JC408 U.S. PTO
10/27/99

DOCKET NO. : REFH-0112 **PATENT**

JC553 U.S. PTO
60/161640
10/27/99

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:

Frederick Herz, Paul Labys and
David Parkes

Serial No.: Unknown **Group Art Unit:** Unknown

Filing Date: Herewith **Examiner:** Not yet assigned

For: SECURE DATA INTERCHANGE

60161640 - 102799

EXPRESS MAIL LABEL NO: EL531271660US
DATE OF DEPOSIT: October 27, 1999

Box Provisional
Assistant Commissioner for Patents
Washington DC 20231

Sir:

PATENT APPLICATION TRANSMITTAL LETTER

Transmitted herewith for filing, please find

☐ A Utility Patent Application under 37 C.F.R. 1.53(b).

It is a continuing application, as follows:

☐ continuation ☐ divisional ☐ continuation-in-part of prior application number ___/_____.

☒ A Provisional Patent Application under 37 C.F.R. 1.53(c).

☐ A Design Patent Application (submitted in duplicate).

Including the following:

☒ Provisional Application Cover Sheet.

☒ New or Revised Specification, including pages 1 to 150 containing:

☒ Specification

☐ Claims

☐ Abstract

☐ Substitute Specification, including Claims and Abstract.

☐ The present application is a continuation application of Application No. _____ filed _____. The present application includes the Specification of the parent application which has been revised in accordance with the amendments filed in the parent application. Since none of those amendments incorporate new matter into the parent application, the present revised Specification also does not include new matter.

☐ The present application is a continuation application of Application No. _____ filed _____, which in turn is a continuation-in-part of Application No. _____ filed _____. The present application includes the Specification of the parent application which has been revised in accordance with the amendments filed in the parent application. Although the amendments in the parent C-I-P application may have incorporated new matter, since those are the only revisions included in the present application, the present application includes no new matter in relation to the parent application.

☐ A copy of earlier application Serial No. _____ Filed _____, including Specification, Claims and Abstract (pages 1 - @@), to which no new matter has been added TOGETHER WITH a copy of the executed oath or declaration for such earlier application and all drawings and appendices. Such earlier application is hereby incorporated into the present application by reference.

☐ Please enter the following amendment to the Specification under the Cross-Reference to Related Applications section (or create such a section) : "This Application:
☐ is a continuation of ☐ is a divisional of ☐ claims benefit of U.S. provisional Application Serial No. _____filed _____
_____
_____
_____.

6265540.1027799

☐ Signed Statement attached deleting inventor(s) named in the prior application.

☐ A Preliminary Amendment.

☒ 12 Sheets of ☐ Formal ☒ Informal Drawings.

☐ Petition to Accept Photographic Drawings.

☐ Petition Fee

☐ An ☐ Executed ☐ Unexecuted Declaration or Oath and Power of Attorney.

☐ An Associate Power of Attorney.

☐ An ☐ Executed ☐ Copy of Executed Assignment of the Invention to _____ _____ .

☐ A Recordation Form Cover Sheet.

☐ Recordation Fee - $40.00.

☐ The prior application is assigned of record to _____

☐ Priority is claimed under 35 U.S.C. § 119 of Patent Application No. _____ filed _____ in _____ (country).

☐ A Certified Copy of each of the above applications for which priority is claimed:

☐ is enclosed.

☐ has been filed in prior application Serial No. _____ filed _____ .

☐ An ☐ Executed or ☐ Copy of Executed Earlier Statement Claiming Small Entity Status under 37 C.F.R. 1.9 and 1.27

☐ is enclosed.

☐ has been filed in prior application Serial No. _____ filed _____ , said status is still proper and desired in present case.

☐ Diskette Containing DNA/Amino Acid Sequence Information.

☐ Statement to Support Submission of DNA/Amino Acid Sequence Information.

FEE CALCULATION:

☐ Cancel in this application original claims _____ of the prior application before calculating the filing fee. (At least one original independent claim must be retained for filing purposes.)

			SMALL ENTITY		NOT SMALL ENTITY	
			RATE	FEE	RATE	FEE
PROVISIONAL APPLICATION			$75.00	$	$150.00	$150.00
DESIGN APPLICATION			$155.00	$	$310.00	$
UTILITY APPLICATIONS BASE FEE			$380.00	$	$760.00	$
UTILITY APPLICATION; ALL CLAIMS CALCULATED AFTER ENTRY OF ALL AMENDMENTS						
	No. Filed	No. Extra				
TOTAL CLAIMS	0 - 20 =		$9 each	$	$18 each	$
INDEP. CLAIMS	0 - 3 =		$39 each	$	$78 each	$
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM			$130	$	$260	$
ADDITIONAL FILING FEE				$		$
TOTAL FILING FEE DUE				$		$150.00

☒ A Check is enclosed in the amount of $ 150.00 .

☒ The Commissioner is authorized to charge payment of the following fees and to refund any overpayment associated with this communication or during the pendency of this application to deposit account 23-3050. This sheet is provided in duplicate.

☐ The foregoing amount due.

☒ Any additional filing fees required, including fees for the presentation of extra claims under 37 C.F.R. 1.16.

☒ Any additional patent application processing fees under 37 C.F.R. 1.17 or 1.20(d).

☐ The issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance.

☒ The Commissioner is hereby requested to grant an extension of time for the appropriate length of time, should one be necessary, in connection with this filing or any future filing submitted to the U.S. Patent and Trademark Office in the above-

identified application during the pendency of this application. The Commissioner is further authorized to charge any fees related to any such extension of time to deposit account 23-3050. This sheet is provided in duplicate.

SHOULD ANY DEFICIENCIES APPEAR with respect to this application, including deficiencies in payment of fees, missing parts of the application or otherwise, the United States Patent and Trademark Office is respectfully requested to promptly notify the undersigned.

Date: October 27, 1999

Michael P. Dunnam, Esquire
Registration No. 32,611

Woodcock Washburn Kurtz
Mackiewicz & Norris LLP
One Liberty Place - 46th Floor
Philadelphia PA 19103
Telephone: (215) 568-3100
Facsimile: (215) 568-3439

© 1997 WWKMN

09465450 - 102796

identified application during the pendency of this application. The Commissioner is further authorized to charge any fees related to any such extension of time to deposit account 23-3050. This sheet is provided in duplicate.

SHOULD ANY DEFICIENCIES APPEAR with respect to this application, including deficiencies in payment of fees, missing parts of the application or otherwise, the United States Patent and Trademark Office is respectfully requested to promptly notify the undersigned.

Date: October 27, 1999

John W. Caldwell, Esquire
Registration No. 28,937

Woodcock Washburn Kurtz
Mackiewicz & Norris LLP
One Liberty Place - 46th Floor
Philadelphia PA 19103
Telephone: (215) 568-3100
Facsimile: (215) 568-3439

© 1997 WWKMN

DOCKET NO. REFH-0112

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c).

Docket Number	REFH-0112	Type a plus sign (+) inside this box→	+

INVENTOR(S)/APPLICANT(S)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)	CITIZENSHIP (if known)
Herz Labys Parkes	Frederick Paul David		Warrington, Pennsylvania Carrboro, North Carolina Philadelphia, Pennsylvania	

TITLE OF THE INVENTION (280 characters max)

SECURE DATA INTERCHANGE

CORRESPONDENCE ADDRESS

Attorney Name: Michael P. Dunnam, Esquire
WOODCOCK WASHBURN KURTZ
MACKIEWICZ & NORRIS LLP
One Liberty Place - 46th Floor
Philadelphia

STATE	PA	ZIP CODE	19103	COUNTRY	U.S.A.

ENCLOSED APPLICATION PARTS *(check all that apply)*

X	Specification *Number of Pages* 150	_	Small Entity Statement
X	Drawing(s) *Number of Sheets* 12	X	Other *(specify)* Return postcard

METHOD OF PAYMENT *(check one)*

X	A check or money order is enclosed to cover the Provisional filing fee	PROVISIONAL FILING FEE AMOUNT ($)	$150.00
_	The Commissioner is hereby authorized to charge filing fee and credit Deposit Account Number: **23-3050**		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

XX No.

_ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE _____

Date: October 27, 1999

TYPED or PRINTED NAME John W. Caldwell, Esquire

REGISTRATION NO. 28,937 (if appropriate)

_ Additional inventors are being named on separately numbered sheets attached hereto.

PROVISIONAL APPLICATION FILING ONLY

Date: October 22, 1999

Secure Data Interchange

Field of Invention

This invention relates to systems that provide personalized information, profiling, automated matchmaking and information exchange, within a framework that protects privacy and allows information collection and profiling within a carefully controlled environment. Secure Data Interchange addresses the direct conflict between the goal of personalization and the need for privacy, preventing the exchange and collection of information without knowledge and consent. The system provides the technical structure for an advance in the current state-of-the-art, supporting controlled collection and exchange of information between, and relating to, autonomous (and possibly self-interested) agents within a distributed environment. Agents may represent any party with individual goals, autonomy of control, and a persistent identity. We describe a system that combines four key techniques: distributed management and protection of multiple identities, controlled release of information to protect privacy while allowing other parties to leverage data and enhance services, distributed personalization and/or customization without the release of personal information, and protected-identity data mining techniques. Detailed transactional and profile information about each agent within the system is maintained on distributed trusted computer systems, that are managed and controlled by the agents themselves.

GOOGLE-ADS-0000105

Problem

The invention of Secure Data Interchange relates to a wide-range of problems, from a number of different domains. We outline a number of problems below:

- User-to-Vendor Electronic Commerce.

  The recent explosion of electronic commerce, in particular Internet-based individual-to-business electronic commerce, presents new opportunities for automated personalized information delivery and the automated customization of products and services. This type of personalization is very valuable to vendors because it can increase sales volumes, and enable cross-selling and up-selling of goods and services, and even support price discrimination. Personalization is also useful to customers when it correctly identifies the requirements and preferences of a customer, because it can reduce search cost and enhance the "shopping experience". Perhaps a customer can find the good or service (i.e. desirable price/quality/feature tradeoff) that he/she wants more quickly than without personalization, or receive information about an interesting new product or service that he/she did not know about.

The basis for these new services is that Internet-based "shop fronts" can be individualized on a per-customer basis, dynamically and in real-time. Traditional high-street shops must offer the same store layout to every customer, because the layout is physical, although some level of personalized service can be achieved through well-trained sales assistants, that act as a "guide" for a customer within a store. On-line "shop fronts" are virtual, and configurable at negligible cost to the customer or the vendor, assuming that computation is cheap and fast.

Furthermore, Internet-based electronic commerce can allow business to collect vast amounts of consumer information, because customers interact through a computer-based interface. Customers can be monitored as they browse a Web site for products and services. Information such as the search-terms that users enter into a search engine, the links that users follow, and the length of time spent on each page, can all provide an insight into the current goal of a customer, i.e. the type of product that he/she wants. When combined across different sessions, and with similar information about the browsing and purchasing habits of other customers, the information can be folded into a long-term view of the preferences and needs of a customer.

Moreover, new network connectivity enables different vendors to exchange profiles for common customers, either statically or dynamically, in order to build broad and detailed profiles across vendor domains. There exist many potentially powerful synergies between the data sets that are collected by different vendors, that can be leveraged to provide appropriate services and products to customers. When analyzed with the proper statistical tools these data sets can reveal fundamental patterns in the behavior of users, and enable a vendor to provide appropriate information to a user. Furthermore, access to user-profiles collected by other vendors can enable vendors to provide focused information delivery to first-time users, and also cross-market services with other appropriate vendors.

Vendors would find benefit in sharing data with other vendors; this would deepen their understanding of their customers' behaviors and preferences, especially if certain customers were traceable across several data sets. Users (individuals) would benefit from sharing data with other users (individuals). This is already evident in the popularity of news groups and web pages catering to individuals with shared interests. By learning what other people with similar tastes and preferences have discovered and enjoyed, a user can sidestep information overload in the search for personally satisfying information.

Vendors and users can benefit from each other. An obvious example would be in the use of collaborative filtering for the marketing of targeted promotions; rather than being deluged with coupons and advertisements that are of absolutely no interest, a user would benefit by being presented with advertising that is highly relevant. In the process, the vendor would increase advertising response rates, boosting overall efficiency. Users could also benefit from the personalization of content at vendors' web pages, and well-focused banner advertisements at other web sites that they visit.

The problem is that a user wants controlled personalization, in the sense that it might not be desirable for information about every on-line transaction that a user performs, every on-line document that a user reads, and every web page that a user visits, and demographic information, to be available to every business that the user interacts with, in the virtual and physical world.

- Business-to-business Electronic Commerce

The Internet provides businesses with network connectivity with other business, both competitors and partners. This connectivity allows businesses to exchange information about customers (dynamically or statically), in order to identify potential new customers, build better profiles for existing customers, and up-sell/cross-sell products and services in real-time. The problem with this exchange of information (that can include swaps, sells, and rental access) is that businesses need to (a) protect the privacy of their customers; (b) prevent information release to competitors, either directly or through third-parties.

- Privacy-protected Identification of Synergies/Matches

  There are many scenarios where autonomous parties, for example individuals and businesses, would like to identify synergies with other parties, in interests, goals, or business capabilities – but without revealing sensitive information, and without allowing a deluge of unsolicited messages from many other agents, including agents that are bad matches. The problem is to provide information that enables matches, without allowing bad matches and abuse of information – i.e. within a very carefully controlled environment.

- Automated Construction of Targeted Mailing Lists from Controlled User Profile Information

  Businesses often use mass-mailing of brochures/information as a technique to reach new customers. The problem with mailing lists, and bulk-mailing, is that it is an inefficient technique to solicit new business, because the advertised product or service will be inappropriate for the vast majority of recipients. The availability of vast amounts of detailed information about individuals allows more precise techniques to identify new customer prospects than has traditionally been available, for example static rules that use demographic information to rate individual prospects can be replaced by dynamic data mining techniques that leverage more detailed and complex data about individuals. This is only part of the solution, because individuals may still have strong preferences about the amount and type of information that they receive. A system that manages unsolicited information delivery must protect the privacy of individuals, while allowing other parties to identify good matches.

- Focused Banner Advertising/Content provision

  Internet-based media sites have followed preceding formats in generating revenue from advertising, with content to users often provided free-of-charge. The business model is similar to that in newspapers, magazines, and television, where circulation and audience/readership demographics are used to drive revenue. Electronic media presents new opportunities for media-based business: for example multimedia techniques and interactivity, personalized delivery of information, and personalized targeting of advertising.

  The problem – as before, is to acquire and leverage information about the preferences and interests of a user, within a system that protects user privacy (i.e. controls the collection and exchange of information about users, and controls the use that is made of that information). A further problem is to extrapolate information from a large corpus of data about an individual user.

- Credential-based Introductions, Contracting and Messaging-systems.

  There are many situations where individual parties, for example individuals or businesses, require introductions to credentialed individuals and/or businesses, with the aim of building a new relationship or making a new contract. Consider for example business associations, where credentials about non-bankruptcy, and no previous attempts to defraud could be important. Consider social introductions, where individuals might be concerned about past criminal activities of new contacts. In the domain of automobiles, we could consider a system that identifies other automobiles in the physical location of a vehicle that have recently been involved in an accident.

  The problem is to manage certificates within a system where users can maintain multiple identities, and to protect the release of certificates without suitable provisions for terms-of-use and criteria for request.

Solution

The above problems are solved, and a technical advance achieved, by the system of Secure Data Interchange. The Secure Data Interchange system enables information about bilateral and multilateral interactions between multiple persistent parties to be exchanged and leveraged within

an environment that uses a combination of techniques to control access to information, release of information, and matching of information back to parties. The system addresses the fundamental conflict that exists between rights of privacy and efficiency gains from better bilateral exchange of profile/preference information.

The system of Secure Data Interchange allows personalization and customization of information, products, and services, without the loss-of-privacy that can easily occur when vast amounts of information can be collected by, and exchange between, on-line vendors and businesses. An example of the current problem is "Cookie" technology, that enables multiple vendors to automatically exchange information about an individual without either the consent or knowledge of that individual.

The novel technical solutions used within SDI to achieve personalization and protect information about users, are:

(1) Controlled randomization of profile information.
(2) Consumer-side personalization instead of provider-side personalization. Instead of passing profile information to a provider and receiving personalized information in return, providers pass personalization methods and generic information to consumers, and consumers run the methods locally, using local profile information.
(3) Releasing information with metadata to control who can access it, and what it can be used for. The information is stored and exchanged within a secure data environment, with a system that controls the flow of information between different parties based on credentials.

Sensitive - Confidential

The system is based upon a core system of cryptographic-pseudonymity, and additional techniques to support that pseudonymity that are infrastructure dependent, for example HTTP-packet "stripping" on the Internet, to remove the IP address from originating packages.

Part of our solution is to allow autonomous parties to control and maintain detailed profile information locally, and carefully control the release of that information. When two parties interact they can do so under anonymous identities, persistent pseudonymous identities, or true identities. When combined with a technique to randomize information that could be used to identify an individual, this is a fundamental methodology for preventing exchange of personal information between third parties.

Another part of our solution is to allow the personalization/customization of services on the local computer system of the recipient of the personalized service, leveraging locally stored profile information. The provider supplies a description of the technique that should be used to personalize information, together with a complete description of its services and/or products. The recipient takes the description of the services/products, the technique to personalize information, and its own profile information, and performs the customization.

Furthermore, privacy-protected information about many parties can be collected and aggregated to construct new models of behavior with techniques such as collaborative filtering. We allow data about bilateral transactions to be aggregated (across multiple transactions of the same type and between the same parties), without revealing any information about individual parties.

We also allow detailed transaction-based information to be combined with broader information from other sources, but only as permitted by the parties to which the information relates.

When describing the application of SDI to electronic commerce we also introduce necessary ancillary systems, that are essential to supporting full e-commerce functionality within an identity-protected system (e.g. banking, mailing facilities).

1. Definitions.

In its most general form, the system of Secure Data Interchange is for a system of multiple autonomous parties, involved in multilateral communication. We refer to the parties as "agents" in general. We restrict the system

description to bilateral communication between agents, without loss of generality because any multilateral (multicast or broadcast) can be implemented as a set of bilateral communications. In any bilateral communication there are two parties, the sender and the receiver (and the parties can dynamically change over a communication session). However, in general, one of the agents needs to gain particular information, goods, or services from the other agent. We refer to this agent as the *consumer* agent. The other agent is the *provider* agent. Communication is not unidirectional, from provider to consumer, because the consumer must inform the provider about its needs and preferences. This information flow, from the consumer to the provider, is inefficient in many standard commerce domains--- but supported, for example with "customer satisfaction surveys". In electronic commerce a provider is more able to customize its interaction with a consumer, on the basis of information provided by the consumer--- for example, information that is implicit in its profile.

Provider. The providers in SDI are agents that provide other agents with information (e.g. news, current affairs, etc.), services (e.g. flight bookings), or products (e.g. computer products). Providers have identities, and can interact within SDI under anonymous, pseudonymous and true identities. The agents that represent providers within SDI are autonomous (cannot be manipulated directly by other agents or SDI), and follow actions consistent with their preferences, abilities and resources.

Consumer. The consumers in SDI are agents that need information, products, and/or services from other agents, in particular from provider agents. Like providers, consumers have identities, and can interact within SDI under anonymous, pseudonymous, and true identities. The agents that represent consumers within SDI are autonomous, and follow actions consistent with consumer preferences and goals.

SDI-0960-97-CON

Agent Computers. Consumers and providers are represented within SDI with dedicated computational resources, agent computers, that have permanent memory, computational resources, and network connectivity. For example, agent computers can be client machines (typical for consumers), or server machines (typical for providers), or machines can be portable devices, for example hand-held devices or other mobile devices. The agent computers can implement a subset of the functionality of the SDI system, for example the pseudonymity functionality, randomization, and consumer-side personalization. Agent-to-agent interaction does not need not be the sole method of interaction between agents, but the scope of the system of Secure Data Interchange is limited to electronic transactions between agent computers.

Certificates. Individuals might maintain certificates about their capabilities and past behavior, while vendors might register with a yellow pages service that identifies their business interests.

Identity: Agents interact under different identities, that can be anonymous, pseudonymous, or true-identities. With an **anonymous identity** the information provided by the agent, and the actions performed, cannot be traced back to the agent except from information provided by the agent that can be used to reveal its true identity (e.g. through revealing personal information). An anonymous identity can only be used to communicate with a single agent, only for a single session (i.e. until the anonymous agent terminates the interaction), and only through lines of communication designated by the agent. A **pseudonymous identity** allows other parties to communicate with the agent, and can either be active over multiple sessions (a **persistent** identity), or over a single session (a **temporary** identity). Similar to an anonymous identity, the main difference is that multiple agents can communicate with an agent under its pseudonym, and that it can be persistent. A pseudonymous identity still does not allow any connection to be made to the true identity of an agent, unless through information revealed (indirectly or directly) by the agent. An anonymous identity with every agent is equivalent to using a unique temporary pseudonymous identity with every agent. Finally, a **true identity** declares information that relates the agent to its identity in the physical world, for example the name of the vendor, or the social security number of an individual user. A true identity is by its nature persistent.

Individuals might transact with multiple vendors, and desire personalized interactions with each vendor without the vendors having the information necessary to exchange transaction information into a single profile. A vendor may be an information provider, or a provider of goods, and might not desire to interact anonymously or pseudonymously with other agents, because a vendor's identity can have significant value.

Profile: A profile is associated with each agent identity. A profile declares information that is relevant to the actions/preferences/goals of an agent, such as the transactions that it has performed with other agents, information that relates to the true identity of the agent – such as a salary range or education level for an individual, or a business

interest declaration by a vendor, information provided by the agent (such as its preferences for a particular type of product, etc.) and other information that has been compiled based on observing the behavior of the agent (e.g. physical location for a mobile user, such as a user in a vehicle, or trace of recent web pages visited for a user that is browsing the Internet.) In general, profiles contain any information that might be useful to another agent during future interactions. Profiles for agent identities can be maintained simultaneously by multiple agents, and they need not all contain the same information, or even agree.

SDI-Proxy: An SDI-proxy refers to the software that runs on top of an agent's computer device, and configures that device for Secure Data Interchange. The proxy is designed to act as an intermediary in all transactions between agents, ideally in a way that agents can migrate seamlessly towards an SDI-enabled system from current systems.

Identity-management Policy. An agent can define an identity-management policy, that automatically configures an identity for an agent when it initiates a new interaction with another agent. The policy will select an appropriate identity on the basis of the nature of the interaction and the information that the agent has about the other agent.

2. System Architecture

The basic architecture for SDI is a system of agents (with computational devices), connected via a network (Internet, wireless, or otherwise) to other agents. Every agent has an associated true identity and profile, and manages its interactions with other agents via the functionality provided by SDI.

Confidential - Draft v0.9

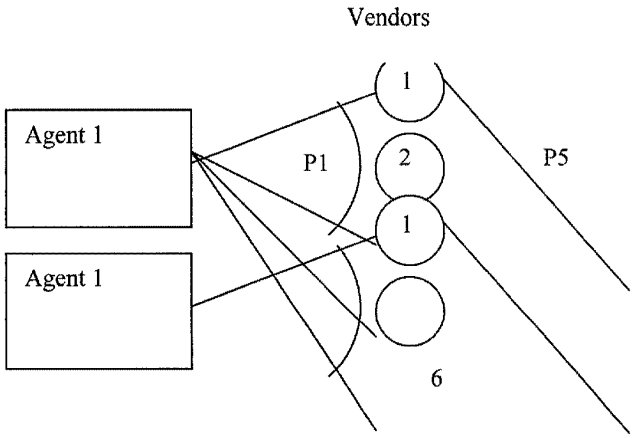
We assume in the following description that agents interact bilaterally with other agents, but there is nothing to prevent agents interacting multilaterally with other agents under pseudonymous or true identities. For the sake of simplicity, we can view such a multilateral interaction as a set of parallel bilateral interactions.

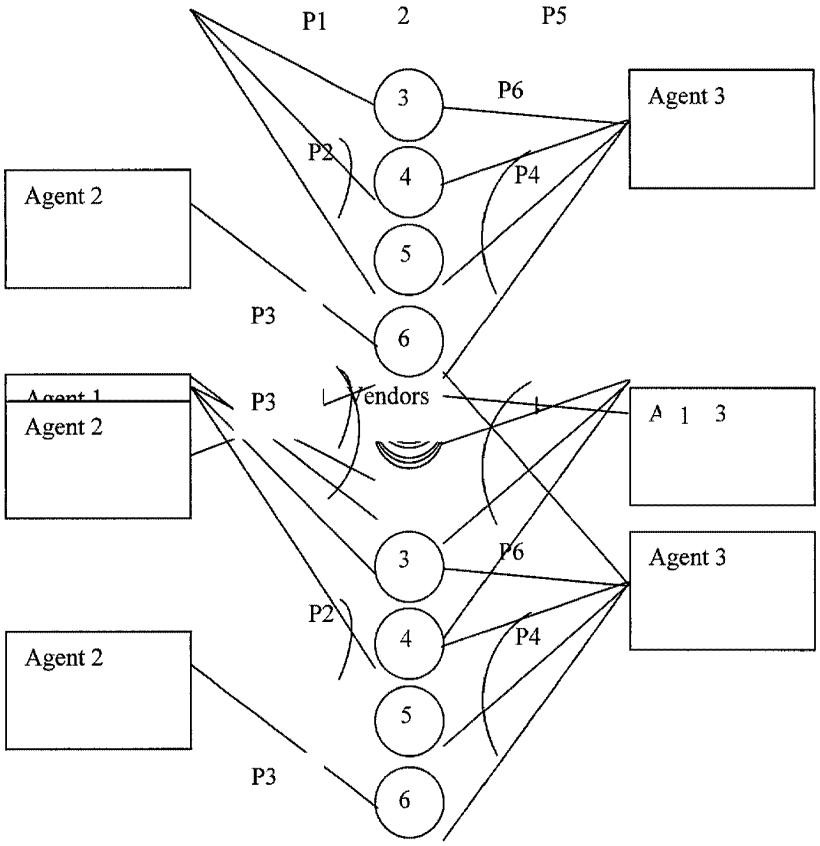
The system of SDI allows an agent to create new anonymous or pseudonymous identities on-the-fly, as it interacts with different agents. Agents can define an identity-management policy that is implemented on the agent computer, and determines appropriate identities for each interaction. At one extreme, with anonymous interactions with every other agent, agents can maintain complete control over profile information, because the only information that is available to another agent at the end of the session is information revealed to that agent during the session, and that information cannot be associated with the agent in the future because the agent used an anonymous identity.

Every time an agent initiates a new interaction, new information is available about the agent, that can be compiled by the agent's own computer device, and also by the computer device of the agent with which it interacts. Agents can exchange any information with any other agent, and privacy is protected by the information that agents reveal in the first place, and the identities that agents use.

Agent computer devices can release selected profile information to other agents. Agent computers can also process incoming information with rules (or programs) provided by another information and according to the local profile information that an agent maintains about itself. Selected information is prevented to the agent, that is personalized on the basis of the mobile code pushed to the agent, and the agent's profile.

The SDI architecture allows for third-parties that provide support for ancillary services, such as certificate management. In the basic system we assume that each consumer agent maintains persistent pseudonymous relationships with the other agents, under pseudonyms that are not necessarily unique. 1





The idea is that each agent has a pseudonymous relationship with a number of vendors, and uses the same pseudonym with some, depending on the agent. For example, agent 1 has a relationship under pseudonym P1 with vendors 1 and 3, and under P2 with vendors 4 and 5. Agent 2 has a relationship under pseudonym P3 with vendor 6. Agent 3 has a relationship under pseudonyms P5 and P6 with vendor 1 and 3, and pseudonym P4 with vendors 4, 5 and 6.

Given this identity management framework, agents are careful to reveal information that will not break the pseudonymity, but agents require personalization. Three key techniques are leveraged: (1) agents submit randomized and anonymous information to a cross-vendor database, to allow cross vendor attributes to be leveraged in customization; (2) vendors provide agents with personalization methods that can use cross vendor attributes to display appropriate information and product mixes; (3) agents release some information to vendors, randomized as necessary to protect their pseudonymity.

2. *Architectural Overview*

In this section we provide an overview of the invention, describing both its top-level architecture, and describing the key technologies that we use to support dynamic and powerful data synergies within a framework that explicitly protects the privacy of **users**. Figure 1 shows the top-level architecture.

The system of Secure Data Interchange is a technical solution to the problem presented by the conflicting goals of providing personalized information/products/solicitations to **users** (beneficial to **users** and vendors) and the rights/desires of individuals to privacy in transactions. In particular, a **user** might desire to control the vendors that have a lot of information about its preferences, and prevent certain vendors/entities from gaining information about particular transactions. A **user** might also like to be able to "keep control" of information that is potentially valuable to vendors and advertisers, and allow only restricted access to that information—and extract a payment for such access.

This is the **user**-centric model of SDI. Similarly, vendors would like to be able to exchange profile/market-research type information with other vendors, but not competitors, and require a secure system for protecting information that they make available.

The main data structure within SDI is the profile. A profile is a record of pertinent information about a **user**, for example demographic information, information about web pages browsed, content information from web pages browsed, recent on-line transactions, etc. Within SDI we associate profiles with an OWNER, and sign them with a ALLOWED-USE key to indicate what the owner has restricted the use of profile information to. The main technique used to support privacy within SDI is a pseudonym, that can fully represent a **user** in all transactions without providing any information about a **user**'s true identity.

Profiles may be located on client-machines, with **users**; at intermediate level SDI servers, for example at ISP servers; at vendor-level SDI servers; and also in a central SDI server. **Users** define privacy policies on their client machine, and the client machine ensures that all **user**-vendor interactions are consistent with a **user**'s policy, and that they cannot be abused— even by a malicious vendor. Client-level proxies can monitor **users** can build profiles (for example, as **users** browse and interact with vendors), and vendor-level proxies can monitor **users** that interact with the same pseudonym in multiple sessions. Both client-level proxies and vendor-level proxies can provide profiles to the central SDI server, with associated conditions of use. The central SDI server can aggregate and validate new profiling information that is received. Client-level and vendor-level proxies can periodically request profile updates.

Google - 4056356

2.1 Main Modules

Users interact with the system of Secure Data Interchange through clients, that are general purpose computers with memory and a connection to a network of other computers (clients and servers), for example via the Internet. The **User** represents an individual that is interested in performing multiple on-line and off-line transactions, with multiple vendors and other **users**, within the managed data interchange and privacy framework of SDI. Each **user** physically interacts with the Internet via his/her host machine. The host machine need not be the same machine for all sessions, and could include a **user**'s home PC, work PC, and also any portable devices that the **user** has configured for SDI. The host machine runs client software, which will include an Internet browser, and also the client-level SDI proxy server, or appropriate browser plug-ins to make the browser SDI-enabled. A host machine is in general a computer with a CPU, disk storage, a network connection, and main memory. We assume that the **user** trusts the host machine and client software.

Vendors interact with the system of Secure Data Interchange through servers, that are general purpose computers with memory and a connection to a network of other computers (clients and servers), for example via the Internet.

In addition to clients and servers, there are SDI modules that run on the machines, termed "client-level proxies" and "vendor-level proxies". The client-level proxy is a core component that interprets **user** messages and makes sure that all interactions with vendors satisfy a **user**'s privacy and data-use policies. The vendor-level proxy enables vendors to interact with other key SDI entities: client-level proxies and the central SDI database.

Another SDI module resides on a gateway between **user**'s and the Internet, to protect the pseudonymity of a **user** further. For example, when the **user**'s client machine resides on an intranet (eg that of an ISP) then there is an ISP-level proxy server at the gateway to the Internet that ensures no identifying information is provided to servers. The ISP-level proxy servers also support pseudonymous email addresses for **users**, and may maintain a database of profiles for **user** pseudonyms in their **user**-base.

The other main SDI module is a central SDI server, that maintains detailed records of profiles for every **user**'s pseudonym. Having all the information available for access at a unified level allows extensive data mining and collaborative filtering techniques to be applied, but still without violating **user** privacy and data use policies. In fact, the profile information can be physically distributed, for example located on the SDI-level proxies for each **user** base.

For example, it is possible to provide recommendations to a **user** under one of his/her pseudonyms through collaborative filtering techniques, even without any other information about the **user**—just from similar profiles from other **users**. Similarly, it is possible to recommend new prospects to vendors from the **user** base of other vendors, without actually providing vendors with profile information about prospective customers. Furthermore, the system of SDI can ensure that vendors can only provide impressions to **users** if it is allowed within a **user**'s privacy policy.

The function of the central SDI server is to make as much use of profile information that is collected (and authorized) by a **user** for a particular pseudonym, by analysis of the profile information. The goal is not to try to augment the profile with more data—for example by adding demographic information or purchasing transaction records from other vendors. This is impossible within SDI because pseudonyms cannot be associated with a real-world identity unless authorized by a **user**, and information provided in the profile for a pseudonym is carefully filtered to prevent **user** identification.

2.2 Overview of System

The client-level proxy manages all interactions that a **user** has with other **users** and vendors. Essentially, the proxy interprets a **user**'s messages to other **users** and vendors, and makes sure that a **user**'s privacy and data-**user** policies are followed. The proxy also maintains up-to-date profiles for **users**, and allows vendors to personalize information that they provide to **users** on the client machine, without receiving access to a **user**'s profile. The proxy is also authorized to automate authentication to vendor servers, and release of certain types of information.

A key mechanism used within Secure Data Interchange is that of pseudonymous interactions. This allows **users** to maintain long-term relationships with vendors, and release personal information to vendors, without any other party being able to use the information—is a system of "pseudonyms". Essentially, pseudonyms allow a **user** to separate its real-life identity from its identity with another **user**, or a vendor. The client-level proxy is also careful not to provide a vendor with any information that would compromise a **user**'s identity. Pseudonyms provide a very useful middle-ground between total anonymity and complete disclosure. Vendors can still keep useful records of their transactions with a single **user**, because the system maintains a persistent pseudonym for each **user** with the same vendor.

Users can however choose to interact with each vendor in the system under a different pseudonym, to protect their identity and prevent information transfer across vendors. The control of pseudonyms provides **users** with a method to control the exchange of profile information between vendors.

A pseudonym allows a **user** to maintain a number of persistent relationships with different vendors or groups of vendors, with complete assurance that the vendors cannot use the pseudonyms themselves to build a profile of the **user** from the **user**'s sessions across different vendors. A pseudonym provides: Identification, Authentication, Encryption, and Contact information. Pseudonyms are triples: (Pseudonym ID, Private Key, Pseudonym e-mail address).

Personal information collected through transactions across multiple vendors by the same **user**, under different pseudonyms is protected from transfer between vendors. Even when vendors themselves maintain local records of the transactions that they have performed with **users**, and response to queries that they have sent to **users**, there is no way of providing information in a form that another vendor can combine with its own information and use to personalize its service to the same **user**. This is because no two pseudonyms can be linked except at the (trusted) client machine. In addition, "randomized aggregates" are used when providing information to a vendor that might allow two vendors to link their pseudonyms. For example, the zip code, age, hair color, and profession of a **user** might all be useful information for a vendor—but providing all that information accurately to more than one vendor could allow the identity of a **user** to be compromised and vendors to exchange information.

Pseudonymous profiles are still useful in the aggregate, as part of a collaborative filtering system, even when noise has been added to some fields. The system of SDI has a centralized database of profiles and pseudonyms, that can be used for data mining and other collaborative filtering techniques. For example, suppose a **user** has a pseudonym that he/she uses to interact with all on-line book stores. Then the centralized database can perform collaborative filtering using the profile associated with that pseudonym and other profiles in the database to make personalized recommendations, for example on behalf a vendor that pays to receive such a service.

A vendor that belongs to SDI can use the profile associated with a **user**'s pseudonym to provide a personalized interaction with the **user**. In one version the profile is released to a vendor, and the vendor can push personalized information to the **user**. In another version, the vendor can push generic information to the **user**'s client, for personalization on the client. It remains important that the vendor does not learn the true identity of a **user** if the **user** is to prevent a vendor from providing information on the **user** to other vendors that know the true identity of the **user**. The **user**'s personal information is only safe when it is not possible to associate information with anything except for the pseudonym of a **user**, and only one vendor can interact with the pseudonym (through SDI).

We also allow the release of anonymous/pseudonymous profiles to vendors to allow a vendor to provide "in-house" collaborative filtering, without the ability to target any of the **users**. The pseudonyms are only useful within SDI. Furthermore, because a **user**'s communication channels with vendors is controlled within SDI, we can provide vendors with "rights" to solicit **users**, and rented (non-transferable) solicitation rights.

In a vendor-centric model of SDI we allow vendors to provide information (e.g. profiles about their **user**-base) to other vendors within a framework that prevents competitors from gaining a competitive advantage. This is possible because one role of the central SDI database is to act as a clearinghouse for profile information, that matches vendors and suggests data synergies. The central SDI database will not release profile information to competitors, according to rules that are designated by the submitting vendor.

Furthermore, even when a **user** designates that a group of vendors can exchange information on the basis of **user**-vendor transactions, one variation provides each vendor with a unique pseudonym. The only way that vendors can take advantage of profile information from **user** interactions with other vendors is via the central SDI database, that has a link between the pseudonyms that a **user** has with each vendor.

The system of SDI can track a **user**'s browsing behavior, by cooperation between the client and the ISP-level proxy. The client-level proxy releases sequences of URLs that a **user** browsers, associated with the pseudonym under-which a **user** surfs at any particular time.

Google Ex. 1029, p. 10 of 39

2.3 Ancillary Systems

SDI provides ancillary systems that are necessary to support a pseudonymous e-commerce system, for example independent certifying authorities, that are able to validate "once-in-a-lifetime" pseudonyms, that allow a **user** to prove that it only has one pseudonym for all interactions with a vendor. Certifying authorities can also issue certify properties of the **user** that owns a pseudonym, for example his/her credit worthiness, his/her age, his/her nationality— without requiring that the **user** identify themselves to a vendor or another **user**, and without breaking a link between pseudonym and real-world ID (even with the Certifying authority itself). The issued patent includes a description of anonymous payment mechanisms and a secure certificate management system, that we incorporate in this patent by reference. **Users** can gain new credentials through recognized certifying authorities under one pseudonym, and transfer the certificate to another pseudonym when transacting with another server.

Similarly, SDI supports a module that allows e-mail to be sent to pseudonyms.

SDI also places personal information (eg mailing address of **user**, credit card number, ...) etc with a trusted third party that has an agent-relationship with the **user**. When a vendor wants to (for example mail a physical item to a **user**) it receives certification from the TTP, and then provides that to another TTP with an agent-relationship with the **user** that mails the item to the mailing address, without the vendor itself ever receiving information about the mailing address.

Legacy/demographic information is integrated with pseudonymous profiles, even when SDI has no way to identify a **user**'s real-life identity with a profile, through client-side updates. SDI provides clients with new information, and clients update profile information for pseudonyms with randomized versions of the new information as they choose.

Finally, within SDI there is a method to allow **users** to receive compensation, in the form of rebates and electronic cash, in return from revealing personal information to vendors (even information that is slightly randomized and cannot be shared with other vendors).

The technique of blinded signatures is used extensively within our system to reduce the amount of information that third parties, and SDI, have about **users**. For example, the technique allows a **user** to create public/private key pairs without giving a certifying party knowledge about the key pair generated. No party within SDI is able to build a dossier that links the pseudonyms of **users** without explicit information provided by the **user**. The **user** has an absolute method to prevent vendors from generating combined **user profiles**. Blinded signatures are described by D Chaum (??), and incorporated here by reference. Originally developed for the purposes of anonymous digital cash, they readily extend to general certificates. It is not sufficient for a bank to sign a number, with a "$1" signature, because the bank knows the number it has signed—and can trace the cash. Digital signatures allow the bank to be "blinded" and then sign, and the recipient can remove the blinding factor from the signed number to receive a valid signed number. This ensures **user** privacy when spending digital cash.

2.4 Main Data Flows

TO CENTRAL SDI SERVER (not via ISP)

Figure 3 illustrates the basic data flows from the client-level SDI proxy and the vendor-level SDI proxy towards the central SDI data warehouse. The vendor-level SDI proxy only knows the pseudonym of a **user** when it executes sessions with a **user** and a particular pseudonym, along with what ever basic profile information and other explicit information the **user** provides the vendor's server with.

The client-level proxy server logs all web surfing activity that a **user** engages in under each pseudonym, but does not have all information about the interaction of each **user** within a session with a single vendor. The client-level proxy also knows the pseudonyms that are equivalent for a **user**, and can provide more integrated information than vendors.

FROM CENTRAL SDI SERVER (to client and vendor)

Figures 4 and 5 illustrate data flow from the SDI data warehouse server to the client-level and vendor-level proxy servers. The updated **user profiles** are periodically requested by client-level proxy servers (pull), to enable the **users** to maintain an up-to-date profile for each pseudonym that they operate. The update to the vendor-level SDI proxies occurs according to one of two main modes, as illustrated in Figure 6. A vendor may have sensitive information that it is not willing to release to the central SDI server. In this case, Figure 6 (b) the central SDI server releases profiling information to the vendor-level SDI proxy, where analysis is performed. In Figure 6 (a) the vendor releases all relevant information to the central SDI server, and receives updated profile information.

Figures 4 and 5 show the preferred mode of profile releases under our architecture. The Secure Data Interchange server releases pseudonym profiles to **users**, when verified requests are received from the proxy server that represents the pseudonyms. The SDI server also releases encrypted profiles to the proxy server, either during daily updates, or when a request for update is received from a proxy server. The proxy server maintains an encrypted profile for each pseudonym in its **user**-base. The profiles are encrypted with a private key of the SDI server. When the proxy server relays a message from a **user** to a vendor it also augments the message with this encrypted profile, so that the vendor can make use of profile information about the **user** if the vendor has purchased the analysis ability from SDI. SDI will supply analysis capabilities to vendors, that allow vendors to gain the benefit of analyzing profiles for new **users**, or **users** currently in their **user**-base, but without violating the privacy rights of **users**-for vendors can access only the results of the analysis, not the profile itself.

EP 0 674 06 A2

2.5 Underlying Cryptographic Infrastructure

All messages sent between **users** and vendors can be encrypted to prevent anyone other than the intended recipient from being able to read them. There are many technical solutions to this problem, including asymmetric public key/private key schemes such as the RSA encryption technique (although, **users** would need a unique key pair for each pseudonym). New "light" security protocols, such as using asymmetric key cryptography for initial validation, followed by shared key encryption are attractive—especially when the asymmetric key infrastructure is inefficient when **users** need a different key pair for each pseudonym anyway.
We use cryptographic techniques to "digitally" sign messages, in order to validate information contained within a message. A digital signature is computed through encryption with a private key, known only to the certifier, but the signature can be verified with the corresponding public key. This provides a recipient with a high degree of confidence that the message was indeed generated as claimed. An example technology for generating signatures, or "message digests", is MD5.

3. Identity Management

Identity management refers to the system that an agent uses to control its identity during interactions with other agents. The novelty of identity management within SDI is that it is pushed to agent computers, and does not require a third party. The core technology used to support identity management is cryptographic.

Identity management *complements* "profile management", and is necessary to support the profile-management functionality provided within SDI. Similarly, without careful profile management, identity management is redundant. For example: consider an agent that interacts with a unique pseudonymous identity with other agents, but does not use profile management--- and releases the same credit card number to every agent. The credit card number, or other similar information, is enough to "break" the pseudonymity of an agent. Similarly, consider an agent that interacts with a unique profile with other agents, but does not use identity management--- and informs all agents that it is Jason's agent. The other agents can "break" the profile management, and exchange profile information to form a portfolio of information about Jason.

Identity management allows agents to interact under different identities, depending on their privacy preferences and information that they have about other agents. SDI supports four types of identity: anonymous, temporary pseudonymous, permanent pseudonymous, and true-identity. Agent computer devices, enabled with an SDI proxy, determine an identity type dynamically, based on an agent-defined identity-management policy. The policy will select an identity based on information about the other agent in a transaction, and the nature of the interaction. Identity management, coupled with the ability to add noise to information provided, allows agents to control the ability of other agents to collect and exchange information. In the descriptions that follow we assume that agents are adding enough noise/using enough privacy-protection devices to support the privacy structure provided by the profile-management policies.

The ability to support personalized interactions between agents on the basis of each identity-management policy depends on the computational resources available to agents, and the cost structures/incentives for provision of information.

> Need to think about this carefully – what are the implications of choosing each type of policy. For example, compare an anonymous policy with a true identity policy, what are the differences? One difference is that a vendor might be able to leverage more advertising revenue etc. if it can augment profiling information about a user to the greatest extent possible. This might not be possible if the vendor has to rely on user-side processing of an advertisement selection policy.

We allow vendors to provide different services based on the identity that a user provides. The system of SDI **notifies** the other agent in an interaction about the type of identity (i.e. anonymous, pseudonymous) that is provided. For example, a vendor might provide a more complete service to a user that provides a persistent pseudonym than a user that provides an anonymous pseudonym because a persistent pseudonym might allow the vendor to support more profitable advertising techniques.

We outline some identity management policies abstractly, before suggesting concrete implementation techniques. These are different levels of **identity privacy:**

- Absolute Identity Privacy

At the highest level of privacy an agent interacts anonymously with every other agent so that agents cannot learn about the agent across sessions, and cannot personalize future interactions. Furthermore, information about the transaction with the agent, cannot violate a user's privacy when provided to other agents, so long as the agent does not release any identifying information during its transaction. In an Internet environment this mode of interaction provides similar functionality to the Cookie technology that allows vendors to maintain states while interacting with a user, for example tracking requests from the same user and providing relevant personalized content in response (such as items in the users "shopping basket").

- High Identity Privacy

At the next level of privacy, an agent interacts with every agent under a unique persistent pseudonymous identity. This prevents another agent exchanging information about the agent with other agents, unless the agent reveals other identifying information. However, this policy does allow agents to receive personalized information over multiple sessions with the same agent. Another agent can personalize information, but only on the basis of previous transactions with the agent. In an Internet environment this mode of interaction provides similar functionality to a desired side-effect of Cookie technology, that allows vendors to track customers across multiple sessions.

- Medium Identity Privacy

A possible identity-management policy to provide a medium level of privacy for agents allows agents to use the same persistent pseudonymous profile with groups of other agents, perhaps segmented according to the type of task that an agent is performing. For example, an agent might specify one identity for whenever it is purchasing books, so that any book seller agent that it approaches can access profile information that relates to the agent's previous book purchases and queries (so long as someone will provide/exchange/sell/rent that information to the agent). As another example, an agent might require that a different pseudonym is used for business-related transactions and leisure-related transactions, to allow more precise personalization. Finally, consider a family that interacts with other agents through the same SDI proxy. Each member of the family might choose to use a different pseudonymous persistent identity for all of his/her on-line transactions. With this level of privacy agents still control of the ability of other agents to build portfolios of information.

- Medium-Low Identity Privacy

This level of privacy might allow a user to interact with other agents under a single pseudonym once trust has been established, but anonymously until that stage has been reached.

- Low Identity Privacy

This level of privacy is the default in most current systems. The true identity of an agent is made available to most agents that it interacts with, for example via its email address or credit card number. At this level of privacy the system of Secure Data Interchange provides no control over user privacy, because we cannot prevent agents from exchanging information that relates to transactions they have performed with the same agent, and we cannot prevent agents from sending unsolicited communications to agents.

3.1 Supporting Infrastructure: Identity Management System

In this section we provide more details to enable one possible implementation of an identity management system.

Identity Management for Customers in a Customer-to-Vendor E-commerce domain. One mode of identity-management could adjust the type of identity selected for the agent (the individual) on the basis of the business-type of the vendor. This requires that vendors are certified within particular business domains, by a trusted third-party. The third party is responsible for auditing vendors. Let L1, L2, ... , LN denote different business types, such that each vendor is associated with a set of labels to identify its business interests, for example vendor 1 might sell

books and compact disks, and be characterized as L(V1) = {L1, L2} where L1 is for printed material, and L2 is for music.

Given this, an identity-management policy is a mapping from vendor business labels to agent identity. Possible ways to construct the mapping include a rule-based system, where rules may fire in parallel, and have associated priorities. We could include rules such as:

A. Use a unique persistent pseudonym for every vendor with at least one label from the set L'

B. Use the same persistent pseudonym for every vendor with at least one label from the set L''

C. Use an anonymous pseudonym for every vendor that sells credit cards.

The user needs to define sets of labels, and identity management actions for each rule. The rule-based engine can be used to **verify** that an action is defined for every vendor type, using straightforward rule-validation techniques, such as consistency and completeness checking.

Sony Exhibit 1005 - Page 14 of 95

3.2 Example Implementation Technique for Pseudonymous and Anonymous Identities.

This section will describe the cryptographic techniques that can be used to implement pseudonymous and anonymous identities.

3.3 Certificate Management

This section will discuss certificate management as it relates to identity management. Certificates are associated with an agent, and need to be transferable to each agent identity. They also need to be granted without revealing any information about an agent's portfolio of identities to a certificate management agency.

3.4 Supporting Anonymous and Pseudonymous Identities within an Internet based system.

In this section we discuss the necessary infrastructure to support anonymous and pseudonymous identities within an Internet-based SDI system. For example, we require that the HTTP messages are "flushed" of any ISP information, that messages are routed through a common gateway, etc.

3.5 Supporting Anonymous and Pseudonymous Identities within an E-commerce system.

Secure Data Interchange applied to electronic commerce requires careful consideration of the functions that are required within e-commerce, for example arrangements for anonymous payment and anonymous mailing of goods. We describe possible ways to implement these features in this section.

4. Randomized Aggregates

The method of randomized aggregates is a novel technique to support profile management, and support data privacy. The technique allows agents to release information that is related to an interaction with another agent, without allowing that agent to pass the information to another agent, that might use it for another purpose.
In particular, within a pseudonymous interaction, the information cannot be used to break an agent's pseudonymity because it is slightly different from other information provided (under a different pseudonym) to other agents, and/or sufficiently similar to other information provided by different agents. However, the information is only adjusted by as much as necessary to protect the data-privacy of an agent, and hopefully not enough to prevent it having value in personalization.

Agents need to be careful about the content of information that they release to other agents. For example, if agents are keen to protect their true identity when interacting anonymously or pseudonymously with other agents, they should not

reveal any personal information, such as company street location, personal email address, date of birth, credit card number, etc.

We can describe a number of high-level **data privacy** policies that can be supported with randomized aggregates:

- High Data Privacy
  At this level, agents release no profile information (e.g. demographic information, mailing address, etc.) and the only information that is available to another agent is the information that is provided from the transaction itself.

- Medium Data Privacy
  Provide a certain amount of randomized profile information to another agent in the process of a transaction, when that information might help to make the transaction more efficient (for example, allowing a vendor to suggest more suitable products to a customer). Perhaps provide other information in a non-randomized form, when it does not identify the agent.

- Low Data Privacy
  Be careful not to provide identifying information, such as street address, or name. However, only randomize other information that is very sensitive, such as the exact salary earned in the previous year, car make and year.

With randomized aggregates it is possible to protect the privacy of agents with anonymous and pseudonymous identities and allow personalization and leverage of large collections of data. The idea is to add **perturbation** to make it very unlikely that a user's identity can be discovered, but still possible to use the data (a) to provide a personalized service to an agent; (b) within a data mining application that looks for trends and patterns within a large corpus of information.

0016450-0025950

Randomized aggregates provide a cheaper and more secure alternative to cryptographic techniques, such as secure-function-evaluation for providing information without compromising privacy. Furthermore, randomized data is secure to computational attacks and the loss or theft of private keys-because we degrade the data, and make access to any one data item virtually useless: it is an information theoretic technique.

Randomized data from a system of agents can be aggregated to obtain the average value accurately, as the number of agents increases, for noise perturbations with zero mean. However, information about an information remains noisy. Traditional cryptographic methods are capable of solving the problem of revealing only aggregate information while concealing individual information. Methods exist for computing aggregates or other values from encrypted information without first decrypting this information. (Such methods are dealt with in an area of cryptography known as secure function evaluation.) However, the general-purpose nature of these methods makes them unnecessarily cumbersome for the limited problem at hand. In particular, the communication and computation requirements of these methods when applied to the problem of aggregation result in an unacceptable overhead on the system.

There is also an additional problem with such cryptographic techniques. Secret information can be compromised by successful computational attacks on the cryptographic scheme or by the loss/theft of private keys. Such problems are present in all uses of cryptography. Nevertheless cryptography is used where it is the best alternative. However, once again, for the limited problem at hand our solution is safe both from computational attacks as well as from private keys being compromised.

Transactional information about agent sessions, and fields of records completed, search queries, etc. are also modified by the addition of a "noise" term. If the field is numeric and continuous then an additive noise term can be chosen, sampled for example from a Normal distribution. If the data is discrete then a new discrete element is randomly selected in a close neighborhood of the original

point. With multiattribute data, different, independently chosen noise terms are used for each field of a record that needs to be perturbed by the addition of noise.

The amount of noise is computed by the submitting agent to provide an appropriate degree of confidence that an adversary with exact information about its attributes cannot identify the individual. For example, in a Normal distribution, the standard deviation that is required depends on the number of agents that have provided similar information, and how sparse the information is in its domain. The standard deviation is "tuned" to provide the appropriate level of privacy. We do not require that every agent uses the same noise distributions. The choice can be made autonomously.

Adding noise to discrete values is a little more complicated, because the new value must remain feasible. For example, in randomizing the name of a CD the new name must be the name of another CD, not some "made up" name. We need to add noise to make data elements "close" to the accurate values. With discrete data, such as the name of an artist, "close" must be defined within the correct metric. The appropriate metric is such that a "close" value shares many of the same characteristics. For example, it is not appropriate to assign a close value on the basis of a shared last letter in the first name, but it is appropriate to assign a close value on the basis of an artist from the same genre of music --- from a "semantic cluster".

We can still perform correlation across fields with randomization, so long as the randomization does not destroy any trends between fields. Randomized data is marked as such within SDI, and labeled with the degree of degradation, so that SDI can be aware of the number of records to get relevant accuracy levels, and can report accuracy to customers. There is a tradeoff between the level of privacy protection and the level of aggregation at which responses to queries become accurate.

HP 2004-0960

4.1 Technical Details

As a simple but illustrative example, consider a consumer agent C_1 with profile information x, where x is a continuous real number. The agent interacts with two different provider agents, P_1 and P_2, and desires to interact pseduonymously with the agents, so that P_1 does not know that the agent also interacts with P_2, and vice-versa. The problem is to release randomized information about x to P_1 and P_2 so that the provider agents cannot determine that the consumer agent is the same agent.

We will follow an adversary argument, and assume that somehow agent P_2 knows the *true* value of x for agent P_1. Given this assumption, the challenge is to select a randomized x' from a distribution f(x), based on the true value x, so that it is unlikely that P_2 and P_1 can break the identity of C_1. Consumer agent C_1 will announce its randomized value x', and also the distribution f(x) that was used to generate x'. In a later section we disclose a method to allow verification that x' is generated from the distribution.

The problem is to select f(x), such that $p(C\_1, x' \mid x, f(x)) < \text{epsilon}$, where epsilon is a small positive number that is agent C_1's desired privacy level. For example, with epsilon = 0.01, the probability that the value x' is submitted by agent C_1, given that the agent has true value x and uses distribution f(x'), is no greater than 0.01.

A distribution f(x') and a randomized value x' generates a distribution over true value x. For example, if the distribution is f(x') = N(x, 1), i.e. with mean x and standard deviation 1, and the new value x' = 1.2, then the distribution of values x that generate x', given f(x'), is f(x) = N(1.2, 1).

We perform strong (pessimistic analysis). Assume an adversary that knows that an agent in the set of agents that have submitted information to agent P_1 has true value x. Each agent submits a randomized data point to agent P_1, with along with the distribution that was used to generate the randomized value. This information allows the probability that a particular agent has true value x, given that some agent has the true value, and that agents have submitted randomized values, to be computed. The goal is for a new agent, agent A_1, given information about the current data submitted to agent P_1, to decide on an appropriate level of randomization to use to protect itself.

Each agent submits two pieces of information to agent P_1, denoted [x', f()], where x' is the randomized value of x for the agent, and f() is the distribution that was used to generate the randomized value.

Suppose the adversary knows that one of the agents that has submitted information to P_1 has true value x. We can use Bayes rules to compute the probability that the agent that submitted data [x', f()] has true value x, given the data submitted by other agents. Consider agent C_1, that has submitted data to P_1, then Pr(C_1 | x) = Pr(C_1) Pr(x | C_1) / Pr(x), where Pr(C_1) is the prior probability that a data point was submitted by agent C_1, Pr(x | C_1) is the probability that the agent has true value x, and Pr(x) is the probability that one of the agents has true value x. The probability Pr(C_1) is computed trivially, as 1/N, where N agents have submitted data. The probability Pr(x | C_1) can be computed from the data submitted by agent 1, x' and f(), and the probability Pr(x) is computed as the sum of Pr(C_i) Pr(x | C_i) over all agents.

Given this analysis, for which we present an example below, the submitting agent C_1 can decide on f() on the basis of a worst-case analysis. Suppose that an adversary knows the true value of its x, then choose f() such that in the worst case randomized value generated, Pr(C_1 | x) < epsilon, where epsilon is some small positive number that represents the degree of protection that the agent requires.

The worst-case outcome occurs when Pr(x | C_1) is maximized. Assuming that the distribution f() places most probability on the outcome x' = x, then this occurs when x' = x. Therefore we propose the following procedure to select an appropriate level of noise to add to a data point:

1. Choose a distribution family (e.g. Normal, Uniform, etc.), and let Par denote the parameters that define a specific distribution.
2. Choose a level of protection, epsilon, where 0 < epsilon < 1.
3. Request Pr(x) from the provider agent P_1, the probability that a current agent has the true value x.
4. Compute p(Par) = max_{x'} Pr(x | x'), where x' is generated from the distribution. I.e. find the worst-case outcome of the randomization, typically when the randomized value remains equal to the true value. The value of the worst-case probability that the agent has true value x given distribution f() and randomized value x' is parameterized in terms of the parameters of the distribution, Par.
5. Select parameters Par such that the Pr(C_1 | x) is below epsilon, even in the worst case.

Notice that we assume, in step (3), that the agent can request the probability that some agent has true value x, given the current information that has been submitted to the provider P_1. Clearly, this must be done using a privacy-protected protocol, because the agent does not want to reveal x to P_1. Before working through a numeric example, we present here a number of ways to determine Pr(x):

a) The provider can maintain publicly available aggregate information, about the probability that an agent in its population has true value x, for a range of different x values. An agent that wants to submit information, can now lookup Pr(x) anonymously in a table.

b) Allow anonymous querying of the provider, from a trusted third party that aggregates queries from many users, and time-delays them, so that the provider cannot draw any inferences.

We also require that the provider maintains the data information provided by consumers. A trusted third-party can be used to verify provider practices.

Confidential - 4025695

Numeric Example

a) Every agent submits information from a uniform noise distribution, centered around its true value. The decision variable is the range of the uniform distribution.

Suppose agent 1 submits data point [4, 6]. This represents that it has a randomized value x_1 = 5, and used a uniform distribution U(-1, +1) to generate its randomized value. Similarly, suppose agent 2 submits data point [4,6], agent 3 submits [3,5], agent 4 submits [3,6] and agent 5 submits [2,4].

Given that an adversary knows that agent X has true value 2.5, and is a member of the database. The adversary can be certain that agent 5 is agent X because Pr(A_5 | x = 2.5) = Pr(A_5) Pr(x = 2.5 | [2,4]) / Pr(x = 2.5) = 0.2 * (1/2) / (0.2 * 1/2) = 1. In comparison, if the adversary knows that agent Y has true value 5, then Pr(A_1 | x = 5) = 0.2 * (1/2)/ (0.2 * ½ + 0.2 * ½ + 0.2 * 1/3) = 0.375, and similarly for agent A_2. The probability that agent Y is agent A_4 is 0.25.

Now, a new agent, agent Z wants to inform the provider about its value x_Z. The agent has x = 5, and wants to be sure that an adversary that knows that agent Z has x =5, cannot determine its identity from its randomized value with greater probability than epsilon = 0.1. The agent Z needs to decide on a range for its uniform distribution, d_Z. Given d_z, and labeling the agent A_6 in the database, the probability Pr(A_6 | x = 5), i.e. the probability that agent A_6 has true value x = 5, given its randomized value is the same whatever the randomized value is because the distribution is uniform, and depends on d_z but not on x'. Pr(A_6 | x = 5) = 1/6 * (1/d_z) / (1/6 * 1/2 + 1/6 * 1/2 + 1/6 * 1/3 + 1/6 * 1/d_z) = 1/ (4/3 * d_z + 1). This is less than epsilon for d_z > 6.75. Therefore, so long as agent Z generates its randomized value from a uniform distribution centered on x with range 7, then it can be sure than an adversary cannot determine its identity with probability greater than 0.1.

b) Every agent generates noise from a uniform distribution, centered on the true value. Let sd_i denote the standard deviation selected by agent C_i for its noise distribution. The analysis is slightly more complicated, because the Pr(x | A_i) is now a function of the position of x within the distribution, not just whether it is in range, as was the case with uniform

distributions. Agent Z chooses the standard deviation for its Normal distribution by assuming that the randomized value x' = x, because in this case Pr(A_6 | x) is maximized, because Pr(x | A_6) is maximized, and the other terms are invariant. This is the worst-case outcome, placing the maximum probability density on the agent's true value x.

The previous analysis assumes that an adversary knows only a single piece of information about one of the agents, and that the agents have continuous valued data. Allow agents to select the distributions and parameters they use. Also, there is clearly a bootstrapping problem. Also, notice that the rule has the right behavior—the more data present in the system then the more accuracy an agent can use to submit its own information, for the same privacy level epsilon. How do we do the bootstrapping of the database? Also, notice that as agents continue to submit information agents already in the system get more secure--- need to guarantee that information is never removed from the system.

Variations

1. Submitting approximate information

The first variation that we consider is an alternative technique to randomize information. Instead of adding noise, that maps each data value to a new value, a provider can provide a number of prototype values--- that are sufficiently broad to allow agents to retain anonymity. Agents truthfully reveal the prototype value (or bucket) into which their true value lies, and the prototype values must be selected broad enough by the provider that there are enough agents with each value.

The previous analysis carries forward, but is simpler. In this case, an adversary that knows that an agent has true value x, first finds the prototype that corresponds to the value. The probability that an agent has the true value is the same for all agents that have posted values with the prototype value, and equal to 1/N_p where N_p is the number of agents with the prototype value. The decision for agents about submitting information is not simplified: if (1/N_p) < epsilon for the prototype that the agent's value maps into, then submit the data (as the prototype value), otherwise, submit no data.

In order to support agents with different levels of privacy-sensitivity (different epsilon levels), then a provider can allow agents to select a degree of accuracy when submitting information--- providing nested buckets. For example, a consumer agent might choose to specify that its true value can lie in one of a number of buckets, or equivalently in a single super bucket. The number of agents present in a value is counted as the number of agents that specify that value, and a weighted count for agents that specify more than one value, depending on how many values they specify.

