools or WordPerfect. The same is true for any other semi-structured or structured data source, including databases, SQL, spreadsheets, email, audio, images, video, and application configuration files. Images overlayed on MS Word text can be used to fully mask the existence of content. DigitalDoors exports that insecure format into plaintext for any subsequent processing; it can also optionally create a monetary valuation on the raw conversion from the MS Word file before that conversion in order to capture the differential value of that visible and hidden meatadata.

AA.9.2.2 Differential Valuation

[2004] DigitalDoors Secure parses and classifies that plaintext document and optionally creates a monetary valuation for the baseline information. Selection and extraction of crown jewels creates a redacted version and optionally creates another monetary valuation. The differences between a monetary valuation of the information at any point in this workflow expresses differing risk and valuation. It is not static. It is however, reproducible. It is also dynamic in that as dictionary valuation, organizational missions, and policy evolves, the monetization of any particular information set will represent that current state. Differentials over time will represent fundamental decay in information or the generation of new organizational knowledge and wisdom. This is relevant to reflect the development over of time of a trade secret and the resulting increase in asset value of the trade secret.

[2005] The value of information is inversely related to its distribution. The more widespread and well-known information is, the less unique it is and the less valuable it is. When information is controlled naturally or artificially, it becomes scarce and hence more valuable. Copyrights, patents, and

trade secrets enforce the scarcity of the information or control the utilization of the information. These are both forms of scarcity, trade secret being natural and copyrights and patents providing an artificial usage limitation. Whatever the methods used to create this scarcity, monetization is increases with scarcity. Granularity of information breaks the context of the information and artificially creates formlessness and this scarcity, also increasing the value of the information. The the security of information is proportional to the scarcity of it.

[2006] DigitalDoors reflects the value of information based on many defined attributes, including sensitivity level, user, group, organization, role, mission, location, purpose, scarcity (in particular through the measurement of monetary valuation), and also other soft properties. In addition, DigitalDoors creates a baseline redaction and versions which can be reconstituted not all, some, or in full. This reconstitution, as described in other papers, fully supports granular content control such that the resonstitution can be driven by the specific authorized user, the group, the category, by sensitivity ;level and compartments, or as fine-grained as per item-level extraction. This reconstitution contains different information than another reconstitution with different item reinsertion, and therefore, should represent a different information risk and demonstrate a different monetary valuation. DigitalDoors fully supports differential monetary valuation to whatever level of granularity is supported in the baseline or version redaction and the corresponding reconstitutions.

AA.9.2.3 Conclusion

[2007] In summary, DigitalDoors generates monetary valuations for raw, redacted, versioned, and partial or full reconstitutions. The differentials between an information set in any two different states represents both the monetary value from exposure or leakage and the difference between the two states in terms of the extraction item-level extraction and contextual structure. Therefore, not only is the

original information evaluated, a redaction is also evaluated, full reconstitution or reconstitution of the format-transformed information, as are other specific and intermediate expressions between the original, the restoration of the original, and anything in between. Assetization requires a level of granularity for robustness and utility.

AA.9.3.0 The Intangible assets production model

[2008] Security is the necessary to sustain the competitive advantage provided by the wealth production model of the information revolution. Without protection of the elements enhancing the basic production model, this information revolution production model reverts to either an agricultural model or industrial model for production of commodities. There is nothing wrong with the production of commodities, but it is not sustainable and does not yield the range of returns available to complex, information-based monopolistic services.

[2009] The differences between the information revolution model with its post-industrial enhancements is the input of intangibles, the output of intangibles, and the multiplier effect from the intangibles on the actual efficiency or specialization of the manufacturing process. The multiplier effect enables higher margins, lower production costs, or monopolistic production presumptive on the enforcement of patent or the protections of trade secrets. The intangibles are most often data or restricted knowledge or the special case of a patent. The post-industrial production model reverts to commoditization with the expiration of patents. It also reverts with the exposure of data and trade secrets.

[2010] A sustainable competitive advantage enables product or service pricing margins that exceed those of commodities or provides a means to produce commodities at prices below the competition. It is an enviable position that enables better profits, growth potential, and a future.

[2011] The only sustainable advantages for agricultural or industrial models are based in government-granted monolopies of land, resources, markets, or regulation. Few markets are available today with such opportunities, so most institutions make these opportunities themselves. The opportunities in the information revolution are based solely on data and knowledge (a form of restricted data), and these are used to create a competitive advantage. The opportunity to create such an advantage is available only so long as the data and knowledge is unique and restricted, so every should be made to retain this advantage.

[2012] However, some knowledge becomes obvious, such as methods, so a monopolistic protection is granted with a time-limited patent. Often a patent is too little and too short, so institutions hope or effect data and knowledge limitation through security. The only venue for maintaining that restricted knowledge is security. It is only sustainable for the duration and effectiveness of that security. So while security does not augment wealth creation, provide inputs or outputs to the production model or generate ROI or contribute to EPS, it is essential and necessary for the continued sustainability of a competitive advantage.

AA.9.4.0 Monetary information valuation

[2013] Monetary information valuation is an important strategic, legal, financial, operational, and managerial concern especially when 95% of organizational value is reflected within intangible assets. It is essential to assign monetary value to information so that it can be integrated within the strategic and traditional roles for managerial control to bridge this divide. It is also essential to realize the inverse valuation of information based on scarcity. The more information is controlled for its uniqueness, the more valuable that information. Distribution and downstream information sharing, while essential for workflows, erodes value.

[2014] Information is intangible by definition, and managing and accounting for intangible is exempted from most functional activities. Information is also hard to define, hard to discover, hard to classify, and nearly impossible to relate to purpose and value. DigitalDoors systems and methods overcome these limitations with flexible, robust and automatic ways to discover information, classify it, assess ownership and potential, and assign an asset valuation that can be integrated in the traditional and conservative control methods. DigitalDoors values information statically and creates dynamic differentials subject to attribute perspectives and reflection of internal or external systemic changes with time.

[2015] The fundamental difference between tangibles and intangibles is that intangibles tend to be unique rather than commoditized. While there are many ways to assign monetary value to information, the DigitalDoors assetization and monetary valuation methodology strives to make this functionally simple, and in particular, automatic. Automation is required to support the volumes of information seen today and the obvious industry-wide exponential growth experienced by information warehouses. This valuation method is primarily automatic, driven by the DigitalDoors Secure application, but flexible enough to support most organizational requirements and allow for manual overrides and revisions. Value is also adjusted automatically as the information within the warehouses changes due to mission revision or just because things change over time, and because production purpose adjusts to reflect the competitive environment.

[2016] The takeaway from this paper should be that random information is unlikely to assert a meaningful or functionally-relevant message. Therefore, primitive content analysis (with white lists and black lists) must be augmented with context and concept analyses to measure such other factors as authentication, integrity, source, method, chain-of-custody, utility, purpose, perspective, functionality,

relevancy, and connection. This specific DigitalDoors measurement system and method supports the proxy recognition of ownership, utility, purpose, frequency, source, method, and other significant organizational behaviors and requirements. In addition, this method recognizes static assignment, information lifecycle management, and the dynamic revaluation of information as it relates within information warehouses and the larger strategic organizational workflow as legal, financial, and operational agendas advance to reflect new challenges.

AA.9.4.1 Monetary Valuation Methodology

[2017] Note that significant cultural and semantic limitations complicate the monetary valuation of intangible property, in particular information, and more specifically, patents and trade secrets. A number of basic assumptions should be met. These assumptions reflect information as data streams consisting of content, context, and concept, a lexicon referenced frequently by DigitalDoors. Specifically:

1. the whole of an information set should have more value than the sum of its separate parts

2. the assemblage or information production should create more value than the parts and its liquidation value

3. the parts should somehow relate to each other and reflect a progressive, significant, and culturally or functionally meaningful narrative

[2018] These assumptions define that a data stream is not merely the random collection of high-value dictionary entities. These assumptions define that a data stream has a coherent purpose, utility, and/or imparted knowledge. The assumptions also represent that a structured data set or other arbitrary data streams contain a culturally-relevant message or that the message relate to a functional process in a meaningful way. Note that the code, format, structure, metadata, symbolism within a data stream might

not be meaningful to every person or just any workflow, but if this stream has a matching recipient, it qualifies as a viable document. Under the concepts of Claude E. Shannon, information theory regards information as only those symbols that are certain to the receiver.

[2019] A unifying theory known as information theory was developed and became the subject of intensive research. Claude E. Shannon, whose initial ideas appeared in the article "The Mathematical Theory of Communication" in the Bell System Technical Journal (1948) defined information to include the messages occurring in any of the standard communications media, such as telegraphy, radio, or television, and the signals involved in electronic computers, servomechanism systems, and other data-processing devices. The theory is even applied to the signals appearing in the nerve networks of humans and other animals. The signals or messages do not have to be meaningful in any ordinary sense.

[2020] The chief concern of information theory is to discover mathematical laws governing systems designed to communicate or manipulate information. It sets up quantitative measures of information and of the capacity of various systems to transmit, store, and otherwise process information. Some of the problems treated are related to finding the best methods of using various available communication systems and the best methods for separating the wanted information, or signal, from the extraneous information, or noise. Another problem is the setting of upper bounds on what it is possible to achieve with a given information-carrying medium (often called an information channel). While the central results are chiefly of interest to communication engineers, some of the concepts have been adopted and found useful in such fields as psychology and linguistics. The boundaries of information theory are quite vague. The theory overlaps heavily with communication theory but is more oriented toward the fundamental limitations on the processing and communication of information and less oriented toward the detailed operation of the devices employed.).

[2021] Nothing in this theory precludes obfuscating the valid message (with hiding, dispersion, encryption, etc.) so as to mask the source, recipient, or potential value. This obfuscation just raises the possibility than a message is viable only from within a specific perspective context. Indeed, holograms can hide vast amounts of information only retrievable within a precise perspective illumination angle and light frequency. The frequency- and angle-encoded information will protect the hologram against attacks completely or partially unless the refraction angle is guessed and the radiation frequency is a whole number multiple of the recording frequency. It is a great security method but not common because of the technical complexity. Padded or hidden messages qualify, but separating value from the chaff in such obfuscated delivery can be an intellectual challenge for all but the recipient.

[2022] Any valuation system is liable to gaming against this system. For example, populating a nearly meaningless document with gratuitous inclusions or repetitions of high-value words. Analysis of a data stream for relative statistical balance of content, context, and concept is an effective and efficient method to vet the integrity, validity, and aggregation of parts into a coherent whole. See the whitepaper A Valuation Basis for additional and detailed assessment of the assumptions, restrictions, and practical methods for assessing information and ascribing relative contextual value to information. Hefty rewards encourage gaming against the system, such as incentives to share information among agencies, boost the search engine placement, increase the number of hits from search engines, pad the frequencies of reference or copies of a document, or valuate an intangible production method far above any reasonable economic value.

[2023] Two factors play into gaming the system. First, an information measurement system must be simple, understandable, extensible, and open. Second, the measurement system must be robust enough to rate all types of data streams in any format, structure, and encoding method with any amount

of metadata (whether internal in-channel or external out-of-channel) and be robust enough to protect itself from schemes to skew the results. If the current state of security shows us anything, these two factors not only are conflicting, they provide the very venues for attack. Converting open systems to closed undermines scientific methods and also gives special dispensation to those that can expose the hidden methods. They will and do come to light, and the more at economic incentive, the sooner these are undermined. The system will be attacked for certain and it is managerially responsible to review automated measurements for abuse, even with huge data warehouses and the acceleration exponential growth in data storage. Intangible information is the overwhelming market value of an organization because of the changes from the information revolution.

[2024] Augmentation of simple measurement with context and concept analyses to measure such other factors as authentication, integrity, source, method, chain-of-custody, utility, purpose, perspective, functionality, relevancy, and connection protects the system from attacks and review of outliers and assessment of content relevancy to context and concept is always effective. Discontinuities, nonsense, grammatical errors, lack of meaning, randomness, statistical distribution, and excessive values are clues to attacks. In particular, randomness, statistical distribution, and excessive valuation are automated and require limited manual review. Violation of configurable weights and thresholds kick out the outliers as discrepancy reports.

[2025] The difference between measuring tangible and intangible assets is arguably complicated. Inventory count (or frequency) is not always a good, unassailable, or effective measurement statistic for intangible assets. Although ten units of data content exceeds one unit of data content, it often does not impart more value. One, some, or all ten units might have no value at all or relevance together, so aggregating it yield no value at all or even less value than one, some, or any of the ten units alone. It

could be disconnected or repetitive, so valuation is fixed no matter the size, padding, or replication. Intangibles are typically unique, so aggregation of intangibles either yields no contextual value or a new (disconnected and far higher) value with limited relationship to the individual parts alone.

[2026] On the other hand, aggregation, inference, and interaction of data as a whole do represent more than the some of the parts. Weaving together key concepts for a practical message or functional purpose is the essence of knowledge creation and the leads to the monetization of the aggregation of intangibles as valuable assets. There is one last consideration. This assemblage of intangibles provides monetization value subject to a perspective. One group within an organization might have no interest or incentive to use this assemblage whereas another group sees this as an opportunity to increase the efficiency or lower the cost of a production process. Likewise, assemblage might well have no internal utility, but might represent a significant potential for sale as a product to others or a service for license. Measurement must transcend frequency count to include measurements of context and concept with additional relevancy to the core mission or other needs as discovered.

[2027] Consider that a bin filled with #12 size 10/32 screws (quantity 14,000) is worth more than a single #12 size 10/32 screw (worth $0.03). An inventory valuation is the part count times the unit value, in this case 14,000 x 0.03. However, repetitions of information do not reliably represent any or even a natural linear increase in value. For example, a database with one thousand entries for "bin Laden" should not be assessed with a value of 1000 times the value of the single "bin Laden" reference. Inclusions and repetitions might indicate higher value, although this is neither necessary nor sufficient. Although simple, frequency counts can be misleading. Sometimes it is contraindicative to value as exclusivity and maintaining exclusivity forces greater value. In some cases, the true information does not even indicate the content at all and presumes that the message recipient is intelligent enough and on

the same conceptual wavelength to comprehend the missing context. In fact, the missing content and the omitted context specifically provides that authentication. Quality becomes integral to quantity. Quality is a function of content, context, and concept.

[2028] Search engine optimization (SEO) techniques pervert classification and tagging techniques to promote an agenda different from the intended search engine mission; these same techniques and the corresponding techniques to detect this game playing and thwart it are relevant for monetary information valuation. It is important to assess whether the presented content reference is related to anything, is consistent with links to any utilitarian context, or conveys any concept. Discrepancies between content, context, and concept, the relative lack of any one of them, or imbalances of relevance distinguish a breach in the previously listed assessment assumptions. Content without context has no value. Context not yielding a concept might have some small value but lacks any implied or explicit utility. This is indicative of flash but no substance. Intellectual property and other intangibles have little value without context and consistent concept.

[2029] Specifically, documents (inclusive or all files, messages, field-oriented data sets, structured and unstructured formats, metadata repositories, and textual, audio, or video streams) are processed for content, context, and concept. These results are aggregated and adjusted by user-configurable monetary values and functional descriptions. These are frequency counts, but mediated frequency counts. Although the monetary valuation of information is intended to support accounting activities and support the assetization of intangibles, be cautious with raw frequency counts. Bean counters like this type of information, but is always suspect. Repetitions of information, copies of information, suspect information, counterfeit or fraudulent information are easily constructed and skew frequencies. The chain of custody along with the sources and methods for data collection and production

is very important.

AA.9.4.2 Weighting Content, Context, and Concept

[2030] Weights are applied to reflect the best or desired internal understanding of organizational requirements, recognition of internal trade secret archives, and reflection of the complex but evolving intangible property basis. The details are described as follows, and are also defined by the included Excel spreadsheet and report samples. Every white list, black list, and gray list dictionary entry and all entries in My Word List, My Group List, My Category List, and My Policy List are assigned a base monetary value. This value can be static, reflect occurrence frequency counts, or comprise an algebraic formula. For example, "bin Laden" can be assigned a fixed value of $200 if that entry appears once and many times with a document, $200 each for each and subsequent occurrences, or a formula of $100 squared for each occurrence reflecting an increasing focus in value as the occurrence counts increase.

[2031] The value is variable and is configurable through administration control. The formula is also variable and configurable, but generally uniform for content and context occurrences for simplicity of administration and maintenance. DigitalDoors does not support conditional and discontinuous functions at this time, but is working on implementing this extension specifically to address the shortcomings in straight frequency counts because it looks promising to improve weighting options. Game playing with SEO and gratuitous inclusion or repetition of keywords is muted by weights that assess the contextual basis for this content. Content without relevant context or content that does not represent or sustain a concept is gratuitous.

[2032] Because "bin Laden" is a name item by category, reflects a specific mission or role, additional values can replace or augment the basic dictionary value assignment. Weights ranging from 0 to 100% stress the context importance of content or applicability of that specific information. This

value can be static, reflect occurrence frequency counts, or comprise a formula. For example, "bin Laden" can be assigned a fixed value of $110 if that name category appears, $200 each for the first each subsequent occurrence, or a formula of $125 and $100 squared for each additional category occurrence reflecting an increasing focus in value as the category occurrence counts increase. The value is variable and is configurable through administration control.

[2033] Recognition of "bin Laden" as a family name instead of as the known terrorist is part of the core DigitalDoors multitier hierarchical overlapping classification system, and a reflection of the need to account for and balance the relationships for context, context, and concept. The specific instance of relevant classification will determine the monetary valuation based on whether that name is recognized as a family name or a terrorist, and the monetary value or risk assessment will correspondingly vary based directly on that specific assessment and utility. Different dictionary entries set different monetary values and different classification assignments will drive different formulaic assetizations. Two different documents or even copies of the same document repurposed by different users might yield different values. This is both confusing but liberating because it enables the accurate reflection of content and context value based on intent, purpose, and specific production utilization. For example, one user might want to assess the ROI and thus clout of the Bin Laden Construction Corporation over the Saudi political structure, whereas a different user might want to trace all sightings and locations of the terrorist. User, role, mission as well as other classifications will rightly drive the monetization of the same documents to different values for these different uses.

[2034] DigitalDoors categorization on a conceptual level is robust and supports sensitivity level and the other multitier hierarchical overlapping classifications. These values are counted for all classifications and can be presented as raw values, tearline rollups, document rollup, and selection

distribution throughout the user, group or entire organization information warehouses. The frequency counts are assigned a base monetary value. For example, 6 items are tagged as [TS] but rolled up to 4 [TS] tearlines and establishes the overall sensitivity level of the document as [TS]. A group definition (My Group List) could well recognize these [TS] items as lesser [S] or [C] or even [U] and downgrade the outcomes accordingly, and assign lower monetization values. It is a built-in method of weighting frequency counts and assessing semantic meaning.

[2035] Likewise, the other active sensitivity levels are also aggregated, although these values are subordinate to [TS] settings until such time as the redacted document is distributed as [S], [C], or [U] variants or delivered as a standard DigitalDoors base document and reconstructed through granular content control. Both sensitivity levels and monetary value for variable content and sensitivity level distributions are fully monetized at a level consistent with and reflective of none, some, or all granular content control.

[2036] Note that [U] classification is legally necessary to sustain a trade secret claim. This demonstrates that not all information is a trade secret, that some is merely operational in nature, and that there is a valid and functional distinction between what is secret from what is not. It is not specifically necessary to use these or other labels for tagging value. However, courts typically want to see functional distinctions and a working balance between operational information and trade secrets in order to support the claim for trade secrets. Claims that everything is a trade secret, or claims that that do not include recognition that some secrets are more valuable than others is often a reason that courts dismiss trade secret cases. It is also natural to perceive and gravitate toward a value chain with secret distinctions. Furthermore, differentiation and potential monetary distinction enhances efforts to assess competitive potential and create effective strategy.

[2037] In addition, specific keywords are compared throughout the current state of the warehouse for raw frequency, weighted and summarized for all occurrences of matching keywords. Note that keywords support the multitier hierarchical overlapping classifications, with the result that assessments will apply the specific and dynamic classification and classification valuation to each occurrence. This method recognizes static assessment, information lifecycle management, and the dynamic evolving revaluation of information as it relates within information warehouses and the larger strategic organization workflow as legal, financial, and operational agendas advance to reflect new challenges. As information within the warehouses or the strategic mission of the organization changes, reevaluation of information value is automatic and dynamic. This reevaluation can be limited to periodic review or represent an ongoing and dynamic automated activity approaching current information inventory asset value on a continuous basis.

[2038] These conceptual values can be static, reflect occurrence frequency counts, or comprise a formula. For example, [TS] can be assigned a fixed value of $20 if that entry appears one or more than once, $20 each for the first and each additional occurrence, or a formula of $20 and $100 squared for each additional occurrence reflecting an increasing focus in value as the occurrence counts increase. The value is variable and is configurable through administration control. Similar valuation is calculated for raw counts, tearline rollups, document rollup, and keyword occurrences for all active sensitivity levels. Weights (0 to 100%) are applied for all active sensitivity levels and for the different frequency counts. The weights determine the relative importance of this conceptual information and its monetary value. The different weights and the user-defined asset valuation functions reflect the best internal understanding of the legal, marketing, and finance groups, and can evolve over time.

[2039] It is important to note that this methodology provide relative monetary or risk

measurement of trade secrets, redacted [TS], redacted [S], and fully redacted information released for patent filings, copyright filings, or marketing literature. This enables assessment of assetized intangible property, the changes over time to the intangible property including the addition of new information or information leaks that must be published to conform to SOX regulatory compliance.

[2040] It is more interesting to view [TS] and [S] in terms of statistical values of scarcity. See the white paper, Classification as Proxy for Scarcity for an analysis of how fuzzy tags can be represented by more exacting numerical values and not just discrete values. Specifically, when information is presented in terms of scarcity, the weighting becomes more useful and more flexible. Frequency counts, or inverse frequency counts in effect representing the scarcity of a particular term, one context from a universe of many possibilities, or an unique concept, become useful to plug into measurement formulae. It is important to note that scarcity is contextual based, and while information might in truth not be scarce, unknown, controlled, private, or secured, its relevance to a situation might be unknown and controlled, thus establishing a specific context for the scarcity and a reason to protect that scarcity.

[2041] This extends the discrete limitations of tagged sensitivity levels to a granular range with more flexible values, and certainly more reflective of issues that might be top secret for a specific organization and a current mission but irrelevant and public knowledge to a different one. This problem of reflexivity and security context is address in the white paper, The Failure of Tag Reflexivity for Information Sharing, is resolved by measurements of scarcity and the monetary valuation of information specific to context.

[2042] It is important to note additionally that the state of the warehouse changes, more documents are discovered, new documents are included in the warehouses, documents are purged to reflect operational requirements or shifts in competitive strategy. Secured, assetized, can be reevaluated

for monetary value. The resulting document value is dynamic and can be automatically updated to reflect these fundamental changes. This conforms to accounting needs to reflect the fluctuating values of IP inventory and dynamic representation of core SOX-stipulated valuation of material organization assets.

[2043] The valuation process is simplified almost to the extreme. Only four formulae define the concept for a sensitivity level. This facility does not need to be active. Only a single relative weight defines a sensitivity level. Only four weights define the relative importance of the frequency counts and keyword distributions for that sensitivity level. There is only a single relative weight for each context classification, of which there are six recognized types in total. Weighting dictionary entries is simple, except for the density of acronyms, words, and phrases included in it. The English dictionary contains about one million entries.

[2044] However, the value of entries defaults naturally to zero dollars, but as an organization defines acronyms, words, and phrases of risk or monetary value and establishes an entries preeminence through My Group List, specific base values can be assigned. Content frequency value is simplified by a single formula, whether by occurrence or occurrence frequency formula.

[2045] In operation with the likely and traditional four sensitivity levels of [TS], [S], [C], and [U] (by default is nothing assigned because it is value-neutral), there are only 14 global weights in three sets that must equal 100%, 16 global sensitivity level formulae, 6 classification formulae, perhaps 10 relevant categories in need of a base monetary assignment, and a single dictionary formula. A typical organization is likely to define 10 to 40 values for dictionary entries and modify this as needed only over time. A person experienced with pro forma spreadsheets or forecasting would understand these formula and weights almost immediately and would understand the opportunity to generate monetary value from intangible information and establish the foundation for creating asset transactions that impact the

financial balance sheets and any specific accounts in the chart of accounts. These formulae and weights approximate many of the natural and required virtues for information measurement and the monetary assetization of intangible property for balance sheets assets.

AA.9.4.3 Discovering and Monetizing the Intangibles

[2046] DigitalDoors asserts that many intangibles such as applications, data sets, workflows, cultural knowledge, ways of doing business, relationships, reputation, as well as the more accepted trade secrets, patents, brands, and copyrights are expressions of information. See the paper Commonality of Data for an explanation of this concept. The differences between these types of information are material only insofar as legal, financial, and functional distinctions can be made. These differences can be reflected by attributed and properties defined by or ascribed to the information and how the information is expressed as a process, distribution method, service, or product. This assertion is important for monetary information valuation because it functionally simplifies (standardizes) the representation of information. For example, format, structure, and encoding become attributes of risk, purpose, and value. Likewise, content, context, and concept become attributes of information. These attributes do not alter the information in any fundamental way, they just make it possible to evaluate it.

[2047] As a result of these circumstances, intangible information is dynamic. Traditional accounting has the tools to represent intangibles as assets but has been slow to adopt historical recording of intangible asset value. SEC unified reporting requirements, FAS regulations, GAAP rulings, international pressures to unify the rules, and the percentages of market valuation reflected by intangibles encourage consideration for assetizing intangibles. In addition, the worldwide pressures to unify intellectual property rights, create uniform and enforceable patent monopolies, and patent rule changes increase the need to recognize trade secrets and provide mechanisms for monetization.

AA.9.4.4 Static and Dynamic

[2048] Measurement of the monetary value of information must include both static and dynamic aspects. Measurement must be static for two reasons, reproducibility and integrity. Reprocessing of the same source document in the same time frame should yield reproducible static results. Results that become assetized as a historical functional transaction must have some stability in order to reverse or revise these transactions within GAAP. Financial accounting requires that integrity. However, traditional financial accounting has been unable to bridge the divide between tangible assets and intangible ones even though the predominant value of any organization is now hidden in these intangibles. As a result, financial accounting must advance and include these assets.

[2049] The complication is that intangibles do not have a static value. Tangibles do not really have a static value either, but in general that value does not fluctuate greatly. Financial accounting represents fluctuations for tangibles as inventory write-downs, adjustments, depreciation, depletions, and one-time charges. The expectation is that these fluctuations are represented as one-time events. The accounting profession is very conservative, to the point even now within this paradigmic information revolution of managing only 5% of the full value of the organization. Five percent is an interesting number. It is the usual and customary GAAP threshold under which accounting measurements and errors are not considered material. Is it possible, any way conceivable, to conclude that organizations are no longer in control over the true material reflection of the organization market valuation?

[2050] Intangibles are dynamic both in raw value and as monetary asset valuations. The fundamental aspect of intangibles is that they are unique. Market valuation with comparables is suboptimal. Valuation is not static because the relevancy and uniqueness is dynamic. A value production method could be enhanced with a new process and lose all existing value because it has been usurped

by the newer method. Uniqueness can be undermined by fraud, counterfeiting, exposure, replication, distribution, or information decay. New information could override and invalidate old information, and often does. Mores, market desire, and the fundamental societal state of knowledge can invalidate existing information and render it worthless or be incorporated to enhance existing information or aggregate with that existing information to create new.

[2051] To meet the dynamic needs for intangible asset evaluation, DigitalDoors enables the monetary valuation to evolve. The dictionary entries, the organization policies, user assignments, and content, context, and concept weights and measures are variable. These are configurable to reflect any change, negative or positive or merely incidental. However, when these changes are stacked, the power of aggregation, inference, and interaction is clear. The results are automated and reflect the evolution of the information and its purpose or functionality; the user does not have to do anything or do any manual review or adjustments. As a result, reprocessing a document will yield different results over time. This differential can be applied for asset valuation adjustments, such as one-time charges for complete devaluation, periodic depletion, and assignment of transfer value when information is applied to production or production method enhancements.

[2052] Dynamic intangible valuations yield a more active importance. Differentials can be delivered as reports, encouraging the strategic direction of business or responses to the competitive climate change. Differentials, reflecting an assemblage of information into new forms or new conclusions represent the opportunity to capitalize on the creation of a sustainable competitive advantage with R&D, product line adjustments, or new organizational direction.

AA.9.4.5 Conclusion

[2053] This specific DigitalDoors measurement system and method supports the proxy

recognition of ownership, utility, purpose, frequency, source, method, and other significant organizational behaviors and requirements. As such, DigitalDoors values information statically and creates dynamic differentials subject to attribute perspectives and reflection of internal or external systemic changes that occur over time. In addition, this method recognizes static assignment, information lifecycle management, and the dynamic revaluation of information as it relates within information warehouses and the larger strategic organizational workflow as legal, financial, and operational agendas advance to reflect new challenges. The resulting monetary value of intangibles and intangible information provides insight into the state and direction of business. Monetary information valuation is an important strategic, legal, financial, operational, and managerial concern especially when 95% of organizational value is reflected by intangible assets. Intangible valuation should not be overlooked as incompatible with conservate and traditional managerial accounting techniques. It becomes essential to assign monetary value to information so that it can be integrated within the strategic and traditional roles for managerial control to bridge this divide and reflect the fundamental changes brought by the information revolution.

AA.9.5.0 Assetitization of intangible content

[2054] Intangible assets must be discovered, recognized, measured, and expressed as a viable monetary value and accounting asset entry. When intangible assets become commoditized in this way, they can be managed and manipulated, and produced and delivered like any other tangible Assets. When 95% or more of our economy is hidden in intangible value or reflected as an estimate market valuation markup ratio over a book value, there seems to a significant failure in the representation itself and the functional accounting processes that create such semantic, economic, and functional disparity.

[2055] This disparity is clearly a business opportunity. It is, in fact, a strategic sustainable

competitive advantage that should be exploited to create wealth, a means to stave off global commoditization, and new barriers to market entry. Many assets are presumed to be public, free, available to anyone able to take and utilize them, whereas in fact they are part of elusive capital markets and complex production processes. They are not free unless the default is to ignore them, their value, and their potential to create intellectual property.

[2056] Part of this complexity is establishing how an asset becomes definable and can be separated from innate built-in ownership so that it becomes transferable to others, to other parties, to be included in wealth production processes, or sold. When assets are captured and ownership is then asserted, they become property, critical data property, intellectual capital, systems and methods, proprietary and confidential, copyrights, and court-established trade secrets. This capture process is called the assetization of an intangible, whether that intangible be a characteric asset, intellectual capital, functional organizational procedures, or formal intangible information assets like copyrights, trademarks, patents, and trade secrets.

AA.9.5.1 Definitions

[2057] The capitalizations of asset (generic) and Asset (functionally specific) is purely to differentiate a generic from the more specific formal accounting definition only within this white paper.

[2058] An asset is any resource, facility, characteristic, or trait of value. Such an asset is an innate property that is difficult to transfer to another person or party. An asset is any useful or valuable quality, person, or thing. It is an advantage or resource. For example, "she proved herself an asset to the company." It is a valuable item that is owned. It is a spy working in his or her own country and controlled by the adversary that provides results of unknown quality, quantity, and duration before exposure. An asset clearly has an intangible jeune se qua lacking clarity and definition, a clearcut

valuation or measurement basis. It is a value without value, a product or service without standardized categorization and evaluation. Such an asset is difficult to transfer or sell to another party.

[2059] An Asset is is a resource controlled by the enterprise as a result of past events and from which future economic benefits are expected to flow to the enterprise, as defined by the IFRS. The probable future benefit (NPV or net present value) involves a capacity, singly or in combination with other assets, in the case of profit-oriented enterprises, to contribute directly or indirectly to future net cash flows, and, in the case of not-for-profit organizations, to provide services. An Asset is the entity that can control access to the benefit; and, the transaction or event giving rise to the entity's right to, or control of, the benefit has already occurred. Similarly, in economics an Asset is any form in which wealth can be held and transferred. It is any item of economic value, especially that which could be converted to cash.

[2060] Examples of formal Assets are cash, securities, accounts receivable, inventory, office equipment, real estate, a car, and other property. On a balance sheet, Assets are equal to the sum of liabilities, common stock, preferred stock, and retained earnings. From an accounting perspective, Assets are divided into the following categories: current assets (cash and other liquid items), long-term assets (real estate, plant, equipment), prepaid and deferred assets (expenditures for future). Equity is a special Asset denoting the residual the value of Assets after all outstanding demands are met that defines the monetary property left to shareholders or owners; it is the remaining book value.

[2061] Realistically, an asset is a resource lacking management while Asset is under management. The process to manage an asset is not much different than creating a categorization for such a resource. This enables choices, accumulation of like or related assets, consolidation of assets necessary to a project or production process, and a structure to measure the value of an asset or its

contribution to productivity. When such an resource is assetized, this enables the asset to have monetary liquidity, handles for functional utility, and an entré into standard organizational practices.

AA.9.5.2 Conclusion

[2062] In conclusion the world must bring intangible assets into the same financial accounting management environment as more tangible items so that they can be developed, traded, and exploited for the greatest ROI or EPS and recognize that if security has any ROI, its immediate value is trivial and short but it has a very long tail that can amount to a lot of monetary valuation. More than that, the world also needs to discover and recognize potential assets to create long term value as capitalized financial assets. Capitalization of assets aside, distinction is essential in this competitive world, and the conversion from intangible asset to tangible balance sheet asset entry is a real business opportunity, with DigitalDoors assetization as a significant formal basis for conversion of intangible information into formalized property of great value.

AA.9.6.0 Attribution of monetary value to information

[2063] The big question is the worth of information. There is no one answer, no static value, and no common etiology for this. Information valuation and monetization is mostly relative rather than absolute because it is contingent on a large number of important but reflexive factors. The academic literature asserts that this value is significant and increasingly important. In fact, government commerce reports establish the market value of organizations is overwhelmingly comprised with intangibles (60 to 95% of the market value not collateralized by recognized hard tangibles).

[2064] (US Commerce Department 2005 Market Report.)\

[2065] Similarly, studies show long term capital investment in intangibles is ten times greater than investment in tangibles as businesses shift from production methods that are completely industrial

to ones with increasing informational content (US Department of State, Capital Expenditures and Investment 2007.)

[2066] There is also a serious but unresearched area of the time-value of information and its distribution of value. Specifically, information might have little current value but a long tail of value as time goes to infinity. This certainly would explain the limited return on investment (ROI) or contribution to EPS (earnings per share) for security or the general inability to forecast one. Traditional security protects infrastructure and equipment with a short half-life and little little to address information risk. Perhaps the distribution of this information risk explains the current attention to perimeter barriers and point solutions.

[2067] However, the literature is curiously lacking in detailed methods for assetizing informational value. IAS and FASB accounting policy assert the predominance of intangibles for capital production but struggle to codify the assets and processes into a balance sheet monetization. This is probably because most factors are relative rather than absolute. Factors such as volatility, productive life span, and alternative market assessment adversely undermine these efforts. Furthermore, the application, workflow, granularity, ownership, stewardship, and security of information increase the perceived variability so that conservative finance experts are unwilling to embrace assetization of intangibles. There is a firm recognition for this philosophy but only a weak commitment to date.

AA.9.6.1 Monetization

[2068] The issue is not what, but when and how. DigitalDoors answers the when – now – and the how – through granular content analysis of content, context, and concept – as part of the DigitalDoors Secure process. It is also automated so that large data warehouses can be uniformly tagged, assessed for related value, and aggregated into portfolios. A viable automated process accelerates

acceptance and commitment.

[2069] Many aspects of knowledge discovery, security, and information sharing, as well as intangible valuation is predicated by a uniform process. This process includes the steps for discovering the location, source, method, scope, depth, purpose, time, ownership and custodianship of information and creating a granular content taxonomy for granular content control. DigitalDoors expands unidimentional data tagging into a far more robust space. This etiology also provides the basis for assessing a functional value for information. It is equally true for all other types, formats, and structures of data. DigitalDoors is cognizant of the complexities to how information applies to workflow.

[2070] It is important to recognize that assessing a functional value for information is not simple nor straight-forward. Information value depends on access, usage, and opportunity. Multiple values can and do exist for the same things, all of which are contingent on perspective and applicable to different circumstances. For example, identity information is often valueless to its owner until there are problems that prevent normal financial activities, while in contrast, this information holds unlimited value to a criminal who can use it to create malicious trust scenarios. These alternative contexts create asymmetric values, zero to its true owner, unlimited to someone else leveraging it. DigitalDoors fully allows the recognition of these alternate asymmetric and competitive values, as well as their aggregation into a representative risk/benefit portfolio for an assessment of internal, external, market value, and leakage values.

[2071] Internal value is as likely as external value, as is value derived from market forces or net present value of anticipated cash flows. Static value, and even value based on a single utilization, is insufficient to establish any real data value that could include its acquisition cost, its contribution to gross profit, or its risk value when exposed to the wrong parties. For example, a trade secret defining a

production process might contribute to the gross sales volume whereas the exposure of this secret could completely erode the value to the going organization. Thus, if the contributory value is 10% of gross sales with 50% profit margin in an instrustry that values business with 6 times multiple on income, this factors into an overall 30% market value of the business. However, the leakage value is 100% market value of the business. The trade secret example has two reflexive values, one a contribution value and one a risk value established through competitive leakage; they are not equivalent and differ greatly.

[2072] DigitalDoors clearly shows that information is not a closed set and therefore has value dependent on external factors such as perspective, reflexive value, time frame, relevance to other data sets and inferences. Furthermore, information can be roughly partitioned into elements of content alone, context, and concept, each of which establishes a basis for value alone but also additional value in conjunction with these other aspects. Clearly, there is no simple, linear, or independent function to assign value.

AA.9.6.2 Contextual Conflicts

[2073] The value of information is curiously cumulative. Ten instances of a word or phrase reinforces the context or concept or its relevance to a target of interest. Perhaps one is more value than ten, or perhaps only one indicates a random or non-causal connection that diminishes relevance. Sometimes, the real point is not even mentioned specifically aloud, but is nonetheless there is concept. In direct contrast, instances are not random and information is not gathered randomly and fairly. Therefore, cumulative instances are not always inferential and can also play against the system. This is obvious with techniques used for search engine optimization (SEO) when known words or phrases of value are inserted into titles, initial paragraphs, or hidden metadata (like comments) to taint results for a conflicting agenda. This is called key or keyword poisoning and it requires defense for all

categorization and valuation operations.

[2074] The same holds true for all aspects of information valuation. This is also in evidence as incentives are used to encourage information sharing. People have a personal benefit to overstate the value of information under their control and so enhance its dissemination. This has the adverse and perverse effect of making the information less valuable. DigitalDoors has noted elsewhere that the basis for sustaining information value is through rarity, security, and granular content control.

AA.9.6.3 DigitalDoors Methodology

[2075] DigitalDoors recognizes this complexity for valuation and has defined information within semantic groupings. This includes dictionary entries (acronyms, words, and phrases), unknowns, and user-defined and situationally-dependent entries. DigitalDoors recognizes more complex semantic collections of entries that are combined into a group, a category, or a policy, akin to latent semantic content. However, this amorphous concept is definable in terms recognizable to the ordinary person as targets of interest through groups, categories, or policies. It is a practical and functional simplification that extends into a useful valuation system. It is also the basis for successful efforts to model physical observations with gaussian normal or logarithmic distribution functions.

[2076] DigitalDoors has therefore defined these simplified categories for valuation, and for each is represented by a fixed numerical value and an algebraic functional value. Both can be used simultaneously or a choice of one alone. It is a matter of setting active values. This means a numerical value must be non-zero and a functional value other than "not defined". Blank entries are explicitly replaced as "not defined" automatically. The value designation could be USD, EUR, 'info monetary units', or any other basis.

[2077] Gaussian distributions reflect natural distributions in the real world. Although the concept

is simple, the equation is a power of e and is a complex and unnatural contruct. Simple equations do not always yield simple or correct results. If this seems complex -- the process of modeling distributions, displacements, values, and rooting the results into some intangible value reality DigitalDoors will work with you to understand modeling and to craft sensible equations. Note that the process allows for notations explaining the equations that can be as functionality utilitarian. For example, one equation could be defined as the equation to reflect a high value for 8 to 12 instances and limited or no value for 1 to 7 instances or oversaturation above 12 to minimize the effects from keyword poisoning.

[2078] The current implementation of monetary value and monetary value functions is applicable to dictionary entries, such as My Group List, My Category List, My Policy List, and all sensitivity level tagging. The structure for assigning monetary value and monetary value functions is identical throughout DIGITALDOORS. It is the same structure and forms to assign value for dictionary entries, sensitivity levels, tearlines, groups, categories, and policies. It is also granular, so that values can be seen for each aspect and aggregated into valuations reflective of content, context, concept, single entries, tearlines, or complete documents.

[2079] Aggregation is supported for both raw and weighted and filtered sets, and additionally accessible for seven types of granular content control accessible by user selection of a selection bar.

[2080] For example, an internal manufacturing process would not register as a HIPAA or GLB value because it is not semantically relevant to those policies but it would register as a trade secret process and as a leakage risk. All semantic groupings are completely configurable for each organization through administrator access. Semantic titles are arbitrary with representation and context arbitrary and fully configurable to the needs of the organization. DigitalDoors does not impose weights or basis.

[2081] This granularity is applicable because a document with 10 instances of "bin Laden"

valued 0 for each "bin Laden" does not preclude a more complex semantic assessment of value when "bin Laden" appears with named terrorisim allies, certain location venues, and other linked high-value targets or material through any or all of My Group List, My Category List, or My Policy List semantic classifications. Value could be both a fixed monetary value and a function, or just one of these for each semantic classifcation. This filters simplify loading of a document with a more fine-tuned comprehension of its value. This facility is available for groups, categories, and policies, as noted previously.

AA.9.6.4 Weights

[2082] Any arbitrary data stream is not completely content, nor will it be completely concept either. Some, it also likely to reflect context. As such, all monetary values (except the raw composite values) are weighted to 100% with a split between content, context, and concept. The administrator establishes this distribution. Monetary value is aggregated through various semantic assessments, and this weighting is one more control against single factors skewing or user explicitly tainting the assetization process.

[2083] Both sets of weights are in force simultaneously so that it is difficult for anyone, or any process, or any source, or any method, to skew monetary values in any other but an unbiased and fair method.

[2084] The application of sensitivity-level weights and categorical weights is also useful for scaling results into the realm of reality. First attempts at valuing documents either generates a portfolio with a value this a fraction of any acceptable range or set values so high as to be laughable. This science requires some judicious application of values and weights to create ranges that are practical and realistic.

AA.9.6.5 Hierarcharchial Assignments and Enhancements

[2085] The large basis dictionary with nearly 1,000,000 entries is unwieldy to manage one word at time. Additionally, additions of new words, imports of foreign languages, and infusion with entries specific to an organization means an enormous amount of setup. Efforts are underway to respond to user requests to create enhancements for setting monetary values based on hierarchical relationships and sensitivity levels. This is likely to include the uniform setting of all entries reflected by a policy, category, or group, or the valuation of individual entries as a function of the multitier hierarchical overlapping tag structure.

AA.9.6.6 Conclusion

[2086] In conclusion DigitalDoors answers the when – now – and the how – for intellectual property monetization and intangible assestization through granular content analysis of content, context, and concept – as part of the Secure process. DigitalDoors provides methods to value the parts of the whole, the whole of the parts, subsets of parts, and the semantic value of relevant aggregations and inferences. There are many complex methods for measuring the value of information in both linear and more asymmetric methodologies, but the inclusive and exclusive configurable basis of a monetary value with a function offers enormous flexibility.

[2087] DigitalDoors continues to explore the finer aspects of meaning, both absolute and relative, contextually based, as a wealth generation process and also as a leakage risk assessment. Valuation is always specific to need and intent. A valuation system should be robust enough for any legitimate goal. As the need for simultaneity of both value and risk is better understood, DigitalDoors is likely to expand monetary calculations to include new outputs and reports beyond the nine currently available, perhaps including simultaneous multivariate evaluations of value for profit generation and contribution, competitive leakage, and privacy loss. DigitalDoors is also working to integrate the data

feeds and data streams into portfolios of value that reflect the aggregation of patents into portfolios and accumulation of trade secrets into a coherent body of knowledge and sustainable competitive value. It is not a merely the matter of assessing the need for recognition of the functional and financial value of intellectual property and intangible assets but also the methods and automation of these methods to effect any viable change.

AA.9.7.0 Information Lifecycle Valuation

[2088] After an intangible information resource has been discovered, classified, connected with other like information, planning for its exploitation begins with expectations for its assetization and capital generation.

Important questions include:

Table: Questions for valuation of capital generation

What type of asset should it be or classified as?

What are its acquisition, development, and legal or organizational costs while bringing the information to market?

What are anticipated revenues?

How long with the asset be productive?

What can be done to increase its net present value?

How can the return be exploited sooner rather than later?

It the intangible security-sensitive and how can it be protected?

How long can it be protected?

Can its lifespan be increased?

Is it likely that competitors will develop a better trade secret or patent and when will it erode the

market for this asset?

What happens economically if our trade secret is leaked or reverse engineered?

What happens economically if a patent is reexamined and current licensing payments are suspended?

What happens if generics do not violate a patent?

If the expected lifespan more condusive to a patent rather than a trade secret?

Can a trade secret format be maintained realistically for the expected lifespan?

Is the rate of return better as a patent, trade secret, or other intellectual property?

Is the intangible concept abstract enough to withstand attempts to infer its secrets or legitimately reverse-engineer it?

What development paths might generate spinoff patents or trademarks?

What is a reasonable assetization value?

What is the likelihood of portfolio creation?

When will the separate portfolio assets affect the value stream?

What is the valuation of a portfolio of such assets?

What is the asset NPV?

What is the asset ROI?

What is the asset contribution to EPS (earnings per share)?

What is the asset IRR?

What is the rollup of the asset value stream?

What inflection points, deviations, or cusps jeopardize the modeling?

How sensitive is the results to minor or sudden projection changes?

What happens when the accounting assumptions are restated?

How stable is the value stream or sensitive to risk shocks?

[2089] DigitalDoors has (built) an information lifecycle management valuation tool derived from its internal methodogy, the ILM Evaluator, to automate this analysis and answer these questions. This tool performs economic simulations and ongoing monitoring of intellection capital and property.

1. ILM Evaluator integrates with the DigitalDoors ILM Asset Database (as generated from DataHound search tool) to define the management characteristics of the intangible asset. This includes name, type of asset, potential utility, etc... (see S&M description of the ILM asset database). This basic template is saved.

2. A cash flow template is generated, either a generic projection or one more precisely defined by the asset database details or prior organizational experience. This projection shows initial sunk costs, current R&D costs, project R&D costs, legal expenses, plus a cash stream associated with either internal exploitation or external SPV licensing streams. See the SPV model for overhead and cost templates.

3. A time scale is created that establish a base time frame including birth date, current cycle date, decision points, lifespan. The granularity of the time scale is set (as days, weeks, months, or years.)

4. Overall range and scale values can be set to establish the overall shape, profitability, relative monetary valuations to the simulation.

5. The user now manipulates the model data control values. They include... These are stored in a grid of financial values that is modifiable and dynamic. The grid is basically a speadsheet of time/value financial data. Model attributes are defined separate by menus, lists, and settings. These include lifespan, time until income, type of asset, market disruptions and events, sensitivity of resize settings, valuation scale basis, change sensitivity level .... The user also can manipulate the grid values mechanically by a mouse, touchscreen, or similar input device by deflecting the cusp points or resizing the physical

representations of the financial values in the grid. In order words, the user can change the value in the grid by elongating or shortening the time value bar graph associated with the particular grid entry with hand pointer. Recalculation is immediate (or can be disabled) for NPV, rolled up cash value, IRR, ROI, breakeven, etc...

6. The user can insert events that represent cusps, inflection points, and unexpected market or legal deviations. This can include loss of asset, exposure of secrecy, patent reexamination, patent revocation, court decisions, adverse market events, depreciation or depletion of asset value, asset value revaluation, total asset loss, etc from a dropdown list of standard and user-defined events. Effects of changes to asset valuation are reflected by immediate recalculations. This values are reflected in the grid as well as the graphs and color-coding of the bar graphs and rolled up current asset valuation.

7. Data values are saved. Actual observation data can be added to the grid to contract forecast with actual experience, generate deviations from expectations, and forecast future values based on adjustments to the initial forecast.

8. Deviations can be generated to use as accounting adjustments to restate financial reports.

9. What if scenarios can be applied to base forecast, observational data, or forecast from observational data, to generate alternate outcomes. These results show deviation from base forecast, sensitivity of the model to shocks, or several different scenaries can be combined to compare against the base forecast to generate alpha and beta values for establishing a statistical integrity and overall risk of the simulation.

10. A risk profile of likelihood for each deviation scenario can be constructed and applied against the ROI, NPV, IRR, etc for scenario to establish a weighted mathematical forecast.

11. Actual accounting data, as generated, can be automatically matched to the approprate model to generate actual deviations from forecast with reports showing the observed change from forecast. These

reports can be routed as defined.

12. Preclude and indicate false assertions or workflow choices, such the attempts to assert a Trade secret after filing a patent.

13. Lockout assertion of trade secret where legal basis precludes secrecy (as with drug development and approvals)

14. Invoke DataHound or manual efforts to explore for prior art locally or externally (offline).

A. Multiple simulations can be overlapped and aggregated into a portfolio showing cumulative valuations, aggregated portfolio risks, long-term expenses and asset valuation.

AA.9.8.0 Security Gaps: Binding Infrastructure with Process with User with Data .The need for a new intangibles production model

[2090] Traditional security utilizes firewalls, logins, role limitations, multifactor biometrics, digital signatures, and encryption. They are expression of perimeter control. If these technologies would work, security would be no problem. If variations on the themes of existing techniques and compilations of existing technology would work, security would not be an issue. You would not need this white paper.

[2091] Apparently, security is a problem. We are unable to respond to the exponential growth in security lapses and exploits. There are so many of them, so many more of them as we shift to an intangible economy, and an explosion of combinations as the complexity of infrastructure and process increases while the amount of users and data grows exponentially. A reactive response is too little too late while falling further behind the known persisting problems. Gaps pop up undermining newer business models and become known even as preexisting flaws are exploited. Gaps are everywhere, and because they are all seem individualized and unique with no threat of commonality, security implementation is specific rather than patterned on some fundamental overarching paradigm.

[2092] Infrastructure is exploited. Process is exploited. Users are exploited. The data is an input and derivative of the infrastructure, process, and users, and it gets exploited. We patch infrastructure and get a data leak, user exposure, or a process failure. Security is applied to each new problem separately with an every accelerating lack of effectiveness. There is no single model for security that currently binds together the infrastructure with the process and users with data. There are massive and growing gaps in our attempts to plug the perimeter, fix the flaws in the process, protect users from fraud of many types, and retain the data that is making the information revolution possible.

[2093] Here are some reasons why security is broken. You have probably heard that security is a process and not a technology. It is a rationalization why security fails and why more security effort is needed. This is not really a legitimate explanation but more of an excuse for our failure to recognize an overarching paradigm. Realistically, the classical (Marxist) production model that forms the basis for wealth creation has changed. The agricultural model has given way to the industrial one, and even that has changed into the information model. The inputs, production methods, and outputs that represent wealth have changed and become more concentrated and amorphous. Three points seem critical to this social revolution. First, the production model is different. Second, the inputs and outputs are different. Third, wealth and its format is different.

[2094] These differences alter the social structure and the methods for producing (and retaining) wealth. They also alter the infrastructure required to protect this wealth. Security is no longer about wealth on the hoof, in the ground, or in a warehouse. Rather, wealth is significantly different because it is represented by the intangible property of money (as a proxy for tangibles) and increasingly by information. Information is significantly different because this intangible property can be used without even using it up, exchanging the principle basis, or trading it.

[2095] The whole and sum of the parts of information are not the same, leading to opportunities for aggregation, inference, and interaction, creating derivatives in wealth likely greater than the whole or the sum of the parts. Information alters the multiplier effect of the production model, and security for infrastructure or the process does not reflect the absolute need for security for these parts. Information can be used, multiplied, or copied so long as the legal ownership, custodianship, and uniqueness is preserved. Like all things intangible, protection for the status, availability, reputation, and accessibility is everything. This is an expansion in information that requires an expansion in protective security, not addressable through traditional perimeter-based methods.

[2096] Previously, security was legitimate when the range, the livestock on the range, the produce in the fields, or the processed product in the warehouse were protected within that perimeter. Perimeter security worked with gaps addressed by brands and stamps for ownership. There were just a few gaps in the perimeter, and there were resolved by intuitive ad hoc one-for-one solutions. For example, stamps for items produced as commodities or in enough multiples are differentiated by serial numbers, and serial numbers provide controls when product is distributed through this perimeter. When the perimeter was not sufficient to prevent insertion of virus inside the perimeter, antivirus tools augmented the gap through the perimeter. However, these gaps are not few, and they are not staying at a few. These gaps overwhelm any stability of perimeter security or a rational expectation of reestablishing a viable security perimeter. Production no longer occurs within any semblance of a sustainable perimeter. Just-in-time production, complex products and services, subassembly through a distribution organization of keretsu, and outsourced raw materials create an infrastructure exposed to threats without rational expectation of an implementable perimeter.\

[2097] The production process itself has changed while inputs and outputs are changed. The

production process is no longer monolithic, because the efficiency of the process is driven by multiplier effects and the inputs are catalysts which might not even be used up. These multiplier effects and catalysts are intangibles, generally information in the form of restricted knowledge. For example, the information necessary to create marquis-label car is not used on other cars and can be used again and again without substantial loss. Were this information or production process to be stolen and used by another manufacturer, this would constitute a breach in protection for the status, availability, reputation, and accessibility of this intangible information. This is called infringement or theft of trade secrets, where the marquis, trade secrets, knowledge, multiplier effects, catalysts, and data are all valuable forms of intangible information.

[2098] The production process includes such modern services as extended credit card credit transactions to make a purchase at a retail store or over the Internet at an intangible store. The purchase could be tangible, or increasingly the purchase could itself be a service to perform some result or simply be the delivery of a software application, necessary data, or a MP3 song. The delivery of intangibles from an intangible store, produced by intangible manufacturing processes, paid for with an intangible service payment creates the very difficult problem of securing an infrastructure with embedded processes with unseen merchants and buyers using data that is intangible. There is not even a good model to define this infrastructure or string of loosely-coupled processes, let alone provide a unified view to effectively espouse a security model. No wonder why there are gaps in the infrastructure, gaps in the process, and scurrilous breaches that leak data.

[2099] Because this production process in this new information revolution is not easily or even remotely confined within a perimeter, a process, to a defined set of vetted users, or even within a consistent set of data, the traditional security methods with perimeters, logins, role limitations,

multifactor biometrics, digital signatures, and encryption simply do not work. There are just too many gaps and every variation is different enough to preclude a one-for-one event security solution. There is no location similarity, no repetitions, no exact processes that are consistent from event to event. Furthermore, compilations of events create unanticipated channels for losses of both tangible and intangible value. Not only are patches ineffective against the purported gaps, frequently these patches combine to create new risks and threats or undermine the functional integrity of the infrastructure and break processes.

[2100] However, these processes and the encompassing infrastructure have self-similarity. The patterns are necessary for automation and the expansion of the evolution in this information revolution. Without these patterns, there could be no self-service, no automation, no computerization. These patterns are becoming increasingly complication and interrelated -- and the recognition of these complexities -- also open unanticipated channels for both losses of both tangible and intangible value.

[2101] This is a situation with exponential growth in the volume of business, exponential growth in the complexity for manufacturing of the product or service, exponential growth in the delivery process, and exponential growth in the complexity of the intangible infrastructure.

[2102] The relevant patterns are bigger than retail versus manufacturing versus financial process versus service. The patterns are production processes reflected in the simplified classical production model. Retail is a production process. Manufacturing is a production process. Financial creation is a production process. Service is a production process, whether washing a car, delivering a package, or fixing a computer. Inputs of labor, capital, and knowledge drive a manufacturing process creating the outputs of products or services. Knowledge is a multiplier on the efficiency of the manufacturing process. Often this multiplier is a trade secret, a patent, or specialized knowledge that sustains a

competitive advantage reflected in higher prices, lower input costs, lower production costs and efficiencies (including byproducts), or a monopolistic exclusivity on the outputs.

[2103] This production model can be reflected by a workflow, drafted as a blueprint, and reduced into a data set. The production model becomes the expression of the data set. Variations on this theme are variations in the expression of the same data set. Of course, we do not have a 'simple' production model; it is extended with information inputs as source, the outputs of information as product, service, or byproduct (and waste), and the efficiency of the production process is moderated by the multiplier effects of knowledge-driven enhancements. The tangible aspects of the 'simple' production model are almost mute, and soon irrelevant. Only data alone will represent a differentiating and critical component of the production process.

