

UNITED STATES PROVISINAL PATENT APPLICATION

For

A CYBER SECURITY SYSTEM

By

John Boyer, Carl Salji, Tim Bazalgette, Fraser Greenlee, Dickon Humphrey, Alexander
Johnsen, Marko Marsenic, Philip Sellars, Steven Haworth, Stephen Pickman, Rob
Hutchinson, Matt Dunn, and Jack Stockdale

Prepared by

RUTAN & TUCKER, LLP

18575 Jamboree Road

9TH Floor

Irvine, CA 92612

(650) 320-1500

Attorney Docket No.: 034306-0024PRO

NOTICE OF COPYRIGHT

[0001] A portion of this disclosure contains material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the material subject to copyright protection as it appears in the United States Patent & Trademark Office's patent file or records, but otherwise reserves all copyright rights whatsoever.

FIELD

[0002] Cyber security and in an embodiment use of Artificial Intelligence in cyber security.

BACKGROUND

[0003] Cybersecurity attacks have become a pervasive problem for enterprises as many computing devices and other resources have been subjected to attack and compromised. A “cyberattack” constitutes a threat to security of an enterprise (e.g., enterprise network, one or more computing devices connected to the enterprise network, or the like). As an example, the cyberattack may be a cybersecurity threat against the enterprise network, one or more computing devices connected to the enterprise network, stored or in-flight data accessible over the enterprise network, and/or other enterprise-based resources. This security threat may involve malware (malicious software) introduced into a computing device or into the network. The security threat may originate from an external endpoint or an internal entity (e.g., a negligent or rogue authorized user). The security threats may represent malicious or criminal activity, ranging from theft of credential to even a nation-state attack, where the source initiating or causing the security threat is commonly referred to as a “malicious” source. Conventional cybersecurity products are commonly used to detect and prioritize cybersecurity threats (hereinafter, “cyber threats”) against the enterprise, and to determine preventive and/or remedial actions for the enterprise in response to those cyber threats.

SUMMARY

[0004] Methods, systems, and apparatus are disclosed for an Artificial Intelligence based cyber security system. The cyber security system uses large language models (and the associated generative AI-creating algorithms that can generate new content based on patterns learned from existing data, compositely referred to as 'LLMs' herein) in example aspects as follow. 1) The cyber security system can use the LLM to suggest remediation actions to restore/perform disaster recovery after a cyber attack by an advanced persistent threat. 2) The cyber security system can use an LLM in front of the autonomous response engine/module to mitigate an ongoing cyber-attack with the cyber-attack autonomous response engine/module to make informed remediation decisions. 3) The cyber security system can use an LLM trained to act as a first-line support for client queries about either 1) how the different cyber security products and services work and troubleshooting issues with those products and services as well as provide 2) cyber analyst answers regarding cyber security techniques and current cyber threat landscapes and current active cyber threats. 4) The cyber security system can provide the LLM with an API specification so that the LLM can understand how to make API requests and translate questions into API queries into the different components of the cyber security system to explain the AI's decision process. 5) The cyber security system can provide the LLM with a list of all breaches and their triggers and then ask the LLM to provide a summary and prioritization of the breaches based on the parameters known about those breaches. 6) The cyber security system can train the LLM on the search syntax for an Advanced Search function offered by the cyber security system to allow users to query for detailed log data in a friendly human-speech way. 7) The cyber security system can feed an LLM the history of cyber security breaches/models breaches and their severity scores as well as the current cyber threats trending currently and have the LLM produce both a summary of these breaches and the trends as well as to provide recommendations to help a user of a cyber security appliance to improve their system in light of the current model breaches and cyber threat trends. 8) The cyber security system can train a LLM to generate software code to generate code that creates data visualizations, such as graphs, charts, etc. to show cyber security breaches, user's activity, and current cyber threat trends. 9) The cyber security system can have the LLM create JSON documents and complex filters. 10) The cyber security system can use LLMs to enhance cyber

security measures by simulating attack scenarios and facilitating customized training. 11) The cyber security system can use LLMs to act as assistants that submit data to third-party services using pre-engineered prompts on how to integrate with that third party. 12) The cyber security system can use an LLM to train on web responses and certificate information, and use the resulting embeddings to search for similar services. 13) The cyber security system can make use of LLMs to make intelligent user combination across platforms. 14) The cyber security system can use of different LLMs and human language identification for automatic translation of content within the user interface, enabling localization. 15) The cyber security system can address generative AI concerns with the introduction of AI Models that help protect data privacy and Intellectual Property. 16) The cyber security system can incorporate additional LLM and generative AI uses into the components of the cyber security system. 17) The cyber security system can have the LLM perform many additional tasks as discussed herein.

[0005] These and other features of the design provided herein can be better understood with reference to the drawings, description, and claims, all of which form the disclosure of this patent application.

DRAWINGS

[0006] The drawings refer to some embodiments of the design provided herein in which:

[0007] Figure 1 illustrates an embodiment of a block diagram of an example Artificial Intelligence based cyber security appliance.

[0008] Figure 2 illustrates an embodiment of the AI based cyber security appliance plugging in as an appliance platform to protect a system.

[0009] Figure 3 illustrates a graph of an embodiment of an example chain of unusual behavior for, in this example, the email activities as well as IT activities deviating from a normal pattern of life for this user and/or device in connection with the rest of the network under analysis.

[0010] Figure 4 illustrates an example Artificial Intelligence based cyber security system using a cyber threat analyst module to protect an example network.

[0011] Figure 5 illustrates a block diagram of an embodiment of one or more computing devices that can be a part of the Artificial Intelligence based cyber security system for an embodiment of the current design discussed herein.

[0012] Figure 6 illustrates a block diagram of an embodiment of the prediction engine with Artificial Intelligence-based simulations conducted in the prediction engine by constructing a graph of nodes of the system being protected (e.g. a network including i) the physical devices connecting to the network, any virtualize machines of the network, user accounts in the network, email accounts in the network, etc. as well as ii) connections and pathways through the network) to create a simulated version of the system to be tested.

[0013] Figure 7 illustrates a diagram of an embodiment of a cyber threat prediction engine and its Artificial Intelligence-based simulations constructing a graph of nodes in an example network and simulating how the cyberattack might likely progress in the future tailored with an innate understanding of a normal behavior of the nodes in the system being protected and a current operational state of each node in the graph of the protected system during simulations of cyberattacks.

[0014] Figure 8 illustrates a diagram of an embodiment of the cyber threat prediction engine and its Artificial Intelligence-based simulations constructing an example graph of nodes in an example network and simulating how the cyberattack might likely progress in the future tailored with an innate understanding of a normal behavior of the nodes in the system being protected and a current operational state of each node in the graph of the protected system during simulations of cyberattacks.

[0015] While the design is subject to various modifications, equivalents, and alternative forms, specific embodiments thereof have been shown by way of example in the drawings and will now be described in detail. It should be understood that the design is not limited to the particular embodiments disclosed, but – on the contrary – the intention is to cover all modifications, equivalents, and alternative forms using the specific embodiments.

DESCRIPTION

[0016] In the following description, numerous specific details are set forth, such as examples of specific data signals, named components, number of servers in a system, etc., in order to provide a thorough understanding of the present design. It will be apparent, however, to one of ordinary skill in the art that the present design can be practiced without these specific details. In other instances, well known components or methods have not been described in detail but rather in a block diagram in order to avoid unnecessarily obscuring the present design. Further, specific numeric references such as a first server, can be made. However, the specific numeric reference should not be interpreted as a literal sequential order but rather interpreted that the first server is different than a second server. Thus, the specific details set forth are merely exemplary. Also, the features implemented in one embodiment may be implemented in another embodiment where logically possible. The specific details can be varied from and still be contemplated to be within the spirit and scope of the present design. The term coupled is defined as meaning connected either directly to the component or indirectly to the component through another component.

[0017] The cyber security system can include several components configured to communicate and cooperate with each other including the cyber security appliance with its cyber threat detect engine, the cyber threat autonomous response engine, the cyber-attack prediction simulator, the cyber-attack restoration engine, the artificial intelligence-based cyber security analyst, and other components. Large language models (LLMs) that can understand and generate human-like language and the associated generative AI-creating algorithms that can generate new content based on patterns learned from existing data are used to enhance cyber security measures across various aspects of the field. The Large language models (LLMs) can be AI algorithms that have been trained on a large amount of text-based data, typically scraped from the open internet such as webpages and sources such as scientific research, books, forums or social media posts. The Generative AI can be an AI system that's primary function is to generate new content (e.g., images, music, speech, code, video, text, etc.). Large language models (e.g., GPT-3, PaLM, LLaMA) can be the underlying technology behind many powerful generative AI systems today (e.g., ChatGPT, Bard). The cyber security system uses large language models (and the associated generative AI-creating algorithms that can generate new

content based on patterns learned from existing data, compositely referred to as 'LLMs' herein) in example aspects as follow. 1) The cyber security system can use the LLM to suggest remediation actions to restore/perform disaster recovery after a cyber attack by an advanced persistent threat. 2) The cyber security system can use an LLM in front of the autonomous response engine/module to mitigate an ongoing cyber-attack with the cyber-attack autonomous response engine/module to make informed remediation decisions. 3) The cyber security system can use an LLM trained to act as a first-line support for client queries about either 1) how the different cyber security products and services work and troubleshooting issues with those products and services as well as provide 2) cyber analyst answers regarding cyber security techniques and current cyber threat landscapes and current active cyber threats. 4) The cyber security system can provide the LLM with an API specification so that the LLM can understand how to make API requests and translate questions into API queries into the different components of the cyber security system to explain the AI's decision process. 5) The cyber security system can provide the LLM with a list of all breaches and their triggers and then ask the LLM to provide a summary and prioritization of the breaches based on the parameters known about those breaches. 6) The cyber security system can train the LLM on the search syntax for an Advanced Search function offered by the cyber security system to allow users to query for detailed log data in a friendly human-speech way. 7) The cyber security system can feed an LLM the history of cyber security breaches/models breaches and their severity scores as well as the current cyber threats trending currently and have the LLM produce both a summary of these breaches and the trends as well as to provide recommendations to help a user of a cyber security appliance to improve their system in light of the current model breaches and cyber threat trends. 8) The cyber security system can train a LLM to generate software code to generate code that creates data visualizations, such as graphs, charts, etc. to show cyber security breaches, user's activity, and current cyber threat trends. 9) The cyber security system can have the LLM create JSON documents and complex filters. 10) The cyber security system can use LLMs to enhance cyber security measures by simulating attack scenarios and facilitating customized training. 11) The cyber security system can use LLMs to act as assistants that submit data to third-party services using pre-engineered prompts on how to integrate

with that third party. 12) The cyber security system can use an LLM to train on web responses and certificate information, and use the resulting embeddings to search for similar services. 13) The cyber security system can make use of LLMs to make intelligent user combination across platforms. 14) The cyber security system can use of different LLMs and human language identification for automatic translation of content within the user interface, enabling localization. 15) The cyber security system can address generative AI concerns with the introduction of AI Models that help protect data privacy and Intellectual Property. 16) The cyber security system can incorporate additional LLM and generative AI uses into the components of the cyber security system. 17) The cyber security system can have the LLM perform many additional tasks as discussed herein. Thus, for example, the LLMs can be utilized to suggest remediation actions in the aftermath of cyber-attacks, aid in mitigating ongoing attacks by making informed decisions, act as first-line support for client queries, explain AI decision processes in cyber security platforms, summarize and prioritize breaches, assist in querying detailed log data, provide recommendations based on breach summaries and threat trends, generate code for data visualizations; create JSON documents with complex filters. By leveraging the capabilities of LLMs, organizations can strengthen their cyber security defenses and improve response mechanisms to counter advanced persistent threats effectively.

[0018] Thus, LLMs can be employed in conjunction with autonomous response systems equipped with a Respond engine, enabling informed decision-making during an ongoing cyber-attack. This integration empowers organizations to swiftly mitigate the attack by leveraging the LLM's knowledge and expertise. Moreover, by training an LLM to act as a first line of support, clients can receive assistance in understanding cybersecurity products and troubleshooting issues they may encounter. Additionally, the LLM can provide valuable cyber analyst answers, shedding light on various cyber security techniques and the current threat landscape. Furthermore, organizations can harness the capabilities of LLMs by providing them with API specifications, enabling the LLM to understand how to make API requests and translate user questions into API queries within an AI-based cyber security platform. This functionality helps explain the decision-making process of the AI system, enhancing transparency and trust. LLMs can also assist

in analyzing breaches and their triggers by providing a summary and prioritization of these incidents based on available parameters. This aids in effective breach response and resource allocation. Additionally, training an LLM on the search syntax of tools such as the cyber security appliance Advanced Search enables users to query for detailed log data in a human-friendly manner, simplifying the process of extracting relevant information for investigations. By feeding an LLM with the history of cyber security breaches and model breaches, along with severity scores and current threat trends, the LLM can generate summaries and recommendations for users of cyber security appliances. This information empowers users to proactively improve their system's security posture in light of prevailing model breaches and emerging threat trends. Furthermore, LLMs can be trained to generate software code that creates data visualizations, such as graphs and charts, showcasing cyber security breaches, user activity, and current cyber threat trends. This capability simplifies the presentation of complex data, enabling stakeholders to grasp key insights quickly. Finally, LLMs can be leveraged to create JSON documents with complex filters, facilitating advanced data processing and analysis in cyber security workflows. In conclusion, through the utilization of LLMs, organizations can enhance their cyber security measures, effectively defend against advanced threats, and strengthen their overall resilience in the face of evolving cyber threats.

[0019] The advent of large language models (LLMs) has opened up new possibilities in bolstering cyber security practices.

[0020] The cyber security system can use the LLM to Suggest Remediation Actions to restore/perform disaster recovery after a cyber attack by an advanced persistent threat.

[0021] One such application lies in utilizing LLMs to suggest remediation actions for restoring and performing disaster recovery after cyber-attacks orchestrated by advanced persistent threats (APTs). By analyzing the attack vectors and understanding the intricacies of the breach, LLMs can provide valuable insights and guidance in recovering critical systems. For example, the cyber security system can use an LLM to recommend Cloud remediation steps.

[0022] The cyber security appliance, its cloud module, its SaaS module, and its autonomous response module can cooperate with the LLM.

[0023] For example, the large language model can be used to recommend remediations for misconfigurations in a user's cloud account by asking questions such as:

[0024] "I have an S3 bucket call production-sales-files-1' that is open to the public, please give me steps to resolve this as well as an AWS-cli command and references."

[0025] This same process could be applied to almost any 'Alert' within the cyber security system where the system provides the context and the recommended questions. As another example: A SaaS User in the platform "Box" has been seen logging in from an unusual IP address expected to be malicious, recommend steps to defend against this in the quickest possible way as well as references and hardening practices for the future.

[0026] Thus, the cyber security system can use the LLM to Suggest Remediation Actions:

[0027] After a cyber-attack by an advanced persistent threat, an organization can leverage an LLM to suggest remediation actions. For example, the LLM can analyze the attack's characteristics, identify compromised systems or vulnerabilities, and provide recommendations on patching, system hardening, and security policy updates. It can offer specific steps to restore affected services and mitigate the risk of future attacks. By utilizing the vast knowledge base of an LLM, organizations can efficiently recover from cyber-attacks and strengthen their security posture.

[0028] The cyber security system can use an LLM to recommend Cloud remediation steps.

[0029] The cyber security system can use an LLM in front of the Autonomous Response engine/module. Integrating an LLM in front of an autonomous response system equipped with a RESPOND engine enables effective decision-making during an ongoing cyber-attack. The LLM can analyze real-time data from various sources, such as network logs, intrusion detection systems, and threat intelligence feeds. Based on this information, the LLM can provide recommendations on containment measures, threat hunting strategies, or adjusting security controls. This collaborative approach empowers the autonomous

response system to make informed decisions, accelerating the response time and minimizing the impact of the attack.

[0030] As part of the cyber security Cloud Security, Using an LLM in front of our autonomous response to mitigate an ongoing cyber-attack with a RESPOND engine to make informed remediation decisions.

[0031] The cyber security system can use a large language model to make intelligent decisions by using the vast amounts of context we can provide. The more context we provide the better the decision. The more information about how the actions will work the better the choice.

[0032] For example, the following prompt:

[0033] The cyber security system has seen a login from a rare IP in Russia as well as regular events from Cambridge given the following possible actions <SET OF ACTIONS> which would be the best one to take. Explain why you made that decision.