Example. Suppose that a consumer agent decides whether or not to reveal its salary, on the basis of information that has been revealed by other agents. Suppose the provider allows agents to state that their salary is in bucket <$20000, $20000-40000, $40000-100000, >$100000, and that the number of agents with each prototype value is currently 10, 12, 16, 5. Now, suppose a new agent, with salary $200000 wants to provide this information, and has a privacy threshold epsilon = 0.1. The prototype values prevent an agent from stating the high value (> $100000) because there are less than 10 agents in the bucket. However, if the provider also allows agents an option, for example

"my value is in bucket 40,000 – 100,000 or >100000", then the agent can state that information safely. In this case, the probability that the agent is an agent with salary $200,000 is computed as its weight in the top valued bucket, divided by the total weight, i.e. 0.5 / 5.5 < 0.1

The main difference between this technique and the previous technique is that it is simpler to implement, because probability distributions must neither be sent to the provider, or maintained by the provider. However, the information is represented with a static accuracy, that is hard to improve as the amount of data submitted increases. This occurs automatically in the standard version.

2. Handling Discrete Data

The problem with discrete data is that *continuous* noise cannot be added to a data value, because this is not well defined. Instead, an agent must map a discrete value to one of a set of "close" discrete values. The original technique works in approximately the same way, except that agents use discrete probability distributions to randomize data points. It is also necessary to define what "close" is in the space of discrete values.

For example, suppose that an agent wishes to reveal the name of the artist that recorded the last compact disk that a consumer purchased. An artist's name is best viewed as a point in "artist space", and therefore as a discrete value. It does not make sense to change a random set of letters, because the new "name" will not be the name of a valid artist. Instead, the concept of noise is to randomly choose a new name close to the current name, i.e. choose a new feasible location in artist space. One reasonable solution for the names of recording artists is to define a neighborhood of artists that are close to the original artist, and select a new artist from within the neighborhood with equal probability. The metric that defines how close artists must be computed using a system that is common knowledge to the provider of information and the consumer agent, because the provider of information must be able to compute the distribution over true artist name, given a randomized name and probability distribution. We can define the probability distribution with a *threshold* in the distance d(name1, name2) metric, and the tag 'uniform' to indicate that each artist within the neighborhood was assigned an equal probability weight. A metric for artist names can be computed using clustering techniques, where artists are clustered according to the buying habits of consumers that purchase them.

Semantic clustering that enables useful randomization of discrete field can be automated when goods are frequent purchase, high volume goods-where individuals purchase goods on multiple occasions, and more than one of the family of goods on a single occasion. However, it is possible that high price, low volume goods, should be randomized on the basis of expert analysis (for example new cars, computers...) -- where an expert can extract key features of a purchase, and represent the purchase generically using either a single prototype good, or one of a set of approximately equivalent goods.

Another approach to discrete data is to apply the "approximation" technique, that allows agents to truthfully reveal approximate information about information. For example, the provider can map

artists to "clusters" of similar artists, such as "trance" or "hard house", and allow a submitting agent to state the cluster of the artist for the CD that it just purchased, instead of the name of the artist.

3. Selecting an appropriate level of randomization when agents submit bundles of information.

Another important variation is for agents that submit more than one piece of information to a provider. We refer to this as a "bundle" of information. Intuitively, when agents submit a bundle of information they open up more possibilities for adversaries to determine their identity, because a bundle of *independent* information reveals more about a user than a single piece of information.

The technique is essentially the same--- a submitting agent needs to know the current probability that an agent in the database has true information equal to its profile. For example, consider an agent that has already submitted approximate information to a provider about its home Zip code, annual salary, and model/year of car. The agent needs to decide on an appropriate level of randomization for new information about its profession. The required level of randomization depends on the number of other agents in the system that could have the same salary, Zip code, car, and profession, given the randomized information that they have submitted.

The problem is more difficult because: (a) agents might not all submit the same bundles of information, i.e. some agents might provide more information than others; and (b) each piece of information might be dependent on other pieces. The task is to *compute* the probability that another agent has bundle B .

Here is the framework for a reasonable solution:

First, if agents submit a randomized bundle of information, each piece of information is randomized independently of the other pieces. For example, if agent 1 has provided information about its value for features f_a and f_b, as x'_1,a and x'_1,b, then x'_1,a was generated by applying a random perturbation to its true value x_1,a for feature a and x'_1,b was generated by applying an independent random perturbation to its true value for feature b. This allows the probability p(x | agent 1) to be computed for values x of features a and b as the product of the two marginal random distributions.

Second, to compute the probability of (x_a, x_b) for an agent that has only provided information about one of the features, assume that the value for feature b is independent of the value for feature a, and multiply by the *a priori* probability of x_b, over all agents that have provided information about their value for feature b.

Finally, an agent submitting information about the pair of features x_a and x_b can determine the maximum amount of probability weight that it can apply to its true values for the features and still provide enough protection against an adversary with true information about both values.

Example. Consider feature a, 'year of birth', and feature b 'car model'. Suppose that agent 1 has submitted car model [VW Beatle, 0.1] to indicate that it selected a random car model from within a proximity threshold of 0.1; agent 2 has submitted car mode [Mercury Mistique, 0.05] and year of birth [1962 1968] to indicate that the year is somewhere in that range; and agent 3 that has submitted year of birth [1965 1975]. Now, agent Z, with true information (VW Beatle, 1972), computes a randomized data entry as agent 4, based on a privacy threshold of epsilon = 0.4 for an

adversary with correct information about agent Z. Assuming that the Mistique is not within a threshold of 0.05 of the Beatle, agent 2 has placed on probability on its true car model being the Beetle. Also, assume that the proximity threshold of 0.1 on the Beetle implies that agent 1 has a Beetle car with Pr(Beetle | Agent_1) = 0.2. Finally, the probability that an agent over the entire population has a Beetle is 0.1. Furthermore, the probability that an agent in the population was born in 1972 is 1/(7 + 11) = 1/18. Therefore, agent Z must place probability P_z on (Beetle, 1972) such that P_z / (1/18 * 0.2) < 0.4, i.e. P_z < 1/225. To support this, the agent can randomize each data point with probability (1/225)^(1/2) = 1/15, i.e. randomize 1972 with a uniform distribution of years [-7, +7], and select a threshold on cars greater than 0.1 to give a smaller probability than the 0.2 achieved by agent 1 with a proximity of 0.1.

4. Validating that a Query is Safe—A Technique for Bootstrapping

The problem with the system as described is that it has a bootstrapping problem. Initially agents will not provide any information to a database, because an adversary can determine their identity with certainty. For example, if an agent with age 18 is deciding a level of randomization, and there is currently a handful of ages in the database: { [24, 28], [34, 38], [42, 46], [42, 44] } then the agent cannot submit a randomized value and distribution, because the range of implied true ages will always include '18', and be the only approximate information to include '18'. The agent cannot provide its age to the database.

We propose a *validation trick* as a solution to this problem. An agent can also define its *privacy level* epsilon when it submits information, for example ([16, 20], 0.2), where [16, 20] is the range of possibilities for an agent's true age, and 0.2 is the desired privacy level. Given an adversary that makes the query "What is the probability that agent 1 has true age 18", then the system will only report that agent 1 might have true age 18 if the probability is *less than 0.2*, because 0.2 is the agent's desired privacy level. Therefore, the agent can submit the information, and be certain that the information will not be used to break its identity. Of course, this requires that the submitting agent trusts the provider agent that requests information and stores data points.

A submitting agent can choose a suitable level of randomization, and state its privacy level. A lot of randomization will allow the agent's data to be provided to the database more quickly, but the agent might prefer to wait longer and provide more accurate information--- for example if it is paid according to the accuracy of the information, or receives a better quality of personalization for more accurate information.

A similar solution can use a trusted third party. The trusted third party acts as a temporary "holding station", that can be used to boot strap the system. The third party collects information from agents, and when all of the information, or a subset of the information is "safe", it can submit the data to the provider agent. The measure of "safe" is defined by the information in the database and the privacy level specified by each submitting agent. When the data is finally submitted it does not need to include any associated privacy level, and the data can be used to respond to arbitrary queries without compromising the privacy of an agent.

4.2 Data Mining From Noisy Data

Suppose that we want to determine the average income level of people purchasing camcorders. A provider agent has collected information from agents about recent purchases of electronic goods. An agent that submits information can provide randomized information about the electronic good it purchased, and about its salary. The provider agent now wants to compute the average salary for agents that have recently purchased a camcorder. With enough information points the provider can compute an accurate value for the average salary for agents that state they have purchased a camcorder. The agents might have purchased another electronic good, depending on the degree of randomization that they use when submitting data. However, as the data in the system increases agents will be able to submit more accurate information with the same degree of privacy protection.

As another unrealistic but illuminating example, consider the following scenario: Suppose an organization receives monetary contributions from a number of individuals, who for reasons of privacy, wish to conceal the amount of their contributions. Suppose that the total of all contributions is to be public knowledge. The randomization scheme works as follows: When an individual contributes x dollars, she chooses a random number r according to a probability distribution f( ) (with mean 0) and sends x+r and f( ) to the aggregate database. By allowing widely dispersed values of r, we can ensure that an eavesdropper who obtains the value of x+r has very little information about x. However, if a large number of individuals register their contributions in this manner, then the total of the values sent to the aggregate database will be very close to the true total contribution.

We can still perform correlation across fields with randomized data, so long as the randomization does not destroy any trends between fields. Randomized data is marked as such within SDI, and labeled with the degree of degradation, so that SDI can be aware of the number of records to get relevant accuracy levels. There is a tradeoff between the level of privacy protection and the level of aggregation at which responses to queries become accurate.

Randomized data can be useful to build models about consumer behavior, so long as there is enough information to support accurate modeling despite the level of randomization. The value of data to a requesting vendor can be estimated by:

1) the number of users matching the relevancy criteria.
(2) the degree of statistical confidence associated with it.
(3) the effective measurable degree of benefit resulting from previously disclosed similar data (if applicable, with similar statistical confidence) as to that of the present data to be disclosed.

4.3 Verifying Noise Levels

We can use a cryptographic technique to verify the distribution of noise that is added to data - and also to enable replay. A vendor must keep a record of the non-randomized data that is supplied. When generating a random perturbation, the vendor uses a one-way function f on the object X to generate a seed for a pseudo-random number generator. The pseudo-random number generator then generates a sequence of random numbers that are used to create the random perturbation from a well-defined algorithm. SDI can uses this technique to verify randomization, and audit a profile-updating agent. "Playback ability" - the ability to reconstruct the original record from a noisy version of that record is important for a number of purposes. An individual may want to obtain proof of a transaction for legal purposes and law enforcement agencies with appropriate warrants might want to examine original records. Each client-level SDI proxy S maintains a trapdoor function f (such as the RSA encryption-decryption function). When adding noise to a record the proxy uses the fixed fields of the record as argument x and computes the inverse of f when applied to x. (Note that third parties that do not have access to the trapdoor secret will not be able to compute this inverse function.) The proxy then uses this value as the seed for a pseudo-random number generator and uses the bits produced by the generator as the random bits used to produce the noise. With this scheme it is possible for the agent database to "playback" the noise perturbation and produce the original record from the noisy record. This playback scheme is optional and may be used in an application if the feature is desired.

4.4 Applications of Randomized Aggregate Technique

Secure Function Evaluation is useful to identify the value of information without revealing information, but it is computationally expensive. Not only must the inputs for the function to be evaluated be transmitted over the network, that is the full representation profiles, but the functions must also be evaluated cryptographically.

Sometimes aggregate information can be used to compute information value, without details of individual records. Consider the following scenario: agent X wants aggregate information about a population of agents that do not include agents of type T. The agent can request aggregate information about this population, although the population selection might be performed to include some agents that are possibly of type T, because the type of an agent could be know only probabilistically. A vendor might request aggregate information about users for which it already has data. This is possible when users have the same pseudonym with multiple vendors. On the basis of the aggregate information the vendors can then decide on mutually beneficial terms of disclosure, and the disclosing vendor can send the requesting vendor the raw information, non-aggregated.

Serial No. 09/459,466 - Docket No. YOR9-1999-0555

5. Safe (Privacy-protected) Personalization

This section will describe the ability to provide personalization through a technique where information is processed at the agent's local computer device, using instructions from the other agent, and the agent's local profile information.

The usual method of personalizing information flow from provider to consumer is where:

1) The consumer describes its preferences to the provider.
2) The provider uses a local method to decide on the appropriate information to provide to the consumer.
3) The provider pushes personalized information to the consumer.

The problem with the above methodology is that the provider learns information about the preferences of the consumer, perhaps more information than the consumer trusts the provider with--- given that the provider might sell the information to other agents, or use the information to determine the identity of the consumer.

We propose the following novel technique:

1) The provider pushes generic information to the consumer.
2) The provider pushes personalization methods to the consumer.
3) The consumer agent performs personalization on the generic information, using the methods provided and its local preferences and profile information.
4) The consumer responds to the provider on the basis of the personalized information that it receives.

The above technique allows the same level of personalization, but does not provide any preference information about the user's preferences to the provider, beyond information that is implicit in how the consumer follows up to the provider, on the basis of the personalized information received.

5.1 Example Implementation: Leveraging XML and Java Techniques

6. Privacy-Protected Data Mining

This section will describe the technique that allows data-mining and model-building while supporting privacy protection by allowing agents to submit profile information anonymously to a data warehouse.

7. Modes of Profile Release / Agent Profiling / Implicit Agent Information

The privacy protection provided by SDI does not assume that there is any technique available to prevent agents from releasing information that they have about other agents.
The main method that protects the privacy of agents is through careful identity management, and careful release of sensitive information that could be used to make inferences about the identity of an agent.

Any party to a transaction is assumed to be able to release information about the transaction to any other party. For example, consider an example from an electronic commerce domain, with an interaction under a unique persistent pseudonym between an individual and an on-line travel agent. Suppose during one session the individual requests flight information for a particular schedule, and then interacts with another travel agent under a different pseudonym. If the individual requests information about the same schedule from that travel agent, and the travel agents share transactional information about their clients, then the agents can attempt to make an inference that it is the same individual – and then leverage previous transactional history.

However, additional infrastructure could support tighter control over information exchange, for example requiring that agents enter into contracts that state that they are authorized to release information about a transaction with an agent, authorized to release only randomized information, or authorized to release any information. Rules could be supported within an economic framework (with penalties for abuse), or a legal framework. Such rules would extend the basic system of SDI.

In this section we will also allow a method that will allow third parties to approach agents with requests to augment their profile information with demographic and other information collected off-line, that can enhance value.

Implicit agent information includes **clickstream data**, and the policy for the management of clickstream data will be discussed in this section.

3. *The Client-level Proxy*

The client-level proxy, implemented as a client program running on the **user**'s client machine is responsible for managing all data transfer between the client (meaning the client machine and the **user**) and vendors, or other **user**s. In particular, a key function of the client-level proxy is to implement a pseudonym-management policy for a **user**—that is able to exert complete control over the ability of vendors to compile data on **user**s. The client-level proxy also negotiates privacy and data-use practices with vendor level proxies. The proxy also allows **user**s to control the addition of demographic and other personal information to electronic profiles, and adds noise to data fields to protect **user** identity. Figure 2 shows the client-side view of the Secure Data Interchange.

The client-level proxy maintains profile information for a **user**'s collection of pseudonyms, and allows the **user** to view and challenge profile information. The proxy also provides a rule-based interface to allow a **user** to select appropriate privacy/personalization policies.

The client-level proxy also retrieves pseudonymous e-mail for a **user** from pseudonymous e-mail boxes, and maintains shared keys for use with each vendor that the **user** interacts with. Finally, the proxy can personalized generic information provided by a server according to rules provided by a server and profile information stored at the client, and filter/validate incoming e-mail messages.

The proxy also provides ancillary services, such as automatic **user**-verification with web pages, a "secure cookie" system to allow servers to maintain stateful interactions with **user**s without allowing the identity of **user**s to be compromised by "flags" that are left on the client and retrieved by another vendor later.

The primary mechanism that protects the identity of a **user** across multiple vendors and service providers is the ability to interact pseudonymously with vendors. The **user** can choose a unique pseudonym for each third party with which

he/she interacts, and be absolutely certain that he/she is the only party that knows his/her true identity, other than the trusted proxy server where the pseudonyms are registered. There is no way that a vendor can know anything about the transactions that a **user** has had with other vendors under alternate pseudonyms unless the **user** chooses to disclose the equivalence of pseudonyms, or use the same pseudonym across multiple vendors.

3.1 Initialization

The Client-level SDI proxy server runs on the **user**'s client machine, and acts as an intermediary between the **user** and the Internet, intercepting all outgoing and incoming messages. Given that the **user** runs a standard Internet browser, e.g. Netscape or Internet Explorer, the proxy can be implemented as a plug-in got the browser, integrated directly into the browser, or downloaded as Java (or some-other platform-independent) code. The browser is configured to use the SDI proxy as its proxy, and the SDI proxy itself connects through the ISP-level (or other intranet gateway) proxy server to the Internet.

A new **user** must down-load the client-level proxy server that will run on his/her local host machine (or an SDI-enabled browser), and configure his/her browser to connect to the first-level proxy server. Furthermore, the **user** must be located on an SDI-enabled network, where the Internet Service Provider has a SDI second-level proxy server at the gateway to the Internet. A new **user** connects to the main Secure Data Interchange through his/her browser (configured through the first-level proxy server), by entering the SDI URL (e.g. http://www.sdi.com). The log-in page will prompt for a new **user**-name and password.

Sovereign Data Interchange

3.2 New-**user** Registration

The client proceeds to automatically generate a unique SDI **user** ID code, and provide information about the **user** to a central SDI database -- although this information will not be linked to a **user**'s pseudonyms. A flow-chart for the process of registering a new **user** with SDI is show in Figure 10.

When a **user** first registers with SDI the **user** provides the client-level proxy with personal information, such as its name, mailing address, and e-mail address (at a minimum). The client-level proxy registers the **user** with the central SDI server, providing the server with the name, address and e-mail address of the **user**. Other basic **user** information could include demographic information, for example a **user**s job, marital status etc. The **user** can configure his/her client-level SDI proxy to release some of this information automatically to vendors.

At this stage the central SDI server must verify the identity of the **user**, and also check that the **user** is not already registered with SDI. The method for verifying the identity of a **user** could include requesting that the **user** provides his/her social security number, or some other institutional solution that is used for this purpose. In the future we could envisage an electronic system for such an identity procedure, but the method might require for the **user** to execute this initial step in person with the presentation of a recognized photo ID. The central SDI **user** ID server maintains a database of all **user**s that are registered with SDI, and checks that that the **user** is not already registered with the system of secure data interchange.

When a new **user** registers with SDI the **user** must create a unique public key/private key pair. This key pair can be generated only once for a person, and although the central SDI **user** ID server does not know the key pair, the server can verify that a key pair is only generated once - because a new **user** must present proof of identity to establish an account. The unique key pair is used by the client-level SDI proxies to generate new pseudonyms for **user**s, and to verify (when necessary) that a **user** has only one persistent pseudonym for each vendor.

The client-level proxy now generates a unique **user** identifier, UUID. This is blinded, and signed by the central SDI server so long as the identity of the **user** can be validated. The technique of blinded signatures is discussed in the Appendix. The client-level proxy now removes the blinding factor, and has a signed UUID that it uses when it is necessary to generate new pseudonyms and request new certificates. The central SDI server that validates new **user**s has a public/private key pair for the purpose of validation, eg. (PKSDI, SKSDI).

The central SDI proxy also provides the **user** with a signed certificate of some universal identifier, such as its Social Security Number, that the **user** can use to generate other certificates from "**User** Certifying Authorities".

Furthermore, the UUID acts as a public-key, and the client-level proxy also generated a private-key. The client-level proxy can now sign messages with its private-key, and provide the signed to UUID, to verify that (1) the UUID represents a validated **user**; (2) it is the client-level proxy authorized to act for the **user**, because it has the private-key associated with the UUID. The technique of public key/private key cryptography is discussed in the Appendix. The client-level SDI proxy uses the private key to authenticate messages that it sends to other modules within SDI, such as Pseudonym administering servers.

The unique **user** ID for a **user** does not carry any information about the **user**, its sole purpose is to provide a unique identity.

The central SDI server does not know the unique **user** ID that is associated with a **user** because we use the method of blinded signatures, but nevertheless a third party or another SDI module receives a guarantee that the **user** ID has been validated by the central SDI **user** ID server, and that only one **user** ID was certified for the **user**. This is a useful property because the system of SDI is absolutely and unconditionally secure from security violations that could lead to the security of a **user**s identity to be compromised.

3.3 Selecting Privacy and Profile Management Policies

The next stage in registering a new **user** with SDI is to establish privacy and profile management policies. We describe a "pseudonym management policy", and a "profile management policy". A **user** can define how he/she wishes to interact with various classes of vendors (depending on the nature of the business that the vendor is engaged in), the kinds of uses to which the transactional information that a vendor collects can be put to, and the amount of information that a vendor is authorized to release. The **user** can also specify a "basic profile" that the **user** is willing to release to any vendor, irrespective of the vendor's policies. The profile-management policy is further broken down into the "data-release" policy and the "use-of-data" policy. The client-level proxy manages a **user**'s interactions with vendors, to keep them within desired policies.

Google Exhibit 1009 - Page 27 of 96

3.31 Pseudonym Management Policies

Pseudonym management, coupled with the ability to add noise to information provided, allows **user**s to exercise complete control over the ability of vendors to collect and exchange information about them. The client-level proxy contains a rule-based interface that allows the **user** to define a pseudonym-management policy.

Abstract Policy Hierarchy:

Level 0 (Highest)

At the highest level of privacy a **user** chooses to interact anonymously with every vendor, so that vendors cannot even personalize its service to a **user** over an extended interaction, because the vendor will never know who the **user** is. An anonymous pseudonym is simply a one-time traditional pseudonym, where a PAS does not need to check that a **user** does not already have a pseudonym for a vendor.

Level 1

A **user** can choose to interact with every on-line server under a unique pseudonym. This completely prevents a vendor from knowing anything more about the **user** than the information that can be inferred from the interaction itself. So long as that information does not identify the **user**, a vendor cannot access any information from external databases (such as demographic information), and a vendor cannot access any information collected by other vendors about the **user**. The vendor can still personalize information that is displayed to the **user**, but only on the basis of its own historical information for that **user**.

Level 2

At level 2 a **user** can choose to share pseudonyms for groups of vendors (for example a **user** might choose the same identity for all vendors that sell music and books), and receive personalized service from each vendor on the basis of her extended profile.

The system of SDI allows **users** to maintain "ownership" of the dossier of information by: (1) providing a different pseudonym to each vendor, but providing the central SDI database with information about which pseudonyms are equivalent; (2) allowing vendors to take advantage of the combined profile when providing information/products to a **user** by allowing personalization on the client machine but without release of the complete profile for the group of vendors to a vendor.

Level 3

At level 3 the **user** can interact with every vendor under the same pseudonym, or anonymously with some vendors until enough trust has been established.

Level 4

At the lowest level of privacy protection (none) the **user** can simply use a unique ID that is linked to his/her true identity to interact with every vendor. At this level the **user** has no control over the dossiers of information that can be collected by groups of cooperating vendors.

Figure 6 illustrates level 2 - which is the preferred mode of interaction. The client-level SDI proxy for a **user** submits a request for information to a vendor server, including the pseudonym of the **user**. The vendor's server can perform some personalization at the server level, based on data that it has accumulated from previous interactions with the **user**'s pseudonym, and also push generic information and rules for processing the information to the client machine, where information is processed according to a **user**'s extended profile for the group of vendors to which the vendor belongs. The rules could be implemented as Java code or Javascript, and the generic information pushed as XML documents, for example.

3.32 Implementation Details

The central Secure Data Interchange server categorizes vendors that register with SDI, by assigning labels from a set of classifiers, that indicates the business of the vendor (i.e. the services and products that a vendor provides). The set of classifiers might include: music goods, news media, vacation packages, groceries, clothes. We denote the classifiers abstractly as L1, L2, ., LN. There is also a label for vendors that are not currently registered with SDI, denoted L0. Given the labels, each vendor has an associated set of labels, denoted L(Vj), for example L(V4) = {L1, L4, L12}.

The **user** is able to configure an appropriate pseudonym-management policy at the first-level proxy level that runs on his/her host machine. The **user** assigns a pseudonym-management action to different vendor-classes. The set of actions that SDI provides can include, but are not limited to:

A. Anonymous interaction with vendors in this class (i.e. use a different pseudonym every time the **user** enters the Vendors' sites.)
B. One pseudonym per vendor in this class, but the same pseudonym for all visits to the same vendor.
C. One pseudonym for all vendors in this class, and the same pseudonym at all times.

Pseudonym-management action A provides stronger data-privacy than action B, than action C, but action C provides the most opportunity to vendors to exchange information on the **user** and provide personalized and informed service to the **user**. This is a basic tradeoff that the **user** must make - between privacy and personalization.

A policy maps a vendor to a "management action", i.e. with Vendor V I will use the same persistent pseudonym for all interactions, but it will be unique to that vendor.

The vendor's are grouped according to their "business classifier" labels. For example, group 1 might contain all business that sell books or CDs, group 2 all businesses that sell computer hardware, etc. Each vendor group has an associated management action, e.g. vendors in group 1 (books/CDs) can share the same pseudonym for the **user**—and exchange information as provided by the **user** and the vendors to the central SDI server. Other groups could be defined by the labels that vendors DO NOT have, e.g. group 3 could be vendors that do not sell credit cards.

Clearly, it is possible for a vendor to be categorized into more than one group, within these rules. For example a vendor that sells books and Compact Disks would be placed in the group of vendors that sells books, and a group of vendors that sells CDs. When this occurs the rule-base must choose the most appropriate group, for example on the basis of profiling a group of vendors according to the profiles of their **user**-base. Alternatively, it is possible to "partition" an individual business into different core businesses that fit cleanly into a single group, and then ensure that a **user**'s interactions respect the current group that the vendor is a representative of.

The **user** can choose a mapping from labels to groups that best fits his/her own privacy needs, and then continue by assigning pseudonym management actions to each group. For example, the **user** might decide that all vendors that sell flights, books, and music should be assigned the same pseudonym for all transactions (i.e. pseudonym action C), while all vendors that sell financing, loans, and credit cards should be assigned to the same pseudonym for all transactions (i.e. pseudonym action C), but to a different pseudonym than the first class of vendors. Similarly, a **user** might require a single pseudonym for each vendor that is not classified by SDI (pseudonym action B), and anonymous pseudonyms for vendors that are classified as suppliers of adult material.

We allow vendors to provide different service levels, depending on the pseudonym action that a **user** chooses. For example, a vendor might desire to provide a better level of service to a **user** that interacts with a persistent pseudonym than a **user** that interacts with an anonymous pseudonym. The first-level proxy sends a token that certifies the type of pseudonym-action policy that a **user** has chosen when a **user** connects to a vendor's site. This is explained in more detail in Section below.

3.4 Profile management policy

The basic mechanism that a **user** has to prevent vendors exchanging information with other vendors about his/her transactions is the pseudonym-management policy. However, there is still a chance that a **user**'s identity can "leak" to another vendor if the **user** has performed a transaction or provided information that reveals the **user**'s identity. For example, if a **user** purchases a flight from Philadelphia to San Francisco on UB 004 that arrives in San Francisco at 11.50am on 3/10/99 from one vendor, and then immediately purchases a car rental from San Francisco International, indicating arrival on flight UB 004 from another vendor, then there is a high probability that it is the same **user**. If the first vendor notifies all major vendors that offer car rentals that it has just completed a sale of a flight ticket to a **user** with pseudonym C1, and another vendor sells an appropriate car rental to **user** with pseudonym C121, then the two vendors now have some positive evidence that the two pseudonyms might belong to the same **user**, and in future can try to cross-sell appropriate products.

3.41 VENDOR-SIDE PROFILE MANAGEMENT

This is where the role of the "profile management policy" is important. A **user** can assign a "data-release" action to each class of vendor, depending on the types of actions that a **user** does with a vendor. There is no way to prevent a vendor releasing the details of a transaction with a **user** to other vendors, and if the details allow other vendors to identify the **user**, then the **user**s privacy can be compromised. However, we can control the type of information that a vendor provides to the central SDI server, for exchange with other vendors.

To help with this process, the central SDI server classifies the type of service/goods that a vendor provides according to whether they are "anonymous" high-volume goods such as compact disks or newspaper articles, or "non-anonymous", low-volume, possibly personalized goods, such as cars, flights, or property. A vendor must initially enter a contract with a **user** about the type of information that it can release to other parties, i.e. the choice of data- release actions can include, but is not limited to:

A. Release no information.
B. Release randomized information.
C. Release all information.

The vendor-level SDI proxy is able to randomize transaction information automatically if option (B) is selected, to allow the vendor to submit profile information to the central SDI server without breaking the **user**'s privacy requirements.

The vendor must provide the central SDI server with its certificate of agreement with the **user**, when it submits information. All profile information can be verified by the **user**, and in the randomized section we also describe a technique to validate randomization of profile information. The client-level proxy maintains profile information for each of a **user**'s pseudonyms, and allows a **user** to view and modify (challenge) profile information that has been provided by vendors and other parties. The proxy requests periodic profile updates from the central SDI server, providing validation via the PID and associated private key that the proxy is authorized to receive profile information.

For example, a **user** might choose that all vendors that sell flights can only release randomized information to the central SDI server or other vendors, while vendors that sell compact disks can release complete information on transactions to the central SDI server and other vendors (although this information will only relate to the actions of the **user** under the same pseudonym).

Google Inc. v. Hood 2/9

3.42 CLIENT-SIDE PROFILE MANAGEMENT

Orthogonal to pseudonym management is profile management, which determines how a client-level proxy will release profile information to vendors.

Level 0 (Highest)

Release no information.

Level 1

Release randomized information. The information is randomized before release to prevent any compromise of the **user**'s identity from profile information that is too specific, while enabling useful personalization at the server level, and enabling useful analysis at the central SDI server.

Level 2

Release non-randomized profile information for a **user** under the relevant **user** pseudonym.

We ensure the accuracy and content of profile, given that updates can be made by any third party that is privy to information by allowing **user**s (individuals) to specify privacy constraints that vendors must uphold in reporting information to other vendors and/or SDI

3.43 CLICKSTREAM DATA (CLIENT-SIDE POLICY)

Clickstream data must be logged at the client, because of proxy caching. This data is released periodically to servers. The client-level proxy server that runs on a **user**'s host machine is in a unique position of being able to monitor the **user** across different pseudonyms and across different vendors' sites. For example the proxy-server can monitor:

1. Clickstream data across different vendors' sites and pseudonyms.
2. Data that is displayed to the **user** (text, names of graphics objects.)
3. Input provided by the **user** at the keyboard of the host machine.

We term all of this data "clickstream" data because it is gathered by passively observing the actions of the **user**, and not by direct question-and-response. The clickstream data policies that are available to a **user** can include, but are not limited to:

A. Release no information.
B. Only release data on the URLs of the most recent sites visited.
C. Release data about the URLs of the most recent sites visited, and the information displayed to the **user**.
D. Release data about the URLs, the information displayed, and the information entered by the **user**.

In addition, the clickstream data policy restricts the depth of information that is provided, i.e. for how many previous sites, and can also restrict the data released to data that was collected under the current pseudonym of the **user**. The data that is released can also be randomized in the same way as explicit data is randomized, for example by removing time-stamp information.

3.5 Certificate Management

User certifying authorities issue certificates that certify properties about the **user** that is represented under a particular pseudonym. This service is important to a pseudonymous electronic commerce environment because vendors will require guarantees about certain properties of **users** that they do business with, such as the age and credit worthiness of a **user**. Although outside of the basic SDI framework, the existence of such authorities is assumed in the description of the basic operation of SDI.

When a **user** needs certificate C(P) for pseudonym P, a **user** can use the following steps, to gain a certificate without providing a **User** Certifying Authority with any new information. The **user** sends a message to the Certifying authority, containing its SSN certificate, that it received on initial registration with SDI. The **User** certifying authority can then check the certificate, and if it is valid and the **user** can be certified for the requested property, signs a blinded message to certify the new property—for a one-time pseudonym specially generated by the **user**. This certificate can then be unblinded, transferred to certify other pseudonyms that the **user** holds, and used for certification purposes. Certifying Authority servers have key pairs, for example (PKCA,k, SKCA,k), which again are using for certification and encryption purposes.

The certificate can be trusted so long as the certifying authority keeps its private key secure. To give an example, a trusted third party can exist to certify that a **user** represented with a particular pseudonym is above 18 years old. The third party can maintain a private key/public key pair (SK_18, PK_18) and sign a message that includes the pseudonym of the **user** with its private key to generate a certificate, C_18(U), that will assure other parties that the **user** is 18 years old. This certificate can then be requested by other vendors and information providers, and checked for validity within a public key infrastructure that maintains a faithful copy of the public key of the certifying trusted third party.

The first-level proxy server is responsible for certificate management. When a certificate is required by a vendor, then the proxy server checks whether the certificate has been issued to the **user** under one its pseudonyms. If this is the case, then the proxy server will simply transfer the certificate to the appropriate pseudonym for the **user** with the current vendor, using a technique taught by D. Chaum. There are various certifying authorities within the Secure Data Interchange system, that are able to use a verified social security number to provide certification about the **user**. When a new **user** registers with SDI, the central SDI server provides the following social security number certificate to the **user**, S((SSN, PKSSNC), SKSDI), that links the social security number of the **user** to a public key that was established for this purpose.

When the **user** needs a new certificate, the first-level proxy server creates a new key pair, using the method described above to certify that this is the first certificate of this kind that the **user** has requested. The first-level proxy server receives a validated new public key, S(PKP, SKCAS), that it sends to the Certifying authority. The first-level proxy server now generates a blinded certificate number, B(CERTP), and transfers the social security certificate that is signed by the central SDI server to its new pseudonym. The first-level proxy server forms the message M=(B(CERTP), S((SSN, PKP), SKSDI)), signs it with its secret key SKP, and sends it to the Certifying authority.

The certifying authority now verifies the public key, PKP, and then verifies that the **user** has the appropriate secret key, i.e. checks that the message M is correctly signed. The certifying authority continues by verifying that the **user** with pseudonym P has social security number SSN. Then the authority determines whether the **user** with social security number SSN has the correct property, and if it does, signs an association between the blinded certificate number that was provided by the first-level proxy server of the **user** and the public key of the **user**'s pseudonym.

Finally, the first-level proxy server can remove the blinding factor, and it is left with a signed certificate that associate the public key for its new pseudonym, PKP, with the certificate number. The certificate is represented CP = S((CERTP, PKP), SKCA).

The certifying authority now knows that the **user** with pseudonym PKP has been issued with a certificate, and furthermore, the authority knows the social security number of that is related to the pseudonym. However, the **user** will never need to use the pseudonym again, because it can transfer certificates issued in one pseudonym to certificates issued in another valid pseudonym using a method taught by D. Chaum. The Certifying authority therefore gained no information, because it already knew the relation between social security number and property that it is certifying. For example, we can transfer certificate CP to certificate CQ = S((CERTP, PKQ), SKCA).

3.51 Pseudonymous Interaction with a Vendor

The proxy can validate that it represents a particular **user** under pseudonym P by sending a message to a vendor with the signed PID, and signing a challenge provided by the vendor. The vendor can validate that the signature corresponds to the PID. See the Appendix for a discussion of this challenge-response mechanism.

3.6 Generating a New Pseudonym

The fundamental model of interaction between a **user** and a vendor under SDI will be pseudonymous. When the pseudonym management policy of **user** requires that a new pseudonym be generated for a **user**, it is necessary to validate the new pseudonym to verify to a vendor that the **user** has a unique pseudonym for its site.

A **user** can interact under different pseudonyms with each server, as dictated by her privacy and profiling policies, and the declared policies of vendors with which she interacts. Each pseudonym is associated with a signed PID, (pseudonym ID), a private key that is useful to validate that the PID has not been stolen. The client-proxy also maintains a shared key with each vendor, because symmetric encryption/decryption is cheaper than asymmetric public key/private key encryption.

There does not exist a central database of public and private keys for **users**, or **user** pseudonyms. We use the method of blinded signatures to certify **user**-generated key pairs, so that only the client-level proxy servers can link PIDs to **User** identities. Client-level SDI proxies generate new PIDs, that are unique with a high degree of probability, and blinded before authenticated for use within the system of Secure Data Interchange.

A key feature of the system for administering pseudonyms is that the pseudonym administering authorities cannot build dossiers of the pseudonyms that are authorized for each **user**, because **users** submit "blinded" PIDS to be validated. The only information that a PAS has is for each unique **user** ID, what vendors has the **user** registered with. This information cannot be linked to the real (physical world) identity of the **user** because the central SDI **user** ID server has no information about the **user** IDs that are authorized for each **user** that registers with SDI (because the UUID is blinded before signed by the central SDI server).

To generate a new pseudonym for a particular vendor the proxy requests a new PID from the "Pseudonym Administering Server" that has a trusted-agent relationship with the vendor. The proxy provides the PAS with the signed UUID, and if the PAS can verify that the UUID has not already applied for a PID for use with this vendor then the PAS will sign a blinded PID provided by the proxy, with a signature that indicates the vendor that it is valid for—and make a record that a PID has been authorized for **user** with UUID and vendor V.

Each Pseudonym administering server has a public key / private key pair (PKPAS, SKPAS) for each Vendor for which it validates new pseudonyms. A PAS will sign the public key of a pseudonym using the PAS private key associated with the particular vendor. In effect, the operational pseudonym that a **user** uses is the triple (S(PK, SPAS), SK, IP), representing a pseudonym that is authorized for a particular vendor server.

There is no information compromise here, other than it becomes possible to construct the set of vendors that a **user** with UUID has applied for PIDs with. However, the only entity that knows the true identity of the **user** with UUID is the client-level proxy, because the UUID was blinded before signed by the central SDI server when a new **user** is registered.

In this way it is possible to provide vendors with guarantees that PIDs are once-in-a-lifetime for a **user**, so that vendors can continue to achieve at least the same level of personalization as they do without SDI in an SDI system. In this way SDI can guarantee once-in-a-lifetime pseudonyms to those vendors that require persistent interactions with **users**.

The PID and associated private key can be used to validate an initial information exchange with a vendor, but it is also possible to perform follow up message exchange using a shared key pair—this is more efficient to implement that an asynchronous key pair cryptographic solution. Messages can be encrypted with the shared key, that only the **user** and the vendor know. This (1) validates that the message is from the sender; (2) ensures that only the intended recipient can read the message.

3.61 Implementation Details

Figure 11 shows a flow-chart for creating a new "relationship" with a vendor. The Pseudonym Administering servers (PAS) provide for validation of new pseudonyms. Each vendor selects a PAS that will be responsible for managing pseudonyms for its domain. When a client-level SDI proxy requires a new pseudonym for a **user** UUID, the proxy does the following steps:

1. Request the URL of the PAS from the server for the new vendor that the **user** wants to initiate a persistent relationship with.
2. Send a message to the PAS with the tag "New Pseudonym", the URL of the vendor, and its validated UUID. Also be prepared to answer a challenge/response with the private key associated with UUID.

The PAS for the vendor then checks that it has authorization to administer pseudonyms for the vendor. The PAS performs the following steps:

1. Check that the vendor URL corresponds to a vendor for which it has pseudonym-management authority.
2. Verify that the UUID is correctly validated for use within SDI (indicating that the **user** is a member of SDI).
3. Verify that the client proxy server represents the **user**, through a challenge-response sequence that requires that the proxy has the private key for UUID.
4. Look up the **user** ID in the database of **user** IDs that the PAS maintains to check that a pseudonym has not been authorized for this **user** and this vendor.
5. Send a message to the **user**, indicating either (OK, or DENY).

Given that the client-level SDI proxy receives an OK response, the client-level SDI proxy then generates a new PID, and private key, and blinds the PID for validation:

3. Generate a new key pair, (PID, SK).
4. Blind the public key, and send a message to the PAS with the tag "Request Certification", the URL of the vendor, the signed unique **user** ID, and the blinded public key, B(PID)
5. Receives a signed copy of the blinded public key in return, and removes the blinding factor to obtain S(PID, SKPAS).
6. Generate a new pseudonym, from the components, (S(PID, SKPAS), SK, EMAIL), where EMAIL is a new e-mail address for the pseudonym.

The public key that represents the new pseudonym is signed with the private key of the PAS that relates to the particular vendor. This enables a vendor that receives the pseudonym to validate that the pseudonym is unique for the **user**, to enable persistent interactions across multiple sessions.

3.7 Personalization of information/Manage e-mail

The client-level proxy can provide vendors with "certificates" to enable them to send e-mail to the **user** (under a particular pseudonym). Outgoing e-mail includes a certificate that a vendor can use to reply, and a pseudonym ID in place of the standard "from" field. The certificate is of the form S(M, SKP), where P is the pseudonym of the **user**, and M = (PKP, PK*V). The vendor includes the signed certificate when replying to the **user**.

Exhibit 1029 - Page 33 of 50

3.8 Automatic release of personal information

Maintains certificates as well.

3.9 Enhancing profile information

Receives update requests from SDI server. allowing **users** to access any profile information that is stored in the SDI server, and request changes to that information if it is not accurate, or if the **user** does not want content to be available to third parties. This information is only released to **users** via the proxy server that is authorized to represent the **user** for the relevant pseudonyms.

3.10 Manage iAmWorthIt module/Negotiation

The proxy server determines what kind of site in being visited, from SDI credentials that are provided to a vendor's SDI proxy server, and either embedded in web documents, or provided in prenegotiation. The proxy then determines an appropriate type of interaction level with the vendor, depending on the profile and credentials of the vendor, the type of information required by the vendor, and the **user**'s privacy policy.

The system of Secure Data Interchange allows **user**'s to receive compensation, what we will call "community dollars", in return for providing information to vendors. The architecture supports a negotiation between **user**'s proxy servers and vendor's proxy servers, to strike a deal about information use and compensation. In some cases the exact nature of an offer may not be anticipated, and the **user** can be contacted directly.

The vendor's host level proxy server is then granted the permission by the client level proxy to receive certain appropriate information. The vendor can use profile information about a **user** (for a particular pseudonym) to present appropriate services, products, and prices, including custom priced items and promotional offers (as suggested in the co-pending application entitled "System for Customized Prices and Promotions").

3.11 Remote Retrieval of Profiles

The client-side SDI proxy is designed to be configurable from a remote database. This enables a **user** to maintain a persistent SDI profile across different client machines, for example at work and at home. The profile, pseudonym and key information that represents a **user** in its interaction with SDI-enabled vendors and information providers during a session can be saved, and then encrypted and stored in a remote database for **user** name and password access. Alternative technologies include smart card techniques, where the data is stored on a portable device in encrypted form.

3.12 Use-of-data Policy (Access Control Policies)

The Secure Data Interchange system also allows a **user** to place constraints on how the vendor can interact with the **user**, and also on what uses can be made of the information that the vendor releases to the central SDI server and other vendors in the system. The **user** can specify whether or not the vendor can send electronic and/or physical solicitations to the **user**, and the **user** can also specify whether data that is released can be used for (any of): solicitation by other vendors, personalization of service should the **user** visit a vendor's site. For example, a **user** might require that any information that is released by a vendor to the central SDI server, and then possibly exchanged with other vendors, is only used to personalize the service and products that a vendor offers to the **user** should the **user** visit the vendors site, and is not used for electronic or physical solicitations. The use-of-data policy is augmented to the pseudonym-management policy, and defined over classes of vendors.

SDI enables a **user** to allow limited and secure data sharing between vendor types of its choice. i.e. if I use the same pseudonym for multiple vendors, I am still not saying-OK you can all share my data (although I might choose to say that). I can say-OK, you can use my data to personalize the service that you offer me without knowing what I did with other vendors. Also, vendors themselves can say-here is the data, allow other vendors who are not competitors to access the data, and refine their models, but not actually access details.

3.13 Dynamic Privacy Management

When a **user** clicks to a new URL, then the first-level proxy server first checks its local cache, to determine whether it already has the vendor's classification certificates. The proxy server requests the classification certificates and public key certificate from the vendor if the vendor is not already in its cache. The proxy server continues by verifying the integrity of the certificates, and checking that the vendor matches the ID enclosed in the public key certificate.

The proxy server continues by looking up the pseudonym-management policy for the vendor, and checking whether an appropriate pseudonym already exists in its local database. If the vendor belongs to a new class of vendors that do not yet have a pseudonym, or the vendor belongs to a class of vendors that require one pseudonym each, or the vendor belongs to a class of vendors that require anonymous interaction, then the first-level proxy server continues by generating a new pseudonym.

When a **user** interacts with a vendor under a persistent pseudonym then the **user** must have the pseudonym certified by a Pseudonym Administering server, to certify that this is the only pseudonym that the **user** has registered for the vendor. Each vendor selects a Pseudonym Administering server that will certify all of its pseudonyms. If the **user** already has a pseudonym for the class of vendors that the vendor belongs to, but the pseudonym has not been certified for this vendor, then the first-level proxy server will have the pseudonym certified. Similarly, if the **user** requires a new pseudonym for this vendor, then the first-level proxy server will generate a new key pair and have the public key certified by the Pseudonym Administering server.

The first-level proxy server performs the following steps (we use the technique of blinded signatures so that the Pseudonym Administering server does not know the public key that it certifies):

1. Generate a new key pair, (PKP, SKP) - or - lookup an existing key pair that will form a pseudonym for this vendor.
2. Blind the public key, by multiplication with a random number, and form the message M=(S(PK*U, SKSDI), S(B(PKP), SK*U), S(PK*V, SKSDI)).
3. Encrypt the message with the public key of the Pseudonym Administering server, and send the message.

The Pseudonym administering server checks that the unique public key of the **user** is signed correctly by the central SDI server, and that the **user** has the private key associated with that public key by using the public key to verify that the blinded pseudonym public key is correctly signed. Finally, the Pseudonym Administering server checks that the public key of the vendor is signed correctly by the central SDI server, and then checks that the **user** does not already have a pseudonym for the vendor, in a database that it maintains of which unique **user** PKs have requested pseudonyms for each vendor.

If everything is OK, then the Pseudonym administering server signs the blinded public key for the **user**'s new pseudonym using a key pair that it maintains for the vendor, and returns the signed blinded key, S(B(PKP), SKPAS,V) to the first-level proxy server.

The final step in the protocol is for the first-level proxy to remove the blinding factor, so that the **user** now has a new key pair, (PKP, SKP), with the public key certified by the appropriate Pseudonym administering server private key, S(PKP, SKPAS,V), to demonstrate that this is the only pseudonym that the **user** has for the vendor.

The first level proxy server now connects to the vendor in secure-SDI mode, by sending the vendor the certified public key for the **user** under this pseudonym that will identify the **user** to the vendor. The proxy server continues by sending the data privacy policy for the **user** with this vendor, signed with the private key for the pseudonym. This serves three main purposes: (1) it demonstrates to the vendor that the **user** is indeed the owner of the public key PKP, because the **user** has the secret key that is associated with it; (2) it informs the vendor about the data-privacy policies that the **user** requires; (3) whenever the vendor submits information about transactions with this **user** to the central SDI server it must also submit this certificate to verify that it is following the **user**'s guidelines.

3.14 Dynamic Profile Management

When a **user** connects to a site and provides a certified public key, the first-level proxy server also provides a time-stamped certificate of connection, S((PK*V, T), SKP), where T is the current time, PK*V is the unique public key of the vendor, and SKP is the secret key of the **user** for the pseudonym that it uses with the vendor. This "connection certificate" is used by the vendor to request a profile-release from the central SDI server.

In addition, the first-level proxy server also provides a basic-profile to the vendor. The basic profile for a **user** contains no identifying information, but can contain whatever general information the **user** is happy to release across all pseudonyms, such as the **user**'s age, nationality, state, sex. This is the basic profile that is configured by the **user** during initial registration with SDI.

The vendor provides this certificate to the central SDI server, and requests profile information about the **user** with pseudonym P that is stored within the central SDI server, and is authorized by the **user** to be released. If the vendor is authorized to receive dynamic profiling information, such as the recent web footprints of the **user**, and the material that the **user** has been reading, and the physical location of the **user**, then this information is released to a vendor when the **user** connects to a new site, according to the dynamic profile policy of the **user**.

3.15 Merging off-line and on-line data

Figure 12 shows a flow-chart for how the central SDI server can request that new information be merged with a **user**'s on-line profiles.

The central SDI server can associate off-line information about a **user** with a **user**'s on line pseudonymous profile, even though the central server does not know the **user**'s pseudonym IDs. This can only be done with the **user**'s consent, and may also involve appropriate compensation. Within the system of iamworthit (see section XX) we can credit **users** for both off-line and on-line information.

Merging a marketing database with SDI usr-profiles can be useful both to initialize the database, for example when asking a **user** questions to generate an accurate **user profile** rapidly and efficiently. Off-line data can also add useful richness to on-line profiling information, which may be largely contextual and low on details/factual information. For example, off-line data can include information such as whether a **user** owns a car, rents an apartment, has house insurance, life insurance etc.

SDI can also extrapolate correlations to other **user profiles**, on the basis of common SDI-profiles, for example using statistical techniques.

It is often the case that individual customers appear in some databases, but not in others. Under normal circumstances, an analyst working across different databases would be faced with a large number of incomplete customer records, each with gaps corresponding to the fields of the databases to which they don't belong. A solution to this problem is offered by SDI, which is capable of drawing correlations between different databases - this information can be used to generate predictions to fill in the gaps of incomplete customer records. The result is a full set of customer records that can be meaningfully sorted or filtered by any of the combined fields, and which can now be handled as a unified set of data, suitable for use by standard database analysis systems.

In a typical example, SDI might be used to combine a demographic database, such as the one offered by the Econometrics Corporation, with a commercial database, such as the one offered by Claritas. The Econometrics database consists of 180 million different customer records, but at a fairly coarse-grained level of detail, consisting of such information as age, gender, family status, location (at the state, city, or zip code level), and personal income. In comparison, Claritas offers a smaller base of customers, but includes information of arguably higher quality, since it breaks customers down to the geocode (sub-neighborhood) level, and includes much more detailed information on personal spending habits across hundreds of different purchase categories. A logical reason to combine these databases would be to supplement information about customers in the vastly broader demographics dataset with particular predictions about their personal preferences and likely commercial spending habits. One could imagine using this augmented data set to support a web site that instantly customizes itself to new visitors' preferences. Since the number of records in the Econometrics database is equivalent to roughly 72% of the population of the United States, it is likely

that most first-time visitors to the site will already have a "thumbnail sketch" in the system, and can thus be greeted with an page appropriately configured to their personal tastes.

The technical details of the combination process (which have been described elsewhere in the patent) to a large degree depend on the amount of overlap between the databases, that is, the number of customer records which are shared in common.

Suppose the demographic databases' fields are coded (x1, . , xn), and the commercial databases' fields are coded (y1, ., yn). Suppose further that customers in set A appear only in the demographic database, customers in set B appear only in the commercial database, and customers in set C appear in both.

The process of supplementing the fields of customers A depends completely on the derivation of the distribution f(y1, ., yn | x1, . , xn), which describes the correlation of fields in the commercial database on fields in the demographic database. As previously discussed in the patent, different techniques may be used to create this distribution, depending on the size and variety of C.

As a concrete example, one could imagine that set C includes customers from rural areas. The demographic database would reveal that, although their incomes aren't huge relative to the national average, they tend to spend a lot of it (i.e. are active consumers), have large families, and purchase large vehicles. The commercial database might show that they enjoying hunting magazines and Ford trucks. If they live inland, they buy hunting equipment, if they live near the ocean, fishing equipment.

If these trends are dominant in set C, they will impact the distribution function. Thus, when a browser from a small town in Texas with a typical income pattern visits the automated website, he could be greeted with discounts on truck accessories and a small sidebar with news on the hunting season. On the other hand, a visitor from a small town in Maine might be given the same truck discounts, but would have news on the fishing season.

Although the demographic dataset is arguably the weaker of the two in terms of content, the fact that it contains even a small amount of information on most people in America makes it very valuable for handling first-time visitors, since most of them will appear in it. By using SDI to leverage the more detailed information in the commercial database, we are able to supplement the rough demographic data with predicted commercial preferences. This allows us to construct more detailed thumbnail sketches for each customer, allowing our reception of first-time visitors to be much more appropriate (since knowing personal hobbies or interests tells us much more about a person than general income level).

Exhibit 1029

4. The ISP-level Proxy Server

The ISP-level proxy server is positioned just behind the firewall of the **user**'s local dial-up network (ISP or Intranet). The proxy provides protection for **users** operating under pseudonyms from point-to-point attacks and HTTP header-tracking by stripping HTTP header-information and forwarding HTTP packets on to their destination with no information other than their source at the ISP-level proxy server. The ISP-level proxy also supports pseudonymous e-mail, between **users**, and between **users** and vendors.

Figure 2 shows a couple of **users** connected to clients, that are in turn connected to the Internet through a local intranet, such as the network of an Internet Service Provider (ISP).

The proxy "washes" outgoing messages of any information that would compromise a **user**'s pseudonymity, for example the "referral" field that contains the previous URL of a **user** in a HTTP message. HTTP messages also leak other information, for example browser software on a **user**'s client machine, the operating system and a **user**'s IP address.

4.1 Support for Pseudonymous electronic mail

A **user** can receive electronic mail through the PID and associated IP address of the ISP-level proxy server.

The preferred implementation of this system allows the **user** to periodically check for new mail. The client-level proxy gains access to the mail box that is associated with a pseudonym by providing a correct response (signature) to an ISP-generated challenge. Notice that with this solution, the ISP-level proxy has no way to connect the pseudonyms of a

user, so long as the **user**'s client is not identified in its messages to the ISP-level proxy server other than by the PID that the proxy makes a request for.

We can extend this mechanism using a technique taught in the Lucent Personalized Web Assistant. The LPWA provides for a sequential access mechanism to the mailboxes that belong to a **user** through a one-way function that takes the **user**'s SDI log-in name and password, and an integer from 1 to N, and computes the mailbox location. The mail server does not need to maintain a list of pseudonyms for each **user**, because the **user** is able to efficiently access all of its mailboxes sequentially as a function of other information.

Another variation, that relies on the **user** placing trust in the ISP-level proxy server, provides the ISP-level proxy with the e-mail address for each pseudonym. This push method is more efficient, because the ISP proxy and the client proxy communicate only when new messages arrive, but provides the ISP proxy with information to compute all the pseudonyms for a single **user**—probably undesirable.

4.2 Support for anonymous profile-based search

The anonymous profile-based search allows a **user** to release her profile as an addition to a query term to an general search engine (such as Altavista or Yahoo), and have SDI perform additional filtering of the results of the search to refine the pages returned on the basis of their profiles and the **user**'s profile. This is an example of how SDI leverages existing Internet technologies. Figure 9 shows how the anonymous profile based search can be implemented on the ISP-level SDI proxy.

GOOG-BOOKS-002795

4.3 Maintain **User profiles**

4.4 Support Electronic Payment/Physical mailing solutions

5. Vendor-level SDI Proxy

The vendors are represented with servers that provide an on-line electronic commerce store-front for the services or products that a vendor provides to **users**. The servers may connect through their own pseudonymizing proxy servers, but more likely a vendor will not require pseudonymous interactions with **users**, and will be happy for **users** to build dossiers on interactions with the same vendor. The vendors make use of the support for pseudonymous electronic commerce functionality that the Secure Data Interchange system supports - for example, pseudonymous payment mechanisms and physical mailing of products and letters to pseudonymous **users**. The vendors subscribe to the services of SDI, and can receive payments for **user** information that they provide to the interchange.

Vendors connect to **users** and other vendors through SDI proxy servers. The vendor-level SDI proxy can profile **users**, periodically release information to the central SDI database, and support various statistical generation modules for analysis of profiles collected and requested from other sources. Figure 1 shows a typical vendor that is connected to SDI through a server and vendor-level proxy server.

5.1 Profile Building

The main function of the vendor-level SDI proxy is to build a deep profile of **users** as they interact with a vendor

5.2 Profile Release

and manage the release of information to the central SDI data warehouse. Figures 3, 4, 5 and 6 show patterns of data flow to and from the vendor.

5.3 Request Profile Updates/ Cache metatags/ Manage target object profiles.

5.4 Markup Web documents with Profile information and Personalization commands

A new vendor may choose to operate under different pseudonyms for different **user**s or classes of **user**s, but in general we will assume that a vendor has a unique key pair, and that it is happy to assume the public key of its key pair as an identity, in addition to its more traditional corporate identity. In what follows we assume that the vendor will operate under a single public key with all **user**s, although the extension to multiple pseudonyms is a trivial extension, just using the same technology as described for **user**s.

When a new vendor wishes to register with SDI, the vendor configures a first-level proxy server to manage its SDI functionality. The first-level proxy server is similar to the proxy server configured by new **user**s. Initially the proxy server will generate a new key pair for the vendor, (PKV, SKV), and submit the public key to the central SDI server for signing, along with an identifying code, such as the IP address from which the vendor operates, ID. The vendor receives S((PKV,ID), SKSDI) in return. In addition, the vendor is audited by SDI and is certified according to the class of products or services that it provides, receiving message S(M, SKSDI), where M=(L(V), PKV). The vendor is also certified according to the anonymity of the service or goods that it provides, receiving either an anonymous certificate (e.g. if it deals in high-volume, non-personalized goods), S(M, SKSDI) where M=(PKV, ANON), or receiving an non-anonymous certificate (e.g. if it deals in low-volume, personalized goods), S(M, SKSDI) where M=(PKV, NON_ANON). The vendor will provide these three certificates whenever a **user** connects to the page and the **user** does not already have the certificates cached for the vendor.

A vendor's data privacy policy can also usefully include the placement of restrictions upon any portions or all of their data disclosed to the interchange, such as which other vendors (or vendor types) are entitled to receive the advertising benefits and what types of advertising benefits derived from his/her data are they entitled to; for what advertising purpose can an "entitled" vendor use it for (if it is not deemed "confidential" the disclosing vendor may request beforehand a copy of the solicitation to his/her customers); limiting use of its data exclusively to other vendors who through their data contributions can provide significant reciprocal benefit to them (often occurring in dealing with competitors); physically confining the data to the interchange (without transferring it to the other vendor); data may only be transferred to non-competitors; advertising solicitations that result from the data is of a form that is non-competitive; recipients of data must remain non-competitors.

For example, a vendor can request that the commercial domain of the requesting vendor not match certain vendor types. Vendors must register their business type with the data interchange, and violators can be punished. Alternatively, data can be constrained only for vendors that are improving existing data models, but not to vendors that are seeking new customer prospects. Disclosing vendors may also require that their advertising campaign and possibly even their corporate identities remain secret. Because SDI is a comprehensive repository of statistical data (and customer rating information across a variety of criteria) about sites, it is able to deploy data mining techniques, provide highly detailed statistical information and verification as to the nature of each vendor's business and its customers. Thus helping SDI to better identify, reclassify and enforce this data privacy policies. Provided that the vendor has agreed to release such information to another type specific vendor, this information can serve as additional useful information (particularly quality and other rating credentials about vendors) to help vendors do a better job of selecting vendors by which they establish their data privacy policies.

6. Central SDI Server

Also shown in Figure 1 is the central SDI server, "SDI data warehouse server". The central SDI server maintains extended profile information for each **user**, at a pseudonym level, and profile information for each vendor. The central SDI server also supports cross-vendor and single-vendor personalization tools, such as multi-attribute collaborative filtering techniques. The data warehouse operates on a trusted and secure server. The server is able to provide information to vendors that enables models to be enhanced, and synergies between data sets realized, often without providing vendors with direct access to the actual **user profile**s (unless the **user** and submitting vendor has authorized such a disclosure.)

The central SDI-server maintains profile information for each pseudonym of every **user** registered with SDI, aggregated from information provided by multiple sources (e.g. vendors and clients). The server contains modules to perform data mining and collaborative filtering on profiles, and release results.

The SDI server ensures the integrity of data, and prevents data being used for unauthorized purposes. The server performs collaborative filtering analysis on the large data sets, and releases vendor-specific model enhancements. The central SDI server will also provide targeted mailing to a virtual mailing list of targeted **user**s, without releasing the pseudonyms to a vendor. The main SDI server also maintains an encrypted portfolio of pseudonyms and personal information for each registered **user**, and is configured to release that information to client-level proxy servers when a **user** logs-on to SDI. Another important roles of the SDI server is the categorization of vendors' business interests and data-handling policies.

6.1 Profile Updates

Figure 3 shows modes of data flow to the central SDI server.

SDI will accept updates to profile for pseudonym P by communication from the client-level SDI proxy server associated with pseudonym P. A **user** can log new information about one of its pseudonyms by sending a request R, with profile updates, to the central Secure Data Interchange server. The **user**'s identity is verified using public-key cryptography, managed at the client-level proxy. A **user** can also reveal the equivalence of a set of pseudonyms to SDI.

The Secure Data Interchange server will accept updates to the profile of a pseudonym from any party in the system that has registered with SDI. The **user** can request to see the profile for any of its pseudonyms at any time, to check that the information accurately reflects his/her behavior, interactions and transactions with vendors and other parties. A **user** request for a profile is initiated by a message M from the **user**. The message M contains a one-time key D, with which SDI will encrypt the profile. Only the **user** knows the key. The message M is encrypted with the public key of SDI, and then sent to SDI via the **user**'s trusted proxy server.

SDI verifies that the request for profile disclosure has been received from the authorized proxy server for pseudonym P, and then sends the profile, encrypted with the one-time key D, to the **user** via the proxy server.

SDI records a vendor ID when a party other than the **user** makes an update to the profile of a **user**'s pseudonym. The vendor must be registered with SDI, and make a signed commitment that it will only register information that is accurate and collected in good faith. Whenever a **user** challenges information, SDI can trace the vendor that provided the information.

6.2 Maintaining Data-use Guidelines

6.3 Use of Data

Vendors can check in data to SDI in any of the following forms:

i. Data or Randomized Data, with/without pseudonym
ii. Aggregate information on profiles in **user**-base, with/without pseudonyms

Suppose that a **user** has requested that a vendor does not reveal any information about the transactions that a vendor has executed with the **user** (under a pseudonym). The vendor can still reveal randomized information that will in no way identify a **user** (even though the **user** interacts under a pseudonym). We can perform system-wide collaborative filtering on randomized information (details below).

SDI can reveal information to vendors in a number of forms:

i. Data or randomized data, with /without pseudonyms
ii. Analysis results: For example

if a customer is interested in product X, suggest product Y
if a customer with profile type A hits your site, suggest

product Y
if a customer with profile type A is interested in product
X, suggest product Y
if customer P hits your site, suggest product Y
if customer P hits your site, and is interested in product
X, suggest product Y

The SDI server can provide analysis based on the profile of a **user** with pseudonym P without requiring that the vendor can access pseudonym P by providing a secure function evaluation procedure, that analyses the encrypted profile and provides results (details below).

The SDI server does not actually reveal any private information about the **user** to the vendor, so long as the results of the analysis are restricted to recommending products that the vendor has that might be just what the **user** is looking for. We protect the privacy of **user**s while providing personalization. The options above provide for a whole range of personalization, from very coarse-grained, to very fine-grained, but it is all dynamic and secure. Furthermore, we can provide the ability of a vendor to personalize service to **user**s that hit his/her site, but prevent the vendor from soliciting **user**s.