[2104] The fundamental pattern becomes the data. Effective security is applied against that data rather than the expressions of the production model. The infrastructure and process, the users and the incidental derivative data are reflections of the fundamental data pattern. Because there are so many variations, so many different expressions, we are reacting to an infinite expression of these security flaws. If we see the pattern as the data and apply security to that pattern rather than the infinite expressions of the data, we are changing the security paradigm and actually have a chance to plug the security gaps with a unified method. Infrastructure, process, users, and the data we process for the users, the process and within the infrastructure are incidental to the data patterns that are relevant to a paradigm security model.

[2105] Security can be designed to respond to these new patterns of complexity, the compilation of events, and the preexisting defects in the infrastructure and process. Reactive response is too little too late.

[2106] We are not producing commodities that are identical, delivered identically, for identical people, and made from identical inputs. We are producing within a bigger structure of patterns, and it is these patterns that provide a hope for defining and implementing security. For all the apparent differences, there are important self-similar patterns which enable a new model for security.

[2107] The self-similarity of the infrastructure, the process, the users, and the production inputs is the intangible data itself. The information revolution is about the information, the data, not the physical things that enable it or that are expressed from the data. If we want security for this revolution, we must protect its basis. That is information at every level. Infrastructure has a basis in data. Process has a basis in data. Users are represented by data. It is all data. While we sit at keyboard in front of a computer and display that connects to a network and the larger Internet with virtualized processes and storage, every tangible entity is an expression of the data path and geared to the creation, differentiation, and retention of that data.

[2108] Everything is represented by designs or data. Even the designs are expressions of data. Even a tangible product like Lexus is a reflection of a Toyota vehicle with a different marquis that is expressed from data. The production process is a controlled just-in-time workflow where data indicates the types of vehicles, its color, and its options. The Lexus and Toyota vehicles are different at an intangible level, one is high end and the other not. Yet they are the same and produced from differentiable patterns based on different data. They are self-similar and routed from the data-driven production of inputs of material and intangibles into parts that are controlled by data into larger subassemblies, assembled through information-driven process.

[2109] The transactional and delivery processes are self-similar too, even for products and services as diverse as vehicles and songs. Sales of a vehicle requires payment and/or credit transfer.

Delivery of vehicle -- and it matters not whether it is a Lexus or a Toyota -- is based on a target delivery location. Effecting that delivery could well require a train, a ship, and a trailer truck. A communications channel arranges the separate physical steps. Sales of an MP3 song requires payment and/or credit transfer. Delivering that MP3 song also requires a target delivery location, although delivery is direct by a communications channel but it could entail creation of a personalized CD or DVD and delivery by FedEx. Scope, scale, and legalities separate these two transactional and delivery processes without breaking the self-similar model.

[2110] However, the point is that the intangible reduction of the transactional and delivery process is data. Payment and target delivery data is one type of data and it is self-similar to design data, production data, and all other types of data. Structured and unstructured data are the same. Attributes distinguish one from another, but the recognition that all is data, and all data is a reflection of data establishes a baseline for applying unified security proactively and universally. Dissimilarities in data structure, format, encoding, presentation, markup with tags and XML, or impression into a relational database are details of data not relevant to perceiving data as the fundamental building block for security. Is not that data is different – of course it is different – but data is the uniform target for applying new methods for security and at the data level itself rather than at any levels resulting from the expression of the data. It becomes infinitely harder to protect a blueprint as an expression of data rather than the data itself that becomes the blueprint. It becomes infinitely hard to protect a manufacturing process defined as a trade secret rather the data that expresses the trade secret that enables the manufacturing process.

[2111] In other words, security is not symmetric. We can protect the data. We can not protect the expression of the data once expressed in all its myriad forms. We can redact the expression into data and protect the data. All traditional security efforts have sought to protect the expression of the data, the

infrastructure, the equiment, the computers, the communications, the workflows, the process, but the very variable and infinite variety effectively precludes security. Obviously perimeter security and point solutions provide protection one-for-one against designated threats. These solutions might even be simple, cost-effective, intuitive, practical, and desirable as a first-level defense. However, they do not scale to existing expression of risk and threats or the accelerating level of threats expression with integration, increased complexity, and the fundamental loss of a definitive perimeter. Security can be provided to the data but absolutely to the reflections of the data, not to expressions of the data, and not to the infinite variability of its process.

[2112] For all the apparent differences, there are important self-similar patterns. All these patterns are expressions of data. Security is failing at the tangible level of traditional security utilizes perimeters, logins, role limitations, multifactor biometrics, digital signatures, and encryption. The data is seen as the subject of actions rather than as the fundamental substrate. Data is not derivative. Data is fundamental. Data is where security starts. Data is where the new security model starts. Protect the data to protect the expression of data, but do not expect to protect the expression of the data once expressed from the data.

AA.9.9.0 Intangible Trade secrets

[2113] Information security is usually abstract, a potential, or something that happens to other people. Information security is often a stepchild to the perceived reality of getting the job done. Potential is not reality, just the possibility that something can go wrong, an abstraction without existence until something goes wrong and it is too late because such losses are irretrievable. As a result, information security is not taken seriously, is not part of the budget, and is not perceived for the competitive advantage that it really provides.

[2114] There is a powerful example worth considering, the 120-year history of Coca Cola and its perpetual battle in information security to preserve a trade secret, the recipe for the syrup. Merchandise 7x is the purported "secret ingredient" in Coca-Cola and has apparently remained a secret since its formulation in 1886. Coca Cola is not an abstract potential; this company exists primarily because of the value of this intellectual property and the ongoing security of that recipe, and represents the growing value of content in modern products and services. The continuity of the Coca Cola is dependent on the security of the intellectual property, and in particular, the control over that information. It is a competitive advantage. This is an example of the value of information for continuity, and explains why we cannot be complacent over security, in particular the security of logical and digital information.

AA.9.9.1 Background

[2115] The Coca-Cola formula is The Coca-Cola Company's secret recipe for Coca-Cola. As a publicity marketing strategy started by David W. Woodruff, the company presents the formula as a closely held trade secret known only to a few employees.

[2116] Apparently, the secret defined the raw ingredients, the method of production, and the ratio of ingredients that define the characteristic flavor. Diet and other special products vary slightly in taste, as do copycats. Although reverse engineering and taste tests over the years have evolved, the uniqueness of Coca Cola is still perceived by the marketplace and customers worldwide.

[2117] Regardless of the truth, and professing that extent recipes are not the real thing, is sufficient to claim trade secret when other efforts are made to preserve the secrecy of that trade secret. The trade secret is part of the culture and represents a clear signal to other organizations that information and secret recipes or systems and methods not published or patented are clearly an integral component to most products and services and deserve the type of protection honed by Coca Cola.

[2118] Efforts to protect the legality as well as the actuality of the trade secret go back 120 years. Only two persons at a time know the formula. The written version is stored in a safety deposit vault at the Trust Co. Bank in Atlanta GA.

[2119] Secret formula, trademarks, names, and market protection separate from competitors Pepsi and Royal Crown, as well as house-label cola knock-offs.

[2120] The formula is mixed into a syrup, a highly concentrated mixture of flavors, which is mixed only in five production facilities throughout the world. This is then distributed to local bottling companies to mix with carbonated water and other minor ingredients, to ensure that the secret is kept very tight.

AA.10.0.0 Introduction to DigitalDoors Dictionary

[2121] The definitions herein are used throughout DigitalDoors patents, white papers, marketing literature, and user manuals. The definitions describe DigitalDoors meaning, usage, and intent. These definitions differentiate the DigitalDoors sprocesses from any prior, similar, or developing art. Lower-case terms represent generic usage whereas capitalized ones denote formal glossary references to common expressions.

[2122] In many cases, the recognition of information as an asset and the processes to discover, classify, and control it requires an entirely new vocabulary with explicit definitions, including some new words. Other researchers are discovering the possibility of granular content control for security applications, although use similar words and phrases but with distinctly different meanings. Because generic definitions or even public ones, like Wikipedia, do not always convey operative and technical preciseness, these definitions are included for precision and relevance to knowledge management, data security, and information sharing. The DigitalDoors literature and our personnel deliberately use these

terms with these meanings to establish a consistent lexicon. These meanings do change and they are updated as needed.

keywords: glossary, dictionary, lexicon, classification, categorization, taxonomy, ontology, definition, term, meaning, explanation, index, key, keyword, surrogate key, content,

concept, context, semantic content, information control, granular content control.

Access - The process or event to get into a computer system, service, or network, by dialing a phone number, providing a password or token, logging on a network, such as the Internet, to retrieve data or use services.

AA.10.1.0 DigitalDoors Dictionary

[2123] Below each keywors of the dictionary appears the definition:

access control

[2124] A system or network function that determines whether the user of a particular resource or process is permitted to employ that resource or process.

Adjacency- list model

[2125] A method for showing the expanded drill-down hierarchies in generally normalized data that consisted of a column for the primary reference key and another column for the secondary data in the relationship. It is a direct implementation in a table of the Adjacency List mode of a graph, but note that this structure is not normalized. This structure is relevant for establishing social networking relatedness. See also Betweenness, Closeness, and Connectedness.

Aggregation

[2126] The process of pooling information or data mining to construct a profile that can identify an individual, reveal hidden sources and methods, or otherwise breach the integrity of categorical

security by discovering other inherent informational taxonomies, ontologies, or contextual associations.

air gap

[2127] A method for security enabled by lack of connectivity, lack of workflow integration, or the sequestration of sensitive information from normal business processes. It is the literal space created surrounding the perimeter of the resources of interest preventing functional access; this space is not bridged with communication channels and access is ad hoc but scrupulously controlled. See also silo.

anchor

[2128] 1. An indicator of an object so that its position relative to some other object remains the same during repagination. 2. The name of the HTML code used to specify a reference to another document.

anonymity

[2129] A method, desire, or process to maintain personal, business, or organizational identity without releasing it to others.

anonymized data

[2130] 1. Data stripped of all personally identifiable information (PII) by safe harbor categorization. 2. Clinical, provider, insurance, or operational data pooled, aggregated, or otherwise accumulated without personal information to render the data still useful but supposedly without sufficient designations to reidentify the individuals represented. This technology is often used to protect the identities of participants in medical studies or other research, but nevertheless this often fails to anticipate the power of statistical methods, aggregation, inference, and interaction to pierce the veil of individual privacy.

annualized loss expectancy

[2131] An accounting and actuarial evaluation used to consider the forecast damage from adverse events like catastrophes, thefts, data breaches, PII exposure with fines and legal expenses, and mitigation or remediation required to stop the damage, compensate those effected, and restore production. Abbreviated as ALE.

approximation measurement

[2132] An evaluation used to consider possibilities. See also qualitative measurement and quantitative measurement.

authenticity

[2133] The process of validating that a person, document, process, or other workflow is originated from whom and where it says it is from, that the content is valid, and that the content or person has not been fraudulently counterfeited.

authorization

[2134] The ability of a user to access a feature of a software application, system, network, or other infrastructure.

authentication

[2135] 1. A security procedure that verifies that a user is he who identifies himself as, or that data or a digital signature is what it is it purports to be. 2. The process of validating that a user or process is who or what that user or process provides as a name and/or purpose.

authorization

[2136] The process of allowing access to resources only to those permitted to use them or otherwise granted authority to them.

authorization escalation

[2137] The misuse of access controls, scope, range, mislabeling, and classification biases extend access permissions beyond intent and expectation.

back door

[2138] A means of gaining access to a program or system by bypassing its security controls. Programmers often build back doors into systems under development so that they can fix bugs. If the back door becomes known to anyone other than the programmer, or if it is not removed before the software is released, it becomes a security risk. Also called trap door.

base redacted document

See Objective Security Compliance for Universal Distribution.

Bayesian inference

The process of deducing new statistical facts from existing statistical facts without a priori knowledge.

Bell-LaPadula Model

[2139] An academic model for enforcing access control in government and military. The model is based around the idea of mandatory access control and is widely referenced and employed in the design and implementation of secured network access. Abbreviated as either BLP or BLM.

Bell-LaPadula Model Compliance

[2140] Strict adherence to the BLP or BLM for need-to-know, sensitivity levels, and compartmentalization.

betweenness

[2141] The measure of centrality to the number of standard, direct formal, or informal lines of communication among a group of people, events, resources, or other arbitrary objects. This structure is relevant for establishing social networking relatedness. See also Closeness and Connectedness.

binary information delivery

[2142] The process of delivery partial information on separate channels and/or by separate delivery methods to mitigate the risk of interception, security failure, or improper access.

biometric authentication

[2143] The process of confirming identity while using various techniques of voiceprint, fingerprint, retinal scan, or other physical test.

biometric identification

[2144] Validation of a person by a physical or behavioral characteristic (such as the way they sign their name, their fingerprint or the marks on the iris of their eye).

black list

[2145] A register of entities that are not acceptable and are being denied a particular privilege, service, or being categorized for attention. A black list might also include entities of interest or entities requiring additional attention that are not necessary 'unsafe.' Contrast to Safe List and white list. See also gray list.

cascade of authentication

[2146] A multi-tier process of login, authentication, biometrics authentication, challenges, and identification as friend or foe on a timed, random, or periodic basis.

category

[2147] The assessment and assignment of one semantic description for data item. See also Label and Tag.

categorization

[2148] The process of assessment and assignment of (one or more) descriptions for (one or more)

data item(s). The application of categorization with descriptive results is a synonym for applying a tag or a label to a data item, or otherwise establishing an entry for taxonomy or ontology. Synonym for classification.

certificate

[2149] Any symbol or token providing proof of attendance, completion, experience, or performance and this is sometimes used as a method to authenticate integrity or authorize access. A certificate has limited attributes or properties unless some system is enabled to give credence to this symbol or token.

certificate-based encryption

[2150] A system in which a certificate authority uses ID-based cryptography to produce a certificate. This system gives the users both implicit and explicit certification, the certificate can be used as a conventional certificate (for signatures, etc.), but also implicitly for the purpose of encryption, which is usually symmetric for performance reasons. Microsoft has built this functionality into SQL 2005.

certification

The process of obtaining a certificate.

checksum compression

See Winnowing.

chaffing

A process of inserting nonsense within data to obscure its meaning.

chain of custody

[2151] Generally a legal term defining a process that establishes each person having custody of the evidence, thereby establishing continuity of possession, and proof of integrity of the evidence

collected (Osterburg & Ward, 1992). However, the security extrapolation of this terminology is that data has been under control and the control can be traced and audited. This refers to the chronological documentation, and/or paper trail, showing the seizure, custody, control, transfer, analysis, and disposition of evidence, physical or electronic. Because evidence can be used in court to convict persons of crimes, it must be handled in a scrupulously careful manner to avoid later allegations of tampering or misconduct which can compromise the case of the prosecution toward acquittal or to overturning a guilty verdict upon appeal. The idea behind recoding the chain of custody is to establish that the alleged evidence is in fact related to the alleged crime, rather than, for example, having been planted fraudulently to make someone appear guilty.

Chaos

[2152] A state of disorder, infinite space, or void. It is often misused to imply a situation of that is beyond human control, as from the Greek Χάος, rather than the true mathematical meaning of emptiness. Note that chaos and randomness are not synonyms but rather mathematically related.

Cipher

[2153] The result of any method of cloaking information by means of word, symbol, compaction, substitution, replacement, reordering, or mathematical encryption.

Ciphering

[2154] The processing of data under a secret key, using a slide rule, substitution, or word replacement to differentiate it from encryption, in such a way that the original data can only be determined by a recipient in possession of a secret key. The application of a specific algorithm to data so as to alter the appearance of the data and make it incomprehensible to those who might attempt to 'steal' the information.

Classification

[2155] The process of assessment and assignment of (one or more) descriptions for (one or more) data item(s). The application of categorization with descriptive results is a synonym for applying a tag or a label to a data item, or otherwise establishing an entry for taxonomy or ontology. Synonym for categorization.

Clearance

[2156] The categorization of information by sensitivity level, compartment, group, user, action, mission, and other descriptive or functional assignments.

Closeness

[2157] The measure of path length or frequency of communication given the number of standard, direct formal, or informal lines of communication among a group of people, events, resources, or other arbitrary objects. This structure is relevant for establishing social networking relatedness. See also betweenness and connectedness.

Command execution

[2158] The command line, parameterization of commands or passed values in functions, or insertion of values into applications, process, and workflow creating an inevitable expression of the security risk.

Compartmentalization

[2159] The storage of data in isolated sets.

Compliance

[2160] The process applied to systems or departments at corporations and public agencies to ensure that personnel are aware of and take steps to comply with relevant laws and regulations. See also

malicious compliance.

Concept

[2161] An abstraction or general idea inferred or derived from specific instances of information ased on a cognitive knowledge of the context and content. This can include business rules, formats, constructs, structures, abstractions, attributes, properties, categories, labels and tags, notations, metadata, and/or the definition of a process or computer application to define information usage and intent. A concept can also be a partial-formed idea that is still incomplete.

Conditional tagging

[2162] A theory and practice of tagging documents (not used by DIGITALDOORS) based on the sensitivity level content and the one-time upfront assessment of the risk of that document subject to the perceived intent and purpose. This does not consider hidden risks, covert channels, or the context and concept when such information is repurposed.

Conditional information sharing

[2163] DigitalDoors enables the delivery of a baseline document, a generic redaction or universal redaction, that contains no sensitive value and can reconstitute that document in fine-grained steps up to the original document.

Confederated search

[2164] The usage of multiple web-based search engines based on keywords with results combined into a single unified report. Also federated search and metasearch.

Confidentiality

[2165] An expectation and workflow process that ensures that the content of a transmitted message cannot be determined except by the possessor of a key associated with the message, even if the

transmitted bits are observed.

Connectedness

[2166] The number of standard, direct formal, or informal lines of communication among a group of people, events, resources, or other arbitrary objects. This structure is relevant for establishing social networking relatedness. See also betweenness and closeness

Connect-the-dots

[2167] The inspiration that cross-domain inter-agency information sharing (in contrast to endemic departmental isolation and retention of independently-acquired intelligence) might have allowed key personal assembling daily terrorism threat status reports to aggregate and inference the potential for 9/11 attacks before they occurred.

Content

[2168] Anything that fills a newspaper, magazine, data stream, document, video, audio, multimedia presentation, or web site. Content is synonymous with data, but content also includes material represented by labels, tags, structure, metadata, references, links, applets, notations, and out-of-band controls.

Content-addressed storage

[2169] The methods to locate information by content rather than by location or file name using alternative information addressing schemes.

Context

[2170] An abstraction of information to include discourse that surrounds a language unit and helps to determine its interpretation. It is an abstraction of circumstance that sets the facts or interpretations that surround a situation or event. It is also the environment, purpose, usage, or

application of information to a particular scenario or problem set. The non-entropic value of data within a signal, message, file, or other data structure.

Contraction

[2171] The process of removing positional markers, widths, punctuation, and other silent indications of DigitalDoors granular extraction. For example, when key information is extracted, the existence of the missing extractions and tear lines are hidden with a reformatting of the data stream to suppress this information.

Controlled release

[2172] The process of .reconstituting granular content items based on rules in order to enable process continuity while conforming to security, risk management, and compliance.

Convergence

[2173] The act, condition, quality, or fact of approaching a limit, such as a point or limit.

Counterfeiting

[2174] The act of making and distributing an imitation that is made usually with the intent to deceptively represent its content or origins. The word counterfeit most frequently describes forged currency or documents, but can also describe clothing, software, pharmaceuticals, watches, or more recently, cars and motorcycles, especially when this results in patent infringement or trademark infringement. Any analog or digital materials, software, data, signatures, images, and data streams are candidates for such imitation.

Credential

[2175] See certificate. The differences between certificate and credential are both contingent on some system given credence to these symbols or tokens.

Critical and important {data}

[2176] Information defined by the system as data, which, when removed, 1.) render the digital document, digital data stream, or digital environment useless to users, processes, or workflow, 2.) create a substantial loss of economic, functional, or strategic value essential for continuing the vital missions. The crown Jewels. See data element.

Custodianship

[2177] A person, organization, or process that has charge of something, generally without ownership or legal stewardship, and is tasked with is responsible use and maintenance of that resource. See also stewardship.

Cyberwar

Also cyberwarfare. See information warfare.

Data

[2178] Any stream or structured format of information that includes data files, documents, objects, records, record sets, databases, transactions, log histories, images, video, sound, data packets and network frames, E-mail streams, process streams, and other data structures. Although this Latin form is plural, datum is really a single item or element. Realistically, a single bit is data. The difference between data and information is generally defined by the Data, Information, Knowledge, and Wisdom (DIKW) Information Hierarchy that defines data as raw symbols lacking definitive context whereas information is data defined by a specific implied practical context.

[2179] Data, Information, Knowledge, and Wisdom (DIKW) Information Hierarchy Russell Ackoff, a systems theorist and professor of organizational change, stated that the content of the human mind can be classified into five categories:

Data: symbols

[2180] Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions

[2181] Knowledge: application of data and information; answers "how" questions

Understanding: appreciation of "why"

[2182] Wisdom: evaluated understanding. Professor Ackoff indicates that the first four categories relate to the past; they deal with what has been or what is known. Only the fifth category, wisdom, deals with the future because it incorporates vision, innovation, and design. People can create the future wth wisdom rather than just grasp the present and past. Achieving wisdom isn't easy; people must advance successively through the prior categories.

Data-centric

[2183] A process or workflow structured to the value of the sustaining information. In the case of data-centric security, the focus of protection is specifically to the information within the process or workflow rather than to the infrastructure. Although infrastructure is prone to disruptive, surreptitious, and denial of service attacks, the data-centric security model minimize leaks and breaches of the data and emphasizes the protection of the crown jewels. Also called information-centric.

Data concept

See concept.

Data content

See content.

Data context

See context.

Data element

[2184] A selected subset of information that can be assigned a metatag. Usually, a data element will be the smallest piece of information that can be removed from context for establishing a granular content security.

Data mining

[2185] A class of database applications that look for patterns in a group of data that can be used to predict future behavior from the nontrivial extraction of implicit, previously unknown, and potentially useful information. True data mining software doesn't just change the presentation, but actually discovers previously unknown relationships among the data.

Data-to-data interaction

[2186] The formal or informal process of creating new information or formulating new hypotheses through the accumulation of unstructured and structured data often through the reorientation, transformation, or normalization of different types of data into unified formats usually in the presence of a specific trigger or catalyst. A good example from the movie "Oceans Twelve" is the overlaying of building blueprints with subway transit maps by a thief seeking to discover weaknesses in a bank vault. This is specifically different from data mining because the data is not derived from like databases. This is specifically different from aggregation because the superimposition or collection of such information does not influence the revelation. This is also specifically different from inference due to the unlikely overlap or linkage references between the data sources. This notion is analogous to gene-to-gene interaction in the presence of active proteins and complementary binding receptor sites that allow genes from the same, similar or even different biological strains to interact to cause sickness, infection defenses, or functional natural adaptation.

Data Encryption Standard

[2187] An encryption/decryption algorithm defined in FIPS Publication 46. The standard cryptographic algorithm developed by the National Institute of Standards and Technology. Abbreviated as DES.

Data dispersion

[2188] A method for data storage and delivery where the constituents of the whole are contextually disassociated into elements and randomized and/or stored in separate locations requiring a directory or recovery script to restore into initial form. The scattered or diffused storage of critical and important information renders that information valueless without context, nearly impossible to match together correctly, and difficult to locate without a directory. In addition, the scattered or diffuse delivery (often by separate channels and types of delivery methods) of that same data for reconstitution protects the constituents of the whole from useful interception and contextual understanding.

Data evolution

[2189] The normal atrophy of stored data and also the accumulated and aggregation of new data to alter the value or context of existing stored data.

Data governance

[2190] A formal program conforming to legal requirements that includes a governing body or council, a defined set of procedures, and a plan to execute those procedures defining the owners or custodians of the data assets in the enterprise. A policy must be developed that specifies who is accountable for various portions or aspects of the data, including its accuracy, accessibility, consistency, completeness, and updating. Processes must be defined concerning how the data is to be stored, archived, backed up, and protected from mishaps, theft, or attack. A set of standards and procedures

must be developed that defines how the data is to be used by authorized personnel. Finally, a set of controls and audit procedures must be put into place that ensures ongoing compliance with government regulations.

Data hiding

[2191] Any process used to embed data, such as copyrights and messages, within other forms of data such as images, sounds, and video with an imperceptible amount of carrier signal degradation. Data hiding also has another meaning in that normally visible data is truncated, redacted, masked over, masked with substitution, or removed. See also steganography, data masking, data obscurity, data truncation, and redaction.

Data masking

[2192] A type of information redaction in which the meaning, value, or utility of a data stream is altered by substituting some of the value with alternate characters. For example, a credit card number like 4256-3340-5687-1220 could be rendered as ****-****-****-1220 or as 4256-3340-5687-****. Downsides to data masking is that the masked data becomes useful as a key or search basis for reidentifying the unmasked value in full. Also, if you notice, the statistical likelihood of reidentifying by brute testing is only enhance to a small solvable set requiring seconds of real computer time. Note that masking can effect a substitution for the original field or the overlay of a placeholders to hide and cover over the original field. See also De-identification.

Data mining

See data-mining.

Data obfuscation

[2193] A technique for securing information by encryption, removal, extraction, and dispersion

of some or part of a data field or data stream.

Data obscurity

[2194] A questionable technique for securing information by masking, truncating, or hiding some or part of a data field or data stream.

Data partition

[2195] The separation of storage locations (local or remote) into a differentiable references, files, databases, file systems, or directories, hives, etc.

Data spill

[2196] The accidental transmission or display of PII, intangible information, or trade secrets to a third-party.

Data sharing

[2197] A security method based on the concept of the dual keys for the bank safety vault where data is distributed among various people or processes thereby precluding any one party or even most of the parties benefiting from the information without full cooperation of all parties. See also Information Sharing.

Data stewardship

[2198] The non-legal functional formalization of the roles, accountability, and responsibilities for the management of data. This includes the overall management of the availability, usability, integrity, distribution, reuse, and security of the data employed within an agency. This is not to be confused with the legal responsibility afforded by a formal stewardship.

Data stream

[2199] A generic way to describe a single or compound unit of information regardless of format,

structure, delivery method, or storage medium. Our supposition is that everything can be described and defined by a data stream and is a logical representation of the physical. For example, DNA is a data stream that represents the RNA cell manufacturing process. Information, applications, and workflow of all types are delivered as a data stream, and essentially are just the expressions of a data stream.

[2200] Essentially, all machinery, whether cells or cell-driven processes, software, and even hardware are essentially manifestations (literally "expressions") of a data stream. This complicates security issues, but also provides a unitary vision for expressing security as the protection of a data stream and provides a new paradigm for its protection that transcends the fail view of process and perimeter security methods.

Data structure

[2201] A specialized format for organizing and storing data. General data structure types include arrays, fields, files, records, hashes, tables, relational indexes, the hierarchical tree, the document object model, and so on. Any data structure is designed to organize data to suit a specific purpose so that it can be accessed and understood as to its expected field format, purpose, range valuation, source, validity, and access control.

Data tainting

[2202] 1. A bit within each character byte used to indicate whether the byte is valid, like a checksum. 2. The subversion of a data set, data table, or entire database with false, misleading, or misinformation in locations known to the owner to track misuse, copying, or distribution.

Data truncation

[2203] A type of information redaction in which the meaning, value, or utility of a data stream is altered by dropping leading or trailing values. The choice of truncation of the most or least significant

values is based on the utility and function of the original data stream. Truncation can also refer to initial redaction as well as redaction of trailing value. This can be very relevant for URLs, database integration with URLs, or SQL parameters where injection is a serious risk. For example, the URL http:\\www.google.com\search = "security+identification+compliance" loses very little by dropping http:\\ or the detailed *.com site as the format of the search phrase is very well understood. Truncation is common to protect account numbers and other indexes. Removing the search terms or in the case of imbedded SQL removing the user-supplied parameters is more effective truncation.

Decentralized Label Model

[2204] A security model that annotates data to guarantee confidentiality using labels to support fine-granularity in information sharing. Abbreviated as DLM.

Declassification

[2205] The process of releasing formerly secret documents or information.

Decoding

[2206] The process in data security to restore the form of data through a replacement lookup table or decryption method.

Decoy

[2207] Disinformation, misinformation, or concealment used to camouflage the accuracy, integrity, or existence of information.

Decrypting

[2208] The process in data security to restore the form data using a mechanical, mathematical, or other statistical process.

Defense-in-Depth

[2209] The siting of mutually supporting defense positions designed to absorb and progressively weaken attack, prevent initial observations of the whole position by the enemy, and to allow the usage of reserves. Also ,the use of multiple techniques to help mitigate the risk of one component of the defenses being compromised or circumvented.

Degrading

[2210] Not a common use. The process of creating a Secured DIGITALDOORS data stream. See also extraction.

DES

[2211] Data encryption standard. An encryption/decryption algorithm defined in FIPS Publication 46. The standard cryptographic algorithm developed by the National Institute of Standards and Technology and under jurisdiction of AES.

Destination data stream

[2212] Any data, data flow, or structured data delivered as a result of processing.

DigitalDoors

[2213] This company and the self-named suite of software applications that provides security, survivability, privacy, confidentiality, and anonymity by finding, extracting, and dispersing critical and important information.

DigitalDoors Model of Controlled Release

[2214] The unlimited control over the reconstitution of the extracted data into the redacted data stream subject to the rules and purpose driven by security purpose.

Digital Door

[2215] A place where a data stream is marked by anchor, placeholder, or by the null set to

indicate that critical and important information was extracted and where other things can be placed, such as a link, misinformation, disinformation, or advertisements. A specific item of information at risk at a process touchpoint.

Digital tampering

[2216] See counterfeiting. The processing of copying, faking, altering, and generating duplicates or facsimiles to bypass security, falsify credentials, or create a clone market.

Directory

[2217] A container defining the location of recovery information for reconstituting the remainder into a data flow for access at any defined sensitivity level. A directory has several formats that include consolidated, inclusive, exclusive, and granular itemized.

Directory, consolidated

[2218] A directory format that includes all recovery information or the location to the consolidated recovery information. It is consolidated because all recovery for all granular items are referenced.

Directory, exclusive

[2219] A directory format that includes the location to recovery information for a single defined sensitivity level.

Directory, granular itemization

[2220] A directory format that includes the location to recovery information for all granular items providing for reconstitution at any sensitivity level, compartment, category, user, group, mission or combination thereof.

Directory, inclusive

[2221] A directory format that includes the location to recovery information for a single defined sensitivity level for reconstituting the remainder to that sensitivity level where the remainder information includes all granular items up to and including the dominated sensitivity level.

Discovery

See electronic discovery.

Dispersal

[2222] The process of creating the remote primary storage of extracted information according to its clearance.

Dispersion

[2223] 1. The location(s) for the remote primary storage of extracted information according to its clearance. 2. The process of securing, distributing, protecting, and anonymizing data by the extraction of identifying components from the source for separate storage or delivery. 3. A method for data storage and delivery where the constituents of the whole are randomized and/or stored in separate locations requiring a directory or recovery script to restore into initial form. The scattered or diffused storage of critical and important information renders that information valueless without context, nearly impossible to match together correctly, and difficult to locate without a directory. In addition, the scattered or diffuse delivery (often by separate channels and types of delivery methods) of that same data for reconstitution protects the constituents of the whole from useful interception and contextual understanding. Microsoft calls this dissipation.

Dispersive protection

[2224] The process of securing, distributing, protecting, and anonymizing data by the extraction of identifying components from the source for separate storage or delivery. Also, Diasporagraphy.

Display

[2225] The representation of a source data file in is entirety or without extractions, extractions alone, or a reconstitution of the Source Data Streams with extracted objects are represented by nothing, placeholders, substitutions, replacements, barcodes, or serialized tags.

Disinformation

[2226] Purposely incorrect information used to divert attention to other areas in order to stall, buy time, or cause others to make incorrect assumptions.

Divergence

[2227] The act, condition, quality, or fact of moving away in different direction from a common point or limit.

Domain

[2228] The realm for which information and process is defined and under which data stewardship is maintained. This could be defined for the intelligence community, the community of interest, a single agency, a department with an agency, or a defined subset of the intelligence community.

Dominated rollup

[2229] Recovery information up to and including the referenced sensitivity level in a single package. See also recovery or inclusive.

Downgrading

[2230] The process of modification of a secret document such that is becomes unclassified through the process. Also called Sanitizing.

EAI

See Enterprise Application Integration.

Electronic discovery

[2231] Any process in which electronic data is sought, located, secured, and searched with the intent of classification or of using it as evidence in a civil or criminal legal case.

Encoding

[2232] The process in data security to encrypt or change the form of data through a replacement table rendering it unrecognizable and useless to all but intended parties.

Encrypting

[2233] The obfuscation process in data security to transform data using a mechanical, mathematical, or other statistical process rendering it unrecognizable and useless to all but intended parties.

Enigma

[2234] 1. A code. 2. An encoded character. 3. A zero. 4. A problem that cannot be solved; a cipher. 5. A security encryption machine used by the German military in WWII.

Enterprise Application Integration

[2235] The use of software and architectural principles to bring together (integrate) a set of enterprise computer applications. It is an area of computer systems architecture that gained wide recognition from about 2004 onwards. EAI is related to middleware technologies such as message-oriented middleware (MOM) and data representation technologies such as XML. EAI enables data propagation and business process execution throughout the numerous distinct networked applications as if it would be a unique global application. It is a distributed transactional approach and its focus is to support operational business functions such as taking an order, generating an invoice, and shipping a product. The use of middleware to integrate the application programs, data bases and legacy systems

involved in an organization's critical business processes. EAI applications are really an extension of legacy gateway applications, with provisions for tighter coupling with back-end applications than the legacy gateway solutions. See also Service-Oriented Architecture.

Entropy

[2236] A measure of the disorder or unavailability of energy within a closed system. More entropy means less energy available for doing work.

Evaluative measurement

[2237] An evaluation that describes how the quantitative and qualitative results were attained. See also Qualitative Measurement and Quantitative Measurement.

Expression

[2238] A technical cellular biology or biochemistry terminology that describes the functional manifestation of the DNA informational unit into the RNA process. In security, expression represents the risk inherent in any data stream when that information unit is manifested as a process or repurposed in any way. Essentially, the content of any informational unit is neutral in terms of risk and utility until a context is applied.

Extraction

[2239] The process of removing characters, words, phrases, numbers, image and video segments, audio segments, and any other identifying patterns and objects from Source Data Streams creating destination data streams. The streams include streams without the extractions, streams with only the extraction objects, and streams for projecting a presentation or reconstituting the Source Data Streams (through methods of censoring, masking, tagging, logging, tracking...etc.).

Extraction stream

[2240] A destination data stream of objects removed from a Source Data Stream.

Faceted classification

[2241] Content classification system showing slices of information displayed along a single axis of concept (where there are and often are many axes).

Federated search

[2242] The usage of multiple web-based search engines based on keywords with results combined into a single unified report. Also Confederated Search and Metasearch.

Firewall

[2243] 1. A mechanism to protect network stations, subnetworks, and channels from complete failure caused by a single point. 2. A device, mechanism, bridge, router, or gateway, which prevents unauthorized access by hackers, crackers, vandals, and employees from private network services and data. 3. A moat between public data networks (i.e. CompuServe, Internet, and public data carrier networks) and the enterprise network.

Folksonomies

[2244] A method for classification based on a social approach to collaborative taxonomy. Also known as social bookmarking, social indexing, or social classification. Although this method is flexible, it often becomes unmanageable due to classification bloat, unsupported classification subtleties, and differences in category applicability. The sensitivity levels of T, S, C, and U, for example, proliferated in just a way with SBU, EO, IU, and other ill-define and narrowly applied tags.

Formlessness

[2245] The state of information or resource without context or concept. The presumptive utility is distorted to mask, hide, obscure, render uncertain without value or with so many values as to be

overwhelming. Consider formlessness as stealth, ambiguity, and surprise when applied to functional security, and as the critical component within the DigitalDoors security paradigm shift.

Freedom of Information Act

[2246] U.S. Congressional law (1996) providing for the release of government documents, even materials at one time deemed classified or top secret. Abbreviated as FOIA.

Functionally-compliant security

[2247] The application of security that is not disruptive to ongoing processes and workflow, that coexists with sharing and knowledge discovery without detracting from the expectations and results. This is security that is not a Hobbesian choice with performance.

Generic redaction

[2248] The process of granular removal of all content within a data stream that conforms to all DigitalDoors filters and sieves leaving a remainder that unconditionally conforms to unclassified distribution policy. See also Selective Redaction and Universal Redaction.

Generic remainder

[2249] The data stream that is left after granular content is removed from within the data stream. The data stream can show placeholders for each redaction or nothing at all.

Ghost

[2250] The resulting remainder form or template that is generated by extracting from a data stream all its sensitive data. See also Remainder.

Granular content

[2251] The smallest divisible atomic unit of data with information stripped from its format, representation, structure, purpose, context, usage, source, destination, and encoding. Granular content

is unitary information (that could be multi-tiered and indeterminate) but it is also defined within its context and concept of the greater information unit.

Granular Content Control

[2252] TheDigitalDoors paradigm shift that includes the structure of information, the nature of meaning and semiotic content, context and concept, the format and structure of file and information storage, metatag and metadata, and the compound and hierarchical layering of knowledge. See also granular data control.

Granular data

[2253] The selected and extracted item(s) of information removed from content, context, and concept to enable security, privacy, continuity, survivability, and information sharing.

Granular data control

[2254] The DigitalDoors paradigm shift that includes the structure of information, the nature of meaning and semiotic content, context, and concept, the format and structure of file and information storage, metatag and metadata, and the compound and hierarchical layering of knowledge. See Granular Content Control.

Gray list

[2255] The bulk of a data stream that is not categorized and filtered by a black list or white list and specifically requires additional attention. This represents the unknowns that are neither not acceptable nor acceptable because these are not recognizable. The gray list represents an unqualified and unquantified risk that usually must be reduced with methods to assign the unknowns into either the black list or white list.

Gray list taxonomy

[2256] 1. The process of categorizing that part of a data stream that slips through the sieves of black lists and white lists. 2. The process of recognizing the potential for single items trapped by black lists or white lists to be aggregated, inferenced, or permutated to form new contextual items of interest. 3. The classification through proxy processes for data stream items that do not fit into a single category, are dependent on contextual assignment, or fall into multiple categories in general but will have an initial or presumptive classification. Also referred to as gray list analysis.

Green Book

[2257] DigitalDoors proposed structure and recommendation for information sharing and data management for security implementations to the Rainbow series of security documentation based on the granular control control paradigm shift.

Hardening

[2258] A generic term applied to any efforts made to enhance security and survivability without validation of its effectiveness. This term is derived from the effort to reinforce concrete with salt and rebar creating a more resistant material.

Hashing

[2259] A method used in security to protect direct access to resources through a secondary chain of custody. It is an indexing technique in which the value of a key (record identifier) is numerically manipulated to directly calculate either the location of its associated record in a file or the starting of a search for the associated record.

Hybrid encryption

[2260] A method of encryption that combines two or more encryption schemes and includes a combination of symmetric and asymmetric encryption to take advantage of the strengths of each

encryption method, usually applied to centralize the resources needed for encryption, to reduce costs, or to handle the matrix information distribution requirements of PKI.

Identification (of) Friend or Foe

[2261] A military practice to create a call sign, password, or challenge password and countersign pair changing daily (or more frequently) to validate authenticity of another person thereby protecting against spies, intelligence extracted from captured soldiers, and familiarity with military protocols and language. Abbreviated as IFF.

Inference

[2262] The process of deducing new facts from other existing facts. It is operationally a transformation of data with other data by lookup, mining, deduction, or association. Inference also acts on data to create a useful, anticipated, or practical information.

Information

[2263] Synonymous with data. The difference between data and information is generally defined by the Data, Information, Knowledge, and Wisdom (DIKW) Information Hierarchy that defines data as raw symbols lacking definitive context whereas information is data defined by a specific implied practical context. Generally, within DigitalDoors, information will refer to a compound document or stream with an anticipated use that can be parsed into smaller and smaller subsets of structured groups or data elements.

Information-centric

See data-centic.

Information Hierarchy

See the Data, Information, Knowledge, and Wisdom (DIKW) Information Hierarchy.

Information concept

See concept.

Information content

See content.

Information context

See context.

Information extraction

[2264] The four-step process to parse a database, or data stream to define and select granular content, classify that content in a multitier overlapping hierarchical structure with social and associative contextual and conceptual clustering. ???

Information fabric

[2265] A term coined by Forrester Research to define virtualization layer of data storage into a centralized location for efficiency and simplified management, however, it creates a single point of failure.

Information sharing

[2266] A strategic partnering relationship between data suppliers and data consumers is characterized by a willingness to be open, and to share data as well as the benefits resulting from the information sharing. Cultural unwilling, organizational barriers, and threats to power and control undermine traditional information sharing, so much so, that information sharing is ad hoc and represents a coin of the realm in power brokering.

Information Technology Infrastructure Library

[2267] A collection of best practices for information technology service management. The

United Kingdom's Central Computer and Telecommunications Agency (CCTA) created the ITIL in response to growing dependence on information technology for meeting business needs and goals. The ITIL provides businesses with a customizable framework of best practices to achieve quality service and overcome difficulties associated with the growth of IT systems. The ITIL is organized into sets of texts which are defined by related functions: service support, service delivery, managerial, software support, computer operations, security management and environmental. In addition to texts, which can be purchased online, ITIL services and products include training, qualifications, software tools and user groups such as the IT Service Management Forum (itSMF). While owned by the CCTA since the mid-1980s, the ITIL is currently maintained and developed by The National Exam Institute for Informatics (EXIN), a non-proprietary and non-profit organization based in the Netherlands. Abbreviated and commonly references as ITIL.

Informational unit

[2268] The generic terminology for the basic and smallest unit of data or any larger arbitrary unit of data. It is an item, an element, a field, a data stream, a file, a message, a signal, a process, or metadata.

Information warfare

[2269] Those actions intended to protect, exploit, destroy, deny, corrupt information or information resources in order to achieve significant advantage objective, or victory over an adversary. Operations conducted during a time of crisis or conflict to achieve or promote specific adjectives where actions are taken to affect adversary information and information systems while defending one's own.

Six factors in the value of an information resource:

Relevance to concerns of commitment of a player

capabilities of a player

Availability of resources of a player

available of a resource to an adversary

Integrity of a resource

over time resource can increase or decrease in value

This is also called cyberwar or cyberwarfare.

Injection

[2270] The surreptitious insertion of data, triggers, commands, macros, and metadata (all types of generic signal data) into a data stream in order to repurpose, redirect that stream, or attack process and resources. Insertion is often exploited as alteration of HTML headers in web sites, URL parameters, or as SQL injection with fraudulent or multiple stacked commands. Failure to qualify user input before command execution and limit command execution is a large risk within any process. However, the very ability to change and alter a data stream in any form with injection provides a doorway of risk for compromising the intent and purpose of just about every data stream. Alterations of injection or the alteration (as in patching, overwriting, manipulating) or the truncation of parameters. This is just a variation on the same theme of injection. It is a very significant threat because detection, verification, authentication, and prevention is all based somehow on data streams, which all in turn, are at risk for injection.

Intangible asset

[2271] An identifiable non-monetary asset without physical substance that can be recognized, identified, and transferred as a separable entity.

Intellectual capital

[2272] The dynamic effects of individuals' intellect applied as an input to a classical economic

production model to create a multiplier effect to the other inputs of labor and capital. Intellectual capital includes all of the knowledge applied to production efforts, culture, history, practices, and goodwill, that knowledge embodied as part of the actual production output, the systems of command, control, and communications applied to organize the production process. See also intellectual property.

Intellectual property

[2273] That specified part of intellectual capital that is codified as property, generally as patents, brands, trademarks, trade secrets, defined proprietary business practices, and internal organizational workflow.

Intentional security

[2274] The fundamental security that is engineered into the operating system, the applications, the environment, and/or the workflows to address both known and as yet unknown flaws assuming the certainty of an eventual attack.

Interaction

[2275] The formal or informal process of creating new information or formulating new hypotheses through the accumulation of unstructured and structured data often through the reorientation, transformation, or normalization of different types of data into unified formats usually in the presence of a specific trigger or catalyst. See also the more formal Data-to-Data Interaction.

Intrusion detection

[2276] Fancy word for security that provides a firewall, monitors traffic loads, prevents virus attacks, and also tracks sources for users.

Keyword

[2277] Any acronym, word, number, or phrase that is used as a search term, data lookup, or index

value and that has content- or context-value itself.

Keyword poisoning

[2278] The insertion, misuse, or overuse of information with content- or context-value to alter search or indexing results for a purpose that undermines the effectiveness of the primary utility of the search or indexing.

Knowledge exploration

[2279] Any process used to lookup, find, search, index, extract, or discover data within a data stream, document, database, data warehouse, or other storage infrastructure with emphasis on discovery of hidden, undisclosed, relationship-oriented, or contextually-based references.

Label

[2280] An applied designation or categorization for a defined range of information. It can be in-channel, out-of-channel, or bracketing the range to specify the relevance of the application of the label. A label can be the [TS] marking within a tear line, the <start><\end> in HTML, SGML, or XML. Labels can also exist as a hierarchical or referential metadata in separate data streams, as typically handled within DigitalDoors. See also tag.

Labeling

[2281] The process of applying a label or tag based on a formal taxonomy or classification process.

Linear adjacent composites

[2282] The recognition and selection of unknown or uncategorized terms that are in proximity to a known dictionary item and the association and connection of these new terms to the previously recognized one. This is useful for automatically constructing new search or categorization targets based

on related items, specifically categorized words or unknowns that are adjacent to each other can be combined to form new linear combination phrases. See also linear adjacent permutations. More detail is available in the white paper titled, "Adjacency with Permutations and Combinations."

Linear adjacent permutations

[2283] The recognition and selection of unknown or uncategorized terms that are in proximity to a known dictionary item and the association and connection of these new terms to the previously recognized one. This is useful for automatically constructing new search or categorization targets based on related items, specifically categorized words or unknowns that are adjacent to each other can be combined to form new linear combination phrases. See also linear adjacent composites. More detail is available in the white paper titled, "Adjacency with Permutations and Combinations."

Link

[2284] A connection to other resources using imbedded references, most commonly implemented as a universal resource locator (URL) in html sources. XML pointers and other hyperlink mechanisms fulfill the same purpose as a link but also create a reverse link for security, integrity, and expanded reference.

Malicious compliance

[2285] A concept of compliance to the letter of the law of one or more standards but not the spirit of the law with foreknowledge that efforts do not achieve overall goals and yield a negative impact on the organization.

Mark-to-market

[2286] An accounting methodology of assigning a value to a position held in a financial instrument based on the current market price for the instrument or similar instruments, and should

equally apply to intangibles that are monetized through any number of viable valuation methods.

Marker

[2287] A device to indicate a position, relative position, or location that dependent on its purpose or context. Sometimes, the term marker is used to describe a label or tag. See also Label or Tag.

Markle Foundation

[2288] A private organization that works to realize the potential for emerging information and communication technologies to improve people's lives by accelerating the use of these technologies to address critical public needs, particularly in the areas of health and national security.

Markle Report

[2289] There are many Markle Reports created by different task forces. The ones of relevance to information security include "Creating a Trusted Network for Homeland Security" and "Protecting America's Freedom in the Information Age."

Markup

[2290] A confusion terminology usually referring to the eventual intent or manual indication that for inserting a marker. Sometimes, the term markup is the prior manual step to insert a marker, which is also used to describe a label or tag.

Masking

[2291] A process of distributing versions or subsets of data prepared by a redaction, filtering, or extraction process where the deliverable result is missing parts of the data or has placeholders for the data. One example of masked data presentation is XXXX-XXXX-XXXX-2807 for a MasterCard account number that complies with the PCI DSS requirements.

Master data management

[2292] A method of enabling an enterprise to link all of its critical data to one single file, called a master file, which provides a common point of reference. When properly done, MDM streamlines data sharing among personnel and departments. Furthermore, MDM can facilitate computing in multiple system architectures, platforms and applications. The benefits of the MDM paradigm increase as the number and diversity of organizational departments, worker roles and computing applications expand. For this reason, MDM is more likely to be of value to large or complex enterprises than to small, medium-sized or simple ones. When companies merge, the implementation of MDM can minimize confusion and optimize the efficiency of the new, larger organization. Abbreviated as MDM.

Metadata

[2293] Simply information about information or data. The term is usually use in context for providing descriptions about the data stored in a database or a compound hierarchical document (including XML or a MS Word file), such as structure, purpose, format, last updated, or intent. Metadata is usually stored interleaved with visible data or out-of-band to the user data thus creating a significant security risk. Metadata is likely to replicate the full content of the user data, define structure and content for the visible user data, and contain internal and external information references.

Metafilter

[2294] A search term categorization used to reduce the search results based on the compliance to selection categories or sensitivity levels or other contextual bases.

Metainformation

See Metatag.

Metalist

[2295] Not the rock band. A list of lists or a hierarchy of lists. It is often created as a collection

of specific dictionary entries, categories, classifications, types of words and phrases that are combined in order to define a compliance target, such as HIPAA or GLBA. Note that entities prohibited by the HIPAA safe harbor provisions and by GLBA are likely to overlap, such as names, dates, costs, social security numbers, and contact information. The metalist creates a shortcut for the complexity of these entries and provide a means to resolve list overlaps. Metalists are implemented as "profiles" and the user selection of profiles for automated selection operations. It is also critical when reconstituting generic remainders with granular content control for unknown future data repurposing while preserving functional security.

MetaPII

[2296] Anything external to a generical or legal set of PII categories, usually descriptive, that can create PII or in other methods uniquely identify an individual (person, process, property) or approximate the identification of an individual. It is any information that can create aggregation, inferential, or interaction subsets that either physically or statistically confines an individual (person, process, property) within an identifying scope or range.

Metasearch

1. The process of search through multiple other search engines with a combined output from all search engines. See Confederated Search or Federated Search. 2. A DigitalDoors divergent search without preconceptions of range and reach to discover related information from a variety of local sources, databases, data stores, and external resources, both structured and unstructured, and Internet facilities and to apply objective security for cross-domain information sharing to uncover a plot during its incubation. Remote searches initiated by the metasearch that penetrate other domains are secured before results are passed back to the original metasearch and consolidated.

Metasecurity

[2297] Marty's coined term... maybe? Idea is right but is terminology correct, useful or just too cute

[2298] The use of imbedded in-channel information, triggers, and macrocode to sign, validate, authenticate, send messages, protect, and inform on the unanticipated use or misuse of a data stream. Problem of such use of metadata is that it itself opens doors to risk so that the assessment becomes one of assessing not only the basis risk but also the metarisk of the added protection mechanisms.

Metatag

[2299] A special tag that provides information about the tagged, included, or referenced information. A metatag is a tag, but it conveys sensitivity level, purpose, source, usage, categorization, or other pragmatic value beyond content. HTML, XML, and DoD sensitivity level tags are metatags. DigitalDoors creates out-of-channel metatags, and specifically a multitier overlapping hierarchical structure of metatags.

Microperimeter

[2300] The demarcation of a presumed closed subset. The enclosure of a set of resources by a surrounding protection mechanism at the demarcation.

Microperimeterization

[2301] The attempt to create a closed set with a surrounding for the purpose of enabling security, like the process sandbox.

Mimic

[2302] Disinformation, misinformation, or concealment used to camouflage the accuracy, integrity, or existence of information.

Mimicry

[2303] Disinformation, misinformation, or concealment used to camouflage the accuracy, integrity, or existence of information.

Misinformation

[2304] Purposely false, misleading, or deceptive information used to divert attention to other areas in order to cause others to make incorrect decisions and initiate the incorrect actions.

Mislabeling

[2305] A method of purposefully classifying information incorrectly to decrease its perceived value in order to bypass security systems or raise its value for asset valuation fraud.

MultiCryption

[2306] A DigitalDoors trademark for mixed usage of encryption methods and mixed one-time keys with output interspersed (in whole or part) within original data stream. In other words, this process selects the elements within a document to encrypt, encrypts each with a random method using a random one-time key, and inserts the result back into the source document.

Multilateral security

[2307] A concept for implementing security when security interferes or directly conflicts functional or operational interests. Specifically, the release of personally identifiable information is necessary to conduct a credit check or medical insurance review but contrary to its full protection.

Multiple Independent Levels of Security

[2308] An architectural approach to security that represents a layered approach to building the high assurance components required for cross-domain solutions. Abbreviated as MILS. See MLS as the forerunner to MILS.

Multiple Levels of Security

[2309] A DoD and intelligence agency security program that enables information systems to provide capabilities that augment its existing single-level data processing and data communications services over many different and separate sensitivity (security) levels. Data of multiple sensitivity levels are processed and transferred by the system, which also separates the different sensitivity levels and controls access to the data. Abbreviated as MLS. See MILS as the current embodiment of MLS implementation.

Multitier hierarchical tag structure

[2310] The application of multiple and overlapping tags assigns sensitivity level, compartments, categories, purpose, usage, and other pragmatic meaning to granular content. It is multitier because many tags can be applied to the same content to describe different aspects of the content, and it is hierarchical because tags can overlap to define words within phrases within paragraphs within documents or other compound data streams. Traditional DoD tags indicate a non-reflexive sensitivity level (and/or compartment) but do not in any way describe the potential for inter-domain information sharing except perhaps to preclude it, whereas multitier hierarchical tagging enables security compliance with granular information sharing to multiple domains. See also Tag Cloud as a web-enabled ontology for the classification of single information streams for multiple uses.

Multitier hierarchical overlapping tag structure

[2311] The application of multiple and overlapping tags assigns sensitivity level, compartments, categories, purpose, usage, and other pragmatic meaning to granular content. It is multitier because many tags can be applied to the same content to describe different aspects of the content, and it is hierarchical because tags can overlap to define words within phrases within paragraphs within documents or other

compound data streams. The basis for success with granular content control is to establish a comprehensive review of data content, context, and concept as a precursor to .tagging and all subsequent filtering and assessment of purpose. [Ron has issue for meaning of hierarchy in terms of overlap with hierarchy resolution] Traditional DoD tags indicate a non-reflexive sensitivity level (and/or compartment) but do not in any way describe the potential for inter-domain information sharing except perhaps to preclude it, whereas multitier hierarchical tagging enables security compliance with granular information sharing to multiple domains. This specific implementation resolves issues that plague cross-domain semantic interoperability and the complexity if not impossibility for mapping a consistent dictionary and data model for communities of interest (COI) by reducing the problem of information tagging of the gross document to a granular one. See also Tag Cloud as a web-enabled ontology for the classification of single streams for multiple uses.

Negligence

[2312] The failure to act and the legally-assigned responsibility for preventing an event of consequence.

Network

[2313] Any local network, wide-are network, intranet, or Internet that provides an information delivery mechanism and infrastructure for workflow. It provides a distribution channel with marginal cost of zero.

Objective

[2314] A method of dealing with logical and experiential evidence.

Objective security

[2315] A concept for security based on external measurement and substantiation. This is a

difficult ambition for a number of reasons because of a lack of clear and stable security models, the rare inclusion of formlessness into strategic designs, insufficient clarity whether to protect infrastructure or apply a data-centric model, and complexities in defining data and information that requires protection.

Objective security compliance for universal distribution

[2316] A redacted MLS or MILS document with sensitive items selected and extracted for a general distribution. Also a Base Redacted Document.

Ontology

[2317] 1. The assignment of information into one, two, or more categorical distinctions. 2. The applied categorization through the use of label or tag. 3. XML semantics often described by XSL or DTD structures. 4. This philosophical term (from metaphysics) refers to assertions or assumptions about the nature of reality. 5. A classification structure based on a set of types, properties, and relationships. It is an aggregated faceted classification as a complete collection. This is very relevant for computer programming in that ontologies comprise collections of objects.

Open source information

[2318] Intelink-U, formerly known as Open Source Information System (OSIS) is a virtual private network for the exchange of unclassified intelligence information.

[2319] A concept for the sharing of information that inherently exposes secrets through aggregation, inference, and interaction.

Order-sensitivity

[2320] Decryption fails if the order of the application of multiple encryption keys is not reversed. In other words, if the first key applied to data is "A1B34840D4848DF092833FC4B"h, the second encryption key applied is "045BE7435CFA345BB3B56EE00"h, the first decryption key is

"045BE7435CFA345BB3B56EE00"h and the second is "A1B34840D4848DF092833FC4B"h. DigitalDoors is order insensitive, and can be Secured infinite times and restored in any order infinite times or delivered through the Defense-in Depth facilities.

Padding placeholders

[2321] A device to mark a position, relative position, or location for out-of-sequence or partial reconstitution when the recovery information is inclusive, exclusion, granular itemized, or in the rare instance when replication or fault-tolerant failover cannot reconstitute the extracted information. The presentation of this placeholder can also be suppressed with contraction. Padding placeholders can be suppressed as null sets, fixed-length special characters, or random substitutions.

Placeholder

[2322] A device to mark a position, relative position, or location that is independent of its purpose, content, context, or concept. The placeholder can be displayed as special character(s) or images and the actual display can represent the exact size of the extraction, a randomly sized substitution, or a null set so as to suppress delivery of information through position, width, frequency analysis, context,

Patchwork

[2323] 1. Inferior effort to complete or resolve problem. 2. A method of Steganography where a block of data is imbedded as a visible part of an image that might not be so obvious, such as a Microdot.

Perimeter

[2324] The demarcation of a presumed closed set. The enclosure of resources by a surrounding protection mechanism at the demarcation.