[0034] The cyber security system can also pre-train the LLM on examples where malicious activity was successfully ended by RESPOND actions, or on playbooks that exist for incident response, to assist in the decision making and recommendation.

[0035] 1. **Intelligent documentation and Ask the Expert lookups.**

[0036] LLM as First Line Support for Client Queries:

[0037] Training an LLM to act as the first line of support allows clients to receive assistance in understanding cybersecurity products and troubleshooting issues. For instance, clients can inquire about the functionality and configuration of security software, seek guidance on implementing best practices, or troubleshoot common issues. The LLM can provide step-by-step instructions, explain complex concepts in a user-friendly manner, and offer real-time support. This helps alleviate the burden on human support personnel and enables clients to access timely and accurate information.

[0038] the cyber security system can use an LLM trained to act as a first-line support for client queries about either 1) how the different cyber security products and services work and troubleshooting issues with those products and services as well as provide 2) cyber

analyst answers regarding cyber security techniques and current cyber threat landscapes and current active cyber threats.

[0039] An LLM is trained on both the entire the cyber security documentation corpus, on all interactions between the cyber security support and clients (tickets) and all interactions via the “Ask the Expert” service.

[0040] The LLM can then act as a first-line support for client queries about the cyber security, and also about how to investigate or triage threats in their network.

[0041] The LLM can also be used to re-write our own analyst and support responses intelligently to add context before posting (almost like an assistant for our own supporters/analysts).

[0042] 2. Intelligent Search and API

[0043] LLM Explaining AI Decision Process in Cyber Security Platforms:

[0044] By providing an LLM with an API specification, organizations can enhance the transparency of AI-based cybersecurity platforms. The LLM can interpret user queries, translate them into API queries, and then explain how the AI system arrives at its decisions. This functionality enables users to understand the reasoning behind the AI's recommendations or actions. It also helps build trust and confidence in the AI system's capabilities, as users can gain insights into the decision-making process and validate the system's outputs.

[0045] By providing the LLM with an API specification an LLM can understand how to make API requests and translate questions into API queries.

[0046] The user interfaces into the different components of the cyber security system can use the AI to enhance how customers can interact with these products in natural language. This also allows for a natural language interaction with the cyber security product/platform. For example: Given our API docs find me all devices that communicated over TCP in the last 4 days.

[0047] Intelligent Summary of Breaches

[0048] Summarizing and Prioritizing Breaches:

[0049] When provided with a list of breaches and their triggers, an LLM can generate a summary and prioritize them based on specified parameters. For example, the LLM can consider the severity of the breach, the impact on critical systems or data, and the potential risk to the organization. By prioritizing breaches, organizations can allocate their resources more effectively, focusing on addressing the most critical incidents first. The LLM's analysis assists in streamlining breach response efforts and minimizing further damage.

[0050] Summarizing and Prioritizing Breaches:

[0051] When provided with a list of breaches and their triggers, an LLM can generate a summary and prioritize them based on specified parameters. For example, the LLM can consider the severity of the breach, the impact on critical systems or data, and the potential risk to the organization. By prioritizing breaches, organizations can allocate their resources more effectively, focusing on addressing the most critical incidents first. The LLM's analysis assists in streamlining breach response efforts and minimizing further damage.

[0052] Provide the LLM with a list of all breaches and their triggers then ask the LLM to provide a summary and prioritization of the breaches based on the parameters known about those breaches. E.g.:

[0053] Given I have 40 Breaches of severity 60 for OneDrive Sync events that are new for the user, for users x, y, z at times a, b, c, etc. 20 For Large network traffic uploads for devices j, k, l, etc. etc. Provide me with a summary, any possible links between breaches, and also prioritize the order in which I should triage them.

[0054] The LLM would be trained on the cyber security API and be aware of where information can be found, what kind of data each endpoint returns, and the data the cyber security AI Analyst considers relevant for similar activity.

[0055] Using an LLM to create Health checks for the cyber security deployment.

[0056] The cyber security system can provide the LLM with a list of all statuses from cyber security modules and components (via APIs) and ask it to summarize the health of the

deployment. On average we see traffic at a rate of 20000 events/s, then for example, 1 out of 20 SaaS Modules have an error status and have been in that state for 2 weeks.

[0057] Summaries the health of my deployment and what I can do to resolve that.

[0058] Natural Language Interface for the cyber security Advanced Search

[0059] The cyber security system can query detailed log data in a human-speech way. Training an LLM on the search syntax for tools like the cyber security Advanced Search allows users to query detailed log data in a more user-friendly manner. Instead of learning complex search syntax, users can express their queries in natural language, making log analysis more accessible and efficient. The LLM understands the intent behind the query and translates it into the appropriate search syntax, simplifying the process of extracting relevant log information for investigations or threat hunting.

[0060] An LLM approach is trained on the search syntax for the cyber security Advanced Search, allowing users to query for detailed log data in a friendly human-speech way. This would be combined with training on the usual “questions” users ask – i.e., searches – to allow the interface to suggest the next steps in a friendly way.

[0061] Producing summary reports and recommendations

[0062] The cyber security system can provide breach summaries and recommendations. By feeding an LLM with historical cyber security breaches, severity scores, and current threat trends, organizations can generate breach summaries and receive recommendations. The LLM can analyze patterns, identify common vulnerabilities, or attack vectors, and provide actionable insights to improve system security. These recommendations can include implementing specific security controls, conducting vulnerability assessments, or enhancing user awareness through targeted training. The LLM's analysis empowers organizations to proactively strengthen their security defenses.

[0063] Feed an LLM the history of cyber security breaches/models breaches and their severity scores as well as current cyber threats trending currently (e.g., model breaches/respond/AI analyst) and have the LLM produce both a summary of these breaches and trends as well as to provide recommendations to help a user of a cyber

security appliance to improve their system in light of the current model breaches and cyber threat trends. E.g., example prompts:

[0064] Given that this Month we have seen the following autonomous response actions: [block IP 1, block IP 2, block user 3, block user 4] produce me some security recommendations.

[0065] An example might be to fully block certain IP ranges and conditional access policies to the users.

[0066] Summarize John Boy's office365 SaaS events for the last week and decide if his current permissions set which is [X Y Z] is appropriate or whether they can be reduced.

[0067] Produce customized data visualizations.

[0068] The cyber security system can generate code for data visualizations. Training an LLM to generate software code for creating data visualizations related to cyber security breaches, user activity, and threat trends can automate the process of generating informative visuals. The LLM can interpret the desired visualization requirements and generate code snippets in programming languages like Python or JavaScript. This capability simplifies the creation of graphs, charts, or interactive dashboards, enabling security analysts and stakeholders to visualize and comprehend complex data in a more intuitive manner.

[0069] LLMs can retrieve data and write code, given a good enough specification. For an LLM pre-trained on the cyber security API, it can create (and implement) code to produce visualization based upon a natural language prompt. For example:

[0070] Produce me a bar graph that breaks down a user's activity over the last week, including sub-sections of the column breaking down the source IP.

[0071] 1. Creating Custom Modules, Log Ingestion, and Models with an LLM

[0072] Creating JSON Documents with Complex Filters:

[0073] An LLM can be leveraged to create JSON documents with complex filters, facilitating advanced data processing and analysis in cyber security workflows. For instance, the LLM can generate filters based on specific criteria, such as timeframes,

source IP addresses, or specific attack signatures. These JSON documents can then be used to filter and extract relevant data from large data sets or for further analysis using other tools or platforms. The LLM's ability to generate complex filters simplifies the data processing pipeline and enhances the efficiency of cyber security operations.

[0074] LLMs are very capable of creating JSON documents and complex filters if you have a well-defined schema, they can even infer fields you haven't supplied.

[0075] **Models**

[0076] The cyber security system has trained an LLM to create models for the cyber security Cloud Security product. The cyber security system pre-prompts the LLM with training material on how the model is formed, the meaning of each field, how to construct Boolean logic, and some worked examples.

[0077] After, it can demonstrably create new examples given a natural language description of the type of activity we wish to target.

[0078] **Log Ingestion**

[0079] The cyber security telemetry data uses GROK patterns to parse fields. Pre-prompted with a defined spec, example parsing templates, and example pattern and log pairs, the LLM can produce new templates when given example log entries.

[0080] **Modules**

[0081] The format for event parsing and classification for SaaS/Cloud security modules is relatively standard. The classification of events into buckets such as "Admin", "login", and "File Created" is currently done manually. Third-party platforms also often provide detailed API specs of their available endpoints.

[0082] LLMs can therefore be trained on existing module structures and event classification to create new modules, given the API spec and – once authenticated - examples of events.

[0083] Additional ways large language models (LLMs) enhance cyber security measures across various aspects of the cyber security field.

[0084] Attack Simulation and Training Applications: The cyber security system can use LLMs to enhance cyber security measures by simulating attack scenarios and facilitating customized training. For example, LLMs can be trained on a user's style of email writing and formatting to generate plausible phishing emails in their "voice." This approach can be utilized in PREVENT/E2E engagements to create realistic phishing simulations and raise user awareness about potential threats. Additionally, LLMs can be employed to detect anomalies in email communication by comparing the actual emails with those predicted to be typical for a specific user. This comparison helps identify potentially malicious emails that deviate from the user's usual patterns.

[0085] Style transfer for emails to generate plausible phishing.

[0086] Train an LLM on a user's style of email writing (including the underlying formatting) and use it to generate convincing phishing emails in their "voice" as part of PREVENT/E2E engagements.

[0087] Similarly, this can be applied as a detection approach – i.e., how much does this actual email differ from that which would be "predictable" for this user.

[0088] Generative use of a message passing neural network to simulate attacks.

[0089] This form of the neural net is specialized to create graph structures. Therefore, it can be successfully trained on the graph structure of AI Analyst incidents in a customer environment and then used to generate convincing synthetic incidents for the purpose of simulating incident response, tailored to the client environment.

[0090] Using Auto GPT like LLMs to simulate attack scenarios.

[0091] Another application involves using message passing neural networks specialized in creating graph structures. By training these neural networks on the graph structure of AI Analyst incidents in a customer environment, synthetic incidents can be generated. These synthetic incidents simulate real-world attack scenarios tailored to the client's environment, enabling organizations to simulate incident response and test the effectiveness of their security measures.

[0092] Furthermore, LLMs like Auto GPT can be leveraged to simulate attack scenarios by working together and feeding back into themselves. The LLMs can work together and

feed back into themselves to perform long complex tasks, as such, they can simulate a cyber-attacker in an environment. For example, an LLM agent placed on an isolated AWS EC2 instance can execute system commands to determine the operating system, make curl requests to check connectivity, and even attempt to exfiltrate information. The LLM agent on the ec2 instance follows the situation to exfiltrate key information and then as much information to a certain endpoint. The LLM agent on the ec2 instance uses system commands to work out what OS it is on, curl requests to check connectivity, if it knows it is on an AWS ec2 instance might make requests to IMDS for credentials, etc.

[0093] This simulation provides organizations with insights into potential attack vectors and helps in refining their defensive strategies.

[0094] Intelligent Customized Training Scenarios: The cyber security system can use LLMs to generate long-form content customized to meet the specific needs of customers. By providing relevant information, such as connections blocked for suspicious websites or unusual file downloads, an LLM can create training plans and materials tailored to individual users. These materials can help users become more security aware and educate them about potential risks and best practices.

[0095] Thus, the LLMs can take prompts of what the system already knows about the organization and its users to produce long form content that is customized to be relevant to our customers' needs. The more info we can provide, the better the content will be. Customers can review the material before sending. For example, we can feed the LLM data from our modeling and RESPOND, e.g. Given that these users had connections blocked for dodgy websites, these users received the following bad emails, these users had unusual file downloads produce me a training plan and training materials for each user that I can send to help them become more security aware.

[0096] LLM-Driven Assistants for Third-Party Service Integration:

[0097] Various kinds of LLM-driven “assistants” which submit e.g., AIA data to 3rd party services using a pre-engineered prompt.

[0098] LLMs can act as assistants that submit data to third-party services using pre-engineered prompts. For instance, an LLM can be prompted to query OSINT (Open-

Source Intelligence) tools or a customer's threat intelligence tools by connecting to their APIs. Thus, the LLM is pre-prompted to query services like OSINT tools or the customer's own datalakes / threat intel tools, using an understanding of how to connect to their APIs. This enables, for example, the LLM-driven assistant to provide information by checking AI Analyst data against services like Microsoft Defender or obtaining insights from tools like VirusTotal. The chatbot/assistant implementation can then be asked things like "Check this AI Analyst information against Microsoft Defender?" or "What does VirusTotal think about this endpoint?"

[0099] Training an LLM on web responses and certificate information, and using the resulting embeddings to search for similar services.

[0100] The LLM is trained on certificate information and responses (almost akin to a JA3 or similar attempts to "fingerprint"), then is fed data from the wider web to try and find examples of the same service it was trained on (i.e., that fit its predication).

[0101] The cyber security system can make use of intelligent user combinations across platforms (e.g., combining users with similar usernames or presences across different platforms). LLMs can assist in this process by considering contextual data, such as roles, activity times, standard behaviors, and clustered peers. By providing all available information about both users, the LLM can assess whether it is appropriate to combine them or not. This capability extends to devices by considering factors such as hostnames, active times, and locations. LLMs have the advantage of being able to process vast amounts of time-series data and make informed decisions, which can then be corrected by human input if necessary.

[0102] The system can combine users with similar usernames or presences back into a single user, when appropriate, without a human indicating they are the same. For example, is john@slammar the same user as john.b@slimmer, Or are they two users with the same initials?

[0103] In an example, the system feeds the LLM all contextual data available about both users (including roles, activity times, standard behaviors, clustered peers, etc.) and then let the LLM assess whether it would be a good idea to combine them or not.

[0104] This can be extended to devices by considering hostnames, active times, and locations - the power comes from it being able to use any information (e.g., vast amounts of time series data a human cannot comprehend) and decide which it cares about. If it is wrong, then a user can just tell it so and correct the output.

[0105] Full Translations within the Platform: LLMs can be employed for the automatic translation of content within the user interface, enabling localization. This feature is particularly useful for user comments, email content, or other non-hard-coded text. For example, an LLM can be used to translate emails, button texts, team messages, event types, or analyst comments into various languages, enhancing accessibility and facilitating effective communication across language barriers. Thus, for example:

[0106] 1) Translate this email so that my Spanish team can work on it;

[0107] 2) Translate all the button texts to French;

[0108] 3) Translate this Teams message to Spanish;

[0109] 4) Translate all the SaaS event types; and

[0110] 5) Translate this analyst's comment into my language.

[0111] By leveraging these additional capabilities of LLMs, organizations can enhance their cyber security measures, improve user training, simulate realistic attack scenarios, integrate with third-party services, facilitate user data analysis, and enable multilingual functionality within their platforms.

[0112] 15) The cyber security system can address generative AI concerns with the introduction of AI Models that help protect data privacy and Intellectual Property. The cyber security system can address generative AI Concerns with the introduction of AI Models that help protect data privacy and Intellectual Property. Networks protected by cyber security appliances can have employees/users of that network using third party generative AI platforms. The cyber security system can have risk and compliance machine learning models to help CISOs as they balance the opportunity and risk of inadvertent IP loss and data leakage from the use of generative AI and LLM-based tools. In response to growing use of generative AI tools, the cyber security system provides these risk and compliance machine learning models address the increasing risk of IP loss

and data leakage. These new risk and compliance models for the Detect engine and the autonomous Response engine/module make it easier for customers to put guardrails in place to monitor, and when necessary, respond to activity and connections to generative AI and large language model (LLM) tools. The cyber security system can, for example, detect and prevent an upload of data into the example third party generative AI tools.

[0113] New generative AI tools promise increases in productivity and new ways of augmenting human creativity. CISOs must balance the desire to embrace these innovations to boost productivity while managing risk. A corporation or other entities can supply guidance using generative AI tools and other LLMs in the workplace on how to make the most of AI without exacerbating its potential dangers.

[0114] Each enterprise will determine their own appetite for the opportunities versus the risk. The cyber security appliance can provide security personalized to an organization, and user's/CISOs can supply input to leverage the components in the cyber security system to enforce their specific compliance policies. The cyber security system can provide components with the ability to quickly understand and control the use of these AI tools within their organizations. The cyber security system can provide components to counter malicious actors who create cyber threats using these publicly available AI tools to create nuanced and effective cyber-attacks.