SDI can also retain control over information on **user**s by placing cryptographically enforced "expiration dates" on the ability of vendors to analyze profiles that are delivered with requests from **user**s. This allows sophisticated pricing models (see later section on pricing models).

There is a hierarchy of personalization that is available to each vendor in the system. As a base level, a vendor can request that SDI performs data analysis on its current **user** base, and with only transaction information collected at its own site. The next level performs analysis on the vendor's **user** base, but also drawing on profile information from across multiple vendors' sites. The next level allows vendors to personalize service to incoming **user**s on the basis of an analysis of the profiles of the new **user**s. Finally, SDI may provide a list of pseudonyms and recommendations, so that the vendor can target new **user**s.

6.4 Aggregating Profile Information from Clients and Vendors

6.5 Releasing Anonymous Profile Information for SDI-enhancedSearch

Figure 9 illustrates how anonymous profile-based search can be performed on ISP-level SDI proxies, with information provided from the central SDI database.

6.6 Data Mining/Collaborative Filtering Functionality

One of the core purposes of SDI is to provide a common location and format for information that has been gathered from a wide variety of sources and that might require different sorts of analysis. Since its framework is designed to generically handle different types of data and algorithms, SDI can be used as a platform to explore and exploit the rich connections that potentially exist within and across the databases of different vendors and customers.

This section describes in greater detail the types of data, analytical methods, and forms of validation that are available to SDI. The Secure Data Interchange architecture in its preferred implementation integrates the architecture issued U.S. Patent No. 5,754,939 "System for Customized Identification of Desirable Objects" into a system for secure data exchange between multiple parties. The aforementioned patent teaches a method for profiling objects and **user**s over a bi-directional distributed network, such as: an ISP, multiple ISP networks, a Web hosting network, or server software (such as data mining or recommender software) that is linked to a coalition of sites (such as a portal or Internet mall). The current invention, the system of Secure Data Interchange, allows correlations to be identified between vendor's data sets, that allows accurate profiling through the application of statistical methods, without providing vendors with explicit access to the profiles of **user**s – because profiles are provided in anonymized and randomized forms. There are less efficient methods that can be used to identify correlations, for example using customer demographics, and vendor categories, to suggest which vendors might be well placed to form dynamic syndication relationships. With SDI it is possible to leverage as many data sources as are available, about **user**s and the

target objects with which they interact. In fact the degree of the measure of improvement in predicting **user** behavior (or increasing clickthrough) is approximately in direct proportion to the square root of the number of **user profiles** and target profile interest summaries which are known. The emphasis in the aforementioned patent is on the bilateral relationships between vendors and **users**, and the architecture is not designed to support secure and privacy-protected data interchange and analysis across the **user** bases of different vendors.

In the system for SDI we push control of the profile for each **user** to the client software that runs on the machine local to the **user**, and provide for personalization through dynamic processing of information on the **user**'s client machine. We enable vendors to exchange data sets only to the degree that is mandated by **users**, and provide technical solutions to enable significant leverage of data while maintaining **user** privacy.

With SDI we enable second-level proxy servers for vendors to interact, and also make available the **user profile** information that is collected at the client-level and ISP-level SDI proxies to develop more accurate profile information, through the combination of "deep" vendor-specific information with broader network-level information. The ISP-level SDI proxies can also capture some information (in a privacy-protected form) about **users** and vendor sites that do not subscribe to the SDI service. The goal of the architecture is to continue to allow relevant filtering to be applied at sites that a **user** has never even directly visited before, or interacted with before, without that site accessing the profile information of the **user**.

The supporting architecture as stated in the above referenced patent also allows for profiling statistics to be collected and processed in such a manner that the information and vendor servers may both contain and implement the modules for profiling in a distributed manner. In the present invention the profile generation capabilities are implemented at various levels – in particular, at the second-level proxy servers of ISP networks and vendor servers, and within the central SDI data warehouse.

Web page tags include profiles of target objects, **user** quality ratings based upon overall quality as well as other criteria (e.g., value, price, entertaining, informative graphic/visual appeal, etc.) and data mining analysis of each criteria as it corresponds to target objects, target object attributes, **user** attributes (and vice versa), trend analysis statistics, location data (for target objects representing physical or geographical items). **User** information, in addition to profiles, can include data mining and trend analysis statistics, **user** provided ratings for target objects, and resolution credentials.

Vendors can use queries, data mining, menuing, and other techniques (such as described in the aforementioned patent), in order to gain access to desired **user profiles** that the vendor is allowed access to. The information that is stored against a **user**'s pseudonym in the central SDI server is checked against the data-release certificate that the **user** provided to releasing vendors, and is therefore data that the **user** has indicated can be stored on his/her behalf.

6.61 Structure of the Central SDI database

The central SDI server is structured as a relational database, with data that is submitted by vendors and **users** to the server indexed by pseudonym. The basic structure of a data record relates a pseudonym ID to a vector of numbers, representing the profile for that pseudonym. (Pseudonym_id, Profile). We also add tags to indicate the usage-restrictions on the profile information, and whether or not the data is randomized.

There are several types of information which can characterize both **users** and items. SDI is intended to function as the intermediary between a vast web of vendors, on the one hand, and individual consumers, on the other. The major sources of data used by this system are therefore:

1) Demographic. Such data will most likely be elicited by SDI from vendors and consumers when they initially register for the service, and details very general characteristics about them. It will consist of numbers and categorical values (age, zip code, sex, level of education, etc.).

2) Commercial. This is the kind of data any that vendor collects in the course of doing business (especially e-commerce); generally, it links customer codes to purchase items, dates, quantities, and prices. Depending on the nature of the business, this data could be fairly complex, and might well include text. For example, one could imagine that a bookstore, in addition to keeping track of its sales history, collects book reviews, author profiles, and plot summaries.

3) Behavioral (vis-à-vis the Internet). When a customer accesses the Internet through an Internet Service Provider (ISP), his ISP is in a unique position to monitor and record every behavior he exhibits while browsing the World Wide Web, writing email, and uploading/downloading files. This information is fairly complicated, as every site visit involves the creation or consumption of content (with associated sound, images, and text), the generation of queries (when using text-based search engines), and the navigation of hyperlinks. An additional complication is that such data is dynamic, in that differing amounts of time are spent for on-line activities and might well reflect the customer's personal preferences (e.g., a long time spent lingering between two choices in an on-line catalog might indicate an equal level of preference, whereas a speedy exit from a site might indicate general lack of interest).

Further information is provided by XML tags attached to pages browsed by the customer. The mere presence of such tags allows for correlations to be drawn between different web pages (e.g., a common XML tag used by travel-related sites), because it implies similarity. Furthermore, it is conceivable that such tags could encode more refined measures of a web page's content, such as browsers' evaluations of its value. For example, a web page of interest to scale modelers, in addition to having images and text related to model trains, might have an XML tag that shows that other scale modelers have given the web site a "five-star" rating. This page should therefore be given a greater weight when SDI is used to create correlations of interest to model hobbyists.

To fully represent these different sets of information, SDI handles collections of data of the following types:

1) numerical (e.g. an age, price, or period of time)
2) categorical (e.g. a color or musical genre)
3) text

A common task for SDI is to compare and correlate different customers, which might well be represented by mixed collections of numbers, categories, and blocks of text. This is handled by treating each customer c_i as a vector in a space whose coordinates correspond to the fields of data available. In the following description we refer to a customer, but when a **user** interacts with a vendor under a pseudonym, the profile information will only relate to information provided to the central SDI server for that pseudonym.

If there are m numerical pieces of data available, there will be n corresponding coordinates in the data space, $(x_1, \ldots, x_m)$.

For each category i, there will be a corresponding number of values, n_i. Hence, for a color category {red, white, blue}, $n_{color} = 3$. Since each value is assigned its own coordinate, category i is represented as an n_i dimensional vector, y_i. Hence, the total number of dimensions used to describe the full set of n categories $(y_1, \ldots, y_n)$ is

$$\sum_{i=1}^{n} n_i$$

Note that sparse methods are especially useful here, since a categorical vector y_i will typically consist of mostly zeroes, with a single non-zero coordinate representing the categories' value (i.e., we encode the color red, using the previous example, as (1,0,0)). Note also that category vectors with different values are treated as orthogonal by the system.

A final issue is the representation of text. As described in previous related patents, all relevant blocks of text in the database are converted into a dictionary that maps unique strings to the number of times they appear in the database. An appropriate TF/IDF weighting function is chosen and calculated for each of the p words that appear in the dictionary. The full set of text connected to a single customer can thus be represented as the vector $(z_1, \ldots, z_p)$, where each zi equals the number of times the word i appears in text related to the particular customer multiplied by the TF/IDF score assigned to word i.

In summary, when a database describes its customers using a combination of numerical values, categories, and text, customer i can be represented by the vector $c_i=(x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_p)$.

6.62 An Example Profile Vector

Suppose we have a database containing information on customers' ages, their musical preferences (i.e. an answer to a survey asking: "Which do you prefer, Mozart or the Beatles?"), and the contents of the emails they've written. Furthermore, suppose the only salient variables in all the emails written consist of the words "Beatles", "Mozart", and "practice", and that we are using the function

$$TF / IDF(x) = 1/\sqrt{n_x}$$

Where n_x represents the number of times word x appears in the dictionary. So, if the word "Beatles" appears a total of 217 times across the full set of customer emails, $n_{Beatles} = 217$, and

We now want to represent one of the customers in the database; he's a 10-year-old boy who prefers Mozart to the Beatles, and who wrote an email to his friend that mostly describes his attempts at practicing Mozart, but in passing mentions his sister's new Beatles CD. Suppose he uses the word Mozart 2 times (although it appears 456 times in the full database of all customers' emails), the word Beatles 1 time (appears 217 times in database), and the word practice 3 times (appears 77 times in database).

We define the following coordinates:
x_1 = age = 10
y_1 = {Mozart, Beatles) = (1,0)
z_1 = # of times customer uses word "Beatles" x TF/IDF("Beatles") = 1 * 0.067 = 0.067
z_2 = # of times customer uses word "Mozart" x TF/IDF("Mozart") = 2 * 0.047 = 0.094
z_3 = # of times customer uses word "practice" x TF/IDF("practice") = 3 * 0.114 = 0.342

In our example, then, we might encode this boy as customer 1:
$c_1 = (x_1, y_1, z_1, z_2, z_3) = (10, 1, 0, 0.067, 0.094, 0.342)$

6.63 Choosing an Appropriate Level of Data Granularity

We define the term granularity to denote the level of detail available within a given set of data, which is often structured hierarchically. Suppose a grocery store database contains records for a box of flavored gelatin powder. This could be categorized in a variety of ways; moving from the most specific to the most general, we might treat this data point as "12.5 ounce, strawberry flavor, Jello-brand gelatin dessert" (which would be entirely different from "12.5 ounce, banana flavor, Jello-brand gelatin dessert"), or as "12.5 ounce Jello gelatin" (a categorization which would treat as identical the strawberry and banana Jellos), or as "flavored gelatin", or as "dessert", or as "food", or as "grocery".

When analysis is performed on such data, the level of granularity chosen will have a strong effect on the outcome of the analysis. If the level of granularity is too fine-grained, the data will be too sparse, although it could be potentially aggregated to the next highest level of granularity. If the granularity is too coarse, the results of the analysis might be overly general (e.g., a customer would find a collaborative filter useless if the only recommendation it makes for a dessert choice is "go to the grocery section of the store").

Since the level of granularity will have a salient effect on the outcome of an analysis, it should be chosen very carefully, and might well play a factor in pricing when a vendor chooses to sell its data.

6.64 Methods used for data analysis

In order to perform a wide range of analytical tasks, SDI needs to make use of a variety of computational approaches. These are described below, starting with the simplest methods first.

- (1). Standard Database Searches

  Since most of the data will be stored in centralized databases, simple searches, queries, and data filters can be implemented by means of standard SQL commands. Typically, data will be collected or sorted using efficient database calls before being fed through analysis routines; once complete, the results can be fed back out to the database environment for further efficient manipulation.

- (2) Metrics – Measuring the Similarity Between Profile Vectors

  Given two customer (or vendor) profiles, c_i and c_j, it is frequently desirable to know how similar they are. For this purpose, we define the similarity metric $M(c_i, c_j)$ to be a function that takes as input two customer vectors and returns as output a numerical value in the range [0,1]. When two customers c_i and c_j are identical, $M(c_i, c_j)=1$; when they're completely different, $M(c_i, c_j)=0$.

  The problem is somewhat simplified by the fact that we treat all customers as vectors. Given two customer vectors, we can use the correlation between them to serve as our metric:

$$M(A,B) = \cos\theta = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

  Note that θ here represents the angle between the vectors A and B, and that we expect all coordinates of the vectors to be positive (in order for M(A,B) to keep its output in the range [0,1]).

  In more complicated cases, however, a customer vector might contain multiple fields with varying ranges of values. For example, we might have customer vectors of the form $c_i=(age_i, income_i)$, in which the maximum age is 80, but the maximum income is 300,000. In such cases, the coordinates with larger values will dominate the similarity metric, overwhelming any influence that smaller fields might have.

  This requires a normalization of the customer vectors, which can be done in several different ways. One approach would be to scale every coordinate by the maximum observed value, forcing all coordinates to lie between 0 and 1 (again, enforcing the rule that all coordinates must be positive).

$$c_i = \left(\frac{age_i}{\max(age)}, \frac{income_i}{\max(income)} \right)$$

  The only problem with this is that if a coordinate's maximum value is an outlier (being vastly bigger than the typical value), most of the coordinates' values will seem unusually small once they are scaled by the maximum. In such cases, it might be better to scale the values with a "squashing" function such as the sigmoid, which deadens the impact of extreme values; one such configuration would be the following:

SDI-0045-00045

$$\overline{age_i} = \frac{age_i - mean(age)}{\sigma_{age}}$$

$$\overline{income_i} = \frac{income_i - mean(income)}{\sigma_{income}}$$

$$c_i = \left(\frac{e^{\overline{age_i}}}{1+e^{\overline{age_i}}}, \frac{e^{\overline{income_i}}}{1+e^{\overline{income_i}}} \right)$$

Note that the mean and variance of the data points are used to fully normalize them, such that the sigmoid function will spread the values somewhat more evenly between zero and one.

The previous approaches are especially useful for single numerical fields, which might well overwhelm each other if some sort of normalization isn't performed.

A different problem arises for text or large categorical fields, since they can potentially consist of hundreds of coordinates capable of overwhelming the influence of single numerical fields. Suppose we believe the age of a customer is as important as the text of articles read. In such a situation, the thousands of coordinates devoted to the text field would dominate the metric's behavior, negating any influence that age would have on our measure of similarity – clearly not a good situation.

A solution to this would be to find the correlations among the fields taken separately, then average the result. That is, if each customer $c_i = (age_i, text_i)$, where $text_i$ is a vector with a very high number of dimensions, we could define the metric:

$$M(c_i, c_j) = \left(\frac{corr(age_i, age_j) + corr(text_i, text_j)}{2} \right)$$

Where

$$corr(c_i, c_j) = \frac{c_i \cdot c_j}{\|c_i\| \cdot \|c_j\|}$$

The result is a metric that gives equal influence to each field.

- (3) Forming Vectors Into Groups

The process of classification is essential to collaborative filtering, as it allows different vectors to be formed into groups based on some measure of similarity. If we are able to create groups of customer vectors, for example, we can then give individual customers recommendations based on the patterns of their group-mates, who presumably have similar tastes.

K-means Clustering and Nearest Neighbor algorithms are extremely useful for grouping purposes: previous iReactor patents give a full and detailed description of our customized versions. This section gives a brief overview of these methods.

(3.1) Clustering

K-means Clustering is an algorithm used to partition a coordinate space such that all vectors in a given partition are more similar to that partition's vector average (the centroid), than to the centroids of any other partition. It is a process that iterates over the following steps:

0. "Seed" the coordinate space with the initial centroids, which are vectors used to describe the centers of the clusters, in the sense that they are the average of all the vectors currently assigned to the partition. This can be done randomly (assigning centroids random coordinates) if no other information is available, or it can be guided by pre-existing information. For example, if we wish to cluster vectors of music customers, we can use information about musical genres to create initial partitions that correspond to pop, gospel, classical, etc. This will locate the centroids in well-spaced intervals across the coordinate space.

1. Assign vectors to the most similar centroids. This is done for each vector by scanning across all centroids and calculating similarity M(vector, centroid$_i$); once finished, the vector is assigned to the cluster whose centroid has the greatest similarity. In this stage, vectors may switch their allegiance from one centroid to another, if the relative distances to the vector have changed sufficiently since the previous iteration. If no vectors change their allegiance, the iteration process is complete, and the algorithm stops.

2. If the iteration is not complete, recalculate the centroids by setting them equal to the average of those vectors that have been assigned to them. Go back to step 1.

Once the algorithm converges, the vectors are grouped into clusters. The centroids' coordinates as well as the identity of cluster members is useful information that can be passed on to subsequent stages of analysis.

(3.2) Nearest Neighbor

The nearest neighbor algorithm, simply stated, creates a list of those vectors in a database that most resemble a particular target vector. This is accomplished by comparing the target vector, in turn, to every other vector in the database; the similarity between them is recorded, and once the comparison loop is complete the list of similarities is sorted. The top k members of this list are returned as representing those k vectors which most resemble the target.

- (4) Generalizing Across Databases

One of the most useful aspects of SDI is that it allows for inferences to be drawn across different databases through underlying connections in membership or content. An especially strong link can be made between commercial databases if they have customers in common. However, for reasons of privacy, individual customers may choose to use different pseudonyms when dealing with different vendors. This might be preferred by the individuals, but it weakens the inferences that can be made between fields occurring in different databases.

The techniques chosen to infer correlations across different databases will depend on how many pseudonyms are shared in common. At one end of the spectrum, every customer uses a single pseudonym for all transactions,

and makes an appearance in every database. At the opposite end of the spectrum, every customer uses a different pseudonym with every vendor, and may appear in only a single database.

Case 1: All customers use a single pseudonym, and appear in all databases considered.

This is the simplest situation to handle. Since all customers appear in all the databases, the customer vectors' fields are essentially scattered across several locations, but can be easily reconstructed. For each customer, we define a new data vector that concatenates that customer's representation from across the different databases.

Hence, if we are considering databases A, B, …, Z, and customer i appears in each one, we define a new vector $c_i = (c_{Ai}, c_{Bi}, \ldots, c_{Zi})$, where c_{Ai} is customer i's vector in database A. We then proceed as usual, making inferences with these augmented customer vectors.

Case 2: Most customers use a unique pseudonym, and frequently appear in different databases.

In this situation, although we see some connections between the databases, many pseudonyms appear in only a single location. Using Bayesian techniques, however, we can still make predictions for customer vectors across databases.

Suppose we have a set of databases, A, B, …, Z. Taking each database in turn, we cluster it using all available data. Thus, using every record in database A, we group A's customers into clusters

$A_1, A_2, \ldots, A_n$. Taking database B, we create clusters using all of B's information, creating customer clusters B_1, $B_2, \ldots, B_m$, and so forth.

Now, scan both databases for common pseudonyms (representing those customers who have interacted with both vendors under the same pseudonym) and create count variables w_{ij} to represent the number of pseudonyms that appear jointly in A_i and B_j.

$$P\left(B_J \middle| A_I\right) = \frac{P\left(B_J \wedge A_I\right)}{P\left(A_I\right)} = \frac{w_{ij} / total}{\sum_{j=1}^{m} w_{ij} / total}$$

$$total = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}$$

We can now produce the probability that a pseudonym appearing in A_i will appear in B_j:

For example, if we have a database of airline ticket purchases and a database of restaurant visits, we can create clusters, in the first case, of customers who travel to similar destinations, and in the second case, of customers who eat at similar restaurants. Given that a particular customer belongs to a cluster of people who frequent Caribbean restaurants, we can infer which travel packages would most appeal to him based on the linking probabilities, as defined above.

- Multivariate Extensions:

If we have a third database C, and there are a large number of pseudonyms common to A, B, C, the above probabilities can easily be extended. For example, knowing that a customer appears in A_i and B_j, we can calculate the linking probabilities to any C_k:

$$P\left(C_k \middle| A_i \wedge B_j\right) = \frac{P\left(A_i \wedge B_j \wedge C_k\right)}{P\left(A_i \wedge B_j\right)} = \frac{w_{ijk} / total}{\sum_{k=1}^{p} w_{ijk} / total}$$

$$total = \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}$$

Or, if there aren't many pseudonyms that span all three databases, the probability of C_k given that a pseudonym exists in A_i and B_j could be approximated by:

$$P\left(C_k \middle| A_i \wedge B_j\right) = P\left(C_k \middle| A_i\right) \cdot P\left(C_k \middle| B_j\right)$$

Case 3: All customers use several pseudonyms, and none appear in different databases

In this situation, there are no common customer codes that can be used to create links across the databases. However, the mere fact that several databases have been brought together for analysis should imply that there are semantic commonalities in the data.

Although each database contains different fields, it may be the case that those fields deal with related subjects. A human expert, knowledgeable in the content of the databases, the subtleties of the domain, and the overall goal of the analysis (e.g. the creation of recommendations), will be in a position to create a "common-information profile" that spans the databases. In essence, the common-information profile defines a format that allows vectors from different databases to share a common coordinate space.

The idea is this: the expert designs a high-level vector format that embodies the content deemed important for the project goals. Next, for each database he develops a mapping that encodes the database's elements into the generic format. Finally, the desired analysis is performed on the full set of common-information profiles.

Although the expert will have to create completely new fields for the common-information profile, certain types of data will map directly to the common-information format. In particular, if every database contains text (catalogued and counted, for TF/IDF purposes, by accompanying dictionaries), the union of the words will define the text coordinates of the new common-information profile. When word counts are being mapped from their original databases to the new vector, the original TF/IDF weightings may be used, or new TF/IDF weightings may be created (using a dictionary constructed from all the databases' text taken together).

Once analysis has been performed, certain common-information profiles will be grouped together by their shared similarities, although the pseudonyms they represent may have been originally drawn from different databases. Such groups will represent links between different databases, and may be used for predictive purposes (see end of example).

6.65 Example of Cross-database Analysis

Suppose we have the following databases:

A. A travel agency keeps track of tickets sold, and vacation web pages browsed.
B. A bookstore keeps track of books sold, and stores an electronic version of the New York Times Review of Books.
C. A sporting-goods and clothing shop, keeps track of purchase items sold (which includes magazines, for which electronic text exists).

A certain airline wants to promote various vacation packages it has available, which include both European and Caribbean vacations, as well as singles and family packages. Although it has leased rights to databases A,B, and C, it turns out that no customer pseudonyms appear in more than one database at a time – in other words, there are no shared records.

A vacation expert is hired to create a common-information profile. He creates the following information vector:
(list of tropical countries, list of European countries, family score, list of sports, text)

Note that the family score is a numerical value ranging from 1 (young singles) to 10 (many small children), and indicates what kind of person the customer is (a party-oriented student vs. a sedate father of three).

The expert the creates the following mappings:

A. Travel Agency. Link destinations of tickets sold to country fields (i.e., the number of trips to Germany by a customer would be placed in the Germany field of the common-information profile). Link sales of children's tickets, or requests for children's meals, to family score. Put web-page data into text field.

B. Bookstore. Link travel books' text to country lists. For all books purchased by a customer, map text from book reviews into text field.

C. Sporting-Goods store. Map warm-weather clothing (and swim gear) to tropical countries, ski gear to countries with skiing areas. Map sales of toys or children's clothing to high-value family scores, map revealing-bikini and student-discount sales to low-value family scores. Map text from magazines purchased by a customer to text field.

These mappings are then applied to each database, generating a full set of common information profiles. These are then clustered, forming groups that share commonalities.

The expert can now do several things with the results. First of all, he identifies the general "flavor" of each cluster (e.g., families with small children that enjoy winter, Europe, and skiing); the pseudonyms contained within each cluster can then be targeted for vacation packages suitable to their tastes. Secondly, the fact that pseudonyms from different databases have been clustered together allows the expert to plan cross-category marketing. If certain travel-book-buying parents have been grouped together with parents who bought their children swimsuits and scuba toys, it may be that they share a preference for family activities that take place in warm places or by the seashore. Hence, the book-users might be advertised various ocean-related sports goods appropriate for young families, and likewise the swimsuit-users might enjoy getting recommendations for travel books that describe tropical destinations that are especially fun for children. That is, if the goal is to cross-market items from A to customers in C, the most logical source of recommendations would be the people in A who have been grouped with the people in C.

SDI also allows dynamic and first-time personalization for a **user** that visits a vendor's site, on the basis of previous analysis and the **user**'s profile.

6.66 Methods for Validation

The main result of using SDI is the creation of connections between different data points, linking vendors and customers, customers and recommendations. To a large degree, the overall success of an SDI analysis is the relevance of the connections that are inferred from the data. It is often the case that a certain amount of validation is required to determine which analytical approaches are the most successful, given that the analyst has had to choose a particular combination from a wide range of algorithms, data sets, levels of granularity, and parameter settings. The process of validation measures the relative success of a given project, and is used to guide the analyst through further iterations of tuning and adjustment so as to optimize the final results of the analysis.

There are two general approaches, not necessarily mutually exclusive, to validation: the first is fairly quantitative, the second relies more on human expertise and intuition.

- (1) Quantitative Approaches

1.1 Test Against a Validation Set

Suggest the use of a human to infer the most relevant attributes influencing a particular output and the role of experimental design (as suggested) as a way of holding out part of the input data in order to validate the key relevancy of particular attributes. Principal components analysis is also implied in the last paragraph of 1-1.

The goal of validation, in this context, is to measure how successfully SDI makes a prediction, most commonly a recommendation. Before a recommendation system can be used commercially (when it is exposed to actual customers), it is important to make sure that it is using the best possible combination of algorithms, input data, and parameter settings (e.g. TF/IDF tuning). If several different combinations are under consideration, there is a need to gauge the relative predictive accuracy of one approach over another. This can be accomplished by holding out part of the data set, training the recommendation system on the remainder, then evaluating the strength of the recommendations made for the hold-out set.

Suppose we are testing two possible settings for a system that recommends music. We make a copy of the customer purchase records and remove a single purchase at random from each customer – this slightly reduced copy will serve as our training set. We then allow the two rival systems to recommend musical albums for each customer, based on the information in the training set alone. Typically, these recommendations will take the form of a list of items with corresponding numbers that indicate the strength of each recommendation. The relative performance of a set of recommendations can then be gauged by looping across each customer, noting whether or not the system recommended the item that had been held out, and if so adding it to a running total. The system with the highest total can thus be judged the most effective, since it most strongly recommended items that the customers did, in fact, end up purchasing.

Because the result of this type of validation is a quantitative score, it is possible to automate the model selection process. Given a set of analytical approaches (each with its own array of parameter settings), it is possible to loop through the full parameter space (using a grid of evenly spaced numerical values, if needed, to reduce dimensionality), computing a validation score at each iteration. Those combinations of algorithms and parameter settings that demonstrate the best performance could be chosen as the top candidates for the final system configuration, since they do the best job at predicting customer behaviors.

1.2 Dynamic Approach

The problem with the hold-out approach to validation is that it isn't dynamic, since it doesn't reflect the impact that the recommendation system has on the customers once it is implemented, and may be based on data that doesn't contain current trends. After all, it is better to predict what the customer will buy rather than what the customer has bought in the past.

A better approach is to run a controlled experiment against the actual customer base. First, the pool of customers is split at random into different segments. Next, each approach under consideration is used exclusively to make predictions for a given segment. Once the trial period is over, each system is given a score based on how valuable its recommendations turned out to be (this could be measured by total sales generated, for example, or by the number of times a customer made use of a recommendation).

- (2) Human Expert in the Loop

Although quantitative methods can automate the validation process to some degree, at the beginning of many projects there is so much raw input data available and so many decisions that have to be made about the analytical approach that an automated process would have to test a prohibitive number of combinations of data, algorithms, and parameter settings to get optimal results. In such cases, it is useful to employ a human expert who understands the psychology and nature of the particular domain being analyzed.

Such a person will have intuition about what is and isn't relevant for his domain. For example, a movie expert might be called in to work on a movie-recommendation system, for which an immense amount of input data is available. In choosing relevant fields for analysis, the expert's understanding of cinema would lead him to include the director's name and numbers of Oscars awarded, whereas the exact length (in minutes) of the movie would be, in his estimation, irrelevant and therefore excluded.

Once the analysis is complete and recommendations have been made, the expert's opinion (based on a qualitative understanding of the domain) can be used to guide which particular combination of settings, chosen from a list of candidates with detailed test outputs, should be used for the recommendation system.

- (3) Combined Use

There is certainly no reason why both approaches couldn't be used in combination. Many data sets include fields that are extremely noisy or simply irrelevant to a given problem; a human expert can be employed to pare the data set down to a reasonable size and dimensionality, using his domain expertise to create a data model reasonable for the proposed analysis. Next, automated methods can be used to fine-tune the parameter settings and to choose which subsets of the input data are the most useful. Finally, the human analysts called back to qualitatively evaluate the results of the fine-tuning, making the decision to either start a new iteration of the analysis, or to certify that the process is complete and ready for commercial application.

6.7 Conditions on Usage and Privacy

As previously described in this patent, various conditions can be placed on the way in which a set of data may be used (i.e., can the **user** make a personal copy of the dataset?), as well as on the privacy controls put in place. It might well be that a vendor is willing to share only a portion of his database, or that he will release only randomized aggregates in accordance with the level of privacy he has guaranteed his customers.

Although such restrictions could impact the content of the data analyzed by a vendor, as long as it is kept in an SDI-compliant format it can be analyzed by SDI's suite of tools.

However, the data that is stored in the central SDI sever still has tight usage restrictions, as placed on the data by the submitting vendor and the **user** that the data pertains to. For example, the **user** will have specified a use-of-data policy that could restrict the data to be used for only personalization purposes, or only solicitation purposes. In this case the **user** has connected to the site of a vendor, so this use-of-data restriction is irrelevant. The vendor will place additional restrictions on the data, for example the vendor might wish that the data that it submits is never used to personalize the service offered by a direct competitor, with the same classifier label. When this is the case the central SDI server must not release any of this data to the vendor, even if this vendor is not a direct competitor, because there can be no guarantees that the vendor will not pass the data on to a direct competitor.

The data within the central SDI server is therefore in two main classes, data that can be released by SDI because the vendors that have submitted the data have placed no restrictions on the other vendors that can access the data, and data that can never be released by SDI because the vendors that submitted the data have placed restrictions on the other vendors that can access the data. We can still make use of the restricted data when providing profiling information to non-competitor vendors. Periodically the central SDI server performs collaborative filtering for each vendor, on the pseudonym records within its **user** base that contain information that relates to the business of that vendor.

For example, consider a vendor that sells compact discs. The vendor will have submitted profiling information about the each **user** (represented with a pseudonym) that it has done business with. Furthermore, the central SDI server might also contain additional profiling information that relates to the same pseudonyms, that has been submitted by other (non-competitor vendors). Finally, there may be records about other pseudonyms that have purchased music - either in compact disk format or an alternative format - that has been submitted by vendors that are happy to have the information used to help directly competing vendors.

The central SDI server can provide the vendor with the benefit of all of this data without actually releasing data to any of the vendors, by providing the results of the collaborative filtering analysis, restricted to the vendor's own data model. For example, given the connection certificate, that certifies that a **user** with pseudonym P has just connected to the site, the central SDI server can look up the profile of the pseudonym, and make appropriate product recommendations to the vendor - recommendations based on the analysis that has been performed using all of the data that it has available.

The structure of a typical data record contains a number of additional fields to represent profile-usage policies. (Pseudonym, data, L, R, P, S). The 'pseudonym' is the public key and ISP-level proxy server IP address relating to the pseudonym, the 'data' field is the profile information for the pseudonym - in general a sparse vector of numbers, L contains vendor classifiers that are excluded from the data, R is a {0,1} bit that indicates whether any of the data is randomized, P is a {0,1} bit that indicates whether the data can be used for personalization of service to the **user** when he/she visits a site under the pseudonym, and S is a {0,1} bit that indicates whether the data can be used for solicitations to the **user** with the pseudonym.

6.8 Location of Data and Algorithms

Although SDI might have controlled links to vendors' databases, the actual information might not physically reside within the SDI system. One could imagine a vendor requiring the joint analysis of data that includes highly proprietary information (kept for safety behind the company's firewall) with slightly less-critical information belonging to another vendor that is stored by SDI. In this situation, the vendor would have the ability, as with a lending library, to "check out" both the secondary data and relevant algorithms from SDI, and to use them at their own site. Thus, although both data and algorithms might reside either at vendors' home locations or within the SDI system itself, the general analysis will work transparently across these boundaries.

6.9 Other Issues

For security reasons, the contents of databases may be injected with a small amount of noise. This prevents database **user**s from surreptitiously connecting database records to individual customers, yet maintains the quality of inferences made about the database in general.

Although such "noisy" records don't pose too much of a problem for those methods that make generalized inferences, it should be noted that recommendations made for individual customer vectors that have undergone such randomization will be less useful, since predictions are being made for a noisy target.

A final consideration is the reduction of the data vectors' dimensionality (which can be extremely high), since it is harder to make clean inferences about sparse data. There are many standard methods that can be used to achieve this, such as Principal Components Analysis.

Another approach is to adjust the granularity of the data, if at all possible. In a music store analysis, for example, there might be many more album titles that artists (since each artist can produce multiple albums). In such a case, purchases could be recorded by artist rather than by album, greatly reducing the dimension of the customer vectors' purchase space.

7. *Randomized Aggregates: Enhancing User Privacy*

Randomized aggregates provide a cheaper and more secure alternative to cryptographic techniques, such as secure-function-evaluation for providing information without compromising privacy. Even with pseudonyms it is possible that the identity of an individual can be revealed through revealing too much specific information about the transactions/profile of an individual. This occurs when the information places a strong restriction on the set of **user**s that could satisfy the revealed constraints. We enable vendors and **user**s to report information about pseudonymized **user**s in a randomized form that prevents this type of reverse-engineering to identify the **user** behind a pseudonym. We add randomization to data fields so that the **user**'s privacy is protected, but also so that the data still allows accurate aggregation and collaborative filtering/multi-attribute clustering analysis. Randomized data is also secure to computational attacks and the loss or theft of private keys-because we degrade the data, and make access to any one data item virtually useless.

This technique of randomized aggregates enhances **user** privacy guarantees, allows vendors to disclose useful information without violating **user**-privacy requests, servers to reduce system-wide reliance on cryptographic solutions. The basic idea is to add a small noise term to each field of a **user**'s profile. Aggregate data can allow economic evaluation of data without access to the data, although aggregate data can have some inherent value in itself. Randomized aggregates enable: (1) personalization; (2) aggregate statistics; (3) protection against profiling.

In many situations it is sufficient to gather information about the activities of an entire demographic group rather than about any one individual. For example, a VCR dealer might be interested in the chance that a person buying a new television set will purchase a VCR within the next twelve months. However, information about whether any one individual who purchases a TV also goes on to purchase a VCR within a year should not need to be revealed in the process of computing this chance.

Traditional cryptographic methods are capable of solving the problem of revealing only aggregate information while concealing individual information. Methods exist for computing aggregates or other values from encrypted information without first decrypting this information. (Such methods are dealt with in an area of cryptography known as secure function evaluation.) However, the general-purpose nature of these methods makes them unnecessarily cumbersome for the limited problem at hand. In particular, the communication and computation requirements of these methods when applied to the problem of aggregation result in an unacceptable overhead on the system.

There is also an additional problem with such cryptographic techniques. Secret information can be compromised by successful computational attacks on the cryptographic scheme or by the loss/theft of private keys. Such problems are present in all uses of cryptography. Nevertheless cryptography is used where it is the best alternative. However, once again, for the limited problem at hand our solution is safe both from computational attacks as well as from private keys being compromised.

The above-described problems are solved by the use of a new technique that randomizes sensitive data. **User** decisions and other numeric fields of records are modified by the addition of a "noise" term that may be positive or negative. These noise terms are chosen by sampling from carefully chosen probability distributions. Different, independently chosen noise terms are used for each field of the record that needs to be perturbed by the addition of noise. The modified values of these numeric fields are then transmitted to an aggregating database. The net result is that when very

few records have been aggregated the value of the numeric fields are likely to be quite unreliable. However the accuracy of these fields improves gradually as more and more records are accumulated.

As an example, consider the following situation. Suppose that we want to determine the average income level of people purchasing camcorders. In each record there could be a numeric field indicating the type of object that we are dealing with. For example, assume that this field is "1" exactly when the object in question is a camcorder. Now suppose that each of these records has another field that contains income information. If this information has been perturbed by random noise then any individual record in the bucket has income information that is highly unreliable. However, when sufficiently many records are aggregated the income information becomes more accurate! (This is a consequence of a theorem in probability theory called the Law of Large Numbers.)

As another unrealistic but illuminating example, consider the following scenario: Suppose an organization receives monetary contributions from a number of individuals, who for reasons of privacy, wish to conceal the amount of their contributions. Suppose that the total of all contributions is to be public knowledge. The randomization scheme works as follows: When an individual contributes x dollars, she chooses a random number r according to a specified probability distribution (with mean 0) and sends x+r to the aggregate database. By allowing widely dispersed values of r, we can ensure that an eavesdropper who obtains the value of x+r has very little information about x. However, if a large number of individuals register their contributions in this manner, then the total of the values sent to the aggregate database will be very close to the true total contribution. The actual scheme and its variations are more complicated as they have to deal with multiple fields and more sophisticated attacks on privacy. But the basic idea described in this example will be used repeatedly in these schemes. In this example, the total of a set of values was required to be public.

Confidential - HDSV96

7.1 Adding Noise to Fields

A record is a tuple of information containing various fields some of which are numeric, for example, describing the details of a commercial transaction. Noise or perturbation which refers to the random value added to individual numeric fields of records.

If the field is a continuous value, such as salary, then we can add a Normally distributed term with zero mean and a carefully selected standard deviation. The standard deviation is "tuned" to provide a good tradeoff between individual privacy and accuracy of aggregate analysis. For example, if an independent noise term is added to the salary field of a set of **user profiles**, and a vendor requests analysis of the mean salary over the set, then with a large noise term more **users** are required in order to generate an accurate average salary. In contrast, if only a small noise term is added to each salary field, and a third party with knowledge of the salaries of **users** requested data on each **user**, the third party would be able to match the **users** to the data fairly accurately, on the basis of the salary field-particularly for **users** with distinctive salaries (very low or very high).

If the field is one of a discrete set of items, such as the name of a CD that a **user** has purchased, then randomization proceeds slightly differently. In this case, one replaces the identifier with an identifier that is semantically close. For example, another CD of the same genre.

We can still perform correlation across fields with randomization, so long as the randomization does not destroy any trends between fields. Randomized data is marked as such within SDI, and labeled with the degree of degradation, so that SDI can be aware of the number of records to get relevant accuracy levels, and can report accuracy to customers.

We need to add noise to make data elements "close" to the accurate values. With discrete data, such as the name of an artist, "close" must be defined within the correct metric. The appropriate metric is such that a "close" value shares many of the same characteristics. For example, it is not appropriate to assign a close value on the basis of a shared last letter in the first name, but it is appropriate to assign a close value on the basis of an artist from the same genre of music. We call such clusters of artist names "semantic clusters", to imply that they have meaning in the domain. Semantic clustering that enables useful randomization of discrete field can be automated when goods are frequent purchase, high volume goods-where individuals purchase goods on multiple occasions, and more than one of the family of goods on a single occasion. High price, low volume goods, should be randomized on the basis of expert analysis (for example new cars, computers...) -- where an expert can extract key features of a purchase, and represent the purchase generically using either a single prototype good, or one of a set of approximately equivalent goods.

We do not require that every vendor/**user** uses the same distribution for its additive noise term. The choice can be made autonomously. All that is required is that a vendor specifies the degree of randomization, so that SDI can be aware of the type of data that it is aggregating and selling.

We now begin an elaboration of the variations possible to the basic scheme and the assumptions and requirements under which each variation would be appropriate.

Records are associated with **user** pseudonyms, not **user** IDs. All fields can be perturbed with noise, except the pseudonym ID field, and any database query can be transformed into a query on "fuzzy" fields. Queries are dynamically constructed, and it is necessary to be sure that the result set contains enough data points so that a **user**'s privacy is not compromised. Furthermore, we do not normally allow restrictions of pseudonym IDs.

The system can be configured and used in a number of different ways. Each transaction involves one or more parties. Most common transactions involve two parties, the vendor and the customer. It is assumed that the communication between these two parties involving the actual transaction is made secure by the use of a public-key cryptosystem or otherwise. This assumption will be valid in most electronic commerce schemes. Once the transaction is completed, the vendor or the customer is responsible for transmitting a record to the central SDI server. Again it is assumed that this communication is secure.

There are several mathematical issues to be resolved, foremost of which is the choice of probability distribution for noise. The system proposed here will be "tunable' for each application and it will be the application that will determine the appropriate noise distribution. Here we simply describe the issues involved in the choice of the noise distribution. An important requirement on the noise distribution is that it should have expected value 0 since we want the noisy total to converge to the true total as more and more entries are aggregated. The second consideration is the variance or standard deviation of this distribution. If the variance is made large (relative to the actual values involved) then each noisy value reveals little about the true value of a field in a record. Clearly this is desirable in order to preserve privacy. The drawback is that when the variance of the noise distribution is large, a large number of entries have to be aggregated before the sum of the noisy values approaches the sum of the true values. Thus there is a tradeoff between the level of privacy protection and the level of aggregation at which responses to queries become accurate. Once again, each application can determine the point on this trade-off curve most suitable for that application.

The choice of the degree of noise to add to a data field is a tradeoff between protecting the privacy of an individual and allowing robust profiling in the aggregate, and personalization at a pseudonym level. There is some information that can be gained from a "fuzzed" numeric value. Suppose that the noise terms are added from a Gaussian distribution, of unknown variance and zero mean. A third party can gain information on the likelihood that a fuzzed zero-one variable has a real value of one as follows: monitor a stream of fuzzed values and fit the most likely Gaussian noise distribution, or if the data is present in a database, just fit the most likely additive noise distribution. Then, given an individual fuzzed value x' and true value x, and an estimate of the statistics of the additive Gaussian noise distribution, it is trivial to compute the $Pr(X = 1 \mid X' = x')$. For example, if the noise distribution is distributed with mean 0 and variance 2, and the fuzzed value, x', is greater than 3, then it is more likely that the underlying true value, x, is 1 and not 0.

This is less of a problem when (a) the domain of the numeric attribute is large and/or (b) the variance of the noise term is large with respect to the domain and/or variance of the numeric attribute.

We can use a cryptographic technique to verify the distribution of noise that is added to data - and also to enable replay. A vendor must keep a record of the non-randomized data that is supplied. When generating a random perturbation, the vendor uses a one-way function f on the object X to generate a seed for a pseudo-random number generator. The pseudo-random number generator then generates a sequence of random numbers that are used to create the random perturbation from a well-defined algorithm. SDI can uses this technique to verify randomization, and audit a profile-updating agent. "Playback ability" - the ability to reconstruct the original record from a noisy version of that record is important for a number of purposes. An individual may want to obtain proof of a transaction for legal purposes and law enforcement agencies with appropriate warrants might want to examine original records. Each client-level SDI proxy S maintains a trapdoor function f (such as the RSA encryption-decryption function). When adding noise to a record the proxy uses the fixed fields of the record as argument x and computes the inverse of f when applied to x. (Note that third parties that do not have access to the trapdoor secret will not be able to compute this inverse function.) The proxy then uses this value as the seed for a pseudo-random number generator and uses the bits produced by the generator as the random bits used to produce the noise. With this scheme it is possible for the agent database to "playback" the noise perturbation and produce the original record from the noisy record. This playback scheme is optional and may be used in an application if the feature is desired.

7.2 Randomized Aggregates: A cheap alternative to Secure Function Evaluation

Randomized data also enables SDI to release data to third parties for analysis, but still apply its results to non-randomized data. The third parties cannot make use of data because they do not have the pseudonyms, and they cannot identify any of the **user**s. The models that they generate can be sold to SDI, matched with pseudonyms, and then rented to vendors.

There are two uses of randomized aggregates:

(a) Revealing aggregated data from a database does not reveal information that is specific to a particular **user**, and does not reveal information that is detailed enough to be useful to a vendor that wants to perform target advertising. It is information of the form "if you had the details, this information would be useful", and allows some calculation of probable economic value.
(b) When only aggregated data is required, then data can be transmitted securely in "fuzzed" form, and the aggregate date reported accurately given enough records. This is useful when the owner of the aggregated database is not trusted, or communication channels are not secure.

We need to check after a query that the data revealed does not compromise a **user**'s privacy. A check at the time of submission of new data to the database is not sufficient because: (a) initially, no data is secure, even in randomized form - we can avoid a bootstrapping problem by checking the results of queries; (2) there are an infinite number of queries that one would need to anticipate at the time that data is checked in. We can use a statistical test to establish the minimum number of records that will ensure safe revelation of aggregate information, based on statistics/simulation from a real set of data, or statistics over a typical **user** population.

We would like vendors that hold useful information to be able to trade that information. This requires establishing that the information will be useful without revealing it. The classic solution to this would be Secure Function Evaluation. Secure function evaluation of the expected economic value of receiving information from one database to enhance the information present in another database is computationally expensive. Not only must the inputs for the function to be evaluated be transmitted over the network, that is the full representation profiles, but the functions must also be evaluated cryptographically.

An alternative to full secure function evaluation would be for an interested party to request aggregate information. Ideally this aggregate information would allow the economic advantage of the precise details of the new information to be estimated, without allowing the economic advantage to be immediately realized. The problem with using raw aggregated information, such as what fraction of **user**s in your database that bought a TV also bought a VCR within a year, is that the requestor might already have information for many of the same **user**s. What is required is that the requestor also transmits a constraint on the **user** ids that he/she are interested in.

Here is a simple scheme for computing the economic value of new information, for Vendor 1 that is interested in acquiring data from Vendor 2:

i) Vendor 1 sends Vendor 2 Query: What is the average income of the **user**s in your database that have purchased a TV in the past 12 months, but are not in this list of **user**s. (vendor 1 sends vendor 2 a list of **user**s that it already has information about for this field).
ii) Vendor 2 can then send the aggregate information back.
iii) Should vendor 1 be interested in receiving some form of access to information about the aforementioned set of **user**s, then negotiation over terms of contract can proceed.
iv) If the negotiation is successful, then finally vendor 2 can send vendor 1 the information, authenticated in some way, and vendor 1 can check the accuracy of the aggregate information that was reported in step (ii).
v) Payment for the information is made contingent on the accuracy of the information provided in step (ii).

Another interaction could look like this:

i) Vendor 1 says "Do you have attribute X for these **user**s". <sends list of **user** ids>
ii) Vendor 2 reports "Yes, for 89% of them".
iii) Vendor 1 makes a deal with vendor 2, and they agree payment contingent that the data that vendor 1 receives is consistent with the aggregate information provided in step (ii), and verifiable.

Finally, third party certification is necessary to guarantee that vendor 2 is providing truthful data for the unknown fields of the records requested by vendor 1. This aggregate data approach will be useful when aggregate data is sufficient to perform an economic analysis on the utility of new information.

Aggregate information does have value. Consider a vendor that has performed some analysis on sparse data from his/her own database and is seeking to support his/her conclusions by performing a similar analysis on the sparse data of another vendor. This analysis could proceed with aggregate information alone, and have value in confirming trends observed in the vendor's own data. The vendor could use the analysis of data from a similar vendor to draw implications about his/her own database of **user** information. SDI must perform secure evaluation on information in this case. The two different methodologies for the determining the value of information prior to disclosure of that information are:

Aggregate information has no direct value to a requesting vendor, but can be used to assess whether the raw information will be useful.

In this scenario it would seem preferable to avoid the complexity of requiring a trusted third party to perform the evaluation. Instead the aggregate information can be provided directly to a requesting vendor, allowing the vendor to perform his/her own analysis. Aggregated information is always sufficient, because the only way that a vendor could use non-aggregated but randomized information would be to aggregate the information himself. We would never need to transmit randomized records to the requesting vendor, because randomized information is only useful in its aggregated form anyway. For example, the vendor might request aggregate information on the **user**s for which it already has information within its own database, or aggregate information only for new customers. Should the aggregate analysis by the vendor indicate that the information held by the disclosing vendor is valuable, then mutually beneficial terms of disclosure can be agreed between the requesting vendor and the disclosing vendor, and the disclosing vendor can send the requesting vendor that information in a non-randomized form.

Aggregate information has direct value to a requesting vendor, even without the raw data from which it is derived (as described above).

In this case we will require a trusted third party to evaluate the potential value of information held by Vendor B for Vendor A. This will require the specification by vendor A of constraints, and possibly a sophisticated algorithm by which to evaluate the utility of the data in the aggregating database. Note that the variables used within such an evaluation function must be limited to aggregate information (means and sums) because that is the only accurate information that can be derived by the trusted third party from the aggregating database.

The trusted third party responds to the requesting vendor, and the requesting and disclosing vendors are then free to pursue negotiation in order to establish mutually agreeable terms of disclosure. If the data that is finally requested by the requesting vendor is aggregate data, then that can be disclosed directly by the third party. Otherwise disclosure must be between the vendors themselves.

Google Exhibit 1025

7.3 Randomized Aggregates: Further Uses

Mutually agreeable terms of disclosure of that data can be negotiated between the requesting and disclosing vendors. If the data is aggregate form, it is disclosed directly by the third party vendor otherwise, the non-randomized data is disclosed by the appropriate vendor to the requesting vendor. The aggregate database may be used for any of the following purposes:

(1) The use of explicit queries submitted by a given vendor to determine whether another vendor (whose aggregated data resides in the secure database) possesses data which matches the criteria of the external request (typically in conjunction with certain stated constraints, as discussed below).
(2) A comparison evaluation of the vendor's aggregated data to determine whether and if so, which portions of the database (possessed by which vendors) contain data which is of potential relevance to the vendor's database.
(3) The generation of "virtual customer lists" which are determined to be of value to the vendor (through the evaluation process of either #1 or #2 above). The pseudonym proxy server may then be used to target the desired **user**s matching the requested criteria of the vendor based upon terms agreeable to the requestor and data providing vendor based upon a per-impression or per-transaction model (through use of a changeable pseudonym).

(4) If sophisticated formulae and/or esoteric algorithms are used where evaluation/processing of a significant quantity of a given **user**'s data is required, it may be preferable to confidentially convey these formulae and algorithms to the third party. The outputs from pseudonymous **user**s or the list of **user**s which achieve certain predefined outputs may then be conveyed to the vendor.
(5) Disclosure of randomized data by the randomized database to a vendor in order to develop formulae, algorithms and models.

As discussed further below, the value of the data to the vendor may be estimated by:

1) (if the prospective disclosed data is **user**s), the number of **user**s matching the relevancy criteria.
(2) (if the prospective disclosed data is statistical in nature), the degree of statistical confidence associated with it.
(4) (if available), the effective measurable degree of benefit resulting from previously disclosed similar data (if applicable, with similar statistical confidence) as to that of the present data to be disclosed.

We can also determine the number of other "undifferentiatably similar" **user**s that the **user**'s identity can be masked by such that they are unidentifiable from the **user**. This is easily estimated by comparing the randomized data set to a set of real data, developing a randomized version thereof and observing the degree of confidence in being able to predict the true identity of the randomized **user** profile of a given "**user**"(from the randomized data) as confirmed by the actual data set. The present system may also notify the **user** if/when the number of identical **user**s falls below a minimum predefined threshold to ensure this requirement

Doc ID: 6924595.2

8. *Techniques of Secure Function Evaluation*

While randomized aggregates address the **user** privacy concerns associated with enabling vendors to gain full access to **user profile**s for modeling and in order to develop algorithms that fully leverage the new information that can be collected on-line, there is another potential **user** privacy concern. In order to actually deliver ** personalized service ** to an individual **user**, it may not be sufficient to use that individual's randomized profile, or an individual might be reluctant to reveal the randomized profile that is associated with her pseudonym, if it does indeed carry useful information.

A technical solution to this problem is offered through secure function evaluation, where two parties can evaluate a function based on distributed information without gaining any information about the inputs other than what is revealed by the outcome of the evaluation. The result of this method is the secure automatic evaluation of the **user**'s profile using the algorithms and tools originally derived from randomized aggregates. The output is assured to be accurate by guaranteeing that both the **user** data and the vendor's imported function are secured to tampering by either party. The client-level SDI proxy server maintains a **user**'s accurate profile data, while submitting randomized profile information to the central SDI server.

In one variation, the secure function evaluation can be used to determine whether there is data in the databases of other vendors that will be useful to enhance the **user** and object profiles of a vendor. If another vendor's data can be enhanced by the type of data on the vendor's site (and amenable terms are agreed upon), the vendor can then provide the useful portion of his/her data. Secure function evaluation allows the evaluation of data without access to the data, so that a fair contract can be agreed between the parties. The evaluation and exchange (and associated negotiation) of the data, occurs on a trusted server or the SDI proxy server where the relevant data resides. Because this analysis, data collection, compilation and modeling procedure occurs securely and under the control of the third party operator or intermediary operating the distributed trusted server network (e.g. SDI) the integrity of the data as well as secure enforcement of the privacy policies of the parties which possess access privileges to use that data are assured.

Practically speaking however, some of the **user profile** data (e.g. direct mail orders) will be directly collected by the vendor (this is inevitably true for any purchase of physical products whether directly using a store card or mail order with the exception of anonymous physical mail described below) and/or some **user**s may not require that their data be protected by a pseudonymous proxy server. Accordingly, the secure function evaluation may also be used on behalf of the vendor to assess what portions of the data on the **user**'s database are of potential value to the vendor. This potentially could involve such database matching activities as a simple direct comparison of fields of data between the **user** and vendor database in order to determine missing entries and/or fields on the vendor database to more sophisticated inference of data fields and/or leveraging additional relevant statistics for supplementing a sparse matrix

for purposes of predictive **user** modeling purposes. This provides the **user** with a high degree of privacy, as no portion of the **user profile** needs to be revealed to the vendor.

The vendor criteria for selecting both **user**s and appropriate targeted ads (based upon their **user profile**s) from a profile database can be performed entirely autonomously by the use of the present secure function evaluation method. Once the secure function evaluation has identified and qualified the relevant data and also measured the predicted benefits to the vendor, the vendor may possibly choose to negotiate terms for full purchase of the **user** data which is specifically relevant to that vendor. This data can be down-loaded to the vendor's site to directly analyze his/her **user** behavior statistics. The requesting vendor's "representation profile" consists of the complete collection of target object profiles to be refined.

The first step is searching for and identifying data within remote vendor databases which are likely to be the most ideal candidates for cross advertising with that vendor. Secondly, the database(s) which contain valuable data are evaluated for the estimated degree of potential commercial benefit the data could provide. Thirdly, this value estimation and the data privacy policies of the disclosing vendor (based upon his/her interest in providing data to the requesting vendor) form the basis for terms, if any, for the data transaction).

One example of two different types of synergistic data bases include vendor databases and a distributed **user** database (e.g. on the proxy server). Large customer interaction and transaction object profiles on the vendor's site are likely to be much more robust that those collected by a **user** side profiling system distributed over a multi-server network, or even multiple networks. From the vendors perspective the breadth of domains containing products that can be potentially cross correlated from data on the **user**'s database is very valuable. Vendors can hope to reach customers who have visited other similar sites and sites which tend to be visited by similar **user**s. Thus, the ad network is able to utilize the **user** database (containing virtually all of the data about what sites and target objects a **user** accesses) in order to effectively identify other vendors whose products and content are most frequently accessed by the most similar **user**s(and types thereof), suggesting the best target customers for each other's advertisements.

The system can contact the remote vendor requesting secure access to his/her data through secure function evaluation in order to determine the estimated degree of statistical benefit. The vendor may choose to disclose data in randomized form in order to enable the ad network to improve its data model while avoiding disclosure of data for purposes of targeting his/her customers. If the vendor lacks installed cluster code in his database, the system automatically installs the code as part of the secure function evaluation. In order to save overhead of re-installation the next time that a request is made to access the vendors database or to update the present model the system requests the **user** for permanent installation of the secure function evaluation and associated analysis code. This enables the automatic updating of a distributed cluster model in accordance with dynamically changing information and **user** behavior patterns. The cluster code is actually loaded onto the remote site, the analysis by the secure function evaluation may also determine the estimated potential benefit achievable to the remote vendor (or **user**) associated with accepting the request for his/her data. This may be achieved by comparing previous similar situations in which disclosure of data occurred, i.e. the marginal measurable commercial benefits resulting from previous data sets with similar degrees of statistical sparseness including the quantity of pre-existing and newly introduced data within a similar domain or cluster to the present one. The degree of estimated commercial benefit resulting from disclosure of the requested data may be then calculated. A separate estimation is then made if the remote vendor agrees to disclose his/her data (in non-randomized form) for purposes of targeting ads to customers on his/her site (or email list).

Depending upon the existing degree of statistical confidence which is estimated between each target customer of the remote vendor and the target objects of the advertising vendor, for the most part, it is likely that utilizing techniques which improve the statistical confidence between **user**s and target objects of these metrically similar vendors will yield significant commercial benefits by improving the relevancy of the matches. It is also possible for a data privacy policy to be submitted in conjunction with the release of any data. The vendor may desire that certain conditions of release be associated with all or each portion of the data, in the form of explicit restrictions of usage which may be assigned and tagged upon its release by the vendor. Adherence to these restrictions is the responsibility of SDI.

These data privacy policy restrictions may indicate which other vendors/vendor types (if any) may receive the advertising benefits resulting from that data, what portions of that data can be utilized for the benefit of each vendor/type. Also the pricing structure may vary in accordance with the type of data released, and who the vendors are who will ultimately benefit from its incorporation. Additionally, the usage of the data may also be subject to controls set forth by the contributing vendor. For example, applying different restrictions to different portions of his/her data, restricting its modeling benefits to his/her own company exclusively, limiting the benefits to non-competitors or certain vendors, limiting it's usage to only certain advertising types (or purposes), restricting its usage to individual

impressions or sales (on a per impression or per transaction basis). Data may only be used on a per-impression or per-transaction basis.

Because of the restrictions which vendors may place upon the usage (what portion of their data, to whom and/or for what purpose), re-clustering the data model separately according to various degrees of disclosed statistical data can be computationally expensive. We can use the same cluster model for portions of the model with common disclosure policies. Each portion of the vendor's resulting data model with the same disclosure policy is called a "common disclosure unit" or "CDU". Delineating CDUs within the data model involves continuous monitoring and identifying those portions of the data model (attributes and associated metrics). The CDUs, whether common access, limited common access or totally "private", operate seamlessly with CDUs that are subject to less restrictive constraints.

It is also worthwhile to mention that more simple merging techniques may be used for enhancing remote marketing databases by adding or filling in data fields to a sparse data model. This may be explicit data from matching fields, or inferred data, e.g. additional products or attributes which statistically correlate with certain other products or attributes or other **user** attributes. The system can notify vendors of the presence of potentially useful data within the **user** database.

Secure function evaluation can also be used to retrieve target objects from secure private databases (or Web sites) that would otherwise be inaccessible. SDI can identify whether there are any metrically similar target objects in the remote database that match the target objects currently retrieved.

We use the cryptographic technique of secure function evaluation to evaluate the encrypted profiles of **user**s that pseudonymously access a vendor's site, but prevent the vendors from accessing the profiles directly. The profile-analysis package is tuned for an individual vendor by analysis of the vendor's data model in SDI. The analysis is provided to the vendor on a restricted basis by requiring that (a) the proxy servers request a new encrypted profile for each **user** every day; (b) the vendors request a new secure function evaluation module every day (or a new key), so that vendors cannot use the analysis for more days than they pay. Secure function evaluation computes results from encrypted information without decrypting that information.

SDI-1005-0059

9. *Leveraging Existing Standards*

The architectural framework outline above can be implemented with a number of existing technical methods. In this section we outline one possible approach, that uses the Extensible Markup Language (XML) to encode web information with meta-tags (that can represent profile information). The Java virtual machine shipped with current Internet browsers can be used to run personalization code, that determines dynamically the information that should be presented to a **user**.

We can use XML to embed the profiling information for products that are offered by vendors, and information that is offered by information providers, directly into the page itself - with semantic labels that allow client-side processing with a Java engine. The Java Virtual machine, implemented on the browser of a **user**'s client machine, takes as input the XML-tagged page of a vendor, and the locally stored profile information that pertains to the **user**'s current pseudonym, and generates personalized content to display on a monitor local to the **user**.

Compare this architecture to a traditional client-server based solution, where the server produces personalized content for the **user**, content that is pushed to that **user**. Such a system architecture requires that the server has direct and explicit access to profile information about the **user**, information that can then be exchanged with other third parties and is out of control of the **user**. Secure Data Interchange collects and distributes information consistent with **user**- and vendor-defined policies. Information can be explicit data, including transactional information that is collected by parties that are involved in a transaction—such as the product purchased, and demographic information (e.g. gender, zip code, occupation). Information can also be implicit data, that includes click stream data that logs the information that a **user** requests and views, and the time-sequence of hyper-links that a **user** follows as he/she browses across multiple web sites. Profile information is embedded within a web page as metadata, that is data about data - machine readable information that informs an intelligent agent (such as an SDI-enabled browser) about the data that is included in a web document. The eXtensible Markup Language (XML) proposal of the Worldwide Web Consortium working group on SGML provides an ideal standard for representing such information [XML].

XML allows meta-content to be included with documents, machine-readable information that enables documents to be processed by client software. Augmenting web documents with structured information in SDI enables clients to

perform **user** personalization - pushing computation to clients, and allowing greater control over **user**-profiles because profiles do not need to be released from clients. XML can represent rich data structures, and that allows a grammar to be defined for information that allows data to be automatically verified for correctness [SGML].

The ability to embed data within web pages allows client-side processing of information. By embedding profile and location information directly within a web document we can alleviate the bandwidth and computational bottlenecks that can occur at a centralized profile server if profiles are fetched on-the-fly when web pages are downloaded by clients. The origin server (supported by the vendor) requests periodic profile updates from the central SDI server. This duplication of information enables the profile and the page contents to be provided directly from a vendor's server.

There are some potential drawbacks of this approach: (1) the profile information associated with a web page and target objects can be out-of-date; (2) the profile information is available to all clients and proxy servers, not just those that are SDI-enabled; (3) the profile information can be altered. We suggest technical solutions to each of these problems below.