Perimeterization

[2325] The attempt to create a closed set with a surrounding for the purpose of enabling security.

Personally Identifiable Information

[2326] Any an all categories of information that is useful to indicate or de-identify a specific individual. The definition applies to categories of information, like social security number, account numbers, or date of birth, under HIPPA Safe Harbor regulations and other legislation. PII is also physical characteristics, tattoos, histories, likely locations to find an individual or relatives of the individual, or even aggregated categorical information that can be used to expose an individual. PII includes any characteristics that can uniquely identify or force the reidentification of an individual, or statistically denote groups likely to include individuals of interest. PII does not reflect the results of aggregation, inference, or interaction on lesser information pooled to identify the individual. Abbreviated as PII.

[2327] Physical separation

See Separation.

[2328] PII

Abbreviation for Personally Identifiable Information.

[2329] Pooled data

See Aggregation.

[2330] Pooling

See Aggregation.

Plurality

[2331] The assumed precondition that DigitalDoors security software data-oriented access supports input and output to 0, 1, 2 or an infinite number of files, paths, or streams. Source and target definitions are collection-oriented; zero or a single path definition is really the special case and multiple

path source or targets is the usual case. As such, BLM compliance with inclusivity or exclusivity, and granular itemized recovery streams presumes multiple hierarchical files for the active sensitivity levels and granular extractions. These multiple hierarchical files can exist on a single local storage path or dispersed to a virtually-infinite array of remote, networked, or linked UNC or X-linked locations.

Pragmatics

[2332] The practical use of signs by agents or communities of interpretation within particular circumstances and contexts.

Privacy

[2333] 1. A method, desire, or process to maintain personal, business, or corporate identity without releasing it to others. 2. A method to encrypt and hide packet data from eavesdropping. 3. Privacy is a subset of security where access and usage control is function of stewardship.

Projection

[2334] The partial or full representation of a source data file; in toto without extractions; extractions alone; or a reconstitution of the Source Data Streams using the destination data streams (extraction streams and reconstitution streams) where extracted objects are partially or fully represented by nothing, placeholders, substitutions, replacements, barcodes, or serialized tags with a major distinction that the data streams require to reconstitute the Source Data Streams are integrated together from separate network, data, communication, optical, or wireless channels. The termination is the representation itself, with no further access or processing allowed.

Proxy server

[2335] A device that implements the technique and the technique used to cache information on a Web server and acts as an intermediary between a Web client and that Web server. It basically holds

the most commonly and recently used content from the World Wide Web for users in order to provide quicker access and to increase server security.

Pseudo-anonymity

[2336] A method, desire, or process to maintain personal, business, or corporate identity by substitution false, counterfeit, or a lack of information.

Purpose

[2337] The intent, assumption, presumption, known rationale for controlled release of granular content (subject to rules and policy enforcement).

Qualitative measurement

[2338] An evaluation that can be tracked by descriptions of thinking and behavior. See also Quantitative Measurement.

Quantitative measurement

[2339] An evaluation that can be tracked with numbers (e.g., return on investment, revenue generated, costs, absentee days). See also Qualitative Measurement.

Random

[2340] 1. A statistical term referring to a process with outcomes in a defined sample space -- random does not mean chaotic, indiscriminate, or arbitrary. 2. A technical term referring to events with no formal interrelationship; events that occur with equal probability.

Randomness

[2341] The state of being random.

Reconstitution

[2342] The full representation of a Source Data Stream from extraction streams and recovery

streams, in toto, as a display or restored as a data stream for secondary or subsequent processing. This is a verbatim recreation of the Source Data Stream or file. (Xerox calls this reverse redaction, while Microsoft calls this function inline translation in relation to the Digital Asset Store; added 02-28-08).

Recovery

[2343] A container with the extracted information, commands for the complete restoration of the original source, and optional information about the source. This information can be consolidated, inclusive for all dominated sensitivity levels, exclusive to a single sensitivity level, or granularized such that each item is a separate container for each tear line or extraction.

Recovery, consolidated

[2344] A format that includes recovery information for all granular items.

Recovery, exclusive

[2345] A format that includes recovery information for a defined sensitivity level.

Recovery, granular itemization

[2346] A format that includes recovery information for a single specific tear line or extraction item.

Recovery, inclusive

[2347] A format that includes recovery information where the remainder information includes all granular items up to and including the dominated sensitivity level. See also Dominated Rollup.

Recovery, stacked

[2348] A format that includes recovery information where the remainder information includes all granular items up to and including the dominated sensitivity level. See also Dominated Rollup.

Recovery information

[2349] A file or stream containing the extracted information, commands for the complete restoration of the original source, and optional information about the source.

Recovery file

[2350] A file or stream containing the extracted information, commands for the complete restoration of the original source, and optional information about the source.

Recovery script

See Recovery.

Recovery stream

[2351] A destination data stream of objects removed from a Source Data Stream with sufficient information and direction to project a presentation or reconstitute the Source Data Stream. See topic recovery for standards on recovery stream metadata.

Reconstitution

[2352] The process used by DigitalDoors to blend the extracted data back into the source (Secured DigitalDoors) skeleton. The process spans the point where no extracted data is restored to the skeleton to where all is returned. This is controlled by user, authentication, and other controlling modes. Microsoft calls this process inline translation.

Redacted document

See Objective Security Compliance for Universal Distribution.

Redaction

[2353] The process of creating a secured DigitalDoors data stream. See also extraction.

Reference

[2354] A connection, link, or similarity between two different resources.

Reference monitor

See Security Reference Monitor.

Reflexivity

[2355] The condition that exists when a situation or meaning is self-identical to its antecedent, the premises is equal to the consequence, or where a tag, metatag, metadata, formatting, or other situational description adds information to the actual data stream content.

Release risk

[2356] The inverse relationship between the reconstitution of more granular content and the level of risk. The more content that is released, the higher the resulting risk through aggregation, inference, and interaction.

Remainder

[2357] The resulting source after it is secured by DigitalDoors through selection, extraction, and redaction. The result is normally the lowest common denominator suitable for distribution to any party with any defined or unknown sensitivity level. A remainder without content is an universal redaction, Skeleton, or a Ghost.

Remainder contraction

[2358] The resulting source after it is secured by DigitalDoors through selection, extraction, and redaction without placeholders, white space compression, and reformatting to mask extractions

Remote monitoring

[2359] The process of tracking process, security, functionality, and other flows by means of remote data gathering.

Remote visual identification

[2360] A method to identify and authenticate a person based on remote video viewing.

Replication

[2361] The automated copying of base-level output in primary storage location(s) to multiple secondary, tertiary, off-line, and knowledge management sinks, and many other storage locations in order to provide security, survivability, and privacy.

Retrofit security

[2362] Security that is added after the fact in order to address known, perceived, or anticipated flaws from recurring. Retrofit security is rarely proactive simply because it represents a patterned response derived from prior attacks. Retrofit security is binary, providing a single mode of protection, such as access or no access, and delivery of data through the firewall or denial of that delivery.

Risk dynamics

[2363] This represents the forecast of what the risk needs might be under different contexts or future times. It is an estimation of data will decay or increase in risk due to the ever-changing ecosystem.

Risk mitigation

[2364] A security based on the concept that barriers, firewalls, access control, and protection of the infrastructure is impossible and will fail but that other methods can potentially provide the continuous adequate performance of critical services and functions after a successful attack or natural disaster.

Role

[2365] A functional assignment with aggregated need-to-know security access control that is separate from sensitivity level and compartment stratification.

Role-based security

[2366] Access control based on a defined roles rather than on user authorization through sensitivity level and compartment dominance.

Rollup

[2367] The aggregation of recovery information dominated by the sensitivity level and/or compartment. In other words, recovery information for [S] would include [C] and [U].

Safe harbor

[2368] A provision of a statute or regulation that reduces or eliminates a party's liability under the law, on the condition that the party performed its actions in good faith. Legislators include safe-harbor provisions to protect legitimate or excusable violations with due diligence.

Safe list

[2369] A register of entities that are acceptable and require no additional attention. This terminology is included here to contrast the difference between the safe list and a black list that has different connotations based on how it is used. Contrast to black list. Also known more generally as white list. See also gray list.

Sandbox

[2370] The establishment of a presumed closed set within a larger open or presumed closed set for security of process or data. See also microperimeter.

Sanitizing

[2371] The process of modification of a secret document such that is becomes unclassified through the process. Also Downgrading.

Scarcity

[2372] An economic principle describing the natural or artificial shortage of a resource. Scarcity

tends to increase the value of that resource and when it provides unique utility, increase its demand as well. The scarcity of information is a foundation for security, where scarcity is both physical, natural, and artificially created by dispersing the information in pieces so that it no longer represents a functional risk.

Search

[2373] The process that looks through documents for specified keywords and returns a list of the documents where the keywords were found. Keywords are often called search terms.

Search convergence

[2374] The metasearch process of validating, substantiating, and establishing the integrity of the fundamental search.

Search divergence

[2375] The metasearch process of expanding the content, context, and concept beyond the strict scope of the fundamental search.

Search terms

[2376] Single words or phases and, words combined with Boolean terms (such as, and, or, not) used as keys or indexes in unstructured text, structure data stores, or relational databases.

Secure

[2377] The process of creating a Secured DigitalDoors data stream. See also extraction.

Secure maintenance

[2378] The process of setting user, process, system, or other configurations by means of a password-controlled and authenticated security system.

Secured DigitalDoors

[2379] Documents or information that were formerly secret.

Secured DigitalDoors stream

[2380] A destination data stream that represents the Source Data Stream without the extracted objects; where the extracted objects are represented by nothing, placeholders, substitutions, replacements, barcodes, or serialized tags.

Security

[2381] The set of functions concerned with ensuring that only legitimate users of the network can use the network, and only in legitimate ways. Ensuring that private information remains private in an atmosphere where all other information is free. Security also means that viruses are prevented from infecting people's systems.

Sclassification

[2382] 1. The process of determining and designating the specific degree of protection against release that the information requires. 2. The degree of protection so determined. Note: Examples of security classifications are: 'top secret,' 'secret,' and 'confidential' See also Security Label, Sensitivity Level, Compartment, Role, and Mission.

Security dynamics

[2383] This represents the forecast of what the risk and the security needs might be under different contexts or future times. IT is an estimation of data will decay or increase in risk due to the ever-changing ecosystem.

Security label

[2384] A tag representing the sensitivity level of a subject or object, such as its hierarchical classification ('top secret,' 'secret,' and 'confidential') together with any applicable nonhierarchical

security categories (e.g., sensitive 'compartmented' information, critical nuclear weapon design information).

Security level

[2385] The combination of a hierarchical security classification and security categories that represents the sensitivity level of an object or the security clearance of a subject.

Selective redaction

[2386] The process of granular removal of all content within a data stream that conforms to all specified DigitalDoors filters and sieves and is applicable to the creation of DigitalDoors Versions. See also Generic Redaction.

Sensitivity level

[2387] Orange book definition for hierarchical access control based on need-to-know where is information release is limited to a single defined upper bound or dominated by that upper bound. Examples include TS, S, or C for 'top secret,' 'secret,' and 'confidential' which are common notational examples for a wider range of tags that can include "Sensitive But Unclassified" or "release to NATO." Provide security.

security by obscurity

[2388] This is a controversial principle in security engineering, which attempts to use secrecy (of design, implementation, etc.) to provide security Also called security through obscurity.

Security reference monitor

[2389] Any edge security device or control that validates tamper proofing, always available, and has a smallest enough test set to exhaustively invoke. It often is implemented as an access control concept that refers to an abstract machine that mediates all accesses to objects by subjects. A reference

monitor should be (a) complete (i.e., it mediates every access), (b) isolated (i.e., it cannot be modified by other system entities), and (c) verifiable (i.e., small enough to be subjected to analysis and tests to ensure that it is correct). This is the only effective tool known for system design and implementation and thus is the "heart of security" as embodied by the DigitalDoors process.

Selection

[2390] The DigitalDoors process in which electronic data is sought, located, secured, and searched for classification and identification of elements for security through extraction. See also electronic discovery.

Selected and critical information

[2391] Words, text, phrases, terms, sounds, and parts of images designated to be privileged, confidential, identifying of personal or organizational identity, and having distinguished economic worth, where designation occurs specifically by inclusion data sets, parts of speech, sound, or image context through other means of artificial intelligence. See Critical and important data.

Semantic

[2392] Pertaining to the meaning of a word, phrase, audio, video, or other content.

Semantics

[2393] The study of meaning that are expressed in a language, code, or other form of representation. Semantics is contrasted with two other aspects of meaningful expression, namely, syntax and pragmatics.

Semantic trees

[2394] A method of data mining based searching for relevant hits, then creating a tiered ranking of found matches, indexing these matches, and then returning a limited number of the most semantically

relevant results.

Semiotic

[2395] Pertaining to the study of signs and symbols.

Semiotics

[2396] The study of signs and symbols, both individually and grouped in systems. This science of meaning extends beyond the signs and symbols used to communicate to include the intended, unintended, and covert delivery of ideas. It includes the study of how meaning is made and understood in biological, human, and machine communications.

Sensitivity level

[2397] A hierarchical structure of secrecy, security, or value for information usually represented as labels or tags such as top secret and applied as [TS], [S], [C], [SBU] and [U].

Sensitivity level hierarchy

[2398] A reconstitution option to include all dominated sensitivity levels with the target sensitivity level or alternately to exclude all superior and dominated sensitivity levels for a report that includes only the items exactly of the target sensitivity level. The options are either inclusionary or exclusionary.

Separation

[2399] Logical versus physical. Issue blurred. See Physical Separation.

Service-Oriented Architecture

[2400] A method for implementing an Enterprise Application Integration. The principal aspect of Service-Oriented Architecture (SOA) is the separation of the service interface from the implementation so that services can be integrated and consumed by clients without concern for how the

functions and features are actually implemented. SOA can also be viewed as the exposure of a software process as an interface. The application programming extension in effect has been expanded from built-in language functions to DLL libraries of practical calls to larger SOA application assemblies. The disadvantages will be SOA-Hell where rampant versions will break the distributed infrastructure.

Silo

[2401] Any repository that is functionally disconnected from operations, organizational works, or other communities of interest. The lack of connectivity provides a measure of sanctuary that is not security. A silo often is formally protected by an air gap, MLS, or the lack of workflow integration. See also stovepipe.

Situational access control

[2402] Access control that is not merely driven by 1- or 2-authentication but also moderated by the current state of the system, network, or database. For example, access to payroll might not be granted Sunday access, whereas all access might be denied if the administrator has evidence of a break-in.

Skeleton

[2403] \ The resulting remainder form or template that is generated by securing a data stream of all its critical and unimportant data.

SOA

See Service-Oriented Architecture.

Software as a Service

[2404] A model for software delivery where the vendor provides maintenance, daily technical operation, and support for the software provided to their client. SaaS is a model of software delivery rather than a market segment; software can be delivered using this method to any market segment

including home consumers, small business, medium and large business. The key characteristics of SaaS software, include: network-based access to, and management of, commercially available (i.e., not custom) software activities that are managed from central locations rather than at each customer's site, enabling customers to access applications remotely via the Internet application delivery that typically is closer to a one-to-many model (single instance, multi-tenant architecture) than to a one-to-one model, including architecture, pricing, partnering, and management characteristics

Source

[2405] Any data, data flow, or structured (record-oriented, field-oriented, database, metadata, hierarchical, typed, or defined by usage) data delivered and captured for processing. The Source Data Stream may be wiped securely and/or deleted during processing, or saved, encrypted, and relocated. See also Source Data Stream.

Source data stream

[2406] Any data, data flow, or structured data delivered and captured for processing. The Source Data Stream can be destroyed or deleted during processing, or saved with optional relocation or encryption. See also Source.

Stacked

[2407] The aggregation of recovery information dominated by the sensitivity level and/or compartment. In other words, recovery information for [S] would include [C] and [U].

Steganography

[2408] The process of digital watermarking; that is the process of hidden information as background noise within another obvious analog image or sound file for security reasons. Usually, this technology hides small amounts of text, such as a watermark, ownership, or copyright, inside a very

large image file and achieves very low rates of efficiency.

Stewardship

[2409] The responsible legal use (including conservation) of a resources in such a way that takes full and balanced account of the interests of owners, society, future generations, and other species, as well as of private needs, and accepts significant moral and even actionable legal liability to society. See also custodianship.

Stovepipe

[2410] Any antiquated process or repository or even a purpose-specific operation that is functionally disconnected from line-of-business operations, organizational works, or other communities of interest. Isolation, air gap, disinterest, lack of mainstream value provide some sanctuary but no real security. See also silo.

Strikeback

[2411] Security countermeasures initiated after breach or violated that employ methods to trace the source of the intruder and apply destructive means against the intruder platform. Note that strikeback often create more harm to the damaged, neutral, or comprised parties than to the attacker, terrorist, or enemy because violations are often hoaxes or routed through unaware third parties to create plausible deniability.

Strong declassification

[2412] The process of separating or processing (Downgrading or Sanitizing) secret (e.g. top-secret, secret, and classified) documents into those which are definitely unclassified and definitely classified without any margin of error. See also Weak Declassification.

Structure

[2413] A schema, format, categorization, or other taxonomy applied to information to delineate a purpose or usage. A structure is but a single rendition of many possible that can add a context or concept to the information.

Structured data

[2414] 1. The application of a specific tagging system to information predicated upon a specific usage or purpose. 2. Any data stream that contains a formalized layout defining the arrays, fields, format, the hierarchical tree, and the purpose of the data within. 3. Any data stream that includes a taxonomy. Note that the structure can be defined either internally, externally, or combinations thereof by means of data table definitions, tags, metatags, metadata, style sheets, formal requirements, and specific program code that parses and extracts data items. Structured data includes databases, records, field-oriented data sets, HTML, XML, most MS Office documents saved and conforming to a document object model, and almost all packet-delivered messages such as e-mail. Because a data stream with an available data structure can be parsed and individual items can be classified within a formal taxonomy, definition 3 is usually a natural consequence of definition 2. Structured data is not to be confused with data structure, which defines the layout of the structured data. Structured data retains its structure only within context or concept, and in fact can also be perceived as unstructured when the context for a useful structure is not obvious. See also unstructured data.

Subjective

[2415] A method of dealing with perceptive and interpreted evidence.

Suppression

[2416] The process of masking the existence of granular extractions or complete tear lines through the contraction and reformatting of the data stream.

Surveillance

[2417] The process of filtering electronic communication in real-time for breaches of security or potential threats.

Survivability

[2418] The continuous adequate performance of critical services and functions after a successful attack or natural disaster.

Syntax

[2419] The construction of complex signs from simpler signs.

Tag

[2420] An applied designation or categorization for a defined range of information. It can be in-channel, out-of-channel, or bracketing the range to specify the relevance of the application of the tag. A tag can be the [TS] marking within a tear line, the <start><\end> in HTML, SGML, or XML. Tags can also exist as a hierarchical or referential metadata in separate data streams, as typically handled within DigitalDoors. See also label.

Tag cloud

[2421] A collection tags often developed by different and arbitrary users for each's own personal benefit and presented so that visual or utilitarian emphasis corresponds to relative importance.

Tagging

[2422] The process of applying a label or tag based on a classification.

Target

[2423] And word, part of a transaction, data field, phrase, sound byte, part of an image, or segment of a multimedia stream (motion, audio, video, text overlays, etc.) for which the limits of that

data are specifically fixed automatically or by offset with a larger data stream or document. For example, the targets within "George Bush ignored the press corps on immediate military response to Indonesian riots" are likely to be "George Bush" and "Indonesian".

Taxonomy

[2424] 1. A subject-based classification system. 2. A system of applied categorization. 3. A definition of the relationship between categories or nodes of information. It is typically a hierarchical structure used to classify information within a knowledge base. 4. One application of an ontology within the set of possible.

Tear line

[2425] Non-shareable information within a document below any designating point in the document. In other words, a document is segregated into paragraphs of decreasing (downward) sensitivity levels so that it can be distributed in a shortened length to recipients with lower sensitivity levels and greater length to those with higher sensitivity levels.

Topic recovery

[2426] An SGML or XML document used for navigation—by people or by machine—within an information set. The DigitalDoors directory and recovery information falls within the concept of topic recovery. A single information set can include any number of types of source data: for example, HTML files, PDF files, and databases could all be part of one information set. A Topic Recovery (TM) can be either embedded within a document, or may be, itself, a separate document. A Topic Recovery provides a particular view of topics within the information set. There could be, potentially, an infinite number of different TMs for any information set: for example, an information set dealing with the plays of Shakespeare could have a TM for navigation within the perspective of Elizabethan culture, or the

perspective of themes throughout the plays. Each TM explicitly declares topics and provides links to relevant information. Topic recoveries are an ISO standard for the representation and interchange of knowledge, with an emphasis on the 'findability' of information. The standard is formally known as ISO/IEC 13250:2003.

Touchpoint

[2427] Any access point, system, conduit, or workflow that exposes an information item to exposure. This is PCI definition, but recognizes the aggregation, inference, and interaction risk when even secured information is accessible within a workflow and brought through any perimeter barrier, however transient the duration, and creates additional unanticipated risks. A touchpoint is a transient or even permanent breach of perimeter security. A touchpoint is an admission and recognition of a process security weakness. See also digital door.

TPM

See trusted platform module.

Tranquility

[2428] A formal Bell-LaPadula Model (BLM) principle that data access control and documents once assigned to a sensitivity level cannot be reassessed. DigitalDoors both adheres strictly to this policy but also extends information sharing functionality through granular content control with dispersion variations that retain strict security and privacy while enabling cross-domain sharing.

Trap door

[2429] A mathematical operational that is easy to complete whereas the inverse of that function is computational infeasible without extra information such as a key (encryption key). See also back door.

Triple-DES

[2430] A variant of DES that encrypts data or a message using 56-bit DES three consecutive times using a different key for each pass. It sometimes refers to a key size three time longer (168 bits) than the standard 56-bit DES.

Trojan information

[2431] The inclusion of information that initiates a risk, leakage, or exposure when it is upgraded or downgraded through an MLS or any other information sharing system that is either intentional or unintentional.

Trojan horse

[2432] A innocuous-looking software program that may not do anything obvious or visually noticeable but nonetheless designed to disable, bypass security or audit options, damage, or destroy computer hardware or read/write storage systems. Typically used by disgruntled employees to damage computer operations.

Trust exploitation

[2433] This is a description of how an attack works when a dependable resource within a network (both social or technical) takes advantage of its special position. This can be combined with access control flaws to counterfeit the position of a dependable resource as well.

Trusted Granular Content Control™

[2434] The DigitalDoors paradigm shift that includes the structure of information, the nature of meaning and semiotic content, context, and concept, the format and structure of file and information storage, metatag and metadata, and the compound and hierarchical layering of knowledge. Because an individual item within the data stream, thus "granular" data, is assigned to a multitier hierarchical overlapping tag structure, an individual item can be distributed subject to sensitivity levels throughout

a trusted network and to trusted platform modules or workstations. See also Granular Content Control.

Trusted guard

[2435] An automated (software) application that performs reviews of security materials and authorizes release based on security policy.

Trusted Platform Module

[2436] A chip or infrastructure which stores digital keys, certificates and passwords that allow servers to "talk" to each other. The new specs include a new remote attestation feature that allows a third party to check to see if a box has been tampered with or compromised by a hacker attack or a virus. Another feature is key storage, in which encrypted keys help store data. Abbreviated as TPM.

Trusted Network Architecture

[2437] A functional design of a computer network infrastructure that provides end-to-end security over access, authentication, process, transmission, and data control. Abbreviated as TNA.

Trusted Network Infrastructure

See Trusted Network Architecture.

Unauthorized disclosure

[2438] The release and exposure of data without the expressed policy of the security system.

Underground economy

[2439] The network of people that trade in information.

Universal redaction

[2440] The process of granular removal of all content within a data stream that conforms to all Digital;Doors filters and sieves leaving a remainder that unconditionally conforms to unclassified distribution policy. See also generic redaction and selective redaction.

UnSecured Digitaldoors stream

[2441] A destination data stream that represents the Source Data Stream with the extracted objects; where the extracted objects are represented by the original data objects in full or partially; if partially, by nothing, placeholders, substitutions, replacements, barcodes, or serialized tags.

Unstructured data

[2442] 1. Any data stream that that apparently lacks any formalized layout defining the arrays, fields, format, the hierarchical tree, and the purpose of the data within. 2. Any data stream that lacks a taxonomy. However, data streams that appear to be unstructured, such as documents, e-mail, IP voice, video, web documents, and reports, frequently contain tags and metadata and so actually conform to a document object model and therefore are really structured data by definition 1. For the purposes of data mining, search, knowledge management, connecting-the-dots, and library sciences such data streams frequently lack categorization or a formal metadata taxonomy, and therefore are unstructured data by definition 2. By right, all unstructured data might have a structure when the appropriate context or concept is supplied. See also structured data.

View

[2443] A reconstituted source data stream in full or in part.

Viewing

[2444] The process of reconstituting a source data stream in full or in part.

Version

[2445] A redacted MLS or MILS document with items (selected for extraction and dispersion) included up to the allowable level of the defined MLS or MILS sensitivity level.

Versioning

[2446] The automatic or manual redaction and (re)labeling of a document into one, two, or many outputs according to its clearance—in order to enable information sharing with parties holding different and varied clearances.

Versions

[2447] The results from automatic or manual redaction and (re)labeling of a document into one, two, or many outputs according to its clearance—in order to enable information sharing with parties holding different and varied clearances.

Watermark

[2448] 1. A logo, image, or design embossed or added into the paper to indicate its source. 2. A proof attached or interleaved into an electronic file to indicate copyright, ownership, or original material.

Watermarking

[2449] The process of adding a watermark to a data stream (data, text, image, audio, or video, et cetera.)

Weak declassification

[2450] The process of separating or processing (Downgrading or Sanitizing) secret (top-secret, secret, and classified) documents into those which are definitely unclassified, which a reasonable degree of accuracy. See also Strong Declassification.

White list

[2451] A register of entities that are acceptable and require no additional attention. Contrast to black list. See safe list. See also gray list.

Winnowing

[2452] 1. A process of deleting important data from within data to obscure its message. 2. More

traditionally, it means the application of a checksum to validate the integrity of the data.

A.1 Creating An Information Infrastructure

[2453] With respect to unstructured data (including semi-structured data) in information files, the prime considerations in creating an information infrastructure involve a recognition of different service levels, different retention-destruction periods for the files and different protective measures imposed by legal, security and compliance requirements. Confidential data may involve the enterprise, its employees, its customers and/or third parties. Further, the enterprise should gather, control disseminate and utilize its intellectual property (IP) rights in and to the data in unstructured data files. The current mind set relative to an open informational ecosystem, wherein employees of an enterprise, at many levels, openly communicate with enterprise partners, customers, vendors, contractors and other third parties, results in many communications paths wherein confidential data and IP data is subject to dispersal, misuse and theft. Concern over hackers accessing such sensitive data, rogue insiders taking and misusing the data and accidental loss of data by authorized users requires that the enterprise build an information infrastructure to identify, contain and classify the confidential data and IP data.

[2454] Information Rights Management (IR Mgt) can and is presently imposed on data files with the use of data file metadata, tags or labels. For example, the file property may be set at "Read Only" which limits revision rights. However, the content in information files is not subject to many controls. IR Mgt (a) controls the open, print and edit functions; (b) limits access, permissions, and distribution; (c) establishes and permits changes to permission levels; and, (d) provides and audit trail. Since the value of the information changes over the life cycle of the information file, oftentimes due to events and objects outside of the information file itself, it is the content that should be managed and subjected to an information infrastructure.

[2455] The present invention employs an extraction engine which identifies, selects and then extracts the security sensitive content (sec-con) or select content (SC), a classification engine that assists the user (the enterprise's CIO) to classify and organize the content, and a vibrant testing cycle to continually challenge the sec-con store, the sec-con security levels and the quality, breadth, depth and scope of the SC store. The classification engines can be used by the Chief Information Officer ("CIO") to establish information infrastructure policies and re-set the extraction engines to enforce the policies (data acquisition and handling rules) established thereby.

A.2 Sampling and Creation of a Simple Info Infrastructure

[2456] As discussed above, mission critical data is oftentimes captured in structured data compilations such as databases and spreadsheets. Herein, a "spreadsheet" is included in the definition of a database because a spreadsheet is a two dimensional database. Therefore, confidential data in unstructured information files which matches some or all of the data in structure files represents the initial starting point to define mission critical enterprise data. Some of this some of this mission critical enterprise data represents secret or confidential data of various parties (the enterprise, its partners, customers, vendors, contractors and other third parties) and IP data of various parties. Data from the structured files initially can be used as a content filter to extract the same confidential information from unstructured files and store the security sensitive sec-con data in extract stores as explained later herein. A simple classification system (hierarchical taxonomic system) can be established by reviewing the label descriptions on the structured data and then expanding class definitions with the use of the knowledge expander KE search engine described later in section C.2 and C.22, among others. The data and information returned from the KE engine can be used by the CIO (system administrator) to create increasingly broader or higher levels of data classifications. The hierarchical taxonomic system can be

used to build contextual filters and taxonomic filters which can further protect sec-con data and expand the value and quantity of SC data. Once the CIO has established the classes of data, he or she can set policies for (a) storage and retention; (b) extraction - dispersal - access - release; and (c) levels of service and permission levels.

[2457] After removal and secure storage of known security sensitive data from unstructured data sources (using filters designed from structured data sources), a review of unknown data in unstructured files may be appropriate.

A.3 System Tools and Processing Methods for an Information Infrastructure

[2458] FIGs. A-1a and A-1b diagrammatically illustrate a basic system diagram, method and flow chart for an information infrastructure. The infrastructure is a dynamic process and hence the CIO or system administrator can enter the process at many points. Information (Info) infrastructure 10 includes a number of modules or process routines or functions which are described in detail in sections B.1 - D. 14. The prime operational modules or functions are set forth in FIG. A-1 (which is a composite of FIGs. A-1a and A-1b). The prime objective of the info infrastructure is to process data from source document or source information file 12. The source is first processed via an editor module 14 to deconstruct the source doc into its constituent parts. See sections C.14 - C.22, and more critically, section D.1.

[2459] After deconstruction (which may be optional per the system operator's choice), the infrastructure processes data with extraction module 15. Extracted data and remainder data is processed by secure storage Sec-Con module 16 and, if appropriate to the CIO's designed task, the select content SC storage module 17. The extraction and secure store is discussed throughout the specification but is particularly discussed in sections B.1 - B.6. The select content SC store module is discussed in sections

C.1 - C.13. C.20 - C. 22. The use of an editor module in sections C.14 - C.19 is used to process input data which, after such processing, is applied to the content filter modules, the context filter modules and the taxonomic filter modules. The system coordinates storage of data with map module 18.

[2460] Extraction module 15 handles input from content filter module 20, contextual filter module 21 and taxonomic filter module 22. The infrastructure system and process is adaptive therefore the extraction module also outputs data to filter modules 20, 21 and 22. Content filters are discussed in many places in the specification (see particularly sections B.4, and C.12). Since the extraction is employed with the (a) security system and (b) the knowledge expander KE system to generate select content SC data, this dual purpose both feeds data to the filters modules 20, 21 and 22 and accepts data therefrom to define the filters used to extract critical data from the data flow. A contextual filter module 21 is employed in an infrastructure security system in order to assure the CIO-system operator ("sys-admin") that (a) the source document (source doc) is properly scrubbed and (b) to add new words, terms or data objects to the list of security sensitive data objects of concern. See sections B.4, C.13, C.20. The new data objects become content filters for another processing cycle for a new source doc. When configured as a knowledge expander KE infrastructure for select content SC, contextual filters increase the SC data extracted form the source doc and increase the value and informational content of the SC store. The taxonomic filter ("tax-filter") permits the sys-admin to employ a higher conceptual layer on either the security extraction or the SC extraction and compilation. Tax-filters are discussed in sections C.1 - C.22 and particularly in sections C.1, C.11, C.14, C.21 and C.22.

[2461] The sys-admin may use the info infrastructure with an existing classification system and information policy, ruled-based or may have no pre-existing classification system for his or her documents. The infrastructure process 10 includes a classification generator module 24 which employs

an SC extractor, a KE search engine to generate additional supplemental documents (supple docs) which the sys-admin may use the create the hierarchical taxonomic classification system. otherwise, the sys-admin may purchase taxonomic classifications commercially from others. However, the sys-admin must ultimately prioritize the hierarchical taxonomic for the enterprise. The taxonomic system can be created by operation of the modules in sections C.1 through C.23. The KE expander module can be configured to list docs and data entries and this listing can be used by the sys-admin to (a) build the hierarchical taxonomic and (b) prioritize the classes. Once the hierarchical taxonomic classes are gathered, the sys-admin defines which tree, branch and leaf is critical in module 26, then defines (or redefines) security levels, sets secure stores, ids SC, sets SC stores and sets permission levels, service levels, access levels and other policies for the enterprise which policies are translated into a rules-based system for the processing of data. Module 32 applies the rules 32A, 32B, 32C, and 32D to filter modules 20, 21 and 22 and to the access-permission module or control function 42. Data requests 40 from others are input into access, permission, distribution, reconstruction control module 42. Features and functions for this module are discussed in sections B.7, B.8. C.9 - C.14, C.20 - C.23 and in the applications D.2 through D.14. The request for data may be an alarm function or may be a data request form another. The info infrastructure then cycles back to source doc 12, editor module 14 and extractor module 15.

[2462] To test the security levels and system, the sys-admin may activate test module 28. Section C.0 discusses how the KE search engine is employed to test the security system.

[2463] FIG. A-2 diagrammatically illustrates a basic information life cycle process or system wherein secure content (sec-con) and select content (SC) is separately stored per categorical filters operable by the enterprise. Data lifecycle 1 in general begins with an acquisition 4 and process or storage 5a in the enterprise distributed computing system or environment represented by data lifecycle

ring 3a and outer ring 3b. Storage memory units 6a - 8a are part of the distributed computer system of the enterprise. However, the enterprise may contract out the storage facilities. In any event, the enterprise controls storage 6a - 8a. Some storage units 7a - 8a may be combined. However, archive storage 8a may be less expensive (slower process and access time) than the SC storage units 7a - 7e. Further, the filter SC storage units 7a - 7e may be in a single memory, and simply segmented by control parameters. In general, as explained in sections B.1 - B.8 and others (see D.7 - D.9), the memory units 6a, 6b and 6c are distinct and access to these sec-con stores is controlled with software or hardware guards (see section A.23 or D.6) or the extract stores are physically separate. Some data stores may be copy protected.

[2464] Data on the right side of inner enterprise ring 3a is in transit 2a due to acquisition 4 and other high frequency access, distribution and processing aspects (service levels). A high service level may be assigned to data in transit 2a. Data in process section 2b is undergoing manipulation and processing by the enterprise. This may include the use of sec-con filters and extraction to top secret TS store 6a, secret S store 6b and classified C store 6c. Remainder data is kept in the enterprise store in ring 3a.

[2465] Enterprise designated categorical filters are used to store select content SC relative to the category in certain SC stores. These categories cover a wide range of topics which may be of interest to the enterprise. For example, enterprise designated filters screen data in process 2b for various enterprise policies including, but not limited to level of service policies, customer privacy policies, supplier privacy and proprietary informational policies (business privacy issues), enterprise human resource privacy policies, financial data handling policies (see GLB or Gramm-Leach-Bliley Act of 1999), public company reporting policies (see SOX or Sarbanes Oxley Act or Public Company

Accounting and Reform and Investor Protection Act of 2002), health care regulatory policies (see HIPPA or Health Insurance Portability and Accountability Act), technical innovation policies (intellectual property (IP) disclosures, patent disclosure systems, trade secret tracking systems), competitive reporting policies (a collection of all data on the enterprise's prime competitor) and document or data retention policies. The enterprise sets these policies and the CIO or sys-operator fashions filters which either continually or upon certain events or periodically sweep the enterprise data system, locate SC and store the SC in the relevant data bases. As a further example, technical information on a certain product may be important and, in light of the fact that 60 - 80 % of all corporate data is in unstructured documents, the data lifecycle sweep may filter all processed data looking for that certain product. The technical filter content file is expanded to include additional contextual materials by looking at outside or extra-enterprise data, thereby creating aggregated SC data for the special technical data SC filter. With the application of a taxonomic classification system, the product specific tech filter is further expanded to cover higher classifications of product types and lower, more detailed products that relate to the target product. A cross-check operation with IP SC filter and database 7e also provides guidance. If the IP is very valuable, it may be stored in sec-con extract stores 6a - 6c.

[2466] FIG. A-3 diagrammatically illustrates a simplistic dynamic filter and select content (SC) storage coupled to a data processing event. SC filter 45 may be one or more of a content-based filter, contextual filter or a taxonomic classification filter. The SC filter 45 accepts, in a first instance, data input 1A and, in a second instance, further data input 1B. As explained below, data 1A may be the same as data 1B. As an example, a sequential operation is described. Input data 1A is fed to SC filter 45. The SC filter identifies SC content SC-1A and this content is sent to SC database 47. Any type of storage may be used rather than a database, however, some type of indexing is useful. The SC content copied

from data input 1A is added "cc" to SC filter aggregated 48. The SC database uses the SC aggregated data to supplement the SC filter 45 with a summation + function via operational path 49.

[2467] The next process supplies further data input 1B to SC filter 45. The data 1B is processed by the filter 45 and select content SC-1B is copied "cc" to SC filter aggregation file 48 and also copied "cc" to the SC data file or structure 46. The SC aggregate is fed back to the SC filter via route 49. In this manner, the SC filter can be one or more content-based filters, contextual filters or taxonomic classification filters. This is the adaptive and dynamic filter for the categorical filtering discussed above. If data 1A is obtained outside of the enterprise, the SC filter expands. See KE expander system in sections C.2 - C.22. Outside or external data is primarily to expand the filter. It is stored in the SC aggregate file. If the input data 1A is internal or intra-enterprise data, both the SC aggregate file is expanded and the sys-operator needs to know that the SC is an internal document and this internal SC-1B is stored in SC data file 46.

[2468] Data 1B is also applied to control 52. The SC filter 45 generates a result command 51 or an output control signal to indicate the presence or absence of certain SC data in data block 1B. Contextual relevance or taxonomic key words or terms may also trigger filter result output 51. This signal is applied to control 52 which activates data process 54. Some data processes are copy, extract, distribute, archive, or destroy. Further, as discussed below, SC filter 45 may be set as a screening filter which is then used to identify a certain categorical filter which then triggers the data process. Alternatively, the data input 1A or 1B may come with a data process request and the data must be filtered or scrubbed before the data process is carried out by the enterprise computer system.

[2469] Some key operations of the system and process follow. The invention organizes and processes data important to an enterprise operating the distributed computing system. The SC or select

content is represented by one or more predetermined words, characters, images, data elements or data objects. The computing system has a plurality of select content data stores 7a - 7e for respective ones of a plurality of enterprise designated categorical filters which include content-based filters, contextual filters and taxonomic classification filters, all operatively coupled over a communications network. An archive filter 8a operates with data at rest 2c at the top left of lifecycle 1. Aged content is subject to archiving policies or back-up policies. A data input 1A or 1B is processed through at least one activated categorical SC filter 45 to obtain select content, and contextually associated select content and taxonomically associated select content (SC-1a) as aggregated select content. The aggregated select content is stored in the corresponding select content data store 47 and, more specifically, filter aggregate file 48. A data process 54 from the group of data processes including a copy process, a data extract process, a data archive process, a data distribution process and a data destruction process is associated with the activated categorical SC filter 45 and the method and system applies the associated data process 54 to a further data input 1B based upon a result 51 of that further data 1B being processed by the activated categorical SC filter 45 utilizing the aggregated select content data per path 49.

[2470] In another embodiment, the method and system extracts security sensitive content TS or S or C (FIG. A-2) from a data input to obtain extracted security sensitive data for a corresponding security level and remainder data. Remainder data can be considered left-over data which is extraneous to the sec-con or SC. It may be surplus or residue data. In this manner, the remainder data is complementary to the sec-con or SC data. This extracted security sensitive data is stored for the corresponding security level in a respective security sensitive extract store 6a, 6b or 6c. Remainder data is stored in the distributed computer system. Distribution of the data input 1A (FIG. A-3) with respect to said distributed computing system is limited based upon throughput from at least one of the enterprise

designated SC filter 45 (e.g., patient privacy or HIPPA filter linked to SC store 76). The controlled release of corresponding extracted security sensitive data from the respective extract TS or S or C stores 6a - 6c with the associated security clearances for corresponding security levels is permitted by the system. In this manner, security sensitive content is separately stored apart from the select content, which is stored in select content data stores 7a - 7e and 8a.

[2471] The sys-operator may activate ON a categorical filter or the system may be automatic. Prior to activating a designated categorical filter (see GLB SC content filter 7c), the data input 1B is processed through a screening taxonomical classification SC filter 45 which output 51 activates said at least one designated categorical filter for the sensed data input 1B. Several categorical filters may be activated, operating concurrently or sequentially, dependent upon processing speed and memory access, based upon the screening taxonomical classification filter output.

[2472] The further data input 1B may be the same as the initial data input 1A. In other words, an iterative process may be imposed on data 1A/1B. First, the SC filter 45 is used to expand the SC content, maybe to add contextual material, or to add a higher or lower classification SC data per the taxonomic filter. Thereafter, the data 1A/1B is processed again by SC filter 45 with the data process 54 associated therewith.

[2473] Regarding the initial data processing, the further data input 1B may be structured or unstructured data. If structured, the data input should be deconstructed and special attention should be paid to the structured items. In email, a partially structured data item, addressee and the addressor data and the regarding field often are key SC items for the enterprise. If unstructured data is filtered, a deconstruction process should be used. See sections C.14 and D.1, among others. The constituent data elements are then SC filtered by 45.

[2474] Activating the enterprise designated categorical SC filter 45 may be automatic or manual. An automatic activation can be is time-based (2:00 AM every day), distributed computer system condition-based (when data flow falls below x % or process speeds exceed y %), or event-based (see section D.3, D.11, and others which discuss event triggers). Activation of the categorical filter can be linked to a data process request from the group of data processes.

[2475] The data processes include a copy process, a data extract process, a data archive process, a data distribution process and a data destruction process, an inter-enterprise data distribution process, an intra-enterprise data distribution process, a data access request coupled with said data distribution process (a request to send data 1A to a newspaper via email (copy and distribute)), a data distribution denied process, and a limit data distribution process. The data process associated with a data block 1A/1B may trigger one or more SC filters or a discovery of SC in a data block 1B may cause a change in the data process from "full distribute" to limit distribution to key enterprise persons." In litigation, discussions of the lawsuit may be an SC filter topic. The groups of data processes may activate different categorical filters. The result 51 of the further data 1B processed by the activated categorical SC filter 45 indicates the presence or absence of select content or aggregated select content in the data 1B.

[2476] On the output side, the system may impose a classification taxonomy on the stored aggregated select content SC file 48. This enables the sys-op to see the scope of the filter (depth, detail, higher level classifications, orthogonal data fields, etc.). See, for example, section C.22. Preferably, the select content file 46 is stored separate from the aggregated select content 48 in the select content store 47. A separate storage permits the sys-op to see what SC data resides in his or her enterprise as compared with the scope of the filter. The system may impose a classification taxonomy onto the stored select content 46. See FIG. C-3 and C-5. Further, the stored SC 46 may be a copy of the obtained select

content from data 1A/1B or may be a representation thereof in said select content data store. A representation may be a marker or pointer to an index of SC. This process avoids multiple copies of SC storage. Additionally, in order to track the origin of the data 1A/1B, the data input has a source indicator (like a URL or the path to the original document location) and the obtained select content SC-1A or SC-1B, when stored in the select content data store 46, includes a corresponding representative source indicator pointing to or linking to the source of the data. In this manner, the sys-op knows where the SC data came from. Additionally, the process may generate a tag or label responsive to the source indicator of data 1A, and the corresponding representative source indicator. The tag or label can be electronically attached to the source document to indicate SC content, importance of the SC content and category thereof.

A.4 Information Dispersal - Basic

[2477] The dispersal of information into the secure storage or the select storage, with maps to the original or source document, is one of several important aspects of the information infrastructure. The storage processes and modules provide an alternative to prior art information dispersal algorithms. A prior art information dispersal algorithm is discussed in U.S. Patent No. 5,485,474 to Rabin and in an article entitled "Survivable Information Storage Systems" by J. Wylie M. Bigrigg, J. Strunk, G. Ganger, H. Kiliccote, and P. Khosla, published August, 2000 in COMPUTER, pp. 61-67. The present system of information dispersal can be additive to the prior art information dispersal algorithms. The present system and method can be automated to facilitate dispersal of information to remote storages upon detection of certain events. Lastly, the information dispersal techniques can provide for the dispersal of context of the whole data stream while maintaining access to the constituent parts of the content.

A.5 Alternative to Information Dispersal Algorithm

[2478] The invention provides a system and method for dispersal of plain text data and an alternative to usage of prior art information dispersal algorithms. The usage of prior art information dispersal algorithms does not permit the user or sys-admin to view or use data stored within a particular storage. The prior art algorithms permit one to view and have access to data only when data is retrieved from a few stores and then is combined together.

[2479] The present inventive process and system enables use of granular data which is which is stored in a known and accessible storage store (in defined storage locations (sec-con store or SC store). The security risk in the present system can be reduced by storing smaller and more granular pieces of data. Further, after splitting granular data from a source document or data stream, the granular pieces of data may be processed using the prior art information dispersal algorithms.

A.6 Adding Prior Art Information Dispersal Algorithms to Granular Data Dispersion

[2480] The present invention provides a system and method for dispersal of granular data to distributed storage stores and the possibility to add the prior art Information Dispersal Algorithms ("IDA"). The invention provides a system and method for dispersal of data to distributed storage stores – in essence an inherent " built in" security feature. The Information Dispersal Algorithm or IDA delivers a relatively high level of security, however it does not permit nor allow access to data content for search and other Knowledge Management operations.

[2481] On the other hand, the present invention may (a) first granularly extract the data form the source, then (b) use the IDA for securing extracted granular pieces of data, one at a time. This combination brings flexibility to the system as a whole since granular pieces can be reconstituted, one at a time, and released from IDA for knowledge management operations without compromising the

security of the whole document. By securing granular data pieces with the Information Dispersal Algorithm or IDA, the system's granular data parts once reconstituted by the IDA are available in system storage and are stand alone data structures - (encrypted or not). These stand alone data structures and the granular data therein can be read on their own without the need to bring together other data shares. Because extracts can be in plain text or decrypted- and stand in their own data structure, the sys-admin can authorize an advanced search and knowledge management operations through the granular data structure.

[2482] Splitting a data stream and dispersing the slices of the split data to distributed data stores creates a very high level of security. This emulates military tactics of splitting and hiding forces. In words of the great strategist, Sun Tzu, the present process and system creates "disorder from order" (the order being the source document, the disorder being the granular and dispersed data parts) and it creates a formless organization which makes it a very hard target for the attacker. The level of security may be further raised if, at the storage stores, there are strong access control mechanisms for authentication and validation of users. Dispersal of data to distributed storage stores achieves higher security as well as a lower overall security costs.

[2483] Distributed storage stores need less security then a centralized data repository for a number of reasons. First, the distributed storage stores hold only parts of the data and they are of lower interest to an attacker that will need to attack few dispersed stores to get the total content. Second, the stores are scattered and if hidden they call for less security. The need for less security means lower costs; more efficiency and less processing power. Thus dispersal of data to distributed storage stores is inherently "built in", " baked in " security. The split granular data pieces may be sent to distributed storage to be processed in an Information Dispersal Algorithm scheme. Each granular item will be split

into n pieces, which will be sent to distributed storage. Reconstructed of a granular item will be enabled by access to M out of N pieces. A directory/ map connects all granular pieces recreate a partial or original document. Using the Information Dispersal Algorithm (IDA) with the present invention may deliver a much higher degree of security. The data stroes may be correlated to the sec-con or SC content or maybe otherwise designated.

A.7 Automatic Dispersal of Data to Remote Storage Locations

[2484] The present invention provides a system and method for the automatic dispersal of data to remote storage locations. Examples of these systems and methods are found in sections D.2 through D.14. Any document data stream created by a user may be automatically split and transported to distributed storage locations. This may take several forms. One form is, when a file is being closed, it will automatically be split and its pieces sent automatically to distributed local or remote storage locations.

A.8 Dispersal - Breaking Down the Context of Data Streams while Maintaining Access to the Content of its Parts

[2485] The present invention provides a system and method for splitting a document or data stream into granular parts for security and privacy reasons which hides the data stream's context. By splitting a document or data stream into different granular pieces and dispersing the pieces into distributed storage, the context of the original document or data stream is broken. The more splits and slices imposed on the data, the harder it is to inference the content, concept and context of the source document or data stream. The uniqueness of this invention is that each distributed piece and its granular content can be independently accessed for different usage. This storage technique is very different than the Rabin Information Dispersal Algorithm (U.S. Patent No. 5,485,474) in which slices of the original

document are distributed in different storage locations but the content of each slice or piece of data is not accessible.

[2486] The present system and method permits the sys-admin to establish an info infrastructure for data mining a plurality of source documents or information files. See sections C.8, C.17, D.5 and D.13. The sys-admin can also employ the knowledge expander KE search engine to enhance the data mined form source docs. See sections C.2 through C.23.

A.9 Granular Tagging - Basic

[2487] The present method and system permits granular extraction and tagging of data. As a result, the system permits (a) automated granular data level manipulation; (b) a process for creation of automated granular tagging and labeling; (c) a process for creation of automated tagged tear lines; and (d) use and implementation of category tags as placeholders in databases.

A.10 Automated Granular Data Level Manipulation

[2488] The system and method provides for granular data manipulation. All input data is analyzed, optionally categorized and optionally presented in a matrix or other format (a data structure). The analysis involves deconstruction or decompilation of the source doc (see section C.14 - C. 16 and, more particularly, section D.1. Categorization is discussed in sections C.11 and C.21. Different tools enable selection of granular data items for different activities which may include, among other things, their use as: (1) key words to search engines; (2) headings for granular data elements, meta tags for such elements; (3) a base for build up of expansive content (supple docs); and (4) filling in missing information from documents (data mining third party sources.

[2489] Granular data control may be used for various information processing activities listed in the Action Table below. The Action Table, in one embodiment of the invention, is a pull-down menu which enables the user of the system to select the granular action associated with the activity.

Action Table

(1) security
(2) privacy
(3) compliance
(4) archiving
(5) replication
(6) backup
(7) continuity
(8) survivability

(9) broadcasting
(10) alerting

[2490] The granular actions are selected automatically (based upon the initial set-up of the system), manually or a combination of both. Granular items may be connected to other granular items in the matrix / format or connected to other data processing steps external to it. Different data matrixes or data structures in the sec-con store or SC store can be constructed for the granular output of the original, source input data. The original data ad constituent parts thereof may be replicated to form different matrixes based on different categorizations. The selected granular data items will usually have semantic meaning. However it is feasible to combine granular items in different ways which do not have an inherent semantic meaning. This might be done for search and content inference purposes. By categorizing and labeling every granular data element, the invention in essence differently weights different granular content items.

A.11 Process for Creation of Automated Granular Tagging and Labeling

[2491] The invention provides a system and method for automatic creation of tagging and labeling at the granular data level. See sections C.3, C.5, C.6, C.7 and D.1.There is a need for automation in tagging data and content in a document and data stream. the tagging and labeling can be used for security or for data mining or knowledge expansion. The need security tagging includes: Top Secret paragraph, Secret paragraph. Top Secret line, Secret line (tagging plus contextual filter set as a range (paragraph and line). There is a need for tagging at a more granular level such as Top secret word / character, or Secret word/ character. The need for tagging at a very granular level also is impressed upon an enterprise by the requirements of recent laws for removal of sensitive names and sources from documents. There is a need for other types of categorization / labeling / tagging. For example, there is

EX. 1002, Page 869

a need for a Private tag, Public tag, and a need for High Priority, Low Priority, Medium Priority tags. The enterprise policy discussed above, when applied with the use of the present method and system, implements this categorization / labeling / tagging operation. Each item, word, character, image in a data stream (also optionally at a bit and binary level) is monitored, analyzed, categorized and tagged. This is done by filtering, i.e., matching the granular data items against pre-set lists wherein a match yields the right classification tag for the granular data element. The element is marked by creation of a tag for the granular data element. For example "name," "location," and "social security number."

[2492] Following process outlines an automatic tagging system. (1) Within the input data stream, all the data and content element is monitored and categorized by a filter or filters. The filters may be content filters, like a "dirty word" list of others. (2) Every data element is categorized contextually (name location, phone etc). Every element may be also labeled, tagged, and categorized according to other categorizations like security classification, Top Secret. See, for example, the discussion of hierarchical taxonomic in section C.21. (3) Every tag may be checked contextual validity and may be checked and verified by an inference engine.

[2493] By categorizing and labeling every granular data element, the invention in essence different weights to the different granular content items.

A.12 Process for Creation of Automated Tagged Tear Lines

[2494] The invention provides a system and method for automatic creation of automated tear tagged lines or contextual ranges about identified content. There is a need for automation in classifying ranges of content within a document or data stream. The need may be for security ranges like Top Secret paragraph, Secret paragraph etc. as well as for other types of categorization / labeling/ tagging. For example, there may be a need for Private tag, Public tag, High, Low or Medium priority tags. Based on

automatic categorization of each item, word, character, image, in a data stream, a contextual defined range, such as a line a paragraph or a group of words, can be tagged by one of the tags in that grouping. For example, if a paragraph has 15 words with 15 tags, the paragraph may be tagged as a whole based on one or few of the tags. In situations where tags represent a hierarchy - ("Top secret, Secret, ..."), the paragraph can automatically be tagged by the highest level tag. Tagging of the paragraph sets a range or a tear line, which can be selected for different uses including a range for searching for key words for a search operation.

[2495] Following is a process for creating automatic tagging with automatic tear line tagger.(1) Select ranges of content (drop-down menu or pre-set range or manual highlight selection). (2) A range of content means a contiguous content stream, which may be a phrase, line, paragraph, page, chapter etc. (3) Within the defined range, every data element and content element is monitored and categorized by a filter or filters. The filters may be content filters like dirty word list of others. (4) Every data element is categorized contextually (name location, phone etc). (5) Every element may be also labeled tagged and categorized according to other categorizations like security classification. (6) After categorization of every element, a process may select a categorization labile to define the content range. (for example, a paragraph may have 10 words 5 labeled are as Top Secret, 3 labeled as Secret, and 2 words labeled as Confidential. (7) The range might be labeled, tagged as Top Secret. (8) Rules can be set so that when a combination of a few words appears the range, the range group is tagged at their level higher or lower. (9) Every tag's contextual validity maybe checked and verified by an inference engine. By categorizing and labeling every data element, the invention in essence assigns different weights to the different granular content items.

A.13 Category Tags as Placeholders in Databases

[2496] The invention provides for a system and method for substituting tags as placeholders for removed data thereby creating the info infrastructure for unstructured data. For security or privacy reasons, granular data elements are removed from a repository, database, document, or file, etc. Tags are inserted as placeholders to replace the data that was removed to other storage locations. Data mining searches locate the tags instead of data. The benefit of this operational system and process enables data mining that will not produce the removed secured data but it will let the user know that the data does exists and what type of data it is (name, location, social security number, etc). Data mining is possible based on tags organized by categories, not whole information in unstructured data files. As an example, a document may be processed to show: (a) Date birth xxxxxx; (b) Mothers name yyyyy; and (c) School zzzz. The data categories (birth, mother's name, school) is listed, but the granular data elements are in different locations - and the owner of data needs to trigger the release of different granular data elements.

[2497] Further, the tagging and sec-con and SC data extraction can be used to populate a taxonomic database (or non-taxonomic database). The tags or labels are the filed names in the database. The structured data in the database will be the extracted data. Therefore, the sys-admin creates the info infrastructure with the present process.

A.14 Controlled Release - Basic

[2498] The controlled release of information (see reconstruction aspects in sections B.7, B.8 and the applications in sections D.2 - D.14) results in the following unique features and characteristics: Layered build up of data streams, layer upon layer; Granular data release from a third party repository; Creation of modular data components from granular data streams; "Rolling" exposure of granular data streams; Controlled release of extracts after verification; Multi purpose usage of filtering and extraction process; Filtering and copying "Selected data;" and, A combined process of extracting part of the

"selected data" and copying parts of the "selected data."

A.15 Layered Build-up of Data Streams, Layer upon Layer

[2499] The present invention provides a system and method for the controlled release of granular data streams which release builds up, layer by layer, a reconstructed document or data stream. The controlled release of data or granular data streams - layer by layer - forms a new, whole or partial document. A granular data stream is defined as extracted, granular data elements and/or the remainder data filtered from an original data stream. The filter may process the original data stream by content filter or other filtering methods. Controlled release of granular data layers during reconstruction can be done in a document, data stream, and website, database or other format. Each layer is built upon the other one to enable a view of a full or partial document.

The invention provides a system and method for controlled release of data and granular data streams after verification and validation before the release of each layer.