[0115] To complement the cyber security system's core self-learning AI for attack prevention, threat detection, autonomous response, and policy enforcement, the cyber security system has new AI models and capabilities, including its own proprietary large language models, to help customers prepare for and fight back against increasingly sophisticated threats. These models are used across the products in the cyber security's Cyber AI Loop. The Cyber AI Loop provides a continuous end-to-end security capability that can autonomously prevent, detect, and respond to novel, in-progress threats in real time. The cyber security system uses generative AI and LLMs in the arsenal of AI techniques. But the cyber security system uses of generative AI and LLMs are not one-size-fits-all and are applied with guardrails to the right use cases and challenges.

[0116] The cyber security system uses the LLM to prevent example risky uses of generative AI. Some example risky uses of generative AI that could be detected by the

cyber security appliance include as follows. The cyber security system can detect sensitive information in queries submitted to public LLMs. As more organizations produce LLMs, queries stored online or materials uploaded may be hacked, leaked, or accidentally made publicly accessible. This could include potentially user-identifiable or sensitive business information. The cyber security system can detect sensitive information in data to a generative AI tool protected by that deployment. The cyber security system can detect sensitive information in internal code or other material marked as confidential information.

[0117] Some example Risk & Compliance Models can include risk and compliance models used by for example the Detect engine and the Respond engine.

[0118] The risk and compliance models are trained to detect the use of generative AI tools and enable enforcement of policies. In an example, a human security team can use the model editor in the cyber security appliance to specifically detect and respond to generative AI and LLM tools including ChatGPT and others. The new models have analyzed and modeled different generative AI and LLM tools including ChatGPT, GPT-3, Galactica, ERNIE, EXAMPLE and EXAMPLE, etc., to make this even easier.

[0119] The cyber threat detect engine, the cyber threat autonomous response engine, the cyber-attack prediction simulator, the cyber-attack restoration engine, the artificial intelligence-based cyber security analyst can continuously monitor for and autonomously respond to (i.e. block or alert security teams) activity from employees accessing or sending specific types of data to these tools, which will help security teams who are restricting certain use or requiring permissions for employees to use these tools.

[0120] 16) The cyber security system can incorporate additional LLM and generative AI uses into the components of the cyber security system.

[0121] The artificial intelligence-based cyber security analyst can use a transformer-based large language model classifier to assist the cyber threat detect engine, the cyber threat autonomous response engine, the cyber-attack prediction simulator, and the cyber-attack restoration engine to detect and mitigate the cyber threat. The transformer-based large language model classifier can categorize malicious communications based on textual properties. This is used in analysis of natural language content, such as phishing

links, to provide context to inoculation hits. The transformer-based large language model classifier can be trained specifically on security data, as part of the core functionality. To further powerful threat investigation and analysis capabilities, the artificial intelligence-based cyber security analyst now uses an enhanced transformer-based LLM that has been trained on security network and engineering data so that it can better identify anomalous behavior, services and endpoints. This also enables the artificial intelligence-based cyber security analyst to provide even more context to human analysts. Also, the cyber-attack prediction simulator can perform attack path validation with the use of LLM generated attacks in addition to other Natural Language processed derived attacks. The cyber-attack prediction simulator can perform intelligent APT targeting analysis, reporting, and refreshed CVE visualization. Thus, the Attack Path Validation portion can bring LLM-generated attack capabilities alongside existing NLP-derived attack engagements. This allows the cyber-attack prediction simulator to emulate attacks in a wider range of sophistication and complexity to meet customer needs. For example, the LLM generated attack capabilities were added to cyber-attack prediction simulator because in some scenarios the attacks emulated using the NLP models were actually too convincing/sophisticated and customers/users wanted there to be a way to have more basic/recognizable fake cyber-attacks (which is what the LLM-generated models are better at - they meet the need for there to be slightly more generalized, less personalized scenarios). Thus, the cyber-attack prediction simulator can use generative AI to, for example, enhance the creativity, relevancy, and precision of simulated phishing attacks for security training.

[0122] The user interface to the cyber threat autonomous response engine presents options on a Quick Setup process to add a greater visibility over the autonomous state of the Respond engine, with accompanying alerts.

[0123] Next, the cyber security system can train those LLMs with factors such as (what data are they trained on, how was the data selected or cleaned of offensive, inaccurate, or biased data, how is it safeguarded, etc. The cyber security system uses a variety of different data sources and techniques to train its AI models. For example, the enhanced transformer-based LLM classifier can be trained on inoculation data (which is anonymized breaches and methods submitted by customers) as well as additional research data. The

artificial intelligence-based cyber security analyst includes models trained on millions of interactions between human cyber security expert analysts and the other components making up the cyber security system so that it can emulate human thought processes and continuously investigate cyber-threats behind the scenes at machine speed. This allows a diverse approach so that there is no one-size-fits-all approach when it comes to the AI approaches.

[0124] The cyber security system uses a variety of different types of AI across the components in this system, that includes unique self-learning AI models that understand the nuance of each customers' particular network composition and business operations as well as generative AI, large language models, natural language processing models, supervised learning models, and much more.

[0125] More on the artificial intelligence-based cyber security analyst

[0126] The artificial intelligence-based cyber security analyst can use generative AI to augment threat analysis and investigations to handle threat investigations at machine speed and scale. The artificial intelligence-based cyber security analyst is trained on millions of interactions between the human cyber security analysts and the components of the cyber security system so that it can emulate human thought processes and continuously investigate cyber-threats behind the scenes. The artificial intelligence-based cyber security analyst brings forward the highest priority threats for human investigation, providing explanations and context so that your teams have access to relevant details, device information and dates from the start and don't waste time querying the system looking into low priority incidents. This is made possible by many different types of AI, including self-learning AI that understands your unique business patterns of life, as well as other types of AI models including large language models that bring more context to the investigations. In general, the components of the cyber security system use different types of AI. The components of the cyber security system apply the right type of AI to the right use cases, and use many different types of AI to address the multitude of challenges facing cyber security teams today.

[0127] The artificial intelligence-based cyber security analyst can use a large AI language model that can understand human language, language like data, and parameters to make

an analysis and can then generate text or take other action based on the analysis. The artificial intelligence-based cyber security analyst and/or cyber threat detect engine can also have some integrations with other generative AI tools such as ChatGPT, including a tool which will allow customers to monitor whether their employees are accessing ChatGPT and whether that contravenes the customer's internal policies.

[0128] The artificial intelligence-based cyber security analyst and/or cyber threat detect engine can have a natural language interface that will enable every user, regardless of experience level or organization size, to take full advantage of the capabilities of that tool - artificial intelligence-based cyber security analyst and/or cyber threat detect engine. The artificial intelligence-based cyber security analyst tool leverages state-of-the-art natural language processing (NLP) techniques to understand and interpret human language in a cybersecurity context. The artificial intelligence-based cyber security analyst tool utilizes machine learning algorithms and neural networks to process and analyze vast amounts of textual data, allowing it to extract meaningful insights and patterns. By comprehending human language, the tool can effectively interact with users, making it accessible to individuals with varying levels of experience and expertise in cybersecurity.

[0129] The artificial intelligence-based cyber security analyst tool helps human cyber security personnel to ask simple, straightforward questions, and then the artificial intelligence-based cyber security analyst tool provides real-time insight into an organization's risk profile, including its threat landscape, risk level against critical vulnerabilities, current security posture, compliance requirements, cybersecurity performance metrics and much more.

[0130] The artificial intelligence-based cyber security analyst tool helps human cyber security personnel to better understand the threats and risks facing their organization. The artificial intelligence-based cyber security analyst tool also helps Security Analysts make better decisions for cyber threat hunting faster, reducing response time to critical incidents. The artificial intelligence-based cyber security analyst tool provides basic information such as which threat actors are targeting that network, what are the critical vulnerabilities being exploited by these adversaries, what are the top recommended remediation actions for the impacted endpoint, etc.

[0131] Thus, in terms of functionality, the artificial intelligence-based cyber security analyst tool provides a wide range of capabilities to support cybersecurity personnel in their day-to-day operations. One of the artificial intelligence-based cyber security analyst tool's primary roles is to assist in risk assessment by answering questions and providing real-time insights into an organization's risk profile. This includes assessing the threat landscape by identifying potential adversaries and their tactics, evaluating the organization's risk level against critical vulnerabilities, and assessing the overall security posture.

[0132] Furthermore, the artificial intelligence-based cyber security analyst tool aids security analysts in making informed decisions for cyber threat hunting. The artificial intelligence-based cyber security analyst tool equips human security analysts with the necessary information to proactively identify and mitigate potential security incidents. By analyzing data from various sources, such as network logs, intrusion detection systems, and threat intelligence feeds, the tool can highlight indicators of compromise (IOCs) and provide recommendations for remediation actions. This empowers security analysts to respond swiftly to critical incidents and minimize the impact on the organization.

[0133] To streamline security operations, the artificial intelligence-based cyber security analyst tool automates repetitive and mundane tasks, enabling analysts to focus on more critical and complex activities. The artificial intelligence-based cyber security analyst can handle data collection and extraction from multiple sources, reducing the manual effort required to gather relevant information for analysis. Additionally, the tool performs basic threat searches and detection, leveraging its understanding of threat intelligence and historical attack patterns to identify potential security threats.

[0134] The artificial intelligence-based cyber security analyst tool as a Large language model (LLM) is built to incorporate cyber threat knowledge from external data stores, external data sources, as well as from a network's own cyber security appliance. The artificial intelligence-based cyber security analyst tool uses threat intelligence to understand a cyber threat adversary tactics and motivations. The effectiveness of the artificial intelligence-based cyber security analyst tool lies in its ability to access and integrate diverse data sources. The artificial intelligence-based cyber security analyst

can tap into external data stores, such as threat intelligence platforms and vulnerability databases, to enrich its understanding of the threat landscape. Additionally, the tool leverages data from an organization's internal cybersecurity appliance, an intrusion prevention simulator system, and endpoint protection solutions, to gain insights into the specific threats and vulnerabilities faced by the organization. This combination of external and internal data sources allows the artificial intelligence-based cyber security analyst tool to provide a holistic view of the organization's security posture.

[0135] Through continuous improvement and refinement, the artificial intelligence-based cyber security analyst tool ensures that it remains up-to-date with the evolving cybersecurity landscape. It collaborates closely with industry-leading cyber threat hunters, managed detection and response operators, and incident response experts to incorporate their expertise and insights into its analysis. This feedback loop helps enhance the tool's accuracy and effectiveness, enabling it to adapt to emerging threats and new attack vectors.

[0136] In addition to its technical capabilities, the artificial intelligence-based cyber security analyst tool facilitates human-machine collaboration to tackle the challenges posed by adversaries. By leveraging the speed and computational power of AI, combined with human expertise and intuition, security teams can gain a significant advantage. The tool serves as a force multiplier, augmenting the capabilities of security analysts and enabling them to detect and respond to threats more efficiently.

[0137] The artificial intelligence-based cyber security analyst tool understands the importance of adaptability and agility in the face of evolving threats. The artificial intelligence-based cyber security analyst tool employs advanced analytics to prioritize critical vulnerabilities and potential risks, enabling organizations to allocate their resources effectively. By generating and validating new indicators of attack (IOAs), the tool aids in the identification of emerging attack patterns, empowering security teams to proactively defend against evolving threats.

[0138] Ultimately, the artificial intelligence-based cyber security analyst tool recognizes that effective cybersecurity requires a combination of AI-driven analysis and human intelligence. The artificial intelligence-based cyber security analyst tool can incorporate

human insight into the investigative loop. The artificial intelligence-based cyber security analyst tool tracks the adversaries that are constantly innovating and adapting their tactics, and incorporates human expertise and insights into the training data used by the models in the artificial intelligence-based cyber security analyst tool. By blending the strengths of AI and human intelligence, organizations can stay ahead of the threat landscape and proactively defend against emerging cyber threats. The adversary is constantly breaking rules and changing tactics, making it hard for AI to respond without the right data to train the model. This is why human-validated content is critical for AI to perform security use cases and give security teams the advantage over adversaries.

[0139] Additional AI concepts

[0140] 1. Use of Transformers & LSTMs to Model SaaS File:

[0141] A method of identifying anomalous behavior of users in the SaaS environment by determining “normal” file locations accessed by the user or their peer group (where the peer group is also defined by the classifier). The classifier (referred to as CUDAP internally) receives and records the file paths accessed on a SaaS platform by all users and after a critical number, begins to train on this list of file paths and users. The classifier uses a method of machine learning that learns the likelihood of a sequence of words occurring in a specific order that is normally used for other forms of text. The classifier also clusters users based upon similar file path usage/behavior and then adds that into the sequence. The classifier is defined as bi-directional, in that it can output a likely user cluster when given a file path or output likely file paths when given a user or cluster.

[0142] 2. Detecting threats and response from machine learning analysis of text –

[0143] 3. A. Using Bert Text Classification Applied To Incident Response Comments To Automatically Assess The Action Taken For An Event

[0144] Taking comments from internal cyber security analysts and grouping them using a generative probabilistic model of a corpus such as Latent Dirichlet Allocation. Then having assigned categories to these groups, use the labelled data to train a BERT Text Classification ML.

[0145] A. Process

[0146] i. using a generative probabilistic model of a corpus such as Latent Dirichlet Allocation on comments to assign comments into ~5 categories

[0147] ii. Train a pre-trained AI based deep learning Transformer neural network used in language understanding (e.g., BERT - Bidirectional Encoder Representations from Transformers language model)

[0148] iii. BERT learns the language associated with each of the applied categories

[0149] iv. Use the trained BERT to assess the likely category of subsequent incident response comments

[0150] 3.B. Categorization of Threat Intel Indicators of Compromise Based on Comment Text and Associated Tags

[0151] Autogenerate categories into which a particular Indicators of Compromise (IoC) falls based on the tags and associated comment texts in Researcher.

[0152] A. Process

[0153] i. Latent Dirichlet Allocation of comments to assign comments into ~5 categories

[0154] ii. Gather tags from various sources – Researcher, Alienvault, Yeti, VT, Google, UrlHaus

[0155] iii. Association of groups with tags – use labelled data to train a language aware neural network such as BERT.

[0156] iv. Classification of new IoC using BERT by analyzing associated comments/tags of new IoC

[0157] 3. "Training a Large Language Model (LLM) on hostname links" "Using a language model for semantic encodings of URIs"

[0158] Large language models such as BERT are pretrained on a large amount of linguistic data, and have a generic understanding of language that can be fine-tuned for a wide range of specific tasks. URIs are in some ways similar to natural language, containing a range of linguistic elements, but their syntax differs in several respects, meaning that limited results can be achieved by fine-tuning large language models, and

this fine-tuning must be redone for every URI-related task. By making a large dataset of URIs and related information (e.g. titles, page content), a large language model can be trained specifically on the patterns in URIs, creating a model with a generic semantic understanding of URIs, including e.g. that “somesite.com/training.mp4” likely relates to a video file, that “somesite.com/activate.php” will involve the execution of server side code, or that “somesite.com/data?starttime=2021” relates to the extraction of filtered API data. This generic model can then be fine-tuned on a range of specific URI-related tasks, such as looking for malicious URI patterns, analyzing phishing links, or detecting C2 communication, providing more power and versatility.

[0159] The resulting model can encode URLs for search & classification across a huge range of tasks.

[0160] This includes:

[0161] • Identifying new threats by detecting common URI patterns.

[0162] • Categorizing threat types from their URL.

[0163] • Predicting if a URL is likely to be malicious.

[0164] • Searching for similar threats by using their URLs.

[0165] 4. An AI interface for the UI into various cyber security system components

[0166] As an ease-of-use tool – a complex AI model which understands complex textual/voice requests.

[0167] Use a large autoregressive language model trained on a huge volume of code. This would then be provided with our API spec.

[0168] Provided the model is large enough, it will be able to convert user requests “Get the last alert from an iPhone” and convert it into the JavaScript code to open that in the UI.

[0169] The resulting product would be a Siri-style interface for the cyber security UI.

[0170] 5. Automated Analysis of Threat Landscape to influence autonomous response severity

[0171] Part of a normal dev/analyst day at work is keeping up to date with the latest malware and news around cyber security. By doing so we build up a knowledge of what malware exists, what its entry points are and its symptoms once infected. This can be used to modify products, update models, aid in analysis/searching of customer systems and on the marketing side keep on top of what customers worries might be and questions they may have. There are many powerful NLP tools that exist such as BERT which can be adapted for use as text summarizers and key word extractors with intrinsic understanding of context and semantics.