9.1 Periodic Update of Web Page Profile Information

The central SDI server provides profiling information to vendors that subscribe the SDI. A vendor sends a 'Request Profile Update' message to the SDI server, to instruct the SDI server to send new profile information. The SDI server responds with a 'Profile Update' message, that contains updated profile information, generated from explicit and implicit data that it receives from vendors and **users** in accordance with privacy policies. The request-response mechanism can be implemented using the standard HTTP Post/Response mechanism in conjunction with XML message types. For example, the server 'Request Profile Update' message can be represented in XML as:

```
<?XML version = "1.0"?>
<?xml:namespace ns = "http://www.sdi.com" prefix = "SDI" ?>
<!doc>
<SDI:Request> http://www.some_vendor.com </SDI:Request>
```

and the SDI server 'Profile Update' message can be represented in XML as:

```
<?XML version = "1.0"?>
<?xml:namespace ns = "http://www.sdi.com" prefix = "SDI" ?>
<!doc>
<SDI:Update>
      <SDI:Profile>
            <SDI:Item> (1231, 0.453) </SDI:Item>
            <SDI:Item> (1041, 0.034) </SDI:Item>
      </SDI:Profile>
</SDI:Update>
```

An illustrative Document Data Type (DTD) for an SDI:Profile element type is presented in the next section. The XML messages are included in the body of standard HTTP Post/Response messages. We limit the performance degradation caused by out-of-date profile information that is stored within web pages of on-line vendors by associating "out-of-date" time stamps with the profiles that are provided by the central SDI server. This mechanism is similar to the "expiration time" tag of a Netscape Cookie message. The frequency with which profile updates need to occur will depend on the speed with which profile information changes. The "out-of-date" time stamp can be included as an additional element in an SDI:Update message.

Vendors request new profile updates when the current profile information is out-of-date, and more frequently if required (although we allow for a per-update charge). This is a "pull" model for profile-updates. An alternative architecture for profile updates is a "push" model, where SDI periodically sends new profiles to be added to the web pages on a vendor's server. The "pull" model is our preferred model because the responsibility for maintaining current profile information is decentralized, resting with the vendors in the system.

The system as outlined above can be implemented within the current HyperText Transfer Protocol (HTTP), as a sequence of challenge/response pairs between clients and servers. The HTTP Post/Response mechanism allows clients and servers to exchange data, and this data can be an instance of an XML Document Type, within the body of a HTTP message. The HTTP protocol is the underlying mechanism, with SDI messages contained in the body of the HTTP Post and HTTP Response as XML documents. In one variation of SDI the profile of a **user** is maintained on the **user**'s client, and partitioned into separate profiles for each pseudonym that a **user** chooses to maintain. Personalization of products and services (product types, prices, etc.) is performed at the client, through the execution of trusted code that is embedded as a Java applet or as JavaScript within the web document of a vendor. In this way a vendor never receives access to the profile of a **user**, but is nevertheless able to personalize its response to **user**s, even when a **user** first visits a site (on the basis of the profile for a **user** from his/her previous online transactions). Profiles for the target objects of a vendor that enable appropriate objects (representing particular products, or news stories for example) to be presented to a **user** are embedded as XML data within the vendor's web document.

In another variation of SDI personalization is not performed at the client, but either at the ISP-level SDI proxy server or the vendor's server. The location and other profile information that relates to a **user** are pushed to the ISP-level proxy or vendor server when a **user** requests a web page. In the same way as XML allows profile information about web sites and vendor products to be associated with a web document, and profile information to be provided from the central SDI server to a vendor, XML can be used to encode a **user**'s profile. The system of SDI allows for profile and location information to be randomized slightly (and even anonymized) to protect the identity of a **user**, for example when an ISP-level proxy is not trusted.

Confidential

9.2 Example: A Possible XML Representation of a **User profile**

The World Wide Web Consortium (W3C) SGML working group developed XML (extensible markup language) to provide an open and extensible grammar for structured data [XML]. An XML document has an associated schema definition to enable an XML-enabled browser to validate the structure of XML data automatically. A Schema in XML is called a Document Type Definition (DTD), and defines the names of tags, their structure, and their content model. XML allows the DTD for an XML file to be identified through a Universal Resource Indicator [URI] in the header of the file (see below). XML also allows URIs for mobile code resources to be referenced, in order to enable a client to process embedded XML data. An XML document must be well formed, and in order to be well formed the tags must form a tree structure. In addition, the DTD allows the structure of an XML document (an instance) to be validated against a particular schema. Senders and receivers must only send valid SDI files. Each SDI message is a valid XML document.

We provide an example XML instance and part of a Document Type Definition for use within the system of SDI. Profile information, as generated automatically through collaborative filtering techniques (for example, see issued US Patent #5,754,939) can be represented as a list of attribute-value pairs within an XML document. An attribute is defined by a numeric code, and the value defines the weight of the attribute. For example:

```
<?XML version = "1.0"?>
<?xml:namespace ns = "http://www.w3.org/OPS/OPS" prefix = "OPS" ?>
<?xml:namespace ns = "http://www.sdi.com" prefix = "SDI" ?>
<!doc>
<SDI:ProfileData>
   <SDI:Location>
         <SDI:Geocode> 12321561 </SDI:Geocode>
         <SDI:DigiMap> http://www.digimap/?12321561 </SDI:DigiMap>
         <OPS:Zip> 19103 <SDI:/Zip>
   <SDI:/Location>
```

```
<OPS:Demographic>
   <OPS:Gender> F </OPS:Gender>
   <OPS:Age> 26 </OPS:Age>
   <OPS:Income> 50000-75000 </OPS:Income>
</OPS:Demographic>
<SDI:ID>
      <SDI:Pseudonym> P12543 </SDI:Pseudonym>
      <SDI:PublicKey> 12453246129421 </SDI:PublicKey>
</SDI:ID>
<SDI:Profile>
   <SDI:Profile-item> (1242, 0.546) </SDI:Profile-item>
   <SDI:Profile-item> (56,   0.045) </SDI:Profile-item>
</SDI:Profile>
</SDI:ProfileData>
```

The Document Type Definitions for this document are specified in the header, and include URIs to a DTD of the Open Profiling Proposal of the W3C, and also a DTD of the Secure Data Interchange. The OPS DTD is used to boot strap the SDI DTD, providing tags for common profile information, such as 'Gender', 'Age', 'Income', etc. The section of the SDI Document Type Definition that is used in the above XML fragment is presented below. It makes reference to tags defined in the OPS DTD, and the RDF (Resource Description Framework), a W3C proposal to standardize the structure of DTDs for XML documents. XML Name spaces [NS] provide a method for unambiguously identifying the semantics and conventions governing the particular use of property-types by uniquely identifying the governing authority of the vocabulary, for example OPS and SDI in the example above. The URI for a schema can contain a human and machine-readable description of an XML schema.

```
<!ELEMENT SDI:ProfileData (SDI:Location?, OPS:Demographic?, SDI:ID?,
SDI:Profile?) >
<!ELEMENT SDI:Location (SDI:Geocode?, SDI:DigiMap, OPS:Zip?, OPS:Address?) >
<!ELEMENT SDI:ID (OPS:Name?, SDI:PublicKey?, SDI:Pseudonym?) >
<!ELEMENT Profile RDF:list<SDI:Profile-item>  >
<!ELEMENT SDI:Geocode #PCDATA >
<!ELEMENT SDI:Digimap #URI >
<!ELEMENT SDI:PubicKey #PCDATA >
<!ELEMENT SDI:Pseudonym #PCDATA >
<!ELEMENT SDI:Profile-item (SDI:Attribute-ID, SDI:Attribute-value) >
<!ELEMENT SDI:Attribute-ID #PCDATA >
<!ELEMENT SDI:Attribte-value #PCDATA >
```

The tag '#PCDATA' is used here to represent numeric or textual information, '#URI' declares that an instance of element 'SDI:Digimap' must be a valid URI pointer.

9.3 Maintaining the Integrity and Security of Embedded Profile Information

The privacy of information in transit between servers and clients can be assured through standard end-to-end cryptographic solutions that establish a secure session prior to any data exchange, such as Secure Sockets Layer (SSL) that uses X.509 certificates and is supported by current browser technology.

In order to prevent the possibility of individual **users** being bribed by vendors to disclose target object profile data which reflects this type of information, **users** should not be provided access to directly decrypt the metatags for these portions of the target objects profile data, but rather this decryption and release of profile data should be performed securely in conjunction with the functions of the profile processing (profile matching module) upon the client level proxy server rather this decryption and release of profile.

In addition, we prevent unauthorized access of embedded profile information through the encryption of the metadata that is represented within the XML structure of a web page. Profile information can be encrypted using a hierarchy of keys, so that different levels of access to the information may be provided according to the access levels of **users** and

vendors. All **users** that request web pages from SDI-enabled vendors, whether or not the **user** is a member of SDI receive the same profile information. We provide encrypted profiles to vendors in the 'Profile Update' messages from SDI to vendor servers, so that: (a) unauthorized agents cannot tamper with the profiles; (b) the profiles cannot be read by unauthorized agents.

The SDI system supplies a private key to trusted SDI client software, that enables only SDI-enabled clients to access profile information, and only access that information to the extent permitted by privacy policies of **users** and vendors. Different levels of encryption enforce multiple levels of access. Periodically the key pairs are changed to prevent extended attempts at cryptographic attacks. The SDI system uploads the key that provides the correct level of access for a **user** to a **user**'s client, once terms of access and profile management have been agreed. A client can only access embedded information once enabled with a relevant key. Finally, profile information is signed with a digital certificate, to prevent third parties from tampering with profiles for commercial gain.

9.4 Related World Wide Web Consortium Proposals

The proposal for a Meta Content Framework (MCF) suggests a particular structure for the description language for web pages, to enable schema to be shared and re-used [MCF]. This proposal is incorporated into the W3C Resource Description Format standard [RDF]. The proposal for an Open Profiling Standard [OPS] describes a system for profile exchange between two parties, building on XML and MIME standards. The W3C SGML working group has defined XML to provide an open, extensible grammar for structured data. The proposal on privacy and profiling on the Web [PRIVACY] extends the vCard [VC] schema for electronic business cards to include profile information, and suggests that profile information can be stored and managed locally, with Client-server exchange of personal information as required (using the HTTP challenge/response mechanism).

The Resource Description Framework (RDF) enables the encoding, exchange, and reuse of structured metadata. RDF is an application of XML, with additional constraints to allow for DTDs to be published, and interchangeability across different communities. The ability to standardize the declaration of vocabularies will encourage the reuse and extension of semantics among different information communities [Mil98]. RDF is a W3C proposed standard for defining the architecture necessary for supporting web metadata. RDF is an application of XML that imposes structural constraints to provide unambiguous methods of expressing semantics for the consistent encoding, exchange, and machine processing of metadata. RDF additionally provides means for publishing both a human-readable and a machine-processible vocabularies designed to encourage the exchange, use and extension of metadata semantics among disparate information communities.

10. Support for E-Commerce Functionality

10.1 Generation of Mailing Lists

We can use the same profile information that provides focused/personalized service to **users** that hit a site that they have not visited before to form well-targeted mailing lists for vendors. The Secure Data Interchange can form mailing lists in a number of different ways.

First, consider a vendor that wishes to send targeted mail to some of its own **user**-base. When **users** connect to a site they indicate whether or not they are willing to receive electronic mail, and provide a "mail certificate" to a vendor if they are happy to receive mail. The Secure Data Interchange can proceed as follows:

(a) perform analysis for the vendor to determine an appropriate set of **users** to receive the solicitation, based on the information that the vendor provides about what it intends to market, and provide the list of pseudonyms to the vendor for mailing; (b) perform the same analysis, but also forward the communication to the **users** directly.

Now, consider a vendor that wishes to target new **users**, represented with different pseudonyms. **Users** indicate whether the information that a vendor submits about his/her transactions may be used for solicitations, and furthermore vendors indicate the set of business interests that can receive the benefit of information that is submitted to the central SDI server. The SDI server can continue by performing analysis on the relevant subset of the permitted class of data

records that pertains to the product or service that the vendor wishes to model, and generate a list of appropriate pseudonyms. Finally, the SDI server can sell the pseudonyms to the vendor outright, together with a certificate that the vendor can send mail to the pseudonyms, or the SDI server can retain control by sending the mail on behalf of the vendor.

Provide vendors with virtual mailing lists that can be mailed to via the proxy server only. i.e. these customers should be solicited based on our analysis. (could even give summary info., without revealing details about **user**s). Furthermore, do not even reveal data that corresponds to a pseudonym to a vendor because the vendor then has that information about me when I am on his site.

A central data warehouse also enables vendors to identify new potential customers. This process is broken down into a number of steps:

(i) The vendor assesses the value of the information present in the secure data interchange. This computation is performed securely either by revealing randomized aggregates to the vendor to enable its own local analysis, or by allowing the vendor to check data and algorithms into the secure data interchange site for analysis.
(ii) The vendor selects criteria for mailing unsolicited advertisements, and agrees on a pricing model. In this case per-impression pricing is the most obvious pricing model, as it is difficult to monitor when a **user** responds to unsolicited mail per-transaction pricing is difficult. The **user** could be motivated to do this should the Secure Data Interchange promise future returns for recording a successful solicitation with the database.
(iii) Either the data list is released to the vendor for its use, if this is within the selling vendor's data policy, or the data interchange sends mailings on behalf of the purchasing vendor.

Google Exhibit 1016 - Page 66 of 95

10.2 Physical Mail

(a) Vendor to **User**

Figure 13 shows a flow chart for the process of sending physical mail from a vendor to a pseudonymous **user**.

A vendor must hold a "physical mail certificate" to be able to send mail (packages, letters) to a **user** under a pseudonym. The certificate is similar to the "electronic mail certificate", in that it is signed by the private key of the **user**'s pseudonym, and indicates that the vendor with public key P*V can send mail to the **user** (under the pseudonym).

Each **user** has a trusted physical address authority, just as it has a trusted electronic mail authority (the second-level proxy server), that maintains the physical mailing address for each pseudonym. When a vendor has a letter X to mail to **user** with public key PKP, the vendor generates a unique ID for the package, IDX, and sends the ID code and the physical mail certificate to the trusted physical address authority of the **user**.

The physical address authority receives the certificate, S((PKP, PK*V, SEND_MAIL), SKP), that indicates that the vendor is authorized to send mail to the pseudonym, and the packages identify code, signed by the vendor to certify that the vendor holds the secret key that matches the public key in the physical mail certificate.

The vendor then passes the letter X and the signed ID code to a trusted mailer, that supports pseudonymous mailing, and has been certified by the central SDI server as such. The trusted mailer then provides the signed ID code to the physical address authority, signed with the private key of the trusted mailer. The physical address authority verifies that the trusted mailer is a valid service, and releases the real address of the **user** to the mailer. The mailer now has the letter X that the vendor wants to send to the **user** with pseudonym P, and the physical mailing address of the **user** - and the package can be mailed. At not time did the vendor determine the true mailing address of the **user**, unless it works in collusion with the trusted mailer, but the trusted mailer is certified by SDI, and also audited by the chosen physical address authority of the **user**. The address authority will only release addresses to reputable pseudonymous physical mail agents.

We can operate physical mailing lists in the same way, and gain additional security by never releasing the pseudonyms or the mailing addresses to the vendor that has requested the targeted solicitations. We can use a technique that is similar to the technique that we used for virtual mailing lists. The vendor describes its solicitation to the central Secure Data Interchange, which leverages as much data as possible (without violating the privacy policies of any of the **user**s or vendors that are represented within the data). The central

SDI server generates a list of suitable pseudonyms, and then provides a series of unique codes to the vendor, that the vendor can supply to its chosen pseudonymous mailer with the material that is to be mailed. The central SDI server also provides the appropriate address authorities with authorization to release the physical mail addresses to the mailer when presented with the IDs. Notice that at no stage did the vendor have the pseudonyms or the mailing addresses. The parties all have only as much information as is necessary - the vendor needs someway to identify its packages to the pseudonymous mailer. The mailer needs an identifier to present to the address authority, and receives the addresses. The address authority just needs to know what addresses to release and to which third party.

(b) **User** to Vendor mail

The Secure Data Interchange system also provides a mechanism for **user**s to send physical mail to vendors that are registered with SDI with pseudonymous return addresses. In particular, when a **user** sends mail to a vendor, the first-level proxy server provides a tool that: (1) Computes/Looks-up the appropriate pseudonym for the **user** with this vendor. (2) Generates a unique ID, and submits a signed message to the central SDI-server, where the message relates the pseudonym, the vendor, and the ID. (3) Provides the unique ID to the **user**.

The **user** writes the unique ID on the envelope, and mails it to the vendor. Should the vendor wish to reply to the **user**, then the vendor can take the envelope to a pseudonymous mailer, and request that the envelope be mailed appropriately. The pseudonymous mailer verifies the identity of the vendor, and then submits the ID, together with the vendor's signature, and its own signature, to the physical address authority that is maintained by SDI. SDI releases the address to the mailer that can then return the mail.

10.3 Pseudonymous Payment Mechanisms

The Secure Data Interchange architecture must be able to support all the standard electronic commerce functions that we take for granted, but while maintaining pseudonymity for **user**s and following privacy policies. There are various different solutions to this problem.

(a) Anonymous Credit Card Payment

The second-level proxy server can maintain information on the **user**'s credit card information, and perform the following transaction. Whenever a **user** makes a purchase from a vendor, the **user** provides the vendor with authorization to bill $x to his/her credit card account, but anonymously - through the Secure Data Interchange as a middleman. The **user** generates a unique number, Y, and signs a "right to payment" message, M=($x, PKP, PKV, Y), that gives the vendor the right to make a claim for payment of $x from the Secure Data Interchange. The first-level proxy server registers the unique number Y with the second-level proxy server to ensure that the vendor does not spend the money twice, and provides the proxy server with authorization to charge $x to his/her credit card when the request for payment is presented.

When the vendor submits its "right to payment" and proof of identity to the second-level proxy server the proxy server first runs the charge through the **user**'s credit card, and if that clears, runs the charge from the vendor through the account of SDI (which could also be a credit card, or could be operated as electronic cash or some other mechanism for payment).

This "anonymous credit card" payment method has the following properties:

1. The **user**'s credit card pays $x, but does not know who receives the money except that it is going to the Secure Data Interchange.
2. The vendor receives payment for $x, but does not know the **user**'s credit card information, or the **user**'s identity.
3. The Secure Data Interchange incurs no financial risk because it receives payment from the **user** before making payment to the vendor, although there could still be problems if the **user** complains about the quality of the good for example.

This protocol is simpler than full cryptographic anonymous credit card mechanisms because the SDI acts as a rusted third party to both the **user** and the vendor.

(b) Electronic Cash

Electronic cash is anonymous, just like physical cash. The **user** purchases electronic cash from an electronic bank, presenting blinded notes, so that the bank has no record of the note numbers that it issues to the **user**. For example, the **user** generates a new note number, X, and has the bank sign a blinded copy with its $10 signature, S(B(X), SKBANK$10). Then the **user**, or the first-level proxy for the **user**, removes the blinding factor, and can use the electronic cash as tender. Whenever the note changes hands the recipient needs to check with the bank that it has not yet been spent, because notes are easily copied, but not forged.

Electronic cash has the same useful properties as anonymous credit cards, although it is perhaps a little more exotic. In particular, notice that the bank does not know to whom, or for what, payment has been made. and the vendor does not know which **user** made the payment - it just receives the payment. We have minimized the amount of information exchange that takes place between the various parties in the system.

11. Models of Data Release

Data can be released with associated "terms of disclosure", that define: (a) the price or "exchange rate" of data; and (b) the time and usage parameters of access to the data. Possible time and usage parameters include: data may be used indefinitely, a periodically renewable usage contract, a number of impressions contract, a number of transactions contract, as a one-time permanent or temporary exchange, as a time limited privilege, or subject to certain conditions.

Data can be priced according to its accuracy and content. For example, aggregate data might be sold more cheaply than detailed data, and we can sell the right to access data dynamically (on demand), on a per-impression basis, or up-front. Some vendors might wish to purchase the rights to utilize the data. The vendor might also be interested in renting data (for example, when the value of data is uncertain), and purchasing the data outright if it proves valuable.

The central data warehouse of the Secure Data Interchange can operate as a competitive market place for data and information, with the possibility of facilitating the trade of information between vendors that have synergies in their data sets, perhaps with SDI acting as a valuation device to provide information about the likely benefits of an exchange. The form and terms of data exchange will be those that provide greatest benefit to the contributing vendors.

In general, vendors will most likely be interested in purchasing the access to profile information about **users** that hit their web pages dynamically, on a per-impression or even per-transaction basis. This removes much of the uncertainty from the vendor about the coverage and relevance of data that is provided.

The methodology that we promote within the basic SDI architecture, where **users** themselves maintain control over their profiles also provides **users** with ownership of that information, and enables **users** to maintain ownership while providing per-impression access to vendors, who receive the results of using the profile to personalize an interaction but not the profile itself. The profile need never be released by the **user**.

Per-transaction pricing, possible in on-line applications of secure data interchange, offers powerful new modes of business. For example, it is possible to monitor the number of "click-throughs" achieved on a particular banner ad,? and charge the requesting vendor on a per-transaction basis. This provides a self-enforcing structure to a contract. It is in the best interest of a providing vendor, and the secure data interchange, to provide accurate data to enable proper focusing of ads, and also to provide good data analysis tools, because the success of the advertising campaign determines the revenue that they receive. It is as though the providing vendor is working on an "on-commission" basis.

Ideally we would like to sell data to a vendor at the value of that data to the vendor. This should be done in a way that prevents the vendor from selling the data on to another vendor on the black market. One solution to this problem is to "rent" the data to a vendor, but not give a purchasing vendor outright control of the data. The secure data interchange can be used to provide access to the data. For example, instead of physically transferring data to the data warehouse of a purchasing vendor one could provide a mail-distribution service for that vendor at the data interchange. The vendor could specify constraints on the **users** that are to receive a particular solicitation, give the interchange "Mailer" the advert, and then request that the interchange deliver the solicitation.

The interchange could also provide a vendor with the right to run analysis at the interchange, without actually down-loading the data.

12. *Variations on the Basic SDI Architecture*

There may be commercial contexts in which an SDI service can be established where there is already in place a pre-existing trust relationship between multiple vendors and a third party. Such third parties are inherently motivated to provide services to enhance advertising and e-commerce for their existing and potential customers. These third parties may include, for example, web hosts or e-commerce service providers (ESPS) which often have hundreds or thousands of sites which they host, Web portals, information and commerce service manufacturers, advertising and affiliate network services and data analysis and business intelligence tool providers (which includes the business to business application).

A third party may wish to implement an SDI which operates separately and independently from the central SDI service. Alternatively, some of these third parties may install an SDI server on their customer information server. The server may be integrated into an existing advertising service which they operate and maintain, in which case the vendor receives an appropriate fee for data which is exchanged between his/her existing customers, and a reduced fee (which may be split with the central SDI service) for data which is exchanged by/between a member of his/her SDI service and vendors who are members of the central SDI service but not of his/her local SDI service. **Unless or until the cliet-level proxy server becomes a standard or a large critical mass of end-users adopts the service, one of several compelling business models could be used by a vendor to encourage the visitors to his/her site to adoapt the user-centric SDI service based upon the monetary incentives the user may receive for subscribing. Typically the resulting revenues are split between the user, the user-centric SDI service and the vendor (in exchange for promoting the service to their site visitors a share of resulting revenues generated may be necessary).**

5045646.00 - 4925765

12.1 An open SDI system

An ISP level proxy server can contain the **user profile** generation module, profile processing module, **user profile** interest summary generation module and target object generation module which operate in distributed manner. This enables an ISP to independently implement the core functionality of the system without the cooperation of information vendors (Web sites) or their operators (Web hosts) **who opt-out of SDI**. The modules in third-party SDI servers can share information with the modules in network vendor servers. This flexible architecture enables **the user-centric** SDI **service to be implemented (by ISPs or completed independently)** and when available **inter-operating with** the complete data sets available from the information vendors.

SDI can allow third parties to operate their own secure advertising and/or electronic commerce-based product syndication affiliate network (for all customers). **In accordance with the preferred implementation of SDI, these "advertisements" are represented at the item-level and may be distributed across the (proprietary or main) SDI system network presented in the form of ads, affiliate or portal links to purchasables or sites (which may contain target objects as purchasables) and/or simply (transparently integrated) selections on an ecommerce product catalogue.** The **users** that are also subscribed to SDI, can be given highly personalized information for each site or for the network of sites (which could involve an interface which provides site to site links as a "virtual mall"), and a menu interface to these sites which includes the 2 or 3 dimensional personalized menuing features and personalized search facilities as disclosed in the parent description (a "personalized portal"). **Traditionally the term "portal" was exclusively reserved for major search engine/directories such as Yahoo, Lycos, Alta Vista etc.**

However, at the present time, there is a trend in which may Web-sites are becoming what are called "portals" or "community portals" serving a particular ecommerce or content niche or theme. Often highly robust content and/or extensive (usually web-wide) links to relevant information resources and web-sites are accessable from such portals in order to create a "one-stop shop" service to visitors or community members. Additionally, in order to leverage low-cost access to more robust content, a transaction-based or "click-through-based" or transaction based model is also emerging. This model fits particularly well within the current framework of SDI. In particular user profiles provided by SDI-enabled site visitors provides the platform for automatic generation of not only personalized ad banners but also content and portal links, it is reasonable for these community portals as part of their business relationships with the sites and advertisers to which they provide access to share revenues paid by these sites and advertisers with the end-users (which significantly increase the clickthrough by virtue of personalization of information at potentially all varieties at the site resulting from revelation of the profiles from those visitors). As discussed on the parent case, these pages are precached in advance onto the local server and client. ***Within today's internet infrastructure as pages are typically dynamically***

generated, the files are first precached then dynamically generated local to where they are anticipated to be used. **Additionally, high-end community portals typically offer content from external sources which is sometimes syndicated from content providers which are paid by clickthrough or transaction. In this model the user or the site may absorb these fees paid to the content provider (because the content when personalized may appeal to almost any type of user, most typically the site would subtract a certain percentage of the clickthrough or transaction fees paid to the user).** ***As is discussed in the parent patent application, content (e.g. links to specific syndicated items) may also be created based upon the collective user statistics those links are presented to the users for which user profile information is not available (where these page links could be generated on-the-fly based upon the user profile).***

The user-centric SDI service also provides valuable information in not only generating the data model for personalizing site content but in general identifying what type of content to syndicate to the site. In a variation, it may even be possible to employ the statistical techniques used by SDI to dynamically select and generate personalized *pages* in a much more ad-hoc and unrestricted fashion. Using the syndication model, a vendor may agree (based upon their associated privacy policies), to enable other sites to not only syndicate their content which may be contextually relevant to that site, but also actual portions of the site or portal may be linked from other pages on the site which are contextually relevant. These pages may be linked from the portal directory or as hyperlinks which are links to contextually very similar pages (or the vendor may allow these pages to be modified as needed by the site. As a general site development technique also the user's clickthrough patterns are useful in updating or suggesting updates to the model to improve the value of the site. Of course, some vendors privacy policies will restrict the *ability* of some sites which can utilize the content.

Google Exhibit 1029 - Page 70 of 95

12.2 A Closed SDI System for a Syndicated Network

The Web host (or more generally a vendor, a provider and/or operator of server functionality to a variety of information vendors), may also be interested in operating his/her own closed version of SDI. The main SDI server for the closed system can be located on the network vendor servers, or it may reside upon the information vendors servers (as it is operated by that local Web host). For example, an affiliate ad network (including a web host acting in such capacity) could upon installing SDI onto their network enable and enforce the wishes and desires of advertisers (and particularly) sites which are advertised upon with regards to what types of sites and advertisers (respectively) they allow or disallow for purposes of standard or affiliate advertising, in accordance with the methods herein disclosed. The general implementation for determining which this general application for using collective **user** feedback to determine relevant site links was described in the parent issued patent). In this case, end **users** who are subscribed to SDI would receive personalized affiliate links (including product level recommendations for on-site purchases) which have been pooled and profiled at the main SDI server from all SDI vendors (in distributed fashion) and matched with the **user**. **Alternatively, such approach may also be used for ISPs who wish to operate their own closed user-centric SDI.**

12.3 Interoperability Between Local SDI Services

With interoperable (local) SDI services, we can also facilitate the secure enforcement of data sharing policies and transfer of transaction fees between these local SDI services E.g., by/between aggregations of ad networks, syndication networks **ecommerce site, portals** and Web hosts operating virtual portals and advertising/syndication networks with personalization as its primary capability.

In each of these primary example domains, the server operator is financially motivated to sell the SDI services to his/her sites because the transaction based model is used, and the server operator receives the commission on each transaction (or click through) occurring within his/her network of sites. However, if the server operator also integrates his/her local SDI service into the main SDI service (to share **user** lists and impressions and/or space to advertise to these target **users**), s/he can receive a commission (in conjunction with each vendor transacted with) for each advertisement placement or syndicated transaction to or from his/her network.

We can also allow the local server operator to split the transaction fee (normally received from the main SDI service), thus "referral fee" for both the referred customer and the referral of customers (through the placement of outside ads or products on one of his/her sites) or other means of targeting his/her site's existing customers.

Reduced overhead resulting from economies of scale which may likely result in incentives to the local operation, e.g., free installation and operation of his/her local main SDI server by the main SDI service, i.e., as the operational overhead would be cost justified by the shared transaction fees of customer referrals and advertising space coming back to the main SDI service.

This architecture also may be useful and is ideally suited for cross vendor product advertising as through an ad network or product syndication network using affiliate links. In addition to the **user profile** generation module, a target object profile generation module should also reside across the network vendor servers such that it is possible to generate target object profiles for target objects on network vendor servers. Alternatively, **user profiles** and target object profiles are downloaded to the client level proxy which performs collaborative filtering tasks as the **user** browses from site to site.

In both of these cases, the main SDI server can receive **user profile** data generated from the **user profile** generation module located on the ISP-level proxy, and target object profiles generated from the target-object profile generation modules located on the various multiple information vendor servers.

13. Dynamic/Real-time Secure Data Interchange

The **user**-centric SDI model allows **users** to provide personal information on a carefully controlled basis to vendors and other **users**. Furthermore, vendors can implement rules that personalize the information, products, and services provided to **users**—on the basis of personal information that they receive from **users** directly, or have acquired about **users**.

As an extension to this model, we also allow **users**, vendors, and other third parties to associate "meta-information" with other **users** and vendors. This information might be a **user**'s opinion about his/her interaction with another **user**, an annotation that relates to a particular web page, or information about a physical object, for example how to get to the top of a tower. The system of SDI enhances the value of this information by providing a secure environment where **users** can also associate their own profile with the meta-information that they "leave". This allows collaborative filtering techniques to generate appropriate meta-information about an object (**user**, physical object, vendor, web page, etc.) that will be useful to a particular **user**—given that **user**'s own profile.

We define "virtual tags" as any piece of information about an object (physical or virtual). The information may be authored by any party, but annotated accordingly. For example, the appropriate virtual tag provided by a **user** about his/her-self is the pseudonymous profile for that **user**, -- and with SDI only the **user** his/her-self can gain access to the profile (either directly through editing, or indirectly through continuing transactions).

It is useful to implement a "reputation system" within such a virtual community. Initially **users** (under pseudonyms) have no reputation, and their opinion does not count for much, but after every positive interaction (as defined by other parties in an interaction), the "reputation" of a **user** can increase. (see the Kasbah system, MIT). This reputation system is appropriate to a pseudonymous environment. Notice that gaining negative reputations is not useful when **users** can simply change identities.

In one variation we can "block" certain **users** from providing information, when those **users** have negative reputations. Clearly, collaborative filtering or other data mining techniques could usefully allow for reputations when weighting information about an object.

13.1 Autonomous Exchange of Information

Client-level SDI proxies can act as autonomous agents in an architectural variation of SDI, where the "client-level proxy" is co-located with a (physically) mobile **user**, for example on a palm-held computer or head-up display.

In a "match-making" application domain, the goal of **user**-agents is to find other **user**-agents with desired profiles/synergies, and arrange person-person meetings, or business-business meetings/agreements. There are two parties in any exchange of information, and although information exchange will be bidirectional, it is useful to talk about a "requestor" and a "requestee". The key provision provided within SDI is that:

1. Requestors and requestees can communicate anonymously, without revealing (even pseudonymous) identities.

2. A requestor cannot access profile information about a requestee unless authorized to do so (implicitly or explicitly) by the requestee.

Implicit authorization occurs when a requestor can present certificates to verify that it has required attributes to access particular information. Explicit authorization occurs when a requestee provides direct authorization to a particular part of a **user**'s pseudonym.

Essentially there is bidirectional information filtering: the requestor agent will only present certain information to the **user**, information that is relevant; and the requestee will only provide information when a request is judged to be legitimate. Information exchange between agents occurs as part of a multi-step negotiation, until both parties can agree on terms for either a physical meeting (or execution of a deal), or further pseudonymous exchange of information or cooperation.

In this case, using methods taught in the co-pending patent **user**-agents can identify other agents that are "close" through an anonymous matching market, where agents provide their location and a (one-time) identity for contact. The market informs agents when other agents are close. In fact, this "anonymous matching market" can be extended, and agents can provide more than just location information.

A **user**-agent might also broadcast a "persistent" query over the agent-network, for example requesting response from agents with a particular set of attributes, and providing some information. Decisions about what information to exchange are made on the basis of both static and dynamic profile attributes, e.g. standard (historic) profile information, current behavior, current location, recent activity, and credentials that can be presented/denied. LEIA style-behavior attributes can be used to automatically decide on the relevance of new virtual tag information. A requestee might also demand certain credentials to indicate the lack of negative reputation marks, for example that an interaction with the **user** has never received a bad rating. Perhaps a third-party could be used to determine whether the **user**'s know each other (eg www.sixdegrees.com)

We can extend information disclosure to include communications between **user**s and other parties. A **user**-agent might decide to make another agent privy to communication, on the basis of the context of communication, content, location of parties, profile of the third-party, etc.

When a requestee denies a request for information, it may instead provide criteria for data releases. A requestor can respond with a different information request, or a subset of required credentials. Finally, the agents might agree on terms of negotiation and conditions can be anonymously fixed.

Negotiation might FAIL, in the case of missing rules in a rule-set, or a special case that has not been anticipated. In this case direct **user**-**user** interaction might be necessary, although this can be executed anonymously via a real-time anonymizing service.

There are (at least 5 levels) of information disclosure:

1. indicate to another **user** interest
2. release of profile information
3. disclose communication
4. add an individual to a current correspondence session
5. schedule a meeting/strike a deal

The end-result of information exchange could be an agreement to calendar a meeting for some future time and place; and absolute, or pseudonymous revelation of identity.

Initially, an implementation of the data-release policies might allow only manual definitions. However, after an initial "beta testing" phase, a data mining suite could be used to cluster **user**s and generate exemplar data release and data request policies. A system can provide default settings for **user**s, and recommend setting based on **user**s with similar profiles. The **user** can further fine-tune the rules.

Finally, automatic feedback techniques can be useful to adjust rules, for example—when a **user** is especially receptive to particular type of introduction then make such introductions more likely in the future. An intelligent interface system might also suggest refinements to the rules, to automatically cover "patches" where the **user** currently controls interactions. (see the Microsoft Lumiere project).

Google - 2525

Note 1: must next describe **user profile** attributes and rating criteria, e.g. selective access to group ratings and annotations by navigating their attributes. Also automatic generation and comparison ratings between items via rapid profiling.

Note2: Real time (including spoken) communications (either **user** to N-**user**s or **user** to **user**) within a local proximity where the **user**s are "pre-qualified" (e.g. knowledgeable in that particular field, educated, etc.)

13.15 Credentials: Examples and Typical Uses

In accordance with the parent issued patent, various credential issuers are provided for issuing standard and resolution credentials to individuals. Thus certain entities may be entrusted with "legitimate authority" to validate and submit credentials which are issued to the appropriate individuals. Resolution credentials are provided to prove the absence of a quality attribute or behavior (which is often of a negative nature) relating to an individual and is submitted by a third party and typically must be issued on a periodic basis in order to maintain currency. If a resolution credential is not issued (or not renewed) an adjudicating third party is provided which has access rights to both of the parties is provided to resolve resulting disputes (from the subject **user**). The present invention describes how credentials can be issued to **user**s pseudonymously.

A few simple examples of resolution credentials which may be of interest to **user**s (credentials which **user**s may commonly request as a precondition to requesting or accepting requests to be introduced or initiate communication with an outside unknown third party) include:

1) (for business associations) are in good business standing, e.g., have not attempted to defraud other **user**s in the course of common business practices. Or maintain sufficient funds in one's account to perform business activities (as represented by the **user**).
2) (for business interactions or social interactions) are in good standing with the law.
3) are considered "safe" individuals,e. g., have not been rightly accused or criminally convicted of violent acts, (e.g. a store clerk may wish to be made aware when an individual walks into a store which cannot present a resolution credential for not having been convicted of retail theft or robbery.
4) have not been accused by other individuals of inappropriate or antisocial behavior.

Some standard credentials which may be of interest to many **user**s, and which may (as with resolution credentials) be incorporated with the standard settings of the **user**'s data request policy as herein described. A few examples are cited (among countless potential others): profession, awards, honors, alma mater, e. g., Harvard graduate, doctorate degree, etc.

There are a variety of rules which a **user**'s data disclosure policy and data request policy may contain, to control what if any attributes are released, and what credentials are required. A data request policy may state a rule for explicitly notifying the **user** if a particular resolution credential (e.g., indicative of a serious problem or concern)

cannot be presented in response to the
user's disclosure request.

Some **user**s may not wish to disclose specific information about themselves via these standard credentials but instead certain "extracted" more general information may be provided about themselves. For example, instead of a "Harvard grad or Ph.D." there may be, for example, credentials indicating "intellectual" or "prominent intellectual". Or instead of indicating an individual's wealth or value of assets, the credential may indicate "wealthy" or "very wealthy" (typically, depending upon **user**'s wishes this latter credential should also be withheld during initial introductions or subject to some fairly stringent conditional criteria from the other party) and instead replaced with an even more general credential e.g., "prominent" or "influential citizen"). Similarly, an individual's exact profession or scope of work may not be fully disclosed initially but rather a more general definition of his/her profession or perhaps the general field initially in which the **user** works **or his/her particular views on certain social or political issues may instead be substituted with (for example) "socially conscious", "philanthropic", or "moderately liberal".**

Another example of a credential of potential interest may include the profiles of **user**s which a certain individual associates with or is acquainted with. The ability of a third party to gain access to this information, however, is conditional upon the data release policy of that associate's or acquaintance's data (e.g. it could be affected by what is the profile of the common acquaintance to whom that **user** would be disclosed as an associate as well as, importantly,

the profile of the prospective disclosee.) In one variation, the system may simply identify the fact that there are common associates and acquaintances between the two individuals. Again that associate's or acquaintance's data release policy may further control even detection of this fact. It may instead also notify one of the parties of this fact, but request that it not be disclosed to the other party.

In accordance with the parent patent application, rules may be learned regarding certain things that a **user** does (as in ascribing these rules for which messages to send to whom or what **user profiles** and under what circumstances/events surrounding the target **user**). Thus, his/her agent may begin to suggest certain future actions which could be performed in the future upon **user** approval or even automatically. If the **user** has had no previous interaction at all with the system, it may identify which other **user**s of the system the present **user** is most similar, and recommend initial rules. Additional textual attributes can also be leveraged to provide extra criteria, and data mining techniques used to generate more appropriate rules.

Another category of **user** credentials include features that may be inferred implicitly by location/time data captured by LEIA. Such information may reveal a **user**'s likely behavior and activities. These inferences, however, are unavoidably somewhat speculative and inconclusive, thus cannot be substantiated on a valid basis for issuing credentials. The data may be useful in suggesting the present context and circumstances surrounding a **user**.

Additionally, the communications which the **user** may be presently involved in i.e., the content profile of his/her spoken dialogue and/or other "on line communications" may be used and combined with location/time patterns in order to further infer the circumstances, behavior, and present temporal interest of a **user** and/or third party for purposes of employing the **user**'s data disclosure and data request policies.

Google - Doc 9

There are many potential application contexts in which this present architectural framework could be usefully deployed. For example, it would be extremely useful for **user**s who wish to be re-identified or re-contacted after original introduction -- with unique pseudonyms. A **user** might wish to withhold all identifying information from his/her acquaintances and associates, except when required for further information exchange. The **user** can also apply the techniques of randomized aggregates to minimize the chance that one friend or associate will be able to correlate that unique pseudonym with the same individual. A new portion of the use's profile could be combined by **user**s with different information about the **user** or more importantly, a private piece of information under a pseudonym could potentially be associated with his/her identity. The **user** agent might use a statistical algorithm to identify a threshold of how much data can be released (including what types of data and to what degree it must be randomized), in order to prevent the **user**'s identity from being compromised.
Within a location enhanced context, unless the prescribed range of "proximity" to the **user** is quite large, securely protecting the **user**'s identity from malicious third party collusion (for purposes of **combining** unique pseudonyms and/or exchanging data that has been released and entrusted to them) is a harder problem. The system could (most obviously) assume data exchange between the parties will occur and limit the combined disclosure to only that of the most data restricted **user** in a given location/time domain. The system could alternatively, perhaps "space apart" the number of **user**s within a given location/time context who can access more "restricted" **user** data (of course the problem goes away if all the disclosees have similar disclosure restrictions by that **user**).

We allow initial information exchange to be anonymous, such that information that is released as preconditions for release of further information is not useful. Similarly, so long as initial encounters are anonymous there is no need to withold information about them from the **user**.

Credentials can allow **user**s to identify other **user**s that may pose a threat. This identification may be provided vis-a-vie resolution credentials and/or rating (by third parties). e.g. a **user** has not engaged in any serious criminal activity, physically harmed another person, or interacted with other individuals who are unable to produce these resolution credentials. Other credentials may specify the nature of an infringement, and its context and severity (e.g. what was the context of a physical assault? Was it performed during a bar brawl, against a friend, a boss, an elderly person, a child, a family member – or at work? In this case, the **user** agent may, for example, bring to the attention of a prospective employer that the **user** could not present a credential indicating that they had not previously harmed or threatened a former employer. Was it minor or severe? Also, if such individuals (lacking, for example, resolution credential proving the absence of having committed armed robbery) are (or come) within a certain proximity of a **user**, the **user** may wish to program his/her **user** agent to notify the **user**. The same would, of course, apply to a store clerk regarding customers of this sort or to baggage security personnel at an airport. Or, highway patrollers may be interested (e.g., on certain

stretches of highway) in being made aware of vehicles and their locations whose agents are unable to provide a resolution credential proving the absence of a drug conviction.

In a more benign though similar example application to that above, **user**s who are "common individuals" could use the above techniques (through similar co-operation/collusion) to "single out" other individuals who may be of a "high profile" nature (e.g., famous or wealthy) who may otherwise strongly prefer to remain incognito, wherein potential **user** to **user** communications could result in a significant invasion of privacy.

In another application (in accordance with the auto insurance risk determination methods described in co-pending patent application entitled "Applications for Location Enhanced Information Architecture"), an on-board computing device within a **user**'s automobile could identify another automobile lacking, for example, a resolution credential for safe driving. i.e. the on-board **user** agent continuously polls agents in other cars for a "safe driving" credential, and if it fails to receive such a credential it issues a warning to the **user**. As an extension, this location data could be converted into a dynamic 2-D rendering upon the **user**'s windshield (using heads up display technology) in order to thus superimpose a persistent flagging or highlighting of that particular automobile from the driver's visual perspective. Pedestrians **or low enforcement officials (for example)** could also receive instant notification.

We can avoid **user**-identification/privacy violations within such a physical system by fuzzing the information provided to a **user**, for example indicating only a general location, or hiding the individual within other individuals. When highly sensitive information is disclosed, it is important that we can protect the real identity of the **user**.

Some examples may include: health problems, like HIV status, neuroses, sexual abuse as a child, family history of violence, alcoholism or emotional disturbance. However, a prospective partner might wish to receive credentials of a lack of these (negative) qualities before pursuing (non-anonymous) contact. As is described in LEIA, a roaming cellular connection, or GPS, is not essential for providing a **user** identifier. For example, optically-based biometric identification techniques such as iris scanning or combined iris/facial identification techniques may be used among other potential inputs as well. **User**s will be reluctant to release location/time data, even anonymously, when suspicious behavior can be inferred—probably subjectively. Of course, we allow the **user** to control this data release within his/her privacy policy. **Possibly a "clean" behavior resolution credential would provide the desired objective.**

Indeed, if a "significant" threshold of suspicious behavior is exhibited and detected,the information may be accessed by law enforcement officials, through seizure of the decryption key for that data (which includes his/her physical location information) and any additional profile data which is considered of immediate critical relevance to the suspect (or prospective) infraction. Such cryptographic techniques for key seizure from a key escrow are well covered in the literature. If behavior is inferred from LEIA information, then the threshold should allow for statistical errors. There may also be certain circumstances in which key seizure may be required after the fact (at some time in the future).

For example, if/when certain even moderately "suspicious" behavior patterns are detected, it may be possible for the SDI data warehouse to preserve a comprehensive record of that information (and perhaps the record of that **user** which precedes and follows that period of interest). Thus preserving evidence which may later prove useful in contributing evidence towards a conviction, acquittal, e.g., proving that a **user** was not at a particular location/time. A record containing more detailed segments of a **user** with a proven negative or questionable history may be preserved and general location/time features may be abstracted for the remaining portions of the record (thus compressing the record substantially). This may be performed for regular individuals as well, thus retaining key relevant features while discarding the majority of the record which is irrelevant or redundant.

A very important caveat underlying the above architectural framework is that until such time as **user**s become personally "wired" (through miniature computing devices and/or wearable computing devices), the above applications involving the use of resolution credentials within the context of a location enhanced (physical) environment will be hard to implement practically. Automobiles may be an exception,as may be technology which enforces the disclosure of the physical presence of a **user** agent (resident in a device) to other **user** agents within the networked environment.

Finally, we need to protect **user**s from prejudice, on the grounds of religious beliefs, political affiliations, sexuality, and membership of certain organizations. A **user**'s inability to provide resolutions credentials could result in a significant threat of the fundamental elements of their individual freedom and civil rights. Thus credentials proving the absence of such affiliations/memberships or professions should again be banned from use. Thus, an important condition should be provided as a part of the system's design criteria enforcing the recommendation that

users must collectively avoid the use and request of such resolution credentials as part of the **user profile**s

13.17 Additional Applications of Virtual Tags

The following is a list of additional types of application specific information which may, for general or specific applications, be usefully deployed through the use of virtual tags.

(a) User and seller information - as detailed in the parent issued patent, specific details of what users and sellers may be looking to buy or sell respectively may be used to suggest the basis for a potential commercial transaction. The transaction may be large (but not necessarily so, e. g., real-estate, private investment in a small business or public stock). If a physical or on-line interaction with the other party is warranted (e. g., for larger commercial transactions), as is suggested later **as** in the present description, **user**s may identify other **user**s which form the most relevant "match" with their interest. At this point the agents can check for credentials, and then either communicate or calendar a meeting. Similarly, the agents may find the
"best" match of **user**s who happened to be physically proximal to the **user** at that particular time, or at some future time(s)/location(s) which is mutually compatible (similar applications are suggested for matching sales persons with prospective clients, identifying experts to work (individually or collaboratively) on a particular project or problem, to answer a question of an appropriate specialized nature to their area of expert knowledge.) The parent issued patent suggests at a general level these commercial applications. An additional feature described therein involves the use of a decision tree called "Rapid profiling" which can be used in the present context to identify from the most common needs of users and "goods" of sellers in general and the known profile data about each user and seller individually, a list of questions for each party which most briefly and efficiently determines the complete user/seller profile of each party individually.

(b) Medical information, such as medical conditions, medical history, active prescriptions, drug reactions, family history, possibly even genetic pre-dispositions (from a genetic profile). Medical insurance information may also be potentially useful for a prospective qualified accessor to be able to readily access in case of an emergency.

(c) Social Interests Profile Information -- The parent issued patent also suggests the present application at a general level. For a dating application, **user**s may be matched on the basis of their common interests/preferences and perhaps on the basis of certain information reflecting personality, social or cultural behavior/affinities or psychological attributes. On the other hand, for purposes of meeting casual acquaintances, **user**s may be interested in another **user** who shares the above characteristics as well as someone who has recently shared similar experiences and/or personal challenges.

(d) Physical location information -- **User**s or advertisers could, for example: a) Query a pseudonymous **user** database to access profiles that are in close physical proximity and match certain criteria, e.g. live in a certain geographical region, had recently attended a meeting or event (or is planning to attend a particular event) had recently communicated with a friend or associate. In another variation, a **user** could for example, submit a query pertaining to every **user** in a particular physical space, e. g., a room, hotel or convention center, e. g., identify all **user**s present here who attended Internet World, 1995.

(f) Professional Information/Qualifications - As in the application of matching users and sellers, a description of a **user**'s needs or situation with relation to various professional services may be provided as additional data about the **user**. Examples may include: (as above) medical data, professional or business history (as well as legal history) which may be of interest to law firms, accounting firms or various business consultants. Personal , family or emotional difficulties may be of interest to psychologists or family counsellors. Again, **user**s may submit this information as a query for prospective matches, or they may be pseudonymous queries or automatically matched in accordance with criteria specified by the professional. The issued parent patent application also lists additional applications which could as well be relevant within the usage context of virtual tags.

Referral Information - Situations frequently arise in a variety of contexts of human interaction (whether social or professional) in which a **user** may wish to refer the **user** they are in contact with to another individual. Often this occurs in a professional services context which a **user** has a particular need/or other characteristics which make him/her

an appropriate match for the services provided by the other party. Or in a business context, often a **user** will forward a business contact or associate to another colleague who is deemed more appropriate for the particular context and/or scope of business. Likewise, in a personal or social context **user**s may sometimes meet two or more individuals which they observe or perceive share common interests, goals or beliefs or perhaps possess complementary capabilities, knowledge, or characteristics. In each of the above scenarios, virtual tags may provide substantial benefits.

For example, the referring **user** could forward the relevant portion of the profile and identified need of the **user** to the referring party whose **user** agent may determine the acceptability of the request and/or the priority with which a communication or meeting could be scheduled (e.g., as could be automatically arranged by/between the two party' calendaring agents). If the referring party's agent is unable to make a decision or priority assessment for scheduling purposes) on behalf of the **user**, the agent could instead try to contact the individual him/herself for assistance (and statistical feedback to the system's data model). In order for these types of referrals to be performed efficiently, the area of expertise required can be specified, and provisions can be made about the type of referrals that a professional will accept.

(h) Employer/Employee Information - **User**s who are seeking employment (actively or unofficially/passively and employers who are seeking employees for specific responsibilities may benefit under the current scheme. An employer may post a description as part of his/her virtual tag (and that associated with his/her company). His/her employees may also have provided ratings and/or annotations which are further descriptive of his/her personality, leadership/management style and skills, work environment which s/he promotes and overall quality. Previous employees for that position (who may also have either provided information about themselves as deemed appropriate) or may also have provided such information as well as pertaining to the position. If not s/he may allow him/herself to be contacted by the prospective candidate (e.g., in exchange for a fee).

Access Privileges Information - **User**s in an organization are frequently given privileged access to certain files within a corporate intranet but not others. Though there are many ways of profiling **user**s according to their level of access privileges to information, the following example is considered: Based upon the position (e.g., responsibilities and tenure with the organization), **user**s may be "classified" into groups according to different levels of access to confidential information. Virtual tags may be used to extend the capability by providing for immediate disclosure of a **user**'s information access privileges to another employee in real-time and in a physical context. Similarly, it is conceivable in either a professional or non-professional environment that **user**s may wish to maintain "secrets" between each other. Access or restriction to/from a secret may be provided in accordance with a particular attribute of a **user** (e.g., departmental secret, a family secret, personal secret) or by the individual identity of the **user**, e.g., upon contact/communication is privileged or not to a given secret per the restrictions by its originator).

Additional rules could be provided, so that if certain pre-defined events occur, the criteria for release of information associated with the secret are changed. Alternatively, this "information" triggered by an "event" could instead be a message. For example, if a **user** reads or accesses certain information, meets with a certain colleague or friend send message X. This message could be (for e.g.) a request to perform some task relating to part of that information, a reminder to address certain issue(s) while chatting with the colleague etc. or, per the request of an individual's employer or colleague if a given individual (a sales person) meets with **user** X send him/her message Y (which may refer to a previous encounter, experience or fact s/he should know pertaining to **user** X and which may have bearing upon their conversation or professional interaction. (This message could as well be sent by **user** X as a reminder notification to him/herself at the time of the encounter).

In another application, the **user**'s access privileges may be used for granting him/her access to restricted physical areas, (thus, the virtual tag effectively may behave like an "electronic door key"). A variation of the technique may be used for granting access to professional meetings, where information access privileges of **user**s must match the anticipated confidentiality parameters for the scheduled meeting. Another application may include the ability to automatically enable access or restrict access, based on payment of fees, and whether or not an individual is a representative or partner of a competing company.

(i) Contextual Information -- If **user** X is a prospective customer, and performed certain on-line or physical behavior that suggests an interest in a product/service, then the advertiser's agent can automatically send an ad (subject to the **user**'s privacy policy). Another condition which could trigger either a notification to the advertiser or the message

(automatically) could be temperature or weather conditions (which may affect people's purchasing behavior as well as often driving/traveling activities which affects the agent's scheduling and reminder's/follow-up with participants on existing scheduled meetings.

13.2 Example: Match users to sellers

For example, in a mobile sales-force, the system of SDI can first generate an ideal list of prospects, and then help salespeople target products and offers to individuals. We can provide information to salespeople about **user**s, according to the profile of a salesperson (and reputation), and a **user**'s personal terms for data-disclosure. Similarly, a system of SDI in conjunction with the methods taught in co-pending patent allow automatic detection of salespeople close to **user**s (via an anonymous location market). The market allows matches to be made, but does not reveal anything about a **user** that the **user** does not authorize.

For example, in a marketing network, with a commission-based sales force. **User profile**s can also be used to determine **user**-responses to offers and products (see the methods in patent application "**System for the Automatic Determination of Customized Prices and Promotions" User profile**s are generated from interactions with other vendors, salespeople, and current activities/behaviors. SDI allows profiles to be built from extended interactions across multiple vendors, so long as the **user** authorizes the same pseudonym for each vendor.

Similarly, **user**s can themselves use seller profiles, to decide whether or not to interact with a seller, and vendors can use seller profiles to decide how to allocate new customer prospects to sellers. The profile of a sales-person may show correlations between product sell-rate and the type of product, type of **user**, that the sales-person interacts with. Initially seller profiles may not be very well related to sales-performance, but instead based on general SDI-style profiling, and wider (eg. professional) credentials. Later, as a seller gains experience, profiling can be based on a sales-person's track record (and this will subsume other information).

As an extended example, we can also consider a system for handling tasks—e.g. a call center, or a dynamic work-flow system. As new jobs (or calls) arrive into a system, the jobs are automatically routed to the appropriate experts/machines/sales-people according to the ability of the person to perform the task. The ability is measured by a match between the profile of the person and the type of task, for example on historical performance/feedback, and on other relevant attributes, and also through collaborative filtering techniques. The goal of the system is to allocate tasks in the most efficient manner, across a system of experts (or machines).

The problem might also be informational: e.g. find an expert on ancient American civilization for purposes of writing an article, or answering a specific question. Relevant information might include the expert's resume, and the expert's knowledge expertise profile developed from his/her activities in responding to previous queries.

Level of expertise might also include the size of projects performed within a particular specialized area, and relevant education qualifications.

Example: business-business introductions. Another application domain for privacy-protected match-making, where **user**s are anonymous until an agreement is struck is business to business introductions. For example, it might be useful to automatically identify synergies between businesses (e.g. in infrastructure, technology, or product) -- for the purposes of pursuing an advantageous strategic relationship. If the meeting is between two employees of competing companies, then the system of match-making could also ensure that a meeting is predicated on a particular task that does not cause conflicts with their respective companies.

We can also use a query-based system to establish a **user**'s relevance to a particular task, or another **user**—along the lines of the method in patent **"System for Generation of User profiles for a System for Customized Electronic Identification of Desirable Objects"** with the efficient method to profile **user**s.

The system of SDI can also be used as a confidential database for the purposes of generating statistics from sensitive data. For example, as a trusted system, manufacturers might be willing to provide information about their productivity, margins, retainment rates, production efficiencies, yields etc. The central SDI server could generate statistics, globally for the manufacturing sector, and then individually for each manufacturer—as it relates to the information provided by other companies. Similarly, it would be possible to use such a system to compare salaries across different universities. While an individual university might be reluctant to reveal information about its pay-scales to other universities, in the

Exhibit 1029, page 78 of 96

aggregate this information is not sensitive—and a survey on salary can be useful to both employers and job candidates. SDI is used to securely calculate statistics, without revealing any information that might compromise the privacy of a single employer.

Note: Describe secure virtual database involving matching by expertise and sharing of human resources (e.g. virtual work groups). Each member (a commercial interest) has a certain "resource sharing policy" which defines 1) what entities or types of entities s/he would share resources with. 2) If so, on a per-entity or per-entity type basis, what types of their resources (e.g. type of skilled employee and for what TYPE of outsourced task) would the entity share. It is an obvious extension to look at sharing of code, technology, intellectual property. A major challenge and limiting factor being how well informed SDI, the neutral intermediary can be made aware of the needs/requirements of a company such that it can make evaluations entirely on its own regarding highly confidential materials with which it can accurately predict the basis for a deal WITHOUT disclosing to the prospective recipient what the technology or know-how entails (which could compromise the value of that asset should a deal not eventuate).

Further describe an extension/adjunct of the present human/technical resource sharing method whereby corporations (in particular high tech and Internet related) may employ SDI to utilize the above information regarding their human and technology sharing synergies in order to detect and recommend strategic (e.g. equity sharing, merger, acquisition etc.) relationship opportunities between the entities. B to b and even b to c **user centric SDI user behavior analysis** combined with text analysis should also provide revealing clues about what types of companies tend to share similar customers and provide similar (complementary or competitive) products and services which may suggest that such synergies are potentially available (Geographic region may also be detected or inferred. **Of course the vendor centric SDI service may benefit from this information as well as provide information for other vendor identification of competition). Other data resources including electric payment protocols, EDI, automatic check payment, check services, etc. may be useful data resources as well.** Again the disclosure of detailed business information is very helpful and a data release policy defining the parameters for such strategic initiatives may be critical in order to determine what companies may be potential candidates for which initial feelers (of high level information disclosure) would be appropriate to put out to a prospective company to determine mutual interest and/or further basis for expected synergies.

Note: Reference FAQ routing scheme in co-pending patent application, also add a section on using customized prices and promotions scheme to assign a price to a task, query response or virtual work group participation (where potential synergies are identified between the parties).

13.3 Virtual Tags for Annotation/Information Filtering

In this extended application of SDI, we allow **user**s and other third parties to annotate objects (physical and virtual) with meta-information, either to remind themselves about a previous interaction in the future—or as a system of "knowledge learning", where systems of **user**s leave useful information for other **user**s. Information is left in the environment, leaving a trail for other **user**s.

For example, the information that is tagged to an object, referred to as a "virtual tag", can contain a pointer to other relevant information, such as a survey of a film by a third party, or the **user**'s own comments/feedback. For example, a restaurant listing could be annotated with meta-information about the quality of the food and service. Such information, when provided by a wide sample of **user**s, can provide robust information about objects. The information that is used by a particular **user** can be filtered—for example, weighting the opinion of a respected restaurant critic, or weighting the opinion of **user**s with common profiles (when that information is available).

Virtual tags can be assigned to objects with physical locations, and the information triggered based on the physical location of a **user** (using LEIA technology). Virtual tags can be assigned with **expiration** dates or other time-sensitive information. An individual **user** might leave an "action item", for example—next time I return to this object (eg. web page/ vendor) be sure to perform this task, enter this query, check this link for new information. As another example, after a conversation with an SDI-enabled **user** it is possible to tag that **user** with some notes, to remember the conversation the next time the two **user**s meet.

The technical innovation that allows this use of virtual tags, in addition to the protection of privacy, is that we allow **user**s to annotate information to objects that they do not directly own through a system that separates virtual tags from the content that is tagged. In particular, tags can be stored (either at the ISP-level proxy, or main SDI server) for

associated web pages, and exchanged/retrieved automatically when the object is accessed. The virtual tags can be used in conjunction with target-object profiles that are generated through SDI for web pages (and approved by vendors).

Virtual tags can be searched, using relevant terms, locations, or times, and can also contain links to authorative information, such as audio and/or video.

Tags are encrypted, so that only SDI-enabled **users** can access them. Tags are also associated with the pseudonymous ID of the **user** that left the information (although they can be anonymous, an associated profile allows more accurate collaborative filtering techniques). Finally, **users** can leave data-disclosure policies, embedded into tags—to certify the properties of other **users** necessary to release the information. When tags automatically are time-stamped with location, and time, and other information we allow for this information to be "fuzzed", as disclosed in the section on Randomized Aggregates, to protect a **user**'s identity.

In the physical world, implementation of meta-information in a **user**'s physical information, can be viewed via head-up displays, video cam monitors, wearable computing devices, or audio pieces. The information itself can be embedded directly on physical objects, for example on magnetic strips or via. visual encoding techniques—or the appropriate information can be accessed from a secure remote database based on the **user**'s physical location (using LEIA location technology); or bar-codes that provide a universal identifier for an object.

13.4 Information Personalization without Vendor cooperation

Google Exhibit 1025

We can also rely purely on information provided (as virtual tags) by **users** about a web page, or other physical or virtual object, to filter and personalize information that is displayed and recommended to a new **user**. For example, because virtual tags are stored and accessed independently of the source of the object (e.g. not on the vendors own web server), they can be automatically picked up and associated with a page, even without a vendor that is subscribed to SDI.

The **user**'s client-level proxy server can annotate web pages with relevant information as they are generated, for example adding mark-up to relevant information, or suggesting a link that the **user** should read, with notes that may be useful. Similarly, the database of tags can be used for an advanced web-search. **In accordance with the earlier part of this specification in which updated metadata is periodically described to Web pages at the main SDI server, it should be noted that user profile queries, or menu navigations may actuate rules which selectnot only web pages but associated annotations and page links. Where these annotations and/or links may be slated by the original user for certain users or user types or selected automatically when a regularly profiled user accesses the page or requests the annotation (or manual link recommendation) associated with the profile for that original user's profile type. If multiple page link recommendations were provided by multiple users of that given profile, the selection of pages displayed may be determined as a statistical average.**

The ability to personalize/provide recommendations to a **user**, can be used in conjunction with a cache engine, that might be located on an ISP server. For **users** that follow advice, the ranking of recommendations is correlated to the information that the **user** reads, and therefore pre-caching can be more accurate, because the information about **user**-recommendations can be made available to the cache engine. Similarly, this is useful for advanced fetching to the client-level SDI proxy, for example, during idle time. Virtual tags associated with web pages are encoded to: (1) prevent a non-affiliated cache engine from using the tags; (2) prevent a competitive infomediary service from taking advantage of the "community information".