[2500] One aspect of the present invention focuses on release from storage of granular data pieces. The security risk is reduced by storing smaller and more granular pieces. The attacker needs to access few stores to piece together all the content Layers of data of the original document data stream may be released at once or at different times. The granular data stream is released into a file with existing content (remainder, or sanitized file) or creating a new file. See section D.6 for a sanitized document reconstruction. For security reasons and to avoid covert channels, Trojans and other attacks, the granular data layers can be released onto different screens in the same computer or different screens on different computers. See sections B.7 and B.8. The build up can be done in a way that resembles build up of modular components. Different layers can be built up in different sequences.

A.16 Granular Data Release from Third Party Repository

[2501] The present invention provides a system and method for granular data control and release from a third party repository. Users control their own private data- release from 3rd party or government repositories and databases. When a user wants to release information, he or she will release it in a controlled release fashion. A higher level of security and privacy is delivered in situations where the data is split into data streams and stored in different memories and or locations. When the information is dispersed, the chances for leakage are minimal. The user may release different data streams from the different storage locations.

[2502] To deliver a higher level of security the user may chose to route the dispersed data streams in different communication paths. For example: deliver critical information in parts over the phone line, cell line, email and fax.

A.17 Creation of Modular Data Components from Granular Data Streams

[2503] The system and method provides for the creation of modular data components from granular data streams. Splitting original document or data streams into granular data streams enables the user or sys-admin to build and maintain a component infrastructure for internal usage (within the enterprise) or for external usage (to others). A granular data stream is defined as the extract and/or remainder data filtered from an original data stream. The filter may process the original data stream by content filter or other filtering methods.

[2504] The trend of implementing Service Oriented Architecture – SOA, in web applications teaches the need for modular code, the benefits of code reuse and other added value. In the same vein, splitting original documents or data streams into granular pieces and creating a modular framework brings different added value and advantages which include: (1) parts of the data stream can be used by the originator, other parts by other users or machines to construct other data streams; (2) the granular

data streams may be used as tags or metatags; and, (3) the granular data pieces may be used as elements to feed search engines. Modular data components may be granular data streams or pieces, granular data pieces with hooks, granular data pieces with tags etc. Hooks are metadata that relate one data element to another data element or file or data object. The modularity enables data pieces to be easily combined into a larger data structure where they may be reused.

A.18 "Rolling" Exposure of Granular Data Streams

[2505] The invention provides a system and method for flexible content access based on "rolling" granular data exposure with decryption for better workflow. In situations where information must be accessed on an ongoing basis in plain text like search operations, encryption may be a major hindrance. The need to decrypt the encrypted plain text and then to encrypt it again, and do the same loop many times, is a major obstacle to work flow and performance.

The invention introduces a solution based on creation of (1) granular pieces of data (2) a distributed storage framework as a way to deal with the need to encrypt yet not overwhelm the processing and other computing workflow. The system creates granular data pieces out of the original document / data stream. This is done through a process of content analysis, selection, extraction and dispersal to distributed storage.

[2506] The granular pieces may be encrypted. The added value of the granularity effect is that there is no need to decrypt the whole original document at once pieces of granular data can be decrypted separately at the different storage locations.

[2507] Operationally, the system may be pre-set to encrypt and/or granularly extract and store based upon (i) event, e.g., save doc, log OFF, or (ii) time function; or (iii) user activated switch for encrypt and/or granularly extract and store.

A.19 Controlled Release of Extracts after Verification

[2508] The invention provides a system and method for controlled release of extracts after verification and validation. See sections B.7 and B.8 and the applications in sections D.2 - D.14. One aspect of the invention focuses on release from storage of granular data pieces. The security risk is reduced by storing smaller and more granular pieces. The attacker needs to access few stores to piece together all the content. It in noted that after their splitting from a document or data stream, granular pieces of data may be processed using the Information Dispersal Algorithm.

The invention provides a system and method for controlled release of extracts after verification and validation. Controlled release of extracts will take place at once or in stages only after verification and validation of different identifiers including any of the following or their combinations thereof: Identification code (PIN); Biometrics; Past history; Chain of title of data extracts; Past projects user was involved in; Current projects user is involved with; and People user is involved with.

A.20 Multi Purpose Usage of Filtering and Extraction Process

[2509] The present invention provides a system and method for multi purpose usage of the process of: filtering the elements in the data stream, the selected data is extracted, and dispersed to memories and reconstructing process – including an interface therefor. See, for example, section D.13 and the web-based process. The process of document or data stream filtering, extracting selected data, dispersing to memories and reconstructing process is created for different purposes. These purposes may be presented to the user in a pull down menu. The purposes include: (1) Archiving; (2) Compliance with policy and regulations; (3) Security; (4) Data searching / data mining; (5) Survivability; (6) Continuity of operations; (7) Replication/Copying; (8) Backup storage; and, (9) Information sharing. All potential applications or part of them can be presented to users in various interfaces including a drop down menu

or in a different menu including a mouse menu. Presentation of a key or identity card including credit card etc. may be requested to enable access to manual or automatic processing of selected choices. In each case, the original document or original input data stream data is being analyzed at the content level.

[2510] The process includes:(a) the original data or data stream is split into different levels / pieces of data, created with the appropriate tagging and extracted to different memories. (b) split - the different data pieces are transferred to different memories, which will allow reconstruction or multi level reconstruction.

A.21 Filtering and Copying "Selected Data"

[2511] The present invention provides a system and method for of filtering the elements in the data stream, resulting in selected data, creating copies of the selected data and dispersing the copies to memories and reconstructing process – including a user selectable interface for activating such operation. The process of document or data stream filtering, extracting selected data, dispersing to memories and reconstructing process created for different purposes includes:

(a) in each case, the original document or original input data stream data is being analyzed at the content level; (b) the selected pieces of data are copied then (c) appropriately tagged then (d) the different pieces are extracted to different memory stores, which (e) will allow reconstruction or multi level reconstruction.

A.22 Combined Process of Extracting Part of the " Selected Data" and Copying Parts of the "Selected Data"

[2512] The invention includes a system and method for multi purpose usage of the process of: (a) filtering the data elements in the data stream, (b) resulting in selected data, (c) extracting and removing part of the selected data while (d) copying other parts of the selected data for further

processing. Copying selected data, extracting data, dispersing to memories and reconstructing process – are all actuated by a user controlled interface - such as a pull down menu.

[2513] For each of the purposes specified below, the original document or original input data stream data is being analyzed at the content level. The selected pieces of data are either: (1) Copied then appropriately tagged then the different pieces are extracted to different memory stores, which will allow reconstruction or multi level reconstruction. Or (2) Split from the original document the content pieces may be appropriately tagged then the different pieces are extracted to different memory stores, which will allow reconstruction or multi level reconstruction. Presentation of a key or identity card (including credit card data) to enable automatic processing of selected choices.

A.23 Data Mining Search Results Basic - Distribution as Multi Level Security Versions

[2514] The present invention can be configured as a data mining tool for search results which are distributed in multi level security versions, a search- creating automatically key words, a continuous data mining with query refinement; an automatic continuous search – with divergence and convergence, and with distributed data mining.

[2515] The present invention provides a system and method for search results distributed as multi level versions. The KE engine (section C.2 - C.23) can be combined with the multi-level security system per sections D.4 - D.9. There is a need to enable a user with a security clearance level and a specific role to search documents, data streams, data bases and data repositories that are at varying security levels which he is not allowed to access. The problem can be overcome by allowing a search to takes place in the different databases. This invention provides a system and method for multi level scrubbing and distribution of multi level search results

[2516] Search takes place in all data bases- a multi level security guard is positioned at gateways

scrubbing all results generated from the different sources/ databases. User gets results scrubbed to his level of access. The system can be positioned as a guard at the gateway of the search process. The system scrubs the search results or scrubs a copy of the results in such a manner that every user gets a version of the resulting document according to his clearance level, or role. The results of the search are (1) sanitized and versions are created that correspond to the user security level and role or (2) a sanitized remainder document is prepared and granular data will be released into it from storage store or stores, local or remote.

[2517] Presented herein is a multi level filtering and dissemination system and method of search results and search hits. The searching takes place in different security level databases. In situations where the search is conducted by a search engine that delivers at first summaries or hits of documents, the following process will take place: (A) The result / hits pages are opened. (B) Each page is filtered; (C) Selected words are tagged and extracted; and (D) MLS – multi level security is enforced for distribution to the users. Each user sees only what his security clearance level permits him to see. The summary of the search hits may also filtered creating words that are tagged by different security clearances. Each user then will see words in the summary of hits that he is allowed to access at his clearance level

A.24 Data Mining - A Search Creating Automatically Key Words

[2518] The present invention provides a system and method for automated search engine with automated creation of key words. Reference is made to sections C.2 - C.23, particularly C.9, C.10, and C.22, C.23. Documents are the core of the" knowledge economy." Finding documents in an organization or enterprise is a tough challenge because of the ocean of information. The challenge is even harder trying to find the needed information outside the organization, including searching open

source information on the Internet. Since it is hard to locate relevant information studies show that more than 50% of data existing in the organizations is recreated. A filtering system is presented herein to remove and to locate relevant documents / data streams from daunting volume of documents in an organization as well as open source data.

[2519] The system and process can be pre-set to automatically trigger a keyword search in which the keyword / search items are not submitted by a person into the search engine, but submitted by an automatic machine process after a body of content, like a document / paragraph or other data stream, is fed into the machine. The machine analyzes the content of the data stream and, based on preset filtering processes, locates the key words and submits them to a search engine or multiple search engines. Different key words may be added to the search that are not included in the processed data stream. The triggering starting point: A search may be triggered by a user dragging a document / paragraph / line into the search engine, copying content into the search engine, or through an automatic feed of data into the search engine, for example an automatic feed of news articles. Selection of search key words: The system analyses every word character icon image and categorizes them. The system selects the ones, which match the predetermined rule set and routes them to a search engine for a search. Words, characters and icons images that are not recognized by the filtering and categorization process may be categorized as "unknowns" and sent as key words to the search engine. Reduction of relevant search results: Results of the search may be processed (by use of a filter or filters) to reduce noise and deliver most relevant results.

A.25 Data Mining - Continuous Data Mining with Query Refinement

[2520] The present invention provides a system and method for automated non- stop data mining cycles with query refinement. Data mining triggered by key word or document feeding into search

engine. Search results are filtered again to find new keywords so another search will take place. As results are submitted they are automatically filtered by preset filters and new keywords are located and sent to search engine. Selection of key words: All key words may be sent to search engine. It is more practical to implement an n word automatic (possibly manual) selection of most relevant keywords for the search process. Different filters may be used. Categorization filters may be used- if data in originating document (or other process related documents) where granularly categorized – selection of specific categories may be enforced in the filter. For example: "select only names and locations"- This will reduce the number of key words only to those, which are names and locations. Other filters (sieves) may be used for example; "select only items with a high sensitivity level tagging " then only items such as Top Secret will be selected. Selection of range for key words: To focus on creating key words for search from specific content area within the data stream ranges of operation will be set. The range defines an area within a document; data stream within that area key words for search will be located, selected and fed to search engines. Ranges can be created automatically based on categorization of items within the data stream.

[2521] The inventive system and process may also be set for dispersal to storage of a federated data search / mining operation. Filtering options and combinations for sending keywords to search engine include: (1) Word list- a dirty list filter; (2) Group word list- a list of groups of words representing some context possibly a concept; (3) Categories- filter based on selection of specific pre-selected categories (like names, locations); (4) Security level – a filter based on selecting a specific security level (Top Secret tagged content or Secret tagged content); (5) Range – A meta filter determining what content area in a data stream would be the focus of the search.(tear line); (6) Scope- A filter selecting what extracted words in the selected range; and, (7) Other key words for a joint search

may be presented.

A.26 Automatic Continuous Search – with Divergence and Convergence

[2522] The present invention provides a system and method for storage with an automatic data search / mining operations (convergence). Automatic data search and mining operations in a data store is triggered by data residing within the data store. Keywords are selected and sent to search engine or engines. Data search mining is triggered by key word or document fed into search engine. Search results are filtered again to find new keywords so another search will take place. Other inputs triggering an automatic search may be sent from other sources including other storage stores, thereby creating a federated search. The data search mining operation may continue on an ongoing basis. The user has the ability to set the system ON for a continuous, non-stop cycle of filtering keywords and feeding them to search engines. The user may set the system ON for a set number q of search cycles. (e.g., stop after 5 cycles). The user has an ability to control cycles by establishing a time (cycles stop after 30 minutes). All results are filtered to reduce noise and achieve a high level of relevancy. Results may be matched to words in the originating data stream to insure relevancy. Data search results or other input of the storage store may be sold, bartered exchanged with others.

[2523] Currently, data storage is cheap yet there is minimal usage of data in storage. When data in storage is used, the data has added value for the enterprise. The system data mines unstructured data and enables the enterprise to sell search results, or combine newly discovered data with enterprise data, for a synergistic resultant data of higher value. The sale or barter of data may be for a price. A.27 Distributed Data Mining

[2524] The present invention provides a system and method for distributed data mining. Running search operations from dispersed locations both for security as well as better coverage and

expansion of select content SC data. For security reasons, searches should take place from different dispersed locations, in order not to enable a compromise of the data. (for example a compromised insider at Google can know all the subjects that are searched by a person or organization).

A.28 Content Inference Engine - Basic

[2525] The present invention can be configured as a content inference engine with the following features: (a) an inference engine searching for the unknowns; (b) securing information sharing with an inference engine; (c) an automatic web data inference crawler robots; (d) changeable data classifications based on content inference threats; (e) search operations creating new continuously growing digital data streams; (f) distributed searches from different storage node; and, (g) storage of separated data pieces in one location

A.29An Inference Engine - Searching for the Unknowns

[2526] The present invention provides a system and method for Data Inference Engine with Divergence data search focusing on "Unknown" data, that is, data which has not been identified by the initial set of filters used to process the source doc or information file. The objective is to locate hidden data and to infer data therefrom that is identifiable and relevant. The system locates documents, which the are related to the initial document, each other by context or concept. The search takes place in structured as well as unstructured documents. See the KE engine in sections C.2 - C.13 and C.20 - C.23. Data search mining is triggered by key word or document feeding into search engine. Search results are filtered again to find new keywords so another search will take place. Known key words are removed by the filters or given low priority- unknown keywords are sent to the search engine.

[2527] The user can set the system ON for continuous, non-stop cycle of filtering keywords and feeding them to search engines, set number of search cycles (e.g., stop after 5 cycles), and control cycles

by establishing a time period. All results are filtered to reduce noise and achieve a high level of relevancy. Results may be matched to words in the originating data stream to insure relevancy.

A.30 Securing Information Sharing with an Inference Engine

The invention provides a system and method to establish an information sharing environment counter-balanced by inference searching. Information sharing environments may be open to security risks therefore it must be counter-balanced by content inference searching and testing operations that will enable the sys-admin to identify a threat so that the security threat can be reduced. Therefore, search engines / inference engines must work constantly in a plain text environment. Different information sharing projects have been shut down because the data acquisition systems where a threat to privacy. Data mining is a constant threat to privacy and leads to many instances of identity theft. This invention provides a system and method for protecting privacy and security in data mining and information sharing operations.

[2528] The invention protects privacy and security by controlling the access to sensitive content. The sensitive information is defined by the inference engine. Documents and data streams are filtered by the inference engine, granular data is selected, (and may be extracted to distributed stores). Granular pieces of data are released by a controlled mechanism to avoid security and privacy breaches.

A.31 Automatic Web Data Inference Crawler Robots

[2529] The present system is a method for automatic crawling robots- spiders locating data and locating threats to data and minimizing the threats including scrubbing web sites. Automatic agents, robots, mobile code analyze data in documents, data bases and data streams. Using inference capabilities, which may include artificial intelligence, the system locates users' private data which may lead to a threat of compromising his other data his privacy etc.

[2530] The system can be configured to remove the threatening data. Removal of threatening data with the option of reconstructing it back upon presentation of proper identification and security clearance. Automatic robots, mobile code, may be used to crawl web sites, data centers, directories, and data streams on a continuous or part time basis. Consumer reports may be generated and sent confidentially to users. Thereafter, the user may demand that the web sites, data centers, directories, and data streams maintain user data in a confidential manner consistent with data privacy laws.

A.32 Changeable Data Classifications Based on Content Inference Threats

[2531] The present invention provides a system and method for flexible changeable classifications and security classifications, with optional inference penetration testing. A system and method enable the user to change security classifications on the fly. Security levels of a document or data stream are upgraded or down graded based on the results of inference tests.

Due to the overwhelming ocean of data in an organization and in open source data, data classifications, and especially security classification, should be changed on the fly as new data is being integrated discovered and fed from search results. What was "secret" in the past (and important and critical) may now be public information and the value of the data may be lost. Data that was classified at a low priority may need to be upgraded as new information is analyzed. The system can be automated to change classification levels. Classification levels change and tags or labels change as well as security classifications based on results of content inferencing penetration testing. This present invention provides a system and method for on the fly change of classification including security classifications: (A) Data that is classified at a specific security level for example "Secret ' may be used by an attacker an insider or outsider to search that classified data against open source search engines or other databases. The results of such an inference attack may lead to new information about the classified document that

is more secretive than the security level assigned to it. For example, an inference search might yield names and locations that where extracted from the "Secret " document in such a case the base classified document should be classified at a higher security level than "Secret." The system presented will conduct inference penetration test based on the level of sensitivity of the located information the security clearance level will be upgraded.

[2532] Since data is constantly changing, real time analysis must take place on an ongoing 24 hours effort. The user requesting access to the data will receive data not based on the old security clearances of the document but based on the up to the minute update. The system and method may be automated based upon enterprise informational policies.

A.33 Search Operations Creating New Continuously Growing Digital Data Streams

[2533] The present invention provides a system and method for automatic data mining operations creating new intelligent digital data streams / entities. Automatic data search mining operations create a wealth of information. A wide net is cast and many documents data streams are located. A filtering process is used to deliver most relevant results. On going search operations create a digital data growth which can be described as a growing tree with branches that are constantly growing and branching out. In essence new intelligent digital data streams are created on the fly around a "content target." The term "content target" is defined as a "concept" which may grow in different directions based on results of the data search operations. The invention provides a system and method for automatic addition of relevant search results to the "content target." Automatic scanning of data within a computer, data base, or network nodes results in analysis of content. Identification of, among other things,: (1) what are the critical elements of the data and / or; (2) what elements of the data need more relevant data including updated data. Then, the system automatically searches for relevant data to update or to add to the

targeted data. The addition of data can be in the form of an attachment, a URL, a meta tag or data in different formats or linked by different methodologies. The analysis process can be done in various ways including the use of content filters.

A.34 Distributed Searches from Different Storage Nodes

[2534] The present invention provides a system and method for distributed searches and establishes an interaction between distributed storage stores with data mining operations. Distributed storage stores each individually carrying data mining operations can yield a vast body of distributed data and knowledge. An infrastructure for interaction between stores is presented herein. The infrastructure enables sharing of data sharing of key words in order to enable better more comprehensive data search operations. For example, if one store stores data on a subject A and another store stores data on subject B, sharing of key words or data between the stores will enable better key word searches. In essence, searching in different stores (each one with its own subject matter) results in more robust search results.

A.35 Storage of Separated Data Pieces in One Location

[2535] The invention provides a system and method for storing different data extracts in one storage location. The system and method stores extracts of a data stream in different memories within one storage location. There is a major difference between splitting a document or a data stream and placing its parts in one storage location and this invention, which deals with placing extracts of a document or a data stream in one storage location. This invention deals in a situation that a whole data asset was already parsed – and split into a "remainder" and "extracts". What is transferred to one storage location is not all the pieces of a whole document or data assets but partial part of the whole the "extracts." The storage location can be one computer one storage node or other medium that is located

in one physical location. To overcome security risks, the extracts in the different memories within the one storage location may contain protective elements which make it very hard to match the different extracted pieces together to create an original data stream. Those protective elements may include adding a substantial amount of additional data assets that will create chaos. It will be a major obstacle for an attacker to locate the original extracts and piece them together. Another protective measure is to encrypt each of the extracts in the different memories and store their encryption keys in a different storage location. An attacker who will penetrate the extract storage location will need the encryption keys in order to access the content. The encryption key or keys may be split in a secret splitting scheme. In essence, m out of n keys will be needed to access the content of the extracts in the different memories. The storage location can be also software storage without a physically defined location.

A.36 Storage of Granular Data Streams - Basic

[2536] The present system and method can provide for configuration of granular data streams and dispersal to distributed storage stores; shuttling of data and granular data streams between distributed storage stores; replication of granular data streams with minimal security risk; and, anonymizing classified and unclassified storage stores.

A.37 Configuration of Granular Data Streams Dispersal to Distributed Storage Stores

[2537] The present invention provides a system and method for control of data extracts in dispersed storage environment. A granular data stream is defined as the extract and/or remainder data filtered from an original data stream. The filter may process the original data stream by content filter or other filtering methods. Granular data streams are filtered and the results or parts thereof form bigger data streams. The resulting granular data streams may be extracted or copied for further processing and storage. To achieve high levels of efficiency as well as security, a set of configurations is required to

enable maximization of benefits in processing and storing granular data streams a distributed environment. Since granular data streams are part of a bigger data stream- the granular data stream location need to be entered into a map or directory of the full data stream to maintain context. Configurations-may include: how many storage locations are available, what would be transferred to each storage location, and when. The storage nodes may be simple storage nodes, or smart processing storage nodes. Enterprise policies set data processing rules for the info infrastructure based upon who, where, what, when, why, and how much-how little (quantity).

A.38 Granular Data Stream Transport

[2538] The configuration of granular data streams transport to storage may include a selection of what type of data streams will be sent to which storage. The selection may be done randomly, or according to the types and or categories of content, specific key words in the granular data strings are going to be shuttled. The data storage facility may be selected at random or via an algorithm to enhance security. Storage by data class is discussed later. The configuration will also include locations of storage on the network. Special configurations will need to handle searches combined from key words in the different extracts, which might be in the same storage node or on other nodes. The configurations will cover the different possible permutations of combinations of which key words in the extracts may be sent as a combination to search operations and in what order.

A.39 Shuttling Data and Granular Data Streams Between Distributed Storage Stores

[2539] The present invention provides a system and method for data asset interaction in a distributed network. Shuttling of data, including granular data streams, between distributed storage locations for data mining purposes (and other purposes such as add on to other data streams) causes chaos which increases security. Creating network paths / "road infrastructure" between storage stores

so data and granular streams can move (1) when requested, or (2) on a scheduled basis, or (3) upon certain events (see attack application in section D.3).

[2540] The method and system also processes data and granular data streams from different distributed stores. The combined processing of the data may include cleaning the data, extracting, and packaging it for data mining search operations and other purposes. Artificial intelligence may be used in this process. Creating a "chamber" or "compartment " for processing data and granular data streams from different distributed stores is possible with the system. The combined processing of the data may include cleaning the data, extracting, and packaging it for data mining search operations and other purposes. Artificial intelligence may be employed.

A.40 Replication of Granular Data Streams with Minimal Security Risk

[2541] The present invention provides a system and method for granular replication enabling the user access to many copies with minimal security risk. It is common practice that to achieve reliability there is a need to store a few copies of the data. The problem is that the bigger the number of copies made the bigger is the security risk. The invention splits a data stream into granular pieces replicates those pieces and disperses them to distributed storage. The security risk is minimized and can be controlling the size of the granular data pieces and their number.

[2542] The invention also proposes a new architecture for storage on the internet. The invention enables a user to make as many copies as he wants of a document or data stream with minimal amount of security risk If a storage node is attacked a small granular piece will not pose a serious threat. A small granular piece does not convey all the substance of the original document/ data stream. If the replicated piece is small enough the attacker will find it useless because it is out of context. For example, a granular piece of data which is a name only can't create a serious threat because it is out of context. Other stores

need to be attacked successfully to access their data to give context to the small granular data piece. The security risk of having many copies can be reduced by the user decreasing the size of the granular pieces and dispersing the different pieces to different distributed storage store.

A.41 Anonymizing Classified and Unclassified Storage Stores

[2543] The invention also provides a system and method for not labeled, non security classified security storage stores, in essence. incognito storage stores. The storage stores may hold highly sensitive secret data but are not labeled with a security label or identifiable risky tag to avoid attacks. The classification of the storage store may be recorded at a distant location, which can be a registry, an index or a map. Data or granular pieces of data (split from a data stream) are stored in dispersed storage locations including Internet URLs. The different storage locations are not labeled or not labeled in a way that will lead to a possible attack. Sensitive storage stores that are commonly tagged with security or hardened classification at their physical location, like Top Secret or Secret, will not be tagged, tagged in a way that maintains their anonymity. For example, the data is labeled or tagged with a number, in a way that its content will be unknown to potential attackers.

A.42 Controlled Release of Granular Data Streams - Granular Data Manipulation and Control

[2544] The present invention provides a system and method for granular data control and manipulation. Granular data control is defined as: the ability to manage manipulate and control granular pieces of data within a data stream or information file in order to achieve certain results. The result of such a process is the creation of new granular data streams or structured data. A granular data stream is defined as the extract data and/or the remainder data filtered from an original data stream. The filter may process the original data stream by content filter or other filtering methods. The resulting data or data stream can now be used or manipulated for different purposes. The resulting data can be presented

in different formats to enable maximum data manipulation. For example, in one format, the resulting filtered data can be put into a matrix as structured data. Assuming a content filter was employed, and the filter also included tagging or labeling for categories (taxonomic filter), one axis of the data matrix may represent source doc markers and placement markers, the other axis may represent category groups (names, locations, and social security numbers), and the data in the matrix shows the critical sec-con or SC data, that is the granular filtered data elements form the source doc.

[2545] The granular content elements in the matrix can be manipulated manually or automatically based on pre-set rules and or configurations. Granular data manipulation and control enables the user to achieve different tasks including: (a) Privacy; (b) Security; (c) Secured Information sharing- Multi level security / cross domain data sharing; (d) Compliance; (e) Survivability; (f) Continuity; (g) Secured backup; (h) Granular data replication; (i) High relevancy searches; and, (j) Creation of tags and meta tags.

[2546] The process for creating the info infrastructure described above in this Part A (A.1 - A.3) establishes this granular data control and manipulation and use of the system tools to create the structured data and to implement the enterprise policies for tasks a - j with pre-set data processing rules.

A.43 Sanitization Process with Inference Filtering - Basic

[2547] The present method ad system can implement a sanitization process associated with inference filtering of the source doc of information file. This process includes (a) process for sanitizing data on the network for privacy security and survivability; (b) automated sanitization of important data; (c) sanitization of non-important data; (d) release of granular data streams in lieu of access identifiers; (e) release of granular data streams in lieu of consideration; and, (f) secured release of granular data streams to multiple display screens.

A.44 Process for Sanitizing Data on the Network for Privacy Security and Survivability

[2548] The present invention provides a system and method for sanitizing data on the network for different purposes including: privacy security, compliance, information sharing and survivability. The sanitization process includes the following process modules: (1) First stage: filtering the input, original data stream, and locating important content items (sec-con or SC data) that needs to be masked or extracted. (2) Second stage: implementing an inference process to verify if anything can be inferenced from the sanitized data stream.(3) Optional third stage: masking or extracting more data items from the sanitized data stream based on the threat results (inference analysis) of the content inference process. The risk analysis can employ the risk modules from U.S. Patent application serial no. 10/396,088 filed March 25, 2003 which is incorporated herein by reference thereto. Alternatively, the first level output report in FIG. C-4 and the tree or second level output report in FIG. C-5 for the life cycle engine may be employed to assess risk by inference. See sections C.9 - C.13 and C.23.

[2549] The user's network and data storage facilities need to be sanitized to defend against different types of attacks. Granular data items are extracted from the network and stored at dispersed storage locations. See sections B.1 - B.6; see also, sections D.3 - D.6 and D.11. The objective is to make the network " formless" to achieve, from order, a creation of disorder, so the attacker does not know what and where to attack. "Formlessness" -is the answer against virus attacks, Trojans worms / inference attacks, covert channels and insider attacks. The theory is to erase the user's digital foot prints while finding and inferencing the competitor's footprints.

[2550] The extracted granular data streams may be dispersed to storage or to distributed storage stores. The dispersed granular data streams will be tapped or retrieved for (1) reconstruction of the document (full or partial reconstruction); and, (2) to allow more data search needs.

A.45 Automated Sanitization of Important Data

[2551] The present invention presents a method and system for automated sanitization of important content (sec-con or SC data). Data, content including data, and content over the network must be sanitized on ongoing basis to stop insiders attacks, Trojan and spyware, content inference attacks and other damage. Minimal information is exposed with the present system and methods. Optionally, upon presentation of identification by users, layers of data will be reconstructed. For higher security, information search operations need to take place in distributed storage stores / nodes, and users work mostly offline.

A.46 Sanitization of Non-Important Data

[2552] The invention provides a system and method for automatic removal of non-relevant and unneeded content from information files. Automatic scanning and analysis of data within a computer, database, or network nodes is processed. Identifying data that might be not be relevant or needed is located. Extracting that data and removing it to a storage location or assignment of destruction - deletion codes is part of the info infrastructure. Data that is not relevant or not needed may include: (a) out-dated data; (b) data not relevant to the core document or data stream; (c) data not relevant to the user's organizations task and mission; (d) data files having mistakes therein; (e) data that takes too much space; and (f) data that may confuse the users or others. The data removed to storage (extracted and remainder data) may be indexed for reconstruction at a later date.

A.47 Release of Granular Data Streams in With Increasing Levels of Access Identifiers

[2553] The present invention provides a system and method for controlled release of layers of data for different access identifiers. A granular data stream is defined as extracted data and / or the remainder data filtered from an original data stream. The filter may process the original data stream by

content filter or other filtering methods. Controlled release of layers of granular data streams filtered from an original data stream is permitted only after different access identifiers are presented. See sections B.7 - B.8 and Part A herein above. As an example: release first granular layer like a paragraph after presentation of fingerprint, release of second paragraph after presentation of pin number, third paragraph released for retina scan, fourth paragraph released by secret sharing scheme need -- other person to be online with his identification to enable a release.

A.48 Release of Granular Data Streams After Obtaining Consideration

[2554] The present invention provides a system and method controlled release of layers of granular data for different considerations received by the disclosing party or enterprise. A granular data stream is defined as the extract data and / or the remainder data filtered from an original data stream. The filter may process the original data stream by content filter or other filtering methods. Controlled release of granular data streams including data pieces and data extracts is provided layer by layer, after verification different criteria including: identity, past history, verification of chain of title of the granular data stream, past projects user was involved, current projects etc. Release layer by layer for consideration obtained by disclosing party. For example- release first paragraph if user inputs name and social security number. Release second paragraph if requester inputs phone number. Release third paragraph if user inputs name of friends or delivers what he is being asked to deliver. This invention provides also a system and method for controlled release of layers of data or granular data streams for different monetary considerations. See sections D.5 and D.6. Controlled release of layers of data for different monetary considerations. For example: Release of first granular layer like a paragraph after payment of $1.50, release of second paragraph after payment of $2.00, third paragraph released for payment of $25.00, and fourth paragraph released after payment of a negotiated price. This invention

also provides a system and method Controlled release of granular data layers for a combination of different considerations. Controlled release of granular data layers for a combination of different considerations. Each granular layer may be released for a different type of consideration. For example: Release first paragraph for a payment, release second paragraph after biometrics identification, release third paragraph after – filling some information.

A.49 Secured Release of Granular Data Streams into Multiple Screens

[2555] The present invention provides a system and method release of granular data layers different computer screens. A granular data stream is defined as the extract data and / or remainder data filtered from an original data stream. The filter may process the original data stream by content filter or other filtering methods. To avoid Trojans, spy ware, and other attacks, granular data layers of original data stream are released into different screens in one or in separate computers. For example: Top Secret data layers will appear on a screen of one computer while rest of data layers (sanitized document) will appear on another computer screen. See sections B.7 - B.8. This invention also presents an overlay of granular data streams stored in different memories which memories can be in one physical machine or archived to a higher level of security in different machines.

A.50 Ecosystem Based on "Granular Data Control" - Basic

[2556] The present method and system can be used to establish an ecosystem with the following features: (a) an ecosystem for information superiority and survivability- based on "granular data control;" (b) a system and method for minimizing data exposure; and (c) creating " digital doors" in a data stream limiting access the content of the stream or information file.

A.51 Eco System for Information Superiority and Survivability- Based on "Granular Data Control"

[2557] The present invention provides a system and method for an ecosystem for delivering

information for superiority purposes and for data survivability purposes, based on "granular data control." In the digital paperless environment, the interconnectivity of networks and the Internet create new threats that can be compared to a battlefield. To achieve information superiority and data structure survivability, the present system integrates the following concepts into one ecosystem. The network or ecosystem must be able to protect itself and the data it holds. It must conduct its own surveillance and generate information in real time on threats, in essence, a command and control capability. The system needs to be able to defend itself and sustain itself throughout different types of attacks and disasters. The core, basic concept for information superiority and survivability is the need for constant data updating, gathering and its secured sharing. To maintain competitiveness and defenses, ongoing and continuous information searching, data mining and data inferencing and information sharing operations need to take place throughout the organization as well as outside the organization. A combination of some or all of the following processes and methods is the basis for information superiority and organizational survivability: (A) Creation of granular data streams. A granular data stream is defined as the extract data and / or the remainder data filtered from an original data stream. The filter may process the original data stream by content filter or other filtering methods. (B) Creation of content "Formlessness" by dispersing the granular data streams. (C) Dispersal and hiding of the granular data streams. (D) Compartmentalization of the different granular data streams. (E) Access check points at the storage stores / nodes. (F) Reconstruction of granular data streams into the original data stream or parts of it. (G) Multi level reconstruction of data. (H) Replication for continuity and survivability. (I) Constant search – for security and info superiority. (J) Content inference for penetration testing as well as locating relevant information. (K) Secured information sharing. (L) Automatic continuous growth of a data stream by addition of new search resulted data to it.

EX. 1002, Page 897

[2558] To achieve information superiority an organization's data / content assets to constantly grow and be updated in real time. Organizations and individuals data resources should be updated in real time. A comprehensive search and data acquisition and organization is key to the organization - like food and oxygen is key to humans and animals. There is a need for on going search throughout the organization. Every node on the in enterprise's network must be engaged in searching data- in essence supplying oxygen to the existing data assets. To deliver "competitive survivability," the search capability needs to constantly locate relevant information and update the existing data repositories. Presented herein is a system for continuous or part-time search all over the organization assets and outside data resources including the Internet, all with updates in real time. The system and method is an ecosystem for managing constant searches for (1) updating its data repositories in real time, (2) creation of competitive knowledge, and (3) to secure itself from content inference attacks and other attacks and disasters. The interconnectivity of networks and the Internet create a digital battlefield. The answer against virus attacks, Trojans worms/ inference attacks, covert channels employs the present process and method. The present system breaks away from the model of perimeter security and adopts a model that has been used in the battlefield for many years. The invention delivers capabilities to transform order within the data content into disorder making it very hard for an enemy to attack.

[2559] The idea of "formlessness" is one of the core strategies of the battlefield as expressed by Sun Tzu. Formlessness calls for minimization of data, in essence, controlled release of data especially small granular pieces of data. In a formlessness environment, all extracts are dispersed all over in distributed nodes. The Ecosystem is in constant mode of searching to secure itself - against inference attacks - and to gain knowledge. Results of search are needed for security for (1) triggering more extractions, (2) more dispersal of data, and (3) gather more misinformation which confuses the enemy -

within or without the enterprise.

A.52 System and Method for Minimizing Data Exposure

[2560] It is impossible to project and foresee all possible threats to a digital environment. It is impossible to do all permutations of possible inference attacks. Therefore, the invention presents a solution with the minimization of data. The invention provides a system and method for flexible content access based on rolling granular data exposure with decryption for added security. Granular pieces of the original document / data stream are dispersed to distributed storage nodes to enable a controlled secured environment for release of data. The granular data layers can be exposed one at a time decrypted instead of exposure of a total document.

A.53 Creating "Digital Doors" for Access and Controlled Release in a Data Stream

[2561] The present invention presents a "digital door" system and method for creation of doors in a data stream to achieve data control and management via management of granular data. A system and method for controlling data and content within a computer, a network, server or web site by management and control of data and data flows at the granular data level. Content and data are monitored and controlled as they exit through access control points (see FIG. A-1, ans sections A.1 - A.3) as well as when data enters the ecosystem, that is entry into the network, computer, servers, storage or web site. Content and data are monitored by the system as the data is held in storage. The system may remove data from storage through the access control points for security reasons or for other workflow and processing needs. Access controls or passwords and sec clearance protocols may be applied sequentially, of concurrently of subsequently with respect to the extraction of data and storage of data. Hierarchical access controls are oftentimes used. Controlling movement and storage of data by monitoring content, analyzing its granular parts when needed and masking or extracting granular data

streams and when needed and tagging data for infrastructure purposes. Enforcing information rights management (IR Mgt) processes to control movement and storage of the granular data streams is a function of the present system.

[2562] The system and method provides for management of data in a digital paperless interconnected environment. Controlling data is achieved by focusing on the granular elements of data (character, word object, sentence paragraph etc.). When extraction is used, the system adds to "granular data control" the attributes of physical and logical separation of data. The security aspect may be further enhanced by transfer of the extracted granular data streams into distributed storage stores. Creation of access control points or monitors in a data stream means, among other things, extraction of data from a data stream with the intention to reconnect the data extracted in whole or in part. Extracted data can be put in a package storage with or without encryption. The package may be tagged so to allow better management (IR Mgt) of the extracted granular data streams enabling fast reconstruction.

[2563] The process may include: First stage: filtering of the inputted original data stream, locating important content items that need to be masked or extracted. Second stage: implementing an inference process to verify if any thing can be inferenced from the sanitized data stream. Third stage: masking or extracting more data items from the sanitized data stream based on the threat results of the content inference process.

A.54 General Aspects - Basic

[2564] In general, the present process and system has the following additional features and characteristics: (a) automatic creation of a user or object's identification profile; (b) copying of filtered granular data and dispersing; and (c) combining extraction and copying for filtering of filtered granular data.

A.55 Automatic Creation of a User or Object's Identification Profile

[2565] The invention provides a system and method for automatic creation of a user or data object's identification profile. The profile will enable automatic interactions that will locate relevant content for the profile including its update without a need for additional input. The profile may be used as a guard and as a filter to enable or disable various transactions or communications related to the profile. The automatic building of a profile is done in different ways including scanning of data assets within a computer, database, or network nodes. Analyzing the content including use of content filters and selecting patterns which define the profiled subject. The system analyses each data asset, categorized its content, and give weights to the data asset. See the KE engine C.2 - C.23. The data profile includes the type of content, which parties interacted the data assets, at which time and location (audit or log data). Who sent what type of data to the data asset, for what reason? What is the security classification level of the data asset, how many people have accessed the data, was it a private document or a private one etc.

A.56 Copying Filtered Granular Data and Dispersing

[2566] The system and method includes filtering data and copying the granular data results and dispersing them to distributed storage. The process filters documents and data streams and cops resulting granular data items for dispersal to distributed storage locations. Dispersal may be done for different reasons for security, privacy, compliance, backup, continuity, survivability, backup or just close access or plain archiving. Dispersal of copied granular elements can be done to different locations applications search engines. Any resulting filtered granular data may be filtered again producing different levels or different sub groups. The granular results can categorized tagged and sent to different storage locations for possible retrieval, full reconstruction or partial reconstruction. The filtering process

may result in creation of sub groups of granular items, in essence more then one group of granular items. Those groups may be dispersed to distributed storage.

A.57 Combining Extraction and Copying for Filtering of Filtered Granular Data

[2567] A system and method includes features for filtering data, resulting in parts of the granular data being extracted and the other parts being copied based on security threats, compliance requirements and operational needs. Dispersal may be done for different reasons for security, privacy, compliance, backup, continuity, survivability, backup or just close access or plain archiving. Dispersal of copied granular elements and extracted granular elements may be done to different locations applications search engines. Any resulting filtered granular data may be filtered again producing different levels or different sub groups. The granular results can categorized tagged and sent to different storage locations for possible retrieval, full reconstruction or partial reconstruction. The filtering process may result in creation of sub groups of granular items, in essence more then one group of granular items. Those groups may be dispersed to distributed storage.

B.1 Basic Operational Theory of Secure or Select Data Storage

[2568] FIG. B-1A diagrammatically illustrates the basic processes for establishing a secure storage of information, generally identified herein as "data." The secure storage includes an extraction module and a secure storage module. "Data," as used herein, includes any data object, e.g., text, images, icons, data element, moving images, multiple images, data representing sound, video, electronic streams of information, etc. Sound bites and video images may also be extracted data. A source document 100 containing data, sometimes referred to as a "plaintext," is passed through a filter 102. Throughout this specification "source document" also refers to an information file containing data, particularly unstructured data. The module described herein operates on security sensitive content and operates on

"select content," wherein select content is information which is deemed to be important to an enterprise. Select content may also be security sensitive content. In other words, select content has value, such as intellectual property IP value. Security sensitive content is deemed to be important in that its unauthorized release is some how detrimental to the enterprise. Although it is convenient to discuss and understand the invention herein in connection with a plaintext document, the document 100 is a data object. It is not limited to an electronic document representing words. The document 100 represents a data object that may be, e.g., text, images, icons, moving images, multiple images, data representing sound, video etc. The term "data object" as used in the claims is broadly defined as any item that can be represented in an electronic format such that the electronic format can be manipulated by a computer as described herein. The data object, or as discussed herein, the "plaintext" is sent to a filter. Filter 102, in a most basic sense, separates out common text or remainder data 104 from uncommon text, words, characters , icons or data objects. The security sensitive (or select content) words, characters, icons or data objects are separated from remainder or common text 104 as extracted text 106. It should be noted that although the word "text" is utilized with respect to remainder text 104 and extracted text 106, the text is a data object and includes words, phrases, paragraphs, single characters, portions of words, characters, whole or partial images, icons or data objects. In a basic implementation, filter 102 may utilize a dictionary such that words present in the dictionary (common words) are separated from the source plaintext document 100 and placed into remainder document or common data file 104. The uncommon words (extracted-security sensitive (or select content) words), not found in the dictionary, would be placed in an extracted text or extracted data file 106. For example, a business may wish to impose a security system on a contract document such that the names of the contracting parties (not found in the dictionary) and the street names (not found in the dictionary) would be stored in extracted

data text file 106. The common text or remainder data would be stored in remainder data file 104. In the illustrated embodiment, remainder data file 104 also includes place holders which enables the extracted data to be easily inserted or set back into the remainder data file.

B.2 General Operation

[2569] FIG. B-4 diagrammatically illustrates the major components of a flowchart for the data security program and shows an extraction module, a content filter module and a secure storage module which can be modified to provide a select content SC module. Rather than operating a security system, the administrator (CIO) may operate a select content SC system to gather important or potentially important content in one location, with an audit trial to locate the source of the select content or "SC."

It should be noted that this flowchart may be truncated to limit user selection of certain items. The system would be pre-set to contain these features. Step 230 initializes the system. Step 232 enables the user to designate various levels of security for the activity which he or she will soon engage. The system, in step 234, enables the user to define the levels of security parameters. The following Security Table gives some examples of the type of security that may be available to the user.

Security Table

to whom
to where
when (time of day, day of week, month, floating but predetermined time frame)
why (purpose, match purpose to other security parameters or to certain predetermined criteria)
how (through what medium (LAN, WAN, Internet, direct dial link), download to what site or destination)
how long (duration) the reconstruction process will be permitted per each security clearance level
how much (different security levels enable reconstitution of documents and data with different amounts of secure data therein)
timing systems may require synchronization for a standard clock (i.e., atomic clock)

[2570] As an example of a truncated or pre-set program, a client-server system over the Internet may have URLs designating storage sites and an ASP 152 (FIG. B-2) controlling storage. In this pre-set system, the user does not select the sites. The sites may be randomly selected by ASP 152. The ASP may use artificial intelligence AI to locate secure extract data storage sites. AI or inference machines can ascertain (a) traffic on communications channels, (b) storage limit issues, (c) transmission failures in the communications links, and (d) the degree of security necessitated by exterior events, i.e., terrorism alerts, virus alerts, war, data security warnings posted by trusted sources, MicroSoft, Norton, NASA, DoD, CDC, FBI, etc. Higher security alerts trigger the AI configured storage locator and facilitator to locate memory stores in higher secured places. These higher security facilities may be more costly, may be located in more stable countries or on more stable servers and may have greater degrees of encryption capabilities.

[2571] The user, in step 326 can designate the location of the filter, the common storage area for the remainder data, the extraction data storage and potentially multiple data storage areas or segments. The user may enable an AI filter design. Step 238 permits the user to engage or disengage encryption and, if engaged, establish the degree of encryption for the system. Step 240 enables the user to define the parameters of the filter. The user can retrieve a preexisting filter or may define a new filter for each data security session. These filters may consist of dictionaries or any type of compilation of words, characters, icon, data objects or pixel formation or any indication that can be perceived by the computer system. These are called content filters. Granular extraction of data elements (or SC) in a data object may be permitted. Step 242 recognizes that the user either inputs a preexisting plaintext document or types data into the system. In any event, the plaintext document is fed through the filter. Step 246 extracts the security data or SC data from the input document. Step 248 stores the extracted data. The

extracted data may be encrypted prior to storage. Step 250 conducts an error check on the extracted data. This error check is helpful in discerning problems in the storage of the data prior to closing down the data security system. Step 252 stores the common data or the remainder data. Step 254 conducts an error check on the common or remainder data. The decision step 256 determines whether the user has selected a "destroy filter" command. If not, the filter is stored with or without encryption in step 257. If YES, the filter is destroyed with a deletion routine. Typically, deletion is complete erasure of all traces of the file including, in high security systems multiple write-overs or disc reformatting. Step 258 stores a map. The map may be stored locally or remotely as described earlier. This is the map module. The system ends in step 260. All traces of these data elements or objects may be swiped clean or removed from whatever computer system generated the data objects or processed them, other than the memory storage locations. Deletion of data also includes the concept of deletion of data transmission paths, URLs, storage site locations and all temporary memory stores. Deletion of file location in the root directory of hard drive 168 of computer 140 is preferable in high security systems.

[2572] FIG. B-5 diagrammatically illustrates basic flowchart features for the reconstruction process. The reconstruction process, although not shown in the flow chart for the info infrastructure, is a module that runs in conjunction with the access-permission-control module. Step 302 accepts a request to reconstruct the secured data. In a select content system, a search for the SC is initiated. Step 304 queries a local map and the security system or protocol. In a preferred embodiment the user would have to input several passwords, one of them being a local password on computer 140. A local map which may be accessed only through the password, may simply identify the URL of server 152. Decision step 306 determines whether the local password is acceptable. If not, and error step is indicated in step 307, the attempt to log on to the security system is noted in step 309 (an audit trail), and the

system either branches to repeat step 311 or bars the user from further activity in step 313.

[2573] Returning to decision step 306, if the password is locally acceptable, the YES branch is taken and the system executes step 308 which releases a reconstruction request to the common storage facility I-com 154 or A-com 108 (FIGS. B-2 and B-1A and B-1B). The system in step 310 logs the user in, as well as time and date and the data regarding the request. In step 312, a download from the common data storage is provided to RAM 166 or hard drive 168.

[2574] In step 314, a query is made to obtain the remote map from the remote security system. The decision step 316 indicates that the user again successfully inputs his or her security code. If not, error routine 317 is activated, the password failure is noted in step 319 (an audit trial), and the user is given an opportunity to repeat in step 321 or is barred or prohibited from further activity in step 323. In an SC mode, the requester may be required to pay money for the SC data. In the SC mode, a password may or may not be required. If the user has correctly input the security code, the system in step 318 releases the keys (to decrypt) and the map and releases the reconstruction request to the remote storage for the extracted data. This could be computer storage I-ext 156 or computer storage B-ext 110. In step 320, the user's access to the extracted data is logged in along with the time and day and type of data request (audit trail module). In step 322, the system downloads the extracted data into RAM 166 and/or hard drive 168 of computer 140. In step 324, an error routine is operated on the extracted data in order to insure that the extracted data properly matches the common or remainder previously stored. Decision step 326 determines whether the error routine properly generates the correct count or output. If not, the system in step 327 indicates an error, in step 329 the system deletes the common files and the extracted files and the system in step 331 logs in the failed attempt. If the error checking routine on the extracted data is acceptable, the YES branch is taken from decision step 326 and the system, in step 328,

proceeds to display the plaintext document or to integrate the plaintext document pursuant to the security clearance initially input by the user. Step 330 ends this process. The end process may entail encrypting the data again and swiping clean all traces of data objects from the memory stores and computer handling units. Of course, every use of encryption requires decryption of the data prior to reconstruction.

[2575] The system may incorporate various types of security systems or routines.

pass word
pass phrase
multiple choice questions and answers
initial, intermediate and subsequent security clearance routines
biometric security routines (voice, fingerprint, signature, eye or retina scan)

[2576] The reconstruction routines may be interrupted or the security system modules automatically activated or initiated upon the occurrence of externally generated triggers or upon certain predetermined conditions or conditional events. See sections D.12 and D.14, for example. Limited extraction, security clearance, release of data and reconstruction limits may be imposed. Artificial intelligence (AI) engines, inference engines or neural networks may be implemented to vary the permitted level of reconstruction via the security clearances. In other words, the AI system, as applied to reconstruction, may, relatively independent of the filter and storage processes, increase the necessary security levels permitted to access and generate full or partial plaintext recreation.

[2577] The display systems 220, 222 in FIG. B-3 include CRT monitors, LCD screens, projection screens and combinations of those systems.

[2578] The audit trail to monitor reconstruct and reconstruction attempts may include adding a time/date stamp to the remainder data and/or the extracted data prior to storage and a cross-check to the audit trail log during the reconstruction process.

[2579] Placeholders in the remainder document may be:

blank spaces
data symbols or elements "---" or "xxx"
false data
clearly erroneous data "ABC Company" or "Baker"
chaff or hash marks
messages
bar code
serialization data
alerts
links to other data objects
null set indicators "[ ]"
URL or website addresses

[2580] It is believed that the present invention is faster, during reconstruction, than standard encryption techniques, on the order of 100 to 1,000 times faster.

B.3 System Configurations

[2581] FIG. B-2 diagrammatically illustrates a personal computer or PC computer system 140, a second PC or computer 142, and a third PC -3. PCs 140, 142 and PC-3 are connected together via a network 145(LAN or WAN) and are also connected to an input/output device 146 that may be generally described as a router or a server to an outside communications system. The input/output device 146 is connected to a telecommunications system 148 which leads to Internet 150. The Internet is a global computer network. Internet 150 is coupled to a plurality of servers, one of which is server 152. Server 152 may be designated as an application service processor ASP. Internet 150 also includes various computer memory storage devices such as computer storage I-com 154, computer storage I-ext 156 and computer storage map 158. Computer storage enabling the store of extracted data includes a security level clearance module 157. Similarly, map computer storage 158 includes security level clearance module 159.

[2582] As stated earlier, the present data security system can be implemented on a single personal computer 140. In this case, different memory segments or hard drive 168 may be used for A-

com and B-ext. Typically, PCs include a keyboard or data input device 161, a display 163, a central processing unit CPU 165, a video board 167 having video board memory 169, a fixed disc hard drive 168, a RAM 166, and input/output device 164, a removable memory media drive 162a (may be floppy disc) and a removable compact disk (CD) read-write (CD-RW) device or drive 160. Also, storage in EMP or electromagnetic pulse hardened data stores is possible. Print stores rahter than digital stores provides another optional layer of security. The system may include other removable disk drives, tape drives, or flash memory units. Internal units CPU 165, video board 167, hard drive 168, RAM 166 input/output device 164, removable media (optionally floppy) drive 162a and CD-ROM device 160 are all coupled together via an internal bus 171. Bus 171 represents a plurality of buses as is known to persons of ordinary skill in the art.

[2583] One methodology of implementing the present invention utilizes distinct memory segments which may be designated in one or more of the following: hard drive 168, memory in a removable disk in drive 162a, memory in a removable CD disc in CD-RW device 160, and, to a very limited extend, RAM 166. In this manner, the user may select, generally at the outset of the process, that the extracted data memory storage B-ext 110 be stored on a floppy (removable memory) via drive 162a or a CD via CD-RW drive 160. The user can then simply remove the floppy or the CD and carry it with him or her. To reconstruct the data, the operative program, generally discussed above would have access to the floppy or the CD and particularly the memory location of the data on the floppy and the CD in order to reconstruct the entire plaintext document 100 (see FIG. B-1A). Alternatively, different portions of hard drive 168 may store A-com and B-ext. Of course, the computer system may utilize tape drives and memories or flash card, programmable memory.

[2584] In a local area network or wide area network implementation, PC 142 includes memory

similar to memory units described in PC 140 and a memory segment may be set aside in PC 142 separate from the common data or remainder data storage typically placed on hard drive 168 in PC 140. As a further expansion of the present invention, the extracted data (that is, the high security data), may be stored on computer storage I-ext memory unit 156 via Internet 150, telecommunications system 148 and router/server 146. In this manner, the common data or remainder data is stored on hard drive 168 and the highly sensitive data is stored off site in a secured location. Access to that secured location may be limited via security layer 157. If the user implements an encryption system (see encryption e 118 in FIG. B-1A), the extracted data is further secured by the encryption during the transfer from computer 140 through network 145, router/server 146, telecommunication system 148, Internet 150 and ultimately to computer storage I-ext 156.

[2585] The present invention may also be embodied utilizing an Application Service Provider on server 152 and in a client-server network.

[2586] An implementation of the present invention over Internet 150 most likely includes the use of a uniform research locator or URL for map memory computer 158, computer storage I-ext 156, computer storage I-com 158 and ASP server 152. In a client-server environment, server 152 acts as a server generally commanding the operation of client computer 140. Of course, persons of ordinary skill in the art recognize that the server may be located on the local area network 145 rather than being interconnected with Internet 150 as shown in FIG. B-2. The claims appended hereto are meant to cover the alternative embodiments.

[2587] As an example of a client-server or web-based implementation of the present invention, the user at computer 140 may define the filter 102 as described above, and input data (plaintext) via keyboard 161 or load plaintext data from drive 162a or CD-ROM drive 160 into RAM 166. In any

event, whether the plaintext data is input via keyboard 161 or copied or accessed from removable media drive 162a or CD-RW drive 160, the plaintext data is filtered as discussed above in connection with FIG. B-1A. Prior to filtering, it would be appropriate for the user at computer 140 to identify where the remainder data or common data will be stored and where the extracted or high security data would be stored. A simple program may automatically select the secure store location. The system is sufficiently flexible to enable the user to select local storage on different memory segments of PC 140 (hard drive 168, removable (optionally floppy) drive 162a, CD-RW drive 160) or be flexible enough to enable user at computer 140 to designate off site storage of the high security data (extracted data) and/or the common or remainder data. An automatic store routine may only require the user to accept or reject to preferred first security level, second security level and higher security level stores. The off site data storage process may include activating server 152 and enabling the server to take over the process directly from user 140. In other words, the user at computer 140 could call up the URL of the server 152, the server could request certain user information (user name, password), and would request data from the client computer to establish the filter pursuant to input selected by the user. The client computer may (a) filter the plaintext thereat or (b) send the data to the server for filtering. The server could store data either locally on computer 140 or remotely at computer memories 154, 156. After storage of the data at any of these locations, the server 152 may establish a map and store the map in memory location 158. Of course, remainder data (cleansed, plaint-text data)and the map may be stored at ASP 152 or client computer 140. The map, if stored at map storage 158, may be downloaded to the user at computer 140. The filter may be stored at computer 140 or may be stored at a secured location on server 152. Alternatively, the map could be destroyed on user computer 140. The filter could also be destroyed on user computer 140. Of course, the filter could be stored in a fourth remote location (not shown),

different from I-com 154, I-ext 156 and map computer memory 158. Storage of the map and decryption keys is a critical, high security task. Appropriate security measures should be utilized to protect those items. Local removable memory storage on disc in removable drive 162a or disc in CD-RW 160 may be reasonable. All traces of the map, the filter, the encryption key, the extracted data, and possibly the remainder data may be scrubbed or deleted from all computer memories (by write-over or disc reformat routines) other than the "com" and "ext" storage sites. Deletion of all URLs, links, x-pointers, etc. is also recommended for high security applications. Deletion systems are known to persons of ordinary skill in the art. For multiple security levels, multiple web site for storage of cleansed plaintext, first, second, third and higher security level extract text is preferable. Where the community of interest has access to the targeted and protected data via the Internet, multiple secured storage locations, multiple stores for filters, for encryption keys and for maps locating the secured stores is provided by multiple storage locations distributed throughout the Internet.

[2588] To reconstruct the document, the user at computer 140 would be required to call up the URL of server 152 and input the appropriate security code. The server 152 would then call up and download data from various memory locations whether they be memory locations on computer 140 or memory locations I-com 154, I-ext 156 and map memory 158. The system compiles the entirety of the plaintext document by gathering the dispersed components thereof or compiles partial reconstructions for different levels of security. By implementing different security levels, the system is dynamic enough such that server 152 can easily locate the various extracted data levels based upon various security codes representing different security levels, as those codes are input by the user at computer 140. Multiple security codes, at the inception and during the process, may be utilized. The user may be required to input security codes at multiple times during the reconstruction or compilation process. Regeneration

of the source is possible with appropriate sec codes. Likewise, the source, via the processes herein, may be reorganized by using tags, labels and different named storage facilities. Maps tracking locations of sec or SC stores enable the user to reclaim data, reform the data previously stored and reorganize the same.