[0172] Our proposal is to scrape many common news and information sources and use one of these machine learning tools to build up a regularly updating database of: Existing malware, its impact, the wider community perception of it, the patterns of infection that it follows, the symptoms of infection, the entry routes it might take. By cross referencing multiple sources, we can build levels of confidence in our stored data to ignore erroneous reports.

[0173] The second part suggested is building a pattern of life for malicious organizations to make predictions on who will be impacted, at what level and to what extent a company should be worried about it.

[0174] This would be performed by analyzing the patterns of mercenary APTs publicly claimed victims (and statements from the APT groups), the political landscape, dark web content, company age/history/exposure and the general atmosphere/consensus of the hacking scene. This would be content analysis on dark web pages and on social media.

[0175] This information can be used to increase priority warnings for specific industry verticals or activity types. For example, autonomous response actions on “fashionable” approaches can be more stringent. Autonomous response actions to mitigate a cyber threat reactivity on customers known to sit within targeted verticals can be increased.

[0176] 6. Self-supervised Vectorization of Network Entities likely uses word2vec style concepts

[0177] For many years, researchers have been developing contextual sequence prediction in the field of Natural Language Processing in order to allow machines to better

understand human language. In 2013 a team of researchers at Google published a simple but elegant technique named word2vec, which maps words to vectors based on contextual similarity. Put simply, differing words which have the same position within a sentence would be considered “contextually similar”. For example, in the following sentences:

[0178] • The King ruled over the land.,

[0179] • The Queen ruled over the land., and

[0180] • The Lord ruled over the land.

[0181] “King”, “Queen” and “Lord” would be considered contextually identical. In addition to contextual similarity, this algorithm also preserves syntactic and semantic relationships such as king – man + woman = queen.

[0182]

[0183] Far less explored and yet to be exploited to their full potential is the application of these techniques to network data or process execution pathways for security purposes. By constructing “sentences” based on network traffic and applying similar techniques, it is possible to construct and populate a vector space of network entities with a distance metric equivalent to contextual similarity.

[0184] This approach was applied to large quantities of connection data collected from a single cyber security appliance deployment. Many interesting outputs were observed and semantic relationships appear to have been preserved: Google – US + Russia = Yandex. These results suggest a number of potential applications, including:

[0185] • Categorization of hostnames using clustering and bootstrapping.

[0186] • Identification of unusual combinations of network entities that might connote malicious activities such as C2 or DNS poisoning.

[0187] • Contextual analysis of potential IoC’s during threat investigation, via nearest-neighbor entity lookup.

[0188] • The vectorized network entities can be used to create a geopolitical map of the Internet, highlighting the strength of relationships between countries and their organizations.

[0189] 7. Fine-tuned Masked URI Transformer for Fuzzy Matching Malicious URIs

[0190] Specific malware variants often use specific URI patterns in their C2 communication, however these URIs are often sufficiently varied that simple string-based matching (or even native fuzzy matching) is not sufficient to search for instances of these variants. For instance, the two following URIs show patterns typical of LokiBot malware, whereas the third does not, despite being superficially similar in form:

[0191] 1. /weofhwefefoefeifei/Xf419.exe

[0192] 2. /sdfhwelkfjwefjewlieje/Q129x.exe

[0193] 3. /gpu-emulators/nvidia.exe

[0194] These differences and patterns are nonetheless clear to trained humans, and advances in NLP have come a long way to capturing such human insights – in particular, masked language modelling has led to the development of generic large language models such as BERT, which can be fine-tuned on specific linguistic tasks.

[0195] With appropriate tokenization and masked language modelling on a wide set of URIs, a transformer-based large language model could be trained in an essentially unsupervised manner. This resulting model could then be fine-tuned on specific URI-related tasks – for instance, a classifier that takes two URIs, and determines whether they are likely to be associated with the same malware variant. Such a classifier would allow datalakes of URIs to be searched for clusters of malicious URIs, and help determine whether novel URIs fall into known malicious patterns.

[0196] 8. Higher Level Incident Classification and Analysis using Graph Neural Networks

[0197] Recent updates to AI Analyst in version 5.2 allow complex cross-entity incidents to be identified using graph-based analysis (covered in a previous patent). While these resulting incidents successfully capture complex compromises, since they contain

multiple different components, it is often difficult to give a given incident an overall description, or to determine the kind of overall compromise or malware involved.

[0198] The recent development of graph neural networks allows the creation of models that specifically operate on graphs – for instance, models that classify graphs based on their properties. As a result, by training such a model on the subgraphs associated with incidents, it may be possible to classify them, identifying an overall description, highlighting the kinds of overall activity involved, and aiding in the suggestion of remediation steps.

[0199] 9. Chatbots

[0200] 10. Identifying Probable URL Shorteners with a Fine-Tuned Hostname-Specific Large Language Model

[0201] Link shortening services are ubiquitous due to the legitimate need to produce short URLs to fit into the confines of e.g., text messages and tweets. However, since they provide legitimate hosts that redirect to any user-provided endpoint, they are also frequently used for malicious purposes such as phishing. As a result, they frequently appear in alerts and lists of malicious hostnames, however they are not inherently malicious, and so should not be highlighted as malicious hosts in general by services such as deep space/inoculation. When combined with the fact that there are large numbers of link shortening services, with new ones regularly appearing, it would be extremely useful to automatically identify possible link shortening services without having to rely on known lists.

[0202] As covered in previous patent submissions, we have a large language model with a similar architecture to the BERT transformer model, but trained on masking of hostname-hostname links extracted from common crawl data. This model thus has an understanding of not just the form of hostnames, but also what they are likely to link to – and so essentially has a semantic understanding of the role of a hostname in the wider internet. Consequently, this model can be fine-tuned specifically on the task of detecting hostnames associated with link shortening services, leading to a high-performance classifier.

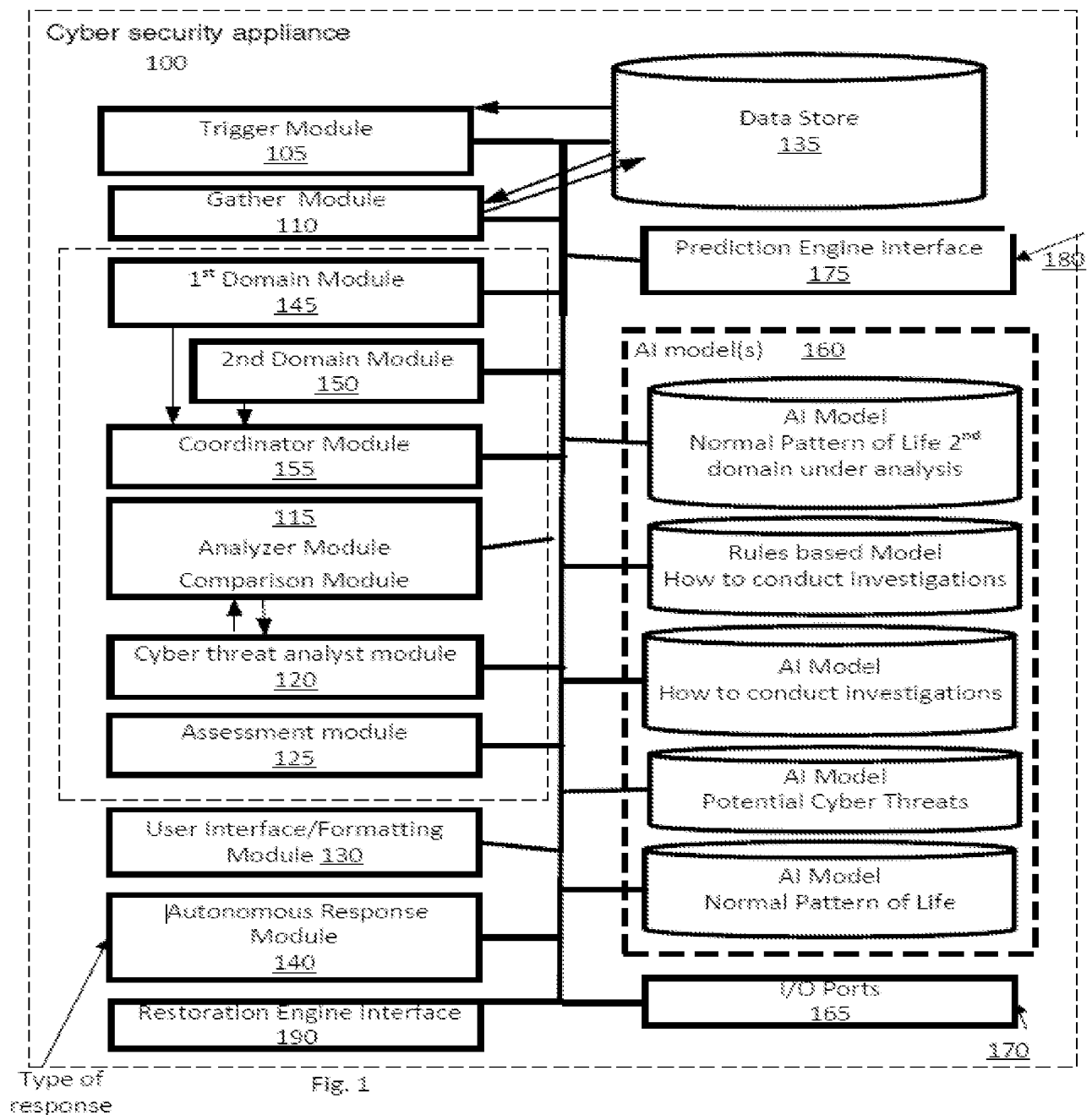
[0203] 11. Email inducement text highlighting

[0204] This product takes in the text of the email and highlights words and phrases which correspond to different types of inducements: redirection, offers, financial details, instructions, and threats. This is to allow customers to see why a specific email has been flagged by AGE security products. The product works in the following way.

[0205] Firstly, several thousand customer emails, which flagged against the AGE security products, were taken. I then labelled phrases within the emails that corresponded to the inducement, set out in discussion with AGE.

[0206] The cyber security system then takes a large transformers model (in this case BERT) and train the model to accurately predict the inducement labels in a named entity recognition task.

[0207] Afterwards, customers will be able to ping the model with an email and then model will return a color highlighted analysis of the text highlighting phrases of interest with the corresponding labels.



[0208] Figure 1 illustrates a block diagram of an embodiment of the AI based cyber security appliance 100 that protects a system, including but not limited to a network/domain, from cyber threats. Various Artificial Intelligence models and modules of the cyber security appliance 100 cooperate to protect one or more networks/domains under analysis from cyber threats. As shown, according to one embodiment of the disclosure, the AI-based cyber security appliance 100 may include a trigger module 105,

a gather module 110, an analyzer module 115, a cyber threat analyst module 120, an assessment module 125, a formatting module 130, a data store 135, an autonomous response module 140, a first (1st) domain module 145, a second (2nd) domain module 150, and a coordinator module 155, one or more AI models 160 (hereinafter, AI model(s)), and/or other modules. The AI model(s) 160 may be trained with machine learning on a normal pattern of life for entities in the network(s)/domain(s) under analysis, with machine learning on cyber threat hypotheses to form and investigate a cyber threat hypothesis on what are a possible set of cyber threats and their characteristics, symptoms, remediations, etc., and/or trained on possible cyber threats including their characteristics and symptoms, a data store, an autonomous response module, a 1st domain module, a 2nd domain module, and a coordinator module.

[0209] The cyber security appliance 100 with the Artificial Intelligence (AI) based cyber security system may protect a network/domain from a cyber threat. In an embodiment, the cyber security appliance 100 can protect all of the devices (e.g., computing devices on the network(s)/domain(s) being monitored by monitoring domain activity including communications). For example, a network domain module (e.g., first domain module 145) may communicate with network sensors to monitor network traffic going to and from the computing devices on the network as well as receive secure communications from software agents embedded in host computing devices/containers. The steps below will detail the activities and functions of several of the components in the cyber security appliance 100.

[0210] The gather module 110 may be configured with one or more process identifier classifiers. Each process identifier classifier may be configured to identify and track one or more processes and/or devices in the network, under analysis, making communication connections. The data store 135 cooperates with the process identifier classifier to collect and maintain historical data of processes and their connections, which is updated over time as the network is in operation. Individual processes may be present in merely one or more domains being monitored. In an example, the process identifier classifier can identify each process running on a given device along with its endpoint connections, which are stored in the data store 135. In addition, a feature classifier can examine and determine features in the data being analyzed into different categories.

[0211] The analyzer module 115 can cooperate with the AI model(s) 160 or other modules in the cyber security appliance 100 to confirm a presence of a cyberattack against one or more domains in an enterprise's system (e.g., see system/enterprise network 50 of Figure 5). A process identifier in the analyzer module 115 can cooperate with the gather module 110 to collect any additional data and metrics to support a possible cyber threat hypothesis. Similarly, the cyber threat analyst module 120 can cooperate with the internal data sources as well as external data sources to collect data in its investigation. More specifically, the cyber threat analyst module 120 can cooperate with the other modules and the AI model(s) 160 in the cyber security appliance 100 to conduct a long-term investigation and/or a more in-depth investigation of potential and emerging cyber threats directed to one or more domains in an enterprise's system. Herein, the cyber threat analyst module 120 and/or the analyzer module 115 can also monitor for other anomalies, such as model breaches, including, for example, deviations for a normal behavior of an entity, and other techniques discussed herein. As an illustrative example, the analyzer module 115 and/or the cyber threat analyst module 120 can cooperate with the AI model(s) 160 trained on potential cyber threats in order to assist in examining and factoring these additional data points that have occurred over a given timeframe to see if a correlation exists between 1) a series of two or more anomalies occurring within that time frame and 2) possible known and unknown cyber threats. The cyber threat analyst module can cooperate with the internal data sources as well as external data sources to collect data in its investigation.

[0212] According to one embodiment of the disclosure, the cyber threat analyst module 120 allows two levels of investigations of a cyber threat that may suggest a potential impending cyberattack. In a first level of investigation, the analyzer module 115 and AI model(s) 160 can rapidly detect and then the autonomous response module 140 will autonomously respond to overt and obvious cyberattacks. However, thousands to millions of low level anomalies occur in a domain under analysis all of the time; and thus, most other systems need to set the threshold of trying to detect a cyberattack by a cyber threat at level higher than the low level anomalies examined by the cyber threat analyst module 120 just to not have too many false positive indications of a cyberattack when one is not actually occurring, as well as to not overwhelm a human cyber security analyst

receiving the alerts with so many notifications of low level anomalies that they just start tuning out those alerts. However, advanced persistent threats attempt to avoid detection by making these low-level anomalies in the system over time during their cyberattack before making their final coup de grâce / ultimate mortal blow against the system (e.g., domain) being protected. The cyber threat analyst module 120 also conducts a second level of investigation over time with the assistance of the AI model(s) 160 trained with machine learning on how to form cyber threat hypotheses and how to conduct investigations for a cyber threat hypothesis that can detect these advanced persistent cyber threats actively trying to avoid detection by looking at one or more of these low-level anomalies as a part of a chain of linked information.

[0213] Note, a data analysis process can be algorithms/scripts written by humans to perform their function discussed herein; and can in various cases use AI classifiers as part of their operation. The cyber threat analyst module 120 forms in conjunction with the AI model(s) 160 trained with machine learning on how to form cyber threat hypotheses and how to conduct investigations for a cyber threat hypothesis investigate hypotheses on what are a possible set of cyber threats. The cyber threat analyst module 120 can also cooperate with the analyzer module 115 with its one or more data analysis processes to conduct an investigation on a possible set of cyber threats hypotheses that would include an anomaly of at least one of i) the abnormal behavior, ii) the suspicious activity, and iii) any combination of both, identified through cooperation with, for example, the AI model(s) 160 trained with machine learning on the normal pattern of life of entities in the system. For example, as shown in Figure 4, the cyber threat analyst module 120 may perform several additional rounds 400 of gathering additional information, including abnormal behavior, over a period of time, in this example, examining data over a 7-day period to determine causal links between the information. The cyber threat analyst module 120 may submit to check and recheck various combinations / a chain of potentially related information, including abnormal behavior of a device/user account under analysis for example, until each of the one or more hypotheses on potential cyber threats are one of 1) refuted, 2) supported, or 3) included in a report that includes details of activities assessed to be relevant activities to the anomaly of interest to the user and that also conveys at least this particular hypothesis was neither supported or refuted. For this

embodiment, a human cyber security analyst is needed to further investigate the anomaly (and/or anomalies) of interest included in the chain of potentially related information.