13.5 Demonstrating Value to Vendors

Vendors that subscribe to SDI **(the vendor centric version)** can provide more attractive offers/products to **users**, based on information about the wider activities/interests of a **user**, on other vendor pages, and in the physical world (of course, only to the extent that this information is authorized by the **user**). Vendors can concatenate information from the client-level proxy, the vendor-level proxy, and the ISP-level proxy. **Such data sets could even be concatenated with that of an Internet backbone**. In particular, wider information is vital for a good first-introduction to a **user**, so

that information and products can be made relevant from the start. Similarly, such information (e.g. the **user** likes this type of music, but has purchased this CD) is very valuable for vendors that sell "content-based" products, for example books, and CDs. Vendors can personalize their service with, for example, a recommender system that interoperates with profile information furnished by SDI.

We can demonstrate this value experimentally, for example we can offer a vendor a free-trial and present personalized information/advertisements to one group of SDI **users (both on the site and if desired also across the network)**, and regular advertisements etc. to another group **without the aid of SDI**. The increase in vendor revenue can be estimated from client-level monitoring of the change in purchase volume achieved with well-focused solicitations *on the vendor's own business*. **What's more, we need not even inform the vendor of these experimental solicitations as it is possible to emulate personalization of ads, purchasables portal links (and other content) on the vendor's site and across the network of SDI sites upon the client level proxy instead of the host level proxy. Other prospective ads and syndicated purchasables could likewise be virtually overlaid on the vendor's site. Thus degree of increased click through rates, transaction rates and syndication revenues could be tabulated for that vendor in advance of his/her subscribing (or even being solicited) to SDI. Portals are also potential beneficiaries of this service in light of their need to better target their partner's ads. Similarly the buyer-centric SDI subscriber's profiles can also be used to virtually personalize Web information so as to emulate the vendor's click-through if deployed personalization engine were able to read his/her profile and detect those preferences of points of advertising presence across the Web whereby comparative click-through measurements could be measured. There are a host of potential "test market" applications ranging from optimal price determination (see co-pending patent application entitled "System for the Automatic Determination of Customized Prices and Promotions" to on-line focus groups for newly unreleased products or services which can be delivered on-line or off-line through on-line solicitations and focus group surveying to targeted audiences of SDI subscribers (for which the user may also receive additional compensation for this feedback & perhaps additional personal profile details). For advertisements which simply (advertise off-line businesses, charging by click through may not properly cost justify the extent of value to the vendors which are providing this data and/or targeted advertising exposure from their respective sales. With some cooperation and trust from the vendor, it may be possible to evaluate immediate increase in a certain portion of off-line sales resulting from on line focused solicitations provided to all or a portion of that vendor's target prospects (and/or customers). WE can look at the on-line profiles of an equal core sample of those respondents and customers which projected the offer compared with those who accepted it by observing the future on line behavior patterns of a small (test) group of users, off-line behavior may be better predicted in advance of the on-line promotion and associated targeting model of related interest is market testing of new products. The following application of iamworthit user data may provide a significantly cheaper alternative to vendors (including small vendors) wishing to perform a cheap and efficient form of market testing for their products (or on-line content or applications). It is possible to run preliminary market tests on sample groups selected on the basis of similarity to other products for which buyer response is known. Random sampling of buyers may be initially performed data mining as applied for marketing research. Based upon this initial response, (as detailed in the parent case) may identify key features respondents which are typical buyers, non-buyers, proponents, non-proponents (through active survey) (for example perceive a particular feature to be particularly beneficial or undesirable or present particular complaint. Accordingly, rapid profiling (as similarly) may enable the automated or semi-automated selection of disclosed segments of users (containing particular attributes) which are likely to have (or may possess), certain response features (buyer response or survey responses) which are of interests to the vendor. (this user selection process may be multi-step and iterative based upon responses from the previous responding user). The ability to the tailor a products targeted marketing strategy to users knowing these correlations could be extremely beneficial in both on-line as well as off-line commercial environments. Additional targeting of modified versions of the product tailored to the best meet the needs of each prospective customer is also of potential value, iamworthit users may benefit by receiving additional value for providing additional personal data and or the ability . Quantifying the ultimate longer-term effects of such advertising for off-line commerce is nonetheless a difficult problem in as much as its benefits are more likely to be evidenced over a longer-term period. Thus, the benefits of revenue sharing will be quite limited in this advertising venue while there may be some useful benefits (which are much more limited via traditional off-line advertising channels performing experimental design testing (targeting differently profiled groups of users with desired solicitations). Of course in the process of selecting the target user profiles significant value may be achieved from data collected from that vendor's own on-line presence the user-centric SDI service which can aggregate and sell a portion of that data as is captured from similar (competitive) vendors.**

Also, we can monitor the performance of vendors that use SDI technology, and provide measurements/metrics for new and prospective vendors (for example on click-through rates and transaction rates) etc. All of this is possible because the system of SDI has access to a **user**'s client-side proxy machine, and the vendors themselves cannot block this collection of data.

NOTE to David -If you agree, it may be clearer o define "infomediary" in conjunction with the iamworthit section as a **user**-centric SDI which typically incorporates the iamworthit and/or community dollars models and go through and replace iamworthit elsewhere in the spec of "infomediary" use in the title "**User** "infomediary

13. 6 Business Contact Organizer

Large business centers can be very confusing places: people are often away from their desks, many different meetings are scheduled and rescheduled during a single day, people are called off to emergencies, and clients drop in for unexpected visits. The inability to coordinate people and information in such situations leads to greatly reduced efficiency.

LEIA offers a unified solution in the face of this chaos; by linking together people's schedule calendars, communications systems, contact information, and location coordinates, a system can be devised to (1) shepherd a visitor (even if unexpected) around the business center, (2) push vital contact information about the visitor to those in the company who will be visited, and (3) adaptively coordinate important meetings within the company.

Confidential - Pending

1. Sensors acquire signals

When a visitor comes to the center, he will:
a) sign in at a front desk; rather than sign a book, he logs in at a terminal that signals his presence to the company LAN.
b) receive a hand-held computer that, while presenting the visitor with useful information, will emit a constant signal enabling LEIA to pinpoint his position at all times.

On the employee side, LEIA can acquire signals from:
a) active badges
b) telephones
c) beepers
d) interactions with terminals

2. Sensors emit Location Identifiers (LID)

All sensors are connected directly to a LAN internal to the company. Only basic security measures need to be enacted.

3. Secure System translates LIDs to User Identifiers (UID)

Since the LAN is internal to the company, direct UIDs can be used and no pseudonymization is needed.

4. UIDs used to access personal profile data.

Registered visitors will likely be profiled in standard contact files containing information about their own company, position, professional interests, previous interactions with the company they're visiting, and basic personal details (e.g., photograph, wife's name).

Employees of the company using this LEIA-based system will have much more information available. In addition to the standard contact information, LEIA will have access to their phone and e-mail logs, daily calendar/schedule, work group and project details, position within the company hierarchy, and work-station notes (e.g., an unexpected "out to lunch" message could be posted on-screen by the employee before they run off).

5. Most appropriate set of information chosen

Although some higher-level intelligence is needed (e.g., for matching interest profiles between employees and visitors), the office-based system could well be run with a combination of a rule-base and optimization routine (which would be capable of coordinating conflicting schedules). Since this version of LEIA will be installed in a fairly controlled environment, the problems of location and timing will be fairly trivial, the former solved by active badges, the latter solved by smart calendar agents.

For the visitor:

If the visit is expected (and previously scheduled), LEIA will double-check the availability of the personnel to be visited and confirm or alter the schedule as needed. If the visit is unexpected, or certain employees are unavailable, LEIA will take into account the purpose of the visit and the visitor's professional interest profile to locate those available employees who would be best suited for a meeting.

Once the visit schedule has been arranged (although it could be adaptively altered during the course of the visit), the visitor will be fed information via the hand-held computer they've been issued. This might give them a visual representation of the building's layout, along with the best route to the next office to be visited (a more advanced building might simply have computer displays in its corridors capable of flashing directional signals to the visitor as he passes them). In addition, it might give the visitor basic information about the person he's about to visit (for example, they might have interests in common, or have gone to the same university).

For the employee to be visited:

Whether or not the visit was previously scheduled, LEIA communicates with the employee's personal calendar agent and double-checks the employee's current location. If the employee is at a fixed location and has some free time, the visit is scheduled. Obviously, his connection to the visitor will determine what information, in addition to the standard contact data, should be pushed to his workstation.

For multiple employees requiring a meeting:

Given that multiple employees have signed up for a meeting, or that a single employee has sent out a signal indicating the need for a meeting and its priority (standard or emergency), LEIA gets to work coordinating the details. Cell phone and beeper signals give employee locations outside the office, active badges signal employees already at the office. Calendar agents are contacted for availability; a standard meeting might depend on people having time slots free, whereas an emergency meeting would be given priority status, and could bump items already scheduled (given that they have been assigned low priority).
An optimal meeting time and place (or teleconnection for remote employees) is set up, and any pre-meeting information is forwarded to the appropriate workstations. ***Co-pending patent application entitled "System and Method for Customized Electronic Identification of Desirable Objects" section entitled "Virtual Communities System" (VCS) describes a useful technique to be deployed in this application. For arranging multi-person meeting or one-on-one particularly within a large organization VCS provides the ability to efficiently leverage profile data about the user whose professional interest profiles best match that of the purpose of the meeting (submitted as a summary description previously referred to by VCS as a "feeler message").***

6. Information Delivered

For the visitor:
Localized radio beacons tap into his hand-held computer, feeding him directional and contact information.

Exhibit 1003

For the employee being visited:
Information about the visitor and scheduling data are fed to his workstation via the LAN. If he is remote, beeper or phone messages could be sent out.

For employees involved with a meeting:
Meeting information and calendar data are sent via the LAN (or by phone if an employee is remote).

AUTOMATED LEARNING OF USER BEHAVIOR AS PART OF AN ENHANCED MEETING SCHEDULER AND CORRESPONDENCE FILTER

In order to utilize LEIA as a user agent which is integral and ubiquitous within the work environment, it is necessary to make its learning as seamless and invisible from the user's perspective as possible.

The next level in applying the capabilities of LEIA is in being able to implicitly learn the context of the user's present temporal activities as well as establish relative priorities of the activities that the user (and others who intend to interface with the user) are engaged in and from the user's previous behavior to similar conditions automatically generate appropriateness functions (or rules) to automate the process of handling meeting schedules and filtering real time requests for correspondence with the user. It is possible to infer:

1. The content profiles which are associated with the user's present activities.

2. (Often obviously) The nature of the user's present activities by observing what sorts of actions the user is engaged in, in the office.

These content profiles and activity indicators may be passively observed by the user's present interactions (and timing thereof) with his/her PC, other smart appliances and data and voice communications which the user engages in (telephonically or from standard face to face dialogues) e.g., suggesting when the user is reading the morning news, checking phone messages, responding to e-mail, engaged in a particular project, etc. If these clues are not observable (or not positively identifiable) cyclical time dependent patterns may be used as implicit indicators of the user's present activities (see the above section entitled "Further Applications of LEIA"). These same methods for utilizing time pattern elements which are used to predict the location of the user may be readily extended to predict or help to substantiate other statistical user variables relating to the types of actions and content related profiles characterizing the user's temporal behavior patterns. The following application of these techniques is used to enhance the automatic meeting scheduler:

Utilizing the identity of the task or request, the activity and content attributes of the task or request profiles (target object profiles) may be developed utilizing the priority of that task relative to the user. This priority is estimated from the priority of the task/request (and its associated attributes) over other tasks/requests by the user. In order to better substantiate the relative importance as inferred regarding these activities, the system should also display the scalar ratings which the user could adjust manually.
Nonetheless, there may be a degree of uncertainty in the user's intentions and LEIA's inference of the user's intentions. Moreover, changeability by the user or LEIA may occur perhaps with increased frequency under certain variables like the identity of the requestor, the time of day, context of the users other present activities or when other meetings and obligations are pending which possess a relatively greater priority with the user (particularly if they often occur spontaneously). In order to take this uncertainty factor into consideration, an overall statistical estimate may then be calculated taking into account the predicted statistical probability of each party ultimately being available for the meeting. This serves the purpose of both refining the automatic scheduling process and informing would be visitors before hand as to the relative importance that their prospective meeting is (or becomes as the meeting time approaches) to the employee as well as the certainty of the meeting. This feature my be particularly useful for impromptu visitors where considerable uncertainty exists as to whether and to what degree the visitor is imposing upon the employee's schedule.

The temporal context of user behavior (using temporal time series analysis in conjunction with use of content analysis and user behavior assessment) may be both important predictors as to whether a prospective meeting of a disclosed purpose is ideal for the employee and at which most likely times. By far its most valuable application

however is in its ability to perform content analysis and user behavior assessment in real time in order to enhance LEIA's ability to determine whether and to what degree a would be unannounced visitor or telephone or intercom caller's unannounced contact is relevant to the employee's present activities and if not, to coordinate with the visitor's agent the time(s) when it would be most relevant to the employees' anticipated activities. For example, there are certain activities during which an employee doesn't wish to be often disturbed e.g., reading the morning news or taking a coffee break, calling home, responding to an important e-mail message, etc. Often within an organization individuals will attempt to speak with one another. As such intended duration affects priority as does the identity of the individual in the organization. Often this variable can be automatically predicted by LEIA. In such a situation the relative priority of the user's pre-existing priority tasks is weighted against the priority of the requested tasks prior to notification. The above application can be readily extended as well to phone calls in which the user could be automatically prompted by LEIA to disclose his/her purpose of the call and/or his/her identity if necessary. This application would be a useful extension to the telephony variation of the e-mail filter described in co-pending patent application entitled "System and Method for Customized Electronic Identification of Desirable Objects" in which the above additional inputs about employees and visitors may provide additional useful feedback in automatically determining the rules dictating how to best handle the call or forward it on to a more appropriate employee on behalf of the employee. If there is a certain degree of uncertainty in the agent's priority determination or assessment of the task affecting the most appropriate rule the employee may observe the rule just prior to its automatic implementation. The present value of the contextual employee activity assessment method if extended to advertising could be significant. For example, the knowledge of the context of the user's present activities could determine what type and when an advertisement targeted to an employee (e.g., a decision maker) will most likely elicit the best response (where the advertisement matches the professional interest profile). As discussed, direct voice, telephony and e-mail dialogs may provide very useful clues (using natural language techniques) as to the particular profile interest summary which the user currently is experiencing at that moment both what types of user interactions and advertising is likely to be most appropriate. Co-pending patent application entitled "System and Method for Customized Electronic Identification of Desirable Objects" suggests as a variation of virtual community creation, the use of present dialogue profiles of users as a means for determining whether/which on-going meeting dialogs within the organization would be most relevant for the user to be introduced to (either physically or virtually). For example, virtual work groups (e.g., over industry-wide extranets) could utilize the above profiling approach in the creation of the group (by watching the employees professional interest summaries with the project profile) as well as in the dynamic adhoc introductions of other experts who have a professional interest profile which matches the content of a particular section of the meeting (which may be gleaned passively or in accordance with a specific query or request to aid a specific task). Co-pending patent application "System and Method for Customized Identification of Desirable Objects" describes how queries may be routed to knowledge domain experts in an organizational environment. LEIA may also identify and prioritize individuals who may be useful candidates to join into a meeting or address a specific request, who are physically proximal to the meeting location.

DOCKET - LEIA

14. Buyer Infomediary: The Iamworthit Model

In this section we describe a key extension of the buyer-centric SDI model, that provides buyers with an additional incentive to provide information to vendors/advertising networks, in addition to the benefits from receiving well-targeted information and products. We introduce a new currency, termed "community dollars", that allows vendors/advertisers to compensate buyers for providing information— but **ties** the compensation to buyers making a purchase, so that: (1) buyers are incentivized to provide information that allows vendors to push relevant advertisements/products; (2) buyers will also be more likely to make purchases at a site **or a coalition of sites** for which they can receive discounts via community dollars; (3) providing buyers with community dollars will increase the number of hits to a site. **Iamworthit's ability to drive competition among vendors has potentially revolutionary implications in many commercial domains** ***in as much as markets are made more efficient by a novel and powerful method of "overlaying" an auction-based platform within the context of standard ecommerce environments (a commercially compelling advantage to buyers which should attract a significant number of subscribers). Applications within*** **business to consumer and business to business domains** ***are suggested within the scope of the below described methodology.***

In addition, we propose the deployment of (1) vis-a-vie the user-centric SDI service, time of purchase competition between vendors as well as sharing between buyers of offer acceptance/rejection information from

either vendors or buyers (2) vis-a-vie the vendor centric SDI service, intelligent extensions of the present collaborative filtering method in order to enable advertisers and vendors vis-a-vie the vendor-centric SDI services to utilize profile data of the buyer (to the extent that this information is provided by the buyer) in order to customize pricing and promotions for that buyer thus optimizing profitability for that buyer in light of his/her user profile data as disclosed to the advertiser or vendor.

14.1 Time-of-purchase Competition

The time of purchase competition feature is designed to be accessable to a user within the context of or *most* any information access mode relating to finding or browsing purchasables on the Internet e.g. Web or portal browsing, receiving email or "push" content, submitting queries (e.g. specific names, categories and/or desired features (or combinations of the above) of products which the user *identified as being or present interest.*

During any of the above information access modes when a **buyer** specifies this "time-of-purchase" competition option to her SDI client proxy, SDI can automatically provide competitors with information about a **buyer's** product or service requirements, and a **buyer's** profile, **as well as on *particular* vendor specific offers the buyer is observing or contemplating** before a **buyer** makes a purchase. This will facilitate competition between vendors, and can lead to better prices and offers for **buyers**. We allow vendors to opt-out of this scheme, and prevent the system of SDI from informing competitors about offers/purchase-requests. Of course, in this case their **buyer** can decide not to visit such sites. The system of time-of-purchase competition enhances vendor competition, and can also help to reduce the costs of entry into a market, because name-recognition becomes less important. New vendors can simply register with the "purchase referral" service, and cherry pick the products that they specialize in.

This is a "next-generation" e-commerce service. Current shop-bots, for example "Junglee" at Amazon.com, www.shoptheweb.amazon.com, provides a static comparison shopping service, using static price information. A **buyer** can specify a product, and receive price information about the product from different suppliers. However, there is no dynamic competition on price or features. The buyer driven service for flights offered by www.priceline.com is more dynamic, in that a seller is found to match the price that a buyer bids, but does not promote competition between sellers. In fact the sellers can make excess profits from the pricing errors made by **buyers**.

We can use profile information, and historical transaction information for similar transactions, together with the customer price/promotion algorithm disclosed in co-pending patent **"System for the Automatic Determination of Customized Prices and Promotions"** to negotiate on a deal with a vendor that will optimize the value to the **buyer**. Profiling of vendors, and **buyer** transactions, can allow **buyers** to avoid making bids that are too high and losing value (airlines in priceline.com can profit from inaccurate **buyer** bids).

We enable vendors with competing products or services to receive automatic notification when a **buyer** is interested in the purchase of a particular product. A vendor can also receive information on the profile of a **buyer**, and the offers made by other vendors; and submit counter-offers to a **buyer** via the **buyer's** SDI-enabled client. The **buyer** can then be presented with a final set of offers, before making a purchase decision. In a variation, the **buyer** may actually initiate the process submitting an initial offer to a vendor or a collection of vendors for an item she/he is interested in (whether in a traditional store-front or even an on-line auction house). iamworthit protects the value of a **buyer**'s profile through competition, because instead of using a profile to practice price discrimination and extract excess profits from a **buyer**, a vendor must use a profile to offer the most appropriate product to the **buyer** at a competitive price.

The iamworthit model provides an environment in which a market is established that is dictated by **buyers** by facilitating competition by/between vendors vis-a-vie the release of information pertaining to offers presented to **buyers** of a disclosed profile (or more typically pseudonymous **user profiles**) by vendors to competitive vendors selling the same products or services. In addition to enabling a more competitive market for electronic commerce, the iamworthit model also increases the value of **buyer** information, that the **buyer** can choose to disclose to vendors to enhance customer targeted promotions and targeted pricing and other on-line marketing strategies, and also to withhold from vendors information in order to elicit optimal offers.

For example, the information that a **buyer** discloses to a vendor could include its sensitivity to discount offers, customer loyalty with other vendors, value responsiveness (bargain driven), high quantity purchases (for only those categories which the **buyer** makes frequent or large purchases).

When a **buyer** selects the SDI time-of-purchase competition option the client software monitors the **buyer's** web actions for queries and browsing related to purchases. The client identifies other vendors with similar products or services, either using a static "web index" that maintains vendors in particular product domains, or through dynamic profile matches between the target object profile of the web site that the **buyer** is currently browsing and target object profiles of the web sites of other SDI-enabled vendors.

Purchasables which are similar or the same to another purchasable on other vendor's web sites must be identified in order to determine which notification of a browsing action or purchasing request by the **buyer** is of relevant interest to those vendors. This may be achieved by solely relying upon static information that classifies vendors into product-domain categories, such as a Web directory **or solely upon meta-data embedded in the pages.** Key terms and other attributes associated with the items, automatic classification and clustering techniques applying usage statistics and content (associative, numeric and textual attributes as described on the parent issued patent) may be further deployed as additional techniques for purposes of identifying similarity at the level of the target objects. **Such a technique utilizing XML for profiling (updating those profiles dynamically) is disclosed above.** Classification and clustering techniques can be deployed to identify similarity between vendors at the level of target objects.

XML tags for products in the future also will contain the product identification codes of individual purchasables. Nonetheless clustering techniques, however, could be used for identifying "similar" products which are not specifically identical but are close enough to elicit comparative interest on the part of the user for purposes of the time-of-purchase vendor competition (e.g. a standard cell phone feature compared with one with a Web browser, email and a universal in-box). If some of the other competing vendors do not respond with counter offers (or *have chosen to opt-out* of the service), it may be advantageous (as motivation for them to actively participate) to ignore existing (static) offers presently on their Web sites, for those items. *The system (at the SDI client-level proxy) could even actively block vendor's sites (from their HTML sources) from the user. Conversely, i*n an alternative variation, the *methods of the present system may also be deployed to* also reveal these competitive items *to the user in the form of "static Offers"* as well (in similar fashion to that of a "shop bot"). If/ when the service is also used for this particular comparison shopping all "similar" items *may be released as well if the user desires this additional feature.* When the service is also used for this particular comparison shopping purpose it is useful if all the features are presented pertaining to the vendor such as geography, available methods (*and speed*) of shipment, order/transaction *methods and* security, user ratings/annotated comments etc. as well as product features exclusively.

Vendors are notified, and provided with the ability to access the profile of the **buyer** ***(as the buyer deems appropriate)***, either with client-level processing or through the release of an anonymous profile to the vendor. Vendors typically will wish to construct offers through a rule-based engine, data-mining techniques, or automatic collaborative filtering techniques, as disclosed in co-pending patent application "System for Automatic Determination of Customized Prices and Promotions" and U.S. Patent #5,754,939, "System for Generation of **User profile**s for a System for Customized Electronic Identification of Desirable Objects" as such techniques may be deployed by the vendor directly or via the Secure Data Interchange representing the interests of the ***vendors***.

User profile information may include a temporal profile of the **buyer**'s present activities, including search terms, recent page navigations, what pages is the **buyer** observing presently (and the profile of this page) or even his/her present physical location as well as the general **user profile**. ***Any*** portion of the above particularly the latter ***two*** may of course be withheld from the vendor). ***Or the buyer may be more generous in providing this information about the various static or dynamic features. This vendor information is of useful advantage by providing them with the ability to target buyers, not only based upon their preferences and interests, but also within the temporal context of when they are most likely to be receptive to offers (thus providing also a benefit/service to the buyer as well as enabling the competing vendors to increase the price point at which their competition begins, knowing that the buyer is in an optimally receptive condition to accept offers for that item). For this reason, it is reasonable to combine the present vendor competition scheme with the buyer-centric SDI service in which vendors target advertise to buyers based upon their profiles. If page view information is provided to vendors (in this case offers) they (or the vendors agents) may target buyers based upon contextual matching of the buyer's current page views with the most relevant matching content of the offer/promotion. However, in this variation, the vendor not the buyer***

is initiating the offer thus the service automatically informs all competitive vendors of the offer (and likewise with all subsequent offers). In the preferred implementation, vendors are also provided with a (client or web-based) rules interface which enables the vendors to input pre-stated rules with which the system may solicit and respond to competitive offers automatically. **In this way data mining may be performed in order for the vendor to determine what a user (or all users) by attribute, tend to best respond to by product, product feature, features or services of the vendor, price, etc.** If pre-stated rules are used to automatically respond to a notification with a competitive offer, the nature and degree of discount is typically determined in accordance with the nature and degree of the original or previous offer and/or the **user profile** as disclosed by the client-level proxy/server to that vendor. In lieu of manually entered rules, co-pending patent application entitled "System for **the Automatic Determination of** Customized Prices and Promotions" or another similar algorithmic methodology may be used as an aid by the vendor in order to automatically determine a competitive offer (or subsequent responses thereto) **as well as an aid to the vendor in selecting optimal rules.** These techniques can also be used **such that** predictive metrics as to selection, price and promotional type, may be **optimally** determined in relation to the individual **buyer** or specific relevant **user profile** attributes, for example, in accordance with a data analysis expert of the vendor (or representing the vendor via SDI) analyzing randomized versions of **user profiles** and randomized aggregate statistics. ***These metrics*** **may be linked with actual behavioral statistics of the buyers on the vendors' web site. In order to further drive competition between vendors, iamworthit provides each competitor details of the final transaction which the buyer chose to accept. Accordingly, in an attempt to better improve their competitive advantage in the competitive bidding process and ultimately presenting "appealing" offers for each individual, each vendor may also wish to perform all of the above data mining operations on all of the competitive offer data as well as the vendor's own data which pertains to those interactive users.**

From the buyer's perspective, there is an elaborate methodology for the automatic determination of the particular buyer information which, if released, will elicit the most optimal offers from vendors which can be determined by utilizing that same type of pricing model which the vendor is likely to use (for example that same method deployed in co-pending patent entitled "System for the Automatic Determination of Customized Prices and Promotions" can instead be tailored to determine which metrics in the user's profile decrease the overall price point and/or increase overall value over an "average" user (with no profile disclosure).

It is even conceivable that vendors could (upon user permission) be allowed to directly identify plus enter into a live discussion with the user during the time-of-purchase competition process.

A variation of the system **was even** is further disclosed **in the above co-pending application** which was intended (according to the **above referenced** co-pending specification) as an electronic assistant to tele-marketers and other sales persons to determine offers and counter-offers which are automatically generated in response to (for example) rejections of the previous offer as well as counter offers by the **buyer**. This dynamic system was originally designed for salespersons to optimize the expected profit from each customer (in view of the general **user profile** and the offer **buyer** responses up to that point in the negotiation). As such, this technique could be readily extended to the current application in which the previous offers up to that point may instead originate from other vendors (instead of a single one), thus the system responses may be affected by the **user profile** as well as the offer response pairs up to that point in the negotiation process.

It is likely that vendors will not compete on price alone, but rather through added-value services such as offering loyalty bonuses, cross-sells, and two-for-one offer and added features **as well as service advantages such as fast delivery, guaranteed service/maintenance, warrantees etc.** Vendors will choose this mode of selling to prevent simple price-comparison at the client. ***Conversely, in some instances*** vendors may attempt to eliminate the features in order to create the perception of a better deal through marginal price reductions, ***then possibly add these features as counter offers to other vendors in which the same price may be maintained. Accordingly the above referenced algorithm as provided by the customized prices and promotions scheme, may offer valuable competitive insights to the vendor in being able to assemble more appealing cross-sells/up-sells and more complex offers (including multiple items which are adjusted by the system to be extremely desirable to the buyer) more effectively than the competition.*** Therefore the client will receive offers from multiple vendors, and after initial filtering of **the large collection of** offers, present a choice set to the **buyer**. **Additionally,** ***not always will an item at a lower price compel the user to select that item.*** **Some competing** ***offers*** **may be for items which in some respects are of less value to**

the buyer (e.g. inferior quality, service or vendor support). Thus, buyers may present ratings or annotations for items which are captured by the ISP level proxy ***which averages the ratings (see the parent case)*** **or if annotations which are also provided by users are preferentially selected and presented based upon the degree of** ***buyer*** **experience,** ***(or particularly comparative experience)*** **across multiple competing products, for example,** ***within the context of buyers whose*** **attributes (which may be important pricing features) are similar to that of the buyer. In order to provide an additional competitive bargaining advantage, this annotated information may, for example, be disclosed to the competing vendors if it provides generally a positive comparison in light of these other vendor's products or services which the user is presently using the service to bargain with.** ***Buyer's clients (client agents) may also further be programmed to monitor and record all previous "opportune" offers provided to other users in the past and notify a user if an offer had previously been made for the same item at a better price or value than that which the present buyer has received through the present competition scheme. The buyer may then request the lower price if necessary with release or further release of certain of his /her profile attributes, if such is further needed by the vendor in order to justify the lower price. If iamworthit detects that discrepancies in the pricing for a given vendor's item occurs in correlation with certain identifiable user profile information (not arbitrarily), it will in the future initially disclose that data of any other which is determined from past buyer interactions to be advantageous for the buyer. Should the vendor exhibit price or offer discrimination (i.e. by having provided more favorable items to other previous buyers, the buyer may demand the lowest price. If the vendor ignores the request and the user has not received (or does not wish to accept) another vendor's offer for a lower price and the iamworthit infomediary service may apply pressure upon the vendor such as creating a buyer coalition and/or making a believable threat (as described below) which has the ability to deny future business to that vendor. If the vendor is able to present an exceptionally attractive offer the buyer may instead wish to stipulate that the offer is valid only if the time of purchase competition feature is not used by the user. iamworthit's statistics based price prediction scheme can then predict if this decision would serve the user's best interests or not based upon the user's purchasing objectives.***

Furthermore, we can allow vendors to offer payment to a client in return for displaying an offer to the **buyer**, and vendors can also bid for space on the **buyer's** web portal which is often represented as a profile associated with a pseudonym **in** conjunction with a description of the ad space. The purchasing decisions of the **buyer** may be performed by an electronic representative of the **buyer's** wishes (as "**buyer** agent") implementing the techniques of pricing/promotion selection algorithms completely autonomously on behalf of the **buyer**. However, the best offer can only be presented to a **buyer** to the extent that the SDI client level software understands a **buyer's** model of "value", and can make appropriate tradeoffs between product features and price (as implicitly inferred by the system through the above suggested techniques or explicitly stated by the **buyer** in advance). Nonetheless, this is a hard problem, and we expect that the **buyer** will often need to make a final product choice decision **which could be a default in which the system would defer judgement to the buyer if it's statistical confidence as to the buyer wishes below a certain threshold.**

The collaborative filtering techniques described in patent "System for Automatic Determination of Customized Objects and Promotions", can allow a **buyer's** client-level proxy server, termed the **buyer** agent in this section, to automatically analyze offers. The system can also be used to send initial offers to vendors, on the basis of historical information about the transactions that have been performed between other **buyers** ***(which include the benefit of his/her complete user profile data)*** and the vendor. Offers can (of course) be sent to a vendor and its competitors. Finally, after offers that are received from vendors are pre-screened, they can be automatically ranked for value—using a combined quality and price metric (again judged within a collaborative filtering framework). ***Buyer feedback e.g. average ratings and annotations from previous buyers may form an additional criteria.*** The goal is to leverage the database of other offers that have been accepted by **buyers** in the past, and form a model of vendors, to determine whether or not a **buyer** has received good offers (i.e. we can exchange information within the system of Secure Data Interchange, and making more information available increases the efficiency of the market). Offers can be filtered and presented to a **buyer** in rank order.

In accordance with the above description (the parent patent application describing the pseudonymous user database) provide a means to equip advertisers with the ability to query, identify and negotiate delivery of a targeted ad to desirable buyers. Desired behavior and other features also reveals buyer data to vendors selectively based upon the determination by the pricing model of which data provides a competitive negotiating advantage to the buyer incorporating offer response information to user profile data elicited by that particular vendor in the past. This database may also incorporate (if advantageous to the buyer) offer information pertaining to previous time of purchase vendor competition negotiations e.g. offers made in the past by vendors for similar products as those

provided by that prospective advertiser (even from the same vendor) which are at least as competitive as offers made to that user or "similar" users. Because of the competitive nature of this environment, vendors will wish to leverage all of this information and using their own pricing algorithms and data mining tools, craft optimal offers through the pseudonymous buyer database. As indicated, the buyer equipped with these same algorithms in combination with all of the historical offer data from previous similarly profiled buyers could anticipate offers which the vendor is likely to make to the user and ultimately be willing to accept. **To this extent the user may be properly equipped with tools in which the user is able to provide useful information about his/her wishes as it pertains to particular types of competitive offers which the system can anticipate in advance given the particular user profile data as disclosed on a vendor by vendor basis. Typically the resulting purchase recommendations are presented as suggestions on the user interface to help the user's assessment of valuing the offer as a "good deal"as well as to accurately represent his/her own preferences. Some users may, however, use this feature as part of a fully automated (agent mediated) buying process which is performed in asynchronous fashion (as described below, this typically is only necessary if the time-of-purchase system is deployed in order to create buyer coalitions prior to negotiating with the vendors. Of course the time of purchase competition enabled buyer will distribute the offers in order to elicit vendor competition.**

If the **buyer** so desires, the client-level proxy can also automatically notify these vendors if/when a particular offer is about to be accepted by the **buyer**. For example, a time delay response in the client-level proxy actually processing the order requests could allow vendors a final opportunity to present another competitive offer to the **buyer**. In another less optimal variation, vendors are notified only upon the **buyer** agreeing to accept an initial offer received. **A secure function within the buyers client residing within the Secure Sockets Layer (SSL) is able to authenticate to the vendor's receiving the notification that the offer is true and untampered by the buyer through the transmission of a secure encrypted message.**

As an additional service to **buyers** the SDI-level proxy server can perform analysis on the offers that a **buyer** receives, through comparison with offers that have been received by other **buyers** with the best offer that has been received by any **buyer** for the same product, and with the typical offer received by a **buyer** with a similar profile to the **buyer**. This can be useful to a **buyer** because it will allow the **buyer** to reject all offers if they are non-competitive. The SDI-level proxy could also automatically identify for **buyers** the profile attributes that promote good offers, and the profile attributes that promote bad offers, as an informational service to enable **buyers** to gain better offers in the future, either through revealing certain information or changing behavior to attain favorable profiles.

Buyers might also form "buyer coalitions", on the basis of automatically detected synergies between their requests **and predicted similarities between them**. This can give buyers more leverage in negotiation with a vendor. **In the automatic evaluation of how "good" an offer is (how competitively it matches the market demand for that coalition's needs) in iamworthit's historical user information, it must also utilize information about the effect that size of the coalition has upon the market's responsiveness. Within the context of the present framework, the simplest version of the system involves creating the coalition around a particular item and, however, it also conceivable that it could be formed at the vendor level in which the coalition members are interested in any item(s) provided by the vendor and vendor competition is elicited by the coalition across all products which its members demonstrate interest in and which are provided by the competitive vendors. Given a sufficiently large subscriber base, buyer coalitions may be automatically created in real-time and in ad-hoc fashion (i.e. from the iamworthit buyers who are interested in a particular type of item at that particular time). Buyer coalitions typically are able to exert *significantly greater* pressure upon vendors the larger they are *(this fact applies as well to multiple company b to b buyer coalitions)*. As such iamworthit may also notify other members which in accordance with their user profiles are likely to have an interest in a particular item (or items) in which a coalition of members have demonstrated interest in order to increase the size of the coalition. Similarly, the aggregate user profile of all members of the coalition may be used in conjunction with collaborative filtering in order to recommend those items which collectively best match the preferences of the coalition collectively and thus are able to provide the best deal for the most items possible to that particular coalition. Thus the coalition model is extended to placing competitive pressure upon vendors across multiple items. iamworthit must also utilize the above modeling techniques in order to establish pricing models based on the value of these different items in light of the total monetary value which the coalition represents to the vendor collectively. In addition to inferring *buyer* interest entirely passively, it is possible for *buyers* to manually edit their profile, e.g. by modifying or actively selecting categories of purchasables of interest which she/he would like to be notified of (e.g. via pager) if/when a buyer coalition for a purchasable of that type is being established. It is further possible to increase the size of buyer coalitions by allowing a certain amount of time for buyer's to respond to offers to**

join that particular coalition (e.g. 24 hours). Typically a couple of days is all the time *that is required* to achieve a substantially maximum *buyer* response. Also, because some vendors may wish to be (automatically) informed if/when substantially large coalitions approach the vendor, there may be an additional marginal price advantage by providing an additional period for vendors to reply. The buyer may wish to allow the iamworthit time of purchase vendor competition system to operate in fully automatic mode whereby certain "acceptance parameters" are provided to enable the buyer agent to act fully or autonomously on behalf of the buyer's stated (or approved) parameters. This technique is likely to provide a central function particularly in most b to b iamworthit commercial *applications.* The b to b commercial domain may exemplify a couple of other novel features (which are not exclusively limited to b to b). For example, because replenishment of products or supplies suggest much greater *repetition and thus* predictability, it is *sometimes* useful to buyers to also leverage as part of the negotiating process an advanced commitment of buyer loyalty over an extended period of time. In the unlikely event of the buyer changing his/her vendor loyalty after considerable value has been provided to the buyer, the vendor may control the right to acquire certain assets of the buyer (e.g. corporate stock) as pre-agreed collateral for the vendor. *Also,* in the b to b domain iamworthit acts essentially as (or in conjunction with) a buyer-side portal to a web-wide vendor resource, though it could also be integrated as part of a seller side-portal interface tool as well.

As an additional service to vendors SDI can provide enhanced profile information, aggregated from other vendors, to enable vendors to provide better focused offers than can be provided on the basis of the profile information directly associated with the pseudonym of a **buyer**. Certain portions of the **user profile** data that is unavailable for direct collection by the vendor (such as information that is collected on other sites including, in particular, competitive vendor sites) may reveal important information which enables the vendor to better target that **buyer**. As such the secure data interchange representing the collective **buyers** may aggregate, analyze and sell this data to the vendor **so long as the release of such information does not negatively affect thepredicted pricing or value levels for that buyer as performed by the above type of pricing algorithm.**

SDI-2015-04-09

Issues Regarding Accessing And Using Medical Records For Advertising And Insurance Quality (an exemplary iamworthit commercial application).

Problem

Increasingly, insurance companies are denying claims for medical services. Hospitalized patients frequently have in-patient days denied, in part or totally, and associated services and studies may also be denied. Days of partial or total denial are especially important for non-DRG (Diagnosis Related Group, Non-Medicare) insurance subscribers. Also in-patient and out-patient testing and services are frequently denied at the discretion of insurance administrators. These denials persist despite consistent efforts by hospitals and hospital personnel to properly document the appropriate protocols to alleviate patients' illnesses. Similarly, out-patient medical care in doctors' offices is not infrequently denied, and to the extent of prohibiting effective diagnosis and treatment. At times, hospitals as well as out-patient office personnel do not have the necessary information to know, prior to treatment, which services are eligible for reimbursement, partially because of the multitude of coverage patterns. The providers must frequently appeal non-payments retrospectively, by letters, further documentation and repeated telephone calls, all of which may or may not be effective, apparently depending on the capriciousness and the judgment of the particular clerk or medical personnel in the carrier's office. These situations have produced enormous pressures to improve communications by extracting detailed data from providers *pertaining to carriers failing to support appropriate and necessary health care delivery and documentation of how these insurance issues have caused definitive health problems. This* data will be used to exert pressure on insurance carriers. In view of these major economic pressures, hospitals and other providers of medical services will definitely be most cooperative in providing detailed billing and clinical information with appropriate release consents, or alternatively through deployment and utilization of the proprietary techniques for privatizing patients' records using a pseudonymous proxy server. This information will be provided by the billing and medical records departments of the hospital and also, in the case of hospital employees, from the Human Resources Department of the hospital for reimbursement for medical services for hospital personnel. The minimal information should include the billing record for those in-hospital days or the specific services/tests denied and the discharge summary of that hospitalization. Based on reviews of this material, further portions of

the in-hospital clinical record may be requested. (Additionally, the iamworthit service provides for the utilization of the above mentioned proxy server for collecting and maintaining pseudonymous patient data whereby advertising revenues from industry advertisers accessing/targeting desired user profiles may be shared with the hospital and potentially also the patient, thus further insisting the hospital to provide this needed patient data).

A compounding problem is the lack of information provided prospectively by insurance companies to hospitals, physicians and other providers. Subscribers and providers have the right to know, in detail, the number and kinds of services covered, and the depth of the insurance coverage including the length of stay. This information is legally accessible and can be made available not only to the patient, but also to his/her physician. *On exemplary commercial need for the present service by employers who wish to reinsure their employees with quality coverage.*

Solution

Coverage Probability Statistics Provided by a Proposed Nationwide Public Data Bank of Health Insurance Carriers

In order to protect the consumer, a solution proposed to address a timely and overwhelming need on the part of consumers to establish a control data bank, accessible to the consumer, which reveals general and specific instances of miscarriages by health insurance carriers of their feduciary obligations to provide timely payment for essential health care services. *Based upon data regarding the insurance company's stated policy and statistical analysis of the ultimately approved or disapproved coverage, performance criteria may be determined pertaining to the percentage likelihood of the insurance company covering certain services and length of hospital stays for future services to patients. In order to provide a robust and comprehensive statistical analysis, it is also required that a certain core sample of non-coverage denied service be collected in order to determine a variety of correlation criteria which affect the probability of coverage in addition to the identity of the carrier, also the type of prescribed medical care for associated illness wherein coverage for hospital stay or treatment is denied. Other correlations are possible, e.g., the medical history of the patient, the family history, etc. Of additional importance, statistics may also be provided which indicate the degree of adherence (integrity) of an insurer to provide coverage and to provide it with the level of depth of coverage which was purported by the insurer according to its standard claims policy.*

The iamworthit insurance service is accordingly designed with two *primary* case objectives:

1) **Provide detailed statistics based on robust historical data regarding realistic expectations which the patient can anticipate regarding acceptance, denial and quality *on* the basis of various types of coverage given the context of the insurance needs and requirements of the buyer *and* based upon the specific *services and depth of coverage purported by the* insurer and any associated medical information regarding the buyer. The service could even anticipate, *by review of* the collective medical statistics the most likely type of *coverage the* buyer may need, based on his/her medical record specifically *within the context of* the associated *determined* probability of coverage and quality thereof *for each insurer* in view of *this* medical history of the buyer. In addition, iamworthit may provide a general informational resource where buyers can access such statistical information by carrier. Data mining tools may enable the buyer to extrapolate the desired statistical correlations as to his/her *type of profile of* medical *or specific features of his/her medical condition* or specific type of coverage and thus *in light of this robust data better equipping the buyer to* determine which insurance provider *and associated policy* is most appropriate for his/her *specific health care coverage* needs;**

2) **This information may be accessed in real time by the buyer *(or for e.g. buyers, employer)* while browsing specific insurers' sites, receiving specific offers by insurers (typically requiring certain medical information) or browsing an insurance retail portal. Typically, buyers of the present service will combine their use of the present browsing-based information with the Time-of-Purchase Vendor Competition Services (described below). Additionally, a rating feature may also be provided which enables the buyer to rate (and subsequent buyers to observe the averaged ratings) the various vendors by various types of criteria relevant to the buyer's experience in receiving acceptance, quality and depth of coverage (as well as how well the insurer adhered to promises for coverage as stated in its standard policy for coverage), o*f course, this feature could be used if for insurance products on-line for later purchase off-line.**

Iamworthit's time-of-purchase vendor competition application to medical insurance (as applies similarly to numerous other ecommerce commercial domains) involves the following protocol:
The iamworthit enabled consumer provides specific needs or requirements as to what types of medical insurance products or areas of coverage are desired. These requests may be submitted either to a portal (or "mall") which connects the buyer to an extensive resource of medical insurance providers, or, alternatively, these requests may be submitted to iamworthit which provides its own default portal to these resources (which likely exists and is marketed as an independent marketplace for insurance purchasers *on its own behalf*). Upon submitting these buyer requests, iamworthit is able to provide numerous competitive insurance providers offering similar insurance products which in this application may be accompanied by certain credentials (such as the buyer's age *and many other clinical parameters).* Additionally, the pseudonymized digital medical records of the buyer (which may also be requested by an insurer, typically in addition or instead of submitting a questionnaire regarding the buyer's health such as certain pre-existing medical conditions). Typically, patient medical records are a prerequisite to submitting an offer. The insurers may then provide specific offers for their respective products. Additionally, the buyer may browse various insurer's sites, providing pseudonymous medical data to the extent (and type) that it is requested and, specifically, to which providers the buyer desires. As offers from insurance providers are collected (in response to requests and/or browsing interactions), the time of purchase vendor competition is elicited whereby each insurance provider is automatically notified as to the most valuable offer(s) provided by a competitor(s) and are given the opportunity to submit counter offers accordingly. It should be noted that the determination of the value may be in accordance with certain buyer prescribed specifications such as price or certain expert determined criteria which consider either the price in combination with completeness and depth of coverage and, if relevant, the level of satisfaction of the buyer's request criteria. Additionally, iamworthit provides the ability for buyers to share the information of offers they have received in accordance with the associated medical profile and/or response to health related questionnaire information, in order to determine if an offer which is submitted by an insurer inappropriately discriminates against the current buyer in terms of price or depth/comprehensiveness of coverage, or where such discrimination occurs in conjunction with buyer medical data which is prohibited by insurance regulatory guidelines as a valid criteria for such discrimination. Iamworthit's insurer coverage statistics (which the buyer may typically wish to observe in conjunction with the present time of purchase vendor competition functionality) are an additional important value determination factor (which may be factored in as an additional weighting criteria as numerically prescribed by the service (or by the buyer) and/or accordingly for purposes of setting a threshold by which insurers become disqualified as candidates for the buyer's business. In a variation, as is provided by the current model, buyers may instead first receive the historical offer information from other "similar" medically profiled buyers in order to submit a bid to that carrier which offered the most relevant and beneficial offer as well as all of the other competitive insurers. As is also provided according to the current scheme, the present iamworthit time of purchase vendor competition facilities may operate fully autonomously on behalf of the buyer (incorporating, if desired, specific request criteria). Thus, optimal offers are provided to the buyer by a "buyer agent". Additionally, the present service may with some modification be provided to buyers dialing into an 800 number. Typically, in this case, however, (as is also the case in the on-line implementation), if the buyer does not have access to his/her pseudonymous medical records, he/she would typically answer a questionnaire for each provider (which would typically consist of an aggregate questionnaire satisfying the requests of all relevant carriers which could then be submitted automatically, or an agent could be programmed to automatically complete the questionnaire as needed.

Commercial Alliances with Regional or National Health Care Organizations - Because of the tremendous restrictions in providing insurance coverage to organizations which provide health care services, it will be greatly to their advantage for patients and physicians to be able to benefit from more fair and competitive insurance coverage by using time of purchase competition and the associated coverage probability service (providing full disclosure of non-coverage and inequities in delivery of coverage). The primary joint marketing entities which will benefit most significantly from this service and thus will be the strongest commercial allies, include independent physicians, medical clinics, large physician groups, pharmacies and perhaps even pharmaceutical companies, providers of medical supplies, hospitals and *home* health providers. Of the above entities listed, physician groups, clinics, and hospitals do a tremendous amount of advertising directly to consumers (and this is becoming increasingly true via the internet for physicians in private practice),

particularly through the direct advertising medium of the internet (e.g., advertising on their web sites as well as email). These organizations, in light of their direct marketing exposure to the consumer which they provide, would likely become powerful allies in assisting the promotion and exposure of the competitive insurance *with* full disclosure of claim information service to their consumers.

Secondly, health care organizations which market to health care professionals, such as physicians and nurses, including pharmaceutical companies, pharmacies, and home health care providers may indirectly benefit by informing their target customers about the above consumer oriented service.

Time-of-Purchase Competition and Quality Ratings of Medical Suppliers

iamworthit's insurance application provides a natural entry into a second commercial application of the service. In a fashion similar to that of Time-of-Purchase Vendor Competition Scheme, for purchasers of insurance products, another useful commercial application of the present service is facilitating the time of purchase vendor competition between medical suppliers and equipment providers. The synergy by which this second commercial application is enhanced is as follows: physicians and hospitals will likely be cooperative allies to the service in providing patient information substantially as required as well as promoting (along with home health care providers) particularly the iamworthit insurer coverage probability service to their consumers. They will also likely utilize such a resource for internal uses and if/when HMO service liability legislation becomes enacted, for legal purposes as well.

To the buyer's advantage, it is possible (as in the insurance application, albeit less likely) that unfair or arbitrary price discrimination may be detected and counteracted by iamworthit's data sharing feature. Common buyers in this environment include hospitals, physician clinics, home health care providers and to a lesser extent HMOs and end-buyers, the nature/types of supplies and equipment which tend to be purchased by each of the above categories of buyers are provided in further detail below. As suggested earlier in this description (and in issued patent "Pseudonymous Server for System for Customized Electronic Identification for Desirable Objects"), a vertical portal as the one described, may be personalized based upon the user profile as disclosed upon accessing that portal in general, at the item level as within a category (or in conjunction with search results). Additionally, vendors may be equipped with price discrimination tools and are in fact highly motivated to form a coalition by entrusting their data and counter offer responses to incoming time of purchase vendor competition offers and bids to SDI (or at least an industry specific counterpart of SDI) for medical suppliers in order to provide limits or "price ceilings" for any given buyer and associated profile and condition relating to competitive offers or bids.

As in the case of almost any category of purchaser, particulary the larger entities, such as, unions, hospital systems and large physician organizations, additional bargaining leverage can be applied to suppliers because of the volume of supplies which can be purchased. The iamworthit model is particularly also likely to facilitate the buyer's leverage when negotiating long-term contracts with suppliers through the buyer-buyer data sharing and real time competition/bidding facilities which iamworthit provides.

Finally, it is worthy to note that as in the insurance application, a rating feature may also be provided which enables the buyer to rate (and subsequent buyers to observe the averaged ratings) the various vendors by various types of criteria such as quality of products, promptness of delivery, quality of service (or as in medical equipment, quality and promptness of technical assistance/repairs), availability of desired inventory etc.

There are a variety of types of health care organizations which could directly benefit from the time of purchase competition and quality rating service of medical suppliers which are
described below:

(a). Physicians Clinics and Large Physicians Groups - Typical types of medical supplies and equipment which these organizations will purchase using the iamworthit medical supplies application include: equipment such as X-ray equipment, ultrasound units, EKG units, EEG units, computer and other office hardware, chemical analyzer units, clinical supplies, office furnishings, pharmaceuticals,etc.

(b). Hospitals and Hospital Systems - Typical medical supplies and equipment which hospitals will accordingly purchase include: MRI units, CatScan Units, X-ray units (including units for special studies and mammograms), radiation therapy units, stress-testing equipment, hematological and chemical testing equipment, surgical supplies, prosthetic supplies, patient-care supplies for medical and surgical floors, IV solutions and pharmaceuticals, physical plant operational and replacement equipment and supplies, etc.

(c). HMOs - Currently, a certain percentage of patient health plans include coverage for pharmaceutical products. These products are typically provided through contract with centralized pharmaceutical suppliers. It is accordingly reasonable to suspect that many of these HMOs will invite the opportunity to provide these supplies through a more open dynamic marketplace environment as is provided by the iamworthit service.

(d). Union Directed Health Care Organizations - Union clinics typically purchase equipment and supplies, similar to private clinics or large physicians offices, as described above. Union private health care organizations would be compelled to incorporate the iamworthit medical supplies application as it similarly could apply to all of the above entities.

(e). Home Health Care Market - Equipment and supplies which are, in part, similar to large physicians clinics as well as some hospital - type supplies, used to provide out-patient care and maintenance services.

Patient Medical Records

SDI-1005-0095

The secure data interchange provides a secure and privacy protected storage and retrieval architecture in which buyers (in this case patients) are able to control the use of their personal profiles (in this case medical record information). Herein, we propose an alternative model which is based upon the assumption and belief that purveyers of this medical information, including hospitals, physician's offices/clinics, labor unions, (to some extent) HMOs are legally permitted to disclose and use their information for advertising purposes so long as a method is provided which insures that their information will not be associated with any identifying information about the patient (name, address, social security number, etc.). If regulatory constraints and/or the purveyor of this information prohibits the use of this information for the above purpose, patient permission could be acquired by patient signature on all appropriate consent forms at the time of admission and directly on the admission form. In one variation, software may be provided to the organization which enables the organization to operate their own pseudonymous proxy server, thus enabling access by advertisers to the pseudonymized patient records. Email and/or telephony-based pseudonymous communications could be readily provided. For direct mail, the address information maintained by the hospital would mandate the hospital to provide direct mail advertising to the patients or to entrust this operation to a pseudonymous physical mail service such as SDI. In the (much) preferred variation, a trusted pseudonymous proxy server operator maintains the patient information on an external database (this would perhaps be SDI or an SDI-like service) in which all of the above pseudonymous communication media could be provided along with pseudonymous direct mail services. For the SDI data entrusted variation, the preferred business model to incentivize these organizations to cooperatively provide this data would involve SDIs privacy policy enforcement capabilities in combination with a revenue sharing model in which revenues from advertisers such as pharmaceutical companies, pharmacies, clinics, HMOs, hospitals (where the purveyor of the data is not a provider of the same competitive services). In the case in which HMOs are the advertisers, the HMO would identify patients who are currently subscribed to competing HMOs and (typically) have a medical history indicative of a good risk to the HMO. The iamworthit insurance competition scheme, as above described, could be implemented within the context of this current solicitation scheme in order to elicit time of purchase competition among HMOs upon the submission of an offer by the original HMO to the pseudonymous patient.

14.2 Community Dollars

We allow **buyers** to receive compensation for providing personal data to vendors, information that has value to vendors because it allows information to be focused (for example relevant ads can be displayed to a **buyer**, based on his/her profile). The system of iamwothit credits **buyers** for information, and provides **buyers** with direct incentives to reveal profile information to vendors.

A vendor can sign up with iamworthit.com and agree to provide only the most restrictive type of community dollars, that can be spent at that vendors site.

Community dollars are the currency that vendors provide in return for the right to provide focused information to **buyers**. Dollars can be general (e.g. for a network of vendors), or very tightly focused (e.g. for a particular product, at a particular time). The **buyer**-centric infomediary acts as a broker, matching **buyers** and vendors. Another key role of the infomediary (e.g. the portal) is to protect the **buyer** from information saturation by controlling the flow of solicitations. (i.e. restrict the number of ads. that a **buyer** sees)

We can use meta-tags to restrict the way that community dollars can be spent. The tag associates the dollar, but the dollar is released within the system of blinded signatures (Chaum) so that a **buyer** that collects dollars over many transactions with different vendors can spend the dollars without compromising his/her private information about pseudonyms.

Dollars can be restricted to a number of vendors, and also restricted in additional ways—i.e. they can only be sent if the **buyer** visits the site through a particular portal, cannot be redeemed at a competitor, are worth a bonus if redeemed with certain vendors, etc.

14.3 The iamworthit community advertising dollars model

The primary objective of the IamWorthIt model is to create a market for information about **buyers**. We allow vendors to pay in "community dollars" for adverts, dollars that can only be spent at that vendor (with the host site of the advert receiving a share of the profits). This provides vendors with the ability to gain long-term customers. Furthermore, so long as the **buyer** agrees to receive advertising from his/her iamworthit subscription offer, community dollars can be replenished at the rate at which advertisers are willing to pay for impressions. This provides **buyers** with an incentive to spend at the vendor's site, because the vendor can monitor (pseudonymously) the **buyer**'s that are sensitive to discounts and other special offers (that are delivered as community dollars).

The main mode of the community dollars advertising model allows vendors to advertise for free, but provide community dollars to **buyers**, that can be spent at some later time. The cost of advertising can be linked to the success of advertising. Moreover, the vendor can direct offers and adverts to particular **user profiles**. The hosting web page receives a share of the vendor's revenue that comes from transactions involving community dollars. The dollars can represent "stored value", such as bonus points, that can be applied to special discounts for offers which are delivered via digital coupons and/or as "straight value" which could be converted directly to purchases thus are equivalent to real dollars at the point of transaction.

The community dollars can be "credits" that can be redeemed as real cash, credits towards discounts, and can be spent across a suite of sites, or limited to one site. The co-pending patent application entitled "System for the Automatic Determination of Customized Prices and Promotions" describes a comprehensive scheme which may be implemented in either on-line or off-line commerce environments. The system enables vendors to deliver a digital message in the form of a promise to a **buyer** (typically on encrypted form for purposes of targeting a **buyer** specifically). This promise is typically a discount for a product, set of products (or all products in stock) or may even include entitlement to special privileges for that **buyer**, thus it is termed a "digital coupon". The community dollars can represent special discounts for a **buyer**.

The **buyer** receives a financial incentive for receiving well-targeted solicitations, while preserving **buyer** privacy within the SDI system. The vendors support the community dollars through advertising revenues and increased sales volume. We can also provide the vendor through which the **buyer** first subscribes a special "first screen" right that allows the vendor to provide a **buyer** with his/her first impression as soon as s/he logs on.

Of course many variations of the community dollars scheme are conceivable and the current description is in no way intended to limit the scope of the claimed invention. For example, the ad revenues generated may instead be apportioned between direct payment which the buyer receives, in real cash, community dollars which the vendor (or ISP) credits the buyer and/or direct payment which the vendor (or ISP) receives as well as iamworthit or any

combination of the above such as exclusively direct payment which is apportioned between the buyer and the vendor (in the absence of the community dollars scheme).

In one variation all community dollars collected by a **buyer** must be spent back at the vendor site at which they originally subscribed (and also the site that hosts the adverts of other vendors). A **buyer** can spend the dollars with any vendors that are site partners of the original site. This provides the vendor an incentive to accept and promote the community dollars concept.

NOTE: DAVID Since the community dollars concept is unquestionably the most critically important piece of iReactor technology I want to be ABSOLUTELY POSITIVE it is comprehensibly disclosed technically.

The value of providing a **buyer** with targeted solicitations is estimated at approximately $300 to $500 per year (based upon $120 per 1000 targeted impressions at approximately 25 impressions per day). Given these significant benefits, a vendor can provide a **buyer** with a significant discount (in the form of community dollars). Vendors benefit from increased sales volumes.

When the price of items is less than the value of dollars, the vendor can limit the amount of discount that is available on any single product, or only allow community dollars to be applied towards customer discounts (which may nevertheless be quite substantial).

As suggested in the aforementioned description regarding the buyer-centric SDI, advertisers may target buyers vis-a-vie negotiations in which they may successfully engage in conjunction with accessing and querying the pseudonymous user profile database. They may also stipulate rules relating to the delivery of ads to buyers in conjunction with performance of certain explicit actions e.g. submission of certain search terms, visitation of pages with certain key words (as a general tendency or on a per-page basis), visitation of certain specific pages or sites, performance of certain location specific behavior (e.g. visitation to a competitor's retail outlet or traveling within the proximity of the vendor's own store), otherwise the presently suggested algorithms may automate most of these manual functions by virtue of its recommendation capability e.g. matching ads to the most appropriate list of buyers or matching ads to metrically similar Web sites which may appear whenever the buyer visits that site(s). Ads may (for example if delivered for off-line purchasables) also take the form of digital coupons which are priced by vendors according to behavioral information which the buyer is willing to disclose to this vendor. Because many buyers are adverse to viewing additional advertising which is "imposed" in the preferred embodiment, any inserted ads delivered by iamworthit can be "turned off" by the buyer voluntarily unless the in kind value which the buyer receives in exchange for the advertising mandates a certain minimum ad impression delivery. Also in light of the above buyer concerns in the preferred embodiment, there are no additional impressions added to the buyer's web browsing experience. Instead, ad blocking technology is deployed to block existing banners replacing them instead with iamworthit ads. **Ad blocking is commonly available. The techniques used are well known in the art. The HTML source for the ad banners may be either called-up from the remote server (such as with an ad network) or alternatively from an ad server (typically purchased by the vendor) on the vendor's own host server. In either case, the ad blocking software typically recognise the HTML source which originates from the ad server. A directory of the HTML source for the various ad servers is maintained with the ad blocking software (and presumably updated e.g. if new servers are added or change their IP addresses). The location/physical dimensions of the portion of the page occupied by the HTML source which the software removes (its "footprint"), may be readily replaced by another replacement banner (by iamworthit) from wherever the proxy server resides (typically on the client but potentially on the network). As a result of a potential difficulty in which the ad server deliberately changes its IP address to avoid recognition, which occurs constantly and dynamically, it may be possible to recognize portions of the page which contain image "features" of ad banners, which are inserted from another HTML source where that HTML source is linked to known vendor sites (which are also linked similarly from other site's ad servers and/or are known via their HTML sources to have previously utilized ad server technology and a "new" HTML source appears in it's stead) etc.).** ***As suggested outside the scope of iamworthit this model may be extended to other media domains e.g. replacing digital TV commercials instead with targeted ads delivered and precached via iamworthit for insertion at appropriate times i.e. during commercial breaks during standard video programming. Another variation could be tailored to pre-loading iamworthit advertising to automobiles for insertion in place of traditional radio commercials.*** **In accordance with the parent patent application, (and as suggested above), email may be a useful targeted ad delivery medium as well. Per the patent case, the system is able to classify email (according to its source and content using implicit or explicit actions of the buyer). Based upon certain desirable confidence threshold settings, the system may automatically delete "span" and replace it with targeted messages provided by iamworthit which the buyer is**

paid to receive pseudonymous physical mail is yet another potential source of targeted mail for which the buyer may be paid in conjunction with the user's voluntary removal from direct mail marketing services which iamworthit could provide in conjunction. iamworthit's buyer compensated targeted advertising model could be extended to a pseudonymous telemarketing service (as suggested above) using one-time call session pseudonymous telephone numbers for telemarketers to reach desired buyers. Finally, human sales persons could be the "ad delivery medium" for iamworthit.

14.4 Preferred Implementation

In the preferred implementation we use an "electronic cash" infrastructure for the community dollar system. A **buyer**'s SDI-enabled client-level proxy stores dollars that the **buyer** receives securely. Dollars are anonymous and non-traceable, so that the **buyer** can maintain a single "bank" of dollars, and aggregate dollars collected across pseudonyms for a single purchase, so long as the purchase satisfies the constraints on the dollars. Each dollar is created using Chaum's blinded signature technique, and also signed with the conditions on its use.

This scheme allows vendors to monitor the offers that **buyer**s respond to, because when a **buyer** presents a community dollar—the dollar can be validated to indicate the type of discount that it is, even if the identity of the dollar (i.e the serial number) is untraceable. SDI provides vendors with guarantees that **buyer**s have once-in-a-lifetime pseudonyms, so redeeming a voucher of a particular type that is redeemable only at vendor V and was issued by vendor V allows vendor V to be sure that the voucher was issued under the same pseudonym, and has not been transferred to another of the **buyer**'s pseudonyms.

In an alternative architecture, the ISP-level SDI proxy, or the web-host for the advertising service, can maintain community dollar "debit" accounts for each **buyer**. This is more limited, because it does not allow **buyer**s to transfer dollars between pseudonyms without compromising privacy (revealing a portfolio of pseudonyms). However, in a scheme where advertisers require that agents have once-in-a-lifetime pseudonyms, and only release community dollars to be redeemed at their own site, this is not limiting. Both of these approaches are useful for "community dollar-enabling" numerous or all sites.