[2589] It should be noted that computer storage 154, 156 and 158 may be located on the same computer or may be located on different computers spread throughout the Internet. If the storage units are different computers spread throughout the Internet, computer storage 154, 156 and 158 would each have their own URL or Uniform Resource Locator. On a LAN, the computer storage 154, 156 and 158 would each have their own addresses for access thereon. In any event, during reconstruction, the server 152 gathers the information and downloads the information into RAM 166 of computer 140. This download may include a first download of the common or remainder data from I-com 154. At a separate time, which may or may not include a decryption routine, the extracted from I-ext 156 is downloaded. Preferably, other than inputting initial security codes and any required or desired intermediate security codes, the system operates automatically without further input from the operator at client computer 140. The download of both data sets may be simultaneous in that the download is not humanly perceivable. This is especially true if storage in different memory locations in PC 140 is utilized.

[2590] The role of server 152 may be expanded or reduced dependent upon the desires of the user and the degree of security necessary. For example, server 152 may only enable separate storage of extracted data in I-ext 156. In this limited role, server 152 would require the input of a proper security code and clearance prior to identifying and enabling the download of extracted data from I-ext 156.

[2591] In an expanded mode, server 152 may be involved in filtering the data, extracting the security sensitive (or select content) words, characters , icons or data objects to obtain extracted data and

remainder data thereat, separately storing the extracted data from the remainder data (extracted data being placed in computer memory I-ext 156 and remainder data being stored in common remainder data memory I-com 154) and then permitting reconstruction via separate or combined downloads of the remainder data and the extracted data into computer 140.

[2592] The innovation is a system and method for automatically or manually controlled selection, extraction, storage, and release of selected and prioritized information. The system extracts selected information from data streams, in computers, computer networks communication devices, and networks, as well as electronic mail systems. The system and method can reside on a single computer, be distributed across multiple platforms, be distributed across multiple networks, or reside as a remote process (known as a hosted application service process in the state of the art).

B.4 Input or Initial Processing Considerations (Basic Filter Modules)

[2593] The security sensitive (or select content) words, characters, icons or data objects may be any word, phrase, letter, character, icon, data object (full or partial), image or whatever, as pre-defined or as established by the user. The user may specifically design the filter, begin with a dictionary (a content filter source) to define common terms, identify any additional security sensitive words, letters, images, icon, data objects, partial versions of the foregoing or any other granular aspect of the plaintext. After defining the filter and accepting the data input, the system filters the plaintext and separates extracted data (security sensitive (or select content) items) from the remainder data. The filter may also include elements of artificial intelligence (AI). For example, the user may select one word as a security word and the AI filter may automatically select all synonymous words. A contextual filter may be thereby designed. The AI filter may enable the user to define a filter in real time at the entry of data via a keyboard. For example, the user may select to secure (i.e., extract and store) some proper names and

may instruct the filter to secure names such as Block, Smythe and Cherry. During input of the plaintext, the system may detect Smith and ask the user if he or she wants to secure (a) all proper names in a common name dictionary collection and/or (b) all names with spellings similar to the filter input data, Block, Smythe and Cherry. As is known in the art, AI typically uses inference engines to define one pathway or to outline a course of action. The filter or extraction engine discussed herein can be configured with AI, inference engines, neural network systems or other automatic systems to carry out the functionality described herein for the dynamic operation of the security system.

[2594] The system and methodology described herein also encompasses parsing the plain text document by bit count, word, word count, page count, line count, paragraph count and parsing based upon any identifiable document characteristic, capital letters, italics, underline, etc. Parsing is a contextual filter using a content as a marker and a range defines by certain file parameters. Algorithms may be implemented to parse the plain text document. The target of the parsing algorithm (a bit count, word, letter, etc.) is equivalent to the "security word, character or icon, data object" (or SC) discussed herein. The parsing occurs with the filtering of the plain text source document 100 and the subsequent storage of extracted data apart from remainder data.

B.5 Storage

[2595] In a basic configuration, the common text or the remainder data is stored in common storage memory 108. The secure storage is similar to the select content SC storage module. This common or remainder data store is identified as A-com generally referring to a segmented memory in a PC or a computer A in a network (LAN or WAN). It should be understood that reference to "remainder data" is simply a short-hand representation of data that is not extracted or filtered by the system. In some cases, remainder data may be nil and all content is removed. Accordingly, "remainder

data" is simply that data which can be viewed, manipulated or further processed by the user inputting or initially processing the data. Remainder data storage 108 may include a confirm storage signal function 111 to send back a confirm storage signal to the data input device generating source plaintext document 100. The extracted data file 106 is stored in a different memory computer storage 110 (B-ext). In a preferred embodiment, memory segment 108 (A-com) is at a different location than computer storage memory segment 110 (B-ext). In a PC embodiment, memory A-com is a different memory segment than memory B-ext. In a networked embodiment, computer storage 108 may be on a different computer as compared with computer storage 110. In an Internet embodiment, common text or cleansed text storage is at one web site (which may be one computer) and the extracted, high security data is stored at another web site, buried web page or other Internet-accessible memory store location. In any event, the remainder text is stored in a memory A-com and the extracted data or high security words, characters , icons or data objects are stored in memory B-ext. After storage of the extracted data in memory 110, a confirmation indicator 113 may be generated to the client computer or the computer handling source plaintext input document 100 (the originating computer system). Data may be stored on any type of computer memory or medium and may include non-volatile memory, such as ROM, flash memory, floppy disk, disk drive memory, CD-ROM, and other permanent storage. Additionally, computer readable medium may include, for example, volatile storage such as RAM, buffers, cache memory, and network circuits.

[2596] As a simple example, the program configured in accordance with the present invention, could automatically detect entry of all credit card numbers types into a user's computer. The filter is set to detect the unique credit card sequence and data string. Assuming that the user's computer is operating a browser and the user is communicating with a server on the Internet, the user's computer would filter

out the credit card number and send the number to a secure storage site. The secure storage site is owned, operated or leased by a trusted party. The extracted data, i.e., the credit card data, is stored at the trusted site. The URL or other identifying data is sent to the vendor from which the user wants to purchase goods and services over the Internet. When the vendor seeks to complete the transaction, the vendor sends a request code to the secure site, the trusted party at the secure extracted data storage site debits the user's credit card account (or otherwise debits the user's bank account) and sends an approval code to the vendor. In this manner, the vendor is never given the user's credit card – the card number is sent to a trusted party automatically by the filter in the security program described herein. The security program may be incorporated in a browser to automatically protect credit card data, personal data (as a method to become anonymous on the Internet), etc. from being deliberately broadcast to others on the Internet or to block others from snooping into the user's personal data while the user communicates over the Internet.

[2597] In a further enhancement of the present invention, the computer or data input device handling source plaintext document 100 may also record the location of A-com 108 and B-ext 110. The location data is called herein a "map." A memory mapping function or module is utilized. The map may be stored in a third memory location 112. Memory location map 112 may be a segment of the memory of the data input computer originating plaintext 100. The map may be encrypted for security reasons.

B.6 Extraction and Storage Enhancements

[2598] As a further enhancement of the present invention, the user, prior to initiating the security system, may be given a choice of filtering out all the uncommon words or words not found in the dictionary and adding certain security sensitive (or select content) words, characters, icons or data

objects to filter 102. The added words or terms are filtered out with the uncommon words. Of course, the user may be required to manually input all security words or download the security word filter from the Internet or another system on the LAN. For security systems having multiple security levels, a plurality of filters would be created, each filter associated with a different security level. Further, multiple security levels would require, in addition to remainder text document or data 104, a plurality of extracted data documents 106. The common or remainder text document or data 104 would still be stored in remainder computer storage A-com 108. However, each extracted data document 106 would be stored in a respective, separate computer memory segment or computer B-ext 110. Separate storage of a plurality of extracted data at multiple, separate locations in B-ext is one of the many important features of the present invention.

[2599] The ability of the program to locate security sensitive (or select content) words or characters can be enhanced by using a telephone book, properly dissected, to identify a collection of last names. Cities and towns and street names can also be identified in this manner. The compilation of last names and cities, towns and streets can be used as a list of critical, security sensitive (or select content) words. The filter is represented by this compilation of words. Similar techniques may be used to create filters for scientific words, or words unique to a certain industry, or country.

[2600] In view of increasing levels of security relating to (a) the storage location A-com; (b) the transfer of remainder text document 104 to memory computer storage A-com 108; (c) the storage of map 112 (possibly encrypted); (d) the creation, storage or transfer of filter 102 (possibly encrypted); (e) the storage of extracted data at memory storage B-ext (whether singular or plural storage sites); and (f) the transfer of extracted data thereto, the system may include an encryption e feature. The encryption e function 115, 117 and 118 is diagrammatically illustrated in FIG. B-1A.

[2601] The program of the present invention can be set to extract critical data (a) when the plaintext or the source document (data object) is created; (b) when the source document or data object is saved; (c) on a periodic basis; (d) automatically; (e) per user command; (f) per ascertainable or programmable event; and (g) a combination of the foregoing. Timing for storage of the extracted data is based on these aspects. Reconstruction of the data object or plaintext may be (a) automatic and substantially transparent to the user; (b) based upon manual entry of security clearance data; (c) periodic; or (d) a combination of the foregoing dependent upon outside events and who is the author of the data object or other definable aspects of the data object, its environment of origination, current and anticipated security threats and its environment of proposed reconstruction. The timing for the extraction, storage and reconstruction is oftentimes dependent upon the level of security required by the user and/or his or her organization.

[2602] The system and method creates a high level of security by automatic selection and removal of critical and prioritized contents from a data objects stream, whether it be a digital document, digital file, database, sound bite, video clip, other structured, or streaming data formats. The system and method enables a controlled release of the extracted data objects, enabling instant display of the instantaneous returned contents, contingent on verification of user identity, access rights, time of operation, location of source and or user, destination of source and or user, and determine threat modes. The system and method delivers high security by removal of the selected prioritized content from memories. The copies and traces of the selected extracted contents are eradicated from the computer memory while the separated extract data stream is transferred to a safe removed storage memory media. The extract, extracts, and any part thereof, will be return transferred to user display as soon as identity and access rights are validated.

[2603] A replacement of the extract (sometimes called a placeholder) can also be substituted on-the-fly to provide updated results, misinformation, dis-information, messages, alerts, links (to reports, data mining, search engines, web sites, and hyperlinks understood in the current art), advertisements, and personalization and customization. The validation can be done instantly by password, challenge questions and answers, remote verification (phone, video, or personal contact with user), or by biometrics means.

[2604] The extraction of data objects within data streams includes words, structured data objects, characters, numbers, bullet points, footnotes, prices, images, sound segments, video segments, and selected digital data packets. The extraction is conducted by separating a source (original) data stream into two or more extracts data streams. The different data object extractions are separated into groups reflecting predefined contextual categories and restitution applications (such as to enable customization and personalization for the same or different users). The modified source (original) stream typically contains the majority of data objects of the source stream, whereas the extract streams contains a minority of the data objects which represent selected and categorized information and information deemed to be of prioritized importance.

[2605] The extracted categorized data objects are separated into one or more contiguous data streams. The extracted data stream or streams are scattered to one or more storage memory memories. The extracted data can be transported and shuttled between different storage or projection apparatus, as directed automatically by various constraints including: security risk criteria of threats and attacks, sources, targets, users, policies, time of day, and threat modes.

[2606] The extracted data, in some cases, is transported to an online removable storage and under extreme security threats to an off-line/off-network, digital or physical vaulted storage. Transport

and shuttle is based on the level of security alert. The use and release of the vaulted extractions is controlled by a set of rules or organizational policy which includes the following options among others: (a) A vaulting of some, all, or specific data object extracts for long or short periods of time. (b) Release of extractions into a display, in which the extracted data objects will reconstitute with the modified original data stream, the original data objects stream or original document. (c) Release of extractions into a projection display in order to project with the modified data stream, the original document while maintaining complete separation of the modified source data stream (the source modified by the extraction of data objects and insertion of placeholders) and the extracted data object streams. (d) Release of extractions into a projection display in order to project a reconstitution of the original document, in other words to create altered versions of the original document, while maintaining complete separation of the modified data stream and the extracted streams. (e) In situations of high security threats, release of extractions into another projection display, PDA, floppy disk, paper document a wireless display, an overlay transparency while maintaining logical and physical separation of delivery streams. This will enable working with a representation of the initial source, but not the initial source itself, while understanding the prevalent, critical extracted information without comprising security to the initial source material by exposing sensitive, identifying, or critical information. (f) The distribution of sources, modified sources, or extracts to remote and distributed viewing devices. (g) Enable the ongoing operation of information delivery and display in defiance of known ongoing or unknown security flaws, breaches, or events compromising the general state of security. (h) The delivery of distinct and separate data streams, delivered on the same or different channels and media, each with minimal, limited, or even substantial usefulness in and by itself, that can be overlaid logically or physically to reconstitute the identifying data stream and display. Separate display devices can be used

to create a composite image or they can be overlaid to physically separate display devices to reconstitute a useful composite display.

[2607] The objective is to create security for the single computer or extended network. When an intruder penetrates preexisting firewalls and other security systems, the data object and streams, digital documents, and digital files which will be valueless and prioritized data objects rendered unidentifiable, the penetration is valueless because the critical strategic information has been exported to a vaulted storage. Competitors or hackers, who learn that a computer or network is protected by the system and method, might decide to attack another target instead. This is comparable to a situation in which a bank robber, who finds out that the bank vault is empty, will most probably look for another bank.

[2608] The system and method has a menu of different options including the ability to extract: (a) All existing databases on the computer or network. (b) All newly loaded, mounted, or integrated data to the computer or network. (c) All plug-in memory devices (temporary or permanent) containing data. (d) All new and imported data to the computer or network. (e) All new work and output created by the computer or network. (f) All data being transported in/out of the computer or network including electronic mail. (g) All data being transmitted in/out of the computer or network including electronic mail.

[2609] The system and method releases the extracted data streams, subject to a controlled-release mechanism and process. The release mechanism is contingent on parameters including; rights to access specific contents, timing criteria, security restrictions, and preset policies. The release of the extracted data objects permits restitution of the source data stream in variations of the source that are full, partial, or modified representations of that source data stream. The release provides for various levels (through

user configuration) of separation between the modified source data stream and the extracted data streams. The system enables the maximum grade of security by means of the option of a visual merged projection of said different data streams, while maintaining a strict physical and logical separation between the data streams.

B.7 Basic Reconstruction

[2610] FIG. B-1B generally diagrammatically illustrates the major features of a reconstruction routine (module) or system. The user, typically at a computer terminal, inputs a reconstruction request 120. The system first executes a security clearance protocol routine 122 in order to determine whether the user has the proper security clearance. The security clearance may be thought of as a security clearance control. If multiple users are permitted access to the documents and those multiple users have different security clearances, the security clearance protocol determines the level of security clearance and, hence, the full or partial reconstruction of the plaintext. The security code input by the user is checked against a security code database or list 124. Clearance is provided in step 126. The location of the map and, hence, the location of the remainder data A-com 108 and extraction is provided to the user's computer in step 128. This may include obtaining a copy of the map 130 showing the location of memory segments in (a) the local computer; (b) the LAN or WAN; or (c) the Internet storage sites. The storage segments are A-com 108 and B-ext 110. The common or remainder data is downloaded or transferred or made available to the user's computer as shown at the output of map location and data step 128. Typically, the extracted or security sensitive (or select content) data from B-ext is downloaded. As described hereinafter, the data can be reconstructed as a complete electronic document in function 130 or may be reconstructed only as a visual reconstruction in step 132. Visual reconstruction is discussed later. Function 130 operates as a compiler to gather the extracted data and remainder data into

a single plaintext document. If the data object represents sound or audio signals, reconstruction and play back may require a speaker output in function block 130. In a telecommunications implementation of the present invention, the input would include a microphone or audio detector (supplemental to the input device for document 100), an analog to digital converter (possibly with a voice to digital converter), the filter, extractor, storage facilities at least for the extracted data, and at the output of the system, a converter to audio and an audio announcer. The recipient of the secured data stream or message would be required to clear a security clearance and possibly obtain a decoding key prior to listening to the entire, decoded message. The key and the security data is separately downloaded to the recipient's device.

[2611] If remainder data in A-com memory 108 and extracted data in B-ext computer memory 110 is encrypted, the reconstruction process includes a decryption step. Encryptors and decryptors are relatively well known by persons of ordinary skill in the art. Further, the filter 102 (FIG. B-1A) may include some encryption routine operating on the data object (plaintext) during the filtering. A simple encryption may include substituting "dummy" text or images for the security words and keeping a pointer to an encryption key document mapping the security words with the dummy words. The filter may be stored or may be destroyed at the option of the user. Storage of the filter impacts the degree of security of the entire data system but storage of the same filter enables the user to reuse the filter at a later time. Encryption of the stored filter increases the security of the data. Creation and storage of map in memory 112 also impacts the degree of security of the system. However, if the filter 102 is destroyed and all copies of the map are destroyed on the user's computer originating plaintext document data 100, and the map is stored offsite in a third computer memory location 112, this offsite map storage may enhance the degree of security of the data. The originating computer processing plaintext 100 may be

scrubbed to remove all reference and copies of the plaintext, remainder text, extracted data map storage data, etc., i.e., a deletion routine may be employed on the data input computer.

B.8 Reconstruction Techniques

[2612] FIG. B-3 diagrammatically illustrates a system diagram for various reconstruction routines. A complete reconstruction is shown as security level path A. This involves an electronic integration of plaintext in step 202 resulting from the complete electronic reconstruction of document 100. For example, a merge may occur between the extracted data and the remainder data or common text data. The document is completely compiled in this process. Placeholders in the remainder document are utilized to locate and insert the extracted data. Most likely, there will be no process controls imposed on the integrated document as shown in step 204. In other words, if the user at computer 140 has the proper security clearance, he or she could download or recreate the entire original source, plaintext document and the user would be entitled to edit the document or change it in any way or copy it and reproduce it.

[2613] The second level of security, path B, results in storage of the common or remainder data in a different memory location on the hard drive 168 as compared with the extracted data. This is noted in step 206. Another words, in a simple example, hard drive 168 or RAM 166 would hold a copy of a remainder data document and another copy of the extracted data document, that is, two separate documents. Since two documents are available in RAM 166 or hard drive 168, these documents are stored in different locations in the memory. In step 208, a map showing the memory location of the common or remainder document and the extracted data document is provided to computer 140. Step 210 commands the processor CPU 165 in computer 140 to interleave the extracted data with the common or remainder data in the video board memory. In this process, the extracted data would

typically have placeholders for the missing remainder data. Otherwise, control codes to locate the extracted data into the remainder data would be executed by CPU 165 to properly place the extracted data into the "visual space" of the placeholders in the remainder data document. The extracted data document may have placeholder for the remainder data. Some type of register between the two image documents may be necessary. The compiler, in this embodiment, gathers the document elements and visually compiles and presents the plaintext to the user.

[2614] FIG. B-3A diagrammatically shows that video board memory 169 is loaded with remainder or common data 1 and a different location of the video memory is loaded with extracted data 1. The next video memory location is loaded with common data 2 and then a different video memory location is loaded with extraction data 2. Since the refresh rate of computer monitor 163 is fast, the display 163 will show the common or the remainder data and then show the extracted data such that the user could not humanly perceive a difference in the document. However, the user could not copy the document from display screen 163 (a "screen shot") since the document is never electronically integrated into a single document. There is only a visual presentation of the combined document by interleaving the extracted data with the common or remainder in the video memory 169. Step 212 notes that the user may be limited in his or her ability to process, edit and store the reconstructed and presented plaintext document.

[2615] Security level path C recognizes in step 214 that the data is stored in different memory or computer locations. In this situation, two video boards, video board A and video board B are shown as board 216 and 218. Video board 216 drives display monitor 220. Video board 218 drives display monitor 222. Display screens 220, 222 are overlaid atop each other. Video board 216 is fed with common or remainder data from the remainder data store (see I-com store 154 in FIG. B-2) and video

board 218 is fed with the extracted data from the extracted data store, for example, I-ext store 156. In this manner, as noted in step 224, the user is presented only with a visual presentation or compilation of the plaintext. Since there was physical separation between video monitor 222 and video monitor 220, there is no electronic integration at all of the plaintext document. Hence, the ability for the user to do any significant editing on the plaintext document is blocked or prohibited because the user only has access to either the data on video board 216 or the video board 218.

[2616] Security level path D shows that the extracted data may be parsed or further separated based on a plurality of security clearances in step 226. Step 228 recognizes that the system can repeat process and security process paths A, B and C only with portions of the extracted data presented to the user based upon the user's security clearance.

C.0 Testing the Security System

[2617] The prime purpose of the data security system is to limit the disclosure of critical data or select content SC data to persons or organizations who may misuse or abuse the data. A testing module supplies this functionality to the infrastructure. With the advent of cheap memory, fast processors and increasingly dynamic search engines and the high speed communications links established by the Internet, the ability of a suspect person or organization (a person/organization which is not permitted to have the SC or secure data) to associate some low level secure or SC data with public source (or accessible private source) data and "discover" the secret or SC presents a significant problem. Therefore, a system to test the secure nature of the SC or secure data system, with inference engines or knowledge expander KE search engines, is beneficial. The following sections C.1 - C.23 discuss such search engines to test the security of the base system described in sections B.1 - B.8.

[2618] Also, the information infrastructure is flexible enough that the system operator (CIO) initially builds the infrastructure using simple filters and simple access and permission polies and rules. The infrastructure becomes more complex due to testing of the secure storage and the SC storage. The higher degrees of complexity may cause the system operator to employ more complex filter (see sections C.11, C.12 and C.13 and employ aggressive deconstruction techniques (see section D.1, among others). Additional complexity is added to the system with further polies or rules which are prioritized and implemented with hierarchical taxonomic classifications. The hierarchical taxonomic classes must be bought, customized or built. The knowledge expander KE module is used in the development of the contextual filter modules, the testing modules, in the taxonomic filter modules and in the classification generator.

C.1 The Etiology of Information

[2619] Security, privacy and information sharing is predicated by the representation of the structure information. The structure has evolved in usage over time from simple linear formats to complex hierarchical trees typified by tags, metadata and modifiers. Although the predominant information delivery and information storage format is in a linear data stream, the internal structure or representations include all possible linear, field defined, metric, tree, compound and combined layouts. In other words, while data is delivered in a linear stream, the complexity of internal structure resolves into specific documented patterns, self-documenting meta data formats like HTML or XML, defined utilitarian and purpose-oriented formats like database management system (DBMS), ODF (open document format) or proprietary document object models (such as the Microsoft DOM model). The combination and recombination of metadata in source documents or data streams complicates finding, location, and expanding one's knowledge base of SC. The issue of the internal format is important to

the regulation, interpretation and application of information.

[2620] As discussed above, the etiology of information involves a complex hierarchical trees of various types of data such as tags, metadata, and modifiers. Specifically to the issue of semiotic (words and objects in language) and pragmatic (words relationship to user) meaning, the internal format of information is important to its regulation, interpretation and the further use and application of the information. A discussion of the differentiation of the data as to content, context and concept is discussed later herein. The abstraction of information is also relevant here. For example, the database scheme DBMS and the DOM style sheets and the internal metadata all modify the encoding, format, structure, purpose and usage of the information. Links and references establish contextual meaning, just as the environment establishes meaning relevant to the multiple granular data in the information stream or document. Metadata can validate or invalidate meaning in the text of the data, for example, a format or a replace meta data element could indicate the complete deletion of material, the connection between topic and footnote, or modification with non-usage of the data. To perceive data only in context without the effects of format, purpose, reference and access rights, potentially misinterpretes the importance of context and concept which may result in a misunderstanding of the information.

C.2 Metasearch Engine for Knowledge Discovery: Search Convergence and Divergence

[2621] As an overview, the system's metasearch engine is a novel and useful service to uncover unknown or concealed relationships, delivery of higher relevancy search results and connecting the dots. Although metasearch is derived from the technology for security and information sharing, it is useful to augment standard Internet search engines because it enables both search divergence for knowledge discovery and search convergence for assessing information integrity, the full process necessary for connecting the dots. Search divergence extends search with aggregation, inference and data-to-data

interaction beyond simple content into the realm of context and concept. The system is a standalone system or can be integrated to process results of other search engines. Presented is a process and a technology for knowledge discovery, security and information sharing that is based on management and control granular data content to secure information while also providing interdomain information sharing. The basic assumption is that all data /content must be monitored, analyzed and categorized at the granular data level for basic security risk management and control. Risk measurement is driven by information recognition, multi-tier overlapping hierarchical meta-tagging technology, and granularization. This same technology, as a standalone deliverable product or service-oriented architecture, is applicable to knowledge discovery, uncovering unknown or concealed relationships, and for connecting the dots. The metasearch process is a serialized and repetitive process of knowledge discovery, usage of the innovation's secure engine, and the application of granular data content, as a standalone service for knowledge discovery. Because the risk measurement is driven by (1) content, (2) context, and (3) concept, it is essential that the system recognize content not only in isolation to itself as a data stream or document but also its value when combined with other private and public sources through aggregation, inference and data-to-data interaction. The metasearch is useful to augment standard Internet search engines because it enables both: (1) search divergence for knowledge discovery and (2) search convergence for assessing information integrity. It completes the process necessary for discovering new knowledge to connect the dots. Although testing for security is discussed above, testing for select content SC is possible. SC is unique information that is of some interest to the enterprise. If the SC store is to be complete and represent a viable and useful knowledge store, the quality of the information in the SC store can be tested to ascertain whether the content is unique or should be supplemented as discussed later herein.

C.3 Initial Process - Metasearch Engine and Knowledge Expander

[2622] The automated metasearch KE search engine takes as source any data stream or document. The metasearch engine module (Figs. C-1, C-2, for example) automatically parses the source for content recognition and assessment of granular content. This granular content is automatically metatagged by the search engine - system for range (contextual relevancy), sensitivity level (prioritized), compartmentalization (sensitivity levels at the same hierarchical level, but laterally or orthogonally disposed to one another, see TS - Navy compared to TS - Army wherein the Navy and the Army are compartments), categories (hierarchical taxonomic classes), relevancy (a type of sensitivity), and other multi-tiered overlapping hierarchical factors. The granularized data content becomes the search terms, while these metatags become metafilters for the knowledge discovery process. These search terms are submitted to any number of extant tools and services (other search engines, such as Google, Yahoo), for locating other relevant supplemental data streams, documents, databases, local, Intranet, Internet, and public or private data stores. Likely tools include Internet search engines, data mining tools, database query engines, data collections, indices and other knowledge management (KM) applications or services, which are generally classified as data structures or data collections herein. Inductive and deductive tools are advantageous too. Most tools can be easily integrated with the system, linked through service-oriented architectures (SOA), or their results piped into the source data steam or document (through Web 2.0 mashups). The metasearch engine also works with standard—a priori—search terms or keywords, like any desktop, web-enabled, or Internet search tool. This search operation may bypass the system's secure assessment of a relevant or representative source data stream or document.

C.4 Simple Metasearch Engine

[2623] The results from the extant tools (the results sometimes called supplemental documents

or supple docs) are logged for auditing, dispersed to maintain a legal chain-of-custody, and aggregated into a single collated result. A List of Abbreviations is set forth in a later section E.1. This collated result (the source, results, reports and links or references) is automatically submitted into the metasearch engine. Inputs include the configuration for either (1) search convergence or (2) search divergence. Search convergence conforms to the standard purpose for using Internet search engines like Google or Alta Vista because it defines, authenticates, and validates the search terms. The first level search results provide for an integrity check for the initial information results in that it does not aggregate or inference the information therein.

[2624] The divergence function, on the other hand, is all about aggregation, inference, and data-to-data interaction because it specifically searches for links, references, relationships, outliers, and social networking associations to the search terms. The divergence function will likely ignore anything overlapping the search terms in order to expand the search. In other words, convergence increases the depth of knowledge where divergence increases the breadth of knowledge represented by the collated result. The divergence function explores new, novel, unknown, and hidden connections. If you want to connect the dots, divergence dredges new but factually uncertain information where convergence thereafter authenticates.

[2625] The system is automatically run. The source (which is now the aggregated result from the extant tool processes) is automatically parsed by the metasearch engine with secondary recognition and assessment of granular content.

C.5 Automatic Metatagging

[2626] This granular content of the source doc is automatically metatagged for: (1) Content categories, (name, location , date, dollar amount etc); (2) sensitivity level, ("Top Secret", "Private" ,

"level 9"); (3) compartmentalization, ("top secret- Finance" "Top secret logistics"); and (4) relevancy, and other mult-itier hierarchical factors.

C.6 Automatic secondary tagging of a "range"

[2627] A secondary level of metatagging may be created to define the "range" or "the area within the document" from which the keywords for the search will be selected. This is a simple contextual select content (SC) or secure content (sec-con) operation. Software modules carry out these functions. To create a higher level of relevancy in the search results the system controls the granular data within the document. Only "areas"/ "ranges" with a high relevancy will become the basis for locating keywords for the next search. Range defines the area or areas within a document (characters from the target selected content, lines away, within the same paragraph, on the same page, etc.), which will be the base and platform for locating key words for the next search cycle. An analogy is the defining the range / area for oil drilling in a suburb of the city of Houston. All the drilling will take place only in that location. As an example of defining a " range," we present the system's creation of automated tear lines.

[2628] Each paragraph (the "range" can be also configured to a "sentence" or "line" or "page") is automatically tagged based on an algorithm. The algorithm may be "tag the paragraph based on the highest sensitivity level assigned to any of its words; or if a group of specific words appear in a paragraph," ("ATTA", "Hamburg" "New York" all tagged as Secret "S" security level or importance) and tag the paragraph at a higher sensitivity level than those words (i.e. Top Secret "TS").

[2629] The granularized selected content becomes the search terms and the metatags become the metafilters for the systems meta search. The automated results of the meta search process is an augmentation of existing knowledge or a breakthrough to unknown or covert relationships. The convergence search vets the integrity of a priori search terms, while the divergence search generates new

search terms and new references. The metasearch is repeated, as configured, for a fixed number of times, for a fixed period of time, until results reach an asymptote, or indefinitely. The metasearch inputs also include the metafilters created prior to the presentation of the aggregated search result. These metafilters define the direction of the search by introducing "fine adjustments" to the search process. In essence the metafilters narrow the scope of all subsequent and serial submissions to the extant tools and services. The search may be configured to include one or a combination of filters as follows:

[2630] (1) Selection of keywords for search based on their categories. The system automatically categorizes each word and character. In one embodiment, there are over 50 different categories or hierarchical taxonomic classes. Categories may include name, location, date, dollar amount, credit card number etc. As an example the system may be configured to feed the search engine with words that where automatically categorized by the system as "locations" and "people" in such an example "Rome" " London" and "Muhammad Atta" may be selected automatically as the keywords for the next search cycle.

[2631] (2) Selection of keywords for search based on their sensitivity level or important or assigned weight. This is the sensitivity level of the hierarchical taxonomic classes. The system may be configured to select as keywords for its next search only data elements that where classified / tagged with a specific sensitivity classification/ tagging. A possible example, is an instruction to use as keywords only words that where given " Top Secret" classification, or "private" classification or assigned with a specific weight (select words that where assigned weight 9 out of 10).

[2632] (3) Selection of keywords for search based on the specific importance of their content. For example, use as keywords, words that belong to a list of specific grouping of words. A list of associated words that convey a concept or a contextual relationship and other features and

compartmentalization. Metafilters for standard search terms, which bypass the initial system's automatic review of a source, are optional.

C.7 MLS, Multi-level Security-Compliant Search and Distributions

[2633] Because the system and all the extant knowledge management tools may run in a secure system high environment, results are likely to be system high too (classified at the same security level). However, the point of any search or knowledge exploration is to share information and distribute it to the users at the edge. The advantage of the system is that the new search terms, the aggregate result, and all intermediate reports and documents are processed for Multi Level MLS-compliant security and information sharing. Each user will get a result based on his security level (for example, TS or S or C or UC). Specifically, the list of new words, phrases, and other content is either automatically assessed or manually reviewed for addition to the system's dictionaries. Ownership sensitivity level, compartment, group, categories, mission, relevancy, and other multitier overlapping hierarchical metatags are applied to each discovery and distributed to users subject to MLS compliance. The aggregate result are secured with multiple MLS-compliant versions and distributed to the edge on a: (1) per sensitivity-level basis version, or (2) a base redacted document with objective security compliance for universal distribution and pushed to the edge with the system's reconstitution services. In other words, information sharing is implemented either through the delivery of: (1) Multi Level Security - MLS-compliant versions, or (2) Through a base redacted document with objective security compliance for universal distribution with reconstitution.

[2634] Reconstitution is supported in whole or in part as: (a) MLS-compliant results, or as (b) Step-wise reconstruction, with defenses-in-depth MLS-compliant results i.e. controlled release layer by layer.

[2635] These two options are also applied to all intermediate results, reports, lists, linked or referenced sources. Of note, all links and references, even access to public data sources can be optionally restricted to minimize subsequent risks from aggregation, inference, and data-to-data interaction. An agency with specific intent and knowledge is able to use aggregation, inference, and data-to-data interaction on public sources to create classified results. For example, if unclassified search terms including "president assassination British visit" returned news clippings and RSS feeds of a forthcoming presidential visit to 10 Downing Street, these public results are still unclassified but the context is clear and may be classified.

[2636] The systems metasearch is a novel and useful standalone service to uncover unknown or concealed relationships, and connect the dots. It is useful to augment standard Internet search engines because it enables both search divergence for knowledge discovery and search convergence for assessing information integrity, the full process necessary for connecting the dots. Search divergence extends search with aggregation, inference and data-to-data interaction beyond simple content into the realm of context and concept.

[2637] One of the main tools in the present knowledge expander (KE) engine is its ability to control granular data with automatic content analysis / filtering and tagging of all the select content SC data and unknown data elements in every document, data stream or input document.

[2638] The knowledge expander KE engine: (1) automatically creates a list of search terms; (2) automatically cleans the noise from the search term list; (3) automatically selects from the list of search terms those that will be sent to the next search cycle (preexisting rules for doing the selection) -example: select for search only the items found which are in the "names category" and "locations category"- from the list of search terms - for example the names "Redhouse," "Kohn," "Miami Beach," and "Ft

Lauderdale" will be sent to the next search cycle; (4) conduct a divergence search - limit set at 200 supplemental documents (supple doc). Each supple doc goes through an automatic content analysis / filtering and tagging of all the data elements in the document. Key words are located in the supple docs by filtering of categories. The process cleans noise from the extract key words from the supple docs. Rules select which key words will be fed into next search cycle. For example, if 1 of the 200 documents had the name "Bob Smith" location "Sarasota" (as long as the rules for selection were not changed the next search will be fed with search terms-"Bob Smith" and "Sarasota") search is expanding - a diverging search. The last process wherein Bob Smith is added to the KE engine is a convergence filtering process because the search results will converge on Bob Smith.

C.8 Benefits of the Metasearch - Knowledge Expander

[2639] Typical search and data mining tools presupposes that the user already knows part of the answer. The user must know how to phrase the question (search terms ir query) in order to get a proper answer. However, this is only partial solution to knowledge management in that does not begin to address the real issues needed to connect the dots. This knowledge management technique as called herein "convergent" because the operator starts with a known entity and the search confirms or denies the basic entity or gathers additional information in depth on the entity topic. The present knowledge expander search, with various modules operating together, automates the convergent process until no additional information is found on the topic.

[2640] The present system and process allows knowledge management as a divergent exploration. The basic starting point is any data stream (RSS, blog, documents, a fire hose of data) or multiple data streams, or a collection of search terms for an automated process. A single term defines a manual process with a primary search term. The automated process redacts the data stream(s) and looks

for content and context of importance creating a list of primary search terms, metadata contextual filters, and noise reduction tools in order to expand the breath of knowledge with new links and references. Alternate spellings, misspellings, partial matches, duplicates, and other fuzzy technology is used to graft depth to a convergent search or prune a divergent search. In the case of a divergent search, the process is specifically exploring for secondary search terms and information that is not part of the primary search terms. Noise and duplicates are pruned. All hits and information located with URLs, X-links, or other pointers is retained for a chain-of-custody source-indicating data, reproducibility, and human intelligence analysis. Consolidated results are insufficient for legal exploration, search warrants, or plans of action, so the retention is critical to build a rationale for action and review the relationships, connections, and networks of people.

[2641] The search is extended with the primary search terms, the secondary search terms, and combinations and permutations thereof. Duplicates and non-duplicates are pruned per the selection of convergence or divergence. Tertiary search terms and information is aggregated in the same manner as the secondary search terms and secondary information. This process would repeat unless manually interrupted, terminated by a timer, manually reviewed midstream and grafted or pruned, or terminated by repetition or lack of new information. When no new information is found both the convergent and divergent search are stopped. The termination rationale is inverted. In the case of convergence, the latest results are only new information that is not wanted and adds no depth to the knowledge. In the case of divergence, the latest results are only old information and represent no new breadth to the knowledge.

C.9 Information Life Cycle Engine

[2642] The information life cycle engine has an input configuration which is saved, an indication of the source of the information and text, metadata, data streams, blogs, RSS (Release Simple

Syndication or Rich Site Summary), or a compound document. The process is a reduction of the information input into the system into its major elements by format, style, type and data type. Compound documents must be taken apart to basic identified types with tags and metadata separated out.

[2643] FIG. C-1 shows an input file 2001 (an information file or data stream) having therein text A, Text B, various tags relative to the information in the text (such as paragraph numbers), metadata associated with items and data objects in the document and image elements. The input file is subjected to a reduction or deconstruction step 2002 which creates an expanded information document which includes source data, and file map of the various elements, text A, text B, metadata, tags, audio and image. Of course, multiple audio files, image files, metadata files, tags etc. may be included in the expanded information document represented at 2004. The Editor as an Information Expander section and following sections provides greater detail of the reduction step 2002.

[2644] Further defining the source as an information file or data stream is important. "Text" is a basic defined type. The information life cycle engine automatically processes the information document. The process includes selection, extraction and categorization. For example, a document having security sensitive words or selected content (sel. cont. or SC in the figures), phrases, images or sounds, identified as secured content by Department of Defense pre-classification must be processed with a step of filtering, identifying tear lines between classified material (contextual filter or parsing technique) and ultimate identification to a base level of classification. Words, phrases, user defined words and a group list of words, phrases, combination and occurrences within a defined range are employed to identify the security (priority) level of the information document being processed. Pattern and categorization, dictionary categorization, and syntactic categorization is employed.

[2645] The information document is also parsed or separated apart by syntax, that is, words, sentences, quotations, parenthesis, other types of textual delineation, and instruction. Complex phrases are resolved for overlapping security levels and complex meanings. Categorization is employed to determine the level of security of the information. Multi tier-overlapping and hierarchical tagging is employed. Any dispute automatically identified by the system upgrades the entire information document 2004 to the next highest security level. The next step in information life cycle engine is to extract and disperse various versions of the document 2004 elements along tear lines or other types of delineation specified by the operator.

[2646] Extracted data objects or elements may be replaced by tags, codes, or null field indicators. The system may create various versions of redacted output documents. The base document, redacted of all critical or security sensitive information, should be stored along with a recovery file. The recovery file may include maps to permit a party with a preferred security clearance to recover completely the base or original source document. A metadata table is created with an intermediate resultant document. Reports are generated showing encoding, dispersion of the data and storage location. Lists are also created showing selections of data redacted from the source document, encoding of the data and synopsis of the data. Additional outputs include a recovery file and a categorized meta search file. Displays are permitted either locally or to test the discovery results.

[2647] FIG. C-2 shows information life cycle search module. At the beginning of the program flow in FIG. C-2, an input function 2006 is noted which includes information regarding the source of the information document or "source document" which source doc may, at the operator's discretion, include search terms 2008, data stream 2010 or a source information document file 2012. The search terms may include one or more SC or select content or security sensitive content (sec-con) which is of

interest to the user. If search terms are input, these terms are supplied to a summation or discrimination function 2014 (AND, OR or NOT AND operator). In other words, the operator (or the system automatically by default) could conduct an AND conjunctive search obtaining additional information from other documents or an OR exclusionary search showing and processing only the items that are identical are substantially similar from source doc and filter outputs. The AND operation may operate on the terms applied to discrimination function 2014 (e.g., the terms from input 2008 must match one output of the simple filter 2018) or may refer to a summation of all terms from input 2008 and simple filter 2018 and metasearch term engine 2020. The OR function has the same options. The NOT AND discriminator operates only to extract and use terms output from metasearch term engine 2020 which do not match search terms 2008 and the output of simple filter 2018. Some select content SC (or sec-con) must fall within the prioritized hierarchical taxonomic class or classes function 2030 in the NOT AND search. The non-matching search term engine is discussed later in connection with the Expansive Unknown Search Engine. A substantial similarity test, as a subroutine in the discriminator 2014, may ignore duplicates and may truncate search terms by eliminating suffixes (and possibly prefixes) from the search term processor. Input IN 2016 recognizes that the operator may provide a manual input into discrimination function 2014. If data stream 2010 or an information source file or document 2012 are input, these information documents may be processed through one or more simple filters 2018 extracting white list terms (inclusive lists) or black list terms (exclusive lists) or terms not found in dictionaries, the result of which is supplied to the summation-differentiation function 2014. See the Simple Filter and Complex Filter sections C.12, C.13 below. In addition, data stream 2010 and document source file 2012 are submitted to a metasearch term engine 2020. The more highly organized metasearch term engine 2020 filters the data stream or source document with a contextual filter 2022, a conceptual filter 2024,

a sensitivity level filter 2026, a hierarchical taxonomic or category analysis 2030 and a parsing algorithm for word count, line, paragraph, or DOL 2032 (contextual filter with range setting module). The order of these metasearch filters may be changed. The filter modules are contextual, semiotic and taxonomic filters. Semiotic include syntactic, semantic and pragmatic features (discussed later in section C.13). Sensitivity level 2026 contemplates potential manual input IN 2028 which establishes the depth or expansive nature of either the contextual filter or conceptual filter. The sensitivity filter and the hierarchical taxonomic filters are used to focus the search in that the operator can select hierarchical taxonomic levels or terms of concern, for example, names of terrorists - priority 1, terrorist location - priority 2, date of source doc - one week - priority 3, temporal (date) - 1 month - priority 4, terrorist organization - priority 5. Each "priority" is a SC selection sensitivity factor (or sec-con factor) and the class of the information is a taxonomic analysis, the result of which is a prioritized hierarchical taxonomic classification system and filter for the search. The result of this meta or more highly organized search term engine 2020 is supplied to the summation-discrimination function 2014.

[2648] The system then operates on convergent or divergent function 2034 which includes, optionally, input IN 2036 from an operator. The convergent or divergent function 2034 enables the operator to only search for converging search terms or select a more divergent search routine. Function 2038 results in a primary set of search terms.

[2649] An element of the information life cycle engine is parsing or deconstructing the original source document or information file. The structure of the source document DOM metadata and compound file format must be simplified and the source document must be broken into its atomic types such as markup data, tags, metadata, links, hyperlinks, references, comment, differing data types, purpose and format. This parsing is discussed later in connection with the DOM data structure in section

C.18, among others. Categorization is applied as a resolution or coding to a single overriding security level. Priority is given to the most complex or highest security level. For example, a document with "Bin Laden" as compared with "Bin" next to "Laden" results in "Bin" being identified at the secret "S" level and "Laden" being identified at the classified "C" level wherein "Bin Laden" is classified at the top secret "TS" level. (The security classes including TS top secret, S secret, C classified and UC unclassified). The resulting document would be identified as top secret TS because "Bin Laden" represents the highest level in that paragraph (format or contextual indicator). Adjacent permutations of a linear nature could also be employed. For example, "Khalid Sheik Mohamed of Syria " results in unknown [Khalid], Sheik [title of middle eastern person], and Mohamed [name] "of" [preposition] and "Syria" [geographic territory]. The resulting analysis results in an up coding of categorization to a next higher security level (e.g., S to TS) due to the contextual range grouping of the SC.

[2650] The system may employ a "My Group" list or profile which identifies key items and triggers a certain result based upon the "My Group" pre-set profile list. The profile is an SC profile of interest or an sec-con profile of interest. The My Group list would include a profile name such as "first strike capacity" and a list of key terms under that profile such as "nuclear, ballistic, submarine" and a "range" such that anytime that key word is found in a paragraph, a certain classification is assigned. The "range" being the paragraph in the source document. The Profile Name could be used as a search term for a divergence search (discussed later) or a link to a dictionary of search terms. The search terms are either content-derived in nature or conceptual in nature. Certain contextually-derived results, see Bin Laden example above, may automatically trigger use of a specially designated My Group search profile. Further, the Profile Name can be linked to another list of synonyms associated with the specific terms nuclear, ballistic, submarine. Contextual and conceptual terms are employed in this manner.

C.10 Information Life Cycle Searches

[2651] The input into the information life cycle search (Fig. C-2) may be one or more search terms, a data stream such as a blog, RSS, or a string of data, or information document (DOM). FIG. C-2 graphically shows the procedure for the information life cycle search. The system pre-processes the input and obtains metasearch terms both in a contextual manner and conceptual manner. These are identified with filters discussed later on. Sensitivity levels are set and the information is categorized both in a contextual manner and a conceptual manner. Ranges of data representing format choices are used such as lines, sentences, DOL's (lines of data) and paragraphs. Thereafter, the information life cycle search engine modifies the search terms and the user inputs either a convergent instruction or a divergent instruction. The resulting search terms are then supplemented according to the convergent or divergent system.

[2652] The search for the primary search terms (convergent or divergent) results in a federated or confederated as well as a distributed supplemental search term group. Hyperlinks, URL, network references, SQL are submitted to one or more of the following: search engines, databases, data warehouses, addressable data elements, artificial intelligence, data mining sources, text storage, method data storage, indexes, libraries, catalogs as well as other data structures.

[2653] The next step involves ascertaining the depth of the search by identifying the number of hits for the first search term and the last search term, the time involved to compile the search, the time involved until no changes, the number of iterations involved as based upon a divergence search as separate from a convergence search. The output is a consolidated search result list.

[2654] FIG. C-3 starts with an input being primary search terms 2038. The search engine module then executes function 2040 which is either a federated, consolidated, or distributed search. A federated

search unites the search terms and additional supplemental documents generated and located from those search terms together. A consolidated search is simply a composite of all the search documents obtained without any relationship or affiliation. A distributed search seeks to widely engage a large number of data structures for the search. Operator input IN 2042 directs the selection of the search function 2040. Search function 2040 operates by sending the primary search terms to additional data collection targets such as search engines 2044 (Google, Yahoo, etc.), databases 2046 (representing a plurality of databases), data warehouses such as dictionaries or other sources 2048, addressable data structures 2049, artificial intelligence or mining operation functions 2050 and other data collections, data structures, indices, etc. The mining operation 2050 may access further tertiary data structures 2052. The search function 2040 applies these primary search terms to text stores 2054, metadata stores 2056, indexes 2058, libraries of words, terms, images, data or data objects 2060 and catalogs 2062. The results are compiled back to search function 2040. Step 2064 is a depth or search depth control in which the operator by input IN 2066 delineates either as a pre-set or a default or an active real-time control the number of hits needed from each of the data structures accessed by search function 2040, the amount of time for the search, the amount of time when no change is found from the retrieved documents, the number of iterations (repetitions), and whether the primary goal is a divergent search or a convergent search. Search depth control 2064 provides control information CNTLR 2068 back to search function 2040. The search results 2070 (supplemental documents or supple docs) are obtained either from output of search 2040 or after the depth control 2064 is applied to the search.

[2655] The third step is to take the secondary search result or the consolidated search result and apply them in a convergent manner or divergent manner. The output is either a convergent compilation of documents which are obtained via the supplemental search terms (adding known SC to the search

term group) or a divergent compilation of documents employing the same search terms. A similar output result can be obtained by using hyperlinks, URLs and other metadata aspects.

[2656] The fourth step (see FIGs. C-4 and C-5) is to build a tree or a table listing the details. The tree shows the hierarchy of the data retrieved in the convergent or divergent search and the table lists the items such as profile, time, username, link, context, etc. FIG. C-4 shows that the primary search term generates a certain supplemental group of documents whereas the secondary search results in a different group of documents and the tertiary search generates a third level group of documents. The number of searches is keyed to the time involved and the criticality of the information sought to be searched. The object of the tree building and table building (FIGs. C-4, C-5) is an exploration of information linked or associated with the initial information provided. Further, a social relationship or connect-the-dots matrix display can be obtained based upon this iterative information life cycle search.

[2657] The output or yield of the information life cycle engine module results in a convergence search wherein additional supplemental information data is obtained to corroborate or validate the initial search terms and initial information document supplied. Therefore, the sec-con is tested or the SC is tested in the respective storage mediums. A divergent search yields more data which is loosely associated or related, coupled or referenced to the initial information document. Initial search terms result in a divergent documents can be expanded upon by the hierarchical tree with a list of details.

[2658] At some point, the operator must prune the hierarchical tree, reduce unnecessary items and augment others. Some of this pruning can be automated with an inference engine or artificial intelligence module. Otherwise, the operator may manually engage in the information life cycle engine.

[2659] The last step is a relationship map between the initial document and the supplemental

documents generated through the convergent or divergent search. The final result is a map of the hierarchical search terms at various levels source, search 1, search 2, search 3 ... search n. See FIG. C-4. A further result is shown in FIG. C-5 wherein a social relationship or process relationship between the various documents is established. Some people would call this connecting the dots or relationship mapping. The relationship mapping changes the degree of separation or similarity between each information document.

[2660] FIG. C-4 begins with the primary search results 2070 and function step 2072 executes a convergent test and a divergent test on the primary search result. The basis 2074 could either be the search results themselves, a search term generation which is search terms extracted from the documents located during the search, hyperlink or x-link and URLs. The convergent test determines whether the search and recovered documents converged into a certain pattern whereas a divergent test shows that the documents, search results, hyperlinks or whatever spread to wider results. The build and display function 2076 takes the primary search terms and builds a search tree and builds a table. Operator input IN 2078 optionally enables the operator to set the scope and format of the search tree and the table. Tree 2080 is a display showing the primary search terms n, n+1, n+2 and secondary search terms m, m+1, m+2 and tertiary search terms o, o+1, o+2, etc. Iterative search function 2084 establishes how many paths or branches are executed in building the primary search tree. This may be one of the input IN 2078 functions for build and display function 2076. The iterative search 2084 may be a number of times counted in branches or may be certain time unit g, h, i, j. Table 2082 is a detail showing the search term including profile, time, username, link and context of the search term. Function 2086 is this first level output report showing the search terms for the information life cycle engine.

[2661] FIG. C-5 begins with the first level output for the search term 2086. Function 2090 is

a convergent display which highlights the same, similar or closely related documents. Also, an output ratio 2092 may show how many of the documents are related by word, term, image segment, hyperlink or metadata factors. Step 2094 is a divergent display which is the opposite of the convergent display. Output factors 2096 may show how many divergent documents having little or no connection to the primary search terms have been discovered by the search engine. Function 2098 permits the operator by input IN 2100 to truncate the search or supplement the search. Function 2102 repeats the search on command of the operator. Step 2104 categorizes the search results by key terms (taxonomic classification) and generates a map. One map is shown by tree 2106 which shows the search results or information documents as document o, o+1, o+2 as well as secondary documents p, p+1. Step 2108 develops a relationship map with associated terms, phrases, documents or hyperlinks. Relational tree 2110 shows that point or document p+1 is related to document o+1, p and o+2. In contrast, document o is not related to any document other than o+1. The relationship tree 2110 is sometimes called a "connect-the-dots" map.

C.11 Search With Prioritized Classifications

[2662] FIG. C-2 shows that the operator or user can input search terms, source document or a data stream 2006 into the system. Search terms function 2008 represents the use of a search term string, document or data stream into the metasearch engine 2020 and through the classification or category analysis filter (hierarchical taxonomic system) function 2030. The sensitivity function 2026 permits the operator or user to pre-select the priority for the hierarchical taxonomic or class system. Alternatively, "My Profile" or pre-set prioritized classes may be used. The user may also select the priority and the class/subclass of the search terms immediately prior to the search. The process, in FIG. C-2, extracts search terms from input data represented by a source document or a data stream, and identifies data

elements in the input data with the classification system corresponding to n priorities therein. In one embodiment, the operator selects either a convergent search or a divergent search in function 2034. Optionally, the convergent - divergent search function 2034 may be omitted. The process then applies the prioritized and classified search terms as primary search terms in FIG. C-3, term function 2038, gathers documents and then resets n priorities in said classification system to m priorities in said classification system. This is part of the depth control function 2064 and control function 2068. Secondary search terms from the input data elements (terms 2008), classified corresponding to m priorities are again used in function 2040 and the search is repeated to gather secondary supplemental documents. The system then applies convergent - divergent test function 2072 (FIG. C-4) on all supplemental documents to find convergent or divergent characteristics of the gathered documents. The search repeats until a search end parameter is met, such as all supplemental documents exceed a predetermined number, a predetermined number of data elements from all supplemental documents fulfill a predetermined number of priorities in said classification system, a lapse of a predetermined time, a predetermined number of data elements from all supplemental documents fulfill a predetermined number of priorities in said classification system, among other search end parameters discussed herein.

[2663] Optionally, the system and process may omit convergence - divergence function input 2034, 2036, and select n priorities from the prioritized hierarchical classification system, extract search terms from input data (functions 2006, 2008, 2010, 2012) represented by a source document or a data stream or a search term string, and identify data elements in said input data with said classification system corresponding to the n priorities. The process then searches data collections with said search terms and gathering supplemental documents based thereon. FIG. C-3. The process then extracts secondary search terms as data elements from the supplemental documents with the classification system

corresponding to n priorities and repeats the search of data collections with the secondary search terms. Secondary supplemental documents are then gathered. See FIG. C-5, truncate, supplement and search function 2098, and repeat search function 2102. Such an optional search ends as described above.

C.12 Simple Filters

[2664] The user, prior to initiating the knowledge expander select content (SC) engine module, may be given a choice of filtering out or identifying all SC data objects or sec-con objects with white lists or black lists or a combination thereof as a content filter. Uncommon words, terms or data elements not found in the dictionary, geographic term lists or name lists and other lists are located with a negative content filters used to discover unknown data elements. The negative content filters can be added to the metasearch term engine such that these unknown words, terms or data elements are tagged "unknown" and assigned to the hierarchical taxonomic class as an unknown data element at the hierarchical taxonomic classification level (assigned to a class of words, terms, etc,. and assigned an SC sensitivity level - that is - a hierarchical marker). The search terms extracted from the source doc are supplemented with white lists or black lists words, terms, etc. (content filters additions) and the supplemented SC or select content words or terms are filtered out to identify the uncommon or unknown words. Of course, the user may be required to manually input all SC words or download the SC word filter from the Internet or another secure network system or LAN. A "profile" of SC with the hierarchical taxonomic markers can be employed by the user-operator. For select content systems having multiple levels of importance (which may be multiple security levels or multiple levels of SC importance in an organization or related to the organization, for example, primary competitors compared with secondary competitors), a plurality of filters would be created, each filter associated with a different SC level. Further, multiple SC levels may require, remainder SC document or data stores (unknown data stores)

and a plurality of extracted data stores. Multiple levels are similar to the MLS security content (sec-con) system described in section D.4 to D.9.

[2665] The ability of the program to locate select content or SC words or characters (or sec-con) can be enhanced by using a telephone book, properly dissected, to identify a collection of last names. Cities and towns and street names can also be identified in this manner. The compilation of last names and cities, towns and streets can be used as a list of critical, SC words. The filter is represented by this compilation of words. Similar techniques may be used to create filters for scientific words, or words unique to a certain industry, or country.

C.13 Complex Filters

[2666] There is a need to construct filters which supplement the initial list or compilation of SC (or sec-con) words, characters, icons and data objects (herein "word" or "data object" or "word/object"). The need arises either due to the fact that the initial SC search term word/object list is incomplete, or that the author of the initial list is concerned that the list is too limited or in order to defeat an attack or an inference engine "reverse engineering" at the resultant expanded SC document. In a test mode, the complex filter is used to determine how "secret" or "special" the sec-con or SC is compared to outside databases and data collections. Further, the incorporation of a filter generator for SC or sec-con search terms enhances the user friendliness of the program. In one embodiment, the program is configured as an editor compiler to screen and build enhanced SC doc or doc collection from a source document. The user selects, at his option, functional aspects which include: compliance with laws (an application of a type of filter, e.g. HIPAA, GLB, Oxley-Sarbanes, EU privacy, executive orders); privacy (another type of filter which locates SC terms, for example, social security numbers, see also, EU policy); search for and supplement filter; pay per view (which enables the user to buy missing sensitive information (for

commercial purposes); survival (which creates a distributed and dispersed copy of the user's source document and other stored documents and items using predetermined storage facilities); security (which triggers the various security routines); and storing (which permits the user to select which of the several storage options the extracted SC data/objects should be employed in the dispersal.