[0214] Returning back to Figure 1, an input from the cyber threat analyst module 120 of a supported hypothesis of a potential cyber threat will trigger the analyzer module 115 to compare, confirm, and send a signal to act upon and mitigate that cyber threat. In contrast, the cyber threat analyst module 120 investigates subtle indicators and/or initially seemingly isolated unusual or suspicious activity such as a worker is logging in after their normal working hours or a simple system misconfiguration has occurred. Most of the investigations conducted by the cyber threat analyst module 120 cooperating with the AI model(s) 160 trained with machine learning on how to form cyber threat hypotheses and how to conduct investigations for a cyber threat hypothesis on unusual or suspicious activities/behavior may not result in a cyber threat hypothesis that is supported but rather most are refuted or simply not supported. Typically, during the investigations, several rounds of data gathering to support or refute the long list of potential cyber threat hypotheses formed by the cyber threat analyst module 120 will occur before the algorithms in the cyber threat analyst module 120 will determine whether a particular cyber threat hypothesis is supported, refuted, or needs further investigation by a human. The rounds of data gathering may build chains of linked low-level indicators of unusual activity along with potential activities that could be within a normal pattern of life for that entity to evaluate the whole chain of activities to support or refute each potential cyber threat hypothesis formed. (See again, for example, Figure 4 and a chain of linked low-level indicators, including abnormal behavior compared to the normal pattern of life for that entity, all under a score of 50 on a threat indicator score). The investigations by the cyber threat analyst module 120 can happen over a relatively long period of time and be far more in depth than the analyzer module 115 which will work with the other modules and AI model(s) 160 to confirm that a cyber threat has in fact been detected.

[0215] The gather module 110 may further extract data from the data store 135 at the request of the cyber threat analyst module 120 and/or analyzer module 115 on each possible hypothetical threat that would include the abnormal behavior or suspicious activity and then can assist to filter that collection of data down to relevant points of data to either 1) support or 2) refute each particular hypothesis of what the cyber threat, the

suspicious activity and/or abnormal behavior relates to. The gather module 110 cooperates with the cyber threat analyst module 120 and/or analyzer module 115 to collect data to support or to refute each of the one or more possible cyber threat hypotheses that could include this abnormal behavior or suspicious activity by cooperating with one or more of the cyber threat hypotheses mechanisms to form and investigate hypotheses on what are a possible set of cyber threats.

[0216] Thus, the cyber threat analyst module 120 is configured to cooperate with the AI model(s) 160 trained with machine learning on how to form cyber threat hypotheses and how to conduct investigations for a cyber threat hypothesis to form and investigate hypotheses on what are a possible set of cyber threats and then can cooperate with the analyzer module 115 with the one or more data analysis processes to confirm the results of the investigation on the possible set of cyber threats hypotheses that would include the at least one of i) the abnormal behavior, ii) the suspicious activity, and iii) any combination of both, identified through cooperation with the AI model(s) 160 trained with machine learning on the normal pattern of life/normal behavior of entities in the domains under analysis.

[0217] Note, in the first level of threat detection, the gather module 110 and the analyzer module 115 cooperate to supply any data and/or metrics requested by the analyzer module 115 cooperating with the AI model(s) 160 trained on possible cyber threats to support or rebut each possible type of cyber threat. Again, the analyzer module 115 can cooperate with the AI model(s) 160 and/or other modules to rapidly detect and then cooperate with the autonomous response module 140 to autonomously respond to overt and obvious cyberattacks, (including ones found to be supported by the cyber threat analyst module 120).

[0218] As a starting point, the AI-based cyber security appliance 100 can use multiple modules, each capable of identifying abnormal behavior and/or suspicious activity against the AI model(s) 160 trained on a normal pattern of life for the entities in the network/domain under analysis, which is supplied to the analyzer module 115 and/or the cyber threat analyst module 120. The analyzer module 115 and/or the cyber threat

analyst module 120 may also receive other inputs such as AI model breaches, AI classifier breaches, etc. a trigger to start an investigation from an external source.

[0219] Many other model breaches of the AI model(s) 160 trained with machine learning on the normal behavior of the system can send an input into the cyber threat analyst module 120 and/or the trigger module 105 to trigger an investigation to start the formation of one or more hypotheses on what are a possible set of cyber threats that could include the initially identified abnormal behavior and/or suspicious activity. Note, a deeper analysis can look at example factors such as i) how long has the endpoint existed or is registered; ii) what kind of certificate is the communication using; iii) is the endpoint on a known good domain or known bad domain or an unknown domain, and if unknown what other information exists such as registrant's name and/or country; iv) how rare; v), etc.

[0220] Note, the cyber threat analyst module 120 cooperating with the AI model(s) 160 trained with machine learning on how to form cyber threat hypotheses and how to conduct investigations for a cyber threat hypothesis in the AI-based cyber security appliance 100 provides an advantage as it reduces the time taken for human led or cyber security investigations, provides an alternative to manpower for small organizations and improves detection (and remediation) capabilities within the cyber security appliance 100.

[0221] The cyber threat analyst module 120, which forms and investigates hypotheses on what are the possible set of cyber threats, can use hypotheses mechanisms including any of 1) one or more of the AI model(s) 160 trained on how human cyber security analysts form cyber threat hypotheses and how to conduct investigations for a cyber threat hypothesis that would include at least an anomaly of interest, 2) one or more scripts outlining how to conduct an investigation on a possible set of cyber threats hypotheses that would include at least the anomaly of interest, 3) one or more rules-based models on how to conduct an investigation on a possible set of cyber threats hypotheses and how to form a possible set of cyber threats hypotheses that would include at least the anomaly of interest, and 4) any combination of these. Again, the AI model(s) 160 trained on 'how to form cyber threat hypotheses and how to conduct investigations for a cyber threat hypothesis' may use supervised machine learning on human-led cyber threat investigations and then steps, data, metrics, and metadata on how to support or to refute

a plurality of the possible cyber threat hypotheses, and then the scripts and rules-based models will include the steps, data, metrics, and metadata on how to support or to refute the plurality of the possible cyber threat hypotheses. The cyber threat analyst module 120 and/or the analyzer module 115 can feed the cyber threat details to the assessment module 125 to generate a threat risk score that indicate a level of severity of the cyber threat.

[0222] According to one embodiment of the disclosure, the assessment module 125 can cooperate with the AI model(s) 160 trained on possible cyber threats to use AI algorithms to identify actual cyber threats and generate threat risk scores based on both the level of confidence that the cyber threat is a viable threat and the severity of the cyber threat (e.g., attack type where ransomware attacks has greater severity than phishing attack; degree of infection; computing devices likely to be targeted, etc.). The threat risk scores be used to rank alerts that may be directed to enterprise or computing device administrators. This risk assessment and ranking is conducted to avoid frequent “false positive” alerts that diminish the degree of reliance/confidence on the cyber security appliance 100.

[0223]

[0224]

INSERT Additional NEW PATENT CONTENT below this is all background on how the overall system works, AI training and operation, computer ops, etc.

[0225]

[0226]

[0227]

[0228]

[0229] **Training of AI pre deployment and then during deployment**

[0230] In step 1, an initial training of the AI model trained on cyber threats can occur using unsupervised learning and/or supervised learning on characteristics and attributes of known potential cyber threats including malware, insider threats, and other kinds of cyber threats that can occur within that domain. Each Artificial Intelligence can be programmed

and configured with the background information to understand and handle particulars, including different types of data, protocols used, types of devices, user accounts, etc. of the system being protected. The Artificial Intelligence pre-deployment can all be trained on the specific machine learning task that they will perform when put into deployment. For example, the AI model, such as AI model(s) 160 or example (hereinafter “AI model(s) 160”), trained on identifying a specific cyber threat learns at least both in the pre-deployment training i) the characteristics and attributes of known potential cyber threats as well as ii) a set of characteristics and attributes of each category of potential cyber threats and their weights assigned on how indicative certain characteristics and attributes correlate to potential cyber threats of that category of threats.

[0231] In this example, one of the AI model(s) 160 trained on identifying a specific cyber threat can be trained with machine learning such as Linear Regression, Regression Trees, Non-Linear Regression, Bayesian Linear Regression, Deep learning, etc. to learn and understand the characteristics and attributes in that category of cyber threats. Later, when in deployment in a domain/network being protected by the cyber security appliance 100, the AI model trained on cyber threats can determine whether a potentially unknown threat has been detected via a number of techniques including an overlap of some of the same characteristics and attributes in that category of cyber threats. The AI model may use unsupervised learning when deployed to better learn newer and updated characteristics of cyberattacks.

[0232] In an embodiment, one or more of the AI models 160 may be trained on a normal pattern of life of entities in the system are self-learning AI model using unsupervised machine learning and machine learning algorithms to analyze patterns and 'learn' what is the 'normal behavior' of the network by analyzing data on the activity on, for example, the network level, at the device level, and at the employee level. The self-learning AI model using unsupervised machine learning understands the system under analysis' normal patterns of life in, for example, a week of being deployed on that system, and grows more bespoke with every passing minute. The AI unsupervised learning model learns patterns from the features in the day-to-day dataset and detecting abnormal data which would not have fallen into the category (cluster) of normal behavior. The self-learning AI model using unsupervised machine learning can simply be placed into an observation mode for

an initial week or two when first deployed on a network/domain in order to establish an initial normal behavior for entities in the network/domain under analysis.

[0233] A deployed AI model trained on a normal pattern of life of entities in the system can be configured to observe the nodes in the system being protected. Training on a normal behavior of entities in the system can occur while monitoring for the first week or two until enough data has been observed to establish a statistically reliable set of normal operations for each node (e.g., user account, device, etc.). Initial training of one or more of the AI models 160 of Figure 1 trained with machine learning on the pattern of life of the entities in the network/domain can occur where each type of network and/or domain will generally have some common typical behavior with each model trained specifically to understand components/devices, protocols, activity level, etc. to that type of network/system/domain. Alternatively, pre-deployment machine learning training of the AI model(s) 160 of Figure 1 trained on a normal pattern of life of entities in the system can occur. Initial training of the AI model(s) 160 trained with machine learning on a behavior of the pattern of life of the entities in the network/domain can occur where each type of network and/or domain will generally have some common typical behavior with each model trained specifically to understand components/devices, protocols, activity level, etc. to that type of network/system/domain. What is normal behavior of each entity within that system can be established either prior to deployment and then adjusted during deployment or alternatively the model can simply be placed into an observation mode for an initial week or two when first deployed on a network/domain in order to establish an initial normal behavior for entities in the network/domain under analysis.

[0234] During deployment, what is considered normal behavior will change as each different entity's behavior changes and will be reflected through the use of unsupervised learning in the model such as various Bayesian techniques, clustering, etc. The AI models 160 can be implemented with various mechanisms such neural networks, decision trees, etc. and combinations of these. Likewise, one or more supervised machine learning AI models 160 may be trained to create possible hypotheses and perform cyber threat investigations on agnostic examples of past historical incidents of detecting a multitude of possible types of cyber threat hypotheses previously analyzed by human cyber security analyst. More on the training of AI models 160 are trained to create

one or more possible hypotheses and perform cyber threat investigations will be discussed later.

[0235] At its core, the self-learning AI models 160 that model the normal behavior (e.g. a normal pattern of life) of entities in the network mathematically characterizes what constitutes 'normal' behavior, based on the analysis of a large number of different measures of a device's network behavior - packet traffic and network activity/processes including server access, data volumes, timings of events, credential use, connection type, volume, and directionality of, for example, uploads/downloads into the network, file type, packet intention, admin activity, resource and information requests, command sent, etc.

[0236] Clustering Methods

[0237] In order to model what should be considered as normal for a device or cloud container, its behavior can be analyzed in the context of other similar entities on the network. The AI models (e.g., AI model(s) 160) can use unsupervised machine learning to algorithmically identify significant groupings, a task which is virtually impossible to do manually. To create a holistic image of the relationships within the network, the AI models and AI classifiers employ a number of different clustering methods, including matrix-based clustering, density-based clustering, and hierarchical clustering techniques. The resulting clusters can then be used, for example, to inform the modeling of the normative behaviors and/or similar groupings.

[0238] The AI models and AI classifiers can employ a large-scale computational approach to understand sparse structure in models of network connectivity based on applying L1-regularization techniques (the lasso method). This allows the artificial intelligence to discover true associations between different elements of a network which can be cast as efficiently solvable convex optimization problems and yield parsimonious models. Various mathematical approaches assist.

[0239] Next, one or more supervised machine learning AI models are trained to create possible hypotheses and how to perform cyber threat investigations on agnostic examples of past historical incidents of detecting a multitude of possible types of cyber threat hypotheses previously analyzed by human cyber security analyst. AI models trained on forming and investigating hypotheses on what are a possible set of cyber

threats can be trained initially with supervised learning. Thus, these AI models can be trained on how to form and investigate hypotheses on what are a possible set of cyber threats and steps to take in supporting or refuting hypotheses. The AI models trained on forming and investigating hypotheses are updated with unsupervised machine learning algorithms when correctly supporting or refuting the hypotheses including what additional collected data proved to be the most useful.

[0240] Next, the various AI models and AI classifiers combine use of unsupervised and supervised machine learning to learn ‘on the job’ – it does not depend upon solely knowledge of previous cyberattacks. The AI models and classifiers combine use of unsupervised and supervised machine learning constantly revises assumptions about behavior, using probabilistic mathematics, which is always up to date on what a current normal behavior is, and not solely reliant on human input. The AI models and classifiers combine use of unsupervised and supervised machine learning on cyber security is capable of seeing hitherto undiscovered cyber events, from a variety of threat sources, which would otherwise have gone unnoticed.

[0241] Next, these cyber threats can include, for example, Insider threat – malicious or accidental, Zero-day attacks – previously unseen, novel exploits, latent vulnerabilities, machine-speed attacks – ransomware and other automated attacks that propagate and/or mutate very quickly, Cloud and SaaS-based attacks, other silent and stealthy attacks advance persistent threats, advanced spear-phishing, etc.

[0242] **Ranking the Cyber threat**

[0243] The assessment module 125 and/or cyber threat analyst module 120 of Figure 1 can cooperate with the AI model(s) 160 trained on possible cyber threats to use AI algorithms to account for ambiguities by distinguishing between the subtly differing levels of evidence that characterize network data. Instead of generating the simple binary outputs ‘malicious’ or ‘benign’, the AI’s mathematical algorithms produce outputs marked with differing degrees of potential threat. This enables users of the system to rank alerts or notifications to the enterprise security administrator in a rigorous manner and prioritize those which most urgently require action. Meanwhile, it also assists to avoid the problem of numerous false positives associated with simply a rule-based approach.

[0244] As discussed in more detail above, the analyzer module 115 and/or cyber threat analyst module 120 can cooperate with the one or more unsupervised AI (machine learning) model 160 trained on the normal pattern of life/normal behavior in order to perform anomaly detection against the actual normal pattern of life for that system to determine whether an anomaly (e.g., the identified abnormal behavior and/or suspicious activity) is malicious or benign. In the operation of the cyber security appliance 100, the emerging cyber threat can be previously unknown, but the emerging threat landscape data 170 representative of the emerging cyber threat shares enough (or does not share enough) in common with the traits from the AI models 160 trained on cyber threats to now be identified as malicious or benign. Note, if later confirmed as malicious, then the AI models 160 trained with machine learning on possible cyber threats can update their training. Likewise, as the cyber security appliance 100 continues to operate, then the one or more AI models trained on a normal pattern of life for each of the entities in the system can be updated and trained with unsupervised machine learning algorithms. The analyzer module 115 can use any number of data analysis processes (discussed more in detail below and including the agent analyzer data analysis process here) to help obtain system data points so that this data can be fed and compared to the one or more AI models trained on a normal pattern of life, as well as the one or more machine learning models trained on potential cyber threats, as well as create and store data points with the connection finger prints.

[0245] The AI model(s) 160 of Figures 1 and 3 can continually learn and train with unsupervised machine learning algorithms on an ongoing basis when deployed in their system that the cyber security appliance 100 is protecting. Thus, learning and training on what is normal behavior for each user, each device, and the system overall and lowering a threshold of what is an anomaly.