A vendor that allows community dollars to be spent does not need to implement a special community dollars/discounts program. The **buyer** can also be issued a special debit account dedicated to community dollars, that permits pseudonymous transactions without revealing a **buyer**'s portfolio of pseudonyms.

A portal site that hosts advertisers and **buyer**s that subscribe to iamWorthIt can mandate that all community dollars are to be spent at sites that advertise on the portal site, and also only when the sites are accessed via the portal site. This technique will increase portal traffic. Portals can be expected to compete in terms of: (a) the fraction of advertising revenue that is turned over to **buyer**s, in return for receiving profile information from **buyer**s; (b) the level of advertising that **buyer**s are exposed to; © the nature of the community dollars "package", i.e. what vendors can the dollars be used at etc. This can be useful to attract niche customers, that have common outlooks, interests, and business needs. The primary goal of the portal is to drive traffic through the portal.

14.5 Commercial Variations

(i) Vendor coalitions

Vendors may choose to form coalitions, to allow **buyer**s to spend community dollars at any "partner" site. Vendors that have similar **buyer** bases can be automatically identified using collaborative filtering. (i.e. determining similarity with the present vendor, from the aggregate vendor preferences of a given vendor's subscribers). Also, these resulting metrics could incorporate predicted online spending by each **buyer** at each site. This could help to narrow the selection of sites the vendor wishes to partner with and/or the selection of these partner sites could be

determined and presented to the **buyer** to even further narrow the selection for each **buyer**. All vendors in a coalition advertise, and provide cross-links and up-links to other vendors.

The coalition model is good for **buyers** that are more likely to find products that they want. Vendors can share the risk of advertising, since dollars provided to one **buyer** by a particular vendor can be redeemed at another vendor. Advertising and community dollars increases sales volume at all vendors in the coalition. **Furthermore, studies on on-line buying behavior suggest that on-line shoppers tend to make purchases across a variety of categories of ecommerce products, thus it is likely that they would also prefer the freedom and flexibility which is associated and can be provided in this way only by deploying a multi-site community dollars scheme.** ***Iamworthit usage statistics are very effective in identifying prospective vendor coalitions which consist of complementary (non-competitive) vendors which tend to attract the same subset of the iamwort***

hits subscriber based and an offer roughly equivalent numbers of customers to the coalition. Alternatively, **personalized coalitions which can be *created* in ad-hoc fashion for each buyer can further serve the buyer's interests by increasing the freedom and flexibility which the buyer often strongly desires. Though practically this can only occur for those (perhaps smaller less commercially significant) vendors which have not been a major coalition consisting of vendors which many buyers tend to use which collectively serve a substantial percentage of the overall user base.**

Thus if the vendors have not established firm partnerships with other vendors, we can even allow vendors to form dynamic and virtual coalitions within SDI, with a potentially unique coalition of vendors for each **buyer**. The coalition may consist of an optimal pool of vendors, as determined by SDI collaborative filtering techniques. The goal in this model is to provide **buyers** with a particular "brand" of community dollars.

We can allow each vendor to retain an exclusive right to advertise to each **buyer**; and also develop a portal for the coalition—that gives advertising prominence to coalition members. Portals will be expected to aggressively promote community dollars. **Buyers** that collect community dollars become loyal return visitors to the portal and its associated vendors. In the case the vendors do not generate the same value we can provide community dollars in proportion to the value that a vendor contributes to a coalition.

We can also provide targeted advertisements for the vendors at the portal, using the **user profile** to focus ads. The categories and links at a portal (that might include a search engine) can be re-prioritized (highlighted and/or re-ranked) in accordance with the **buyer**'s preferences (as described above), and to favor subscribing vendors. **In commercial practice, the motivation for vendors to become coalition members is largely driven by customer partnerships with providers of ecommerce tools and solutions. There are many emerging trends by which these intermediary tool providers could conceivably integrate community dollars. Consider loyalty points (e.g.** www.mypoints.com**) or loyalty bonuses programs (e.g.** www.clickrewards.com**). These points or bonuses could be substantially subsidized by the advertising and deeper benefits passed onto the consumer. Other simpler technology which would compel these vendors to cooperate include shopping basket technology, the emerging standard, ECML, common Web tool and/or tool development solutions, common hosting solutions, common ad delivery systems.**

Vendors pay the portal site to advertise, and the portal provides community dollars to **buyers** in return for privacy-protected profile information. This model does not provide incentives for the portal to provide well-targeted adverts, because there is no direct link between a portal's revenue stream and the vendors' sales volumes.

A portal with community dollars that can only be spent under a single pseudonym at its partner sites also provides an incentive to **buyers** to interact under a single pseudonym—which in turn allows a portal to profile **buyers** across its complete vendor partner network. **Buyers** will access many sites with the same pseudonym. The system of SDI allows vendors to leverage the shared profile information as **buyers** browse web pages and products.

We can also lock **buyers** into a single portal—and a single coalition of vendors—with community dollars that "decay" over time, and must be continually replenished. In this way a **buyer** cannot pick and choose different portals, and different community dollars, but can benefit mainly from high web-browsing volume through a single portal. The value to vendors in terms of consumer lock-in can be considerable.

For example, a coalition of vendors can join to allow a **buyer** unlimited access over all affiliate vendors. The program can be sold through the existing marketing channels of each vendor, as well as through a portal directory of sites for those vendors. Vendors that join can be required to promote the program through their own marketing channels. Additionally, perhaps vendors are selected to cover exclusive physical regions (e.g. in the case of a set of

ski resorts), or exclusive product categories (e.g. in the case of on-line vendors). Vendors can provide a community dollar-for-real dollar exchange, in return for becoming part of a vendor network. Alternatively, perhaps vendors provide an up-front fee, that can be recovered via dollars spent by **buyers** at their own site. Each vendor is obligated to sell the partner network community dollars, but is not necessarily required to promote the other community dollar vendors.

(ii) Enabling "Transaction-based" Revenue-sharing

In a second mode, the vendors provide discounts to **buyers**, and advertising is primarily "free". The only time that vendors incur a cost for providing a **buyer** with an impression is when they achieve a sale—because then the portal site receives a cut of the transaction price. The vendors provide **buyers** with community dollars directly. The dollars, which are stored at the portal site, allow **buyer**-spending to be tracked. This allows the portal to monitor when a sale occurs, not just a hit on a banner ad. With transaction-based revenue of this kind, personalization is critical. In this model the portal with give prominence to adverts from successful sites.

A portal site may forgo payment from a vendor in exchange for the increased click-through from a strong network of community-dollar enabled vendors. Value is credited directly to **buyers** for future redemption at that particular vendor's site. The community dollars provided to a **buyer** can be restricted, such that a **buyer** can only redeem dollars if s/he maintains enough visits to the portal site.

Vendors can offer discounts on their own products directly, instead of providing the portal with money for advertising. The vendor only pays to the extent that its advertisements are well-targeted. The vendor could also request special ad priority. A vendor that presents advertisements to a **buyer** offers the **buyer** discounted promotional offers for products offered by partner vendors, in exchange for subscribing to iamworthit and receiving targeted impressions. These offers are in lieu of community dollars, and can be provided by partner vendors—maybe in exchange for a right to a number of ad deliveries for the vendor's own advertising purposes.

We can also require that **buyers** are automatically routed through a portal when accessing any partner vendor directly. The portal (and therefore the coalition of vendors) then receives exposure each time the **buyer** clicks on an ad (or link) to that vendor.

The portal may also provide benefits (e.g. additional advertising prominence) for sites which are responsible for driving traffic through the portal. Community dollars can be provided whenever the **buyer** accesses a site from the portal. Portals can offer free advertising to e-commerce sites (forgoing advertising fees). The portal provides discounts to **buyers** that purchase a product following a link provided at the portal.

A **buyer** receives the discount by validating a purchase with the portal, and the site agrees to provide the portal with a share of revenue whenever the **buyer** cashes in community dollars in this way (we do not rely on HTTP refer mechanism because that can be blocked and falsified. Furthermore, we do not rely on URL+extension correspondences, also not secure—instead rely on providing **buyer**'s with incentives, and monitoring **buyers** that have earned community dollars.)

(iii) Delivering per-impression dollars

We can deliver community dollars on a per-impression basis, with vendors competing to offer **buyers** high values for being able to present an advert. The existing collaborative-filtering engine at a **buyer**'s SDI client-level proxy can filter ads, and select appropriate offers, using community dollars as just another measure of the usefulness of a message. This is an alternative to providing dollars on a one-off (or even yearly) basis, for consumption via the vendor's site that the **buyer** subscribes to the service.

A hosting site can take a fraction of any dollars provided to a **buyer**. Alternatively, a site can convert the value into the community dollars to provide to the **buyer**, possibly at a preferable basis. The portal might also wish to convert its commission to credits for the **buyer** at any one of its partner vendors, with the stipulation that the **buyer** must access those sites via the portal in order to be able to redeem the credits.

14.5

There are numerous (perhaps nearly endless) example commercial applications in which the time of purchase competition features can be utilized to provide significant benefit to the user. It is suggested the

user can through sharing further achieve additional benefits and information regarding the iamworthit price offer from as many users as possible. Most of the primary commercial domains of electronic merchandise and services provided over Web (b to c e commerce as well as b to b e commerce) are potentially important commercial uses of the service. Based upon the current disclosure, some of the less obvious (nonetheless very important) commercial domains include the following:

1). Travel Packages and Flight and Hotel Rates – Features, amenities, class, activities, entertainment, excursions, etc.

2). Financial Services (e. g. , Loans and Mortgages) - Terms, amortization, interest rates, service fees, etc.

3). Automobiles – Luxury and safety features, warrantees, auto club membership, terms of purchase, etc.

4). B to b Commerce – Numerous commercial applications including standard Web-based environments for b to b e commerce including buyer-side and seller-side e commerce applications , vertical b to b portals and trading floors.

5). Job Positions – Employers posting job positions may be notified by the service of competing job offers which employment candidates are browsing (and ultimately accept if this information is indicated by the candidate). Resumes and, if desired, other profile information is provided to all competitors. Depending upon the profile of the candidate and/or competitive employment opportunities which a particular candidate is observing (or has recently observed), employers may wish to customize offers dynamically on an individual basis.

6). Off-line Vendors – The Location Enhanced Information Architecture (LEIA) enables vendors in the vicinity of where a user is travelling to deliver a targeted ad or particular personalized offer. Or a user may search or navigate directories on the Web or browse a particular (competitive) item. Regardless of the mode of access, the system filters or identifies purchasables by their direct physical proximity to the user (or the user's intended travel path). If the time of purchase competition system uses location proximity of an item as a parameter for vendors to compete with one another (i. e. providing an additional speed/convenience factor for on-line items with an off-line sales presence) the particular vendor's price which is finally agreed upon and accepted by the user can be securely enforced at the point of sale using a digital coupon which is non-transferable and non-reusable may be down-loaded to the user's device, printed or simply securely transmitted to him/her as a unique code redeemable with user identification at the physical point of sale. These techniques are fully disclosed in co-pending patent application entitled "System for the automatic Determination of Customized Prices and Promotions".

6015640-6015769

14.6 Combining Community Dollars with Time of Purchase Vendor Competition

It is worthy to note that the time of purchase vendor competition scheme *is an exemplary incarnation of the* customer centric infomediary model and thus is almost by definition unavoidably in direct opposition to the interests of vendors (for this reason, the time of purchase system is likely to be marketed as an independent service or through ISP partnerships *with user-centric vendors which include among others, browser companies, ISPs and* possibly non-commerce sites (and not through the channel of most commercial web sites). There is a variation of community dollars in which some of the competitive features of the time of purchase competition system can be utilized by vendors to provide a competitive advantage over other competitors while providing the user with a certain degree of vendor competition (albeit within a much more limited scope). In this model, community dollars are implemented according to the preferred disclosure as above described. Each vendor which provides its own particular community dollars implementation) is given the unique ability to exclusively identify, in real time, other competitive product/service offers which their users (who subscribed through their site or portal) are *currently in* the process of observing. In the preferred variation, all other competitive vendors are denied the privilege of engaging in any time of purchase competition activities for that user. *Again, the vendor may request profile data from the user which s/he may wish to optimize the selection/features of the particular product offering as well as the value of the offer. The vendor may then wish to either utilize his/her own pricing/recommendation engine to determine optimal rules to generate counteroffers to those accessed by*

***the user or (optimally) employ the vendor-centric SDI to perform these functions instead.* However in a variation of the scheme, if the user requests a more robust time of purchase competition service, e.g., he/she already subscribed to the time of purchase competition service before receiving the subscription offered to community dollars, he/she may be willing to give the vendor conditional exclusivity which is a first right to provide a counter**

offer. If on the other hand, the user does not accept the counter offer, the system will default to the standard multi-vendor *(open-ended) vendor* competitive scheme. As a component of the preferred business model, which is associated with the present community dollars scheme, the iamworthit service *may wish to establish* a strategic relationship (or possibly *independently* owns) with an advertising network. As described above, banners delivered from all advertising networks or ad servers are blocked at the HTML source and replaced with iamworthit ads. However, with the exception of all banners delivered by the ad network/ad server partner company or alternatively ad banners delivered from sites utilizing the community dollars scheme. Assuming a significant percentage of Web users utilize the community dollars scheme (e.g., greater than 5%), enormous pressure upon the vendor will be applied to either deliver their advertising using the ad network or ad server of the partner company or to adopt the community dollar scheme (thus automatically exempting that site from the community dollars' ad replacement technology). *It is worthy to note that the present modified time of purchase competition scheme can also be adapted for the application to television commercials. In a preferred commercial mode a relationship is created between iamworthit and one of the providers of commercial insertion technology (at the head end). All cable providers not utilizing the particular commercial insertion technology provided by that particular partner provider would be subject to the use of the commercial replacement technology which iamworthit provides by delivering typically (but not exclusively) digital set top terminals streaming (HTML based) video commercials or the same preloaded onto the set top for subsequent display. As indicated above, the system is able to recognize if/when commercial breaks occur during the course of standard video programming through tags which are typically digitally encoded or (in the case of analog TV) transmitted within the vertical blanking interval.*

Petitioner - Ex. 1002 - Page 102 of 196

14.7 Providing Loyalty Bonuses

We can use the client-side SDI proxy to provide vendors with "loyalty guarantees", that are credentials to verify that the **buyer** has executed no transactions with any competitor, under any of its pseudonyms. The client-side SDI proxy is in a unique position to be able to implement this monitoring, because no other system knows a **buyer**'s portfolio of pseudonyms. The **buyer** can present its digital credential when visiting a vendor's site.

A vendor may wish to provide loyalty dollar credit; for example, it would be possible for vendors to offer **buyer**'s credits if the **buyer** is a 100% loyal customer i.e. that she/he did not (over a specified period) do his/her purchases at the site of any competitor. For example, certain types of high value customers could be given considerable value in the form of credits or discounts as a result of demonstrated vendor loyalty.

A competitor could be determined by a simple known vendor list, e.g. through Yahoo or by matching similar products or service descriptions on sites across the web (e.g. using web crawler technology).

The credential can be time-stamped, to prove loyalty. It does not reveal any information about the **buyer**'s other pseudonyms to a vendor, because many pseudonyms will exist that have not made any purchases from a competitor. Upon accessing the vendor's site, this credential may be presented to the vendor. One criteria for the above benefits could be that the **buyer** may visit a competitor site, and engage in interactions; however s/he should not transact with that vendor.

14.8 Off-line Variations

The community dollars may be stored on a portable smart card carried by the **buyer**, in which the community currency was originally loaded by the **buyer**'s own computing device, or alternatively a kiosk or cash register controlled by the vendor or another third party whereby redemption may occur at any location where the subscribed vendor possesses a smart card reader. **For vendors who offer the integration of community dollars into their discount programs (e.g. may be members of a vendor coalition) the buyer may at any vendor terminal, kiosk or his/her PC upload targeted-vendor promotions such as digital coupons which are tailored to the users profile. This profile is downloaded to the vendor and which consists in part of the recent purchases s/he had recently made at the various vendor stores.** In another variation, the community dollars may be coded into a form which is bar code reader enabled and distributed to the **buyer** electronically or potentially if used in conjunction with a traditional loyalty points program, additionally printed for the **buyer** at the vendor's physical location (such as point of sale or kiosk), applied in conjunction with purchases at the vendors physical location. At which point typically a new coupon is reprinted containing the updated secure information pertaining to the **buyer**'s community dollar and/or loyalty points account.

In another variation, a promotion for a yearly allowance of community dollars could be printed as an advertising offer on a magazine coupon, newspaper insertion or direct mail piece which could contain a unique URL (typically the actual URL for the iamworthit community dollars subscription site with a unique post script as the character string ("code") identifying that particular vendor and/or that promotion) from which the **buyer** could subscribe to iamworthit, wherein the unique URL acts as an identifier for that particular vendor's promotional piece from which the **buyer** originally received the offer for his/her own community dollars promotion.

Within SDI we could send physical solicitations to **buyer**s, and allow **buyer**s to access promotions pseudonymously. SDI can target a selected audience for each vendor.

Example: An iamworthit card in accordance with the pseudonymous payment methods described above, such a card could be a direct extension of SDI into the off line environment. **Buyer**s could use this card as an identifier such that when they travel physically from vendor to vendor, their profile data can be readily identified where data pertaining to their own behavior and policy (depending on their data release potentially part of the vendor's user profile data) is retrieved.

If a smart card is used this user profile data may not have to be remotely retrieved but may be stored on local memory on the card itself ***along with the client-based pseudonym proxy server***. In one novel variation, a card is done away with completely by virtue of revolutionary technological breakthroughs in being able to instantly and positively identify **buyer**s biometrically using ***iris*** scanning techniques (which may in a variation be further combined with facial recognition techniques). Many vendors will wish to utilize **user profile** data in order to deliver promotions targeted discounts and promotions (see pending patent "System for Customized Prices and Promotions").

The co-pending application entitled "Location Enhanced Information Architecture" (LEIA) describes an integrated advertising delivery platform which selectively targets **buyer** personalized advertising based upon both the **buyer**'s personal profile and the present location of the **buyer** which may suggest appropriate ads from vendors which are local to the **buyer**, wherein user identifiers (UID's) which could include any of the above identification media provide the essential elements for this **buyer** targeting platform. **In accordance with the iamworthit targeted advertising platform either either advertisers select user profile attributes which they are interested in or the system matches their ads automatically to the most relevant users. The former application may be performed either within the context of the pseudonymous user database which the vendor queries or targeting rules which s/he provides which control the advertising and pages which are dynamically generated for each user based upon his/her user profile. Whether the user's location is detected on a mobile network or a terrestrial network including telephone or cable TV, LEIA's location features are able to provide additional useful features regarding users to the vendor's rules interface or to the pseudonymous user database controlling dynamic page generation at his/her site.**

For example, at a bookstore, we can recommend isles and particular books; at a supermarket, can play music preferences; smart-radio, play appropriate channels in a cab based on target object profiles (as meta-data). As suggested in issued patent "System for Broadcast of and access to Video and Other Data Using Customer Profiles" the appearance of relevant selections can be continuously scanned for, dynamically selected and presented to the **buyer** in the form of "Virtual radio station". Such a system can also be linked to a service for making an instant purchase, or linked to a database (in conjunction with LEIA) to recommend where a **buyer** should physically go to make a purchase. For example, music selections that the **buyer** is presently listening to may be ordered.

Also, it is possible to provide advertising which is targeted to a **buyer** by automatically recognizing pre-existing commercials and replacing them with targeted counterparts. This can be done through the identification of previously played commercials for example, commercials that have been manually identified and classified. Upon recognition, targeted commercials (including those which are targeted by **buyer** location in accordance with LEIA) may be inserted into these spots, and delivered and/or pre-cached through cellular, satellite or radio communications.

At a public phone we can identify a **buyer** with his/her calling card, and deliver targeted advertisements, via the public telephone readout or delivering the targeted ads as audio messages in which server software at the phone switch (an ISP level proxy) recognizes if/when the **buyer** is put on hold and delivers audio and/or audio/video advertising to the **buyer** accordingly.

Targeted discounts and advertisements can be delivered at kiosks, for example using a credit card/smart card/other ID method (e.g. biometric...) Similarly, we can use credit cards to deliver targeted print advertisements on the backs of purchase receipts, e.g., supermarkets or fueling pumps or, alternatively, on a sheet dedicated for a advertising purposes conjunction with public copiers or printers or in another variation, on the cover sheet of incoming faxes which are sent to the **buyer**'s fax machine or in which the **buyer** is otherwise identified automatically from the recipient's name field on the cover sheet.

Similarly, we can deliver targeted advertising and other information through cable TV systems, as described in the issued parent patent application entitled "Broadcast of and access to Video and other data using customer profiles", and co-pending application entitled "Broadcast & System for reduced memory terminals broadly address the use of cable systems as an interactive medium (in a bi-directional network architecture) for purposes of delivering targeted advertising targeted advertising and other information to the consumer based on **user profiles**". In this system customer behavioral data is collected at the digital set top and the upstream channel enables these profiles to be processed at the lead end server. These detailed profiles may then be subsequently transmitted down and stored at the level of the individual set top. The cable environment is a two way interactive medium. The bandwidth allocation is inherently asymmetric. Separate channels can push parallel adverts, which are selected at the set-top-box according to a **buyer**'s profile. Each channel can have associated meta-data to allow matching at the set-top-box. As an alternative variation, full motion advertisements may be down loaded in the form of applets to the digital set top box and displayed to the **buyer** in similar fashion as described above. ***Or full-size commercials or infomercials could even be downloaded and inserted in place of existing commercials on the video stream. Digital tags used for queing for ad insertion technology enable this capability. The preferred commercial implementation of the system is used within the context iamworthit (the buyer infomediary service in which the viewer receives value for advertising received). This method is preferred (in as much as in the following alternative, the cable operator does not need to be in the loop). With that said, in an alternative embodiment, however, requiringthe cooperation of the cable operator various commercials could be broadcast in parallel during commercial breaks. The user profile at the set-top is deployed to select the most appropriate commercial accordingly. It should further be noted that user profiles regarding viewing behavior can be collected with relatively minor up-stream transmission, e.g., to infer whether a buyer is viewing a program the set-top may transmit the viewers current viewing selection, e.g., two minutes after the beginning of each half hour and two minutes before the end of the half hour (and possibly at additional intervals during the viewing segment as well). In the point-to-point access and delivery of personal information as well as the passive monitoring of viewing behavior (for profile generation at the head-end and subsequent user profile delivery for the set-top) because of strict buyer privacy regulations in the cable/satellite industries the use of the pseudonym proxy server architecture (as described in the parent case and integrated into the above referenced case) is extremely important.***

Once interactive television is enabled by access to very large amounts of bandwidth, it will be very feasible from a bandwidth utilization standpoint to transmit more interactive content. Predictive caching is still extremely advantageous in as much as advertising (unlike interactive video real-time on demand video content) tends to be controlled by targeting rules of the advertiser rather than the user who receives value in exchange. Of course, radio or music content may also be delivered in conjunction with preloaded audio commercials. And any of the above content is deliverable over the same networks to other appliances such as the P.C.

14.9 Example: An on-line Gaming Site

Consider an on-line gaming site that has a network of affiliated vendors, that do not pay to advertise, but provide community dollars that can be spent either at the vendors—or at the casino. When **buyer**s lose money at the casino the casino receives real dollars from vendors. The casino is one possible outlet for spending dollars—and a vendor only pays the casino (the host of its ads) if a **buyer** chooses to gamble on the site, and loses its dollars.

The gaming site becomes a portal, with links to partner vendors. Each vendor offers the **buyer** community dollars, that can only be spent back at that site or at the casino. However, the number of dollars which can be spent at the store is substantially less than the number of community dollars which can be spent at the casino. IF the **buyer** accepts the credits, whenever the **buyer** accesses the URL to the store he/she is either automatically routed first to the casino portal or to the vendor site whereby a prominent banner is displayed which is displayed to that particular **buyer** from which the **buyer** can conveniently engage in a casino gaming session.

If the **buyer** loses a substantial amount of community dollars he/she may regain the lost credits by spending a specified amount (in real dollars) at a partner vendor. This provides a safety-net for **buyers**. The cost to the vendor is the cost of the dollars that the **buyer** lost at the casino, and the cost of replenishing the **buyer**'s community dollars (which can be used for further gambling). However, the vendor makes a sale—so the vendor is happy so long as the dollar value is a reasonable discount for the sale.

Clearly, the casino gains substantially through the redemption of these community dollars. The vendors can make an agreement with the casino where they only compensate a fraction of community dollars. A percentage of a **buyer**'s profits at a casino are paid in community dollars, another percentage can be paid in real dollars. Perhaps the casino can also provide vendors with a revenue share.

14.10 Implementation Details

We allow for community dollars that are restricted to particular products, and customized for an individual **buyer**. The dollar-object can contain two parts: the first part is readable to the **buyer**, and indicates the nature and the amount of the discounts to which the credits can be applied. The second part of the message is encrypted, and accessible only to the vendor, and is signed by the vendor to prevent any form of alteration. The information can contain the dollar credit to the **buyer**, the terms and conditions of the community dollars, a dollar amount, the pseudonym ID of the **buyer**, an expiration date, the terms and conditions of discounts and special offers to which community dollars may be applied in combination with a partial cash transaction.

The vendor must check that it has not previously redeemed any piece of community currency with the same identifier; the identity of the **buyer** is correct; the date; the terms and conditions. Some vendors may allow redemption of the community dollars at other vendors' sites.

14.11 Example Implementation within Web Hosting Environment

One of the commercial applications of which community dollars may be used is in offering web sites free web development services and/or free web hosting services. In the former case a Web development tool provider could share in the community dollars revenue stream from iamworthit ad revenues, and accordingly cover the costs of the tool suite. Whenever new upgrades including new development modules are added, the costs could be covered by an incremental increase in the percentage of advertiser transactions. There is considerable economic pressure on web tool providers to provide competitive solutions which are not only rapidly deployable but also extremely robust (typically bundling some form of personalization technology), thus as a result of economic pressure to drive advertising and e-commerce transactions through the site (often through the use of a revenue sharing model which can reduce upfront costs to the vendor substantially). iamworthit through its customer loyalty enhancement capability provided by community dollars can significantly improve the effectiveness of personalization at the site (perhaps more significantly than cookies which often are blocked and all typically not implemented so as to be recognizable across the collection of sites which the tool provider supplies or the web host services). It is believed that the revenue sharing model may be structured to cover both the cost of the tools, web development services as well as web hosting (thus unless he himself offers a complete turn key suite of solutions and services a tool provider which integrated the community dollars concept could be in an extremely competitively advantageous position in attracting strategic relationships with Web developers and Web hosts. The tool provider may for example sell an ad server module or could "automatically subscribe" the user with an opt-out option. The very fact that all of the tool , Web development or hosting provider's sites are exempt from ad blocking/ad replacement technology (barring iamworthit competitors) would be a significant motivating factor in incentivizing sites to utilize their services (e.g. through promotional programs). An emerging large market for Web development is providing these services to individual end-users. Completely free services in this regard would likely mobilize what activity in what is now a dormant, albeit enormous commercial market. For example, tools for developing elaborate Web-wide community portals could have customized community links which are matched to the user's personal profile. Visitors to the site (subscribed to iamworthit) could experience an additional layer of personalization based upon their own user profiles.

Web hosts may also use an additional feature (optionally to significantly drive increased traffic through potentially all sites on their network. This optional feature is a set of links (e.g. along the side of the user's screen). These links are used to point the user to other relevant pages issuing aggregate site usage statistics of their visitors (as described in the parent patent application) iamworthit, user referral links may additionally personalized based upon their user profiles.

14.12. Extension Application

iamworthit in a future version of the system could incorporate a novel optically based medium for delivery of the advertising. The protocol would enable the delivery of user targeted advertising in a non-electronic environment directly to the user from a static source in visible contact of the user. The system incorporates the following aspects: The user wears glasses or contact lenses which contain a polaroid film in which the optical medium is polarized for only certain very narrow and specific wavelengths. Advertisements (or other messages) may be presented from signboards which display from the same surface ads appropriate for every different user profile segment for which a unique message is appropriate. The polaroid film for each user is unique to the user segment to which they belong. Each user segment's optical film filters out all of the particular wavelengths of all other messages except that which is appropriate to the user segment to which that user belongs. Alternatively, the current application may also be relevant to virtual tags as users may provide instead optical messages specifically targeted by a user to other users characterized by their profile features which establishes the entire criteria for the user segment they fall into for which an appropriate message may be targeted.

6654650 - 4622948

14.13 Privacy Enhanced Marketing Applications

There are a number of commercial applications for a system that allows vendors to contact privacy protected lists of **buyers** in the pseudonymous **buyer** data base, wherein the contact, interaction, and business relationship with the vendor occurs under terms of complete **buyer** pseudonymity. Or wherein the **buyer** remains pseudonymous in his/her search and access to qualified professional services. In accordance with the parent patent application the pseudonymous communication may be either email, real-time text communications, voice (such as the pseudonymous telephony or Internet telephony). ***In the case of pseudonymous telephony, instead of a one-time or persistent pseudonymous buyer address, pseudonymous buyer telephone numbers may be used for the third party to reach the buyer under his/her terms.***

For example:

(i) Financial Advice and Financial Planning Services

Often **buyer**s are quite sensitive about the confidentiality of the release of this type of information related to personal financial matters and particularly with certain matters (and perhaps in general) prefer that their financial advisors were unaware of their true identities. Similarly, investment advice or sales communications by stock brokers are another application where similar **buyer** information is typically disclosed.

(ii) Insurance Agents & Brokers

For many types of insurance, (e.g. health, life, casualty) personally sensitive information is disclosed by **buyer**s to their agents and brokers. Initially, before insurance services are purchased, it is possible that useful detailed quotes and/or insurance advice could be provided to a **buyer** pseudonymously.

(iii) Legal advisors

There are a variety of legal disciplines in which the associated legal services delve into highly sensitive personal information (e. g., bankruptcy law, divorce law, criminal law, etc.) Many lawyers also offer to first-time prospective clients a free consult in which such a privacy-enhanced communications system could be initially beneficial to the parties.

(iv) Family Counseling and Psychological Counseling

The parent patent application also suggests these applications which often involve the exchange of highly confidential personal information.

(v) **Medical Consultations Involving Drug Prescriptions**

In this variation, although medical consultation may be conducted anonymously, in order for a physician to prescribe medication, the identity of the patient must be known (within the current regulatory legal requirements).

14.14 Business Models Leveraging Community Dollars

Free or discounted retail products with "niche" partners in each category Free dial-up ISP (as an independent ISP or a service to jointly promote free access with ISPs)
Free Cable and ISP service, Free pay-per-view (note that viewing patterns and the associated content could provide additional valuable **user profile** information)
Free phone service (e.g. advertise subscription service on screen phones or audio ads from pay phones)
Free prepaid calling card
Free print media subscriptions (magazines, newspapers)
Free book clubs
Offer any combination of the above with "deep discounts" for each (this can involve $350 community dollars per **buyer** or it may simply involve certain purchasing limitations per customer). Each vertical niche partner in exchange gets exclusivity within their own respective niches to target advertise to those **buyer**s (e.g. retailers).
Free access to sporting events.
Free credit for casinos
Free lottery tickets
Free charity donations
Discounted hotel lodging
Monetary credit to a credit or debit card (either an iamworthit branded card or provided as a partnership with the card companies.
Monetary credit to a diner's club
Free subscriptions plus credit to retail buyer's clubs (on-line or off-line)
Credit or discounts for book clubs
Free musical concerts, or theater presentations, movies or
access to arcade entertainment
Free access to amusement parts or theme parks
Free golf season passes
Free commission fees for stock trading
Free commission fees for travel booking (if implemented for on-line **buyer**s would be less compelled to search for travel information on-line though go off-line to make their bookings).
Providing a substantial credit in a **buyer**'s electronic wallet in exchange for their downloading the electronic wallet to their client.
Free personal home pages (which community dollars could subsidize a high quality site).

14.15 Iamwothit Marketing Strategies

A number of marketing strategies are worthy to note. These include the following:

1. Allow the ISP to promote free Internet access through traditional means of Web advertising (impressions on ad servers) Web advertising whereby this advertising is effectively traded, directly for (e.g. pop-up) advertising on iamworthit. Furthermore, the ad server would be able to recognize through the associated domain names, the **buyer**s which are coming from a competitor ISP. So long as that ISP is not a partner of iamworthit, the associated **buyer** would be selectively targeted with an offer of this sort "free Internet access" by subscribing to iamworthit".

   Smaller ISP's would be particularly compelled by such offers to their direct target prospects. This is because they are operating on a "thin margin". Furthermore, both they and their small regional counterparts would be

particularly vulnerable to this type of advertising by regional competitors from the same geographical area, during specified period of months of initial usage of the service, the share of profit due iamworthit could instead be committed to purchase additional advertising for the Internet service provider (or the balance of this profit traded-out in the form of additional advertising through the ad server partner).

2. The ad server partner could further become an exclusive partner of iamworthit on the following commercial venture:
Relationships as established with on-line merchants and other e-commerce sites. The vendor actively promote an offer to their customers through both off-line media (using a URL unique to that vendor) and on-line advertising through the ad delivery partner. The offer may say (as an example),"receive three hundred dollars in purchasing credit at Books a Million in exchange for subscribing to iamworthit (or receive five hundred dollars worth of discount credits at Books a Million In addition, as a further benefit to the vendor iamworthit could trade its own advertising impressions with impressions on the ad server for the current offer (in order to reach a different base of **buyers** which are not currently subscribed). A particularly compelling industry for this application is on line travel inasmuch as a plaguing problem to this industry is the fact that many **buyers** use travel sites as an informational resource on available travel deals and packages, but ultimately book their trips directly through a travel agent (thus cutting out the travel site). A three hundred dollar a year travel credit would be a compelling incentive to many **buyers** to modify their current travel booking habits. Affiliate networks are also an ideal channel for these types of promotions because affiliate sites agree to participate (typically) purely based upon the degree of the profit sharing opportunity (which would be significantly larger than most types of transaction - based affiliate advertising).

3. Marketing Network Concept to Sell iamworthit - sites which offer a community dollars promotion could, upon the **buyers** subscribing to iamworthit, additionally offer the **buyer** with a down-loadable client based software which provides a small promotion in conjunction with a link to iamworthit. Each time a recipient of the email subscribes to iamworthit, a percentage of the value of that customer is credited back to the **buyer** in the form of community dollars. Each subscriber resulting from the current subscriber's email (though reduced) provides an additional credit to the original subscriber in accordance with the marketing network business model. If the site originally delivering the promotion is not an e-commerce site, a percentage of advertising revenues resulting from the subscriber (and potentially all resulting subscribers) could be used. It could be applied in the form of iamworthit advertising (or exchanged) for advertising in an ad server.

4. Free Community-based Content (e.g. broadband over the Web) - As an alternative to the community dollars scheme, particularly if a site does not have a principle focus on e-commerce, an attractive proposition to Web-sites could be the creation of premium content which is free to iamworthit subscribers as it would be subsidized entirely by community dollars. Each iamworthit **buyer** would be granted free access privileges to the premium content on all sites which are part of the program. Some content may be purchased and/or reusable, other may be entirely site-specific and novel. This model would be particularly appropriate for community sites which are largely member-based (or for example ISP-member-based communities) where much of their value to members is based upon information and other content which it can provide. It is conceivable that all iamworthit-enabled community sites would enable free access to their content by all other iamworthit customers (though it is possible that they may be mutually restricted if members of competitor communities as desired by the community site). If an ISP service is not already provided, a virtual ISP service could additionally be offered at a substantially reduced price or possible free (depending upon the number of community dollars left over. One could imagine further extending this present network of free content to free content iamworthit subscribers for free access to fee-based television programming or VOD services. In as much as community sites and television channels are becoming different media for delivery of the same information as the number of channels increases, VOD becomes technically enabled, and, most imminently, full motion video can be delivered upon demand over the Web.

5. Free Access To Subscription and Fee-for-use Information of the Web - In addition to the aforementioned free community site content, it would be possible to further provide free and automatic access to fee-based information on the Web. These costs may be able to be covered by the model across all or most sites depending upon the usage characteristics of its **buyers** (e.g. assuming advertiser/community dollars payment to the sites are averaged across **buyers** in accordance with the consumption patterns of the average iamworthit **buyer**). The identity (pseudonymous) of the **buyer** would have to either be disclosed to the site via the proxy or a unique pass code (as required by the site) provided to the **buyer** could be automatically entered upon the **buyer** accessing the fee for use area requiring the code. A directory (portal) of these fee-based sites would be a useful adjunct to subscribers.

It would be possible to offer websites the ability to become Internet service providers where the interface to the ISP home page would essentially be heavily branded to that site or portal. Companies like GTE already offer a "Virtual ISP" service in which the content to the ISP home page is unique to the ISP while the network is provided by the virtual ISP service. This model would be particularly compelling for sites which are largely community oriented and have a potentially loyal customer base. Moreover, interestingly, many of these community sites are offering many of the services and capabilities that a full-blown ISP would offer from its home-page, e. g., a portal interface, links to high-quality content, chat/forums, e-commerce, commerce affiliate links, etc.

6. Bundling iamworhit links with hardware with a PC manufacturer - Many PC manufacturers are now recognizing e-commerce as a very important sales channel. The present model would involve the P.C. manufacturer bundling a link along with a promotion for iamworthit. The promotion would offer the **buyer** cash credit for the **buyer**. The PC manufacturer would also receive exclusive advertising rights to target **buyer**s whose browsing behavior profile qualifies them as a future sales prospect. E.g., instead of cash credit as the profile is identified, the offer could then become modified to offer free hardware or credit to their purchase. Because PCs are highly portable, the advertising targeting techniques described in LEIA could add substantial additional value to advertisers. A similar model could be used for manufacturers of PDAs.
7. Providing a service to iamworthit partner vendors whereby their **buyer**s may be automatically matched with each other based upon the similarities of their **user profile**s. This may be done for purposes of introduction or for introducing **buyer**s into on-line chat or discussion forums. This service may be provided as a virtual service wherein a menu of different forums and chats are displayed and accessible from each iamworthit member site. (the underlying methodology for which is described in co-pending patent application "Virtual Community Service for System for Customized Electronic Identification of Desirable Objects"). In accordance with this specification, a variation of the service involves the process for identifying individuals who most closely match a given category or target object. For example in the context of the present implementation a category or content, merchandise or a purchasable being specially promoted may be the focal point of a discussion forum or chat room, which is automatically organized by the Virtual Community agent. Accordingly, a portal (or in accordance with the present trend) a site with which a portal interface is integrated utilize the present techniques for generating virtual communities for each category or sub-category of content on the portal or for direct access into a forum or chat room which was automatically created around that particular site (as the target object used as the matching criterion). As described, the user may navigate a hierarchical menu of virtual communities which may be constructed automatically according to the methods described which involves communities assigned to category, sub-category, and association with corresponding sites. Ideally in this scheme the portal is actually a "virtual portal" which may be utilized in providing access to the communities across numerous sites (and/or ISP home-pages). Users may also be navigated (at the individual user level) which along with their pseudonymous **user profile** data is subject to their data release policies. In a variation of the above schemes, if there is geographical information which is associated and which is released in accordance with the above individuals and/pr communities (e.g., as may be occurring or scheduled to occur in physical space), LEIA may be employed as a primary (or additional) selection criteria for navigating the present information accordingly.

8. Creation of an iamworthit online multi-store retail site - Instead of selling the community dollars model to on-line vendors, an alternative approach would be to establish a retail presence in a (or potentially multiple) retail niches. The primary business model would be to leverage existing large iamworthit subscriber base (involving the other various types of commercial partners) in order to dedicate a certain percentage of the community dollars (e.g. thirty percent or approximately one hundred fifty dollars per customer) which could only be redeemed at that multi-store retail site (and/or the value of these dollars could be worth more at the retail site). In addition, in this model, the independent advertising initiative of iamworthit would be geared towards community dollar credit of that retail site. It should be noted that, because if other outside competition occurs to the basic iamworthit scheme to a substantial degree there will not be a compelling incentive for **buyer**s to adopt a more restricted form of valve (as retail credits at a particular site), versus accepting the credit from a competitor in the form of cash. Thus this model could provide a viable means for attaining a leading position in one or more on-line retail markets if this competition does not substantially exist.

9. Advertising in Exchange for Equity - A potentially attractive optional form of value, which could be provided to iamworthit customers involves equity shares in companies which advertise to the **buyer** (in lieu of community dollar credit or cash). This scheme is an ideal application for iamworthit in as much as iamworthit customers can

be highly targeted and because many Internet-based start-ups are highly niche community oriented (thus iamworthit customers who are interested in the sites can be efficiently identified and targeted). Moreover advertising is typically very expensive which in the absence of accurate targeting may be of questionable value. It should be noted, however, that because the primary objective is to both find viable prospects and to engender an element of loyalty (which the equity model does). This scheme would be the preferred approach to advertising for sites which do not sell on-line where community dollars would be the preferred loyalty engendering scheme. In order for this model to substantially provide its desired advantages of increased advertising exposure to fledgling web based companies, the iamworthit subscriber base would have to be quite substantial.

10. Loyalty credits for off line retailers - Deliver through the back of sales receipts, kiosks or direct mail or on-line substantial purchase credit to retailers (e.g. grocers') customers, using the aforementioned technique of utilizing a unique URL to identify the vendor and/or promotion from which an iamworthit subscriber originally accessed the iamworthit subscription site (thus identifying for both **buyer** and vendor the appropriate denomination and/or terms of community dollars issued to the **buyer**). In the preferred implementation, a loyalty card is used to identify the **buyer** thus enabling the community dollars value to be provided to the customer at check-out as straight credit or possibly an enhancement to loyalty credit. The **buyer** may also be identified via credit card or alternatively a voucher (or coupon) could be printed from the **buyer**'s computer or from a kiosk which is typically situated near the entrance to the store and which could be activated upon insertion of a loyalty card credit card (or associated authorization code) and could also be used to disclose the **buyer**'s community credit balance. A unique identifier for that voucher or coupon is provided and non-tamperability measures are provided such that the **buyer**'s community dollars account can be appropriately debited upon redemption. Preferably, a pre-determined value is specified on each voucher (which could be predetermined by the service or the **buyer**) or alternatively, the total community dollars balance could be specified on the voucher along with the **buyer**'s name/address and redeemable only upon presentation of valid **buyer** ID.

11. Auto Insurance Application - Co-pending patent application entitled "Applications for a Location Enhanced Information Architecture" describes a location-enhanced framework by which statistical methods are used in order to very efficiently and confidently extrapolate the most relevant attributes in predicting automobile accidents (or the avoidance thereof). The correlations from some of the existing metrics used may be refined using this technique e.g. LEIA is able to accurately determine the number of miles a **buyer** drives per week while the **buyer** will often lie about this, thus the basic model may be refined and more accurate information may be provided on a per-**buyer** basis. The scheme also enables completely new metrics to also be identified and utilized as well which may correlate the attributes location with time. It is conceivable that if a **buyer** provides access to this location-enhanced information by an insurer, that the insurer could in turn offer premiums, discounts or deliver credit to the **buyer** which could be added to monetary credit the **buyer** receives for personal information from iamworthit, for example, an iamworthit implementation which uses LEIA to profile and target **buyers** with ads by their location (e.g. while riding in an automobile).

12. **Free Web hosting – A portion of community dollars may be allocated in a revenue sharing arrangement between the buyer and the web-site from which subscriptions to the service are provided, thus enabling Web hosts to offer free hosting services while also receiving full payment which is iamworthit advertising supported. Iamworthit could also easily use its vast statistical data collected form each site in the hosting network in order to predict which sites tend to be most commonly linked to from the present site (and/or share "similar content"). This technique is described in the parent patent application by establishing these links users can better access relevant information . In one preferred variation, all non- customer's sites are not included in the referral links. This level of targeting is likely to drive significant traffic through the Web hosts' network (and even more so if iamworthit user profiles are transferred to the site upon visitation of an iamworthit subscriber as these similar cross-links can also be personalized to the user.**

13. **Free ecommerce/Web development tools & services**

14. **Free ASP Services – Web-centric applications are becoming an increasing central part of the e-business infrastructure. iamworthit could potentially enable organizations to gain free access to these Web-centric applications in exchange for iamworthit enabling their employees and customers (e.g. which may use some of these applications as well).**

15. **Credit to a User's Credit Card – Many credit cards are tailoring promotions to enhance not only acquisition of market share but also loyalty of its users. The user-centric SDI provides an enhanced (e.g. rebates for expenditures) platform which could provide more efficacious loyalty enhancement marketing strategies for card issuers. For example, a credit card user could be given certain credit towards the card (e.g. via a promotional offer) for becoming an iamworthit subscriber (which is largely ad revenue supported). An additional novel application could involve providing an offer for an additional benefit which could be provided if the user performs all of his/her on-line transactions using the present credit and (i.e. wherever the card is accepted).**

16. **Free Personal Web Development Services**

17. **Free Personal Portal for Individuals – It is reasonable to offer individuals completely free, Web design/development and hosting services which are offered and mass marketed. At the time this patent was written, Web developers were offering such free services in exchange for revenue sharing for advertising product syndication or ecommerce offering on the individual's site. Iamworthit could conceivably be deployed in conjunction with this commercial model whereby either a portion of the user's own iamworthit profiles are used to subsidize (or subsidize in part) the portal and the developer and/or revenues shared from profits from iamworthit subscribers who subscribed from the portal are utilized also (or instead). As such, it is also reasonable for such a developer to provide links on affiliate sites (e.g. access an affiliate network) which offers free individual portals in exchange for subscribing to iamworthit (and perhaps agreeing to offer a similar iamworthit expectation from that portal as well from which the user can also share in a portion of the revenue stream. Such a program could even be structured in its revenue distribution to the subscribers as a multi-level marketing network.**

18. **Multi-level Marketing Network – This economic model for distribution of shared revenues from iamworthit may be implemented several ways. For example, as above suggested, the user's Web portal offering solicitations for iamworthit (as a standard offering or in combination with a free personal portal), could be implemented in a number of different ways, such as ; a). The solicitation is presented upon the user's Web portal. b). The solicitation is made through the traditional sales channel of direct one-o-one correspondence with friends or associates. C). The solicitation is attached to any email messages of the user such solicitation provides a link to enable the user to subscribe.**

19. **Affiliate Network Leveraging eCommerce Partnerships – An affiliate program could be developed by partnering with the affiliate network's advertising. Typically each advertiser provides a purchasable(s) which affiliate sites offer for sale on their sites in exchange for a percentage of the resulting sales. In the community dollars variation, each vendor agrees to tailor an offer for their products/services on the affiliate network which is subsidized by community dollars which the user collects after subscribing to iamworthit. An example solicitation could be "receive $100.00 free credit toward future purchases at booksamillion.com plus $50.00 cash.**

20. **Combining off-line transactions with on-line transactions – An iamworthit pseudonymous credit card (see previous description above) could be provided whereby users collect credit for advertising (e. g., on billing statements for the ad) which the user receives as well as the advertising delivered via iamworthit's on-line (Web or targeted email) as well as pseudonymous physical mail, telemarketing calls using user profile pseudonymizations and one all time or pseudonymous phone numbers. This user profile data consists in part of off-line data and on-line data.**
21.
22.
23.
24. **14.16 Iamworthit Interface Features to Enable User Control Over Personal Privacy**

1) Access to User data via the Pseudonymous User Data base and Direct Site Visits

Users access control policies apply to both the context of the user accessing a vendor's Web site portal or alternatively advertisers accessing his/her profile pseudonymously. The user's data access control policy using identity or feature data of the vendors (e.g. their privacy policies or "vendor profiles") may enable users to provide automated privacy rules. Based upon user privacy rules, various levels of user identity disclosure information may be released to the vendor (or other "user") including full identity (and/or address information) pseudonymous user identity, or complete absence of the user and his/her profile (*true anonymity).* The profile may also have a credential *which is issued by a secure function within the client-level proxy* indicating that the user has been willing to disclose all relevant personal information relevant to the vendor (e.g. in exchange for a reward). The user has the right/ability to limit the specific profile information which is accessible on a specific vendor or vendor type basis. Typically this data exists under separate pseudonyms (For *purposes of preventing data sharing,* each vendor has access to a unique pseudonym for each user *which is not shared with other vendors),* in as much as vendors could conceivably *otherwise* share this information reconstructing portions of the profile which the user intended to be kept separate from the vendor. As is previously described the user may be "alerted" if/when different vendors under a common identifying pseudonym (or identity disclosure) are receiving (collectively) too much information for the user to be the guaranteed of a certain desired level of identity protection among an *aggregation* of similar individuals. *This notification could be triggered by either detection of the user being identifiable among an aggregate group or users which is smaller than that desired by the user or that his/her pseudonym may become associated (excluding a "reasonable" statistical confidence threshold) with another (SDI) vendor's information about that user who knows other profile data about him/her or the true identity of the user.

6046460 - 4025752

Buyer Infomediary:The Iamworthit Model

In this section we disclose technical details for the iamworthit model. The iamworthit model is an extension of Secure Data Interchange to provide buyers with financial incentives and customization within a semi-automated profiling system. Vendors benefit through well-targeted advertising (both push and pull), the ability to customize information and services (even to first-time customers), and access to a large database of information about buyer purchasing habits. Buyers benefit through personalization with privacy, and financial rewards in return for releasing profile information.

System Overview

Buyers interact with vendors through an iamworthit-enabled client computer. The client computer allows buyers to interact pseudonmyously with vendors, accept focused advertisements in return for financial rewards, and exchange profile information in return for customization and financial rewards. The client computer also tracks buyer activity on-line, and periodically releases information to a central database, the iamworthit server. The information for each user is maintained anonymously in the iamworthit server, and randomized before it is released. As such, while it is useful for the purposes of building models of buyers, for example purchasing patterns, and is useful for targeting banner advertisements, the privacy of buyers is protected. A level of randomization is selected to provide a statistical guarantee about the degree of certainty with which a vendor could identify the buyer from the information that it submits, given that the vendor knows a buyer's true information set.

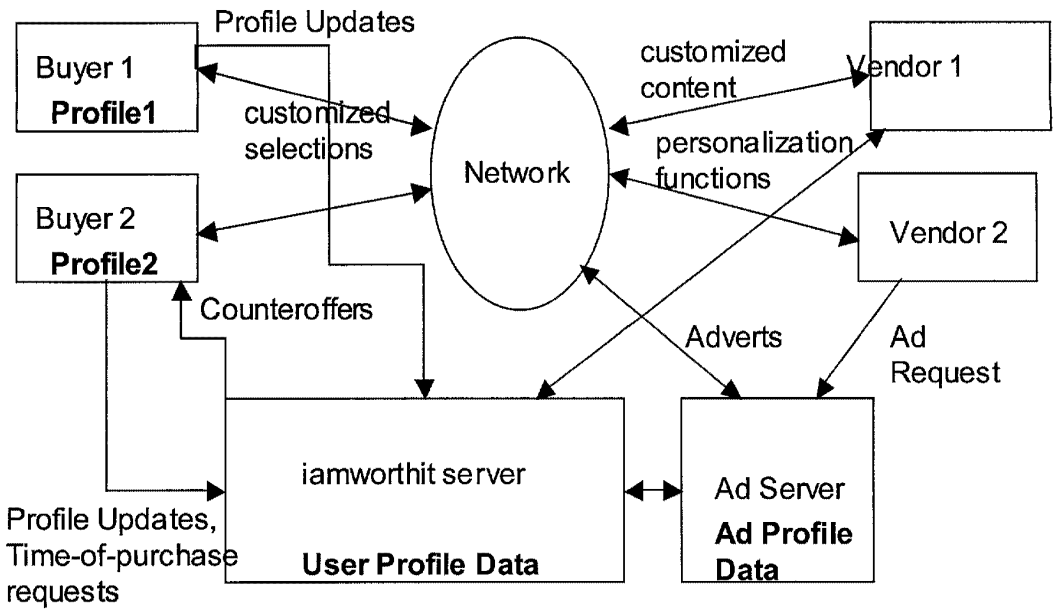


Figure 1. Overall System Architecture.

Figure 1 shows the overall system architecture. The 'buyer' module represents a buyer and an iamworthit-enabled client-side computer. The 'vendor' module represents a vendor and an iamworthit-enabled server. Buyers interact with vendors via the network, with standard SDI functionality to supporting pseudonymous IDs and to control profile release.

Specific to iamworthit is the profile updates from buyer client machines to the iamworthit server. These occur periodically, and profile information (representing web-wide browsing, and detailed vendor-level interactions) is randomized to protect privacy before release. In the basic variation, the iamworthit server maintains a unique one-time pseudonym for each user, so that a user has a single record with a portfolio of information. This ID also allows vendors to push targeted adverts to users, via the iamworthit ID. In general we allow the client-side iamworthit system to submit profile information under multiple identities if a user wishes to maintain multiple "on-line profiles". The iamworthit server also maintains a user acceptance function, that will specify a user's price/relevance tradeoff when receiving advertisements (this is explained in the section on targeted advertisements).

The iamworthit buyer infomediary supports "time-of-purchase" competition that allows vendors to make counteroffers to users that are about to purchase a product or service from another vendor. Time-of-purchase requests can be made by user clients to iamworthit, to request that iamworthit cascades purchase requests onto other vendors, collects responses, and then return them to user clients.

Vendors are provided with access to the iamworthit database for datamining purposes, to build personalization and counteroffer models, and tune advertising models. Vendors can also constract to have iamworthit-enabled advertisements on their web pages, and to push targeted advertisements to users on other pages. The ad server maintains a choice of adverts that "bid" for each user that hits an iamworthit-enabled web page. Adverts and bidding functions are pushed to user clients, and evaluated using local profile information that is maintained for a user.

The system is designed to provide personalization to a user via client-side processing. Vendors can push a "selection function" to a user, that is evaluated on a client machine. The function selects the most appropriate information and

products to display. The client machine computes a selection, and responds to a vendor with its selection, so that the vendor can push personalized content and make customized offers. Each client machine maintains a profile for the buyer. Profiles (anonymous and randomized) are also maintained in the central server for data mining.

Not shown, but also important, are flows of community dollars, that in one implementation are stored as digital cash on a user's client machine. Community dollars provide financial incentives for user's to release profile information to vendors, and can be restricted to be spent in particular ways and with particular vendors.

Iamworthit Buyer Profiling

The iamworthit system maintains a central database of anonymous and randomized information about each buyer that registers with the system. The client-side software is SDI-based, and allows buyers to interact with vendors under multiple pseudonyms. However, it is useful from a profiling perspective to allow users to release information *across pseudonyms* because this allows profiling to leverage cross-web (broad) and single-vendor (deep) information. The iamworthit system allows clients to release information randomized and anonymously, so that users can keep control over privacy. We assume in the following description that each user has a *single* anonymous ID in the iamworthit database, but the system trivially expands to users that maintain more than one ID.

Google Inc. v. Vehicle IP, LLC
IPR2015-00375

The user's iamworthit ID is used for one purpose: to maintain profile information in the database as a unique record, so that replica records do not exist for the same user, and to allow *push* of targeted but unsolicited messages and advertising to users from iamworthit, according to policies stored within the database. The user's iamworthit client will only receive messages from the iamworthit central server.

The client machine for a user periodically sends *update* messages to the iamworthit server, with information about new activity (e.g. web browsing, purchases, adverts not selected, etc.) The iamworthit server maintains information for all users, and allows this information to be used by vendors to build models that enable good personalization of products, services and adverts.

The level of randomization is designed to protect a user's identity to a level of certainty, for an adversary that knows a user's true profile information. As described in the section "randomized aggregates", the client submits the *randomized data point and the distribution that was used to generate the noise.* This allows the central iamworthit server to track the distribution of user profiles in profile space.

A condition of being part of the iamworthit buyer infomediary is that users agree to have profile information pushed to the iamworthit database, to enable profiling and model building for vendors, and targeted push-based advertising.

Real-Time Targeted Advertising

The system is designed to support dynamic personalized advertising, with vendor-side competition for the right to present an advert to a buyer, and buyer side criteria for accepting or refusing an advert. A technical solution is presented to allow vendors to select an appropriate bidding-policy, to maximize revenue from advertising. We use client-side profiling and advertising selection to support targeted advertising with absolute guarantees on the privacy of a buyer. The ad-server sends a *selection* of adverts to a user, and the user's client chooses to display the most appropriate advert. We assume that there is a limited capacity to display an advert to a user, and have vendors *bid* for the right to advertise within an auction, that is implemented on the client –machine of the user.

In overview, the following steps are performed to support targeted advertising to users:

1. Buyer hits the web page of a vendor with an iamworthit advertising link.
2. Vendor pushes content to the buyer (possibly personalized).
3. Vendor contacts iamworthit ad server, requesting personalized advertising and also pushing whatever profile information it already has about the user to support an initial pre-selection of adverts.
4. Ad server pushes a choice of adverts to the user, along with methods to compute the estimated hit rate of the user with each ad, and the value that the vendor will bid to show the ad to the user.
5. Client computer runs its ad selection auction locally, having characterized the relevance of each advert, and the bid that each vendor will make.
6. The client computer proceeds to request graphics for the appropriate ad, and display it to the user.
7. Finally, the buyer receives financial credit for advert.
8. Eventually the buyer's response to the advert is later recorded with the central iamworthit server.

Confidential - Appsys

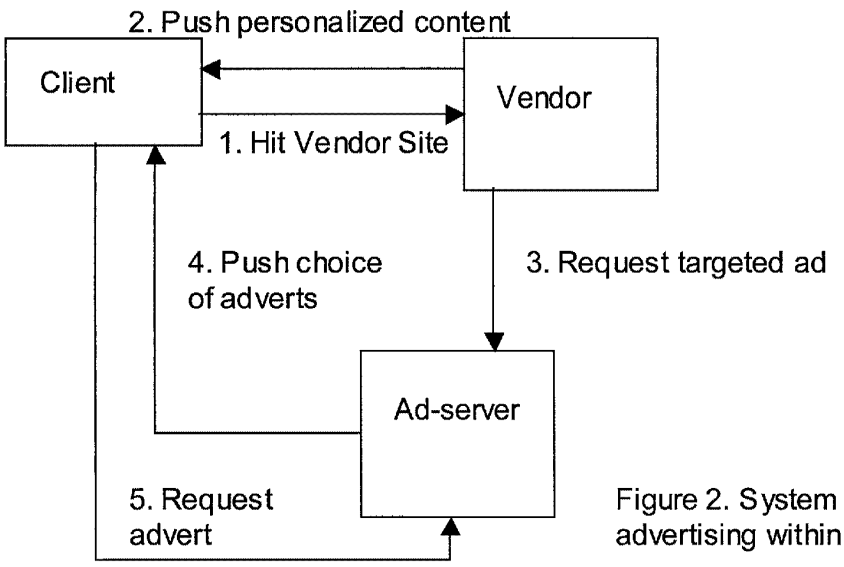


Figure 2. System for targeted advertising within iamworthit.

Figure 2 presents this information in diagrammatic form. Again, notice that the choice of adverts is performed at the client dynamically, using local profile information stored about a user.

The system of iamworthit supports automatic profiling of buyers according to the likelihood that they will respond to an advertisment. The goal is to predict the probability that a buyer will respond to a particular advert, given information about the buyer in its profile. This information is provided to vendors, and allows vendors to compete for the right to present a user with an advert. Furthermore, a client-side selection criteria can also use the probability to decide whether or not to accept an advert, and how much to charge. For example, if a user is likely to respond with 60% probability to an ad then the ad is likely to be revelant and the user will require less compensation than if the ad is likely to be less relevant.

A collaborative filtering system can be used to predict the probability that a user with profile **x** will respond to an advert. The computation is based on historical information for similar users to the advert, and also on the basis of historical information for similar users to similar advertisements if there is little information available about the actual advertisement. The historical information is available in the iamworthit server, stored anonymously for each buyer enrolled in iamworthit, and also randomized to protect the identity of buyers. The basic technique is to select a set of users that have been shown the advert, or a similar advert, and are close in profile attributes to the current buyer. Call this the "relevant set". Given this, the probability Hit(**x, Ad**) that a user with profile **x** will hit an advert with profile **Ad** can be computed as the ratio Hit(**x**) = Num_Hit / Num, where Num_Hit is the number of users in the "relevant set" that responded to the advert, and Num is the total number of users in the relevant set.

Although subject to a certain level of necessary uncertainty, because buyer behavior cannot be predicted perfectly, ultimately it is useful to predict with as high an accuracy as possible whether a buyer will respond. Given a probability Hit(**x, Ad**) that an individual buyer will respond to an advert, a vendor can define a bidding policy. The policy maps Hit(**x, Ad**) to a value to bid for the right to target a buyer.

Assume in this section that the profiling system places buyers into one of several "classes" of buyers, **C**1 to **C**n, given an advert with profile **Ad,** where each class has an associated hit rate, i.e. **x** ε **C**1 ⇒ Hit(**x, Ad**) = Hit(**C**1, **Ad**). Furthermore, assume that the system also predicts the average amount spent by a user that hits on the advert, or the average value to the vendor from a hit, this can again depend on the class of buyer, and can be determined within a profiling system based on historical information about this advert, or about similar adverts. Let Revenue(**C**1, **Ad**) denote the value to a vendor with advert **Ad** of getting a hit with a buyer in class C1. Now, the vendor can determine its expected value for placing an advert to a user in each class, for example using the computation Value(**C**1, **Ad**) = Hit(**C**1, **Ad**) * Revenue(**C**1, **Ad**).

The system of vendor-side advert competition works as follows. Once a user hits a web page with an iamworthit-targeted advert, a choice of adverts are pushed to the user's client from the ad server. Each piece of information about an advert has the following form:

<link to location of advert (graphics and URL of clickthrough destination) , bidding function Bid(**hit, rev**) , profiling function Profile(**x, Ad**), value function Rev(**x, Ad**) >

The profiling function Hit(**x**) = Profile(**x, Ad**) is used to place a buyer into the appropriate class of buyers, and compute the hit probability for a buyer, Hit(**x**) = Hit(**x, Ad**) = Hit(**C, Ad**), where **x** is the profile for the buyer, **C** is the class that the buyer's profile places it within, and Hit(**C, Ad**) is the probability that a buyer in the class will respond to the advert. A possible implementation of the profiling function is to encode it using *prototypical clustercenters* for a buyer in each of the set of buyer class types, and then assign the buyer with profile **x** to the cluster that is closest (in some well-defined metric) to the cluster type. The metric does not need to be linear in each of the dimensions of a user profile, and in particular would be expected to be place more weight on terms that are important to the hit rate in a particular class of buyers.

The above-described implementation of the profiling function has the following useful characteristics. First, it allows the hit probability for a buyer to be computed *at the client machine*, using the profile that is stored at the client for the buyer. The client machine does not need to release the user's profile, and the **user's privacy is assured**. Second, the information necessary to compute the hit probability can be encoded in space **linear in the number of clusters**, which is much more efficient that passing information about every relevant user profile to the client machine. Without computational/communication restrictions one might pass historical information in the form of <Profile, Hit/Miss> pairs for users that have been shown the advert, or shown a similar advert. A nearest-neighbor algorithm could then be used to find the relevant set of profiles for a new profile **x** and compute the expected Hit-rate from the ratio of users with similar profiles that historically hit/missed the advert. We approximate this. The cluster centers are designed to

represent the average hit-rate of "close" profiles, so that the client machine can simply locate the closest cluster center and use that as a proxy for the probability that its user will hit the advert.