[2667] The filter routine or module diagrammatically illustrated in FIG. C-6 is useful in compiling a SC or sec-con search term filter which separates both the sensitive word/objects and contextual and semiotic and taxonomic aspects of the initial list of SC word/objects. The filter works in conjunction with a compilation of data, typically located on a network which could be private or public. In low level SC situations, the filter may access Internet databases to gather additional data for the filter. In more robust SC systems, the filter could access private data bases (one located at the same organization level as the user) and build or compile the additional SC word/objects. The filter program 950 in FIG. C-6 begins with step 952 which compiles the initial list of SC word/objects. In 954, the initial list is supplemented with dictionaries, phone books, corporate records (to obtain subsidiary data and trade names) and thesaurus data. This is a content filter. Each of these represent different compilations of data and the added data is added to the initial list of SC word/objects. In 956 a search is conducted on a network, usually through a search engine, to gather excerpts near and abut the keywords. This is a range or contextual filtering aspect. These keywords are the initial SC word/objects. Statistical algorithms are applied to gather non-common word/objects which are associated with the keywords as found in the additional data compilations. The goal of the adaptive complex filter is to obtain contextual, semiotic and taxonomic words, characters or data objects from the compilation of additional data related to the SC words, characters or data objects. Semiotic is a general philosophical theory of signs and symbols (read language and words and objects) that especially deals with their

function. Semiotics include syntactics, semantics and pragmatics. Syntactics is the formal relationship between signs. Semantics is the meaning of signs and pragmatics is the relationship between signs and their users, such as the relationship of sentences to their environment. Taxonomy is the scientific classification and categorization of items. Therefore as an example, a search through the Internet on Google search engine under "Bin Laden" may show a number of uncommon (non-dictionary words)(content filter-based search) within 200 words of the target "Bin Laden" (a contextual filter-based search). This search string would gather documents from the Google search and copy 200 words on either side of "Bin Laden" and then extract only non-dictionary words into a supplemental SC term list. This type of filter algorithm looks for contextual matters close or near to the target. The search is semiotic and statistical in nature. Additionally, the initial supplemental list would identify the Bin Laden is an Arab and this classification (a taxonomic aspect) can be used to expand the list for the filter. The algorithm may include a simple command to gather all 10 words on either side of Bin Laden. This is a pure contextual search and the "10 word" range or format aspect is a statistical number. From the supplemental list, all pronouns, prepositions and conjunctions may be eliminated. Spiders or robots may be used in the gathering of the contextual and semiotic filter data. The contextual, semiotic and taxonomic words, characters or data objects from the compilation of additional data are all related to the initial list of SC words, characters or data objects.

[2668] Step 958 compiles the adaptive or complex filter. The above noted contextual, semiotic and taxonomic filter is adaptive since it can be used to expand (and potentially contract or reduce) and adapt an existing list of SC or sec-con word/objects to a larger list which better expands the information content of the source document and supplies SC term data to an inference engine. As a test, the sec-con storage is tested with the complex filter to determine whether the security level is appropriate. For an

SC test, the complex filter is used to ascertain wether the SC store should be supplemented with additional outside data or that the H-tax class be changed. Step 959 repeats the filter gathering and compilation for various levels of SC inquiries. Higher SC inquiries require a broader search (1000 uncommon words near Bin Laden and add all Arabic and sub-Asian continent cities). Orthogonal SC groups (those groups having the same level, e.g. S Secret, with each other but being different organizations, e.g, Department of Defense compared to the FBI) often have different methods to keep SC data separate between compartments.

[2669] The adaptive filter can be set to automatically gather additive SC word/objects. The test module may be run periodically. The system, with a basic filter, may identify a SC word in a paragraph being scanned by the initial filter. This SC or sec-con word may be a special word in the existing filter or may be a non-common word not found in the initial filter. The adaptive filter system may then obtain this "unknown" or "special" word (a negative filter in that the word-object is not matched to a word-object having the same content), and conduct a search through a compilation or data base of additional words, etc. Any new word/objects falling within the contextual, semiotic and taxonomic SC words, characters or data objects from the compilation of additional data (database) related to said SC words, characters or data objects are then added to the filter. The expanded filter is then used to supplement the source document.

[2670] Step 960 compiles a supplemental filter with random words, phrases, etc. in order to further defeat an inference engine reverse engineering assault on the SC supplement document matrix. In some sense, the production and use of a random filter is an encryption technique since the resultant filtered product, in order to be understood by others, must be reverse filtered or decrypted to reveal the source doc and enhanced doc matrix at the appropriate SC level. Nonsense words may be added to this

supplemental filter. Step 962 applies the primary filter (with the SC word/objects and the additive word/objects from the contextual et al. filter) to the source document. Step 964 extracts the SC word/objects per SC organizational level. It is noted that several filters are used, on one for each SC level, whether hierarchical or orthogonal. The extracted SC word/objects are stored as a supplemental search term doc and the system gathers supplemental documents (supple docs) which expand the knowledge base about the SC word/object. Step 966 applies the supplemental filter to the supple docs returned per search level1 with SC search terms. The system then repeats the process of SC search term generation on primary retrieved supple docs level 1, generates SC search terms per level 2, and retrieves SC level 2 supple docs. A tertiary search term generation and supple doc retrieval is possible based upon the operator's initial set-up or upon operator input at steps 964, 970. Step 968 stores the supplemental doc to permit information enhancement of the source doc. Step 970 publishes, distributes or pushes the source and all supple docs and search term data to others having a need to know.

C.14 The Editor As An Information Expander

[2671] FIGs. C-7 through C-8D diagrammatically illustrate an editor which may be employed to locate SC or sec-con word/objects in a source document and expand the knowledge base with supple docs. In one embodiment, the editor is a standalone application or a module to add onto other applications for plain text and media creation, editing, and sensitivity SC level tagging. Other types of tagging, wherein the editor supplements the initial group or subset of select content SC or sec-con sensitive words, characters, icons and data objects by categorization, taxonomy classification, privacy, security, compliance, and semiotic meaning, are also available. The editor supports a full range of document management and can be integrated into a unified infrastructure, from creation, editing, document markup, tagging, tag conversion, tag removal, context sensitivity level redaction, context

reconstitution, support for complex process work flows, and expanding the knowledge base by adding supplemental documents (supple docs) to the initial collection of source doc. The architecture assures separation of data from metadata so that no security lapses are introduced into the traditional word processing and document management cycle.

[2672] The Editor automatically and with nominal operator input (after the Editor is initialized) separates the data stream from all markup and tagging word/objects for SC knowledge generation purposes.

[2673] The interlacing of user content with metadata creates significant process, storage, distribution, and workflow security failures that are not resolved with current technologies. Current technologies include encryption, firewalls, intrusion detection, perimeter guards, and locked distribution packages.

[2674] The Editor enables text and media creation. However, all additions, deletions, changes, insertions, and reorganizations and reordering are tracked as metadata that does not become part of the document or information file. The document (information file) as seen and shown to the user represents the deliverable format. Since formatting is metadata, it is not included in the representation. Formatting, such font sizing, colors, font selection, footnotes, headers, subscripts, superscripts, line numbering, indexing, and other features characteristic of standard document preparation can be supported but are represented only as metadata. Tagging, including SC sensitivity level, categorization, taxonomy classification, privacy, security, compliance, and semiotic meaning are also represented only as metadata. This separation of representation from meta-representation is critical for creating the infrastructure for SC knowledge expansion, secure information sharing, privacy, security, and compliance.

[2675] The editor is currently set in a WINDOWS environment. Pulldown menus provide access

to formatting and tagging features. The document, from source, precursor (marked and tagged but not yet filtered or extracted) and resultant final versions for each SC sensitivity level, as seen and represented to the user, is distributed in resultant final form, thereby assuring SC knowledge level compliance. No hierarchical, hidden, encapsulated, linked, associated, or referential information is part of the data stream, file, or storage.

[2676] Metadata (such as formatting, such font sizing, colors, font selection, footnotes, headers, subscripts, superscripts, line numbering, indexing, and other features characteristic of standard document preparation) is usually hidden from the user. This supplemental metadata information contains all markup, tagging, formatting, and process support information for the editing process and enables immediate granular distribution of the data stream subject to the needed SC compliance rules. In other words, the data stream can be automatically processed with other functions to satisfy multiple competing requirements and SC sensitivity levels.

[2677] FIGs. C-7, C-8A, C-8B, C-8C and C - 8D are discussed concurrently herein. FIG. C-7 is a basic flow chart for one embodiment of the Editor. Editor program 972 begins with obtaining the source document 974. Of course, the source document may be any type of document or may be a data stream. Of course, the data stream is typically delimited by start and stop characters or codes. Hence, the term "data stream" is similar to "source document" herein and in the claims. Step or function 976 obtains one or more filters for one or more SC or security sec-con or sensitivity levels. Step 978 screens or processed the source document with the filter(s). For example, the source document in FIG. C-8A in window 991 has text regions 993, 994, 995 and 996. In step 979, the Editor displays, in situ (in the displayed document), the filtered identified SC or sec-con material and conforms the precursor document to the SC or sec-con sensitivity protocols. As shown, the SC or sec-con sensitivity protocols are security

level protocols for the system within which the Editor is employed as an information processing tool. SC or sec-con sensitivity levels 1, 2, 3, etc. correlate to security levels TS, S, C and UC for top secret, secret, classified and unclassified. FIG. C-8B shows that the address data 993 is marked TS (top secret), region 994 is displayed in color A for TS coding (please note that the addressee data may also be so marked) and is "red-lined" or struck out. In an information-expansion system, the process marks, labels or tags each word. This is a content filter and tag regime. The user may manually upgrade or downgrade SC or sec-con levels. Region 995 is displayed as presented in the source document and is labeled U (unclassified) and region 996 is shown in color B, is redlined and is labeled S. Labels TS, S, C (classified) and U are the established security labeling protocol used by the organization employing the Editor. The same tagging for each word or character in the source doc can be used. Other labeling schemes for SC elements may be employed. Color is used to assist the user to select (and in some non-standard cases, deselect) the SC or sec-con data marked by the editor. Redline is used to inform the user that the filter(s) will tag and extract the marked data. Labels are used to permit the entity using the editor to employ standard tear line protocol. Tear line protocol is a contextual tool. Any data beneath a security classification of the user is under the tear line and the data is permitted to be distributed to the lower security cleared user. Of course, electronic distribution of secure data need not use the hard copy or print version of the tear line. However, this nomenclature referring to the tear line is used in the prior art systems.

[2678] Step 980 accepts the user's manual changes (typically upgrades) to the precursor document. These manual changes are displayed, redlined, colored and labeled. Step 982 inserts the SC or sec-con tags (or security label TS, S, C and U for the sample doc in FIGs. C-8A, C-8B, C-8C and C-8D) as discussed above. Step 984 notes that the system takes certain metadata such as author, date-time,

version history, change history, etc. and converts this meta data into ordinary text, marks that data at the necessary security level or SC or sec-con sensitivity level and labels the meta data. Step 986 permits the user to add (or omit) placeholders into the final document. FIG. C-8C shows placeholders as black lines or as XXXXX symbols (or other symbols) wherein the sensitive SC or sec-con text is not shown but some replacement markers are shown. The byline in region 1003 show "sanitized document." The byline 1003 in FIG. C-8B lists the sec-con security level (or SC level of priority) and the color representation.

[2679] In a specific knowledge expander engine, each word, term and character is labeled or tagged with a content filter to show data object type (name, town, credit card number, etc.) and an SC or sec-con sensitivity level (a "tax" or taxonomic classification and sensitivity level). See step 982. The user may add tags (step 979, 980). Meta data is also tagged or labeled.

[2680] Step 988 activates the filter, extracts the sensitive SC or sec-con data and temporarily stores the extracted SC or sec-con data. The expansive knowledge engine also stores negative filter results which are those words, terms or characters, images, not found in the context filters. Step 990 displays the filtered document and the user may view the filtered document at each SC or sec-con level. Therefore, the user, before transmitting a secured email (or letter) doc may look at the source (FIG. C-8A), may look at the TS level (FIG. C-8A) without the redline strike out but with security labels and colors, may look at the T level revealing regions 996 and 994 but not regions 993 and 994 (which are TS coded regions), and look at U versions as shown in FIG. C-8C. Step 992 disperses the extracted data and the remainder data or disperses partial versions of the document (those partial versions formatted and containing only data at or above the target security level (all TS level data (which includes TS, S, C and U data), or all S data (comprising S, C and U) or all C data and U)).

[2681] In step 979, the SC or sec-con level protocol determines whether single words are granularly classified (TS, S, etc.) or whether a line is classified (context filter), or whether an entire paragraph is classified (see FIG. C-8B). If a commercial/privacy filter is used to exclude all social security numbers, the organizational protocol is set at a granular level to exclude just social security numbers. Different group protocols use algorithms to mark, filter and extract adjunctive security sensitive words, characters, icons and data objects near the target SC or sec-con sensitive words, characters, icons and data objects. The SC or sec-con words may be security sensitive words, characters or data objects defined by compliance with law, regulation or policy, privacy, national, organizational or private security concerns. For example, "Bin Laden" is the target sensitive word in FIG. C-8B and this classifies the entire paragraph as TS level. The other words in the paragraph are adjunctive word/objects.

[2682] In a knowledge expander mode, the SC or sec-con filters are applied in a negative manner as follows: (a) in the user set-up, the user establishes taxonomic categories or classifications and sets the selection priority of the classes; (b) the source document or source data stream is broken apart to separate all metadata; (c) the source document (may be a data stream) is processed by a taxonomic filter which tags or labels each word or data element with the taxonomic classification and the sensitivity or priority label (multiple "tax" tags, overlapping "tax" tags and "unknown" tags are possible); (d) after labeling, a content and contextual filter is used to separate out the unknown words or data elements (a negative filter). The resulting negative list of search terms is used in various search engines for both public and private data sources, to compile a compilation of supplemental documents (supple docs) and, thereafter, the supple docs are re-cycled through the H-tax and priority filter, then the content and contextual filter and a secondary supple doc collection is obtained. The primary and secondary supple doc collection

represents the expanded knowledge search not typically found with commonly available search engines and processing techniques.

C.15 Document Object Model (DOM) - Protection and Processing

[2683] The search for expanding the knowledge base from a single source document to a larger compilation of docs has changed from locating content (see the prior art GOOGLE search engine) to expanding the search and search results for concept and context. Sequential text files are the exception rather than the norm. Flat, plain, and sequential files would have disappeared entirely from all but transitional processing steps except for the recent success of HTML web sites and the desire for storage of complex data into sequential XML formats. In spite of the apparent linearity of HTML and XML, in practice these flat files participate in a greater complex hierarchy of structured data mapped by object models. The object models blur the lines between content, concept, and context such that effective security requires a broader stroke than merely encapsulating content with encryption and limiting access with tokens or encrypted certificates.

[2684] Linkages to external files, style sheets, and embedded applications or scripts undermine the simplicity of HTML and XML flat formats and compromise point security. Even structured field or line and record-oriented file formats have given way to more complex data storage models. It is insufficient to view security of content and files in terms of encryption and encapsulation alone. Structured object models mix content with metadata and methods such that non-granular access—that is, either/or barrier-based access through encryption keys, dongles, and passwords—undermines any concept of effective security.

[2685] Furthermore, simplistic document management and access control overlook the multiple purposes for each compound data document and the adverse impact on organizational processes and

work flows. Barrier-based security also fails from any Pacman-style attack, where the barrier, once breached not only provides full access to the once-protected interior also interferes with analysis of the attack and observation of how to prevent the ongoing attack. Granular multi-level control of user data, metadata, data stored through the specifications of a hierarchical data object model, and methods underscores the new security paradigm.

[2686] All data sources important to data process workflow are non-linear, non-sequential, and not standalone in that the data sources are interconnected to or required by other data sources. This includes databases, structured documents, desktop application user files, hierarchies of data structures, and work flows. The most advanced data workflow and the focus of attention is the object-oriented models used in data processing today which comprise a cascade of events rather than a single point operation. This complicates SC or sec-con data expansion activities to promote security, survivability, privacy, confidentiality, and anonymity. The present invention improves the security of complex document object models and interdependent workflow by expanding the knowledge base form a source document, thereby testing the classifications levels and generally expanding the knowledge base of a user form the simple source doc.

[2687] There are only a handful of counterexamples to complex data structures, mostly monolithic file structures and simplistic processes. This includes text files, raw binary image files, and lists. These are typically inputs to older or uncomplicated computer activities; they do not reflect the complexity and interrelationships consistent with and necessary for most critical networked data processing activities. Examples of flat files are text files, binary images, and lists. Plain-text documents are used only as temporarily or as conversion paths for other activities. Binary graphics are employed for their specific simplicity, speed of display, and small size. It should be noted that they (BMP, GIF,

and other formats represent the bulk of web images) are usually stored in an inverted backward last-to-first sequence. List files are rarely important and standalone files are often a temporary part of another process. One of the most ubiquitous of plain-text files, the HTML web page, is rarely a simple text file, but a circular connection to many other like files and one part of a more complex hierarchy. A relative of lists is the field-oriented record structure. This is web page usually a grid-like storage of linear data. However, even a table grid, multi-dimensional indexing, SQL query concept is giving way to object-oriented post-relational database storage methods based on object models in order to augment functionality, speed of performance, cross-platform and application functionality, and compete with easier to use user and developer products. Even the image files are becoming increasingly complex. Hierarchical images formats with vector graphics compress motion and curves into small packages. Examples include Corel Draw, Macromedia Flash, Adobe Photoshop, and Microsoft Photo. These of course contain proprietary and unintentionally-distributed information. Increased reliance on reliable data storage infrastructure and networked storage technologies is enabling the transition to data storage based on object models.

[2688] FIG. C-9 shows the root, branch, and leaf paradigm of this principal data storage structure. See root 1012, content leaf 1014, branches 1016, 1018 and leaf 1020. The object model refers to the layout or the map (a blueprint supplied by the document object model (DOM) vendor) of how the data is potentially stored in what is definitely a linear file. The stored file is the document object structure containing the data whereas the model is the schema representation. The model FIG. C-9 is just a blueprint for an empty data structure.

[2689] The data structure is stored as a binary file populated with data representing a subset of that blueprint. The data file is often referred to as the document binary file so as to make clear that it is

not a plain-text file, not in user-friendly format, and generally readable by an ASCII reader only in discontinuous chunks. The model and the structure are not the same. The model (FIG. C-9) does not represent a security threat in itself; it just represents how to find and define data stored within an actual data structure. It is the data structure in memory (the source document) or stored as a file that is the security threat. Usually, the file containing the data structure gives enough clues to the purpose, methods, and sources unless addressed by a multi-level security scheme attuned to the complexity of the object model. Although this "file" is stored as linear flat file, the extended structures is dependent on the hierarchical collection of potentially infinite branch and leaf references. Despite this complexity, there are clear reasons based on simplicity for this hierarchical structure, not the least of which is flexibility, self-documentation, and backwards/forwards compatibility.

[2690] The subtle differences between a plain-text file, a file containing lightly structured data, the schema, and a file containing data within an object structure becomes very important for security. When files are distributed and those files each contain data within object structures, workflow is complex and cannot be effectively protected with barrier-based security without complicating or disrupting operations. For these reasons, internalized security reflecting leaf content, structural paths, and the mesh of inter-relatedness among the paths, leaves, and external sources becomes the next paradigm for implementing effective content-level and application-level security. Consider the data structure defined by an object model as an organizing container. The contents within can be empty, or collections of containers, with more containers within. It is a security sieve with traditional encryption and the requisite inter-process work flows. The leafs and the security of the leaves does not secure a chain of evidence increasingly necessary in modern data processing activity.

[2691] Enhanced security must reflect this distributed requirement since the data sources are not

single point sources, but complex relational, object-oriented, or hierarchical. In addition, data access and processing is approaching a worldwide distributed infrastructure, and completion transcends single places, times, and events. When the security problem is dispersed, the security solution cannot be monolithic either but must reflect the dispersed distribution and hierarchical complexity of the data and process. Location is not the problem, so metaphorical perimeter walls are not the answer. To treat security too as a monolithic, static, and walled solution when the security problem is granular and dispersed within a flexible time frame misses its true need. Effective data security must reflect five informational attributes in a newer paradigm for security. The five informational attributes are listed below and examples of the attributes are also listed. For each select content SC or sec-con sensitive organization, the data structure must be analyzed and the five attributes must be applied to each root, branch and leaf to ascertain the level of SC or sec-con sensitivity for that item. For example applying a security schema to the SC knowledge expander problem, a TS level may be established by applying the five attributes that all audio files are "security safe" for that SC or sec-con level but these audio files will not be downgraded or released to a lower SC or sec-con level. Therefore the meta data representing the audio file is designated TS. Another example is that all machines at the SC or sec-con level S (3rd level of the TS-S-C-UC schema) are 2004 machines and programs. The organization may set, as a policy, that all MS Office program meta data need not be backward compatible beyond 2004. This organizational protocol then reduces security issues relative to the backward compatibility issue.

Informational Attributes Table for Security

Purpose
Sources and methods
Ownership
Date or timeliness
Content

Purpose Classification - Exemplary Table

.backwards compatibility (purpose: communication across machine platforms
.background color (purpose: visual presentation)
.font size (purpose: visual presentation)
.image
.video
.audio
.version control (purpose: source identification)
.etc.

Sources and Methods Classification - Exemplary Table

.origin plain text
.origin entire document
.image
.video
.audio

Ownership Classification - Exemplary Table

.source, author
.security level initial document
.security level generating modifications to initial document
.hierarchical, orthogonal security classification
Date or Time lines - Exemplary Table
.version control
.source identification (includes all contributing authors supplying modifications)

[2692] These five security attributes reflect not only the data content but also the point processes,

embedded resources, and work flows. In a similar manner, SC or sec-con sensitivity levels can be applied based upon time (temporal) issues, competitor or size of company, type of product (critical, sub-critical, or ancillary), etc.

[2693] This metaphor fractures complex data processing workflow. Traditional search methods point only to defined search terms, maybe with a simple truncating algorithm which deletes "s" and plural suffixes from the defined search terms. A monolithic approach to prior art searching ignores metadata, process, multi-tiered files, delivery, and storage. Data sources are not monolithic, and certainly data is not either. Distributed data, distributed processing, and widespread distribution defeats common search techniques. Access and search techniques need to be granular and multi-level, and represent the five informational attributes presented above.

C.16 Recognizing The Document Object Model (DOM)

[2694] Many word processed documents, such as those created by MS Office, binary documents which contain confidential information, typically metadata. This ranges—from small amounts of information about authorship—to the editing history complete with deletions, reviewer comments, file attributes, and source and routing information—to extraneous baggage from documents previously edited during the same session.

[2695] A multi-faceted SC or sec-con workflow process becomes an issue over control of distribution by document type, recognition and categorization of all user content defined by security exons (discussed later), removal of non-coding or non-activating security introns (discussed later), preparation and distribution by SC or sec-con sensitive levels, content certification and accreditation (C&A) subject to conversion to primitive and certifiable file formats, distribution in print-representative-like packages, with guarded ingress and egress of the source data files. Finally, implementation of a

knowledge expander search through granularity of document node elements by analysis for inclusion and exclusion is a far more effective method, permitting collaboration within a multiple-usage infrastructure.

[2696] It is important to recognize that there are many file types and document structures associated with a document management program such as MS Office. The structures are shared with other applets and competing products. Each program application, such as Word or Excel, create file binaries or binary files with different object structures but interchangeably read/write and import/export each other's file types, embed portions as formatted text or complete objects, or link through remote procedure calls to these other file types. These object model structures are generically called the Document Object Model (DOM). The DOM is another term for an object-oriented data storage package. The purpose for the DOM with hierarchical storage of metadata is three-fold. First, it is useful for backwards and forwards version compatibility. Second, metadata extends the document creation session from one-time event into an ongoing revisional process. Third, metadata provides order and structure otherwise notoriously difficult for inherently free-form and flexible documents.

[2697] Metadata provides backwards and forwards version compatibility. Metadata provides the necessary blueprint, format, and structure retention so documents can be revised in future editing sessions. Format information is part of the DOM metadata, although style sheets and schemas maintained in a different storage channel are valuable in HTML and XML and might aid the future transition to a different DOM.

[2698] It is incorrect to assume a static basis for any document structure, as a monolithic DOS-based file, or as an in-memory object. For example, the Excel DOM can be embedded inside a Word DOM, which selectively can then be pasted as a formatted object into a PowerPoint presentation.

Another concern that arises in almost every document is imports, pastes, and OLE imbedding of other documents and aspects of the object model corresponding to that application type. For example, a base Word document with a spreadsheet and Project waterfall chart now includes editable components referencing a different Office applications with data in a structure referenced by that corresponding application object model, in this case Word, Excel, and Project.

[2699] FIG. C-9 shows each branch or leaf can be replicated indefinitely until reaching the limits of RAM or file size. Each application has a different DOM. FIG. C-10 shows the object model for word documents circa 2004. Because of DOM evolution, with the assertion of backwards and forwards compatibility, realize that some nodes might exist in the binary document file but not every function appears within each published output because it is not used by the author.

[2700] A basic Word document file requires 18 KB on initial saving, but a full 28 KB with edits and deletions, metadata, and redlining. Footnotes, font changes, hidden text, additional changes, headers, and footers, table of content, indexing, an index, macros, .DLL add-ins, .OCX add-ins, and formulae could arbitrarily increase the file size indefinitely. This shows that MS Office security risks are reproducible at any user desktop. A hex editor used in conjunction with an initial raw ASCII file and the corresponding .DOC file also shows risks. ASCII text has only 40 characters despite the directory display of the 1KB FAT32 block. The internal encoding of the .DOC file with initial content, the binary object structure and additional metadata are partially encoded in a padded form of ASCII. The metadata displays the source location of the document, removing possible doubts of file directory structures, security based on location obscurity, and other rational workflow techniques for securing user files within the context of a network infrastructure.

[2701] Microsoft has identified thirteen key categories of metadata: Name; Initials; Organization

name; Name of originating computer (desktop); Name of network server and/or hard drive; File properties and summary information; Non-visible embedded documents; Names of previous authors; Document revisions; Document versions; Template; Hidden text; and Author comments. Some of this metadata is accessible through the Office application menu interface through menus and dialog boxes. The exploitation of this metadata in an knowledge expander search is useful. There are also the document file properties exposed by the Tools/Options pulldown menu and the User Information tab. This is not the complete list of metadata. Other visible metadata with potential search characteristics include: Footnotes; Cross-references; Table of Contents tags; Indexing tags; Hyperlinks; and Smart tags. Expect x-link and x-pointers plus style sheets and schemas within documents saved in the XML format. In addition, other undocumented structures are part of the extended and expanding Office document object models. Consider fields and mail-merge fields, which are markers for information automatically inserted by Office or by a user when opening, saving, printing, or emailing documents. These fields create a built-in facility for careless information disclosure or overt hacking. There are also the document file properties exposed by the File/Properties pulldown menu. This includes: File/properties; General; Summary; Statistics; Contents; and Custom.

[2702] The knowledge expander search also contemplates finding and using other items not specific to MS Office. Techniques for information camouflage can be used in an knowledge expander search. These are equally valid in most any desktop application, and are most relevant to presentation output rather than binary file delivery. Information camouflaged in a source document includes text set to small font sizes, such as 0 or 1, fonts set to type unlikely to be installed on the system which map to symbols or line drawing, PostScript or Unicode font sets with alternate encoding, and font color set to match the paper color or an applied background. White font on white paper hides text, black font on a

black border or shading hides text too. Text can also be hidden with graphics when the graphics are anchored to a specific location congruent with the text. Color games with text and graphics also hides the text. Macros, VBA (Visual Basic Application) codes, VBA add-ins, and applets also represent a search asset. Anything than anyone can imagine as an application can run from within MS Office, productive or destructive. Usually, these bits of code are stored as part of the document metadata. However, they also can be out-of-channel files. Either way, they can be compromised by a new code that overwrites the original. They also can be inserted through fields, formulae, or menu add-ins. Collaborative tools are the most obvious entrée.

[2703] New features in Windows and other Microsoft digital rights management (DRM) applications, such as ORAPI, ADSI, and MS IRM provide for collaboration, resiliency, and complex versioning and backup far beyond the capabilities of MS Office.

C.17 Content

[2704] The differentiation of content within an MS Office document based on initial owner and target distribution is important for search expansion. Some content will be strategic and some tactical. Content of MS Office documents transcends the actual presentation as a printed page, slide, spreadsheet, database report, email message, an index of documents, UML: or project waterfall, or organization chart. Microsoft Corporation is positioning Office as a platform for delivery of new services; it is not just about a PowerPoint presentation or a Word document printed to a facsimile. The DOM is a project plan, with a structure, with components that do things and are sensitive of their own.

[2705] For these reasons, it is important to explore the MS Office DOM factors: Content classification; Tagging; Clearance level; Data mining; Traffic analysis; Inference; Encryption; Digital Signature; Document access linked to Fortezza (an encryption program/system), PC Crypto cards,

smartcards, and n-factor authentication; Granularity; Strategic information; Tactical information; Common Criteria or NIST analysis; Covert channels; and Bell-LaPadula model conformance.

[2706] Content classification with taxonomic classes occurs with tagging for formatting with bold, indexing, and paragraph marking, explicit element tagging for HTML and XML or database and spreadsheet table, field, ranges, row, and column designations, as well as authorship techniques. Formulae and macros define ranges with informational content (contextual algorithms which link content), as well as indicate purpose and intent of the process as well as the target data. When content is tagged at the sideline, as in "eyes-only," or within-the text with any label name for clearance level, as in "<1>," this attests to a SC sensitivity level with an importance factor. For example, a subtotal of employee salaries within a pro forma business plan matched against a list of employee names compared to a bank check ledger gives away each employee's salary level; each document in isolation does not give away information until several are merged and analyzed together. Direct analysis through record relationships and sorting is one type of data mining, human intelligence through inference or statistical inference with set theory or Bayesian methods is yet another. For example, because you know that 6 employees are traveling to a conference in D.C. and two others are not in the office, you can approach a particular person who by inference is manning the station desk with a very specific social engineering attack. OneNote, InfoShare, Net Meeting and/or Live Meeting, Outlook, and Exchange with MS Project also enable workflow routing, group editing, and acceptance sign off. This information becomes part of the document metadata so that traffic analysis shows where the document originated, what changes were made and by whom, how it was routed by username, network, and IP address, who has seen it and has access to it, and all process flow and comments. One of the secure prizes of organization information thus unintentionally published is the names of people within the organization and functional roles.

[2707] Designing a knowledge expander search engine through granularity of MS Office node elements by analysis for inclusion and exclusion is an effective method. Multiple source documents create structure and semiotic meaning not in evidence with subsets. This process breaks the context to prevent useful data mining, routing inferences, and the more powerful semiotic information methods. It allows for the separation of strategic information from the tactical, so that access is granular by role, user, and other discriminators. Many academic and implemented security models are in use today, both as a straw man and for certification processes.

C.18 DOM Process Editor

[2708] Document object model (DOM) source documents, and particularly Office document modules, comprise the blueprints, process, external data sources and linkages, and materials for building the resulting presentation; the presentation content is usually the ultimate end product. The blueprints and process often are immaterial to the presentation and represent proprietary and confidential material. This DOM object model complexity and diverse accessibility creates an opportunity for the knowledge expander search engine.

[2709] Effective DOM and metadata searches use the object hierarchy structure as variously described as a binary tree, category structure, or hive. In any event, the entry point is the root or base, containing a potentially infinite number of subcategories, each with a potentially infinite number of leaf items. See FIG. C-9. The structure can be pruned, deleted, or rearranged. The items represent object-oriented information, from entire subdocuments, to relational databases, layered graphics with vector elements, to simple plain-text, to a single binary numerical element.

[2710] The process requires a parse of all branches to each and every leaf. This process is not recursive, just extensive. Each path is examined for context, each leaf for content, all nodes for external references, and everything must be viewed within the context of sources and methods, not just obvious content. The obvious content is what the user created and sees, but as you now know, that is a minor portion of the data contained within the document object structure. This is a paradigm shift is shown in the hierarchy below:

DOM Deconstruct Table

For each document (the file and structure); Access the root; For each limb; For each branch; For each sub-branch; For each leaf (item); and Process each leaf.

[2711] Preservation of the path to each leaf is important as it defines the access to that data element. The existence and/or null value of the leaf represents a security control point. The model defines, with supplemental external knowledge of the object model, potential search opportunities. The model and the content are not separate from external knowledge of sources and methods. The model and the content are part of a contextual analysis of the source document or data stream itself. The leaf the significant search control point. It is possible to review and alter the contents of the leaf within the context of the purpose of the leaf to retain functional access with multi-level SC or sec-con sensitivity.

C.19 Five Informational Attributes

[2712] The five information attributes of SC or sec-con sensitivity in context to processing include the leaf, purpose, sources and methods, ownership, date or timeliness, and content. The entity must establish protocols which rate or prioritize the five information attributes on each root, branch and leaf in the DOM source document. With the system initialized in this manner, the processing of the DOM document within the parameters of the knowledge expander search is accomplished.

[2713] Purpose: How does the purpose of the leaf provide context, purpose, or informational reference to the document as a whole or the individual leaf? Does it provide source, destination, authorship, viability, validity, verification, or integrity to the document as a whole or the individual leaf? Consider the value of processes imbedded in the document as cell formulae, a help file, or other complex routing wizard. Does it show linkages or references to other documents? What is its status or position within the document? What is its element position, as a headline, footnote, or redlined status? These seemingly minor details transcend actual content but provide clues to the following attributes.

[2714] Sources and Method: Intelligence agencies stress the confidentially of the sources and methods used to gather information. The information itself might or might not be important, but the

ongoing care of the sources and methods is important for future information gathering activities and retention of any status quo until action is initiated. In addition, the viability, validity, verification, or integrity of the document is predicated by the viability, validity, verification, or integrity of the sources and methods used to create it. In terms of the Office document, this type of information is both contextual, leaf content, and metadata. To presume that the SC search uses only doc content at the leaf misses the value of metadata and the inherent value of the object-oriented document format. For example, authorship, source, source dates, editing dates, deletions, redlining, notes, footnotes, MS hidden text, links, and other structural elements in the source doc describe when, how, where, and who created the document. This speaks to the viability, validity, verification, or integrity of the document as a whole, and can compromise past, ongoing, or future data collection efforts and operations.

[2715] Ownership: Ownership is reflected both in leaf-level content—that is obvious when a document is presented or published—but also in the metadata. Ownership is also a characteristic of file storage properties, in ring rights, file storage position, linkages, SMB or network file access rights, and HTML references. Ownership, particular the number of links, the times access and edited, numbers of hits, and the level of churning, suggests the relative importance and merit in the document.

[2716] Date-Timeliness: Date or timeliness reflects currency. The dates, in terms of edit times, access times, and frequencies suggest the relative importance and merit in the document. Touch and other file-level commands can only mask the overt date and timestamp of a file, not its purpose or content, true timeliness, or merit. This information is spread through the metadata and leaf content. In some hierarchical structures, this information is stored in tables or other structures apart from the immediate document root. When a document is a relational data structure, as in Access or SQL, hidden system fields and hidden security data define edit and deletion times. It is also important to recognize

that in databases, records which are deleted by the user are only marked as deleted but persist until the database is purged, packed, cleaned, compressed, or otherwise processed in a maintenance mode. When relational technology with transactional logs and rollback facilities are enabled, data can be recreated or dated despite many types or natural of instigated disasters. This supplemental metadata defines date and timeliness too.

[2717] Security: Security of content can be compared to erecting a barrier around that content and may be viewed as an important SC sensitivity issue. However, when content becomes a collection of simple data elements along with data objects, dispersed and distributed sources, effected by embedded events and triggered methods, a search which ignores the security content characteristic fails to acknowledge that the SC sensitivity may be impacted by such omission.

[2718] While content is king in most search systems, it is not the only critical aspect of a source document. In terms of the knowledge expander processing of an Office document, each leaf must be processed and assessed for its metadata. Note again that each leaf may be another object-oriented structure in its own right or a simple element. It will need to be processed and assessed accordingly.

C.20 Select Content Introns and Exons

[2719] Terminology employed in connection with the operation DNA (deoxyribonucleic acid) provides an appropriate metaphor for the MS Office document object model or any other DOM model. While the DOM is separate from an MS Office binary file, it defines the purpose of that file and maps its activation. The DOM "genes" are expressed into the file binaries only as specifically referenced, and frequently diverge from the pure MS Office application as genes from other OLE (object linking and embedding) applications are embedded into the document. The DOM and the expressed document can mutate for better or worse, and both backwards and forwards the document is adaptable just like DNA,

with unforeseen consequences including the profound security flaws evident within the MS Office workflow.

[2720] In genetics, an intron is any non-coding or non-activating sequence of DNA initially copied into RNA but cut from the final RNA transcript or unknown as to singular or recombinant purposes. Introns are excluded or ignored in the DNA process. An exon is a coding or activating sequence with a known purpose that is actually used or one that is unknown as to purpose but nonetheless still used. DNA is, of course, the blueprint for life. RNA is the functional transcript of the DNA blueprint used for cell division and replication. Exons are the useful portions in the DNA cycle.

[2721] In the object model, the DOM is metaphorically the DNA blueprint for an MS Office document whereas the actual Word, Excel, or Outlook message is an expression of the RNA as a functional transcript. Correspondingly, the SC select content or sec-con intron is any document branch, leaf, or node element with a non-coding, non-activated, or even unknown control utility for the document. From a select content - search (or sec-con - search) standpoint, each and every intron represents a non-qualified element that is a potential search term that may lead to an expanded knowledge expander supple doc compilation. Rather than ignoring SC or sec-con introns as most common search engines do, the current knowledge expander search engine employs the SC or sec-con introns to expand the scope of the search. A SC or sec-con exon is any document branch, leaf, or node element serving an end purpose. Each SC or sec-con exon in a MS Office document becomes a certifiable data element.

[2722] Unless each such SC or sec-con intron and SC or sec-con exon in the source document DOM is vetted for credentials, which vetting includes a prioritized taxonomic label or tag, the

knowledge expander search, whether through conventional search engines or the novel search engines described herein, will be incomplete. This is effective for DOM, HTML. XML, databases, and any structured file binaries. The standard 2-phrase process transforms into a 3-phase process where DOM node elements are coded either as exons or introns, and thereafter processed accordingly for inclusion or exclusion.

[2723] The improved accuracy of 3-phase scanning of documents within the context of an object model is apparent. While the traditional 2-phase method find SC or sec-con introns and SC or sec-con extrons within a source document, it also miscodes several other sequences as introns. The accuracy of such process will always include statistically measurable false negatives and positives, thereby missing true threats and removing non-threats. The 3-phase process improved on the 2-phase process with granular deconstruction of the document and subsequent recoding of both false positives and false negatives to yield a higher rate of accuracy. SC or sec-con introns are added to the search term list in the knowledge expander search engine and are scalable numerically. Better DOM maps mean better intron handling.

[2724] FIG. C-11 shows the General DOM Editor program 1022 in a flow chart form. Step 1024 is the initialization that is employed by the security entity or SC or sec-con sensitivity supervisor to set up the program. Step 1026 obtains the DOM layout or blueprint from the DOM vendor. Step 1028 notes that for each root, branch and leaf in the DOM, the information attributes must be prioritized with the SC select content or sec-con and organizational informational attributes (a taxonomic routine with prioritization). SC or sec-con introns are identified and SC or sec-con exons are classified, placed in a hierarchical structure and labeled. This step, of course, uses a content filter to identify the SC or sec-con exons. A contextual filter or algorithm is used to taxonomically classify SC or sec-con exons at a higher

priority level. For example, when "Bin" is next to "Laden" the SC or sec-con exon combination -- Bin Laden-- is classified TS or the top level for the SC or sec-con H-tax class.

[2725] Step 1030 obtains the source document DOM. Step 1032 maps the DOM source document as a binary file populated with content data and meta data as a subset of the DOM blueprint. Step 1034 identifies SC or sec-con exons and SC or sec-con Introns. Step 1036 labels or tags SC or sec-con exons per the tax class (taxonomic classification) and per priority, to be included in the further processing of the item and identifies SC or sec-con introns to be used as search terms in the knowledge expander search. Multiple tags and overlapping tags are possible and a ruled-based system is employed to upcode multiple tags and overlapping tags to the next higher tax class - priority level. Step 1036 converts SC or sec-con exons at each root, branch and leaf into SC or sec-con level save doc, form and format (for example, a safe DOM template), and generates a save blueprint or map, drops (or alternatively stores in a remainder store) all low priority SC or sec-con exons, and expands the search term list with the SC or sec-con introns and the top group of the SC exons. Step 1038 processes the save DOM doc, that is, the top level SC or sec-con exons and all SC or sec-con introns as search terms through public and/or private databases, indices, search engines, etc. A divergent search uses SC or sec-con intron. A convergent search uses SC or sec-con exon. Supplemental documents are gathered form the search, the filter H- tax class priority tag process is repeated on the supple docs, the SC or sec-con introns are identified and selected top priority SC or sec-con exons are identified and the search and a second tier supple docs compilation is processed. Supple docs are stored and a relational map is generated both with the search terms and the supple docs.

[2726] FIG. C-12A shows a flow chart of a basic application for a DOM Editor -knowledge expander search for MS Office document. Step 1042 initializes the system. Step 1044 obtains the

source DOM document and notes that all processing occurs after a spam and virus filter. Step 1046 notes that the program creates or is provided with a security safe or clean DOM document and map. All levels of SC are labeled or tagged and saved in the "save doc" routine. This hierarchical taxonomic (H-tax) tagging labels all SC or sec-con words, terms, etc. and locates and maps exons and introns in the source doc. Step 1048 notes that a template is opened and SC or sec-con exons and SC or sec-con introns (non-standard SC or sec-con and unknown terms, characters, words, etc.) are copied from the source doc into the clean DOC. A clean .DOC template (Word) or whatever the new document type is opened for the specific application. The Normal.DOC or New Spreadsheet.XLS on the MS Office distribution CD is safe. In Visio, for example, start with a new chart. Older documents saved as templates tend to accumulate unanticipated metadata. If one must use a non-standard template, clean it up. View it as both a printed document, as a binary file, and as a document object. Write protect the templates, or store the templates as non-modifiable networked volume. If the templates are used throughout the organization, create a network store for them and configure each user's installation of MS Office to look there for the templates. Search the network for duplicate copies of these templates and eliminate them.

[2727] If changes from any version of MS Office to another version are made — this can be a regularly upgrade or a even a downgrade—create new documents and cut-and-paste parts of prior documents into new ones. Lose the older files and templates. If you receive or open an Office document in either an older or newer version, create new documents and cut-and-paste parts of prior documents into new ones consistent with the MS Office version that you use.

[2728] Step 1050 disables edit controls and step 1052 copies SC or sec-con exon and SC or sec-con intron content. The point is one must edit cleanly. This is not a single step but rather a process, both

one time and ongoing. Disable versioning in step 1050 to prevent a buildup of past versions of the document. With versioning, prior sessions will be stored as document.doc 1, document.doc 2, and so on. These tend to get entwined with the latest version. If workflow with InfoPath, Outlook, or other collaborative workflow tools creates duplicate copies of the source document file for each user. Step 1050 includes the concept that the system is configured to store a single network copy instead. Preserve an audit trail and backup with a system backup rather than versioning. Disable change tracking in step 1050 to curtail the buildup of additions, deletions, and changes that transcend the publishing intent of the document. If redlining is necessary, establish guidelines for periodically accepting changes to purge change log. Use the command to save the document without the change log with File/Save or File/Save As. Do not use nor rely on fast saves, timed saves, or file recovery after a MS Office crash to purge the dirty metadata. After a crash, reopen the document, save the document under a new name. Close the Office application. Delete the old file precipitating the crash. Rename the new file under the old name. Reopen the Office application.

[2729] Step 1054 locates text in footnotes, hidden text, etc and labels or tags that content as SC or sec-con exons or SC or sec-con introns and maps those data elements to the SC or sec-con save doc and copies the elements into the Clean DOC. Use comments instead of hidden text. It is documented as a feature so it can be found rather than accidentally uncovered. Hidden text with a font color change or font size change looks like an artifact that most users will ignore or delete. Avoid the use of headers, footers, footnotes, endnotes, inserts for table of contents, index and the like. These appear only in the printed output unless specifically viewed from the View pulldown menu. Such links also create a lot of dirty metadata beyond what is visible even during editing that persists until changes are accepted. Remove references from the source document. This is subtle, but very important when documents are

specifically posted or even inadvertently listed on web sites. References include other files, documents, hyperlinks, and other possible embedded formatted materials. These references create the ability to infer quite a lot about the purpose of the document from other related documents in the same directory, by the same authors, and the types of other documents. For example, a spreadsheet stored with a report that is not overtly included in the report suggests that is source material that has not been reviewed with an eye towards privacy, security, or client privilege.

[2730] Paste and copy images, cut text, formatted text, pie charts, record sets, slides, waterfalls, milestones, organizational charts as plain text or an image rather than formatted Office objects. If the embed commend is used, all the metadata baggage from the other Office application is now added to the metadata in the target document. Since that metadata baggage is not native to the target document application, it is inaccessible and truly hidden. Tools, such as Metadata Assistant will not find Excel metadata within a Word Document, Word metadata within an Excel spreadsheet, and none of them within an Outlook note or message.

[2731] Step 1056 notes that a clean map for the meta data cleared DOM document must be created.

[2732] Step 1058 executes the intron search and /or granular search routine, gathers and compiles supple. docs and the processes those supple docs through the granular filters discussed earlier to locate, with the hierarchical taxonomic filter process tags new words, terms, symbols, which are related to the original SC select content or sec-con, thereby expanding the knowledge base of the meaning of the SC or sec-con source document.

C.21 Hierarchical Taxonomic Class Examples

[2733] Various type of classification systems (taxonomic systems) may be employed. For

example, a dictionary classifies all words as nouns, verbs, adverbs, etc. This is one taxonomic system. A prioritized H-tax system classifies nouns into a name subclass and this name subclass may be priority or sensitivity level 1. Geographic locations (class nouns) may be priority 2. The operator of the present knowledge expander search process may place a high priority or sensitivity on "noun" class, thereby converting the simple taxonomic structure into a hierarchical taxonomic system. Identifying categories and subcategories for SC or sec-con sensitive words, etc. or critical data, creating a compilation of pre-existing data, comparing the compiled pre-existing data to the target data and labeling or tagging the SC or sec-con terms is one of many important aspects of the present invention. Table 1 in FIG. C-12B provides categorical identifiers for personal identity. These categories, which may be utilized to identify a person, are not meant to be all encompassing but are mainly provided as examples.

[2734] After categories are identified for the critical data, it is important to create the hierarchical taxonomic system against which the target data is tested. Ranking or qualifying the categories at SC or sec-con sensitivity levels is next. Table 2 in FIG. C-12C is a general attempt to quantify the categories for personal identification from a high risk value beginning with "name" to a low risk value ending with "personal interests". Again, the Ranked Identity Category Table 2 is not meant to be limiting but is meant to be an example. Individuals skilled in identifying a person may alter the ranking of the identity categories in Table 2, FIG. C-12C.

[2735] The present invention can be employed to use a hierarchical taxonomic system for a business. Table 3 in FIG. C-12D set forth below provides examples of categorical identifiers for a manufacturing business. Again, this list is not meant to be exhaustive or complete, but is only provided as an example of the types of categories and subcategories which a manufacturing business would employ in order to establish the risk monitor of the present invention.

[2736] With this list, the manufacturing business may assign a SC or sec-con sensitivity level to each category (class, subclass and sub-subclass) thereby creating a prioritized hierarchical taxonomic system.

C.22 Knowledge Expander (KE) Basic Program

[2737] As discussed earlier, the information life cycle or knowledge expander search module can be configured with many different modules in order to achieve the objective of the operator. The knowledge expander or KE basic program 3001 in FIG. C-13 is one basic implementation of the expander search program. FIG. C-13, In step 3002, the operator or user sets various filters and sets the search targets. The search targets may be publicly available search engines or private data bases, data collections, indices or any item that the user wants the KE search to access. Filters are described earlier as hierarchical taxonomic filters or taggers, content, contextual and other types of filters. Filters are used to identify class, subclass and priority in the hierarchical taxonomic or H-tax system, to apply contextual rules or algorithms ("bomb" within 10 words of "arab"), and content (select content SC or sec-con verses non-matching SC or sec-con). Step 3004 obtains the source data or the source is input by the operator. Data input may be from a client computer in a server-client computer system. The source data may be a data stream, a source document or other item. Step 3004 contemplates that if a document is utilized, the document is deconstructed to its basic data elements and appropriately mapped. See the discussion above in connection with DOM branch, leaf and tree deconstruction. In step 3006, the hierarchical taxonomic classification (H-tax) occurs and each data element in the input document is tagged based upon the classification system which has been prioritized by the user (or the user uses a pre-set prioritized H-tax). A note to deconstruct the source document into data elements (see DOM deconstruction into its object hierarchical structure above) is found in step 3006. In step 3008, for non-matching data elements, a search is conducted through the search targets. Non-matching data elements

form the input doc are those which do not match SC or sec-con. In step 3010, the system gathers the documents from the search (supplemental documents) and repeats the H-tax tagging operation for b number of supple docs. A relationship mapping function and display function is activated if c number of hits occur within the H-tax top priority range. In other words, if the search for non-matching data elements returns 50 select content or SC or sec-con terms and of those 50 SC or sec-con terms in the supple docs, 20 SC or sec-con terms fall within priority ranges 1-5 of the hierarchical taxonomic classification set by the user (n priority H-tax levels), then the system and process maps the results and displays the representations of the H-tax recovered from the supple docs c is less than 20. If less than b number of H-tax tags are noted in the supplemental documents, the system repeats steps 3006 and 3008 and gathers a second tier of supplemental documents.

[2738] Steps 3012, 3014, 3016 are optional. As an option to be set by the user or as an option to be set by the system operator as a default, step 3012 changes the H-tax top range bandwidth from n priorities to be searched to m priorities. This change may be needed to expand the search bandwidth from b to b-10 to b+20. A reduction (m less than n) is needed if too many supple docs are recovered by the search. If too few are recovered, then m is made greater than n priorities in the H-tax. Option 3014 changes the output supplemental document bandwidth to b + y if too few supple docs are recovered. Of course, the output supplemental bandwidth could be reduced with b - y. Optional step 3016 mixes n number of select content search terms with the no-match data element before the search. This effectively is a convergent filter. By adding the known SC or sec-con to the search terms (initially, search terms are generated from the input doc based upon priority H-tax), the search is biased or is set to converge on the known SC or sec-con added to the search term group. By using the no-match (NOT AND) function, the KE search diverges to gather a wider range of docs. Of course, the user may always employ

common search engines in addition to the search plans set forth herein. Step 3018 repeats the process until b supplement documents or c hits (SC or sec-con sensitivity matches) are noted within the H-tax top range. Step 3020 maps, displays, stores and ends the program.

[2739] FIG. C-14 shows an expansive or expansion unknown search engine 3031. Step 3032 initializes the program with preferably user input, sets up the hierarchical taxonomic H-tax classification system along with a priority or sensitivity level assigned to each level and branch and leaf in the H-tax system. Select content SC or sec-con filter modules for content and contextual range or algorithm tests are also set. At least some select content should fall within the prioritized hierarchical taxonomic system. The select content SC or sec-con encompasses words, terms and data objects that the user wants to expand his or her knowledge base with the search engine. The user sets a no-match search bandwidth nms and sets search targets such as search engines, public or private databases, data compilations, indices, data collections and data structures or whatever. Step 3034 obtains a source or data stream or input document. Step 3036 processes each data element (which may include a document deconstruction) with the H-tax, content and context filters, and tags each data element with an H-tax marker. Multiple, singular, and a default "unknown" H-tax classification indicators are linked or tagged on every data element. The system in step 3036 may upgrade or downgrade the priority or sensitivity H-tax level for a particular word, term, etc., due to contextual rules in the context filter. In step 3038, the system conducts a search for nms number of not matching or unknown data elements. If nms is 10, the system takes the top 10 priority H-tax supple SC or sec-con terms obtained from the input doc and uses the top 10 supple SC or sec-con as search terms in the target data collections. The search is conducted through search targets identified in the set-up step 3032. The no-match search 3038 is sometimes identified as step A herein. Step 3040 gathers supplemental documents. Sometimes, step 3040 is step B herein. In

step 3042, the system processes the supplemental documents through the H-tax tagger. The supple docs are classified with the H-tax. This is sometimes step C. In step 3044, the system determines if p hits have been noted in the H-tax top range levels. If YES, the system stops and maps the select content SC or sec-con sensitivity sel. levels to the H-tax mapping structure and displays that SC or sec-con hierarchical representation map to the user. Further, the system in step 3044 maps the supplemental documents to the H-tax map structure and displays that to the user. Sometimes, step 3044 is step D herein. In step 3046, if p number of hits are NOT identified in step 3044, steps A,B,C, D are again executed and repeated unless the repeat cycle is more than R. The nms bandwidth may automatically change (increase) or the n priority H-tax may change to m H-tax levels to increase the supple docs or the hit count for supple SC or sec-con . The same techniques may be used in reverse if too many supple docs are recovered. The system ends after step 3046.

[2740] FIG. C-15 shows knowledge expander KE search options 3051. Step 3052 is an option to add a noise filter after the H-tax tagger step 3036 in the expansive unknown search engine routine 3031. The noise filter reduces the number of unknown data element and operates on the unknown tags for those unknown data elements. For example, terms "and" and "or" and "the" may be deleted by the noise filter. Optional step 3054 adds to step D, if p hits is equal to zero, then the system resets the no-match bandwidth nms + u and repeats step A,B,C,D and repeats the entire process until p hits in step D or R iterative cycles have been noted (counted). In optional step 3056, a convergence filter is added to the expansive, unknown search engines 3031 by adding w select content SC or sec-con search terms to the no-match search step A. In optional step 3058, a divergence filter is added by changing, either increasing or decreasing, nms value if p hits do not exceed p + x. In optional step 3059, a content filter is added to detect temporal relationships such as time and date ranges. Time may be time of length of

search or time may be a time-date range limit on recovered supple docs. The temporal contextual filter (time and date is a matter of context) would be added to step 3036 the H-tax step. In optional step 3060, the H-tax map is output to the user so the user sees the classification and the prioritization of that classification. In step 3062 the operator can change the repeat cycle R, the time the search runs, the number of cycles, and the search target databases, data sets, spreadsheets or public or private search engines. In optional step 3064, the operator can limit the supplemental documents based upon a temporal factor such as time, the total number of SC or sec-con terms located in the supplemental documents, and the total number of supplemental documents.

C.23 Secured Metasearch Engine Based on Granular Data Control

[2741] As an overview, the secured metasearch engine technology is based on granular data control. The engine is part of the KE engine described above in sections C.9 - C.22. The secured metasearch engine technology enables secured searches and distribution of secured search results including secured information sharing with multi level security (MLS) search capabilities across domains. Data is explored not only by content but also context and concept. Search should occur across domains, with information shared and delivered across domains and through the extraction and integration of both private and public data sources including the Internet. Stovepipes and silos, independent efforts at information collection and intelligence analysis, and air gaps or security guards divide information domains and prevent secured information sharing. Centralization and consolidation of data gives rise to data overload, a problem in part addressed by new web-like searching techniques, but also opens the door to security leaks.

[2742] The security of current data mining and search operations as well as the security of the multi level distribution of the search results to users with different access levels is questionable.

Furthermore, the current data mining and search is rigid because the starting point for each and every effort requires an index or search terms and the mining is completely content-based. When connecting the dots, it is essential that search be possible without preconceptions of utility, purpose, or cultural bias. Effective search techniques, called "metasearch" herein, comprises more than web-like textual (and other formats such as audio or video, but also compound meta-formats) search technology. It includes data mining, textual and contextual search, aggregation, inference, and data-to-data interaction all based on granular data control subject to security compliance.

[2743] Metasearch generally means the usage of multiple web-based search engines based on keywords with results combined into a single unified report; this is better described as a confederated search or federated search. Each web-based search engine is rigid in terms of its reliance on preexisting indexes; it is usually fast but limited in terms of most likely usage bias. However, the combined results reinforce the preexisting answers based on the implicit search question.

[2744] In contrast, the Secured Metasearch Engine (sections C.9 - C.22) is designed for searching the unknowns and expands the base of information, as well as conducts the search operations across different domains with different security levels. The Secured Metasearch Engine results provides two distinct and fundamental directions, search convergence and search divergence, both of which are critical to the process of connecting-the-dots. Failure to consider both convergence and divergence leads to incomplete results.

[2745] Search divergence is the process of expanding the content, context, and concept beyond the strict scope of the fundamental search. Search convergence is the process of validating, substantiating, and establishing the integrity of the fundamental search. Divergence is the expansion of the search query or data mining operation. There is a need to find outliers useful to expand the scope

of the users knowledge and infer new information and intelligence. Some information and intelligence operations require exploration without preconception. Herein, outliers with the least priority, relevance, or ranking frequently provide previously unknown references or interconnections that are purposefully hidden. This is a search divergence because the base search query primes the process but effectively we already know the highest ranking results and reject those results while iterating the search for non-ranking items.