[0246] **Anomaly detection/ deviations**

[0247] Anomaly detection can discover unusual data points in your dataset. Anomaly can be a synonym for the word 'outlier'. Anomaly detection (or outlier detection) is the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data. Anomalous activities can be linked to some

kind of problems or rare events. Since there are tons of ways to induce a particular cyberattack, it is very difficult to have information about all these attacks beforehand in a dataset. But, since the majority of the user activity and device activity in the system under analysis is normal, the system overtime captures almost all of the ways which indicate normal behavior. And from the inclusion-exclusion principle, if an activity under scrutiny does not give indications of normal activity, the self-learning AI model using unsupervised machine learning can predict with high confidence that the given activity is anomalous. The AI unsupervised learning model learns patterns from the features in the day-to-day dataset and detecting abnormal data which would not have fallen into the category (cluster) of normal behavior. The goal of the anomaly detection algorithm through the data fed to it is to learn the patterns of a normal activity so that when an anomalous activity occurs, the modules can flag the anomalies through the inclusion-exclusion principle. The goal of the anomaly detection algorithm through the data fed to it is to learn the patterns of a normal activity so that when an anomalous activity occurs, the modules can flag the anomalies through the inclusion-exclusion principle. The cyber threat module can perform its two-level analysis on anomalous behavior and determine correlations.

[0248] In an example, 95% of data in a normal distribution lies within two standard-deviations from the mean. Since the likelihood of anomalies in general is very low, the modules cooperating with the AI model of normal behavior can say with high confidence that data points spread near the mean value are non-anomalous. And since the probability distribution values between mean and two standard-deviations are large enough, the modules cooperating with the AI model of normal behavior can set a value in this example range as a threshold (a parameter that can be tuned over time through the self-learning), where feature values with probability larger than this threshold indicate that the given feature's values are non-anomalous, otherwise it's anomalous. Note, this anomaly detection can determine that a data point is anomalous/non-anomalous on the basis of a particular feature. In reality, the cyber security appliance 100 should not flag a data point as an anomaly based on a single feature. Merely, when a combination of all the probability values for all features for a given data point is calculated can the modules cooperating with the AI model of normal behavior can say with high confidence whether a data point is an anomaly or not.

[0249] Again, the AI models trained on a normal pattern of life of entities in a system (e.g., domain) under analysis may perform the cyber threat detection through a probabilistic change in a normal behavior through the application of, for example, an unsupervised Bayesian mathematical model to detect the behavioral change in computers and computer networks. The Bayesian probabilistic approach can determine periodicity in multiple time series data and identify changes across single and multiple time series data for the purpose of anomalous behavior detection. Please reference US patent 10,701,093 granted June 30th, 2020, titled “Anomaly alert system for cyber threat detection” for an example Bayesian probabilistic approach, which is incorporated by reference in its entirety. In addition, please reference US patent publication number “US2021273958A1 filed February 26, 2021, titled “Multi-stage anomaly detection for process chains in multi-host environments” for another example anomalous behavior detector using a recurrent neural network and a bidirectional long short-term memory (LSTM), which is incorporated by reference in its entirety. In addition, please reference US patent publication number “US2020244673A1, filed April 23, 2019, titled “Multivariate network structure anomaly detector,” which is incorporated by reference in its entirety, for another example anomalous behavior detector with a Multivariate Network and Artificial Intelligence classifiers.

[0250] Next, as discussed further below, during pre-deployment the cyber threat analyst module 120 and the analyzer module 115 can use data analysis processes and cooperate with AI model(s) 160 trained on forming and investigating hypotheses on what are a possible set of cyber threats. In addition, another set of AI models can be trained on how to form and investigate hypotheses on what are a possible set of cyber threats and steps to take in supporting or refuting hypotheses. The AI models trained on forming and investigating hypotheses are updated with unsupervised machine learning algorithms when correctly supporting or refuting the hypotheses including what additional collected data proved to be the most useful.

[0251] Similarly, during deployment, the data analysis processes (discussed herein) used by the analyzer module 115 can use unsupervised machine learning to update the initial training learned during pre-deployment, and then update the training with unsupervised learning algorithms during the cyber security appliance’s 100 deployment in the system

being protected when various different steps to either i) support or ii) refute the possible set of cyber threats hypotheses worked better or worked worse.

[0252] The AI model(s) 160 trained on a normal pattern of life of entities in a domain under analysis may perform the threat detection through a probabilistic change in a normal behavior through the application of, for example, an unsupervised Bayesian mathematical model to detect a behavioral change in computers and computer networks. The Bayesian probabilistic approach can determine periodicity in multiple time series data and identify changes across single and multiple time series data for the purpose of anomalous behavior detection. In an example, a system being protected can include both email and IT network domains under analysis. Thus, email and IT network raw sources of data can be examined along with a large number of derived metrics that each produce time series data for the given metric

[0253] **Additional module interactions**

[0254] Referring back to Figure 1, the gather module 110 cooperates with the data store 135. The data store 135 stores comprehensive logs for network traffic observed. These logs can be filtered with complex logical queries and each IP packet can be interrogated on a vast number of metrics in the network information stored in the data store. Similarly, other domain's communications and data, such as emails, logs, etc. may be collected and stored in the data store 135. The gather module 110 may consist of multiple automatic data gatherers that each look at different aspects of the data depending on the particular hypothesis formed for the analysed event. The data relevant to each type of possible hypothesis can be automatically pulled from additional external and internal sources. Some data is pulled or retrieved by the gather module 110 for each possible hypothesis.

[0255] The data store 135 can store the metrics and previous threat alerts associated with network traffic for a period of time, which is, by default, at least 27 days. This corpus of data is fully searchable. The cyber security appliance 100 works with network probes to monitor network traffic and store and record the data and metadata associated with the network traffic in the data store.

[0256] The gather module 110 may have a process identifier classifier. The process identifier classifier can identify and track each process and device in the network, under analysis, making communication connections. The data store 135 cooperates with the process identifier classifier to collect and maintain historical data of processes and their connections, which is updated over time as the network is in operation. In an example, the process identifier classifier can identify each process running on a given device along with its endpoint connections, which are stored in the data store. Similarly, data from any of the domains under analysis may be collected and compared.

[0257] Examples of domains/networks under analysis being protected can include any of i) an Informational Technology network, ii) an Operational Technology network, iii) a Cloud service, iv) a SaaS service, v) an endpoint device, vi) an email domain, and vii) any combinations of these. A domain module is constructed and coded to interact with and understand a specific domain.

[0258] For instance, the first domain module 145 may operate as an IT network module configured to receive information from and send information to, in this example, IT network-based sensors (i.e., probes, taps, etc.). The first domain module 145 also has algorithms and components configured to understand, in this example, IT network parameters, IT network protocols, IT network activity, and other IT network characteristics of the network under analysis. The second domain module 150 is, in this example, an email module. The second domain module 150 can be an email network module configured to receive information from and send information to, in this example, email-based sensors (i.e., probes, taps, etc.). The second domain module 150 also has algorithms and components configured to understand, in this example, email parameters, email protocols and formats, email activity, and other email characteristics of the network under analysis. Additional domain modules can also collect domain data from another respective domain.

[0259] The coordinator module 155 is configured to work with various machine learning algorithms and relational mechanisms to i) assess, ii) annotate, and/or iii) position in a vector diagram, a directed graph, a relational database, etc., activity including events occurring, for example, in the first domain compared to activity including events occurring

in the second domain. The domain modules can cooperate to exchange and store their information with the data store.

[0260] The process identifier classifier (not shown) in the gather module 110 can cooperate with additional classifiers in each of the domain modules 145/150 to assist in tracking individual processes and associating them with entities in a domain under analysis as well as individual processes and how they relate to each other. The process identifier classifier can cooperate with other trained AI classifiers in the modules to supply useful metadata along with helping to make logical nexuses.

[0261] A feedback loop of cooperation exists between the gather module 110, the analyzer module 115, AI model(s) 160 trained on different aspects of this process, and the cyber threat analyst module 120 to gather information to determine whether a cyber threat is potentially attacking the networks/domains under analysis.

[0262] **Determination of whether something is likely malicious**

[0263] In the following examples the analyzer module 115 and/or cyber threat analyst module 120 can use multiple factors to the determination of whether a process, event, object, entity, etc. is likely malicious.

[0264] In an example, the analyzer module 115 and/or cyber threat analyst module 120 can cooperate with one or more of the AI model(s) 160 trained on certain cyber threats to detect whether the anomalous activity detected, such as suspicious email messages, exhibit traits that may suggest a malicious intent, such as phishing links, scam language, sent from suspicious domains, etc. The analyzer module 115 and/or cyber threat analyst module 120 can also cooperate with one or more of the AI model(s) 160 trained on potential IT based cyber threats to detect whether the anomalous activity detected, such as suspicious IT links, URLs, domains, user activity, etc., may suggest a malicious intent as indicated by the AI models trained on potential IT based cyber threats.

[0265] In the above example, the analyzer module 115 and/or the cyber threat analyst module 120 can cooperate with the one or more AI models 160 trained with machine learning on the normal pattern of life for entities in an email domain under analysis to detect, in this example, anomalous emails which are detected as outside of the usual

pattern of life for each entity, such as a user, email server, etc., of the email network/domain. Likewise, the analyzer module 115 and/or the cyber threat analyst module 120 can cooperate with the one or more AI models trained with machine learning on the normal pattern of life for entities in a second domain under analysis (in this example, an IT network) to detect, in this example, anomalous network activity by user and/or devices in the network, which is detected as outside of the usual pattern of life (e.g. abnormal) for each entity, such as a user or a device, of the second domain's network under analysis.

[0266] Thus, the analyzer module 115 and/or the cyber threat analyst module 120 can be configured with one or more data analysis processes to cooperate with the one or more of the AI model(s) 160 trained with machine learning on the normal pattern of life in the system, to identify an anomaly of at least one of i) the abnormal behavior, ii) the suspicious activity, and iii) the combination of both, from one or more entities in the system. Note, other sources, such as other model breaches, can also identify at least one of i) the abnormal behavior, ii) the suspicious activity, and iii) the combination of both to trigger the investigation.

[0267] Accordingly, during this cyber threat determination process, the analyzer module 115 and/or the cyber threat analyst module 120 can also use AI classifiers that look at the features and determine a potential maliciousness based on commonality or overlap with known characteristics of malicious processes/entities. Many factors including anomalies that include unusual and suspicious behavior, and other indicators of processes and events are examined by the one or more AI models 160 trained on potential cyber threats and/or the AI classifiers looking at specific features for their malicious nature in order to make a determination of whether an individual factor and/or whether a chain of anomalies is determined to be likely malicious.

[0268] Initially, in this example of activity in an IT network analysis, the rare JA3 hash and/or rare user agent connections for this network coming from a new or unusual process are factored just like in the first wireless domain suspicious wireless signals are considered. These are quickly determined by referencing the one or more of the AI model(s) 160 trained with machine learning on the pattern of life of each device and its

associated processes in the system. Next, the analyzer module 115 and/or the cyber threat analyst module 120 can have an external input to ingest threat intelligence from other devices in the network cooperating with the cyber security appliance 100. Next, the analyzer module 115 and/or the cyber threat analyst module 120 can look for other anomalies, such as model breaches, while the AI models trained on potential cyber threats can assist in examining and factoring other anomalies that have occurred over a given timeframe to see if a correlation exists between a series of two or more anomalies occurring within that time frame.

[0269] The analyzer module 115 and/or the cyber threat analyst module 120 can combine these Indicators of Compromise (e.g., unusual network JA3, unusual device JA3, ...) with many other weak indicators to detect the earliest signs of an emerging threat, including previously unknown threats, without using strict blacklists or hard-coded thresholds. However, the AI classifiers can also routinely look at blacklists, etc. to identify maliciousness of features looked at.

[0270] Another example of features may include a deeper analysis of endpoint data. This endpoint data may include domain metadata, which can reveal peculiarities such as one or more indicators of potentially a malicious domain (i.e., its URL). The deeper analysis may assist in confirming an analysis to determine that indeed a cyber threat has been detected. The analyzer module 115 can also look at factors of how rare the endpoint connection is, how old the endpoint is, where geographically the endpoint is located, how a security certificate associated with a communication is verified only by an endpoint device or by an external 3rd party, just to name a few additional factors. The analyzer module 115 (and similarly the cyber threat analyst module 120) can then assign weighting given to these factors in the machine learning that can be supervised based on how strongly that characteristic has been found to match up to actual malicious sites in the training.

[0271] In another AI classifier to find potentially malicious indicators, the agent analyzer data analysis process in the analyzer module 115 and/or cyber threat analyst module 120 may cooperate with the process identifier classifier to identify all of the additional factors of i) are one or more processes running independently of other processes, ii) are the one

or more processes running independent are recent to this network, and iii) are the one or more processes running independent connect to the endpoint, which the endpoint is a rare connection for this network, which are referenced and compared to one or more AI models trained with machine learning on the normal behavior of the pattern of life of the system.

[0272] Note, a user agent, such as a browser, can act as a client in a network protocol used in communications within a client–server distributed computing system. In particular, the Hypertext Transfer Protocol (HTTP) identifies the client software originating (an example user agent) the request, using a user-agent header, even when the client is not operated by a user. Note, this identification can be faked, so it is only a weak indicator of the software on its own, but when compared to other observed user agents on the device, this can be used to identify possible software processes responsible for requests.

[0273] The analyzer module 115 and/or the cyber threat analyst module 120 may use the agent analyzer data analysis process that detects a potentially malicious agent previously unknown to the system to start an investigation on one or more possible cyber threat hypotheses. The determination and output of this step is what are possible cyber threats that can include or be indicated by the identified abnormal behavior and/or identified suspicious activity identified by the agent analyzer data analysis process.

[0274] In an example, the cyber threat analyst module 120 can use the agent analyzer data analysis process and the AI models(s) trained on forming and investigating hypotheses on what are a possible set of cyber threats to use the machine learning and/or set scripts to aid in forming one or more hypotheses to support or refute each hypothesis. The cyber threat analyst module 120 can cooperate with the AI models trained on forming and investigating hypotheses to form an initial set of possible hypotheses, which needs to be intelligently filtered down. The cyber threat analyst module 120 can be configured to use the one or more supervised machine learning models trained on i) agnostic examples of a past history of detection of a multitude of possible types of cyber threat hypotheses previously analyzed by human, who was a cyber security professional, ii) a behavior and input of how a plurality of human cyber security analysts make a decision and analyze a risk level regarding and a probability of a potential cyber threat, iii) steps

to take to conduct an investigation start with anomaly via learning how expert humans tackle investigations into specific real and synthesized cyber threats and then the steps taken by the human cyber security professional to narrow down and identify a potential cyber threat, and iv) what type of data and metrics that were helpful to further support or refute each of the types of cyber threats, in order to determine a likelihood of whether the abnormal behavior and/or suspicious activity is either i) malicious or ii) benign?

[0275] The cyber threat analyst module 120 using AI models, scripts and/or rules based modules is configured to conduct initial investigations regarding the anomaly of interest, collected additional information to form a chain of potentially related/linked information under analysis and then form one or more hypotheses that could have this chain of information that is potentially related/linked under analysis and then gather additional information in order to refute or support each of the one or more hypotheses.