A similar technique can be used to compute the expected revenue from a buyer with profile **x** that hits an advert. In fact, this information can be computed using the same method, simply by associating an expected revenue with each cluster center.

The bidding function **bid** = Bid(**hit, rev**) is used to compute the value of the bid from a vendor to place advert **Ad** with a user with hit rate, **hit,** and revenue given that it hits, **rev**. The client implements an **auction for the right to show an advert to a user**. The auction is a *Vickrey auction* between all the adverts that are passed by the ad server to the client. The Vickrey auction (Vickrey, 1961) is a second-price sealed-bid auction. Given bids **b1, b2, ... , bn** the auction sells the right to show an advert to a buyer to the vendor that bids the highest value for the value of the second-highest bid. The auction is *incentive compatible*, which means that the optimal bidding strategy for a vendor is to reveal its *true value for an advert*. This makes the bidding function easy to implement. The value might be computed by multiplying the hit rate and the expected profit given the revenue. For example, given a profit-margin of 20%, a vendor could bid **bid** = 0.2 * **hit** * **rev**. (or with a further discount if the vendor prefers not to bid all of its profit in advertising). Then, the vendor is sure to show its advert for no more than **bid**, and pay only as much as is necessary to beat out its competitors.

We also allow the buyer to have an *acceptance function*, that computes the required financial compensation to view an advert with an expected hit rate, **hit**. Intuitively, when the hit rate is higher then a buyer might require less payment to be shown an advert, but when the hit rate is low a buyer might require a high payment. Let Accept(**hit**) denote the payment required to view an advert with relevance **hit**.

The client-side advert auction runs as follows:

1. For every advert **Ad,** Compute the hit rate of advert **Ad,** given the profiling function Profile(**x, Ad**) and the buyer's profile **x**, and the expected revenue from showing the advert to the user, **rev**.
2. For every advert **Ad,** with hit rate **hit** = Hit(**x, Ad**), compute the acceptance level **accept** = Accept(Hit) for the buyer and the bid for the vendor with the advert, **bid** = Bid(**hit, rev**). If **accept** > **bid** then reject the bid for advert. If no adverts remain, jump to step 7 (in this case no advert is shown to the user).
3. Given the bids that are not rejected by the client, sort them in order of increasing value, and insert a "buy" bid from the buyer equal to the value of its acceptance level for the advert with the bid of highest value. This buy bid will be less than the bid for that advert, because the advert was not rejected in step 2. However, this buy bid might be greater than the value of the second-highest bid for an advert, and is required to make sure that the price paid by the vendor that wins the auction is greater than the buyer's acceptance level.
4. Accept the advert with the greatest bid, and charge the vendor the price of the second-highest bid.
5. Fetch the graphics and URL for the advert, and display the advert to the user.
6. The client monitors the actions of the buyer, and records (to be later transmitted to the iamworthit server) whether the buyer responds to the advert.
7. STOP.

Here is an example. The advert auction message from the ad server arrives with a choice of three adverts, **Ad**1, **Ad**2, and **Ad**3. Each advert is associated with a profiling function, a bidding function, and a value function. The client machine computes the hit rate for each advert, based on the profiling functions and its local profile for the buyer. Suppose **hit**1 = 30%, **hit**2 = 5% and **hit**3 = 15%. The client machine also computes the expected revenue if the buyer

hits an advert, **rev**1 = $2, **rev**2 = $8, **rev**3 = $3. The hit rate and revenue are used to compute bids for each advert, using the *bidding function*. Suppose that the bids are **bid**1 = $0.50, **bid**2 = $0.70 and **bid**3 = $0.40.

Now, the client also computes the acceptance level for each advert, based on the hit rates predicted within the iamworthit system. Suppose **accept**1 = $0.30, **accept**2 = $1.00 and **accept**3 = $0.30. The bid for advert 2 is rejected because it is below the accept value. The bids for adverts 1 and 3 are accepted. Now, the auction is constructed with bids **bid**1 and **bid**3, and **accept**1 because that is the accept-value for the bid with the highest value that is not rejected. The auction takes the bids (0.50, 0.40, 0.30), and sells the right to advertise to the user to vendor 1 for $0.40 (the value of the bid from vendor 3).

Finally, the client fetches the graphics and URL link information for the advert from vendor 1, and transfers payment from the vendor to the user for the right to show the advert,

It might be necessary to prevent users from falsifying information to maximize their revenue income. In a system of trusted iamworthit clients this is unnecessary because iamworthit can ensure that profile information is faithfully stored and updated. However, in a more open system vendors can require that the local evaluation of the *profiling function* is only computed with validated profile information, that is assured by a trusted third-party. The function itself can be digitally encrypted, so that the local iamworthit client can be sure that it was securely generated.

Google Inc. - Proprietary

Tuning a Bidding Function

Although the optimal strategy for a bidding agent in a single Vickrey auction is to reveal its true value for the right to show an advert to the user, the vendors are bidding over a number of different auctions, and might have a limited advertising budget. The system of iamworthit can be used to compute the expected value of a particular bidding function, by performing *monte carlo* analysis on a policy. To enable targeted *push* advertising the iamworthit server stores the acceptance functions of each subscribed user centrally, in addition to profile information (we repeat that this information is all stored anonymously and randomized, to protect the privacy of an individual.) A vendor can use monte carlo analysis in the following way: take a random sample of users in each class of users with different predicted hit rates, and simulate the auction that runs on a user when the user hits the pages that a vendor advertises with. The auction will reveal the average number of times that an advert is shown to a user of each type, and the average amount that the vendor pays to show the advert.

The analysis can be used to determine that although a particular class of buyers are most likely to hit an advert, there is also a lot of competition to show adverts to buyers in that class, and the average cost to show an advert is high. It might prove more cost-effective to advertise to buyers in cheaper but less relevant classes. The analysis can also be used to check that the average buyers is not setting an acceptance level higher than the vendor's bid, which again would indicate that the policy should target a different set of buyers. A vendor might need to advertise to a less desirable "demographic" of users if it is very expensive to advertise to the most appropriate demographic.

A general system for dynamic customized advertising might allow a vendor to specify a bid for each type of buyer, and also a *maximum budget*, so that the vendor that places adverts can maintain control over its spending.

Choosing an Appropriate Advert to Display to a User

The system that we have described can be extended to allow a vendor to select an advert to display to a user based on the type of profile of a user. For example, a vendor can pay for the right to be one of N vendors that compete in an auction for the right to advertise to users that hit the web page of a particular vendor. We can allow a vendor to not only tune its bid to the profile of a particular user that hits a web page, so that only users with a good fit with the service offered receive the web page, but we can also allow a vendor to tune the advert that it shows to a user. Essentially, a single vendor does not only submit a single *profile, valuation, and bid function* to a user when a user hits a page with an advert, but a vendor also requests that the ad server sends a range of different adverts.

Suppose Vendor **V** pushes adverts **Ad1, Ad2** and **Ad3** to a user. The client-side advertising auction first selects one of these adverts to show to the user, the advert with the *greatest bid price.* The actual price that the vendor pays for the right to show an advert to a user is independent of the advertisement, because it depends on the bids from other vendors. Therefore a vendor maximizes its utility by displaying the advert with the greatest value--- as indicated with its bid price.

The hit rate and value for a single advert can be computed in the same way as before, using off-line datamining in the anonymous and randomized iamworthit central database.

GOOG-AT-MP-0000769

Push-based Advertising/Solicitation

In addition to submitting profile information to the central iamworthit server, clients submit user acceptance functions, that state the amount of payment required by a user to receive an unsolicited advertisement from a vendor (e.g. via email). Vendors cannot contact users directly, because the profiles in the central server are associated with the iamworthit anonymous ID, that only allows contact via the iamworthit server. As described in the section on targeted advertising, the level of compensation that a user requires to receive an advertisement will depend on the relevance of an advert, as computed within the system based on the profiles of other users that have responded positively to an ad. A vendor requests the right to advertise to users by sending an ad to the iamworthit server. The server pushes the advert to relevant users and transfers payments between the vendor and users. This is shown in Figure 3.

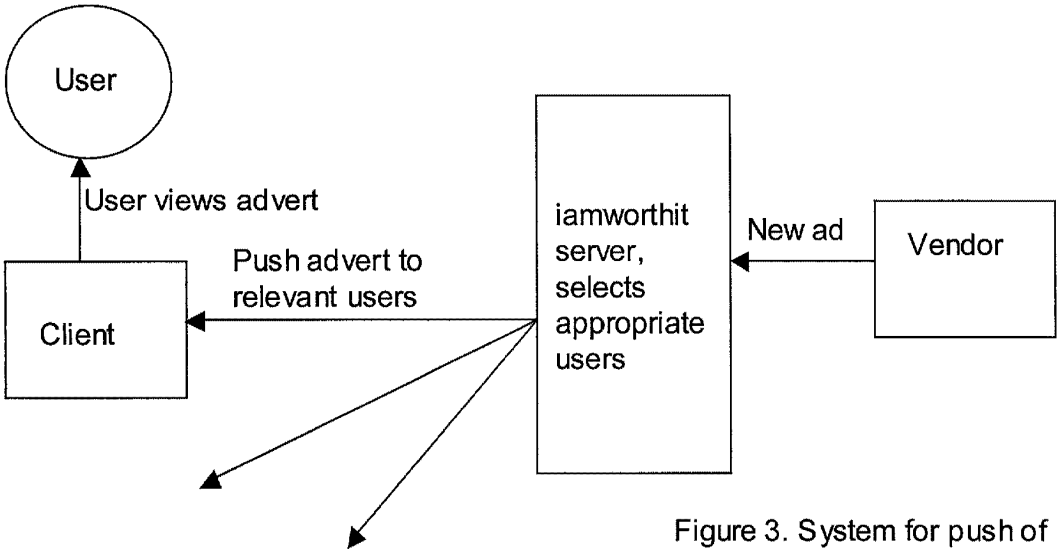


Figure 3. System for push of targeted ads to users in iamworthit.

Suppose Vendor V wants to advertise a new product to relevant users. Vendor V can use the iamworthit database to select users with a high predicted hit rate for the advert, and then set a bid that it is prepared to pay users for the right to provide them with its advert. The vendor submits bids (to the iamworthit server) to selected users, and the iamworthit server evaluates the request with respect to the acceptance function for each user. If successful the vendor pays the minimum value necessary to have its ad accepted by the user, again simulating a Vickrey auction as for the dynamic competition for banner ads described earlier.

For example, suppose a vendor has an advert **Ad** to push to users, and is willing to pay up to $1 to place the advert with users that have a predicted hit rate of greater than 30% on the advert. The vendor first uses the database to identify users with a predicted hit rate that is greater than 30%, and then makes bids to those users. The iamworthit server determines the acceptance level for each user, for the advert **Ad**, and if it is less than $1 then the advert is pushed to the client of the user via the iamworthit server, and a payment of **accept** is made to the user, where **accept < $1** is the amount required for the user to happy to receive the advert.

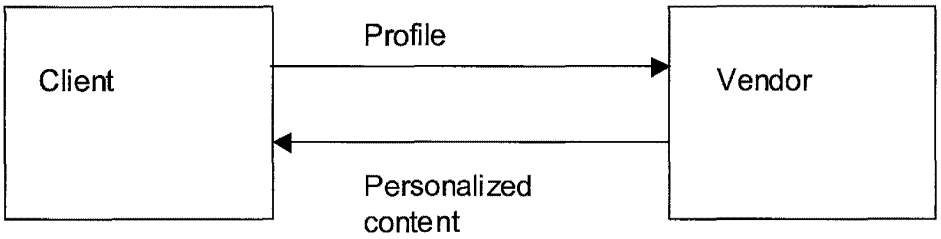
The client records the success/failure of the solicitation to the user, and later updates that information to the iamworthit central database.

A simple extension is to allow users to specify in their advert acceptance policy how many adverts they are prepared to receive a day, so that they are not inundated with too many adverts, even if they receive financial compensation.

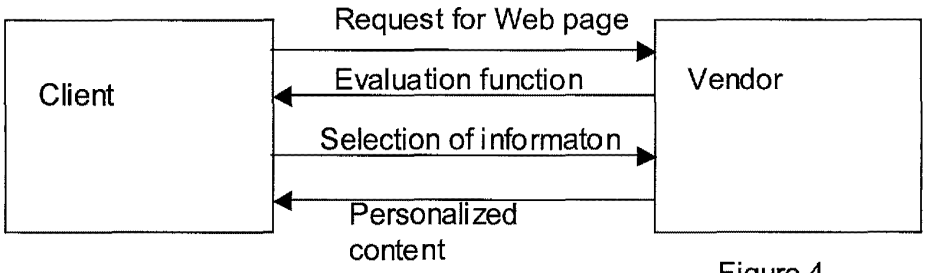
Personalization of Content

A very similar system to the above-disclosed technique for dynamic advert targeting can allow vendors to push personalized content to users without access to the profile of a user. The system will push an "information selection" function to the client, that is tuned by the vendor to select appropriate information based on a user's profile. The

function is evaluated at the client on the basis of local profile information, and the result is returned to the vendor, that can then send appropriate content to the user. An extension can even allow users to request payment in return for allowing a vendor to display personalized content (or vice versa). The traditional method of personalization has a user push profile information to a vendor, and the vendor push personalized information to the user. In our system the vendor pushes its evaluation method to the client, that is computed on the client. Figure 4 compares the system of personalization within iamworthit to that within a traditional system that does not protect a user's profile privacy.



(a) Traditional Method



(b) Iamworthit/SDI Method

Figure 4. Customization of information to users.

Within the iamworthit system we can support one-time pseudonyms for users to vendors, so that vendors can personalized information that they send to a user on the basis of past transactions with that user without cooperation of a user's client. However, the system that we describe allows vendors to also customize information on the basis of a user's transactions with other vendors, browsing activities across the web, and also wider demographic information---all of which can be maintained at a user's client.

The flip-side of the new solution is that the vendor provides the client with its evaluation function, and this might be information that the vendor would rather not release. However, we can protect the function by requiring that the function is evaluated securely within the iamworthit trusted client, so that only the iamworthit client can evaluate the user's choice of information (the result of personalization).

An alternative architecture, that can be readily supported within XML/HTTP is to push *generic* information to the client, together with the evaluation function, so that when the appropriate information has been selected it can be immediately displayed. This removes the final handshake between client and vendor, but can offers a different bandwidth/speed tradeoff.

Finally, we can provide a user with the *option of charging a vendor for the right to push personalized content.* Similar to the advertising auction, a user can maintain an "acceptance value" for the amount of community dollar payment that he/she requires, and a vendor can compute the expected value of personalization versus the expected value of showing generic information, again using datamining in the iamworthit central database. Again, the client for a user tracks the effect of personalization decisions on vendor revenue.

Time-of-Purchase Competition

The iamworthit buyer infomediary system can provide a useful "time-of-purchase" competition service for buyers. The idea is to allow buyers to hold an impromptu "auction" for a particular service or product, just before it is about to commit to purchase from a particular vendor. The system works by maintaining a searchable index of vendors that provide certain products and services, and sending buyer-profile and purchase information to appropriate vendors just before a purchase is made.

Time-of-purchase competition is particularly useful for configurable goods and services, instead of commodity goods. When goods and services are configurable then the vendor that can offer a buyer the best feature-price combination will expect to win the time-of-purchase auction for the right to sell to the buyer.

The key technique is a system for monitoring the interaction between a buyer and a vendor, and notifying other vendors when a purchase is about to take place. The simplest version assumes vendor cooperation, with vendors that *register* with iamworthit's time-of-purchase competition system. This is described below. An alternative, more advanced system, would require more client-side software and automation to classify the type of product or service that a user is requesting from a vendor.

We assume that vendors register with the time-of-purchase competition service, because it will bring more business their way (at the cost of placing them within the system, so that other vendors can compete with them for any sale). When a vendor registers it agrees on a *standard classification system* for products and services, and to state what types of products and services it wishes to enable for time-of-purchase competition (to both receive messages about purchase requests to other vendors, and have messages submitted to other vendors when it is about to make a sale). The classification system encodes the product or service that a user requests, at an appropriate level of detail to allow other vendors to make reasonable counteroffers. For example, if the product is "flights to the UK" then the code might encode the dates that the user wants to travel, and his/her preferred departure and arrival airports.

The iamworthit central server maintains a database of vendor interest sets, so that it can determine from a classification code which vendors will be interested to provide counteroffers.

The system works as follows:

1. User enters into a dialogue with vendor.
2. Vendor makes "final offer" to a user, and the user is about to make a purchase. The vendor has also informed the user of the product/service classification code that encodes relevant information about the type of product the user wants to purchase.
3. User clicks on 'time of purchase competition option' on its local iamworthit-enabled client.

4. Client sends message to the iamworthit server with <User profile information, one-time anonymous ID, product/service classification code>
5. Iamworthit server looks up the product code, and sends the user profile information and anonymous user ID to relevant vendors.
6. Vendors form competitive offers to be shown to the user, based on the user profile and the product code. Vendors send the counteroffers to the iamworthit server, that is able to send them back to the client machine of a user.
7. The user is presented with a set of alternatives, and can make a final purchase decision.
8. The final purchase is completed pseudonymously because the vendor might have received profile information about the user before placing its counteroffer.

In step (4) the user's profile is randomized to protect the identity of the user from an adversary that knows true profile information about the user. The one-time anonymous ID is constructed so that the client-side iamworthit proxy will only accept messages from the iamworthit server, and not from vendors directly. This protects the users from receiving unsolicited messages. The vendors that receive the ID in step (6) can only respond to the iamworthit server, that might forward messages to the buyers if they are relevant to the purchase the user is about to make. In step (8) the final purchases are made psuedonymously, and the true ID of the user is not revealed, to protect the profile information that was released about the user.

Finally, as in all transactions, the client can monitor purchase decisions about offers received from vendors, and update the central iamworthit database with information periodically. The database can be used to allow vendors to provide optimal offers to buyers, based on their profiles.

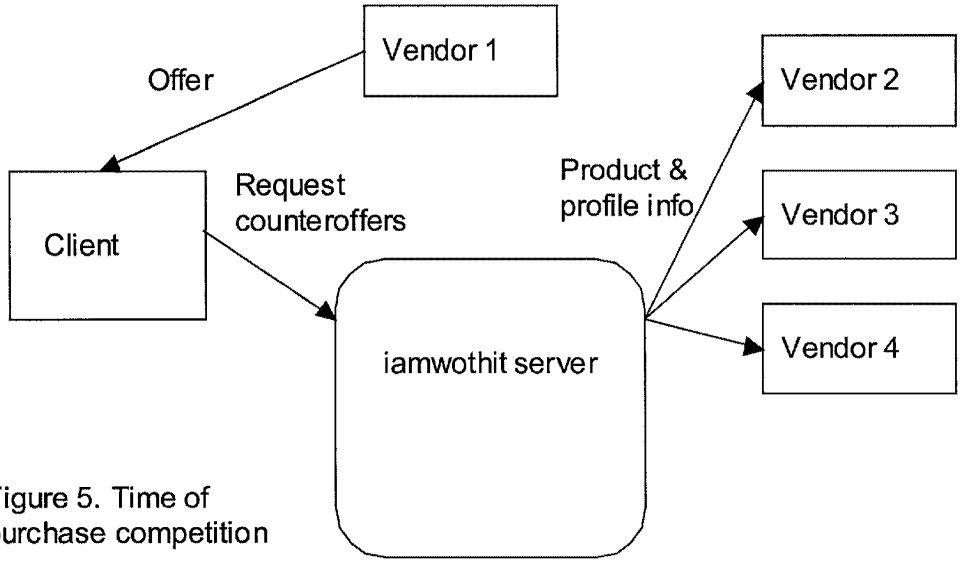


Figure 5. Time of purchase competition

Figure 5 shows the offer process in time-of-purchase competition, with an offer from a vendor that a user interacts with via an iamworthit-enabled client triggering a set of new requests for counteroffers from other vendors. The 'offer' message from Vendor 1 includes a code about what type of product or service the user is requesting, and is passed to the other vendors. The 'request counteroffer' from the client to the iamworthit server contains profile information from

the user to enable the other vendors to make personalized responses. This information is also passed onto the other vendors. After the 'request for offers' are made, the iamworthit server collects responses and forwards them to the client for display to the user.

Virtual User Coalition Generation

The time of purchase competition system can also be used to allow the automatic formation of *user coalitions*, within a computationally efficient system. The coalition formation problem in general is very difficult, because individuals may not have the same goals, and it can be intractable to reach an agreement on the types of outcomes that a coalition should seek, and how gains from trade should be divided between coalition members.

With 'time of purchase' the iamworthit server has information about the current and pending purchase requests from users in the system. The server can *take a position* in the market, and make requests for bulk purchases from vendors for products that appear to be useful to a number of users. The ability to offer a single vendor a number of sales allows significant negotiation leverage, and can be used to get special discounts.

If the server is successful in taking a position in the market for some good that its user base seem to be interested in purchasing, then in addition to returning offers from vendors in the system, the iamworthit system can return an offer for one of the items that it has just purchased in bulk, and should be able to undercut other vendors.

The user is left with a simple accept/decline decision, and the complexity of coalition formation is avoided. The problem is simpler for the iamworthit server because it has an informational advantage as an intermediary. We refer to this as *virtual coalition formation* because no explicit coalition is formed, instead the buyers are dynamically and loosely tied into purchasing as a coalition by making a preferential offer that results from taking a position in the market on behalf of the users.

Community Dollars

The system of iamworthit needs to support "community dollars" that support payment to users in return for advertising and personalized services/products. We suggest community dollars instead of cash because they are more configurable. In particular community dollars can be *restricted* to spending at certain vendors, can have non-linear values, and can be designed to lose value with time. For example, an amount of dollars below a certain threshold might have no value, while more dollars can have more marginal values as they accrue. This can build customer loyalty. Dollars that lose value over time can increase the rate of repeat-purchases. This functionality can be supported within a system of electronic coupons, based on cryptography, as disclosed by Tygar et. Al. (1998).

We need the following functionality, that can be achieved with Chaum's technique of blinded signatures for digital cash. First, we would like a user to be able to receive payment of community dollars from multiple vendors under multiple pseudonyms, but be able to "pay" with the dollars under a single pseudonym. We propose the following system: users receive digital cash in the form of blinded signatures, where the signature represents the value of the cash, and also the restrictions on how it can be spent. The use of "blinded signatures" ensures that the signing party cannot trace the token that it signs. Users are then able to store the tokens locally, and transfer them in an "electronic wallet" to alternative pseudonyms.

A vendor will accept community dollars as part-payment so long as the digital signatures indicate that the dollars are valid for the purchase. An extension to the scheme presents duplication, with real-time verification of tokens.

Eventually, dollars can be exchanged for "regular" cash via an electronic bank, that accepts community dollars (debiting an associated account).

For example, when payment is received in response for displaying an advert, the local SDI client for a user will send a message to the bank for the vendor, requesting the amount of money that the vendor promised to pay to show the advert (this can be certified as the output from a secure Vickey auction).

15.1 Rewarding Users for Profile Information

In fact the competition site would likely pay the user for any additional *user profile* data which is relevant to their own *purchasables* and advertisements and in relation to the amount of marginal value provided by those resulting statistics. There are a variety of different "vendors" which users *(client-level proxy)* may interact *within* such exchange *wherein a percentage of "clickthrough" may be provided to the user as a result with a smaller percentage slated for the user- centric SDI (infomediary) service and the original site from which the user subscribed. As is consistent with the methods taught within the current spec and parent issued patent, the type of user data disclosed and to which vendors (and/or vendor types) is dictated by the user's data disclosure policy. Some of these vendors* include:

1) **The advertiser querying *the* pseudonymous user profile database.**

2) **The user visiting a page with an ad server or ad network (where the clustering or nearest neighbor algorithm interact with the ad server). Ads on that ad server's database, e.g., on the sites most frequently visited by the user are periodically uploaded to SDI such that the ad(s) of highest predicted interest to the user *is* presented upon the user visiting that site. *Preferably the target object profiles of all the ads on the ad server database, as well as the profiling algorithm, which is used by the ad server is properly integrated with that of SDI so that it is possible to convert the usage statistics of the data model used by the ad network (which may include ad server data for users collected across the ad networks) into useful statistics for SDI.***

3) **Content sites, e.g., for personalizing news articles (which are again uploaded to the SDI server each news day if those articles would not otherwise be available on that web site. If they are, as with any web**

content format which is universally accessible, the content may be frequently profiled following retrieval, e.g., by a simple web crawler script.

4) Portal Sites – As detailed in the parent case, the user profile may be presented to the information vendor server containing the portal site. The selection and prioritization of the most relevant personalized content categories may be automatically determined at the client level proxy, which is conveyed to the information vendor server in order to personalize every aspect of the portal interface.

It should be noted that without the vendor's cooperation (with SDI) the interfaces of portal sites and content sites could be automatically modified at the client-level proxy in order to provide the personalization of selections to the user (from the user's perception as if the personalization was dynamically generated on the site itself). However, ad servers/ad networks could not be personalized as easily without privileged access to the ad database (unless the ads from that site were recorded and stored in the ISP-level proxy from other users visiting that site).

As described above the ISP level proxy may also contain metadata relating to specific sites, products or informational content (as an alternative to the metadata residing on the HTML pages, which they describe through cooperation with the vendor). Such metadata includes annotations, average user ratings, according to a variety of attributes, as well as *the ability to browse Web pages with associated annotations and/or user ratings provided from those users who are identified by either: a user-selected cluster, or the user's profile or high or low ratings according to site selected. Endorsements (and ratings by a variety of relevant criteria)* by a variety of types of organizations *may be available as well, and browsing and searching may be performed with these organizational endorsements/ratings as a criteria to bias or filter searches or filter the sites accessible via browsing interface (the same may be also performed for user based ratings as well).*

5) Cache Engines – As an alternative to the profile generation and processing modules **running on the network vendor (ISP) servers, it may operate instead, or in combination as a distributed process upon the caching server in accordance with the method taught by "Broadcast System for Reduced Memory Devices and Asymmetric Networks". (check exact title of this patent) As taught in this patent (in accordance**

with the current applications) the pages predicted to be accessed by the user on the following day are precached in advance. Typically sites which tend to be visited most frequently are assessed for selections which are personally relevant but have never been previously accessed. These would be presented as recommendations to the user if/when he/she visits that site (which is also probabilistically determined). The fact that the recommendation

(whether generated by the site itself or modified at the client level proxy), the user's behavior is positively being reinforced by the fact that the ranking of personal recommendations is prioritized in direct relation with the prioritization scheme for precaching, i.e. personalized recommendations and precached pages are substantially identical and as a result of precaching to the server, or even the client, can be accessed by the user with little or no latency (thus in the ideal embodiment it may even be advantageous to highlight the links which have been precached). ***In a preferred commercial model, the techniques which are above described, that meta data containing profile information on the pages are encrypted such that a cache engine that is used to precache cannot be decrypted and read by certain caching engines which do not possess a desired business relationship with iamworthit. In another implementation, other competitive intermediary services to iamworthit may not be able to decrypt and read these web page profiles in order to present personalized information to the user (even if the competing protocol were pursued, the accuracy of the iamworthit version would be substantially more accurate due to its ability to leverage vendor centric SDI data from the host level proxy.***

6) ***Ecommerce Sites with the iamworthit Profiling System - Vendors are assured that all enabled users (who are subscribed to iamworthit) will receive highly personalized recommendations for their products, services and advertising access from their site. (and because many sites are migrating towards providing also a full-blown niche portal with content, chat and forums,- personalization of this information is enabled as well). In addition to leveraging click-through navigations and transactions captured from their own customers during present and previous visits to their site, iamworthit provides the unique advantage of being able to utilize the user's web-wide profile consisting of a***

variety of attribute types upon accessing the site in order to leverage statistical correlations captured by the integration of the statistical models of the vendor, with that of Iamworthit. If stored on the SDI main server, the iamworthit data captured at the client level or ISP-level proxies is typically concatenated with the data sets of the vendor captured from the host- level proxy (maintained at the vendor's host server or if SDI enabled at the SDI main server)This is particularly important for first-time visitors to the vendors site and particularly within vendor domains in which content-based merchandise (e.g. books, magazines and news subscriptions) are sold. Assuming the number of iamworthit subscribers is substantial enough and assuming that the present business model involves a "relationship" between iamworthit and a vendor of recommender systems and/or (at an arms length) the vendor-centric SDI service, this advantage alone could impose significant pressure upon the vendor to deploy a recommender system(s) which interoperates with iamworthit (or subscriber to SDI). In this application, in particular (though it is also relevant to the other presently suggested informational domains), if the vendor is unwilling to cooperate with SDI, other vendors using the vendor-centric SDI services with whom there exists an agreement to access their data model consisting of target object profiles may be used to make highly accurate dynamic recommendations to the user through automatic generation of "virtual links" which are dynamically generated at the ISP-level proxy for display to the user. In one application, a core sample of subscribers of iamworthit are not presented with these virtual links. The comparative increase in clickthrough and transaction rates with those which are may then be presented to the vendor as compelling evidence of the value to that vendor of integration with the iamworthit service. The same type of statistical experiment could as well be performed also for ads and product syndication links which may be again virtually targeted (at the ISP-level proxy) to iamworthit users based upon their user profiles. In this model the following assumption is made; that as iamworthit's user subscriber base becomes varys, there will be considerable pressure which an advertising vendor may receive as a result or iamworthit's ad replacement technology whereby the adnetwork may wish to negotiate with iamworthit to instead interoperate with iamworthit (for a percentage t the user iamworthit) (to --------- iamworthit's ad blocking/replacement service which affects clickthrough adversely. As the service

operates in distributed fashion, these links may be distributed over multiple other sites on which the targeted ads or products are predicted to reach their target customers. In order to compel the vendor to additionally subscribe to SDI the targeting of these virtual links of product-level items and ads may be further refined by combining statistical data from other vendors which are subscribed to SDI with that of iamworthit. If the other vendors use an ad server or an ad network which cooperates with iamworthit and SDI, i. e. , by accepting its recommendations to profiled and regular users and depending upon what transaction or clickthrough-based percentage the ad server is willing (has negotiated) to provide SDI, iamworthit and the advertising vendor, this statistical data could be used to determine estimated revenues for the vendor if she/he were to 1) interoperate with iamworthit, 2) subscribe to SDI. For purposes of the experiment, this procedure would involve (virtually) "stripping out' existing ads presented by the ad server. If the vendor himself/herself uses a cooperating ad server, estimates may be determined as to how much the vendor could receive from similar revenue shared advertising placed upon his/her own site (again by presenting virtual links to these virtual links to these iamworthit users and comparing user response to those which were not presented with these links. Or (for the benefit of the non-iamworthit subscribed visitor base of that site), the aggregated statistics of these users may be used to suggest item level links which are useful to the site for determining optimal item level standard and affiliate advertising links (through affiliate partnerships) for distribution across the Web (developing and ecommerce enabling these links thus, could provide another potential ecommerce related commercial service provided by SDI). An alternative (parallel) approach to achieve the same objective of convincing the vendor to subscribe to the present services is to identify the basic core engine used by the vendor site and utilizing the techniques above suggested. It is possible to "reverse engineer" the particular data model tuned to that particular vendor's unique data set and/or rules specifications based upon the user inputs and resulting recommendation responses to the iamworthit users from which this data is collected. One preferred implementation providing an ideal commercial marketing strategy for soliciting the recommender engine for vendors involves the following: Because the user data which is captured about the user provides robust detail, not only pertaining to the clickthroughs and transactions of all user subscribed to the system but

SDI-2004-0026795

also personalized dynamically generated page links by a recommendations or rules engine located on the vendor's site, the iamworthit (user side) SDI database in cooperation with the vendor centric SDI service can sell to the vendor centric SDI service or other industry or market research organizations strategic information about the comprehensive behavior activities and user profiles of visitors and customers of these vendors (as is suggested earlier in the spec). Additionally, targeted survey questions may be presented on behalf of these entities in order to extract further information which may be correlated with certain features and attributes of these users. One such method, rapid profiling is detailed in the parent application. Of particular interest to vendors is the ability to utilize SDI's data mining interface for purposes of observing content and product consumption affinities based on user profiles describing Web wide behavior. This interface may also enable vendors to observe comparative click through and/or transaction rates of their competitors, how these criteria are affected by user profile attributes including geographic criteria (if relevant to that vendor) as well as how these criteria, (including even individual multivendor customers) are affected by various types of recommender and rule based engines and further the particular rules and weighted feature correlations used in generating these recommendations (again as described above, these rules and feature relationships may be derived automatically through a core sample of users, who are iamworthit subscribers). This novel informational capability and service of iamworthit provides an ideal framework by which a vendor(s) and/or the vendor centric SDI service (at arm's length) which has a commercial relationship with iamworthit could leverage this iamworthit platform as a commercial advantage in selling vendor centric recommendation products and services. For example, the vendor could utilize the present statistical data to tune certain features and implement certain rules (either directly or with the help of a user centric SDI analyst. Moreover, direct real time response feed back may be collected on these manipulations to the data model in iterative fashion. Thus, in a particular instantiation, the business model could contain a provision which is a requirement for the privilege to access the above imaworthit user-vendor statistics. That is that the vendor must purchase a particular recommendor system(s) or purchase the services provided by the vendor-centric SDI (though this latter type of

Google - 0045940

relationship must absolutely remain at an arm's length in order to prevent the genesis of any potential conflict of interest and between the two SDI services).

For the above reason, it would be highly advantageous for content sites, news sites, and (particularly for *the critical need to optimize* clickthrough) web portals and ad servers to inter-operate with SDI whereby their servers are able to recognize and automatically display the appropriate personalized pages as recommended by the client-level proxy

15. Secure Data Interchange: General Examples

15.1 Assessing the Value of Data

- Plug together sets of data, and measure predictive accuracy.

15.2 Matching Data Across Vendors

- find patterns in common pseudonyms, denoting common areas of interest.
- use catalogues of order codes and item description to find similarities across data sets

15.3 Targeted Recommendations

- describe CDNow style project : similar customers defined by nearest neighbor and clusters; these are used to create recommendations for an individual customer.

15.4 Leveraging Portal Data

- data from portal to leverage data needs for ISP

15.5 Automated Generation of Customized Web Pages

Analyze customers for broad preferences in choice of web pages visited (corporate, Star Trek fan, etc.). This defines the initial look and feel for the page that greets them at their portal (a teen might enjoy lots of bright colors and sound clips, an investor would prefer a more staid design); different "skins" could be created to match the major categories of customers, and would designate both the graphical design and modules available on the page (e.g., a working stock-ticker for an investor, a real-time weather map for a jogger).

The web pages and information most frequently accessed by a customer would be given priority, and a hierarchy of usage could be developed. Since stock prices are of the highest importance to an investor, a ticker reflecting his portfolio value would stream across the top of the page. However, although he enjoys spending his profits on vacations and automobiles, these are only of secondary interest to him (as revealed by his on-line behavior), and so are relegated

to a sub-menu on his web-page. As his usage changes, the priority level assigned to the modules would change as well, so that when a jogger purchases a treadmill for indoor running, his weather reports won't dominate the top-level screen.

Small children could have simplified browsers, with extra-big buttons and access to pages pre-screened by a "web-nanny" service.

SDI would be used in the initial phases to group customers into general categories based on their patterns of their web surfing, and would be used in later phases to adjust the content and style of their portal home-pages (based on what similar customers seem to be enjoying).

SDI can improve upon this rule engine-based model by giving far more information about the users who tend to perform certain actions on their sites. Moreover a second (outside) vendor which is syndicating products or advertising on the vendor's site could (upon further approval of the hosting vendor) use data mining of the hosting vendor's site and based upon explicit actions of users on his/her site and/or the user's profile ascribe rules for more explicit criteria for targeting these remote prospects. Of course, response of users provides a means for iteratively refusing these rules explicitly (or alternatively implicitly).

The Secure Data Interchange clearly offers significant potential value to a vendor licensing rules engine in many ways. As an additional example, many vendors also use a site traffic reporting tool in order to analyze traffic patterns through their sites. It is very helpful for vendors to analyze, for example, how often certain pages become clicked-on or transactions occur or conversions occur, how moving the link may affect (typically improve) these patterns and how certain rules (which may dynamically create links or move links) affects these patterns. Of course, observing the movement patterns of users throughout the site helps the vendor better understand user's perception of the relatedness and natural progression of page accesses throughout the site. This is particularly so if different rules are experimented with such that optimal click-through or transactions occur particularly for key pages or purchasables. Adding to this statistical data SDI user profile information is able to reveal (via data mining) much more robust relationships between pages as these relationships change with user profile features. Accordingly appropriate rules may be provided for individuals based upon this general user profile information which is gleaned from interests and behavior before visiting the site, i.e., either from the user-centric or vendor-centric SDI (while the user visits the vendor's site or alternatively accesses an affiliate site).

Google Exhibit 1029

15.6 Analyzing Affinities

- suppose a vendor has a list of customers, and knows to some degree what web pages they visited after leaving vendor site. A large collection of customers taken from an ISP will contain their web-surfing behavior. Cluster web sites and cluster customers, finding cluster-to-cluster interactions.
- use this information to classify vendor's customers; gives vendor an edge in knowing customers' tastes.

15.7 Personalization of information.

Personalizing information on-the-fly requires that a vendor has a data model, for example that clusters its current **user**-base according to what they are likely to be interested in. Notice that it is only with more information from outside of their domain that **users** can be clustered with respect to their product space. Technology One: Need to be able to update vendor's data models, re-cluster **users**, on the basis of wider information about the **users**, without revealing that information. i.e. just provide the results of the analysis to the vendor, on his current **user**-base. This will not violate privacy, and allow rapid personalization. Technology Two: Need rapid profiling of a new **user**, in a way that does not reveal personal information to the vendor. i.e. can we provide an algorithm that the vendor can run on cryptographically secured profiles that accompany a new **user** to enable that **user** to be categorized. Alternatively, the vendor would need to request a categorization for the new **user** from SDI.

15.8 Enable **user**'s to find appropriate information/products.

This could be more contentious with vendors, for example if we never recommend particular services.

15.9 Ad networks.

We need to allow ad networks to show ads relevant to the current **user** on the page. Approach: each ad network has a set of ads that it is currently running on a page. Can off-line develop a decision tree to decide how to assign a **user** to an ad, given a profile. Again, we need the **user** to come with the profile, and the ad network to have a secure evaluation technique that can run from profile, get answer, without getting any data.

Ad networks. Track **users** across multiple domains. Leverage the **user** database. Fine level control over ads that the **user** sees. The trusted secure data interchange can operate as an "ad network", allowing for the placement of well-focused banner ads to market goods that are relevant to **users** of a particular content site. Electronic banner ads provide the potential for one-to-one marketing, when the advertising agency has information about the **user** that has just hit a site, together with information about what the **user** is doing local to a site. For example a car manufacturer is able to place a focused advertisement to a **user** that has just performed a search for new cars in a search engine, to a **user** that is known to have a large family and a high disposable income.

There are two possible business models. Firstly, an Internet content provider could purchase access to information placed by vendors and **users** within the Secure Data Interchange database. This information may be "rented" for a period of time, and then whenever a **user** visits the site of the content provider (possibly through the pseudonymous proxy server), the provider can query the data interchange for information about the **user**. The Internet content provider sells well-directed advertisements to vendors. Secondly, the data interchange could sell or rent data to an advertising agency directly, providing information in real-time to enable the advertising agency to provide more focus in its banner ads for its clients. "Per-transaction" pricing is a very powerful pricing model that is enabled with on-line banner ads. It is simple to monitor the number of click-throughs that are received at a particular banner, in response to an advertisement. In the off-line world pricing must be based on the number of impressions, or worst still, the number of mailings sent and it is more critical to understand the expected value of a campaign up front.

The proxy server could also act as an "ad network" itself, and sell focused advertisements for vendors, and purchase ad-space on the sites of content providers. The on-line domain provides this unique opportunity for quick experimentation with advertising strategies in order to get feedback on the likely utility of untested approaches. The system can use a hierarchical cluster tree to identify the most revealing items in a dynamically responsive fashion such that the profiles of all of the selections can be generated with the most minimal amount of interactions with the **user** (see "Rapid Profiling" section in issued patent entitled "System & Method for Customized Electronic Identification of Desirable Objects). Thus a more robust statistical model across multiple vendors is established as a result of the **user**'s click through response of these intelligently selected virtual banners as well as other pages which are subsequently navigated through once the remote site is accessed via the banner.

In the preferred approach rapid profiling not only dynamically identifies and presents items which are most revealing of the other items in the collection, it also selects the **users** whose profiles suggest the greatest familiarity with these items (i.e., potentially correlated items). Furthermore, if the system's objective is to find new **users** or **users** who may be interested in the present vendor's other products, products for which little is known, then it will match **users** who are least familiar with exemplar items. The idea is to reveal the most significant data about the **user profile** with respect to the present collection of items of interest. Finally, rapid profiling can use direct explicit queries to determine interest on an item(s) or to collect demographic data on a **user**.

The target object profiles of advertisements on the ad server are matched against the **user profile** in order to automatically present the most relevant recommendation(s). Typically, the client-side proxy requires the host-level proxy to disclose the target object profiles of the products/services sold by the vendor. This data is stored as meta-tags in XML form and is encrypted. This data can be very useful to the **user** in navigation, filtering and search activities in the future or in a variation the ISP - level proxy a party (a neutral server) could store these target object profiles and selectively disclose relevant pieces of them (e.g. genre cross-correlations) to vendors, which are considered according to the disclosing vendor's data disclosure policy acceptable to receive this data. These profiles are not accessible to the client-level proxy but may be disclosed only if there are restrictions within the vendor's data disclosure policy.

In another variation, if the data to be disclosed to the vendor is acceptable to the original vendor but she/he is untrusting of the vendor, the data is received by the host-level proxy (another neutral third party) instead of the vendor, thus providing the disclosing vendor with an additional level of security, assurance about the use of his/her data while enabling the **users** of such a site to access all of the merchandise or content in a completely personalized fashion. Thus these XML tags are stored in association with, but on a separate server from the actual HTML pages stored on the

vendor's site. Additionally, these profiles are constantly updated by **user profile** data conveyed to the host-level server which operates in distributed fashion.

15.10 Personalization of links

Adding value via data collected about this **user** and other **user**s. Not just surfing data, but wider data-purchasing behavior etc.. On-line commerce enables the dynamic personalization of retail for each **user**. A virtual shop floor can be arranged to match the predicted preferences of each **user**. A commercial Internet site can leverage two types of information in order to personalize its product. For a new **user**, that has never before visited the site, it is very advantageous for the site to already know about the preferences of that **user** in order to personalize the goods and services that it offers. The information provided at the secure data interchange, and gathered from the transactions of a **user** with another vendor, is vital for this type of personalization to first-time **user**s. The second type of information that an Internet site can leverage is information that it has collected from previous interactions with the **user**, information that is collected locally to the site.

15.11 Hospital database

Express Mail Label No. EV 339 704 945 US

In one data application of the client level proxy server, the **user profile** includes medical data which is obtained from medical records. (such as from hospitals or physician's medical records or potentially that of a health insurer). Typically, various physician's offices and hospitals which a patient (hereinafter "**user**") has visited over the years contains separate portions of a **user**'s overall medical history, thus these various sources may be combined upon the **user**'s request by downloading this data to the client-level proxy (or preferably, the **user** enters into a contract with those organizations in which all medical data and updates thereof are downloaded by the organization and/or an "agent" to the organization which transmits a request which is digitally signed by the **user** at the client-level proxy server. The origin of the request (the **user**) is authenticated and may be processed by a human or another agent located at the organization's host computer.

Because of the highly sensitive nature of medical data, there are potential **user** privacy advantages in using randomized aggregates. For example, a **user**'s age, medical history of specific relatives (particulars of which could be more generalized) genetic data, numeric values associated with various medical tests, results for which are a numeric value. This data may be of relevance to pharmaceutical companies, alternative medicine vendor and clinics insurance companies hospitals physicians, clinics and home health care providers, the latter three of which may wish to advertise to patient prospects and extend their medical practices. The privacy architecture herein provided is a critical component for enabling access to **user** data by these commercial entities **and is perhaps critical to the extent that users must explicitly authorize the transfer of medical records from either the physical copying or electronic duplication of an associated transfer of such information to an intermediary (SDI) which is trusted by the user.**

15.12 Smart Web browsing

Definitions: exemplar - the profile of target object or (as pertinent to following description), **user profile** which is "most like" the profile of the cluster to which it belong, perhaps a median metric.

The Platform for Privacy Preferences (P3P) provides for the ability to utilize XML meta-tags for purposes of annotating Web pages with comments from previous visitors to those pages. It further suggests that **user**s with an affinity or otherwise associated with a particular category (e.g. of expressed interest) may both identify themselves as well as select a category (e.g. which they are associated with) and observe the annotations of other **user**s associated with that category. One of the divisional applications of the parent case "System for Customized Electronic Identification of Desirable Objects" relating to the automatic creation of virtual communities suggests that **user**s may be automatically assigned to particular communities (e.g. chat groups, forums etc.) which may be either automatically generated and labeled or constructed and labeled manually.

Additionally, **user**s could further be allowed to rate the annotations to the pages, which could be a means by which qualifications of **user**s to provide annotations could be measured by content domain (i.e. cluster). Their comments (and

particularly future comments) could then receive a priority position in the annotated comments available. Future comments from **users** with a poor rating history for a particular content cluster may be deleted. A persistent interface feature on the tool bar or side bar may provide for annotations to also be accessed by **users** selecting certain profile features of **users** as they browse from page to page. E.g. identify the comments of a news article about abortion by **users** who are self identified as advocates of the Women's Rights Movement, ultra conservative senior citizens, teen women or those with a strong interest in alternative medicine or the Catholic Church or identify the annotations relating to Ford Motor Company by general Motors (though in the case of competitor annotations the above description suggests means protective safeguards for the recipient of the annotation).

The parent case further suggests that **users** may actively provide ratings in a completely privacy protected manner according to various criteria of pages they browse. A reasonable extension would further include being able to observe how these ratings relate in accordance with various **user profiles** (or groups which **users** identify themselves in association with). E.g. overall quality, aesthetic appeal, interest in the content, value (if it is a purchasable) etc. Each of these criteria's overall ratings would vary in accordance with the particular **user** groups which is selected by the **user**. Accordingly **users** may submit as a query a **user profile** or **user** cluster profile a page rating criteria a combination thereof and receive a listing of the pages of highest relevance to the search criteria. These may include, for example, the exemplar **user profiles** of **users** who previously most visited the site or the exemplar **user profiles** of the most predominant clusters of **users** who previously visited the site. In addition, the page rating profiles may be also displayed as well as related links which are determined to be most relevant to that collective group of **users** e.g. as statistically estimated from the referral logs **where members of that group tend to go subsequently or (via SDI) come from** or explicitly identified **such as via active page recommendations or** as book marks by those **users** as being of particular relevance or similarity to the present page.

As suggested and described in the parent issued patent application, these clusters may be accessed through querying (search) filtering (where relevant pages and/or annote to relevant page are "pushed" to the **user** as they appear on the Web), and browsing which includes navigating a hierarchical menu of **users** who are classified according to their passive behavior patterns and/or ratings which have been actively submitted as well as automatic hyperlinking (or automatically linking the presently viewed page to its "nearest neighbor"). Based upon the above criteria as selected by the **user**, ratings and annotations may be viewed in addition to these techniques are preferably deployed in the context of the navigational techniques as taught in the parent case.

The above description also describes the use of a hierarchical menu through which groups of **users** may be identified by their profile features (wherein a profile feature could even be a rating criteria itself of for e.g. an opinion via a site survey). These features could also include a **user** selected rating criteria by a certain type or group of **user** or **users** sharing rating similarities. These features could in turn be used to either selectively filter-out content which falls outside of that criteria as the **user** navigates the information by a variety of means (e,g. it could be used to selectively filter items or even categories in the hierarchical menu) or identify if/when pages are otherwise encountered where these **user** rating features are present (or metric features are predominant), thus displaying this **user** statistical information in conjunction with the ratings statistics and/or associated annotations if desired. (NOTE: CLARITY ??)

Other criteria for observing ratings and annotations may be in accordance with those submitted by organizations. As above described, recall that the **user** may also use one or more of the organizational approval criteria to also bias or completely filter selections as the **user** proceeds to navigate the Web. These endorsements may apply to individuals which may either provide annotations, ratings, news group postings, or editorial content.

Vendors delivering targeted on-line advertisements may also be interested in the above information, in particular the profiles of the most exemplary groups of **users** and their affinity ("similarity") toward particular ads, purchased products or product or content categories—as measured against **user** attributes or exemplar **user profiles** of the most relevant clusters of **users** based upon passive behavior and/or active ratings. This information may, of course, may be displayed (or selectively displayed) with the pages for the product (or service) as meta-data in accordance with the vendor's wishes. For vendors, these techniques may of course be applied exclusively within the scope of the web-sites of the vendors concerned (and/or additionally to **users**). For example, the site-specific page view correlations (including time spent viewing each page) in accordance with the dominant clusters, Exemplar **user profiles** and attributes of those **users** are certainly of interest to vendors to which those sites belong as well as affiliate sites on which their advertisements and/or syndicated products are advertised and sold remotely.

15.13 Location Enhanced Delivery System Architecture (LEIA) Enhanced SDI

For an exemplary application of the Secure Data Interchange technology, consider its extension to co-pending patent, entitled "Location Enhanced Delivery System Architecture" (LEIA), we teach a method for matching information providers and information recipients that utilizes location information, in addition to static and dynamic profiling information. The method customizes the information that is displayed on a private or public information device to the real audience in the vicinity of the device, instead of a predicted audience. LEIA collects an extremely detailed and comprehensive information set about the daily activities of a **user**, enabling enhancement of the **user profile** with location information and temporal activity patterns. The co-pending LEIA patent suggests appropriate application environments, for example in a smart home, an office, on a mobile shopping device, and in an automobile. A LEIA-based system stores personal information on **users**.

We can extend LEIA by incorporation with the Secure Data Interchange system that we teach in this patent. SDI enables the **user** to receive the benefits of powerful and well-directed information, but within a system that respects his/her privacy requirements. The interchange acts as a secure data warehouse for **users** and information providers, enabling information providers to target **users** without revealing private information to the providers directly. ***As described in LEIA, the privacy architectures provided for in LEIA and SDI are all the more critical as components of an SDI enhanced LEIA system architecture given the extreme sensitivity of personal location data which must be securely protected in both real-time and within an asynchronous context.***

Co-pending patent, "Location Enhanced Information Delivery System Architecture" (LEIA) customizes information that is displayed to an information recipient based on object profiles and physical location of **users**. Presents the information most relevant to the REAL audience, not a predicted audience ***per se. Because of LEIA's ability to combine user's preference information with information about their current physical locations and deliver advertising which is both of personal and location contextual relevance to the user, LEIA provides a valuable enhancement to the targeted advertising services provided by iamworthit. For example, instead of advertisers being limited to accessing desired preference attributes of users in the pseudonymous user database exclusively they may additionally access prospective target recipients by present (or anticipated future) location parameters of users connected to mobile and even terrestrial communications networks. One application of LEIA involves the integration of the time of purchase competition architecture. In particular, purchasables which a user is browsing with an associated offer price may be either purchasable on-line or off-line. Based upon his/her present physical location, if the user so desires s/he may identify not only other similar items but also that subset of these items sold by bricks and mortar vendors which are physically proximal to the user's current location (thus available for immediate purchase). By introducing location enhanced information to users about their relative similar proxmities to off-line purchasables the time of purchase vendor competition model is further enhanced by placing additional competitive pressure upon those vendors selling similar goods exclusively on-line (which are not immediately available to the user). In this environment if an off-line product offer is ultimately accepted by the user, a digital certificate could be securely conveyed to the user or authentication of his/her identity could be established at time the offer was accepted and upon redemption.***

One application includes "Smart Home Intelligence", where methods are disclosed by which **users**' real-time behavior may be profiled through their movement throughout their home, and specific interactions with the various network enabled appliances throughout the home. Other inputs may include the **user**'s speech patterns (using voice recognition **in combination with** text analysis). It could for example, note the **user**'s speech content patterns in real-time. Such information provides invaluable **clues** as to the **user**'s present activities, mood and interest state and may be processed by the presently described algorithms tuned with location/time features typically using the assistance of human data analyst to identify the key features and correlations. (This information may also provide enhanced information pertinent to the **user**'s general, static preferences as well).

Other extensions of this scheme are also considered e.g. within the context of the **user**'s office, or automobile and pedestrian activities. This application may thus extended the usefulness of the iamworthit model to advertisers in being able to target **users** through the presently anticipated on-line media as well as networked appliances and in either case, based upon the relevant context of **users**' present activities and behavior (and from this potentially their inferred moods or mental states) within their homes and elsewhere. **Additionally (perhaps most importantly), LEIA provides a hi ghly beneficial value to vendors whose customers purchase primarily from the vendor'sbricks and mortal store presence. SDI can identify multiple vendors which share a common physical location. Additionally, LEIA can**

also identify the physical location of buyers on mobile and terrestrial networks and suggest the sharing of off-line prospects for these off-line retailers on the basis of not only interest but physical location.

15.14 Extended Example: Vacation Packages

A vacation package organizer decides to begin a large-scale marketing campaign to target those people who would be the most interested in joining a new Caribbean Cruise. Although the vendor has a database of current customers, it is interested both in increasing the number and suitability of its potential leads.

Interfacing with the secure data interchange with which it is a member, the organizer identifies several possible sources of supplemental data: a LEIA-based travel discussion group, an on-line bookstore, and a Caribbean restaurant. These are found both by browsing through the interchange's internal list of members, and by using SDI-based data analysis tools, used within the interchange to automatically identify entities sharing common characteristics.

The package organizer then contacts each of these entities through the interchange, and negotiates different data-sharing deals: the travel discussion group is willing to exchange full information for a large travel discount, the on-line book store is willing to reveal the pseudonyms of **user**s who have bought travel books in exchange for a per-sale commission, and the restaurant is willing to sell its entire database for a flat fee (and will provide an aggregated data set as a sample).

The vacation package organizer now chooses fairly basic data-mining algorithms to identify the individuals with the greatest potential interest in a Caribbean vacation; however, the organizer does splurge on a new neural network approach developed by a small software company. On a per-sale commission, the Software Company is willing to loan the vacation package organizer use of its data mining code.

First, the organizer decides which data sets to use. The initial results on the restaurant's aggregated data aren't so good (its customers turn out to not be very affluent), so the organizer declines the purchase of the full data set. However, it does agree to the conditions asked by the travel discussion group and the on-line bookstore.

The data provided by the discussion group and on-line bookstore, being in a common format, are moved in a secure fashion to the interchange's processing area, and are acted upon by the data mining tools, which are also in a compatible format. As per the agreement, the interchange forwards discounted Caribbean cruise offers to the members of the discussion group, and forwards standard promotions to targeted individuals in the book store's customer list. A few of these individuals respond favorably; these electronic transfers of money and passed back through the interchange, which slices off a commission for the book store before passing the accepted offers back to the tour organizer, who learns the identities of the customers and can now count them as part of its database.

This protocol specification could even be digitally signed by the "owner" of the data as proof of ownership of the data and its associated restrictions by the owner, i.e., effectively a "digital deed" which is both legal and untamperable by any other party and thus acts as a legally binding proof of ownership and terms/conditions dictating how that data can be used.

16. Future Directions

(i) Digital Ownership Deeds.

A digital-deed is a digitally signed credential issued by a trusted witness of a transaction. It can be useful to have a secure record of a large ticket item that is bought or sold in an on-line transaction. It is convenient to have this verified and maintained, automatically **to provide** increased confidence for members of SDI, particularly those which may wish to have a record or "back-up" verification of large transactions **which may be particularly beneficial,** e.g. within the business to business implementation of SDI or for real-estate transactions, investments in small entities, etc. Additional data may further be associated as part of the digital deed credential, including any restrictions, stipulations (including

time-based), warranties, insurance, **licenses, indemnification,** etc. pertaining to that item. Credentials cannot be tampered with by some other party, **they are** secured using cryptographic techniques.

In the preferred implementation of digital deeds its record and/or any associated conditions to which it is associated (such as herein suggested) may, depending on the user's wishes, be made public information or, alternatively, concealed from all or only certain identified or credentialed users.

In the domain of "information goods", SDI can automatically add an electronic "water mark" to digital information (e.g. a video, a CD, or a book), that associates a license for use, and terms-of-use. In this way (because the water mark cannot be removed, but does not affect the use of the data), the legality of selling/duplicating pure information goods can be verified by inspection of the water mark. Metadata can be used to assign licenses for intellectual property, along with data regarding terms of licenses, assignments, filing and expiration dates, etc. The seller can impose, within this untamperable credential, certain terms and conditions of who the user may be.

Credentials can be used to restrict the ability of minors ***(e.g., via-a-vie the absence of an adult resolution credential)*** to purchase pornographic, or other unsuitable material, and can also be tagged to products that are sold, so that such material cannot be exchanged within SDI without the correct credentials. **The parent patent applications *(use* of "vulgarity"*)* scores could be one modality by which metadata *can* be used. Another could be the method as disclosed by which users are able to rate organizations (or alternatively products or Web pages). XML, for example, on the vendor's site or on the user's client could be programmed to eliminate links to any information which is determined automatically (or manually) via the vendor or the iamworthit service should be made inaccessible to certain individuals based upon their credentials.**

Google - 1005

(ii) Pay for Use Model for Information Access

Credentials can be used to automatically enforce contracts which may be between users and sellers of digital information. For example, such credentials could automatically convey a certain pre-agreed royalty fee on subsequent transactions involving digital goods wherein SDI is automatically notified each time a subsequent transaction occurs involving the use and/or resale of that digital item.

The author of a digital work or application may benefit significantly from this model. Additionally, his/her data release policy could provide restrictions on who may access or use that information and, accordingly, the price which a given individual or organization or (credentialled) types thereof is offered as the royalty fee back to the original author. Alternatively, if an iamworhtit *user* has, for example, traded sufficient advertising credit for free use of particular content which is syndicated through a Web site he/she is visiting, a user credential will identify this fact upon the *users* visiting that site (this credit may be as simple as viewing iamworthit ads while reading the content). An information access control policy which is provided in advance by the user thus *enables* the user agent to automatically negotiate terms and conditions for purchase of ownership or usage rights of all of this digital information. As an example, the vendor centric SDI product syndication network could be modified for the distribution of content. In the advertising application, the site is paid for the delivery of the information to one of their targeted visitors. For content, the site pays for the information delivery rights. The other difference is that the content selection, though targeted, is pre-selected by the vendor typically to assure contextual relevancy in order to enhance existing content on the site. *SDI keeps a record of vendors which are commercially competitive in nature. It is further able to characterize (using implicit or explicit feedback) a would-be user and compare his/her profile with the rule base constituting the would-be seller's information disclosure policy in order to determine whether or not the seller's policy allows the sale of information requested by the perspective purchasing vendor. Typically the licensing vendor quotes a price for the content where this content is identified directly or found through search or matching of a "feaser" (sample) message. The time of purchase competition system may then be employed to identify other similar content (as determined using a text matching scheme). All sites which are members of SDI are equipped within the time of purchase competition system in order to receive offers from prospective users or most typically syndicators (of their content). The content is negotiated by the buying vendor (which involves typically some percentage of revenues from future users accessing it on a pay for use basis, or alternatively on a " free basis" either of which is negotiated involving the time of purchase competition system. Once the content is negotiated and acquired the time of purchase competition scheme may also be emulated for user's accessing this information from the vendor. Similar content to that which the user is currently reviewing or as submitted via a "feaser" summary of what the user is interested in may be identified automatically and the*

associated vendors presented offers or counteroffers for it according to the time-of-purchase competition scheme. **Whether the vendor centric SDI or the individual site directly licenses the content, the user agent may initially request that content while searching or browsing the Web. The user agent may present features of his/her profile which have been historically successful in eliciting a lower price from the vendor, e.g., customer loyalty credentials, first time access to the information.** ***In a variation, if the vendor is willing to accept "similar" content. A contextual comparison of the target object profiles of all "similar" content which is available on the Wen.*** **If the content is determined to be "very similar" to other available content, these vendors may also be notified as well as the competing author (owner). Based upon the response, the agent selects the lowest counter offer submitted by each vendor and/or author (which may be zero cost), the *user* agent may re-circulate the counter bids until there is no further movement in price or the value for the desired information reaches zero. For example, the application (digital item) if conveyed through a transaction with another entity, would through a transaction standard, guarantee, through a cryptographically secure function, that a verification of receipt of the required royalty was received by the encrypted address of the appropriate author. Variations of the protocol could also be applied to a "pay for use" transaction model which is based on an internal "meter" of the application. Similarly, often multiple authors may acquire royalty payments commensurate with their contributions. The iamworthit model also provides extremely user friendly royalty payment solution by providing a "debit" model instead of the traditional real time traditional model. As indicated, *users* are able to leverage iamworthit's highly vendor competitive e commerce infrastructure in order to optimize user value. In addition, targeted advertising which the *user* is credited for (including while using the application) is one of the applications in which *users* are able to receive value at no charge to the *user.* It is believed that as the market evolves *pressures* will support the development of this pay-for-use model of Web content and applications. In the preferred embodiment iamworthit advertising is delivered to the *user* primarily through "replacement advertising" in which traditional ad banners delivered from the sites are stripped using ad blocking technology and replaced with iamworthit ads (thus appealing to most Web users by eliminating the intrusiveness factor of advertising which the *user* is paid to receive). If over the period of months or years, the majority of Web *users* adopt this technology, sites which are content based will need to migrate to some form of transactional model in order to replace lost revenues incurred by blocking of ads of iamworthit users which provides a "wallet" for the user in which microtransactions are performed based on the user's consumption patterns and where iamworthit's advertiser supported user credit is able to substantially subsidize these transactions in as much as iamworthit ads are delivered as the user consumes the content. If the content is very expensive, the cost of the content may accrue faster than the compensation by advertisers. So it is conceivable that a net payment by users may be required in some cases. Short of a mass adoption of infomediary services it is indeed advantageous for individual sites individually to adopt this pay-for-use content consumption model as it would provide a beneficial solution to advertising *revenue* losses incurred as a result of user infomediary services.** ***In addition to advertising another methodology can also be employed in order to allow users to gain credit which may off set the expenses accrued thorough the consumptions of pay for use content. Because content is typically targeted to the user (based on his/her user profile), assuming that user typically consumer user content which is similar to a certain percentage of ads which vendors are willing to pay the user to read, it is possible to not only market this relevant advertising to the content but also pay the user (from the advertiser) for reading that content in accordance with the pay-for-use model (SDI may also determine the effectiveness in the firm of transactions which occur as a result of users who access that information as compared to those who do not access that information.*** **So long as the transaction occurs though the conveyance of an SDI digital deed, the royalty payment (which is effectively assured) need not be for digital goods exclusively.**

6044640-402765

iii) Digital Deeds for Use within a Resale Market

Credentials can also be used to create a "resale market" for potentially any and all items a **user** purchases (which has a potential resale value). The **user** that buys the good can be an "advertiser" in a resale market. The **user** will want to prescribe controls over the personal information made available to interested parties in the resale market. A **user** becomes a potential seller upon having made a purchase for an item (which is documented and recorded with the digital deed).