[2746] The Security Metasearch Engine starts with known keywords and indexes, but accepts primarily information that is new or different from the direct response to the implicit question, that is the base search query. Validation and integrity is not important to divergence. The search engine is looking for all related metadata, references, links, and relationships without regard to accuracy. Exact matches are in effect usually ignored, but all new references are added into a secondary query yielding subsequent results (supple docs and supple search terms). The subsequent results in turn yield yet more new references. This subsequent query yields subsequent results, which is repeated indefinitely. The iterative results create an asymptotic adjacency list model, which may include a networking relatedness. The networking relatedness is often viewed as flow charts showing betweenness, closeness, and connectedness. The networking relatedness may also be a social network relatedness.

[2747] This expansion of the base search query is the process by which the user discovers more dots. Connecting dots the user already knows is pointless and does not expand the user's knowledge; the user wants to expand the set of dots and generate new content, new context, and new concepts. The Secured Metasearch Engine starts with old and/or limited information, but generates new information. The theory of six degrees of separation has some legitimacy, however there is a difference between accidental and active relatedness.

[2748] Divergence is also critical to enabling the Secured Metasearch Engine to foster security by exploring the purpose and usage of information beyond the immediate and obvious preconceptions (i.e. foresee reflexivity) of semantic and pragmatic meaning. See beyond the immediate security perception for other possible cross-domain requirements and enable information sharing. The Secure Metasearch Engines is connecting known dots with new dots, the true concept of "connecting the dots". The new dots lack integrity and substantiation. They are not ignored or dismissed outliers because they can reveal the outlines of previously unknown, undiscovered, unforeseen, uninvestigated and unrevealed situations.

[2749] The search process is really a metasearch, applying techniques of data mining, textual and contextual search, aggregation, inference, and data-to-data interaction, and the integration is the establishment of relationships. However, some of the new dots will be false and noise. At this point, a search convergence is useful to establish likelihood and risk for the new information and integrity and substantiation of the difference between accidental and active relatedness.

[2750] The Convergence Search is described herein. Convergence is the usual web-like search engine or data mining operation. It is rigid. The user starts with known keywords and indexes—search terms—as the base search query and accept hits with the highest relevancy. This information is used aggregately to confirm or deny the integrity of the initial search. The user gets back what the user expects with priority, relevance, or ranking focusing the hits to the content most consistent to the search term context. Outliers are naturally rejected and filtered out. Search convergence might entail a physical inventory, database report verification, an independent party confirmation, or a drone flyover to substantiate the information. The result is in direct response to the implicit question, that is the base search query. The user positions and substantiates the dots the user already knows. The user starts with

old information, and the user gets back old information. However, the benefit provides by the Secured Metasearch Engine is that it contextualizes these convergent results with a higher level of relevancy to the user's specific purposes based on tools which manage and control granular data. Those include tools include: (1) metafilters for automatic categorization; and (2) sensitivity level taxonomy. These tools deliver " fine adjustments " to the search process and find the most appropriate results for the user prioritized according to the highest relevancy.

[2751] The combination of divergence and convergence operations provides additional information. The Secure Metasearch Engines performs three key functions. (1) First, it selects new data elements of value and recognizes unknowns for automated iterative divergent search. (2) Second, the combined divergence search delivers "fine adjustments" to the divergence search process and finds the most appropriate results for the user prioritized according to the highest relevancy. (3) The Secure Meta search Engine categorizes all aggregated results to sensitivity level and enables distribution subject to MLS-compliance cross domain. While the Secure Metasearch Engines primarily focus on unstructured information, it can be integrated with structured data sources too because it is actually easier to extract context and concept from structured data.

[2752] There are multiple starting points to the metasearch process. One or more candidate data streams, news feeds, or documents can be dragged from the desktop into a queue. These items may be monitored for changes and then fed into the Secure Metasearch Engine for granular data selection and metatag categorization with the output result that search terms are generated and metasearch terms are also generated. A document or data stream automatically sets the context of the search and is explored for likely search terms/ key words. No index or bias is necessary to begin the process. The resulting search terms of interest are fed into the next search cycle, which may include also one or more URLs for

further exploration.Indexes and search terms do not define the implicit question or establish parameters for an effective answer because they represent a single dimension of information. Metafilters are used to control the granular data elements in order to fine-tune the search. The metafilters add both context and concepts to the initial search and all subsequent exploration. Such metafilters may be: (1) Categories of the granular data elements in a document ("name" "location" "date" "amount"); (2) sensitivity levels of the data elements ("top Secret" "Private" "level 9"); (3) phrases, ("Mohammed Atta" "United Nations"); and (4) Group Lists, (Groups of words which represent a concept or a context –"Nuclear" + "Submarine").

[2753] Generation of keywords / search terms for conducting the continuing searches is done by controlling of data at the granular data level. Control of data at the granular data level is enabled by combining different filters. The Secure Metasearch Engine recognizes complex content and context by implementing a combination of filters / sieves. Those combinations of different filters/ sieves create different permutations of resulting keywords for the search. The different keywords for the continuing searches are selected automatically based on: (1) the words entered into each filter/ metafilter; and (2) the configuration of the different filters/ metafilters combinations. As an example in one embodiment, all the content of a document is: (1) categorized, i.e. each individual word (or phrase) in the document is tagged by category, "name" "location" "date" etc.; (2) the individual words are also tagged by sensitivity tags "secret " "confidential" "Private" "level 3 sensitivity" etc.; and (3) all the above is presented in a matrix format. The filters will automatically select which words in the matrix will be used as keywords for the continuing search. Different keywords /search terms will be selected based on the different filters or combination of filters used. The entries for such filters may be machine or user defined. The combination and selection of different filters may create different permutation and results.

The filters may include the "Word Lists", "Phrases list", and "Word Group Lists" which limit the range or area from which the keywords may be selected. The ideal combination of filters to control granular content for automatic production of keywords may be found also by trial and error. Artificial intelligence tools may be integrated.

[2754] Linear adjacent selections may be used as keywords. However, unknown content can slip through this sieve without an explicit human intervention, therefore the Secured Metasearch Engine automates this sieve by creating permutations of both categorized and unknown words and phrases. These permutations are used as new search terms. The likelihood of random combinations (included in the class of the NP! problem space) is so low, that emphasis is focused instead on linear adjacent selections. This result in is a small and focused problem space that can be solved on a typical personal computer and ads disproportionate value. For example, a Google search of "Mohammed Atta" returns "Khalid Sheikh Mohammed" as a disconnected string of categorized and unknown words out of the box; that means parts of the phrase "Khalid Sheikh Mohammed" are recognized without any special configuration, but not as a contextual phrase. Permutations of linear adjacent selections do however create the complete connected phrase and automatically submit the complete phrase for a divergent metasearch, thereby creating new and previously unrecognized relationships to connect the-dots.

[2755] The engine can be set for automated contextual pruning with flexibility to discover codes. Many divergent mining and search efforts will expose unknown, deceitfully hidden, and statistically irrelevant relationships. For example, a Google search of "Mohammed Atta" also shows meetings at Aleppo. The relevant thread for Aleppo is that it is Syrian university town with social connections to Hamburg and the 9/11 hijackers associates. However, Aleppo is also a common Lebanese street name, and the original shoemaker character in the Pinocchio story by C. Collodi. The name was changed to

Geppetto in the Disney children's movie. While it is tempting just to prune this track, it is also important to explore associative contexts, because the use of codes can mask more sinister meanings. The 9/11 hijackers communicated within the seemingly innocuous context of weddings, guest lists, and dates. So while contextual pruning is easier after the fact, it can be very important before the fact when the use of codes and false contexts are not so clear.

[2756] In conclusion, although stovepipes and silos, independent efforts at information collection and intelligence analysis, and air gaps or security guards among domains prevent information sharing, the Secure Metasearch Engine enables security with information sharing including Multi Level Security (MLS) search capabilities across domains. It overcomes security leaks consistent with the needs to both protect national security interests but also enable cross-domain information sharing. The Secured Metasearch Engine technology is based on granular data control. It enables secured searches and distribution of secured search results including secured information sharing with Multi Level Security (MLS) search capabilities across domains. Data is explored not only by content but also context and concept. Search should occur across domains, with information shared and delivered across domains and through the extraction and integration of both private and public data sources including the Internet.

D.1 Secure Editor, DOM Editor (General) and DOM Editor -

Basic Applied as Extractor Modules

[2757] The Secure Editor can be used to generate SC or sec-con data for a search and can also be used to locate and extract security or sec-con data and SC data before storage of the sec-con or SC data in the distributed storage sites. The Secure Editor in FIG. D-20 diagrammatically illustrates an editor which may be employed to secure sensitive word/objects in a source document. In a current working embodiment, the secure editor is a standalone application or a module to add into other applications for plain text and

media creation, editing, and sensitivity level tagging. Other types of tagging, wherein the editor supplements the initial group or subset of security sensitive (or select content) words, characters, icons and data objects by categorization, taxonomy classification, privacy, security, compliance, and semiotic meaning, are also available. The editor supports a full range of document management and can be integrated into a unified infrastructure, from creation, editing, document markup, tagging, tag conversion, tag removal, context sensitivity level redaction, context reconstitution, and support for complex process work flows. The architecture assures separation of data from metadata so that no security lapses are introduced into the traditional word processing and document management cycle.

[2758] FIG. D-20 is a basic flow chart for one embodiment of the Secure Editor module. Editor program 972a begins with obtaining the source document 974a. Of course, the source document may be any type of document as explained later herein. Step or function 976a obtains one or more filters for one or more security or sensitivity levels. Step 978 screens or processed the source document with the filter(s). For example, the source document in FIG. B-8A in window 991 has text regions 993, 994, 995 and 996. In step 979, the Secure Editor displays, in situ (in the displayed document), the filtered identified material and conforms the precursor document to the security level protocols for the system within which the Secure Editor is employed as an information processing tool. FIG. B-8B shows that the address data 993 is marked TS (top secret), region 994 is displayed in color A for TS coding (please note that the addressee data may also be so marked) and is "red-lined" or struck out. Region 995 is displayed as presented in the source document and is labeled U (unclassified) and region 996 is shown in color B, is redlined and is labeled S. Labels TS, S, C (classified) and U are the established security labeling protocol used by the organization employing the Secure Editor. Other labeling schemes may be employed. Color is used to assist the user to select (and in some non-standard cases, deselect) the

sensitive data marked by the editor. Redline is used to inform the user that the filter(s) will extract the marked data. Labels are used to permit the entity using the editor to employ standard tear line protocol. Any data beneath a security classification of the user is under the tear line and the data is permitted to be distributed to the lower security cleared user. Of course, electronic distribution of secure data need not use the hard copy or print version of the tear line. However, this nomenclature referring to the tear line is used in the prior art systems.

[2759] Step 980a (FIG. D-20) accepts the user's manual changes (typically upgrades) to the precursor document. These manual changes are displayed, redlined, colored and labeled. Step 982a inserts the security label TS, S, C and U has discussed above. Step 984a notes that the system takes certain meta data such as author, date-time, version history, change history, etc. and converts this meta data into ordinary text, marks that data at the necessary security level and labels the data. Step 986a permits the user to add (or omit) placeholders into the final document. FIG. B-8C shows placeholders as black lines or as XXXXX symbols (or other symbols) wherein the sensitive text is not shown but some replacement markers are shown. The byline in region 1003 shows "sanitized document." The byline 1003 in FIG. B-8B lists the security level and the color representation.

[2760] Step 988a (FIG. D-20) activates the filter, extracts the sensitive or SC data and temporarily stores the extracted data. Step 990a displays the filtered document and the user may view the filtered document at each security level. Therefore, the user, before transmitting a secured email (or letter) may look at the source (FIG. B-8A), may look at the TS level without the redline strike out but with security labels and colors, may look at the T level revealing regions 996 and 994 but not regions 993 and 994 (which are TS coded regions), and look at U versions as shown in FIG. B-8C. Step 992a disperses the extracted data and the remainder data or disperses partial versions of the document (those partial versions

formatted and containing only data at or above the target security level (all TS level data (which includes TS, S, C and U data), or all S data (comprising S, C and U) or all C data and U data)).

[2761] One feature of the present invention is that in step 979a, the security level protocol determines whether single words are granularly classified (TS, S, etc.) or whether a line is classified, or whether an entire paragraph is classified (see FIG. B-8B). If a commercial/privacy filter is used to exclude all social security numbers, the organizational protocol is set at a granular level to exclude just social security numbers. Different group protocols use algorithms to mark, filter and extract adjunctive security sensitive (or select content) words, characters, icons and data objects near the target security sensitive words, characters, icons and data objects. The sensitive words may be security sensitive (or select content) words, characters or data objects defined by compliance with law, regulation or policy, privacy, national, organizational or private security concerns. For example, Bin Laden is the target sensitive word in FIG. B-8B and this classifies the entire paragraph as TS level. The other words in the paragraph are adjunctive word/objects.

[2762] Several steps are prudent to enable MS Office document protection. The first step is have a network guard (see FIG. D-16) that filters all incoming and outgoing traffic for MS Office document files and quarantines them. Spam and virus filtering is necessary to preclude system, resource, and file exploits. URL filtering, quarantine lists, black lists, white lists represent the minimum responsible approach. Ingress files can harbor viruses, etc. Outgress files can harbor privileged information at any and all levels of the DOM. With a means to filter and check every node for purpose, content, metadata, formats, structure, comments, links, and so on, there is no other way to vet the integrity of the file.

[2763] FIG. D-24 shows the General DOM Editor program 1022a in a flow chart form. Step 1024a is the initialization that is employed by the security entity to set up the program. Step 1026a

obtains the DOM layout or blueprint from the DOM vendor. Step 1028a notes that for each root, branch and leaf in the DOM, the information attributes must be prioritized with th the security organizational informational attributes. For example, the audio files of potential Bn Laden voice prints may be critical at TS and S levels but be completely excluded at C and U levels. Therefore, any audio files below C and U are security introns which are excluded or ignored.

[2764] Step 1030a obtains the source document DOM. Step 1032a maps the DOM source document as a binary file populated with content data and meta data as a subset of the DM blueprint. Step 1034a identifies security exons to be included in the further processing of the item and identifies security introns to be ignored or excluded in the processing. Step 1036a converts security exons at each root, branch and leaf int security safe form and format (for example, a safe DOM template), and generates a safe blueprint or map for the precursor DOM and discards all security introns. Step 1038a processes the content of the safe DOM with security filters discussed above. It also processes all data objects (exons) from the safe DOM with granular filters. Data is dispersed as necessary and retrieved and re-assembled as necessary with a safe map.

[2765] FIG. D-25 shows a flow chart of a basic application for a DOM Editor for MS Office document. Step 1042a initializes the system. Step 1044a obtains the source DOM document and notes that all processing occurs after a spam and virus filter. Step 1046a notes that the program creates or is provided with a security safe DOM document and map. Step 1048a notes that a template is opened. A clean .DOT template (Word) or whatever the new document type is opened for the specific application. The Normal.DOT or New Spreadsheet.XLS on the MS Office distribution CD is safe. In Visio, for example, start with a new chart. Older documents saved as templates tend to accumulate unanticipated metadata. If one must use a non-standard template, clean it up. View it as both a printed document, as

a binary file, and as a document object. Write protect the templates, or store the templates as non-modifiable networked volume. If the templates are used throughout the organization, create a network store for them and configure each user's installation of MS Office to look there for the templates. Search the network for duplicate copies of these templates and eliminate them.

[2766] Step 1050a disables edit controls and step 1053a copies content. The point is one must edit cleanly. This is not a single step but rather a process, both one time and ongoing. Disable versioning in step 1050a to prevent a buildup of past versions of the document. With versioning, prior sessions will be stored as document.doc 1, document.doc 2, and so on. These tend to get entwined with the latest version. If workflow with InfoPath, Outlook, or other collaborative workflow tools creates duplicate copies of the source document file for each user. Step 1050a includes the concept that the system is configured to store a single network copy instead. Preserve an audit trail and backup with a system backup rather than versioning. Disable change tracking in step 1050a to curtail the buildup of additions, deletions, and changes that transcend the publishing intent of the document. If redlining is necessary, establish guidelines for periodically accepting changes to purge change log. Use the command to save the document without the change log with File/Save or File/Save As. Do not use nor rely on fast saves, timed saves, or file recovery after a MS Office crash to purge the dirty metadata. After a crash, reopen the document, save the document under a new name. Close the Office application. Delete the old file precipitating the crash. Rename the new file under the old name. Reopen the Office application.

[2767] Step 1054a locates text in footnotes, hidden text, etc and either discards or ignores the subtext because those items are considered security introns or copies them into a safe DOM text form and disables functions, if necessary. Step 1056a notes that a clean map for the security cleared DOM document must be created. Step 1058a executes the editor or granular filter and dispersal routine as

necessary. The distribution or push of partial security cleared versions of the text can be included in step 1058a.

D.2 Applications of the Secure and SC Data Systems

[2768] The secure data storage systems in Part B, combined with the search engine or KE engine in Part C can be deployed in various applications, some of which are described in this Part D.

D.3 Automatic Features

[2769] The system and method described herein may operate substantially automatically, that is, without operator intervention, other than the security clearance function. The clearance function does require some type of operator authentication prior to retrieval of the extracted and remainder data.

[2770] The system and the method may operate automatically in that the plaintext or originating data could be identified by a party desiring security. The system could obtain that data from any data input device (hard drive memory, floppy drive memory, flash card memory, personal data assistant (PDA), or any other type of data input device), filter the data, separate the extracted text or the remainder text, encrypt (or not encrypt) the data, separately store the extract and remainder data (all automatically, that is, without operator intervention). Hence, it is not necessary that the system operate with significant operator or manual intervention. Of course, the system may also operate on a plaintext document or data object that is being created "in real time" by an operator and keyboard, mouse or other type of data input device.

[2771] The automatic operation of the system and the method can be caused by a triggering event. This triggering event may be a security attack (generating a trigger to start the gathering of plaintext, filtering, extraction and storing) or may be any other type of trigger such as a building burglar alarm, door alarm, fire alarm, or virus detection algorithm trigger. The event may be a time of day, week

or month. It may be n seconds after the user stops typing on a keyboard. It may be a timed back-up feature.

D.4 Multiple Security Levels

[2772] Multiple filters may be utilized in the system and in connection with the method. These multiple filters may be useful in the operation of the system with a plurality of security levels. Each filter could filter out different levels of security sensitive (or select content) items and each bundle or group of security sensitive items (from each distinct filter) could be stored at different computer storage locations. Multiple filters, multiple security levels and multiple storage areas may also include multiple encryption routines and decryption routines. Encryption and decryption routines can be related to the level of security of a particular group of data.

[2773] Multiple maps may also be provided for singular or multiple storage of extracted data and remainder data. These maps may or may not indicate the originating point of the data. Maps can be parsed such that an intruder, upon discovery of a single map or map portion, could not locate the storage locations of all piece of the extracted data and remainder data. Maps may also be encrypted. The map may also be stored at a distinct map store location.

[2774] The concept of partial reconstruction also includes the concept that a portion of the plaintext would be reconstructed and the unreconstructed portions of the plaintext could be encrypted or could show blanks or other symbolic indicators. See the placeholder table above. Partial reconstruction of the plaintext also includes a concept that the security sensitive (or select content) items or materials may be subject to different types of encryption. Hence, a single plaintext document may have multiple levels of security and multiple levels of encryption wherein each encryption has a different level of security assigned to it.

[2775] The present invention can also be configured to provide a computer network which transparently establishes and manages the separation of user-based communities of interest. The separation is accomplished by extraction pursuant to security levels, dispersion of data into secure storage facilities (memory stores) and reconstruction based upon the assigned security level. A low level security clearance results in only partial reconstruction of the plain text or source document. These user-based communities of interest are a plurality of users each having respective security clearances. As described above, each successively higher level of security clearance permits the user to see greater degrees of reconstructed plain text obtained from the extracted data stored in extract stores and the remainder data from the remainder stores. By integrating encryption (and necessarily decryption), separation of user-based communities of interest are established such that the users in a particular community are permitted access to some or all of the plain text data based crypto-graphically separated communities and need to know security levels.

[2776] FIGs. D-1 through D-5 are omitted as shown in the drawings.

[2777] FIG. D-6 is an exemplary computer network diagram showing various user communities. It should be noted that select content SC data stores may be grouped for different portions of the same enterprise. Therefore, rather than have security groups as discussed herein, the system could provide SC interest groups in an enterprise. An "SC interest group" is the same as a security level group. The telecommunications network 402 is connected to the server application server provider ASP 452 and to various networks and personal computers or PCs. The PCs may be computer work stations. Network A 404 is coupled to telecommunications network 402 via an input/output unit 406. Network A is coupled to various PCs identified in FIG. D-6 as PC-4, PC-5 and PC-6. Of course, Network A could be coupled to other PCs not illustrated in FIG. D-6. As described earlier, server 452 can facilitate remote

or offsite storage of extract data and remainder data in store 1, store 2 and/or store 3. Further, the map showing the storage location may be encrypted and stored in any one or more of these stores. Also as described earlier, the memory in one of the PCs, for example PC-4, PC-5 could be utilized to store extract data and remainder data from PC-6 and PC-6 can be configured as the input data computer. Hence, the present system and methodology encompasses the concept of local storage and remote storage. On the local level, the storage begins by storing the extract data at different locations in the hard drive of the PC. The next level higher is storing the extract data in removable computer media such as floppy disk, removable tape drives, CDs etc. associated with the PC accepting data or associated with a server on Network A. The next higher level of extract store is storage of the extract data on a server or other computer in a particular network. If PC-6 is designated as the input computer, the extract data may be stored on PC-4. Of course, PC-4 could be designated as the server for Network A.

[2778] PC-7, PC-8 and PC-9 are coupled to telecommunications network 402. Network C 408 and Network B 410 is coupled to communications network 402. The lines, one of which is line 409 extending from Network C 408, represent a plurality of computers or workstations coupled to Network C. Line 411 represents a plurality of workstations or computers coupled to Network B 410. In an e-mail implementation of one embodiment of the present invention, PC-7, PC-8, etc. may represent computerized devices accepting e-mail (personal data assistant, pager, cell phone, etc.). The sender and the e-mail addressee may utilize simple computerized systems to communicated via e-mail. Further, the network may be any telecommunications network including wire, cable, cellular, wireless, satellite, IR or RF systems.

[2779] FIG. D-7a diagrammatically illustrates a flow chart showing the key component steps for the multiple layer security program for the community of users. The "community of interest" system

described herein enables persons and organizations at the same security level to share data on a peer to peer level. Further the security or SC system may operate automatically, with respect to extraction, storage and reconstruction, such that the peer to peer dissemination of data objects is quickly and readily available to all at the same or higher security levels. Step 420 initializes the program. Step 422 enables the user, administrator or system operator to designate multiple levels of security, that is, multiple words, characters, icon, data objects, or whatever, for each security level and further to define encryption for each security level. The designation step 422 also includes identifying the communities of interest and the particular security level and security clearance or SC level for each community of interest. One example of various security levels for communities is set forth below in the Community Security Level Table which is keyed to the computer network diagram of FIG. D-6.

Community Security Level Table

Security level	Community Group
High	PC-7; PC-8
Medium high	all high group plus Network B
Medium	all above plus Network A
Low	all with nominal clearance
Special set medium	PC-7; PC-9; Network B

[2780] Further, designation step 422 will include identifying the words, phrases, icons or data objects subject to security or SC concerns and the potential location of the extract data and, if necessary the remainder data and the degree of encryption. The following Selection Table provides some examples.

Selection Table

Level of encryption/storage	type or category of word or phrase; input specific word, phrase
High, web-based storage	dollar values, names of streets, countries, "Smith" and 5 words about "Smith," "avocado"
Medium high, remote storage	all addresses, all names

Medium network storage	all family names, all client names
Low, encrypt and separate store in local memory	all items not in dictionary

[2781] As an example of various encryption methodologies, the following Encryption Table is illustrative.

Encryption Table

DES, random pad A ("r. pad A")
Huffman, r. pad B
Crypto API, r. pad 7
Two fish, r. pad C-2
Blowfish
RC4
Skipjack
Ghost

[2782] In FIG. D-7a, step 424 executes or enables the security program or SC levels with multiple filters, multiple encryption levels and multiple storage levels. Each one of these filters, encryption levels and storage levels correspond to the security level for the various communities of interest. Step 425 responds to an inquiry from a user to reconstruct the document. Step 426 accesses the user's security clearance and the particular inquiry. Decision 428 determines whether the inquiring party is entitled to full or partial access to the source document. If not, the NO branch is taken and the system, in step 429 adds placeholder substitutions. Step 429 may be optional. If YES, the system reconstruct pursuant to the clearance level in step 430. The following provides an example of multiple level encryption utilizing placeholder substitution.

Example: Multiple Level Encryption

Applicants must be _____ zzxx xx _____xxx ____ _____ _____ citizens and have a high school diploma or equivalent. They must possess a valid subsubsub driver's license and qualify for top SUBWORD_____ clearance.

[2783] With this multiple level encryption, substitutions may be utilized "subword" to indicate to the user with a less than superior security level that a certain word, term or phrase has been extracted and stored by he or she is entitled to know that substitute word, term or phrase has been inserted into the plain text document. Of course, any type of substitution character may be used for the placeholder.

[2784] In step 432, the system displays the plain text in a normal format or utilizing a split or bifurcated video memory or utilizing overlay display screen. FIG. B-3 and the description of that figure set forth above describes the normal display in steps 202, 204, the split video memory display in steps 206, 208, 210 and 212 and the overlay display system in steps 214, 216, 218.

[2785] The system, in step 434, monitors and logs the location of the user making the inquiry, the type of inquiry, the time, day, date, clearance level and access level and logs all modifications to the plain text source document. One example of the log is set forth below in the Security Report Table.

Security Report Table

Privacy Scrubber Report
source file: path\filename
scrubbed file: path\filename-scrub
source file: date, time, size
process: date, time
user: name
system: name
Recovery File
(a) storage location, type of encryption, random key
(b) storage location B....
(c) store C .....
(d) store D .....

[2786] Step 436 enables the security program and parses and extracts the data per the security program, filters the data, extracts it and codes it disperses it and stores it as discussed above. The multiple layer security program ends in step 440.

[2787] The following Security Level Access Placeholder Table is another example of the type

of placeholder substitutions that may be available. The example in the Security Table Access Placeholder Table may be used in conjunction with step 429.

Security Level Access Placeholder Table

[security level 2] intelligence located [security level 4] 20 miles from [security level 4]. He is using the name [security level 4], and dressed as a [security level 4] preacher. With him are his lieutenants, [security level 4] and [security level 4]. He is communicating with the international media through Mr. [security level 4], who resides at [security level 3], [security level 4], [security level 4]. Telephone is [security level 1] and Facsimile is [security level 1].

[2788] It should be noted that in order to reconstruct some or all of the plain text source data, some or all of the subsets of extracted data from the extract stores will be utilized dependent upon the respective security level of the inquiring party or user.

D.5 Sharing Data with Different Security Levels - Data Mining

[2789] The present invention can be configured to overcome obstacles to intelligence sharing and data sharing between parties by enabling the parties to identify granular critical data and control the release the granular critical electronic data subject to a sharing arrangement with other parties. In some instances, the controlled release process is designed to implement an agreed upon plan to share secured data or SC data based upon arms length negotiations between the parties. The invention enables a party to release specific granular data such as a name, address, or date without releasing the entire "classified" document. In a commercial context, this is akin to data mining in that the inquiring party seeks limited data (not the entire data file, record or document) and is willing to pay for the "mined" data. As an example of a security intelligence system, a local police chief may release granular critical data about a suspect to a federal agency, when in return the federal authority will release further intelligence "mined" or obtained for the secured data storage, about the suspect. The controlled release of data from the higher security level party (the FBI) may be an intelligence document or a granular part of it (a partial

reconstruction provided to the local police). The rational behind this implementation of the invention is that there are many obstacles for sharing intelligence and information. There are even many more hurdles when it comes to sharing of raw intelligence. The invention creates a leveled playing field in which the different parties must share and exchange information in order to achieve their objectives.

[2790] The invention can be configured to resolve the major challenges facing government by enabling sharing of information between its different organizations in relationship to fighting terrorism. The invention for example can enable organizations, connected to the Homeland Security Department, to search data bases of various other government, state and local organizations, eliminating the fear of the "source" organizations, owning or controlling the source or plaintext documents that their proprietary data or granular critical data is released without their specific permission. The invention enables open negotiations between the parties regarding what data to release and for what consideration. When several organizations are seeking access to a specific document, the invention and can allow a controlled release of different granular data to different parties for different considerations and benchmarks.

[2791] The invention's mechanism of controlled release of the located document/data enables other parties to search their documents without the fear that sensitive information will be released to the searching party. This invention is designed to foster sharing of documentation between different parties, taking into consideration the need to limit the access of other parties to the total content of the owner's document.

[2792] The invention is a machine and process and its purposes and advantages may be as follows: (a) To automatically control selection of data objects within a data stream and release them in a controlled method only to authorized parties. (b) To automatically separate data objects within a data stream into two or more digital data streams according to the importance and categorization of contents,

through extraction and removal of the prioritized content and its replacement by appropriate placeholders. (c) To automatically control selected contents in E-mail, and enable its release in a controlled method only to authorized parties. (d) To enable users to leverage the growth in computer and telecommunications connectivity and electronic commerce by reducing security risks. (e) To enable users to release documents, digital files, and data streams into closed and opened digital networks with the confidence that important, identifying, and critical contents in that documents, digital files, and data streams is secure and will be seen only by authorized parties. (f) To enable real time simultaneous customization and personalization of selected contents within a data stream to different parties, allowing instant display of the selected content or part of it based on, and tailored made to the status of the user or receiving party. (g) To secure the important and critical contents of a document or digital file by transporting said contents into a separated data stream and removing said data stream to a removed storage memory, while eradicating any copies, temporary caches, or traces of the removed extracts on the original computer or machine. (h) To enable instant return transfer to the display or to another display all or part of extracted content instantly with verification of authorized user. (i) To create a projection of the original document, digital file, data objects within a data stream, or variations of it through combined projection of the splinted data streams, while maintaining separation between the data streams. (j) To create an alternative method for security, instead of encryption, which is secure, cost effective, less time-consuming, and flexible. (k) To enable automatic timed removal of specific content items, automatically or manually selected from a document, digital file, or data objects within a data stream. (l) To enable an automatic timed reconstruction (reconstitution) of the said document, digital file, or data objects within a data stream.

[2793] Another object of this invention is as a system and method for automatically creating

customized and personalized versions of a document, data object, or data stream. In real time, simultaneous versions of the original are created and altered, then disseminated based on the status of the different users and their access privileges. The system and method enables content management and control by automatically locating content items prioritized by importance, transporting them to a secure memory, and releasing them under explicit controls or preset rules.

[2794] Another object of the invention is as a system and method for control, analysis and management of important and prioritized information within documents, files, data object, and data streams. The system and method, enables the processing of all data objects at the time in which they are created or imported into the system. The early stage processing, enables early stage inventorying of prioritized contents as well as early stage pattern recognition. Extracting critical information, such as credit card numbers, last names, first names, social security numbers, phones numbers, transaction dollar amounts and addresses, enables the system and method to aggregate data in categories and analyze the data in different optional methodologies including pattern recognition.

[2795] Another object of the invention is as a system and method for comprehensive monitoring of various activities including business activities in real time. With this level of detail, the system and method becomes a management information tool and information/data command and control center. The said system and method can include an alert system, which in effect creates a real time apparatus for command and control of the systems activities. In real time, and at any point in time, the user can get a comprehensive view of different activities including: (a) How many transactions are being processed, their content, their context, identity of the involved parties identity, their profiles, and the personnel involved. (b) How much money is being transacted. (c) When, in terms of dates, relevant to the transaction. (d) Where, in terms of geographical location, the transactions are taking place. (e) Where, in terms of geographical

location, monies or goods are being transferred. (f) Which departments in the organization are involved.

D.6 Multilevel Security through Sanitization with Reconstruction of Sanitized Content

[2796] A multilevel security (MLS) or SC level technology secures the targeted, filtered content with extraction and dispersal to storage, bypassing the use of classification labels, in order to achieve stronger security of the source document or data or better understanding of the SC material. During the process of developing security or SC technologies for defending critical infrastructure, it was discovered that the business model was too complex and there was a need to redefine and create new systems and methods for doing business. As a result, one aspect of the present invention provides a system and codifies methods and business processes to automatically identify, extract, store critical data (as an input security system) and permit reconstruction of critical data only in the presence of certain security clearances (as the output of the security system) or as established by the SC policy release rules.

[2797] The invention is a method and process to establish a stronger multilevel security (or MLS) architecture and product, than is currently available. The invention introduces multilevel security through sanitization of critical content of a source or plaintext document (or data object) with the unique ability to reconstruct all or part of the original document in conformance to the classification level of the user. A user with top classification may view the entire document, while a user with a lower level classification will view a sanitized document, tailor made automatically for his clearance level. The invention secures the targeted filtered content of a document, file, or data stream, through extraction and dispersal to storage, bypassing the common use of classification labels in order to achieve stronger security. The invention enables secure document storage and secure message transfers between users and networks with different security classification levels while protecting the information on a need to know basis.

[2798] Currently multilevel security MLS systems are using multiple PCs for each user, and using

physically separate systems for processing data at each classification level. The inventive system, in several embodiments, eliminates the need for the use of multiple computers. All the documents in the user's PC are automatically secured with a granular classification process generally described above with identification of special security data, extraction from the source document or data object, and then separate storage of the security data. The classified granular content is dispersed to different secure, distributed storage locations. The classification level of a user will determine his right and ability to access and release the stored critical extracted content from the various storage locations for reconstruction. A user with top classification will view the entire document, while a user with a lower level classification will view a sanitized document, tailor made automatically for his clearance level.

[2799] Types of government security levels are: Top Secret (TS); Secret (S); Confidential (C); and Unclassified (UC). Business identifies security levels as: Restricted to Management (R, for example, attorney-client privilege); Proprietary (P); Sensitive (S); and Public (P). These MLS security levels may be supplemented with "need to know" classification labels, organizational limits (Army, Navy, DoD) and time limits. Prior art security systems identified each file with: owner, size, date and time of creation and security attributes. The Bell Lapadula (BPL) security model uses concepts such as domination of the MLS security level over both a process and the subject (a data object). Some examples of various processes are read, execute, overwrite, append, write, kill (delete), etc. Some examples of process rules under the BPL model are: NRU - No Read Up (a lower security level cannot read a document at a higher security level); NWD - No Write Down (a higher level cannot write down to a lower MLS level).

[2800] The invention herein does not use the "classification labels" of the prior art. Instead it creates a situation in which the user gets access rights to specific distributed storage locations based upon his MLS level, each access right can be classified with a different classification level. With respect to the editor

described later herein, security labels (for example ("e.g."), TS, S, C and UC labels) are added or inserted into the filtered but not disassembled document. As explained later, the insertion of these SL labels conforms the current inventive system to the prior art methodology and protocol. However, the current inventive system does not use the SL labels for processing purposes. The current system uses a granular or filter approach to make secure the sensitive data in a particular document. SC labels, matching the relevancy of the SC data may be employed rather than security level tags.

[2801] FIG. D-7b diagrammatically illustrates a multiple level security system accessed by users having different security clearances (which also represents a data mining system and operation). A MLS SC data system uses different SC policy rules permitting access to different SC levels of interest. Source data 100 passes through security program 200. Critical, important data objects or elements are extracted and dispersed into storage 801. In the illustrated embodiment, storage 801 has four security levels SL1 - SL4, level SL4 being the most secure data requiring the highest security clearance. Levels SL1 - SL4 may refer to SC relevancy levels or organizational levels. Between each level is an MLS or multiple level security guard. The guard (physical or software configured) limits transfer of data objects there between. Upon a request or inquiry from user 1, 2 or 3, each having a security clearance s1, s2 or s3, respectively, the query or request for access to data Q1, Q2, or Q3 is sent to security clearance process 803. The process 803 detects and confirms the user's clearance level and passes a cleared query to storage 801. Cleared data (an entire document/data object or a portion thereof or simply one secured data (i.e., a name)), is sent as Data 1, 2 or 3 to clearance process 803. If clearance is still valid, data 1, 2 or 3 is sent to the respective user.

[2802] FIG. D-7b can be a data mining system in that the user is permitted to mine the "cleared" data from storage 801. Data mining may be a monetary charge associated with the clearance function in process 803.

[2803] In a secured system, the documents in the user's PC may be in "declassified" to his security level 99.9% of the time. The "declassified" or available documents are reconstituted through a controlled release of the critical data from storage, and re-classified only when the user presents his identification and his classification level is being verified. Reclassification is automatic at the user's PC. The result is that the user's PC or workstation can operate in classified and un-classified modes. It is unclassified when the documents are declassified and when the documents are reconstituted the user is working in a classified mode.

[2804] The invention introduces a new paradigm whereby computers that are classified as secret or top secret, in actuality will contain 99.9% of the time declassified documents. This capability strengthens substantially the security of such classified systems.

[2805] The invention can resolve the major challenges facing government in enabling sharing of information between its different organizations in relationship to conducting military operations as well as fighting terrorism. The invention for example can enable organizations connected to the Department of Defense (DOD) or the Homeland Security Department to search into data bases of various other government, state and local organizations, eliminating the fear of the organizations owning the documents that their proprietary data or granular critical data would be released without their specific permission. The invention's mechanism of controlled release of the located document/data enables other parties to search their documents without the fear that sensitive information will be released to the searching party. This invention is designed to foster sharing of documentation between different parties, taking into consideration the need to limit the access of other parties to the total content of the owner's document. The invention enables overcoming the obstacles of existing multiple level security MLS systems by enabling sharing of sensitive data, and granular data between parties in a much more flexible way which also enables much greater access to information not enabled by the current MLS systems. The invention includes a controlled release

mechanism for release of data in conformance to benchmarks, which can include submitting of access identification, the giving of consideration, submitting of other information, etc.

[2806] The invention creates better collaboration between users and organizations based on a better flow of information. It enables better efficiency enabling easier communication between users and networks with different levels of classification while maintaining the highest levels of security. The invention enables a much better management of documents in storage and in transport including e-mail. The invention introduces automation to the sanitization process and an automatic reconstruction process. The automation will avoid human error both intentionally as well as unintentionally. The automation will enable a substantial reduction in costs, furthermore the ability to create a multilevel security environment in one PC or workstation will save costs of purchasing operating and maintaining multiple machines as is the current practice.

[2807] The challenge of many organizations is in getting mission critical and time sensitive information speedily to the users who need it. In many cases the needed non- classified or low-level classified information is stored in systems but is not provided to the user who needs it, because the information is in documents which are highly classified. This creates situations in which users are unable to access information, which they need to accomplish their tasks, because of a technological classification barrier. This over classification of information results in hampering critical tasks and activities, as well as creating system redundancies inefficiencies. The DoD (Department of Defense) multiple level security (MLS) was based upon the Bell-Lapadula (BPL) Model. Many believe that the BLP security model is superior to other models. The Bell-Lapadula Model and the existing MLS uses labels to classify users and subject matter. A professional attacker will use his efforts to change or damage the labels in- order to compromise the machines secured information.

[2808] The architecture or the present invention extracts and physically separates data whereby

content is being recognized not by labels by automatically based on the semantic content of the plaintext. In some embodiments of the present invention, labels are added to the granular, filtered document to conform to known organizational protocols. These labels are displayed but not used by the system in processing.

[2809] In the DoD's MLS, data of multiple security levels are processed and transferred by the system, which separates the varying security levels and controls access to the data. In the prior art MLS system, some applications process only one level of data at a time, (for example, when a user edits a document with a word processing tool, the data in the document are treated as if they were a single level, the classification of the document itself). Other applications treat individual data elements at their actual levels. For example, a word processor enforces paragraph and page MLS classification labels, or an MLS data base brings together data elements of different security levels to allow an analyst a multilevel view of the information.

[2810] The vulnerabilities of MLS: The components in the MLS system contain the data in their memories and disks, and the data could be compromised if adequate physical security is not maintained. An attacker who gets access to the system might be able to locate the data or its copies. MLS guards control the flow of information across security boundaries. These MLS guards are known. One concern with the Bell-Lapadula Model and the existing MLS is the use of labels to classify users and subject matter. A professional attacker will use all his efforts to change or damage the labels in- order to compromise the machines secured information. The invention introduces an architecture whereby content is being recognized not by labels by automatically based on the semantic contents of the plain text. The invention sanitizes and enables reconstitution upon valid authentication. It is the only architecture and system which enables both sanitization and reconstitution according to user's verified access identification. The conventional way of classifying documents with high classification (TS), limits the low level clearance users (C) from accessing substantially

un-classified information "granular data" which is in the classified document. Furthermore, the invention enables maximum sharing of unclassified information which lies dormant in classified documents. Top security- secret information is dispersed to distributed storage in many locations. The invention is designed to avoid any one point of failure. The theory behind the architecture is the creation of substantial lines of defense in depth. The attacker will need to break through many obstacles before accessing all the dispersed data of the document. Additional levels of security are provided with multi-type encryption. The system and process introduces the capability to encrypt different parts of a document with different types of encryption. Multi type encryption creates a major barrier to an attacker. Should he wish to break the encryption, he would need many super computers. Should the attacker look for implementation mistakes, even if he finds few, he will still not get access to the total plain text. The inventive system provides flexibility. The system and process delivers flexibility to accommodate changing circumstances. By controlling the level of the granularity, the user can boost the level of security according to changing circumstances. For example, if a competitor becomes a partner the user enables him access to more storage locations, by changing the matrix.

[2811] The system and process integrates the Internet for dispersal and hiding of contents. If a party needs more information it could be released granularly. There is no need to release the whole secret document. The system and process does not use labeling but rather extracts the critical to storage (bu the system may label sensitive text to conform to known protocols). The system avoids situations, in which, attackers may manipulate the labels or the labeling system. Furthermore, the release of information is based on changing circumstances (time, location-GPS, event).

[2812] The invention is a machine and process and its purposes and advantages may be as follows: (a) To automatically control selection of data objects within a data stream and release them in a controlled method only to authorized parties. (b) To automatically separate data objects within a data stream into two

or more digital data streams according to the importance and categorization of contents, through extraction and removal of the prioritized content and its replacement by appropriate placeholders. (c) To automatically control selected contents in E-mail, and enable its release in a controlled method only to authorized parties. (d) To enable users to leverage the growth in computer and telecommunications connectivity and electronic commerce by reducing security risks. (e) To enable users to release documents, digital files, and data streams into closed and opened digital networks with the confidence that important, identifying, and critical contents in that documents, digital files, and data streams is secure and will be seen only by authorized parties. (f) To enable real time simultaneous customization and personalization of selected contents within a data stream to different parties, allowing instant display of the selected content or part of it based on, and tailored made to the status of the user or receiving party. (g) To secure the important and critical contents of a document or digital file by transporting said contents into a separated data stream and removing said data stream to a removed storage memory, while eradicating any copies, temporary caches, or traces of the removed extracts on the original computer or machine. (h) To enable instant return transfer to the display or to another display all or part of extracted content instantly with verification of authorized user. (i) To create a projection of the original document, digital file, data objects within a data stream, or variations of it through combined projection of the splinted data streams, while maintaining separation between the data streams. (j) To create an alternative method for security, instead of encryption, which is secure, cost effective, less time-consuming, and flexible. (k) To enable automatic timed removal of specific content items, automatically or manually selected from a document, digital file, or data objects within a data stream. (l) To enable an automatic timed reconstruction (reconstitution) of the said document, digital file, or data objects within a data stream.

[2813] The invention differs from the current implementations of multilevel security MLS systems based on the Bell-Lapadula Model, and the prior art use of labels to classify users and subject matter. A

professional attacker will use all his efforts to change or damage the labels in- order to compromise the machines secured information. The present invention introduces an architecture whereby content is being recognized not by labels by automatically based on the semantic contents of the plain text. The invention enables overcoming the obstacles of existing multiple level security systems by enabling sharing of sensitive data, and granular data between parties in a much more flexible way which also enables much greater access to information not enabled by the current MLS systems. The invention includes a controlled release mechanism for release of data in conformance to benchmarks, which can include submitting of access identification, the giving of consideration, submitting of other information, etc. The invention creates better collaboration between users and organizations based on a better flow of information. It enables better efficiency enabling easier communication between users and networks with different levels of classification while maintaining the highest levels of security. The invention enables a much better management of documents in storage and in transport including e-mail. The invention introduces automation to the sanitization process and an automatic reconstruction process. The automation will avoid human error both intentionally as well as unintentionally. The automation will enable a substantial reduction in costs, furthermore the ability to create a multilevel security environment in one PC or workstation will save costs of purchasing operating and ,maintaining multiple machines as is the current practice.

D.7 Multiple Independent Levels of Security (MILS)

[2814] FIGs. D-16, D-17 and D-18 diagrammatically illustrate a computer system configured as a multiple independent levels of security (MILS) system. Although the general operation and layout of the MILS system is well known, the incorporation of the inventive system, that is, granular filtration, extraction and re-assembly is unique to the inventive system. In the following MILS system 910, the dispersion and retrieval operation of the present application is discussed using, as an example, one of the most common

environments for protecting classified data, that is, Multiple Independent Levels of Sensitivity (MILS). The MLS system, discussed in connection with security levels, can be implemented with SC based upon either relevancy or based upon enterprise units who have permission (policy rights) to access the SC data.

[2815] As is known in a MILS configuration, each level (TS - top secret; S - secret, U - unclassified) of classified data is isolated from other levels by confining it to set of components dedicated to a single classification level. Data labels are not used within the system, since its components are Commercial-Off-The-Shelf (COTS) products that are not able to handle data labels. The level of the system TS, S or U (top secret, secret or unclassified) establishes an implied sensitivity label. Data is explicitly labeled only in guards 926, 932 (the Dispersal and Re-assembly guard computers) and other MLS devices that connect the MILS system to systems at other sensitivity levels, that connect the TS system to the S system and to the U system. Data transfer between levels is accomplished manually (sneaker net), or through a few, high assurance and closely protected MLS devices, such as guards, digital diodes, etc. A user with the authority to access multiple levels of data is required to use a separate set of interface equipment to access each MILS system. In some cases, keyboard-video-mouse (KVM) switches are permitted.

[2816] Figure D-16 shows an implementation of the present invention in a MILS environment. Two similarly configured domains 911, 915 are shown (upper region consisting of sub-networks 912, 913 and 914 and lower network 915), presumably under different administrative control. Each domain has three networks, one for each of three classification levels. Each network includes a plurality of workstations (only one station being shown in the figure), a Dispersion and Reassembly (D&R) server 926, 932, (a MLS device connected to all networks), and database servers TS, S and U data servers (928, 929 and 930 in the upper domain and 934 in the lower domain) to support the D&R server's proxy document server functionality.

[2817] The user interface components of the present invention reside on the user workstations W St-

U (920), W St-S (918) and W St-TS (916). Dispersion and reassembly functionality is hosted in the D&R servers 926, 932. Digital signatures (encryption) protects the integrity of user data between the user interface at 916, 918, 920 and the D&R server 926. Encryption provides a level of discretionary access controls that prevents disclosure in cases where others have the appropriate clearance ut lack the formal "need to know" level to view the classified information.

[2818] The present example discusses an e-mail and a file server to help clarify its operation in a MILS environment. The system may be configured to handle other documents, images, etc. In the e-mail example, a Top Secret user 916 will prepare a multi-level message and send it to a Secret user 922b, and the recipient 922b will only be able to read the "Secret Level" and below parts. In the file server example, a Top Secret user 916 will prepare a multi-level document and post it to a file server, then a Secret user 922b will retrieve it, getting only the Secret and below parts. These are both downgrading examples. Movement of data within a level and to higher level conforms to existing security rules, for example, Bell-LaPadulla rules.

[2819] The path an e-mail message takes from a Top Secret workstation 916 in one domain 911 to a Secret workstation 922b in the other domain 915. It would work similarly if both sender and receiver were in the same domain; the local D&R server 926 would perform the functions of both D&R servers 926, 932 in this example. The D&R server 926 hosts an e-mail server that is customized to perform D&R functions, as described below.

[2820] An e-mail originator on a Top Secret workstation 916 composes an e-mail using MS Outlook. Originator marks sections of the message with TS, S, and U levels (explained herein), and the system on workstation 916 compartment tags the email document using a combination of the automatic features described earlier and manual tagging (user based), as allowed by the security policy of the accrediting organization. Originator 916 digitally signs and sends the message to the mail server on the D&R system

926.

[2821] The Disperser component of the D&R mail server 926 verifies that the originator 916, identified by e-mail address and authenticated by the digital signature key, has the authority to downgrade messages to the requested levels. The Disperser within Server 926 separates the message according to the originator's tags. The Disperser writes the message sections tagged higher than Unclassified to the Secret and Top Secret servers 929, 928. The unclassified part becomes the base message map stored in server 930 and contains pointers to the higher-level components. The pointers map the re-assembly path when the document is fully or partially re-assembled. This base message map is the only step in the process that requires downgrading. A guard 926 is used to provide increased assurance that the downgrading is done correctly (see FIG. D-17, guard 936). The originator's identity is provided to the guard as part of the downgrade credentials. The Disperser forwards the base message and message map, to the D&R e-mail server 932 of the addressee, using its Top Secret network path c. Other paths 924a, b and c link the same security levels in either domain. Using the TS path prevents downgrading if the target server 932 is not a D&R server. Alternately, the D&R servers 926, 932 could authenticate each other before transferring messages.

[2822] The target D&R e-mail server 932 determines the clearance of the addressee 922b from its network address and looks up the addressee's compartment authorization in its own tables. It then removes links from the base message and map to all message components that are not dominated by the addressee (as defined by Bell-LaPadula). The target D&R e-mail server 932 then retrieves the data for the remaining message components from S database server 929 via guard server 926, constructs or re-assembles the message that the addressee is cleared to read (secret level, not top secret level), and places the message in the addressee's inbox 922b that is at the security level of the highest component in the message. These steps are executed for each addressee. The addressees 922b connect to their local D&R e-mail servers 932 and

download unread e-mail.

[2823] A similar method is used for document serving, such as would be used for file service, web service, ftp service, etc. The document creators tag their documents, sign them digitally, and post them to a D&R file server 926, 932, which is actually a proxy using storage in the database servers 928,929, 930, 934, and others not numbered. The server 926, 932 disperses the message components to its database servers on the MILS networks.

[2824] The following describes the path a document takes from its creation (or modification) in a Top Secret workstation 922a in one domain 915 to a reader on a Secret workstation 918 in another domain 911. The D&R server 932 hosts file servers that are customized to perform D&R functions, as described. The document posting process follows: A document is created on a Top Secret workstation 922a using MS Word, Excel, etc. The originator 922a marks sections of the document with TS, S, and U, and compartment tags using a combination of automatic and manual tagging features, as allowed by the organization's security policy. Originator at 922a digitally signs and sends the document to the file server on the D&R system 932.

[2825] The Disperser component of the D&R file server 932 verifies that the originator 922a, identified by the digital signature, has the authority to downgrade documents to the requested levels. The Disperser in server 932 separates the document according to the originator's tags, creating a base document at the lowest tagged level (U - Db or S - Db) that contains pointers to the more classified components (TS - Db 934). The base document is encrypted using an algorithm appropriate to its level. The Disperser writes the document sections to the file servers (U - Db or S - Db or TS - Db 934) according to their tags. This is the only place in the system where downgrading is performed. A guard 932 is in the downgrade path to provide additional assurance that data is not leaked. The user identity derived from the originator's 922a digital signature on the tagged message will be provided to the guard 932 as part of the downgrade

credentials. The disperser places references to the document in the directory of each file server 932, 926 for which a version of the document can be built (e.g., if there is no unclassified data in the document, a reference to the document will not be placed in the unclassified directory).

[2826] The document retrieval process follows. A user on a workstation 916 logs onto its local D&R proxy document server 926. The strength of user authentication is application specific and determined during system accreditation. The user 916 locates a document and requests that it be downloaded for reading, specifying a requested sensitivity level and compartment. The Re-assembler component of the D&R server 926 loads and decrypts the base document. The Re-assembler in server 926 verifies that the requestor 916 dominates the requested security level and compartment, based on the level of the network over which the request was received and the contents of its compartment authorization table. The Re-assembler in server 926 constructs the document to the authorized security and compartment level. The Re-assembler provides the document to the requestor. The re-assembly function does not violate Bell-LaPadula and does not require downgrade authority.

D.8 MLS Server and Workstation Components

[2827] FIGs. D-17 and D-18 diagrammatically illustrate the server and workstation. Figure D-17 shows the D&R Server 926 hosted on a DigitalNet CTS-400 system, which is currently evaluated. It consists of the CTX-400 hardware and the STOP 6.0 operating system. The disperser part 937 of the D&R server 926 has the most security critical functions, since it must move data contrary to the Bell-LaPadula security policy. It maintains a table of user downgrade authorizations 938 keyed by the originator's public digital signature key 939 and downgrade authority 941. A guard 936 is also included, which may employ a version of automated tagging system of the present invention to identify sensitive data. The double check of the security label tag for each part of the message/document provided by the guard 936 is particularly important since the

EX. 1002, Page 1027

downgrade labels are applied in a single-level system.

[2828] The re-assembly side (in re-assembler 942) requires no policy violation, but requires its own table of user compartment authorizations 944 because the MILS systems do not have the ability to label data. The table 944 is keyed by e-mail address (for e-mail routing) or user identity 945 from the user authentication process (for file service). Authorized compartments 946 are coupled to email user id 945.

[2829] The components of the user workstations are shown in Figure D-18. A MILS workstation 916 is a single-level component, that is, it is coupled to a single security level network. The user interface will in some cases be installed on existing workstations 916 and in other cases new workstations procured for this use. The workstations include an operating system Windows 2000, a graphical user interface Windows GUI, WS Office as a document processor, digital signature system rated at EAL 4 and a rated user interface.

D.9 Flexibility of the Present MLS Approach

[2830] The late-binding techniques or retrieval and reassembly features used in the present invention is a compelling feature, because it provides solutions to some of the most significant problems of sharing data in modern warfare and international coalition operations. A single comprehensive document can include data sensitive to many different environments, yet by tagging its components correctly, maximally useful versions can be provided to readers with widely different authorizations, and without having to identify them all ahead of time.

[2831] For example, in a coalition environment, countries or classes of countries, or agencies within countries can be assigned non-hierarchical and hierarchical labels and then a single document, suitably tagged, can be securely distributed to all of them using a properly configured and administrated infrastructure of the present invention.

[2832] In the presently configured MILS embodiment, the tagging protocol is, at a minimum,

confirmed by the guard in the D&R servers. In prior art systems, only the guard tags documents. In the presently configured system, the thin client applications program on the workstation tags the document segments but the guard confirms this preliminary tagging. One feature which seems to be important in certain security systems is that the re-assembly map be cleaned or scrubbed to match the then current security level where the map is stored. This theory, carried forward, would include the concept that the re-assembly map, stored in each version of the secured document, only point to the next level storage location. Hence, if all portions TS extracts, S extracts, C (classified) extracts and U (remainder) document are dispersed into, for example TS 928, S 929, C (not shown) and U 930, then the document in U 930 only includes a pointer to the location of C extracts and the C extracts only includes a pointer to the S extracts and the S extracts includes only a pointer to the TS extract location. The downgrade discussed above refers to the location of the higher secured extract.

D.10 Multiple Extraction Filters and Application Outline

[2833] There is a need to construct filters which supplement the initial list or compilation of security sensitive (sec-con) or select content (SC) words, characters, icons and data objects (herein "word/objects"). The need arises either due to the fact that the initial security or SC word/object list is incomplete, or that the author of the initial list is concerned that the list is too limited or in order to defeat a attack or an inference engine "reverse engineering" the sanitized document and ascertaining not only the filter (a type of code) but also the sensitive word/object removed from the source document. Further, the incorporation of a filter generator enhances the current user friendliness of the program. In its current embodiment, the program is configured as an editor to screen and sanitize a source document. The user selects, at his option, functional aspects which include: compliance with laws (an application of a type of filter, e.g. HIPAA, GLB, Oxley-Sarbanes, EU privacy, executive orders); privacy (another type of filter which excludes, for example, social

security numbers, see also, EU policy); search for and supplement filter; pay per view (which enables the user to buy missing sensitive information (for commercial purposes); survival (which creates a distributed and dispersed copy of the user's document and other stored documents and items using predetermined storage facilities); security (which triggers the various security routine discussed herein); and storing (which permits the user to select which of the several storage options the extracted sensitive or SC data/objects should be employed in the dispersal.