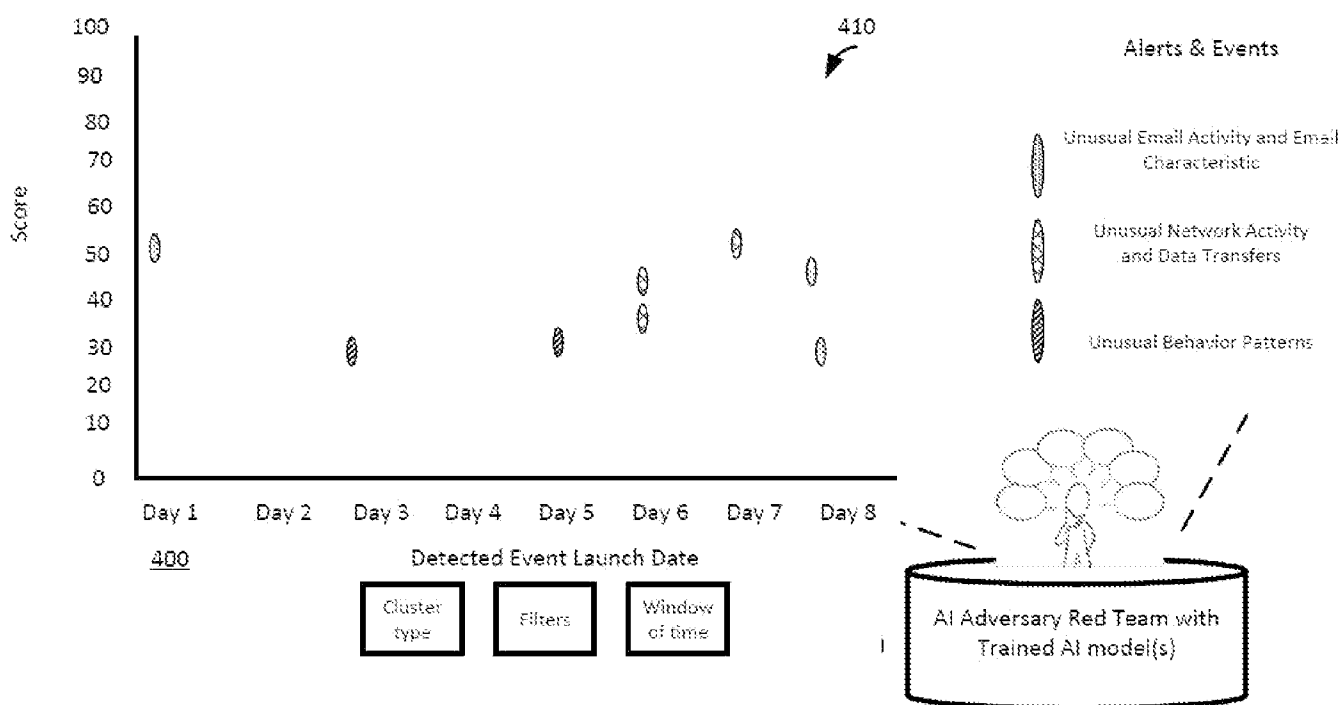


Figure 3

[0276] In an example, a behavioral pattern analysis for identifying what are the unusual behaviors of the network/system/device/user under analysis by the AI (machine learning) models may be as follows. The coordinator module 155 can tie the alerts, activities, and events from, in this example, the email domain to the alerts, activities, and events from

the IT network domain. As shown in Figure 3, a graph 410 of an embodiment of an example chain of unusual behavior for, in this example, the email activities as well as IT activities deviating from a normal pattern of life for this user and/or device in connection with the rest of the system/network under analysis. The cyber threat analyst module 120 and/or analyzer module 115 can cooperate with one or more AI (machine learning) models. The one or more AI (machine learning) models are trained and otherwise configured with mathematical algorithms to infer, for the cyber-threat analysis, 'what is possibly happening with the chain of distinct alerts, activities, and/or events, which came from the unusual pattern,' and then assign a threat risk associated with that distinct item of the chain of alerts and/or events forming the unusual pattern. The unusual pattern can be determined by examining initially what activities/events/alerts that do not fall within the window of what is the normal pattern of life for that network/system/device/user under analysis can be analysed to determine whether that activity is unusual or suspicious. A chain of related activity that can include both unusual activity and activity within a pattern of normal life for that entity can be formed and checked against individual cyber threat hypothesis to determine whether that pattern is indicative of a behavior of a malicious actor – human, program, or other threat. The cyber threat analyst module 120 can go back and pull in some of the normal activities to help support or refute a possible hypothesis of whether that pattern is indicative of a behavior of a malicious actor.

[0277] An illustrative example of a behavioral pattern included in the chain is shown in the graph over a time frame of, an example, 7 days. The cyber threat analyst module 120 detects a chain of anomalous behavior of unusual data transfers three times, unusual characteristics in email messages in the monitored system three times which seem to have some causal link to the unusual data transfers. Likewise, twice unusual credentials attempted the unusual behavior of trying to gain access to sensitive areas or malicious IP addresses and the user associated with the unusual credentials trying unusual behavior has a causal link to at least one of those three email messages with unusual characteristics. Again, the cyber security appliance 100 can go back and pull in some of the normal activities to help support or refute a possible hypothesis of whether that pattern is indicative of a behavior of a malicious actor. The analyzer module 115 of Figure 1 can cooperate with one or more models trained on cyber threats and their behavior to try to

determine if a potential cyber threat is causing these unusual behaviors. The cyber threat analyst module 120 can put data and entities into 1) a directed graph and nodes in that graph that are overlapping or close in distance have a good possibility of being related in some manner, 2) a vector diagram, 3) relational database, and 4) other relational techniques that will at least be examined to assist in creating the chain of related activity connected by causal links, such as similar time, similar entity and/or type of entity involved, similar activity, etc., under analysis. If the pattern of behaviors under analysis is believed to be indicative of a malicious actor, then a score of how confident is the system in this assessment of identifying whether the unusual pattern was caused by a malicious actor is created. Next, also assigned is a threat level score or probability indicative of what level of threat does this malicious actor pose. Lastly, the cyber security appliance 100 is configurable in a user interface, by a user, enabling what type of automatic response actions, if any, the cyber security appliance 100 may take when different types of cyber threats, indicated by the pattern of behaviors under analysis, that are equal to or above a configurable level of threat posed by this malicious actor.

[0278] Referring still to Figure 1, the autonomous response module 140 is configured to take one or more autonomous mitigation actions to mitigate the cyber threat during the cyberattack by the cyber threat. The autonomous response module 140 can reference an AI model trained to track a normal pattern of life for each node of the protected system to perform an autonomous act of, for example, restricting a potentially compromised node having i) an actual indication of compromise and/or ii) merely adjacent to a known compromised node, to merely take actions that are within that node's normal pattern of life to mitigate the cyber threat.

[0279] The chain of the individual alerts, activities, and events that form the pattern including one or more unusual or suspicious activities into a distinct item for cyber-threat analysis of that chain of distinct alerts, activities, and/or events. The cyber-threat module may reference the one or more machine learning models trained on, in this example, e-mail threats to identify similar characteristics from the individual alerts and/or events forming the distinct item made up of the chain of alerts and/or events forming the unusual pattern.

[0280] Cyber Threat Assessment and Autonomous Actions

[0281] In the next step, the analyzer module 115 and/or cyber threat analyst module 120 generates one or more supported possible cyber threat hypotheses from the possible set of cyber threat hypotheses. The analyzer module 115 generates the supporting data and details of why each individual hypothesis is supported or not. The analyzer module 115 can also generate one or more possible cyber threat hypotheses and the supporting data and details of why they were refuted.

[0282] In general, the analyzer module 115 cooperates with the following three sources. The analyzer module 115 cooperates with the one or more of the AI model(s) 160 trained on cyber threats to determine whether an anomaly such as the abnormal behavior and/or suspicious activity is either 1) malicious or 2) benign when the potential cyber threat under analysis is previously unknown to the cyber security appliance 100. The analyzer module 115 cooperates with one or more of the AI model(s) 160 trained on a normal pattern of life of entities in the network under analysis. The analyzer module 115 cooperates with various AI-trained classifiers. With all of these sources, when they input information that indicates a potential cyber threat that is i) severe enough to cause real harm to the network under analysis and/or ii) a close match to known cyber threats, then the analyzer module can make a final determination to confirm that a cyber threat likely exists and send that cyber threat to the assessment module to assess the threat score associated with that cyber threat. Certain model breaches will always trigger a potential cyber threat that the analyzer will compare and confirm the cyber threat.

[0283] In the next step, the assessment module 125 with the AI classifiers is configured to cooperate with the analyzer module 115. The analyzer module 115 supplies the identity of the supported possible cyber threat hypotheses from the possible set of cyber threat hypotheses to the assessment module 125. The assessment module 125 with the AI classifiers cooperates with the one or more of the AI model(s) 160 trained on possible cyber threats can make a determination on whether a cyber threat exists and what level of severity is associated with that cyber threat. The assessment module 125 with the AI classifiers cooperates with one or more of the AI model(s) 160 trained on possible cyber threats in order assign a numerical assessment of a given cyber threat hypothesis that

was found likely to be supported by the analyzer module 115 with the one or more data analysis processes, via the abnormal behavior, the suspicious activity, or the collection of system data points. The assessment module 125 with the AI classifiers output can be a score (ranked number system, probability, etc.) that a given identified process is likely a malicious process. The assessment module 125 with the AI classifiers can be configured to assign a numerical assessment, such as a probability, of a given cyber threat hypothesis that is supported and a threat level posed by that cyber threat hypothesis which was found likely to be supported by the analyzer module 115, which includes the abnormal behavior or suspicious activity as well as one or more of the collection of system data points, with the one or more AI models trained on possible cyber threats.

[0284] The cyber threat analyst module 120 in the AI-based cyber security appliance 100 component provides an advantage over competitors' products as it reduces the time taken for cyber security investigations, provides an alternative to manpower for small organizations and improves detection (and remediation) capabilities within the cyber security appliance 100. The AI-based, cyber threat analyst module 120 performs its own computation of threat and identifies interesting network events with one or more processors. These methods of detection and identification of threat all add to the above capabilities that make the cyber threat analyst module 120 a desirable part of the cyber security appliance 100. The cyber threat analyst module 120 offers a method of prioritizing which is not just a summary or highest score alert of an event evaluated by itself equals the worst and prevents more complex attacks being missed because their composite parts/individual threats only produced low-level alerts.

[0285] The AI classifiers can be part of the assessment module 125, which scores the outputs of the analyzer module 115. Again, as for the other AI classifiers discussed, the AI classifier can be coded to take in multiple pieces of information about an entity, object, and/or thing and based on its training and then output a prediction about the entity, object, or thing. Given one or more inputs, the AI classifier model will try to predict the value of one or more outcomes. The AI classifiers cooperate with the range of data analysis processes that produce features for the AI classifiers. The various techniques cooperating here allow anomaly detection and assessment of a cyber threat level posed

by a given anomaly; but more importantly, an overall cyber threat level posed by a series/chain of correlated anomalies under analysis.

[0286] In the next step, the formatting module 130 can generate an output such as a printed or electronic report with the relevant data. The formatting module 130 can cooperate with both the analyzer module 115 and the assessment module 125 depending on what the user wants to be reported. The formatting module 130 is configured to format, present a rank for, and output one or more supported possible cyber threat hypotheses from the assessment module into a formalized report, from a one or more report templates populated with the data for that incident. The formatting module 130 is configured to format, present a rank for, and output one or more detected cyber threats from the analyzer module or from the assessment module into a formalized report, from a one or more report templates populated with the data for that incident. Many different types of formalized report templates exist to be populated with data and can be outputted in an easily understandable format for a human user's consumption. The formalized report on the template is outputted for a human user's consumption in a medium of any of 1) printable report, 2) presented digitally on a user interface, 3) in a machine-readable format for further use in machine-learning reinforcement and refinement, or 4) any combination of the three. The formatting module 130 is further configured to generate a textual write up of an incident report in the formalized report for a wide range of breaches of normal behavior, used by the AI models trained with machine learning on the normal behavior of the system, based on analyzing previous reports with one or more models trained with machine learning on assessing and populating relevant data into the incident report corresponding to each possible cyber threat. The formatting module 130 can generate a threat incident report in the formalized report from a multitude of a dynamic human-supplied and/or machine created templates corresponding to different types of cyber threats, each template corresponding to different types of cyber threats that vary in format, style, and standard fields in the multitude of templates. The formatting module 130 can populate a given template with relevant data, graphs, or other information as appropriate in various specified fields, along with a ranking of a likelihood of whether that hypothesis cyber threat is supported and its threat severity level for each of the supported cyber threat hypotheses, and then output the formatted threat incident report with the

ranking of each supported cyber threat hypothesis, which is presented digitally on the user interface and/or printed as the printable report.

[0287] In the next step, the assessment module 125 with the AI classifiers, once armed with the knowledge that malicious activity is likely occurring/is associated with a given process from the analyzer module 115, then cooperates with the autonomous response module 140 to take an autonomous action such as i) deny access in or out of the device or the network ii) shutdown activities involving a detected malicious agent, iii) restrict devices and/or user's to merely operate within their particular normal pattern of life, iv) remove some user privileges/permissions associated with the compromised user account, etc.

[0288] The autonomous response module 140, rather than a human taking an action, can be configured to cause one or more rapid autonomous actions in response to be taken to counter the cyber threat. A user interface for the response module can program the autonomous response module 140 i) to merely make a suggested response to take to counter the cyber threat that will be presented on a display screen and/or sent by a notice to an enterprise security administrator for explicit authorization when the cyber threat is detected or ii) to autonomously take a response to counter the cyber threat without a need for a human to approve the response when the cyber threat is detected. The autonomous response module 140 will then send a notice of the autonomous response as well as display the autonomous response taken on the display screen. Example autonomous responses may include cut off connections, shutdown devices, change the privileges of users, delete and remove malicious links in emails, slow down a transfer rate, cooperate with other security devices such as a firewall to trigger its autonomous actions, and other autonomous actions against the devices and/or users. The autonomous response module 140 uses one or more of the AI model(s) 160 that are configured to intelligently work with other third-party defense systems in that customer's network against threats. The autonomous response module 140 can send its own protocol commands to devices and/or take actions on its own. In addition, the autonomous response module 140 uses the one or more of the AI model(s) 160 to orchestrate with other third-party defense systems to create a unified defense response against a detected threat within or external to that customer's network. The autonomous response module 140 can be an

autonomous self-learning digital response coordinator that is trained specifically to control and reconfigure the actions of traditional legacy computer defenses (e.g., firewalls, switches, proxy servers, etc.) to contain threats propagated by, or enabled by, networks and the internet. The cyber threat analyst module 120 and/or assessment module 125 can cooperate with the autonomous response module 140 to cause one or more autonomous actions in response to be taken to counter the cyber threat, improves computing devices in the system by limiting an impact of the cyber threat from consuming unauthorized CPU cycles, memory space, and power consumption in the computing devices via responding to the cyber threat without waiting for some human intervention. The trigger module 105, analyzer module 115, assessment module 125, the cyber threat analyst module 120, and formatting module 130 cooperate to improve the analysis and formalized report generation with less repetition to consume CPU cycles with greater efficiency than humans repetitively going through these steps and re-duplicating steps to filter and rank the one or more supported possible cyber threat hypotheses from the possible set of cyber threat hypotheses.

[0289] Figure 2 illustrates an embodiment of the AI based cyber security appliance 100 plugging in as an appliance platform to protect a system. The cyber security appliance 100 is part of an enterprise network 230, which may further include one or more computing devices 240 such as database servers 250, web servers 260, networking devices 270 (e.g., bridge, switch, router, load-balancers, gateways, and/or firewalls), endpoint devices with connectivity to resources within the enterprise network 230 as well as a publicly accessible network 280 (e.g., the Internet). The endpoint devices 270 may include, but are not limited or restricted to desktop computers, laptops, smart phones, tablets, wearables, smart appliances, or the like. The security controls operate as probes and detectors that are configured to monitor, for example, network-based activity (e.g., email activity, TCP/IP communications, text or Short Message Service (SMS) activity, etc.) and computing device activity (e.g., download activity based on volume, day, time of day, etc.); credential update/modification activity (e.g., credential changes, failed access attempts to a resource, etc.); and/or resource activity (e.g., attempted/successful accesses to enterprise resources, etc.). The security controls provide the monitored data (or a version of the monitored data) as input into the modules of the cyber security appliance 100 to determine what is occurring in each domain individually.