She/he is automatically asked upon transaction (by software on the client-level or ISP-level proxy) whether she wishes to have his/her item listed as a potential purchasable. If no, she/he is asked if/when at a later times she/he may change his/her mind. A typical price range for that type of item is presented and a question is asked about an approximate price range she/he would be interested in selling for (though a stated range is optional since the preferred variation involves

the use of a bidding scheme). The **user** may wish to control identifying information about themselves, i.e. they may wish to remain anonymous or pseudonymous. The current seller can however be provided with assurances about the **user**, before entering into a commercial relationship.

Regarding prospective future users, so long as the future transaction occurs on line (or conceivably even if it occurs off-line), a **user** credential can be a required stipulation to ownership of a given product. For example, some vendors may wish to impose a restriction which states that any product which they sell may not be sold to a competitor (based on stated characteristics and/or an explicit list of competitive vendors), or this will prevent an on-line black market for age-restricted goods. Such goods, e.g. information goods, can be protected at source—and water-marked to prevent unidentifiable duplication. **For more valuable or sensitive information, resolution credentials may help to further demonstrate the facts implicitly or explicitly about the user which may suggest (though not guarantee) that s/he is not likely to be, for example, working behind the scenes on behalf of one of those competitors.** Similarly, **users** (e.g. owners of kittens) may have certain personal interests whereby items or personal or sentimental value which are sold should be owned exclusively by certain types of **users**.

iv) Assigning Irrefutable Authority to SDI in Order to Optimize Bargaining Leverage for Users

Another set of examples provides for a means by which a user or entity is able to assign irrefutable authority to SDI (an effective trustee) which assumes from the user or entity absolute decision-making authority with regards to enforcement of particular future conditions or actions *pertaining* to that asset which typically constitute a believable threat.

One example may be the sale of corporate assets, whereby the ability to impose certain restrictions on these assets may provide negotiating leverage at the present time for the seller. The seller may state that if a presently submitted offer under the stated terms is not accepted by X (future) date, the offer **and/or perhaps any subsequent offer** will be **irrevocably** cancelled. The seller can credibly commit to this negotiation strategy, via software credentials. SDI acts as a trusted mediator.

A seller could also make a believable threat to place a penalty on all users, except for on some commercial entities, through irrefutable statements. The excused entity could be a competitor to the prospective user. **Such competitor if formidable enough for the user could even be granted certain benefits as part of the threat, e.g. a first right of refusal, purchase option, or, for example, rights as to use or benefit from the asset, e.g., as with technology, human resources *or* intellectual property.** The seller can also commit to a time limit, such that the seller cannot reverse the time limit. In general, the ability to make irrefutable claims improves the efficiency of negotiation.

Other examples could further involve entrusting the decision-making authority for actual conditions and terms of irrefutable claims to SDI which stipulates these irrefutable terms and conditions as part of the negotiation process. Typically in this case, SDI is further entrusted with the power to negotiate on behalf of the entity with at least a certain degree of autonomy. This approach could also provide an additional advantage to individual users or entities by enabling them (for political reasons) to remain at more of an arms-length from such a sensitive type negotiation process involving believable threats.

v) User Coalitions

An interesting variation of this example involves the situation in which multiple entities represented by SDI users may be negotiating with the same other entity (or individual). It may be for the same purchasable. Alternatively, it may involve different purchasables or the represented group may even be a combination of users and sellers who happen to be presently dealing with the same entity. Accordingly it is important for SDI to achieve *an* understanding as to the individual negotiating parameters as well as an assessment of a market demand model which characterizes the needs and objectives of that entity with regards to the particular prospective transactions being negotiated which SDI is mediating on behalf of each party *which is captured through the main SDI server.* As such it is *often* possible to thus *attempt to* predict the minimal acceptable terms of an offer which would meet those conditions. *The* negotiating leverage from all parties *is* based upon terms which affect multiple (or perhaps all) of the deals collectively. SDI thus has the power to significantly strengthen the negotiating leverage *based upon* a market demand model which characterizes that entity thus benefiting the entrusted parties on the whole. The statistical data used for market demand models from the user-centric or vendor-centric SDI *is* collected about the other parties which *the* parties are negotiating with.

One useful source for estimating this predicted market demand curve ***which they represent. For example it is possible for end-users as a coalition of users to be automatically formed for this purpose. A user-centric SDI is then assigned to the coalition in order to (a) ascertain from user profile information (in addition to subsequent active query or recommended offer feed back/approval information about terms which the coalition would likely wish to receive from the vendor (subject to negotiation) (b) ascertain from vendor behavior relating to past users and coalitions terms which would be reasonable to expect from the vendor in order to anticipate reasonable offers/counter offers to present to the vendor.***
The user-centric SDI may even detect from the user profiles the possibility (followed-up by a confirmation query to the users) that the users may be willing to commit in advance for a specified period of time to certain customer performance criteria e.g. in exchange for not only guaranteed discounts but also an equity stake in the company which may increase over time and where continued ownership of which is contingent upon its sustained user performance as agreed upon in advance for the specified period of time. Assuming sufficient willingness, the coalition as directed by the user centric SDI will submit the recommended offer to the vendor along with its competitors using time of purchase competition. The coalition may even incorporate a believable threat (upon approval of its members) (e.g., that it will not deal with the vendor in the future if the terms of its request (such as long-term discount privileges or conversely promise to give the vendor certain competitive advantages in future user coalition negotiations which may include future SDI coalitions.

As suggested above, such statistical modeling technique as ***disclosed*** **in co-pending patent entitled "System for the Generation of Customized Prices and Promotions" as well as other well established traditional vendor site to demand modeling techniques can be used (depending upon who the represented party is) for the benefit of either user or seller. As further suggested above user coalitions (represented by the user-side SDI (as in the b to b application) and opposing seller coalitions (represented by the seller side SDI) may be organized automatically by the appropriate SDI service. It is also particularly appropriate to utilize SDI's predictive tool suite as a means of predicting the most likely prospects for a given coalition, i.e., matching users who have explicitly indicated an interest or have engaged in negotiations or past dealings with a given vendor. As indicated the recommended users may be targeted** ***either*** **with the same items or completely different items, the latter case in which SDI will try to negotiate a lower price based upon the aggregate sales form the coalition, not the volume of sales for one particular item per se, as in the first instance. In the present application, however, because negotiating pressure is being applied by a collection of entities upon (typically) one entity (or possibly alliance of entities), an SDI (which may typically be** ***spawned*** **from the main SDI service) representing the coalition is created in ad-hoc fashion. The other entity may accordingly wish to acquire its own SDI to also represent its commercial interests. Each SDI may represent multiple parties. No SDI must of course ever represent an entity which is negotiating with another entity which it represents. Coalitions are typically arranged automatically by SDI based upon observed synergies in user and seller activities which SDI observes. In fact the predictive capabilities of SDI could be used to model (1) the anticipated number of cooperating entities which are predicted to join the coalition and (2) from that information, the predicted terms and price which could be anticipated for the user coalition or (for the seller coalition). The anticipated price and volume of sales which could be anticipated if the entity were to accept SDI's invitation to participate in the coalition based upon the terms and conditions thereof as predicted to be most advantageous (collectively) according to SDI.**

There are several other applications in which user coalitions of iamworthit customers may be used which include the following:

1) ***Users coalitions leveraging control over the vendor's suppliers. One of the roles of the iamworthit infomediary service could be extended to helping users not only identify vendors with whom they are most likely to do business with (vis-a-vie identifying the most desirable products/services for each user individually as the basis of the formation of the coalition), but also identifying of the other companies which do business with the user recommended vendors which companies in particular may present a potential investment opportunity to the user. Or alternatively to determine which suppliers are potential candidates to do business with a company (or preferably multiple companies) which sell products which are of interest to the user.*** **This SDI application is one of several conceivable applications by which the user is able through the intelligence gathering capability of SDI to identify potential opportunities which could potentially benefit the user which are also within control of the vendor wherein commercial pressure may be exerted upon the vendor via the coalitions. In this particular application,** ***assuming***

that the supplier company is viable, an additional qualification as an investment opportunity to the user is how early is the development stage at that time. Additionally, vendors may be selected on the basis of their ability to provide products/services which match the customer's preferences and currently do business with suppliers which are prime candidates for investment (i.e., are early stage, preferably in need of capital investment and importantly are doing business with as many other vendors as possible which sell products/services, in non-competing fashion which are presently or predictively of determined interest to the user. In our implementation, if the vendor is early stage (i.e., represents a reasonable investment opportunity for the (coalition members), the coalition may guarantee the vendor of certain customer loyalty commitments which are guaranteed (upon forfeiture if the terms of loyalty are not lived up to), certain shares of stock in the vendor's company. One criteria for the selection of this vendor includes the particular suppliers which he/she is currently doing business with or would be willing to switch to in exchange for the coalition's business which is typically a long-term commitment. In conjunction the selection of vendors which the coalition approaches is based upon the willingness of other competitive vendors to conform to the wishes of the coalition with regards to the vendors' selection of suppliers typically. Several multivendor (supplier) relationships currently exist. And the objective of iamworthit is to identify suppliers which are well qualified to satisfy the needs of the vendors which are candidates for their businesses and which similarly provide significant growth potential, if these vendors switch over to these suppliers. If the coalition's offer is successful, it will invest in the supplier(s), prior to its public announcement of having obtained the vendor(s) as a customer(s). If the supplier is in substantial need of investment (e.g. a private placement), it is even conceivable that the coalition may be compensated by the supplier in the form of monetary compensation or preferably free or discounted product from the vendor for which they provide their product/services. This may further assure user loyalty from the coalition and is typically limited to an initial period needed to entice the coalition members that the investment in the supplier is worth while in light of the rather high risk in such an early stage company. **Other more elaborate applications are also conceivable, for example, more general requests for knowledge and influence over particular executive decisions which are not highly confidential which may provide benefits to the members of the coalition e.g. in addition other types of partnering decisions with other vendors also potentially to strategic product development directions and revisions which provide the coalition optimal personal satisfaction and value for their purchasing commitments. It is even conceivable that through the aid of vendor credentials (for identification purposes) user coalitions could bargain with vendors using "restricted currency" which stipulates its use must be for particular uses. (or its value for any other users is diminished) (such as vendors in which the coalition has an investment state).**

In a variation, these restrictions could "encourage" subsequent transacting parties of the currency (e.g. the employees of the company) to buy from particular vendors which the coalition support (e.g. in conjunction with or equity stake), perhaps thus encouraging the employees join the coalition as well. Data mining (performed by SDI) will also indicate existing vendor relationship opportunities for investment between/within (respectively) different vendors as well as the flow of currency through sequences of transactions in which these complex vendor synergies exist even indirectly via one (or multiple) removed vendor-vendor and user-vendor relationships.

COALITIONS CREATED FOR PURPOSES OF ELICITING CHANGE ON THE PART OF VENDORS

In addition to applying market pressures, one other example in which entities may be automatically introduced into coalitions using SDI may involve SDI automatically identifying certain situations or complaints from users or entities in which there may be common ground for these parties to collectively apply political or legal pressure upon another entity from which the problem arises, or in one example to apply legal *tactics* for purposes of reclaiming damages from the infringing party. For example, members of SDI could be asked to disclose particular instances of these types of legal infractions even if they are relatively minor (e.g. simply *a short fall in acceptable quality of misrepresentations through misleading advertising etc. may be relatively minor inconveniences* and/or a site may be provided as well for non-SDI members (e.g. www.class.action.com. A *similar application tailored specifically to the Federal Government could also be provided (e.g.,* www.government *industries.com.* In one example, SDI's ability to provide annotations excluding complaints and ratings pertaining to particular organizations and individuals is complemented by the current scheme by introducing the ability of users to apply pressure to the organization or individual as a coalition in order to more effectively

motivate the desired change accordingly. To this end other tactics may be used, for example, a threat to collectively sponsor negative advertising or editorials. In fact, the above technique for presenting a believable threat could be utilized in this case and backed up by an insurance policy which provides in addition to the guaranties that the threatened action will be carried out, for example, by X date unless a particular condition is met, also provides sufficient funds to follow-through accordingly and in one variation further instill the motivation on behalf of the parties by actually *insuring* the outcome of the *litigation* (where the threatened (and guaranteed) consequence is sufficiently great that the entity is much rather proved to settle thus reducing exposure significantly.

Within the present system SDI provides the framework by which appropriate users and sellers may be matched together. It also enables a methodology by which the user interests are protected through the use of matching of sellers offers to competitive vendors (using iamworthit).

(vi) Monitor Privacy Violations

We can also monitor sites to determine privacy violations. A human analyst can review the stated privacy policies of a Website, and monitor information that is used about an individual, from the personalization that a **user** receives. If the practices are consistent with stated policies then a credential can be issued. Otherwise, it may be possible to passively observe the degree (e.g. frequency and nature) of the violation, and what is the particular privacy violation. This can be done in combination with any history and details of privacy litigation (e.g. damages). This information could be very useful to a privacy insurer in determining at any given time, what sites are insurable and the associated risks of insuring them.

This information may also be used to indicate to iamworthit **user**s (e.g. statically via a "black list" or dynamically while browsing) what sites violate **user** privacy, what have been the nature and (potential) extent of violations to other **user**s as delivered via the client-level or ISP-level proxy. The subscribers may choose to adjust their data disclosure policy settings to assume (for example) single site pseudonymity or complete anonymity whenever visiting those sites and only certain profile information (or none) may wish to be released accordingly. This information release benefits also iamworthit in as much as in the preferred implementation, it automatically provides its **user**s with privacy insurance, paid from a share of ad revenues (unless the **user** electively opts out).

We can also provide reports to **user**s, including specific information relating to their own interactions and transactions with a vendor. The information may include pageview statistics, time spent viewing the site or its pages by visitor or collectively, transactions and amounts transacted. In a variation, the service may provide **user** specific information regarding perks, benefits and bonuses (e.g. community dollars/discounts) which the **user** is entitled to**. Of critical importance is the fact that privacy violations detected on sites can be determined for purposes delivering to all iamworthit users via the proxy server all sites which have violated user privacy and the nature of severity of the violation of ----------credentials may be issued based upon the adherences of sites to monitoring their privacy policies as well as the nature of their policy. Buyer coalitions may also be automatically notified in order to take collective action (collectively) against the violating vendor e.g. on tine form of class action suits negative publicity.**

(vii) Vendor-centric Tools

Finally, metrics are useful within electronic commerce—for example, how many hits does an advertisement receive, what is the profile of **user**s that hit the advertisement, how well does targeted advertising perform, etc. This information is available within SDI, through client-level proxy monitoring, and can be released or aggregated, according to **user** data-release policies.

There are a variety of different auditing, validation and reporting services which can conceivably be performed and, furthermore, this data can be analyzed and converted into digital credentials which may be useful to sites. ***Or these* credentials may be** issued directly to the site for the benefit of validating the associated information to visitors to the site (these credentials could alternatively be displayed virtually via the ISP-level proxy).

Auditing and validating click-through rates of various on-line vendors and advertising services. Data collected by iamworthit at the ISP-level proxy server and by the vendor-centric SDI service at the host-level proxy may provide a

form of a reporting service in which SDI could be used to perform on-line usage and transaction measurements across a variety of sites (including competitive sites, personalization enabled, and targeted ad delivery sites etc.)

Existing reporting tools are typically implemented on a vendor's own site, however, the present scheme which monitors the **user** at the **client**-level proxy is valuable for purposes of suggesting correlations of visitation and behavior relating to sites external to the vendor with those simply providing raw aggregates statistics regarding click-through clusters and associated traffic volume on that particular site sometimes as a fraction of time.

The primary goal of SDI is to protect the privacy of **users**, while providing incentives for **users** to provide information (within a controlled environment), and allowing **users** to reveal personalization within a controlled environment. However, it is conceivable that vendors might try to break the system, and make links between **users** across different vendors. Relevant information could include: the timing of access to and from sites, the time between clicks, typing cadence, and cursor movement. The system of SDI disables the Netscape cookie mechanism, removes referral tags from HTTP messages, and anonymizes routing information—however if any one of these systems is not in place (perhaps a **user** is not subscribed to SDI) then vendors can clearly use all of these to gain additional tracking accuracy.

The Vendor-centric version of SDI (unlike the **user**-centric version i.e. iamworthit) may wish to make optimal use of information which is collected from an unidentified visitor's click stream in order to make inferred "guesses" about who the **user** might actually be (e.g. a **user** who is unidentified by cookies, digital certificates, their client-level proxy or their customer account or credit card number). The vendor-centric version of SDI may utilize the data it collects from **users** across multiple vendors in order to make certain inferences about the identity of that individual.

A **user**'s overall profile (available to all vendors) may reveal clues about the identity of **the user**, if it is not sufficiently randomized.

** With vendor cooperation, referral across sites can be used—in conjunction with timing clues. This can be very valuable information, in conjunction with general profile information, similar typing, click rates, interests, etc., i.e., vendor cooperation can recreate some of the information that is lost when the HTTP referral log is carefully "washed" via the SDI proxy server.

A robust **user profile**, across multiple vendors, is very valuable. Such a profile can be very revealing of a **user**'s identity, allowing a more complete profile of the **user**'s interest. Many of the same techniques described in SDI to extrapolate **user** preferences (including data mining, collaborative filtering and text analysis) **even the feature of time associated with browsing site visits or profile fluctuations (or mood)** may uncover unique identifying features.

Additionally, the fact that this information may be correlated with other unique identifying features, such as dynamic manual features of click stream, (time between clicks), cursor movements and typing cadence may enable the construction of a rather complete picture of the **user** and his/her identity. (e.g. using neural nets **or non-linear kernal regression techniques.**

A **user** could also be presented with questions to answer, to establish identity—combining a certain amount of real data from a **user**'s SDI proxy with the vendor's profiles of **users**.

This can help to positively identify a **user** across sites.

Multiple cooperating sites may use the present techniques to predict, both when a **user** is likely to have returned to his/her site, and when a **user** is likely to have accessed another vendor site. This can allow a vendor to update a **user**'s profile (with cooperation with another vendor), on the basis of what a **user** does at the subsequent site. Profiles can be updated statistically to allow for uncertainty.

Cookies cannot be leveraged under the standard SDI model, because they are disabled.

Theoretically, all vendors could, cooperate in the present initiative (to identify and comprehensively profile the **user** without his/her consent) while not authorizing SDI to do anything further with that data (for data sharing purposes). If each vendor logs the path of **users** (under pseudonyms) as they pass from site to site, it could be possible to use data-mining techniques (with timing patterns) to make statistical predictions about **user** pseudonym portfolios. However, data is likely to be noisy, and the amount of data that is necessary is huge, especially when many **users** may hit a vendor's site simultaneously. **A couple of redeeming factors are the fact that users frequently return to sites, they have visited before. To the extent that the quantity and diversity of information available for consumption is very large (such as a diverse selection of books, grocery, department store selections or portal links etc., the specific combinations of features comprising the user's profile are more likely to be *fairly* unique for any given individual. Of course, even if a particular user or *pseudonym* cannot be singled out with a reasonable level of**

statistical confidence, the user profiles which are generated on-the-fly as the user browses the site may provide a fair amount of indication to the users interests as they pertain to that particular site. Of course these features will provide considerably less insight *into* preferences relating to general web content or correlations with other vertical commercial or content domains (with the possible exception that the site is a Web portal directory to more general commercial vendors and contents).

Unknown Inserts:

Insert No. 1

1) Investment Scheme using a Reputation System

The present system involves an SDI based framework by which users are evaluated by "peers" according to a large range of quality and feature-related attributes which create a highly robust and accurate profile of the user. Users may select different pseudonyms in which their profiles exist, however, upon user request or approval a unified picture of the user can be reconstructed as needed for particular applications such as that of the present. The present scheme further embodies a variation of a reputation system (which is well known in the art) variety of different criteria which are provided to anticipate future achievements and benchmarks of monetary success of the user. These types of predictions often depend upon future variety of very complex interacting variables. Often these variables are complex and difficult at best to use as predictive criteria for eventual future outcomes (such as the user's personal professional success or business success). The human evaluators thus must be very well qualified to provide legitimate opinions on a user by user basis on the basis of these criteria of determined predictive relevance. The evaluators themselves may be screened by other highly qualified individuals as being "sufficiently skilled" to perform the current evaluations. Characteristics of the various individuals which are relevant to the predictions are used as characteristics to in order to group the individuals into a common investment package. Individuals of common characteristics with future potential opportunities which are characterized similarly may be grouped together into an investment pool (in order to diversify the risk to the investors, (a "mutual fund" of sorts). Typically, the individual is in present need of some capital and is willing to share a percentage (typically) of the success from an anticipated future successful achievements, professionally and/or in business ventures.

In a variation, the above scheme provides a very novel means of capitalizing commercial ventures (including very small, early stage and/or highly speculative in nature). The above capabilities are enabled by leveraging the two most unique and key aspects of the system (1) the ability to use experts to collect highly detailed present data using well proven predictive features (2) the ability to utilize the Net to identify numerous very similar investment opportunities or which have numerous very similar opportunity/risk related characteristics (i.e. from an investment perspective).

Insert No. 2

An Ideas Market

In the consumer driven market place there is a significant unrecognized opportunity in being able to harness the intellectual capital of the consumer market. Historically, (in a non-networked environment) leveraging of such knowledge has been impractical and virtually infeasible due to the inability to identify and measure this resource of human skills (and thus match this skill with in the appropriate matching problem), the particular types of problems which companies may be in need of new ideas and solutions as well as the lack of a trusted

intermediary which can validate on an impartial basis the authentic contributions of the provider of the information (as not having been claimed by the recipient after the fact as having been previously conceived before). Of course, there are remaining obstacles such as the fact that important problems are sometimes maintained confidential within a company (not even disclosed to some its own internal staff). With the emerging of extranets a certain amount of inter-organizational knowledge leveraging is becoming more readily achievable. A secondary advantage over an extranet is because of its ability to act as a trusted intermediary. It can validate the fact that that the individual is only dealing with one commercial entity and not its competitors (at least formally). Thus unlike extranet based knowledge leveraging it is conceivable a trust relationship enabling a certain amount of privileged information disclosure is at least theoretically possible. In the preferred embodiment, SDI retains a record as to the particular problem/solution sets that were successfully or unsuccessfully provided by users in the past in order to improve predictions as to their abilities to successfully address the same types of problems/needs in the future as well as to more accurately determine the "similarity" of different and (perhaps often apparently) disparate problems from the standpoint of the user's ability to solve problems based upon knowledge of their ability to successfully (or unsuccessfully) solve others SDI is entrusted with the knowledge of particular needs of organizations (it may, for example, cooperate with Think Task companies). These needs are commonly oriented ideas (though not exclusively so), for example, suggestions or ideas for product development, marketing, a business model or concept (such as a new business direction), content creation, technology or even potential commercial customers or strategic partners. In lieu of identifying particular problems to which an organization may be desiring a solution, users may instead provide their own ideas and/or solutions to needs or problems which they perceive. SDI vis-a-vie its in-depth knowledge of the internal needs and problems of numerous entities is thus able to determine the relevance of the suggested solutions or idea or alternatively identify a particular organization which is likely to possess matching need which is relevant.

Defendant - Exhibit 1005

17.0 Infrastructure for Smart Home Intelligence – (a Location Enhanced User centric application)

The concept of a "smart home" is not a new one: Even in the 1950's futurists discussed the possibility of imbedding intelligent systems into a house, relieving the homeowners of such boring daily tasks as watering plants or adjusting the thermostat. Such a system could easily be implemented today using control computers running simple sets of IF-THEN rules, but might soon prove unlivable because of its inability to adapt to new situations. For example, a garden party could easily be ruined if inadvertently scheduled at the same time as a scheduled lawn-watering.

LEIA provides a much more sophisticated alternative. Given a house that incorporates an internal local area network (LAN) linking household appliances, sensors, communication devices, and home computers, LEIA could easily form the basis for a truly intelligent and adaptive home environment capable of handling daily chores, security, and information delivery.

Firstly, LEIA could easily subsume the tasks traditionally slated for use in a smart home; however, its ability to infer the location and needs of family members would greatly enhance the handling of these tasks. For example, the above-mentioned garden party would not be spoiled because a calendar agent would inform the watering system that a party is taking place on the lawn. Or, being set in a power-saving mode, LEIA could light only those rooms currently being used; as a family member moved around the house, lights would be turned off and on, tracking the person's movement so that only their immediate surroundings are lit.

Secondly, LEIA could use its ability to predict informational needs (in a manner very similar to that used for the Automotive Information Delivery System) in the house, customized to the personal tastes of different family members. One could imagine recipes being delivered audibly in the kitchen, or a television program that follows a family member from room-to-room as he strolls around the house.

Finally, LEIA would have the sophistication needed to deal with intelligent appliances, telling the swimming pool to clean itself, for example, an hour before the commuter returns home.

LEIA provides a unified infrastructure for the increasingly sophisticated systems being incorporated into modern homes. It would act as the central manager, making sure that basic household tasks, security, and informational needs all work in a coordinated and intelligent manner, adapting smoothly to the uncertainties of daily life.

1. Sensors acquire signals from an individual user

A smart home would have several different types of sensors:

i) "Passive" detectors --
*motion detectors
*sound detectors (including voiceprint and speech analysis module)
*Active Badge detectors
*Window and door sensors (are they open/closed?)

ii) sampled communications
*telephone
*e-mail
*radio/TV
*computer usage

iii) appliances
*alarm clock
*light switches
*stove
*dish washer/clothes washer
*water taps

2. Sensors emit Location Identifiers (LID)

Since all of the sensors above could easily be wired into a LAN, no radio transmission takes place and only simple encryption of signals is required (so that a burglar can't intercept signals by tapping into the side of the house).

3. Secure System translates LIDs to User Identifiers (UID)

In this application, UIDs can correspond directly to registered family members or frequent guests. Pseudonymization is needed only if an interface with commercial systems outside the house is required.

4. UIDs used to access personal profile data.

Three major types of personal data are useful for running the smart home:
i) Explicit rules for routine tasks (E.g., desired schedule for lawn watering)
ii) Expressed preferences (E.g., a desire for morning news about international events)
iii) Historical database of previous interactions between the family and the house (This will allow LEIA to learn daily routines and family habits).

5. Most appropriate set of information chosen

There are two major classes of data that need to be served by LEIA: (1) control signals for household systems and appliances, and (2) information desired by household members.

To a large extent, household systems can be appropriately controlled using a traditional rule-base, although a fuzzy rule-base would be inherently more flexible (IF <the bedroom is somewhat chilly> AND <it's almost time to wake up> THEN <substantially increase the bedroom's heat>). Many of these rules would be explicitly programmed by the family from the beginning; others could be learned using inductive methods.

Although most of the household systems could run unsupervised using such a rule base, LEIA would remain the ultimate arbiter, and could override the rule-based system in exceptional cases. For example, if the family suddenly has to go out-of-town, LEIA can infer their absence and inform the household systems to go on energy-saving modes.

LEIA's core functionalities in this application are very similar to those outlined in the case of the automotive information delivery systems: certain sets of sensor readings will prove useful for inferring the identity of individuals within the house; non-parametric analysis of historic data would allow for the prediction of preferred information content.

It should be noted that certain state variables will have a major impact on this analysis. The date, time, mood (inferred from sounds levels and voice content), and current composition and location of people in the house should be carefully conditioned on. Strong daily routines will emerge (for example, a typical working day), but exceptions will also be noted (when the teenaged son is home alone, and in an exuberant mood, play pop music LOUDLY).

Advertising Content could also be targeted at the most appropriate times and locations. Strong statistical clues could be extracted from what other variables (including time) ads (which match the general profile interest summary for the user) elicit the greatest response. Also, an important source as to the context of the user's present attention focus is text attributes associated with the natural language content of the user's communications (e.g., e-mail, language spoken face to face or over the telephone).

6. Information Delivered

LEIA commands household appliances and systems, monitoring their proper functioning. It constantly adapts the commands to be appropriate to the current state of the household.

LEIA also delivers programming and media to household members -- where they need it and when they need it.

LEIA delivers advertisements to household members where they need it and when they need it.

An extension of the System for art Home Intelligence is its ability to determine the temporal context of user behavior in a location enhanced environment for providing a couple of unique and useful applications includes the following: A miniature camera device which is ceiling-mounted is positioned at a location in the user's home (or alternatively office, etc.) at a location where the user typically conducts reading activities (e. g. in a reading room or living room). This device ideally also has the (dual secondary) function of detecting the location and (inferred) behavior of the user (thus could where used replace and provide more robust information about the user. It could for example match particular behavior patterns with identity information which is gleaned from the movement of an identified individual moving throughout the house from room to room leaving " identity clues" such a voice signature or unique behavioral clues or through the use of the ceiling-mounted miniature came biometrics such as facial image patterns and/or retinal patterns (using retinal scanning techniques). As the primary utility of the device, the user may wish to read information (e.g. newspaper, magazine or book). If the user activates the functionality of the system, the miniature camera device captures data from the particular page(s) which the user is observing, transmits this data (e.g. using an RF signal) to the user's local terminal device (e. g. set-top box, P.C., radio or other). In one variation, this transmitted data consists of a unique identifier code for the particular page of the printed material. In the other variation, this data consists of textual data. Within the scope of the present network architecture, the parent patent application describes how such textual data can be converted into target object profiles and through the use of the profile processing module " similar" target object may be retrieved from t electron media distribution system (cached locally or over the network) and delivered and displayed at the local user terminal device for viewing bye user. Such media (which may be tailored for delivery at a TV set-top) may include streaming video/audio, multi-media files or video-on-demand which are delivered in near real time upon activation of the request and which are of matching similarity to the print media viewed by the user presently. Alternatively, the data captured by the ceiling-,mounted miniature camera device relating to the particular pages of print media presently observed by the user may consist exclusively of an identifier code a portion of which is used by the system for purposes of identification of the code (or alternatively such identification may be determine by its physical coordinates on the page itself). The other portion of the code is a unique identifier of the particular page

which may be used to identify manually by a content expert whereby select print material s and particularly their associated electronic designed and customized for use within the system context of the present convergent media application. It is apparent that such a system could also be usefully deployed to befit advertisers, e. g. presenting advertisements electronically to correspond with particular printed content or matching printed ads with corresponding electronic counterparts. In a variation, as consistent with the iamworthit commercial model, users could be provided credit for the ads which they receive

17. References

[APPEL] "A P3P Preference Exchange Language (APPEL)." World Wide Web Consortium Working Draft (Work in progress), 1998. URL: http://www.w3.org/TR/WD-P3P-preferences

[CKR97] Connolly, D., R. Khare, and A. Rifkin. "The Evolution of Web Documents: The Ascent of XML." *World Wide Web Journal* 2 (Fall 1997): 119-128.

[IRDF] Lassila, O. "Introduction to RDF Metadata." W3C NOTE 1997-11-13. URL:http://www.w3.org/TR/NOTE-rdf-simple-intro

[KRR97] Khare, R., Rohit, and A. Rifkin. "X Marks the Spot: eXtensible Markup Language opens the door to a motherlode of automated Web applications." *IEEE Internet Computing* 1 (July/August 1997): 78-87

[KR97] Khare, R., and A. Rifkin. "Capturing the State of Distributed Systems with XML." *World Wide Web Journal* 2 (Fall 1997):207-218.

[MCF] "Meta Content Framework Using XML." Submitted to W3C, June 1997.

URL: http://www.w3.org/TR/NOTE-MCF-XML-970624

[MCFXML] Guha, R.V., and T. Bray. "Meta Content Framework using XML." Submitted toW3C, June 1997. URL: http://www.w3.org/TR/NOTE-MCF-XML/

[Mil98] Miller, E. "An Introduction to the Resource Description Framework." *D-Lib Magazine* May 1998.

[NS] Bray, T., D. Hollander, and A. Layman. "Name Spaces in XML." W3C Note, January 1998.

URL:http://www.w3.org/TR/1998/NOTE-xml-names-0119

[OPS] "Proposal for an Open Profiling Standard." Submitted to W3C, June 1997. URL:http://www.w3.org/TR/NOTE-OPS-FrameWork.html

[PRIVACY] "Privacy and Profiling on the Web." Submitted to W3C, June 1997.

URL: http://www.w3.org/TR/NOTE-Web-privacy.html

[SCHEMA] The W3C RDF Schema Working Group,

URL:http://www.w3.org/TR/WE-RDF-Schema/

[SGML] *Information Processing—Text and Office Systems—Standard Generalized Markup*

Language (SGML), International Organization for Standardization, Rdf No. ISO 8879:1986, 1986.

[SPEC] "Resource Description Framework (RDF) Model and Syntax."

URL:http://www.w3.org/RDF/Group/WD-rdf-syntax/

[URI1] *RFC 1738: Uniform Resource Locators (URL)*. IETF RFC1738 IETF (Internet Engineering Task Force, ed. T. Berners-Lee, L. Masinter, and M. McCahill. 1994.

[URI2] *RFC 1808: Relative Uniform Resource Locators*. IETF RFC1808 IETF (Internet Engineering Task Force). ed. R. Fielding, 1995.

[VC] *vCard Home Page*, URL:http://www.imc.org/pdi

[W3C] *The World Wide Web Consortium Home Page*, URL:http://www.w3.org/

[XML1] Bray, T., J. Paoli, and C. M. Sperberg-McQueen. "Extensible Markup Language (XML): Part I. Syntax", World Wide Web Consortium Working Draft (Work in progress), August 1997. URL:http://www.w3.org/TR/WD-xml-lang.html

[XML2] The W3C XML Extensible Markup Language Working Group Home Page,

URL:http://www.w3.org/XML/

SDL-1053938

CONFIDENTIAL

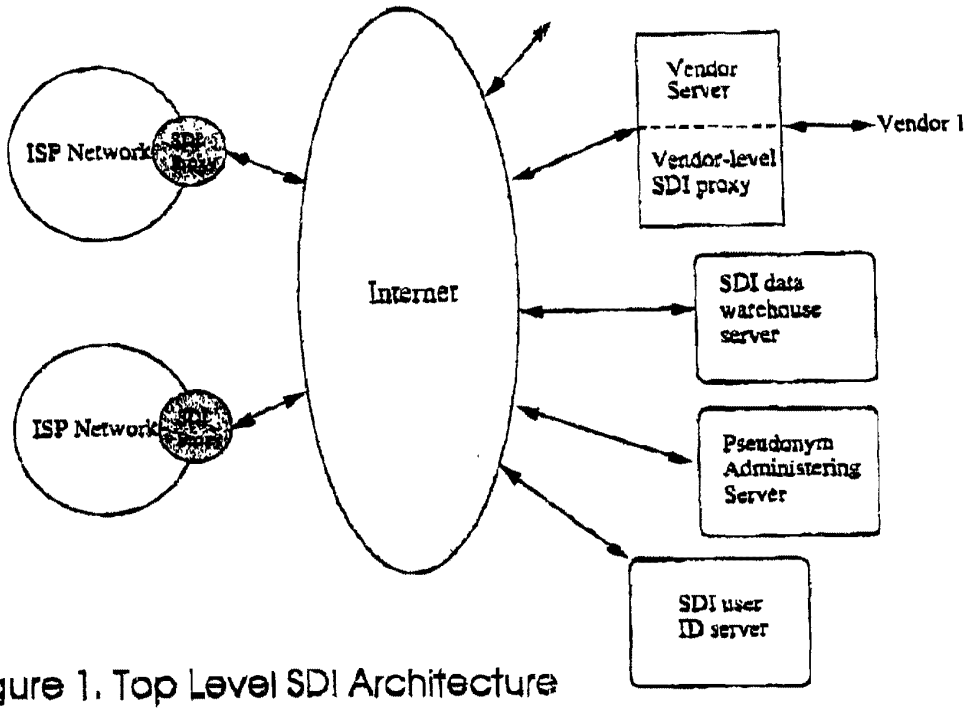


Figure 1. Top Level SDI Architecture

CONFIDENTIAL

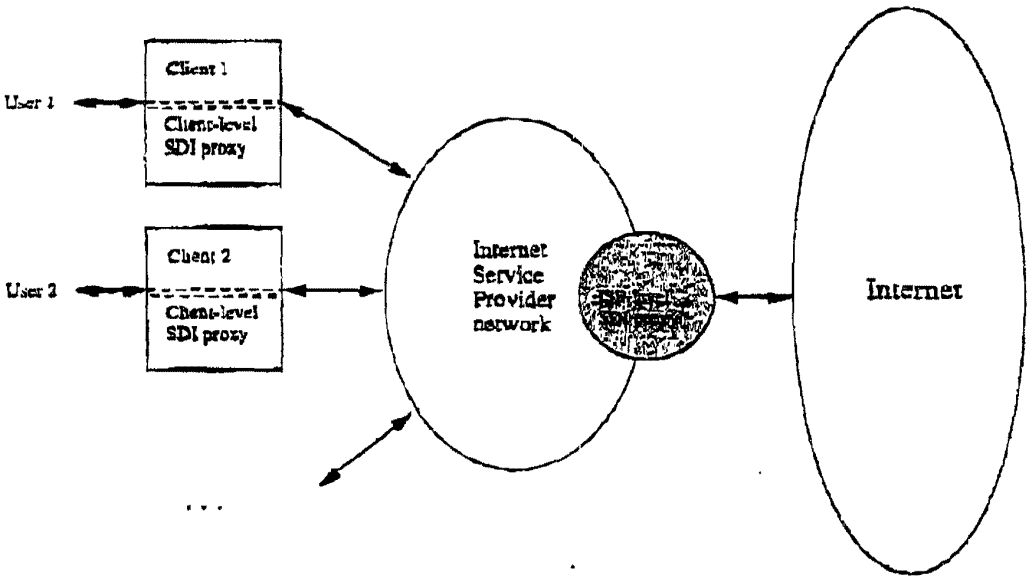
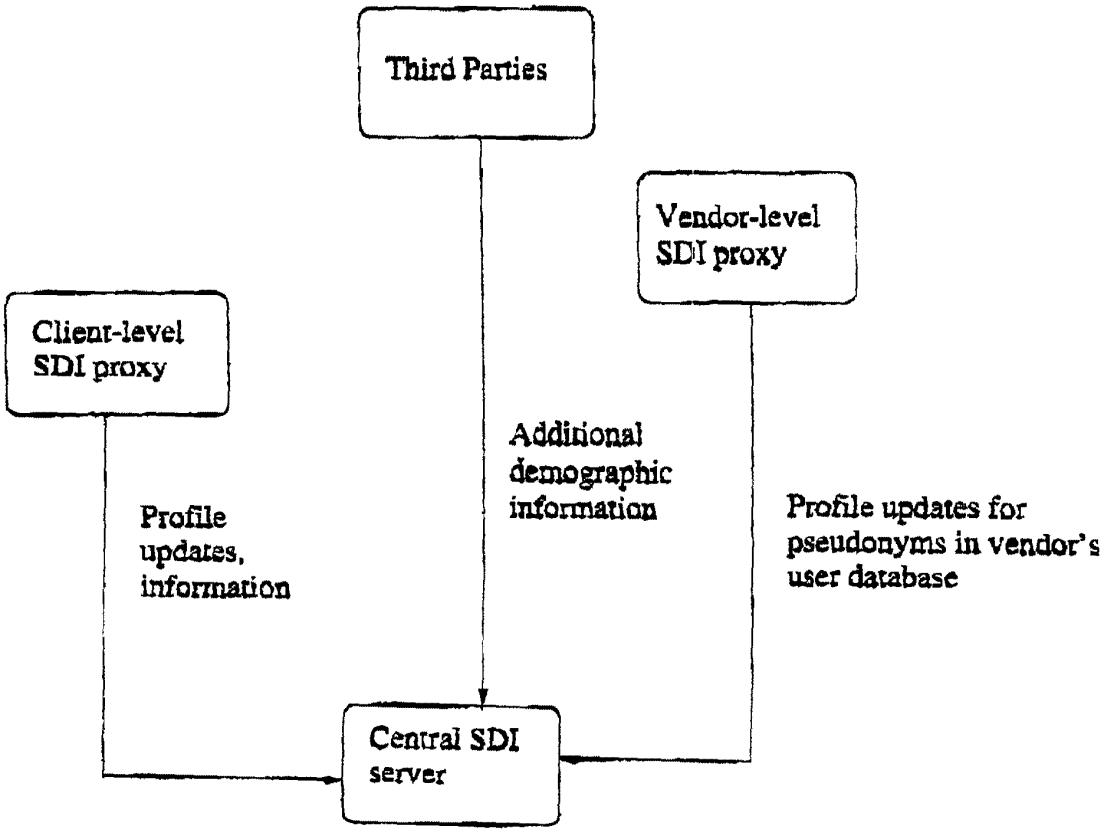


Figure 2. Client-level SDI and ISP-level SDI proxies

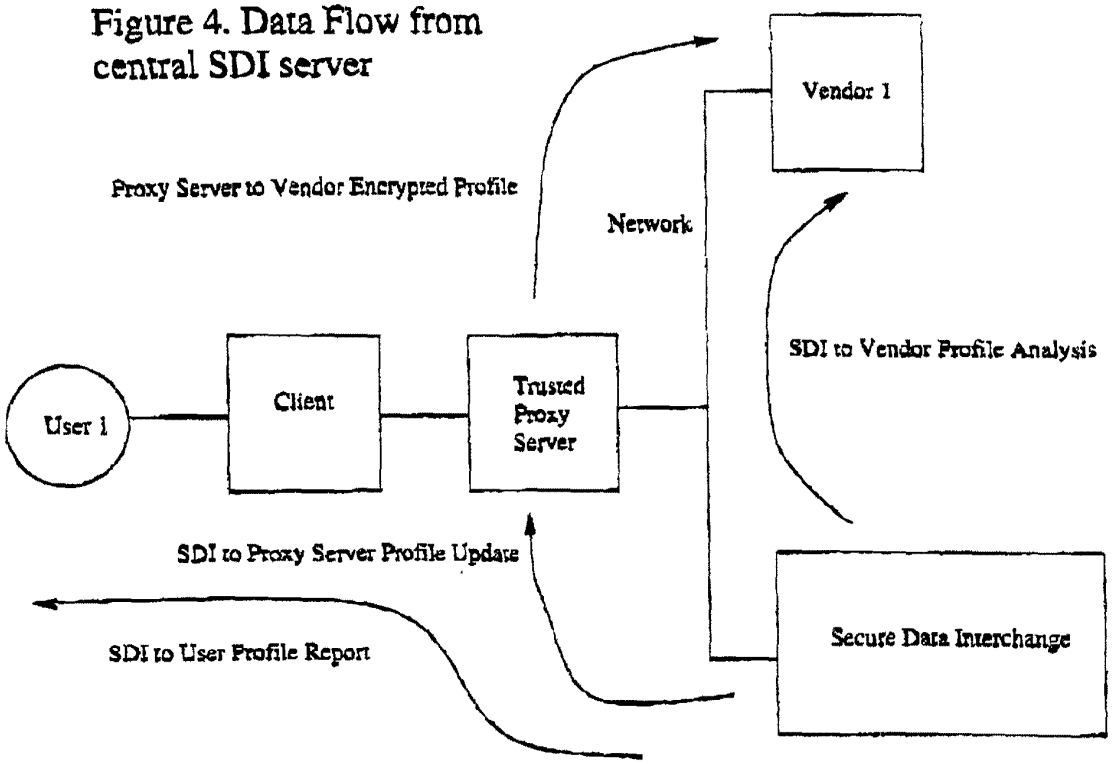
Sprint-Exhibit 1056

Figure 3. Central SDI server Profile Updates



Declassified - EO 13526

Figure 4. Data Flow from central SDI server



SRI-SDI-000396

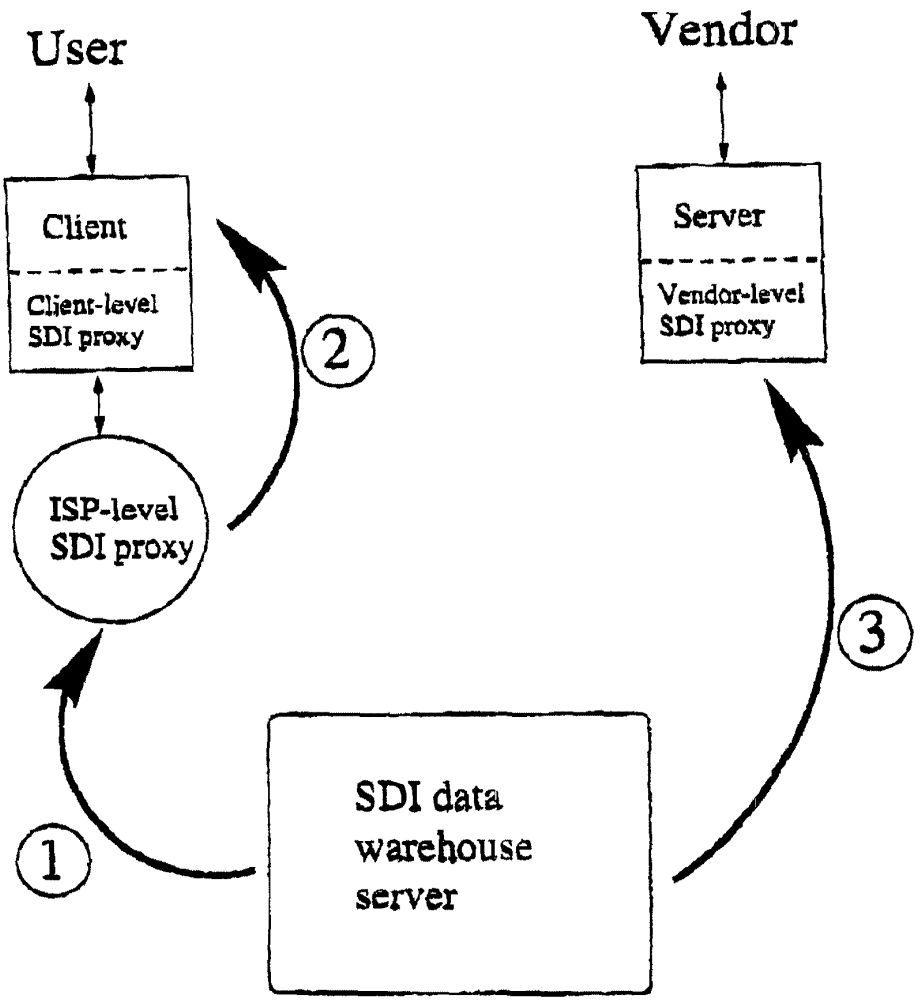


Figure 5. Data flow from central SDI server

DOE/EIA-0548-1998

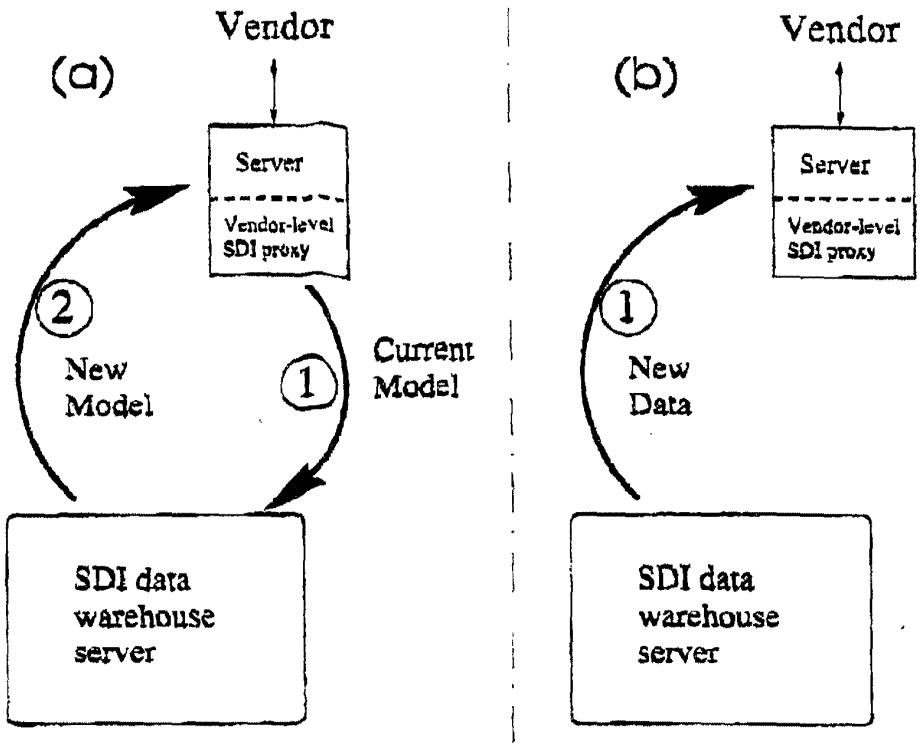


Figure 6. Updating Vendor Profiles: (i) Model-based update; (ii) Data-based update.

SDI-0155-1999

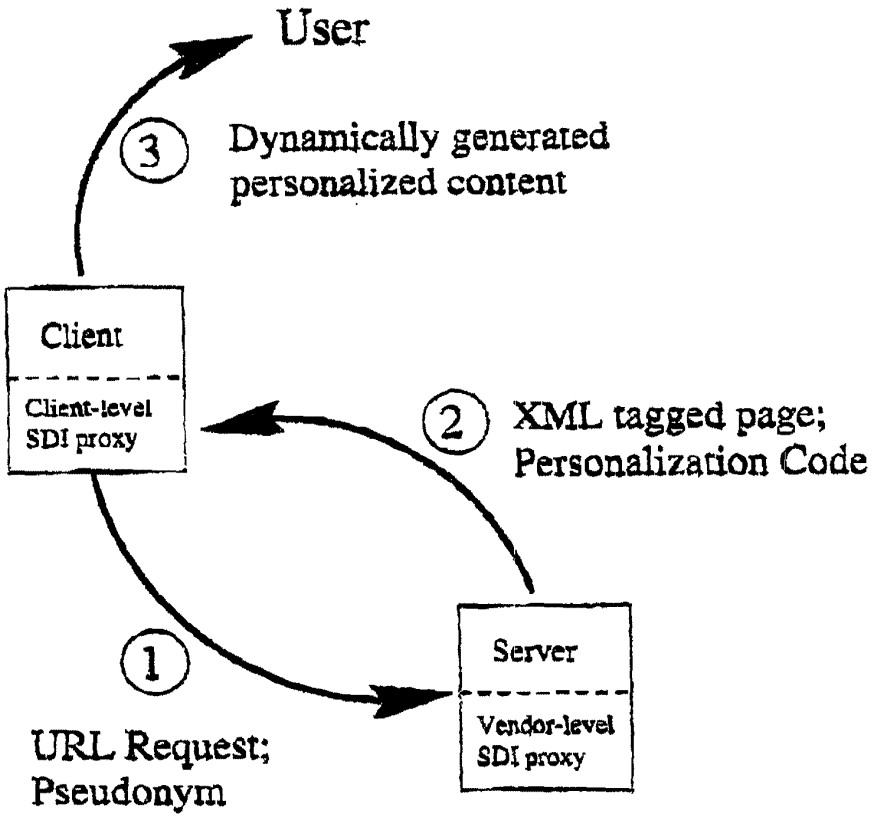


Figure 7. Client-based Personalization of Information

Google Exhibit 1015

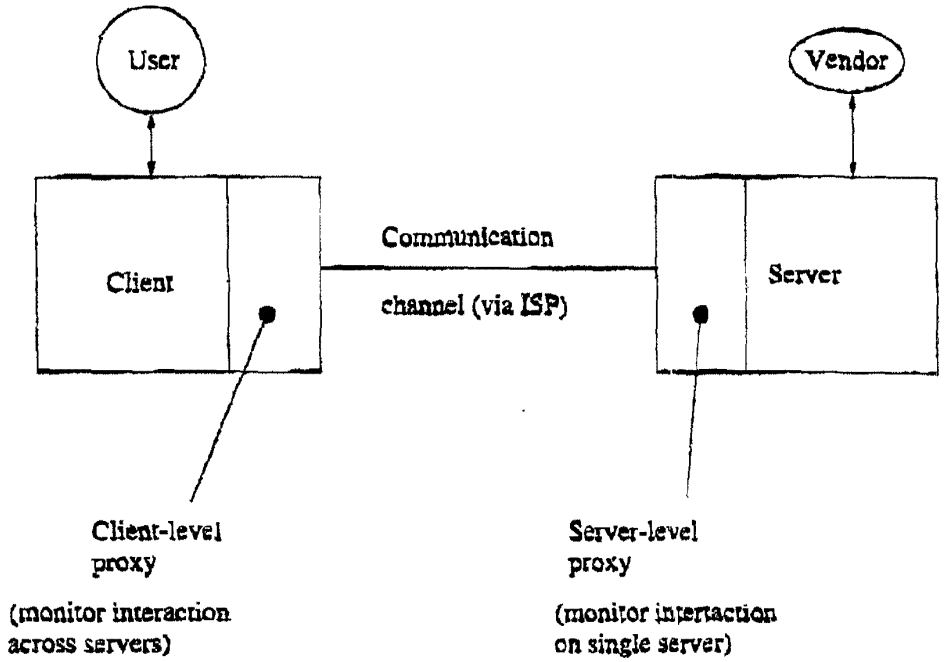


Figure 8. Monitoring a User's browsing: combining vendor-level and client-level information.

Confidential - Highly Confidential

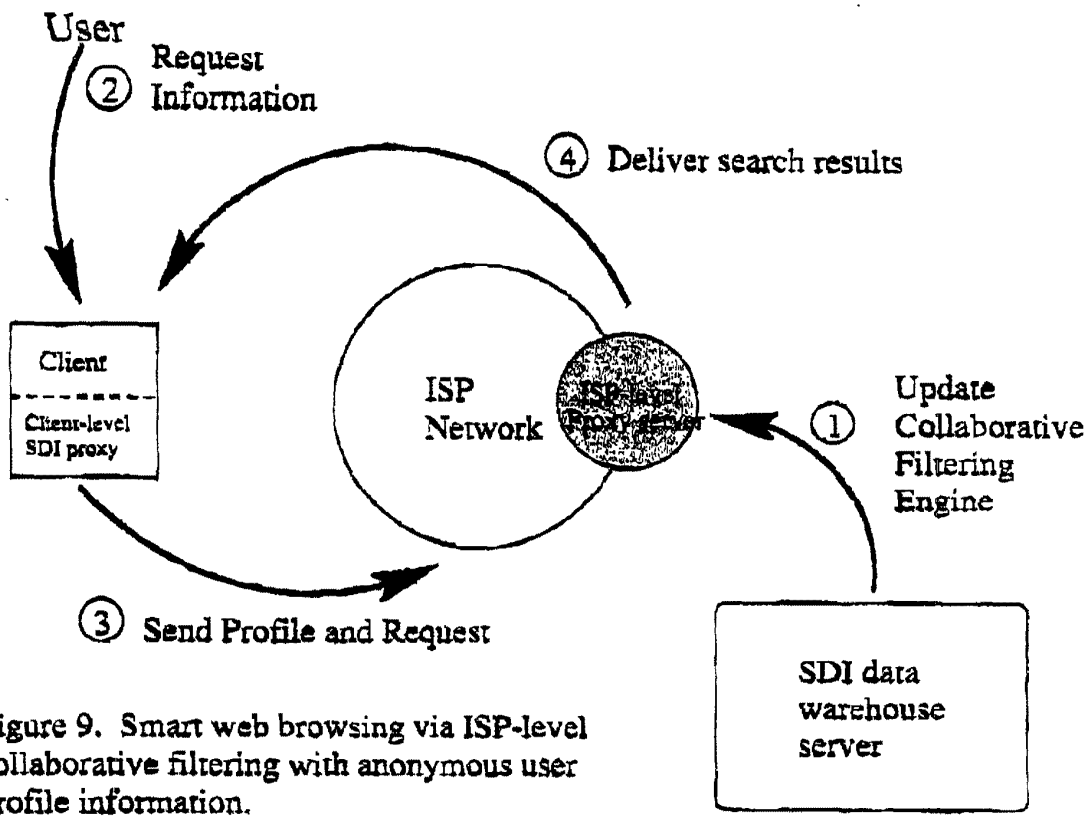


Figure 9. Smart web browsing via ISP-level collaborative filtering with anonymous user profile information.

Google Exhibit 1015 - Page 158

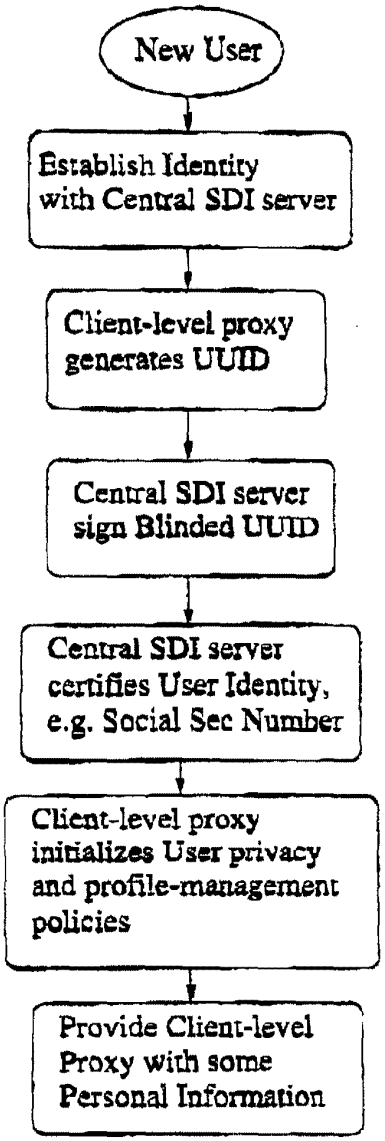


Figure 10. Registering a new user with the Secure Data Interchange.

SDE0054.003

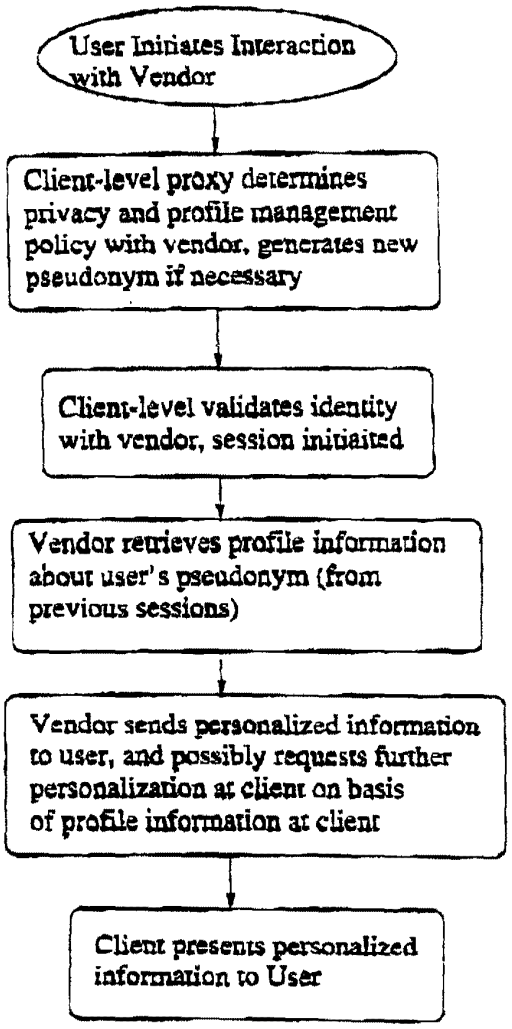


Figure 11. Enabling contact with a new Vendor within a SDI web browsing session.

Confidential - 2005765

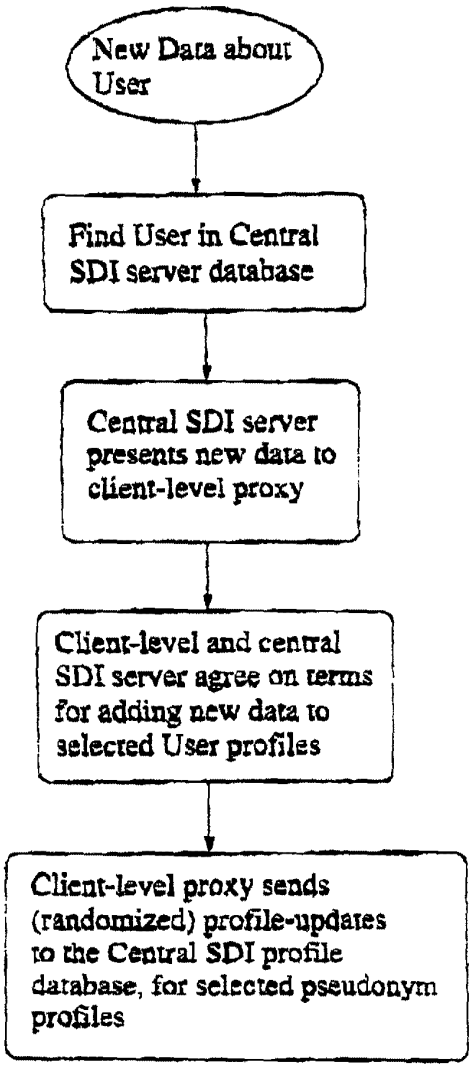


Figure 12. Merging off-line with on-line profile information.

Petitioner - Ex. 1009

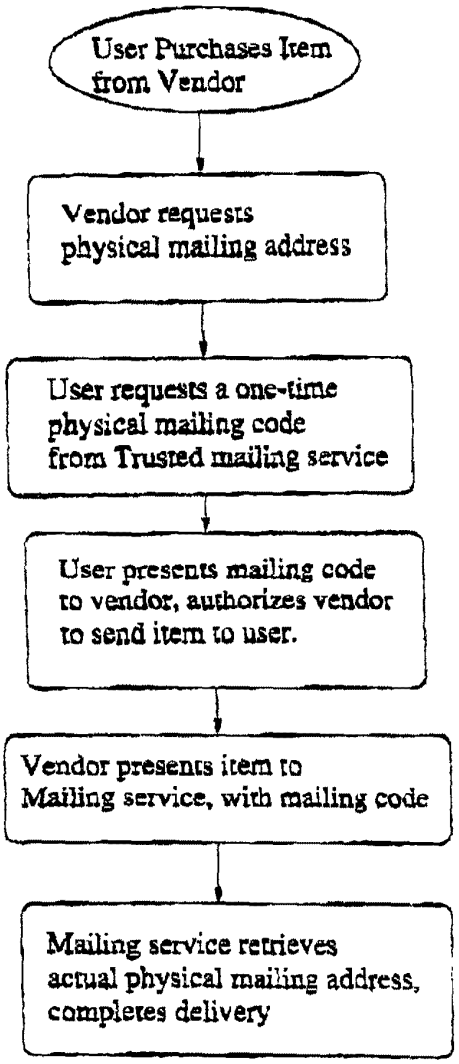


Figure 13. A system for pseudonymous physical mail.

Sensitive - Confidential