[2834] The filter routine diagrammatically illustrated in FIG. D-19 is useful in compiling a filter which separates both the sensitive word/objects and contextual and semiotic and taxonomic aspects of the initial list of security sensitive (or select content) word/objects. The filter works in conjunction with a compilation of data, typically located on a network which could be private or public. In low level security or SC situations, the filter may access Internet databases to gather additional data for the filter. In more secure or SC systems, the filter could access a secure or SC data base (one located at the same security level as the user) and build or compile the additional word/objects. The filter program 950a in FIG. D-19 begins with step 952a which compiles the initial list of security sensitive (or select content) word/objects. In 954a, the initial list is supplemented with dictionaries, phone books, corporate records (to obtain subsidiary data and trade names) and thesaurus data. Each of these represent different compilations of data and the added data is added to the initial list of sensitive word/objects. In 956a, a search is conducted on a network, usually through a search engine, to gather excerpts near and abut the keywords. These keywords are the initial sensitive word/objects. Statistical algorithms are applied to gather non-common word/objects which are associate with the keywords as found in the additional data compilations. The goal of the adaptive filter is to obtain contextual, semiotic and taxonomic words, characters or data objects from the compilation of additional data related to the security sensitive (or select content) words, characters or data objects. Semiotic

is a general philosophical theory of signs and symbols (read language and words and objects) that especially deals with their function. Semiotics include syntactics, semantics and pragmatics. Syntactics is the formal relationship between signs. Semantics is the meaning of signs and pragmatics is the relationship between signs and their users, such as the relationship of sentences to their environment. Taxonomy is the scientific classification and categorization of items. Therefore as an example, a search through the Internet on Google search engine under "Bin Laden" may shw a number of uncommon (non-dictionary words) within 200 words of the target "Bin Laden." This search string would gather documents form the Google search and copy 200 words on either side of "Bin Laden" and then extract only non-dictionary words into a supplemental list. This type of filter algorithm looks for contextual matters close or near to the target. The search is semiotic and statistical in nature. Additionally, the initial supplemental list would identify the Bin Laden is an arab and this classification (a taxonomic aspect) can be used to expand the list for the filter. The algorithm may include a simple command to gather all 10 words on either side of Bin Laden. This is a pure contextual search and the "10 word" aspect is a statistical number. From the supplemental list, all pronouns, prepositions and conjunctions may be eliminated. Spiders or robots may be used in the gathering of the contextual and semiotic filter data. The contextual, semiotic and taxonomic words, characters or data objects from the compilation of additional data is all related to the initial list of security sensitive (or select content) words, characters or data objects.

[2835] Step 958a compiles the adaptive filter. The above noted contextual, semiotic and taxonomic filter is adaptive since it can be used to expand (and potentially contract or reduce) and adapt an existing list of sensitive word/objects to a larger list which better protects the source document and inhibits the operation of an inference engine. Step 959a repeats the filter gathering and compilation for various levels of security. Higher security may require a broader search (1000 uncommon words near Bin Laden and add all Arabic and

sub-Asian continent cities). Orthogonal security groups (those groups having the same level, e.g. S Secret, with each other but being different organizations, e.g, Department of Defense compared to the FBI) often have different methods to keep data secret between compartments.

[2836] The adaptive filter can be set to automatically gather additive sensitive word/objects. The system, with a basic filter, may identify a sensitive word in a paragraph being scanned by the initial filter. This sensitive word may be a special word in the existing filter or may be a non-common word not found in the initial filter. The adaptive filter system may then obtain this "unknown" or "special" word, and conduct a search through a compilation or data base of additional words, etc. Any new word/objects falling within the contextual, semiotic and taxonomic words, characters or data objects from the compilation of additional data (database) related to said security sensitive (or select content) words, characters or data objects are then added to the filter. The expanded filter is then used to screen the source document.

[2837] Step 960a compiles a supplemental filter with random words, phrases, etc. in order to further defeat an inference engine reverse engineering assault on the secured and sanitized document. In some sense, the production and use of a random filter is an encryption technique since the resultant filtered product, in order to be understood by others, must be reverse filtered or decrypted to reveal the document at the appropriate security level. Nonsense words may be added to this supplemental filter. Step 962a applies the primary filter (with the security word/objects and the additive word/objects from the contextual et al. filter) to the source document. Step 964a extracts the sensitive word/objects per security level. It is noted that several filters are used, on one for each security level, whether hierarchical or orthogonal. The extracted word/objects are stored or th partially extracted document per security level is stored in the corresponding security cleared data base or storage. Step 966a applies the supplemental filter to the remainder or lowest classified document. Step 968a stores the supplemental random filter to permit the low level user to decrypt

the document. Step 970a publishes, distributes or pushes the document to others having a need to know. The pointer to the location of the supplemental filter decoder is encrypted and stored in the filtered remainder document. This permits the low level person to decode the remainder document.

D.11 Multiple Security Features for Data

[2838] FIG. D-9 diagrammatically illustrates a flowchart showing the key components of a multiple encryption program using multiple types of encryption in one document or data object. Multiple levels, types or modes of encryption are utilized in the same document or data object to enable securing data and transparently managing the separation of user-based communities of interest based upon crypto-graphically separated, need to know security levels. These security levels are associated with a plurality of encryption types or with different cipher keys using the same encryption. An example of a multiple level encrypted document is shown above in the Multiple Level Encryption sample. Different levels or modes or types of encryption are listed in the Encryption Table above.

[2839] Step 510 in FIG. D-9 initializes the system by organizing different security levels with different encryption types and cipher keys. Also, the program sets filters to create the multiple encryption or ML document or data object. Step 512 filters the document or data object. Step 514 encrypts the extracted data for each security level. These steps 510, 512 and 514 utilize many of the routines discussed above in connection with FIGS. B-4 and D-7a, steps 232, 234, 236, 238, 240, 422 and 424. Step 516 recognizes that the secured document or data object may be stored for later use (with associated multiple decryption), published, distributed, or otherwise utilized to achieve the primary purpose of the document, i.e., to communicate information or to safely store security critical information. Step 518 permits the user, with the proper security clearance to retrieve the document or data object. Step 520 illustrates that the user must retrieve his or her cipher key to decode all or a portion of the ML encrypted document or data object. This

step may be manual which engages the user to into certain codes or may be automatic such that the user's computer automatically, without operator input, decodes all or part of the document or data object. Step 522 decrypts the document pursuant to the user's security clearance. Step 524 recognizes that the user may review, re-publish, store, comment on, re-encrypt or otherwise deal and handle the full or partially decoded document or data object. The program ends or otherwise continues with other programs set forth herein. It should be noted that storage of the extracted data may be included in the flow path of the program in FIG. D-9 is necessary.

[2840] FIG. D-10 diagrammatically illustrates a chart showing the key components of the parsing, dispersion, multiple storage and reconstruction (under security clearance) of data. Document or data object 100, in function element 550, is created or obtained by the input computer device. The document is stored in a normal manner in customary data store 552. A parsing algorithm function 554 is utilized in parsing step 556. The parsing algorithm, as stated earlier, targets the plaintext document or data object 100 and splits, cuts and segments (that is, parses) the document by bit count, word, word count, page, line count, paragraph count, any identifiable document or icon characteristic, or other identifiable feature such as capital letters, italics, underline, etc. Hence, the parsed document 100 constitutes at least remainder data and data which is extracted or parsed or segmented out. A plurality of data extracts may be obtained. The parsed data (which is both the extract data and remainder data) is then dispersed into storage facilities data store DS 1, 2, 3, 4, etc. Preferably, the parsed documents are encrypted as shown by "e" in FIG. D-10. In order to facilitate the potential reconstitution of document 100, a map is stored in a map storage 558. Hence, the dispersement 560 largely spreads out or distributes the parsed document 100 to a plurality of memories in the distributed computer system. These memories may be removable memory devices (floppy disc, removable tape drive, CDs) or may be more fixed devices such as hard drives, Internet storage facilities, etc. Preferably, the map

is also encrypted.

[2841] Reconstruction step 562 enables a person with the appropriate security to obtain the map from map storage 558, decode the map, gather the dispersed, parsed segments of document 100 and compile the document. This is noted in function 564.

[2842] Since the original document 100 is stored in a customary manner in data storage 552, the parsed document stored in multiple data storage units DS1-DS4 provides a unique backup for document 100. The algorithm can employ many different mathematical constructions but is, in the current embodiment, primarily based upon one or more of a bit count, a word, a word count, a page count, a line count, a paragraph count, and identifiable document characteristic, and identifiable word characteristic, and identifiable icon characteristic and identifiable data object characteristic, capital letters, italics, and underline found in the plaintext document or data object. Further, the parsing algorithm can generate different security levels wherein parsed segments are stored at different storage facilities having various degrees of security clearance. This establishes a hierarchy of data storage units and corresponding degrees of security clearances. The parsing algorithm may identify unique words or strings of data, i.e., credit card numbers. The hierarchy of security clearances may involve first a password, second a biometric confirmation such as a voice match and a third highly unique biometric characteristic such as a fingerprint or retinal scan. The parsing system enables a large distribution of data in a secured environment. In this manner, if the original data object 100 at customary data storage 552 is destroyed, a person with an appropriate security clearance can reconstitute the original data document 100 due to the secured parsing and dispersal of document 100 through data storage units DS1-DS4 and map storage 558. The parsing may occur on a granular level. In particular, the parsing may occur on a financial document in electronic form.

Financial Document Table

Startcode; Abel, Robert, NMI; 100567; TRANSFER803; To8900586943; FROM3897622891; $700.00; end code

[2843] In the Financial Document Table, the start code and end code is typically represented by a digital code unique to the communications channel, the name on the account has no middle initial (NMI) and the various words "transfer 803" and "to 8900586943" and the words "from" and "$" are represented by predefined numeric or alpha numeric codes. The electronic financial document complies with an established protocol. In any event, financial documents are often times transmitted through electronic communications and telecommunications channels. The present invention, in one embodiment, enables a higher level of security by parsing the financial document or data stream. Further, a higher level of security may be employed by extracting identified text or characters and storing the extracted text as discussed above in connection with FIGS. B-1A, B-1B and B-2.

[2844] To some extent, the present system can also be utilized for key management and encryption systems.

[2845] In a broad sense, the parsing methodology disclosed herein is not based upon the separation of critical versus non-critical or classified versus non-classified security information. The primary focus of the parsing methodology is (1) automatic transparent parsing of data content into granular data groups which are thereafter dispersed to different storage locations in order to maintain a very high level of security with or without encryption; (2) dispersal of the segmented data to different storage locations each which, potentially, demand additional identification or security clearance prior to the release of the stored segmented data, including, possibly, the creation of a digital bureaucracy, in order to hinder or circumvent digital attacks on the plaintext document or data object; (3) proposing and implementing a system wherein the user has a very basic appliance since most of the user's data is stored both locally (customary data storage 552; FIG. D-

10) and parsed and stored in a distributed system (DS1-DS4) and wherein an important asset is the map stored in map location 558; (4) enabling an institutional system to parse highly confidential information and extract the same in granular form and disperse the same throughout the Internet or other storage locations with or without encryption without compromising the document's security privacy and integrity.

[2846] The process involves parsing the documents or content into granular data groups and optionally creating small groups of data wherein the data segments cannot be recognized even to the level of providing 2-4 data objects in each file; dispersing the granular data groups into different storage locations; creation of a map of dispersal to the different storage locations (wherein the map is secured and encrypted and stored); and reconstructing the documents or data content. The reconstruction utilizes the map of dispersed and distributed storage and requires the presentation of security clearances such as passwords, biometric information and/or physical identifiers for access at the storage level and potentially at all the other data storage sites. The data is compartmentalized through distributed storage and sometimes requires separate security clearance. This need for presenting additional security clearance at different storage locations (DS1-DS4) creates a digital bureaucratic process which enhances the security level of the entire system. The selection and extraction of data and dispersal of that data to select storage locations can be established under different criteria. For example, one level of criteria extracts last name, address and social security numbers. Another criteria extracts every other line, every third word, etc. The parsing algorithm can utilize random selection or systematic selection as long as the parsing algorithm is documented and utilized in reconstruct step 562. The parsing algorithm may be stored with map and map store 558 or may be stored separately. An additional feature, as discussed above, involves utilizing place holders or adding substitute content to the remainder data of the parsed document 100. The use of place holders and substitute content may be thought of as an algorithm for the parsing. By using place holders and substitute data, private or

highly confidential data is masked insuring privacy, security, and confidentiality. The ability to parse the information and/or extract security information is important for financial transactions. The transactions which require account numbers (see Financial Document Table above) are useless without the account numbers. The security of the account numbers, whether identified and extracted or severely parsed and segmented, stored and reconstituted under security clearances, is enhanced by the present system.

[2847] To achieve a very high level of security, the system can optionally incorporate a two-man key system. The system automatically separates the selected data stream into one or more data groups and extracts one or more of these data groups and disperses them into data storage DS1-DS4. To release the extracted data groups and/or critical content, the reconstruct step 562 may require two persons submitting identification credentials or security clearances. This two-man key method is a further protection against identity theft and insider attacks. The two-men key system can be implemented on a regular basis or on an emergency basis when there is need for a higher level of security.

[2848] Financial documents sometimes include substantial amounts of numerical data such as financial projections, balance sheets, electronic funds transfer messages, etc. It should be noted that the extraction may be based upon a particular item such a digit and a nine digit number representing money or may be parsed automatically based upon some parsing facility. Of course, the financial document may also be viewed as a data stream with delimiters ";" separating fields in the data stream. The parsing algorithm may work on the data in each field as well as different fields in the entire data stream.

[2849] Most storage facility systems require a map in order to reconstruct the original plaintext document 100. The map may be encrypted and may require a secret key sharing scheme for access thereto. Further, the map may be a physical map (a printout) or may be stored on a removable data storage medium, rather than be an electronic representation. In some instances, a map is not necessary. For example, if the

security data or the parsed or segmented data were automatically stored on a floppy disc, the originator of plaintext document 100 could move the floppy disc from the computer system thereby physically safeguarding the security data or the segmented, parsed data. Without the disc, another person or the originator of plaintext document 100 could not reconstitute the document. The originator may deliver the floppy disc to another in order to permit reconstitution. The same is true regarding removable tapes and CD-ROMs.

[2850] Advantages of the present parsing system, methodology and program, include the ability to connect to unsecured networks without adversely affecting the overall security of the plaintext document 100; less dependence on existing security system including fire walls; the reduction of the requirement to keep daily updates regarding vulnerabilities of the computer system originating plaintext document 100; the security of plaintext document 100 is not dependent upon the number of access points into the network or number of users located on the network originating plaintext document 100; there is no damage to the parsed and stored backup version of plaintext document 100 if new security systems are installed wrong or misconfigured and there is no damage if system administrators turn OFF the existing security systems or improperly install or operate the security systems.

[2851] The parsing system can operate as a main security operation or an emergency backup system or as a customary backup system. The plaintext source document or data object may be preserved with or without encryption, or destroyed as a further data security step. The parsing and dispersement of data protects plaintext document 100 and insures the survivability of plaintext document 100 if the system originating plaintext document 100 comes under significant electronic or physical attack. That is, if customary data storage 552 is destroyed electronically or physically, the survivability of data in the plaintext document 100 is established by the present system. The storage of granular data groups most likely would

defeat any attempt to view the entire content of plaintext document 100. Only verified user users with a confirmed security clearances or identifications verified at reconstruct step 562 and in data storage sites DS1-DS4 are permitted to reconstruct plaintext document 100. Further, the parsing of the system can be triggered based upon an electronic attack, an electronic hack or a physical environmental detection scheme. This system immediately protects of the critical data plaintext document 100 with a transparent, automatic parsing, dispersal and storage system.

[2852] It should be noted that various aspects of the methodology and program described above in connection with FIGS. B-1A through B-4 and D-6 through D-9 can be incorporated into the parsing methodology and program in order to enhance or modify the system.

D.12 Adaptive Data Security - Event Responsive System

[2853] The present invention can also be configured as an adaptive security program which adapts and adjusts the security provisions based upon intrusion into a particular network or attempts to electronically attack or hack into that network or successful hack events. Programs are available to track electronic attacks or hacking attempts. One of these programs is manufactured by Cisco and identified as the Cisco Intrusion Detection System (IDS). The Cisco IDS system can work on a server or on PCs in a network. The Cisco IDS is an electronic intrusion detector, or an electronic attack detector or a hacking monitor. The hack or attack monitor is software loaded into a designated computer.

[2854] The output of the electronic attack or hacking monitor loaded into PC 142 (FIG. B-2) for example, or loaded into PC-6 acting as a server for Network A 404 in FIG. D-6, generates a plurality of attack warnings. The attack warnings progressively and incrementally indicate the severity and degree of intrusion and hacking attacks directed to the computer system. The following Security Level Table illustrates an example of various responses to increasing levels of attacks. These increasing security responses include

engaging the filter and extracting critical data and storing it locally; the next level involves storing the critical data on removable storage media; the next higher level involves offsite storage of all security data; the subsequent security alert results in multiple offsite storage for multiple levels of security or critical data and the highest level involves offsite storage of both common data (remainder data) and security data. Of course, other combinations responsive to the hack attack may be provided. The electronic attack monitor may use artificial intelligence AI to (a) assess the severity of the attack, (b) plan an appropriate "secure data" response, (c) select the degree of filter, extraction and/or encryption, and (d) locate secure extract data storage sites. AI or inference machines can ascertain (a) traffic on communications channels, both intra and inter network, (b) storage limit issues, (c) transmission failures in the communications links, and (d) the degree of security necessitated by exterior events, i.e., terrorism alerts, virus alerts, war, data security warnings posted by trusted sources, MicroSoft, Norton, NASA, DoD, CDC, FBI, etc. Higher security alerts trigger the AI security monitor to heighten the security level (or to decrease that security level in view of a reduction or withdrawal of an electronic attack). Aspects of AI systems, inference engines and neural networks are discussed above in conjunction with the AI configured filter. These AI aspects can be utilized with an AI configured security sensor.

Security Level Table

Attack (low threat level) Level One
    engage filter
    local storage - disk drive
    encrypt map
Attack (moderate threat level) Level Two
    same as Level One but use removable storage media (local)
Attack (nominal attack) Level Three
    Engage higher level filter
    Off site storage, single storage for all security data
Attack (moderate attack) Level Four
    Multiple off site storage, multiple levels of security data
Attack (severe attack) Level Five

[2855] Hence, the filtering of data is based upon respective ones of the plurality of attack or hack warnings and the extraction of data and degree of extraction is dependent upon respective ones of the plurality of attack - hack warnings. Storage of the extracted data and the remainder data is also based upon the degree of attack which is reflected in the attack - hack warning issued by the monitor.

[2856] FIG. D-8 diagrammatically illustrates a flow chart showing the key components of the adaptive security program adaptable to various levels of hacker of electronic attacks. Step 460 senses all intrusions and attempts, that is, electronic attacks, hack attacks or hacking actions on a computer or a computer network. This step is equivalent to the output of the attack - hack monitor. Step 462 assesses the current network performance, adjusts the storage location for the extract data (the location of the extract store), the encryption level (the degree of encryption) and the storage of the map showing the extract data storage (if necessary) and storage of remainder data, if necessary given the severity of the attack. For example, during high utilization of the computer network (high utilization in a server computer in a server-client environment), local storage of extracted data may be preferable as compared with offsite storage of critical data. However, if the attack occurs during non-working hours, the performance of the network is very high, and the security system could utilize all the resources in the computer network to achieve the security goal of safe guarding the data during the attack. System resources include processing resources (for encryption/decryption), bandwidth resources to store extract data and any other resources that are critical for the utilization of the security system described herein. Decision step 464 determines whether a threat or attack as occurred. If not, the system takes the NO branch returns to step 460. If YES, the system in step 466 assigns an attack level or a hack warning level to the threat or attack. The system in decision step 468, monitors the network during the attack. If the network performance or the computer performance does not

change, the YES branch is taken. If the computer performance or network performance changes based upon or during the attack, the NO branch is taken and the system returns to step 466 which reassigns an attack level or a warning level to the next higher or significantly higher warning levels.

[2857] After decision step 468, the system executes step 470 which assigns the security level and implements the security program based upon the attack. It should be noted that the administrator establishes the degree of security level, the encryption, the extract store and remainder store (if necessary) for various levels of attacks or hack warnings. The security level assigned to a particular attack warning is implemented in step 470. Decision step 472 determines whether the security program's communication path is clear. For offsite storage of extract and/or remainder data, a communication path is important. If the path is blocked or compromised by the attack, the NO branch is taken and the system in step 473 reassigns the security level to a next higher level or a different, safer security level and returns to step 470. If the security and communications path is clear, the YES branch is taken from decision step 472 and, in step 474, the system maintains the security program. Decision step 476 determines whether sufficient time has passed from the attack. If not, the system loops to step 474. If YES, the system executes step 478 which either permits reconstruction of the user operating the plain text or source document or automatically reconstructs those documents that were filtered, parsed, extracted, and subject to outside storage. The system ends in step 480. To provide additional security, the attack monitor can be configured to monitor security warnings from trusted parties such as MicroSoft, Norton, NASA, DoD, CDC, FBI, etc. Emails or electronic communications from trusted parties can trigger higher levels of security. The attack monitor described above can be configured to accept messages from trusted parties. These messages are equivalent to detecting an electronic attack.

[2858] Further, the attack - hack monitor can be configured to monitor and assess other environmental conditions such as fire, power failure, equipment failure, unauthorized physical entry into the

building, plant, or computer room. These exterior threats or events are monitored by the attack monitor since they may quickly develop into an electronic attack on the secured data retained by the computer system. In response to these exterior events, the attack monitor generates corresponding attack warnings similar in nature to the hack attack warnings discussed above.

[2859] There are various methodologies that may be utilized in the adaptive system. The tables in FIGS. D-8 A (Standard Automatic Defenses Matrix Table), D-8B (Optional Automatic Defenses Matrix Table), D-8C (Security Meter Module Table), D-8D (Normal Work Mode Table), D-8E (Threat Made Table), and D-8F (Attack Mode Table) sets forth these various security methodologies.

[2860] Another object of the system and method is to enhance the survivability of a system, network, or an organization through distribution of critical information. The objective is to enable a network or organization to carry on its critical missions even while under attacked or damaged. Survivability is the ability of a system to execute its mission and provide critical operational services during and after a successful intrusion or damage. Providing critical operational services includes maintaining availability of information and data such as credit card numbers, names, phone numbers, transaction amounts, shipment details without compromising the security of the information and data.

[2861] The invention is designed to enable a network to adapt to ongoing attack and react in a way that permits critical missions to continue. With the current state of the art, when firewalls or other security measures are compromised, no real obstacles curtail or hinder intruders. The system and method is very adaptable and flexible to provide additional layers of security, privacy, anonymity, redundancy, and backup through the selection, extraction, storage, transportation, and reconstruction processes. The dynamic architecture of the invention enables it to conduct an automatic real time configuration of its extraction/transport/recovery activities, in response to the challenge of attacks.

[2862] The invention's survivability modes enable: (a) Presetting of rules for computer or network functioning under attack or alert. (b) An automatic assessment of damage and automatic reaction to enable functionality of critical missions.

D.13 Email, Web-based and Other Types of Applications

[2863] FIGS. D-11A and D-11B diagrammatically illustrate a flowchart showing the key components of one embodiment of the present invention, that is, an e-mail security system. A similar system may be deployed to capture select content Sc data in real time FIG. D-11A is linked to FIG. D-11B via jump points D-11-A and D-11-B. The method of securing e-mail data operates on a distributed computer system which at least includes a remote memory designated as an extract store. Of course, the extract store may comprise a plurality of extract stores operative in conjunction with a plurality of security clearance levels. The extract stores may be SC stores. A singular security or SC relevancy level is identified in FIG. D-11A. Further, the e-mail may be subject to a parsing algorithm which, as discussed above, is generally independent of the identification of security sensitive (or select content) data. However, with respect to the parsing aspect of the present invention, the original e-mail data is split into extracted data and remainder data and the extracted data is stored in an extract store. Hence, the parsing algorithm operates essentially independent of the content whereas the secured e-mail program operates based upon content identification. Although FIGS. D-11A and D-11B primarily relate to identification of security or SC data, the same is true regarding the use of securing e-mail data with a parsing algorithm.

[2864] The e-mail security or SC system begins with step 602 wherein the system or program is turned ON or is activated. Step 603 recognizes that the user originating plaintext document 100 (not shown) has set a security or SC filter identifying one or more security sensitive (or select content SC relevant) words, characters or icons. In step 604, the user composes the e-mail representative of plaintext document 100. In

step 606, the user selects the "send" command in the typical e-mail program. As is customary, the system in step 608 conducts a spell checking routine prior to sending the e-mail. In step 610, the system conducts a security or SC check on the plaintext document or composed e-mail generated in step 604. The filter is used in step 604. In step 612, security or SC words are highlighted or distinguished in the e-mail prior to the actual sending of the e-mail to the addressee. This step 612 is optional. In step 614, the user selects the security or SC words for data to be extracted out. The highlighting step facilitates this selection. In step 616, the system extracts the security or SC data and, preferably, in step 618, the security or SC data is encrypted. Step 618 is optional. In a parsing application to secure e-mail, the parsing algorithm operates automatically at step 610 thereby eliminating steps 612 and 614. The extracting step 616 simply represents that the segmented data obtained from the original plaintext e-mail generated at step 604 is separated from remainder data.

[2865] After encryption step 618, the e-mail security or SC system generally operates in one of three manners. Other systems may be formulated based upon the systems and subsystems discussed herein. In one methodology, a second e-mail is created (see step 629), in a second methodology the secured or SC data in encrypted form is attached or appended to the original e-mail containing remainder data (step 621) or, in a third methodology, the encrypted security or SC data is simply added to or inserted into the end of the remainder data of the e-mail (step 623). The methodology of generating a second e-mail is initially discussed.

[2866] A second e-mail having encrypted security or SC data is created in step 620. Further, the system in step 622 adds a hyperlink to the remainder data in the original e-mail created in step 604. The hyperlink presents a pointer for the addressee to a secured application service provider ASP or SC application service provider ASP. See the discussion of FIG. B-2 above. The ASP represents a data storage facility for the secured or SC e-mail data. In step 624, the remainder data from the original e-mail is sent to the addressee in a normal manner. This step also includes the concept that the second e-mail containing the encrypted

security or SC data is sent to the ASP. In step 626, the addressee receives the remainder e-mail which includes a hyperlink to the secured or SC data ASP. The system jumps at jump step D-11-A from FIG. D-11-A to FIG. D-11-B.

[2867] In step 628, the addressee receives the remainder e-mail, visits the ASP via the hyperlink and clears the security levels at the secured ASP. In the SC system, the SC data is stored in the SC storage and this data capture is logged into the system. In step 630, the secured or SC data ASP obtains a map for each secured or SC data e-mail (since the original e-mail may be broken up into a plurality of extracted, secured or SC data e-mails) obtains all secured or SC data e-mail and decrypts the same. In step 632, the secured or SC ASP downloads the secured or SC data as an e-mail to the addressee. In step 634, the addressee system compiles the original plaintext e-mail 100. In the SC mode, the SC data in added to the remainder data and the user-viewer is permitted to see additional SC content associated with that relevancy level. In this sense, the SC email is a knowledge expander permitted the user-viewer to enhance his or her knowledge with additional SC content. A reconstruction program may be necessary to decode the secured or SC data and insert the data into the document via the placeholders.

[2868] Optionally, the decryption could occur at the recipient's e-mail device somewhat prior to the reconstitution of the e-mail plaintext document 100 during step 634. This requires the addressee to have the encryption routine and the correct key or decrypt code. The e-mail security or SC system described above may include many of the features discussed earlier in connection with the other figures. For example, both the security or SC data and the remainder e-mail data can be encrypted prior to transmission to the addressee and the secured or SC data ASP. The encryption may include multiple levels of encryption and decryption may require multiple levels of security clearance. The encryption may be mixed in the remainder e-mail. Partial as well as full reconstruction is enabled as discussed above in connection with FIG. B-3.

[2869] From the senders or originator's viewpoint, the e-mail facility described herein facilitates the storage of the extracted data at one or more secured or SC sites.

[2870] Another implementation of the secured or SC e-mail system attaches the encrypted and secured or SC data to the remainder e-mail data as indicated in step 621. For an SC email system, the addressee may add the SC sent with the email to the addressee's SC database. E-mail attachments are well known. Alternatively, the encrypted secured or SC data may be embedded or copied in encrypted form at the end of the remainder data in the original e-mail as indicated in step 623. In either case, in step 625, the e-mail is sent to the addressee. In step 627, the addressee opens the attachment. In step 629, the system of the recipient decrypts the secured or SC data attachment or the embedded data attachment. In step 631, the recipient's system integrates the now decrypted secured or SC data with the remainder data. Of course, this a compilation step. Place holders or other position indicators are customarily utilized. Appending the encrypted security or SC data is generally equivalent to attaching a file to the original e-mail which constitutes, after extraction, the remainder data. Including the encrypted security or SC data is adding the security or SC data to the original e-mail at a predetermined location (either the top of the e-mail, the bottom of the e-mail or some predetermined line number). All SC data may be added to the addressee's SC store. Relevancy of the SC is based upon SC relevancy markers transmitted with the email, similar to the security levels for the MLS secure email system.

[2871] It should be appreciated that the e-mail or SC security system may work automatically or may be selected manually by the user. The highlighting or special distinguishing manner for the security or SC words in step 612 is optional. By highlighting the security or SC words, the user may select or deselect those words for extraction. At the addressee's side, the addressee's system may be configured to automatically seek out the secured or SC data ASP, enter security clearance data, download the secure or SC data and integrate

the secure or SC data in the remainder data e-mail (or SC data structure). The present invention contemplates automatic as well as manual steps in steps 626, 628, 630, 632 and 634. The hyperlink with the original remainder e-mail essentially maps the remainder data to the secured or SC data and the remote storage locations handling the secure or SC data. Multiple security clearances may be required of the recipient or addressee. The e-mail system can be combined with other features of the security or SC system discussed above such as multiple security data locations, secret key sharing schemes, multiple encryption of the data in a single document, multiple security clearance levels required for a plurality of storage facilities, the two man key system, automation of key management and a plurality of levels of access to the data such as partial reconstruction in step 634 and full reconstruction.

[2872] FIGS. D-12A and D-12B diagrammatically illustrate a flowchart showing the key components of one embodiment of the system and the invention which implements the security or SC system on a web browser. Jump point 12-A links FIG. D-12A to FIG. D-12B. The system, at step 700 is ON. The filters establishing either the parsing or the identification of security or SC data are established in the filter set step 701. In step 702, the user inputs data into open field of an HTML display page which the user has previously downloaded from a web server. In step 704, the user may select "secure now" turning ON the system (or ON SC operator) or the system may automatically be ON such that the filter is screening all the data input by the user in the open field. In step 706, the system scans all the open field data, locates security or SC data and extracts security or SC data. In step 708, place holders are added to replace the extracted security or SC data in the remainder data and a hyperlink is added to the open field remainder data providing a link to the secure or SC data ASP. In the SC system, the placeholders show the addressee critical terms subject to the SC knowledge expander feature. In step 710, the user selects the "send button" or any other indicator on the HTML page triggering an operation which transmits the open field data (which is now

remainder data) to the web server. In step 712, the web server and particularly the common gateway interface (CGI) receives the remainder data fields, identifies the place holders in the data and the hyperlink to the secure or SC data ASP. In step 714, the web server receiving the data from user's browser goes to the secure or SC data ASP, inputs and clears any security level, and obtains the secured or SC data. In step 716, the web server reconstructs the open field data which generally is represented by plaintext document 100. In step 718, the web server processes the data as necessary. Many of the features discussed above in connection with the above figures may be implemented on the browser system.

[2873] The credit card scrubber or financial data scrubber operates in a similar manner to the email and browser data security system described above. A similar system can be designed to comply with privacy laws or other governmental regulations. A credit card number is similar to a social security number and patent name, address and phone are common personal identifiers that some privacy laws require extraction and special handling. The credit card or financial data or social security number scrubber (herein collectively "CC scrubber") typically operates on a defined sequence of numbers. For example, if a credit card number is 17 digits, whenever the email or browser security system or program detects17 sequential numerical digits (a pre-set filter), a pop-up window may appear enabling the user to select or turn ON the scrubber. If ON, the data security program strips or parses the credit card number and sends, for example, five of the 17 digits to a secure store. In an SC system, the CC data may be data mined by affiliated companies. Placeholders or substitute characters may be inserted into the remainder CC data. To reconstitute the entire CC data, the intended recipient would be required to pass security clearance levels at the secure store. An SC system requires SC relevancy clearances set by the policy administrator. Of course, the CC scrubber could be set to detect bank account numbers, personal or business account holder names, pre-set passwords, etc. In an OFF state, the CC scrubber would let pass the CC number, account number or pre-set data stream or string. The

user may select (i) always ON; (ii) pop-up window, select ON or OFF per transaction; (iii) pop-up window to select OFF (default being ON); or (iv) always OFF but minor reminder (audible sound, icon appearance, etc.) of data security risk. The CC scrubber may encrypt the extracted data for security. Other visual ques may rather than a pop-up window may be used (for example, a drop down menu). The scrubber can also be deployed on wireless devices to scrub sensitive data such as credit card and other financial data.

[2874] FIG. D-13 diagrammatically shows several revenue systems which may be employed with the data security systems described herein. Many types of revenue systems may be employed in conjunction with the present invention. FIG. D-13 shows two basic systems, one at the data input stage and the second at the data output or reconstruction phase. Release of the reconstructed document or portions thereof are based upon security clearance or SC relevancy and compensation. "Compensation" includes an exchange of information, particularly for the SC storage system. Within each revenue subsystem are two types of revenue generators, an advertising revenue generator and a user charge generator. The user charge system contemplates charging or assessing a fee to the user's employer or organization. Therefore, the system operator may select up to four (4) revenue generation systems (ads at the input, charges at the input, ads at the output and charges at the output). It is well known that vendors selling goods and services over the Internet are willing to pay a certain percentage of their sales revenue to other entities referring customers to the vendor's web sites. The concept of display ads in FIG. D-13 includes this revenue stream. The system operator may choose all, one, several or none of these revenue systems to be deployed in conjunction with the data security system described earlier herein. Other revenue system may also be utilized. The steps in the revenue system described herein may be reorganized to attain higher consumer and user acceptance and/or to maximize the revenue to the system operator. Swapping SC content may include an evaluator module that tests the "proposed" SC data file and determines whether the proposed SC data file is worthy or being exchanged for an existing SC data file.

[2875] Decision step 730 determines whether the system is deployed at the data input phase or not.

It is clear that the system operator may utilize the data reconstruction revenue system and hence the decision step 730 is not necessary. If the data input system is employed, step 732 displays the ad to the user. The user may be uploading a complete document to an application server on the Internet or may be using a application service provider on the Internet or an private LAN to secure his or her data. The display ad 732 step enables the user to click on the ad and visit the vendor, thereby potentially generating a referral fee. See referral fee branch 757. Step 734 requires password clearance. Step 736 processes the document or data object with the security or SC system. The user may input the document real time or input it to the application server or may upload the complete document to the server. Alternatively, the ad could be buried in the email or application program run on the user's computer and the user would be shown an ad and given a link to the vendor's Internet site. Selecting the link points the user's browser to the vendor's site.

[2876] Step 738 shows display ad 2 to the user thereby potentially generating referral revenue for the system operator. Step 740 notes that the user exits the revenue system. Step 742 determines whether the system charges the user for the security or SC service. If YES, the program processes the charge in step 745 (charge systems are known). If NO, the system ends or returns to other programs in step 747.

[2877] The NO branch from determination step 730 leads to the receipt of a reconstruction request by the user in step 750. Step 752 determines whether the user will be charged. If YES, the system executes step 745. If NO, the system displays the ad 1 in step 754. Referral generation is noted by branch 757 from step 754. In step 756, the user's password is subject to clearance. In step 758, the user's request is processed, the document or data object is reconstructed (fully or partially as described earlier), and in step 759 the system displays ad 2. In step 762, the user's activity is logged in to the system. Step 764 determines whether the charge to the user is reduced (because he or she viewed the ads) and if not, the system ends in step 747, if YES, the system processes the charge in step 745. Alternatively, the user may be shown display ads and/or charged for services upon storage of extracted data. Step 750 includes this concept.

D.14 Portable Computing Device Environment

EX. 1002, Page 1052

[2878] The invention can be applied to portable computing devices to secure files and data objects in such devices. The invention extracts, disperses, via a controlled release of data segments to storage locations, and permits reconstruction utilizing security protocols to provide a security or SC system for data based upon the location of the portable device, typically detected by a global position signal generator (GPS) or based upon triangulation data from several broadcast points. In an SC system, the sensitive data may be made available to the user-viewer for a period of time or as long as the user is in a geographic territory. Beyond that time-space limit, the SC data is removed from the user's portable data device. Scrubbing security or SC icons from maps, credit card data or financial data from text, a data object or data stream is part of the portable security system.

[2879] As used herein, the term "portable computing device" means a laptop computer, a PC with a movable feature, such as a PC mounted in a car, plane, truck or trailer, PDAs or personal data assistants, mobile or cellular phones configured with a memory, a processor and some type of GPS or locator system to determine where the phone or cellular unit is located within a territory and digital pagers having similar electronic systems.

[2880] The present invention can be linked with a location sensing circuit, such as a global position sensor or system (GPS) or other type of location sensing system, such as a system which utilizes triangulated signals. The concept is a location based access oriented security (or SC relevancy) such as an automated trigger (which activates the security or SC program discussed hereinabove when the portable computing device is beyond a predetermined region); an automated safety system; a trip wire; an interlock; a method to disable systems, activity or access to data; and means to limit functionality or access in whole or in granular parts. The portable security system operates on text, data objects, images or other digitally configured data objects. Security access (or relevancy access) is limited by a location way point (in relation to a reference point) or a calculated range (using satellite GPS, high altitude services, or earth-based range finding GLS (geographic location services)) about a way point with physical means or mathematical calculations to define

a geographic area by equations or geometric shapes or aggregated ranges (the shapes including rectangles, solids, cubes, circles, oval, spherical region or other areas defined by algorithms). Physical and logical access or entry control to weapons, devices, vehicles, computers, equipment, tools, data, networks, local access, remote access beyond a physical location (reference point), can be enabled or disabled with the system of the present invention. The regions (sometimes identified as a single predetermined region or a plurality of predetermined regions), can consist of complex definitions of three dimensional areas of arbitrary shape and sizes, as long as those regions can be defined by algorithms. The region can also be defined as an area circumscribed internally by a perimeter or by an area external to that perimeter. In other words, access can be denied if the portable device is within a certain region as compared with denying access when the device is beyond a predetermined regions. The claims are meant to cover both situations.

[2881] FIG. D-14 diagrammatically illustrates a portable computing device 810 at location B. The portable computing device 810 includes, in the illustrated embodiment, a GPS system (or a receiver system) 812 coupled to a bus 814 and further coupled to memory 816, a processor 818 and an input/output system 820. Input/output 820 is coupled to, among other things, a key board or key pad, a display, and possibly a transmitter and receiver subsystem. As is known, GPS Systems detect satellite positioning signals and generate an output indicative of the location of the GPS system. In the illustrated embodiment, this location is location B in FIG. D-14.

[2882] A simple implementation of the present security or SC system provides that upon detection of d1 from location A, defined by building 822, certain security events occur, e.g., automatic extraction and a denial of reconstruction rights. In one example, GPS subsystem 812 continually monitors the location of portable device 810. When the location of device 810 exceeds a predetermined distance (d1-limit), the program operating in memory 816, operable by processor 818, either extracts data and stores the extracted data as discussed in detail above or prohibits reconstruction of data as requested by the operator of portable device 810. Alternatively, automatic extraction may occur without prohibiting reconstruction due to device

810 being located beyond the predetermined region d1-limit. The portable computing device 810 in FIG. D-14 may have many other electronic components such as those shown in FIG. B-2 in connection with computer 165. Alternatively, the security or SC system can be configured in a reverse manner such that the extraction of security or SC information is triggered when portable 810 is within a predetermined region (less than d1-max) close to location A and building 822, that is, the security or SC system disclosed above is triggered to extract information when distance d1 is less than d1-max.

[2883] The security or SC system can also be configured such that GPS or locator system 812 detects a variable distance such as distance d2 between truck 824 and location B of portable device 810. In this sense, the location of portable device 810 is obtained by GPS circuit 812 and further some type of communications must be established between truck 824 at location C and the portable device 810. For example, the receiver coupled to input/output 820 receives this information from location of truck 824 and location C. This reference location C is then processed in conjunction with the location data from GPS circuit 812 by processor 818 and memory 816. The same results as discussed above in conjunction with fixed reference location A can be achieved with a variable reference location C. Truck 826 at variable location D enables the system to provide an additional level of security. In other words, within distance d2 (d2-limit), the operator of portable device 810 may be able to reconstruct information upon request. However, if portable device 810 intrudes upon or is less than distance d3 (d3-max) the distance between location B and location D, the security system may trigger an immediate extraction routine thereby disbursing, on a granular basis, the secured or SC words, data objects or whatever and further prohibit reconstruction. Of course, the security or SC system could be configured simply to extract the information and permit reconstruction. Otherwise, the security or SC system could be configured to simply extract information and prohibit reconstruction. In this manner, the security or SC system discussed in conjunction with the portable computing device 810 can have multiple triggers or location established events enabling the security or SC program to extract security or SC information or disabling the security or SC program to prohibit

reconstruction based upon a fixed location A or one or more variable locations C, D.

[2884] Another configuration of the present invention utilizes triangulation to obtain location B for the portable computing device 810. In a triangulation situation, the receiver system 812 for the locator receives signals from one or more fixed locations, and preferably three locations diagrammatically illustrated by tower T1, T2 and T3 in FIG. D-14. The triangulation of signals to obtain location B is known in the art. However, the combination of such location position coupled to extraction of security or SC data and/or reconstruction of security or SC data based upon certain location parameters is part of the present invention.

[2885] For illustration purposes only, the present security or SC system for the portable computing device 810 can remotely store extracted security or SC information. Remote store 828 coupled to receiving tower T1 illustrates this concept.

[2886] FIG. D-15 diagrammatically illustrates a basic flow chart for the portable security or SC system program utilized in connection with portable computing device 810. Decision step 830 determines whether a certain event has occurred. This event may include a power ON for the portable computing device, may be a "save document" command, may be a screen ON event or may be a timed function. For example, if the portable computing device 810 is continually ON, the program may periodically poll the locator circuit (GPS 812) and determine whether location B is within or beyond the predetermined regions (d-max or d-limit). Step 832 activates the GPS or the triangulation circuit to obtain current location data. This may include locating variable locations C, D. Step 833 obtains territory limits for various security levels or SC relevancy levels. As discussed earlier, security or SC level SL1 is public or non-confidential information, security or SC SL2 confidential or proprietary information, level SL3 is secret information and level SL4 is top secret or SC information. This system can be configured such that various territories or predetermined regions correspond to respective ones of the security or SC levels SL1-SL4.

[2887] Decision step 834 determines whether location B of portable computing device 810 is within or without the predetermined territory limits or predetermined region. If YES, the system determines in

EX. 1002, Page 1056

decision step 836 whether the user has initiated a reconstruction request. If not, the system returns to a point preceding decision step 830, the detect event function or step. If YES, the system, in step 838, reconstructs the document only if location B is beyond a predetermined region. Of course, a negative operation could occur in that reconstruction would be prohibited if location B was within a predetermined region. The claims appended hereto are meant to cover both within a region and without a region and independently, extract or permit reconstruction. Step 840 secures the reconstructed file again as necessary.

[2888] Returning to decision step 834, if location B is not within the predetermined regions or territories defined by security or SC levels SL1-SL4, the NO branch is taken and decision step 842 determines whether portable computing device 810 has any unsecured files. If YES, the system executes step 844 which is extract and store the security or SC sensitive words, data objects etc. in accordance with security or SC levels SL2-SL4. As stated above, the storage could be on media in a local drive or can be remotely distributed to memory segments designated as a remote extract store. If the NO branch is taken from decision step 842, the system executes decision step 846 which determines whether the user has requested a reconstruction of data. If not, the program ends or returns to the event detection step 830. If YES, the system executes step 848 which determines whether a lower security or SC clearance is available within the current territory, determines whether the user has the proper pass code to access the reconstruction and process the reconstruction or whether the system prohibits all reconstruction. Partial reconstruction for lower security or SC items may be permitted. For example, reconstruction at top secret or SC level SL4 may be prohibited when distance d2 is greater than d2-limit but reconstruction at a lower security or SC level such as confidential level SL2 may be permitted beyond limit d2-limit. In this sense, the present invention can be configured to generate extraction for various security or SC levels at various predetermined regions based upon a fixed reference point or a variable reference point. Alternatively, reconstruction can be permitted or denied based on a plurality of security or SC levels and a plurality of corresponding regions or distances. The term "mobile predetermined region" is sometimes utilized in conjunction with variable regions d2 and d3.

[2889] As an example, the data object retained by portable computing device 810 may be a map having security or SC sensitive icons on the map. These icons are extracted if location B is less than a predetermined defined distance d3-limit between variable location D and location B. If location B is beyond d3-minimum, the map can be viewed by the operator on portable device 810. If location B is less than distance d3-minimum, the security or SC icons are removed from the map. In a similar sense, security or SC sensitive credit card characters can be extracted from plain text documents carried on portable computing device 810 when device 810 is a certain distance beyond d1-limit from fixed reference point A. This location A may be a bank headquarters. Encrypting and decrypting the data based upon the geographic event is also contemplated by the present invention. Of course, portable device 810 may be a plurality of portable devices linked via a hard wire network or via a wireless network. The same security or SC program disclosed above in herein can be utilized with one computer or a series of computers. Further, portable computing device 810 can include a plurality of memory segments (see FIG. B-3A) and may include a plurality of display screens as discussed above in conjunction with FIG. B-3. The extraction and storage and reconstruction of streaming data is possible as is operation on voice data. Additionally, the portable computing device may set off an audible and/or visual alarm prior to extraction of data. For variable territories or predetermined regions, step 832 or 833 may include gathering information regarding the variable location of vehicles 824, 826 prior to determining the territorial limits for various security or SC levels SL2, SL3 and SL4.

E.1 General System Comments

[2890] It is important to know that the embodiments illustrated herein and described herein are only examples of the many advantageous uses of the innovative teachings set forth herein. In general, statements made in the specification of the present application do not necessarily limit any of the various claimed inventions. Moreover, some statements may apply to some inventive features but not to others. In general, unless otherwise indicated, singular elements may be in the plural and vice versa with no loss of generality. In the drawings, like numerals refer to like parts or features throughout the several views. The section titles

are not meant to limit the detailed description of the system and process described therein.

[2891] The present invention could be produced in hardware or software, or in a combination of hardware and software, and these implementations would be known to one of ordinary skill in the art. The system, or method, according to the inventive principles as disclosed in connection with the preferred embodiment, may be produced in a single computer system having separate elements or means for performing the individual functions or steps described or claimed or one or more elements or means combining the performance of any of the functions or steps disclosed or claimed, or may be arranged in a distributed computer system, interconnected by any suitable means as would be known by one of ordinary skill in the art.

[2892] According to the inventive principles as disclosed in connection with the preferred embodiment, the invention and the inventive principles are not limited to any particular kind of computer system but may be used with any general purpose computer, as would be known to one of ordinary skill in the art, arranged to perform the functions described and the method steps described. The operations of such a computer, as described above, may be according to a computer program contained on a medium for use in the operation or control of the computer as would be known to one of ordinary skill in the art. The computer medium which may be used to hold or contain the computer program product, may be a fixture of the computer such as an embedded memory or may be on a transportable medium such as a disk, as would be known to one of ordinary skill in the art.

[2893] The invention is not limited to any particular computer program or logic or language, or instruction but may be practiced with any such suitable program, logic or language, or instructions as would be known to one of ordinary skill in the art. Without limiting the principles of the disclosed invention any such computing system can include, inter alia, at least a computer readable medium allowing a computer to read data, instructions, messages or message packets, and other computer readable information from the computer readable medium. The computer readable medium may include non-volatile memory, such as

ROM, flash memory, floppy disk, disk drive memory, CD-ROM, and other permanent storage. Additionally, a computer readable medium may include, for example, volatile storage such as RAM, buffers, cache memory, and network circuits. Furthermore, the computer readable medium may include computer readable information in a transitory state medium such as a network link and/or a network interface, including a wired network or a wireless network, that allow a computer to read such computer readable information.

[2894] In the drawings, and sometimes in the specification, reference is made to certain abbreviations. The following Abbreviations Table provides a correspondence between the abbreviations and the item or feature.

Abbreviations Table

A-com	computer or memory store for common or remainder data
ASP	application service provider - server on a network
B-ext	computer or memory store for extracted data
bd	board
CD-RW	compact disk drive with read/write feature for CD disk
CIO	Chief Information Officer or system administrator
comm.	communications, typically telecommunications
comp	computer
CPU	central processing unit
DB or db	data base or structured data file
D & R	dispersion and re-construct or re-assemble
Distrib	distribution or transmission of data (information) files
DS	data storage
doc	document
dr	drive, e.g., computer hard drive
e	encryption
ext-data	extracted data
F	floppy computer drive or flash drive memory
H-tax	hierarchical taxonomic system, a prioritized classification system
I/O	input/output
I-com	Internet storage for common or remainder data
id	identify
I-ext	Internet storage for extracted data
KE	knowledge expander, such as a knowledge expander search
loc	location
MLS	multilevel security
mem	memory
mod	module or software-firmware-hardware function
obj	object, for example, a data object

EX. 1002, Page 1060

opt	optional
PC	personal computer or any other type of general computer
pgm	program
re	regarding or relating to
recon	reconstruct
rel	release
req	request
rev	review
SC	select content -- e.g., words, terms, images, sound that is of particular interest to the user
sec	security
sec-con	security sensitive content
sec. level	TS top secret, S secret, C classified, UC unclassified
sel	SC sensitivity level, sometimes SC sel. level
sel cont	select content
sel levels	SC levels, hierarchical taxonomic classification levels of SC
sch	search
SL	security level (sometimes S1 for sec. Level 1, S2 is Level 2, etc., also, for example, TS is Top Secret, S is Secret, C is Classified, U is Unclassified))
supple	supplemental
supple doc	supplemental document of data object
sys	system
sys-admin	CIO or system operator
t	time
tax	taxonomic or classification system
tele-com	telecommunications system or network
unkn	unknown item or data element or data object
URL	Uniform Resource Locator, x pointer, or other network locator
W St	computer work station

[2895] "Data," as used herein, includes any data object, e.g., text, image, icons, audio, video, still images, etc. and data is sometimes referred to herein as a "data object." A source document is either a document, any data structure, or a data stream. Since a data stream has a start bit or term and an end bit or term, the data stream is structured data, and reference to a "document" refers to any document, data structure, or data stream. Likewise a "supplemental document" is any document, data structure, or data stream. The select content (SC) is any critical or security sensitive word, character, image, or data object as pre-defined or as established by the user. The user may specifically design the filter, begin with a dictionary to define common terms, identify any additional SC words, letters, images, icons, partial versions of the foregoing or

any other granular aspect of the source document.

[2896] FIG. C-16 diagrammatically illustrates a personal computer or PC computer system 140, a second PC or computer 142, and a third PC -3. Similar numerals designate similar items in FIGs. C-16 and B-2. PCs 140, 142 and PC-3 are connected together via a network 145 (LAN or WAN) and connected to one or more private databases or data structures which are represented by Db A. The LAN 145 is also connected to an input/output device 146 that may be generally described as a router or an I/O device to an outside communications system. The input/output device 146 is connected to a telecommunications system 148 which leads to Internet 150. The Internet is a global computer network. Internet 150 is coupled to a plurality of servers, one of which is server 152. Server 152 may be designated as an application service processor ASP. The ASP provides various tools to the user on computer 142 such a Google search through other data bases, indices and data structures. Internet 150 also includes various computer memory storage devices such as computer storage B 154a, computer storage C 156a accessible through Gate C 157a (via password or other control device) and computer storage D 158a accessible via Gate D 159a. Access to computer storage C and D is permitted via, in one instance, security level clearance modules or Gate C and D 157a, 159a. Maps to the computer stores C and D may require security level clearance.

[2897] Typically, PCs include a keyboard or data input device 161, a display 163, a central processing unit CPU 165, a video board 167 having video board memory 169, a fixed disc hard drive 168, a RAM 166, and input/output device 164, a removable memory F media (flash or floppy) drive 162 and a removable compact disk (CD) read-write (CD-RW) device or drive 160. The system may include other removable disk drives, tape drives, or flash memory units. Internal units CPU 165, video board 167, hard drive 168, RAM 166 input/output device 164, floppy drive 162 and CD-ROM device 160 are all coupled together via an internal bus 171. Bus 171 represents a plurality of buses as is known to persons of ordinary skill in the art.

[2898] One methodology of implementing the present invention utilizes filters on the PC or on

EX. 1002, Page 1062

removable memory F. The present invention may also be embodied utilizing an Application Service Provider (ASP) on server 152 and in a client-server network. The user or operator on the PC 142 calls up an ASP on system 152 and operates the KE process on the computer system 152. Filters and rules may be located with a uniform research locator or URL to find filters, data collections, target files from computer store B, C and D. In a client-server environment, server 152 acts as a server generally cooperating with data operations with client computer 140. Of course, persons of ordinary skill in the art recognize that the server may be located on the local area network 145 rather than being interconnected with Internet 150 as shown in FIG. C-16. The claims appended hereto are meant to cover the alternative embodiments.

[2899] As an example of a client-server or web-based implementation of the present invention, the user at computer 140 may define the SC 2018 as described above, and input data via keyboard 161 or load source data from F drive 162 or CD-ROM drive 160 into RAM 166. Alternatively, the data stream on LAN 145 may be monitored by PC 140. In any event, whether the source data is input via keyboard 161 or copied or accessed in or from F drive 162 or CD-RW drive 160, the source data is filtered as discussed above in connection with FIGs. C-1 - C-2. Prior to filtering, it is appropriate for the user at computer 140 to identify the hierarchical taxonomic system and the location of SC filters and My Group Profiles. Off site data storage and processes are available permitting activation of server 152 and enabling the server to take over the process directly from user 140. In other words, the user at computer 140 could call up the URL of the server 152, the server could request certain user information (user name, password), and would request data from the client computer to establish the SC filter and H-tax classes pursuant to input selected by the user. The client compute may (a) filter the source doc thereat or (b) send the data to the server for filtering. The server could store data either locally on computer 140 or remotely at computer memories 154a, 156a. Either the PC or the server could conduct the knowledge expander search and return the compiled data, supple docs, and map results to the PC.

[2900] The claims appended hereto are meant to cover modifications and changes within the scope

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☒ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:**

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

and spirit of the present invention.

[2901] What is claimed is:

1. A method of organizing and processing data in a distributed computing system having select content represented by one or more predetermined words, characters, images, data elements or data objects, said distributed computing system having a plurality of select content data stores in a server cloud for respective ones of a plurality of security designated data and scarcity designated data, each with respective access controls thereat, said plurality of select content data stores and server cloud operatively coupled over a communications network, the method of organizing and processing data comprising:

extracting and storing said security designated data and said scarcity designated data in respective select content data stores of said server cloud;

withdrawing some or all of said security designated data and said scarcity designated data in said respective select content data stores only in the presence of said respective access controls applied to said select content data stores in said server cloud.

2. A method of organizing and processing data as claimed in claim 1 including applying a monetization algorithm to said security designated data and said scarcity designated data based upon non-secure availability of said security designated data and said scarcity designated data and generating and assigning either a financial value to said security designated data and said scarcity designated data or generating and assigning a risk-release value to said security designated data and said scarcity designated data.

3. A distributed computer system for organizing and processing data having select content represented by one or more predetermined words, characters, images, data elements or data objects, said distributed computing system having a plurality of select content data stores in a server cloud for respective ones of a plurality of security designated data and scarcity designated data, each with respective access controls thereat, said plurality of select content data stores and server cloud operatively coupled over a communications

Ex. 1002, Page 1065

network, comprising:

means for identifying plurality of select content data stores for respective ones of a plurality of security designated data and scarcity designated data in said server cloud,

a processor activating at least one of said select content data stores in a server cloud there by permitting access to said select content data stores and respective security designated data and scarcity designated data based upon an application of one or more of said access controls thereat;

said processor having means for applying a reconstruction data process employing said respective access control to combine one or more of said security designated data and scarcity designated data in said server cloud.

4. A distributed computer system for organizing and processing data as claimed in claim 3 including means for applying a monetization algorithm to said security designated data and said scarcity designated data based upon non-secure availability of said security designated data and said scarcity designated data and means for generating and assigning either a financial value to said security designated data and said scarcity designated data or generating and assigning a risk-release value to said security designated data and said scarcity designated data.

Best Available Copy

Abstract

Distributed computer system processes data having select content (SC) represented by one or more predetermined words, characters, etc. The system has a plurality of SC data stores in a server cloud for respective security designated (Sec-D) data and scarcity designated (S-D) data, each with respective access controls thereat. The SC data stores are is operatively coupled on a network. An identification module for identifying SC data stores for the Sec-D data and S-D data in the server cloud. A processor activates an SC data stores in the server cloud thereby permitting access to the SC data stores and respective Sec-D data and S-D data based upon an application of one or more of said access controls thereat. The processor has a reconstruction module operating as a data process employing the respective access controls to combine one or more of the Sec-D data and S-D data.

G:\RCK\CLIENTS\Redlich\Patents\6851-08-CIP-pa.wpd

EX. 1002, Page 1067