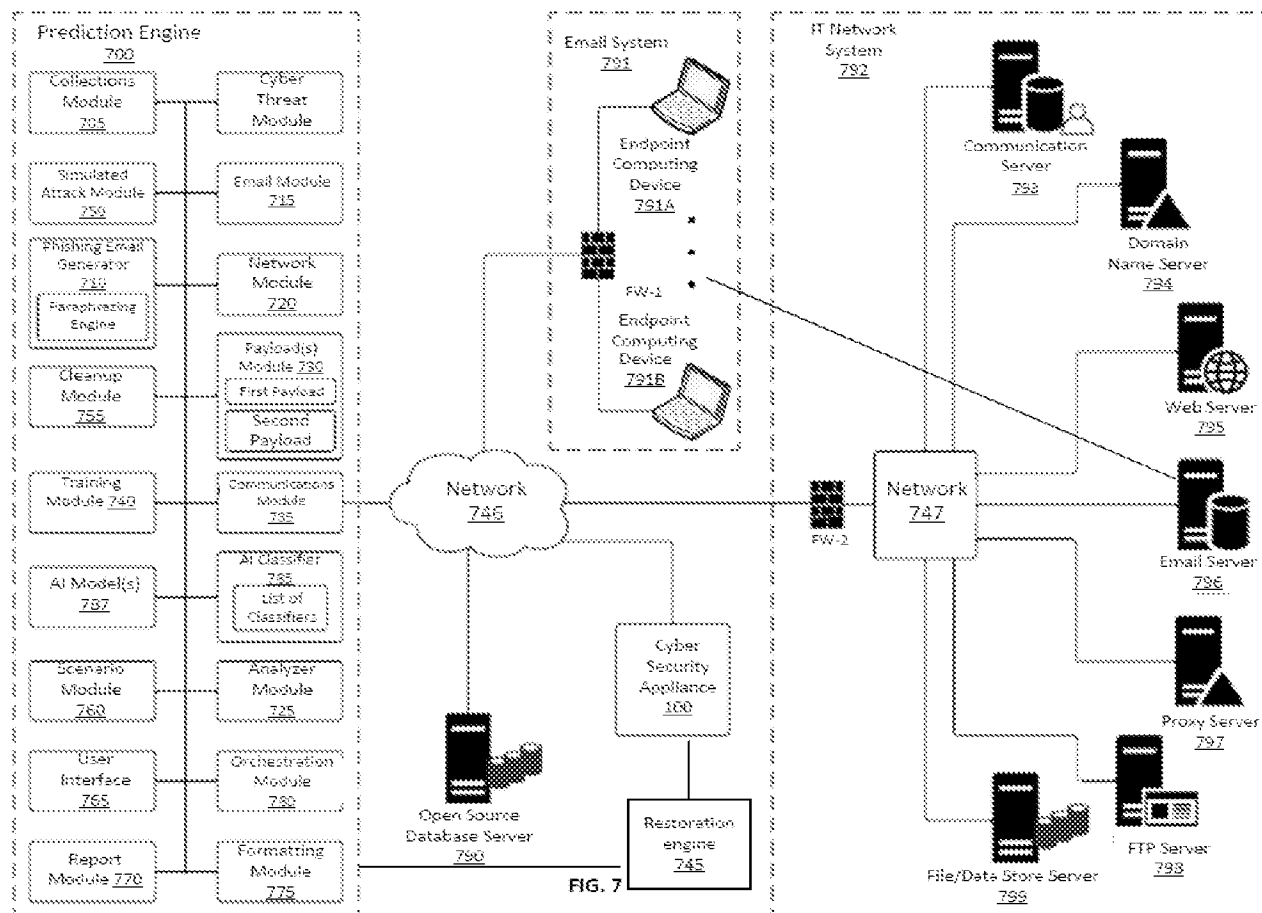


FIG. 7

[0290] PREDICTION ENGINE AND RESTORATION ENGINE

[0291] Overall, the cyber security appliance 100 and its modules use Artificial Intelligence algorithms configured and trained to perform a first machine-learned task of detecting the cyber threat as well as the autonomous response module 140 can use a combination of user configurable settings on actions to take to mitigate a detected cyber threat, a default set of actions to take to mitigate a detected cyber threat, and Artificial Intelligence algorithms configured and trained to perform a second machine-learned task of taking one or more mitigation actions to mitigate the cyber threat. A cyber security restoration engine 190 deployed in the cyber security appliance 100 uses Artificial Intelligence algorithms configured and trained to perform a third machine-learned task of remediating the system/network being protected back to a trusted operational state. The prediction engine 700 of Figure 7 uses Artificial Intelligence algorithms configured and trained to perform a fourth machine-learned task of Artificial Intelligence-based simulations of

cyberattacks to assist in determining 1) how a simulated cyberattack might occur in the system being protected, and 2) how to use the simulated cyberattack information to preempt possible escalations of an ongoing actual cyberattack.

[0292] Referring now to Figure 7, an exemplary block diagram of embodiment of the prediction engine 700 is shown. The prediction engine 700 conducts Artificial Intelligence-based simulations by constructing a graph of nodes of the system being protected (e.g., a network including (a) the physical devices connecting to the network, any virtualized instances of the network, user accounts in the network, email accounts in the network, etc. as well as (b) connections and pathways through the network) to create a virtualized instance of the network to be tested. As shown in Figure 7, the various cooperating modules residing in the prediction engine 700 may include, but are not limited to, a collections module 705, a cyberattack generator (e.g. phishing email generator) 710, an email module 715, a network module 720, an analyzer module 725, a payloads module 730 with first and second payloads, a communication module 735, a training module 740, a simulated attack module 750, a cleanup module 744, a scenario module 760, a user interface 765, a reporting module 770, a formatting module 775, an orchestration module 780, and/or an AI classifier 785 with a list of specified classifiers.

[0293] The simulated attack module 750 in the prediction engine 700 may be implemented via i) a simulator to model the system being protected and/or ii) a clone creator to spin up a virtual network and create a virtual clone of the system being protected configured to pen-test one or more defenses provided by the cyber security appliance 100. The prediction engine 700 may include and cooperate with one or more AI models 787 trained with machine learning on the contextual knowledge of the organization, such as those in the cyber security appliance 100 or have its own separate model trained with machine learning on the contextual knowledge of the organization and each user's and device's normal pattern of behavior. These trained AI models 787 may be configured to identify data points from the contextual knowledge of the organization and its entities, which may include, but is not limited to, language-based data, email/network connectivity and behavior pattern data, and/or historic knowledgebase data. The prediction engine 700 may use the trained AI models 787 to cooperate with

one or more AI classifier(s) 785 by producing a list of specific organization-based classifiers for the AI classifier(s) 785.

[0294] The simulated attack module 750 by cooperating with the other modules in the prediction engine 700 is further configured to calculate and run one or more hypothetical simulations of a possible cyberattack and/or of an actual ongoing cyberattack from a cyber threat through an attack pathway through the system being protected. The prediction engine 700 is further configured to calculate, based at least in part on the results of the one or more hypothetical simulations of a possible cyberattack and/or of an actual ongoing cyberattack from a cyber threat through an attack pathway through the system being protected, a risk score for each node (e.g. each device, user account, etc.), the threat risk score being indicative of a possible severity of the compromise and /or chance of compromise prior to an autonomous response action is taken in response to an actual cyberattack of the cyber incident.

[0295] The simulated attack module 750 is configured to initially create the network being protected in a simulated or virtual device environment. Additionally, the orchestration module 780 and communications module 735 may be configured to cooperate with the cyber security appliance 100 to securely obtain specific data about specific users, devices, and entities in specific networks of for this specific organization. The training module 740 and simulated attack module 750 in the prediction engine 700 use the obtained specific data to generate one or more specific cyberattacks, such as a phishing email, tailored to those specific users, devices, and/or entities of the specific organization. Many different cyberattacks can be simulated by the AI red team module but a phishing email attack will be used as an example cyberattack.

[0296] The prediction engine 700 is communicatively coupled to the cyber security appliance 100, an open source (OS) database server 790, an email system 791 with one or more endpoint computing devices 791A-B, and a network system 792 with one or more entities 793-799, and a restoration engine 745 over one or more networks 746/747. The cyber security appliance 100 may cooperate with the prediction engine 700 to initiate a pen-test in the form of, for example, a software attack, which generates a customized, for example, phishing email to spoof one or more specific users/devices/entities of an

organization in an email/network defense system and then looks for any security vulnerabilities, risks, threats, and/or weaknesses potentially gaining access to one or more features and data of that specific user/device/entity.

[0297] The prediction engine 700 may be customized and/or driven by a centralized AI using and/or modelling a smart awareness of a variety of specific historical email/network behavior patterns and communications of a specific organization's hierarchy within a specific organization. Such AI modelling may be trained and derived through machine learning and the understanding of the organization itself based on: (i) a variety of OS materials such as any OS materials collected from the OS database server 790 and (ii) its historical awareness of any specific email/network connectivity and behavior patterns to target for that organization as part of an offensive (or attacking) security approach. The training module 740 can contain for reference a database of cyberattack scenarios as well as restoration response scenarios by the restoration engine 745 stored in the database.

[0298] The prediction engine 700 may use the orchestration module 780 to implement and orchestrate this offensive approach all the way from an initial social engineering attack at an earlier stage of the pentest to a subsequent payload delivery attack at a later stage of the pentest and so on. The prediction engine 700 is configured to: (i) intelligently initiate a customized cyberattack on the components, for example, in the IT network and email system 791; as well as (ii) subsequently generating a report to highlight and/or raise awareness of one or more key areas of vulnerabilities and/or risks for that organization after observing the intelligently initiated attack (e.g., such key areas may be formatted and reported in a way tailored for that organization using both the formatting and reporting modules, as described below); and (iii) then allow that enterprise (e.g., organization) to be trained on that attack and its impact on those specific security postures, thereby allowing that organization to go in directly to mitigate and improve those compromised security postures going forward, as well as iv) during an actual cyberattack, obtain and ingest data known on the cyberattack, run simulations, and then supply information, for example, to the autonomous response module in the cyber security appliance to mitigate the actual cyberattack.

[0299] The prediction engine 700 may cooperate with the cyber security appliance 100 to provide feedback on any successful attacks and detections. For example, in the event that the prediction engine 700 is successful in pentesting any of the organization's entities in the email and network defense systems 791/792, the prediction engine 700 may be configured to at least provide the cyber security appliance 100 (and/or any other predetermined entities) with any feedback on the successful pentest as well as any specifics regarding the processes used for that successful pentest, such as providing feedback on the specific attack vectors, scenarios, targeted entities, characteristics of the customized phishing emails, payloads, and contextual data, etc., that were used.

[0300] The simulated attack module 750 in the prediction engine 700 may be configured with an attack path modeling component (not shown), which is programmed to work out the key paths and devices in a network via running cyberattacks on a simulated or virtual device version of the network under analysis incorporating metrics that feed into that modeling by running simulated cyberattacks on the particulars known about this specific network being protected by the cyber security appliance 100. The attack modeling has been programmed with the knowledge of a layout and connection pattern of each particular network device in a network and a number of connections and/or hops to other network devices in the network. Also, how important a particular device (a key importance) can be determined by the function of that network device, the user(s) associated with that network device, the location of the device within the network and a number of connections and/or hops to other important devices in the network. The attack path modeling component ingests the information for the purposes of modeling and simulating a potential attack against the network and routes that an attacker would take through the network. The attack path modeling component can be constructed with information to i) understand an importance of network nodes in the network compared to other network nodes in the network, and ii) to determine key pathways within the network and vulnerable network nodes in the network that a cyberattack would use during the cyberattack, via modeling the cyberattack on at least one of 1) a simulated device version and 2) a virtual device version of the network under analysis.

[0301] Figure 8 illustrates a diagram of an embodiment of the cyber security appliance and its Artificial Intelligence-based simulations constructing an example graph of nodes

in an example network and simulating how the cyberattack might likely progress in the future tailored with an innate understanding of a normal behavior of the nodes in the system being protected and a current operational state of each node in the graph of the protected system during simulations of cyberattacks. The prediction engine 700 plots the attack path through the nodes and estimated times to reach critical nodes in the network. The cyberattack simulation modeling is run to identify the routes, difficulty, and time periods from certain entry nodes to certain key servers. The simulations of the cyber attack by the cyber threat from one compromised entry point such as Device n to a key server can take 10 days, 100 days, etc. depending on normal behavior of Device n, security settings of the network devices in the different nodes of the network, etc.

[0302] The attack path modeling component in the simulated attack module 750 cooperating with the other modules in the prediction engine 700 are configured to determine the key pathways within the network and the vulnerable network nodes in the network that the cyberattack would use during the cyberattack, via the modeling of the cyberattack on at least one of 1) the simulated device version and 2) the virtual device version of the network under analysis via using the actual detected vulnerabilities of each network node, a predicted frequency of remediation of those vulnerabilities within a specific network device in the network without a notice from the restoration engine 745, and an importance of the key network nodes with the actual vulnerabilities compared to other network nodes in the network.

[0303] The modules essentially seed the attack path modeling component with weakness scores that provide current data, customized to each user account and/or network device, which then allows the artificial intelligence running the attack path simulation to choose entry network nodes into the network with more accuracy as well as plot the attack path through the nodes and estimated times to reach critical nodes in the network much more accurately based on the actual current operational condition of the many user accounts and network devices in the network. The attack simulation modeling can be run to identify the routes, difficulty, and time periods from certain entry nodes to certain key servers.

[0304] Note, the cyber threat analyst module 120 in the cyber security appliance 100 of Figure 1 as well as the prediction engine 700 of Figure 7 may use any unusual, detected

behavior deviating from the normal behavior and then build a sequence/chain of unusual behavior and the causal links between the sequence/chain of unusual behavior to detect any potential cyber threats. For example, as shown in Figure 4, the cyber security appliance 100 and the prediction engine 700 may determine the unusual patterns by analyzing i) what activities/events/alerts that fall outside of the window of what is the normal pattern of life for that network/system/entity/device/user under analysis; and (ii) then pulling in and analyzing the pattern of the behavior of the activities/events/alerts that are in the normal pattern of life but also connect to the indicators for a possible cyberattack, to determine whether that pattern is indicative of a behavior of a malicious actor, such as a human, program, and/or any other cyber harmful threat.

[0305] The prediction engine 700 and its Artificial Intelligence-based simulations use artificial intelligence to cooperate with the restoration engine 745 to assist in choosing one or more remediation actions to perform on nodes affected by the cyberattack back to a trusted operational state while still mitigating the cyber threat during an ongoing cyberattack based on effects determined through the simulation of possible remediation actions to perform and their effects on the nodes making up the system being protected and preempt possible escalations of the cyberattack while restoring one or more nodes back to a trusted operational state. Thus, for example, the restoration engine 745 restores the one or more nodes in the protected system by cooperating with any of 1) an AI model trained to model a normal pattern of life for each node in the protected system, 2) an AI model trained on what are a possible set of cyber threats and their characteristics and symptoms to identify the cyber threat (e.g. malicious actor/device/file) that is causing a particular node to behave abnormally (e.g. malicious behavior) and fall outside of that node's normal pattern of life, and 3) the autonomous response module 140.

[0306] The restoration engine 745 can reference both i) a database of restoration response scenarios stored in the database and ii) a prediction engine 700 configured to run AI-based simulations and use the operational state of each node in the graph of the protected system during simulations of cyberattacks on the protected system to restore 1) each node compromised by the cyber threat and 2) promote protection of the corresponding nodes adjacent to a compromised node in the graph of the protected system.

[0307] The restoration engine 745 can prioritize among the one or more nodes to restore, which nodes to remediate and an order of the nodes to remediate, based on two or more factors including i) a dependency order needed for the recovery efforts, ii) an importance of a particular recovered node compared to other nodes in the system being protected, iii) a level of compromise of a particular node contemplated to be restored, iv) an urgency to recover that node compared to whether containment of the cyber threat was successful, v) a list of a most important things in the protected system to recover earliest, and vi) factoring in a result of a cyberattack simulation being run during the cyberattack by the prediction engine 700 to predict a likely result regarding the cyberattack when that node is restored.

[0308] An interactive response loop exists between the restoration engine 745, the cyber security appliance 100, and the prediction engine 700. The restoration engine 745, the cyber security appliance 100, and the prediction engine 700 can be configured to cooperate to combine an understanding of normal operations of the nodes making up the devices and users in the system being protected by the cyber security appliance 100, an understanding emerging cyber threats, an ability to contain those emerging cyber threats, and a restoration of the nodes of the system to heal the system with an adaptive feedback between the multiple AI-based engines in light of simulations of the cyberattack to predict what might occur in the nodes in the system based on the progression of the attack so far, mitigation actions taken to contain those emerging cyber threats and remediation actions taken to heal the nodes using the simulated cyberattack information. The multiple AI-based engines have communication hooks in between them to exchange a significant amount of behavioral metrics including data between the multiple AI-based engines to work in together to provide an overall cyber threat response.

[0309] The cyber security appliance 100 and its modules use Artificial Intelligence algorithms configured and trained to perform a first machine-learned task of detecting the cyber threat as well as the autonomous response module 140 can use a combination of user configurable settings on actions to take to mitigate a detected cyber threat, a default set of actions to take to mitigate a detected cyber threat, and Artificial Intelligence algorithms configured and trained to perform a second machine-learned task of taking one or more mitigation actions to mitigate the cyber threat. The restoration engine 745

uses Artificial Intelligence algorithms configured and trained to perform a third machine-learned task of remediating the system/network being protected back to a trusted operational state. The prediction engine 700 uses Artificial Intelligence algorithms configured and trained to perform a fourth machine-learned task of AI-based simulations of cyberattacks to assist in determining 1) how a simulated cyberattack might occur in the system being protected, and 2) how to use the simulated cyberattack information to preempt possible escalations of an ongoing actual cyberattack. In an example, the autonomous response module 140 uses its intelligence to cooperate with the prediction engine 700 and its AI-based simulations to choose and initiate an initial set of one or more mitigation actions indicated as a preferred targeted initial response to the detected cyber threat by autonomously initiating those mitigation actions to defend against the detected cyber threat, rather than a human taking an action.

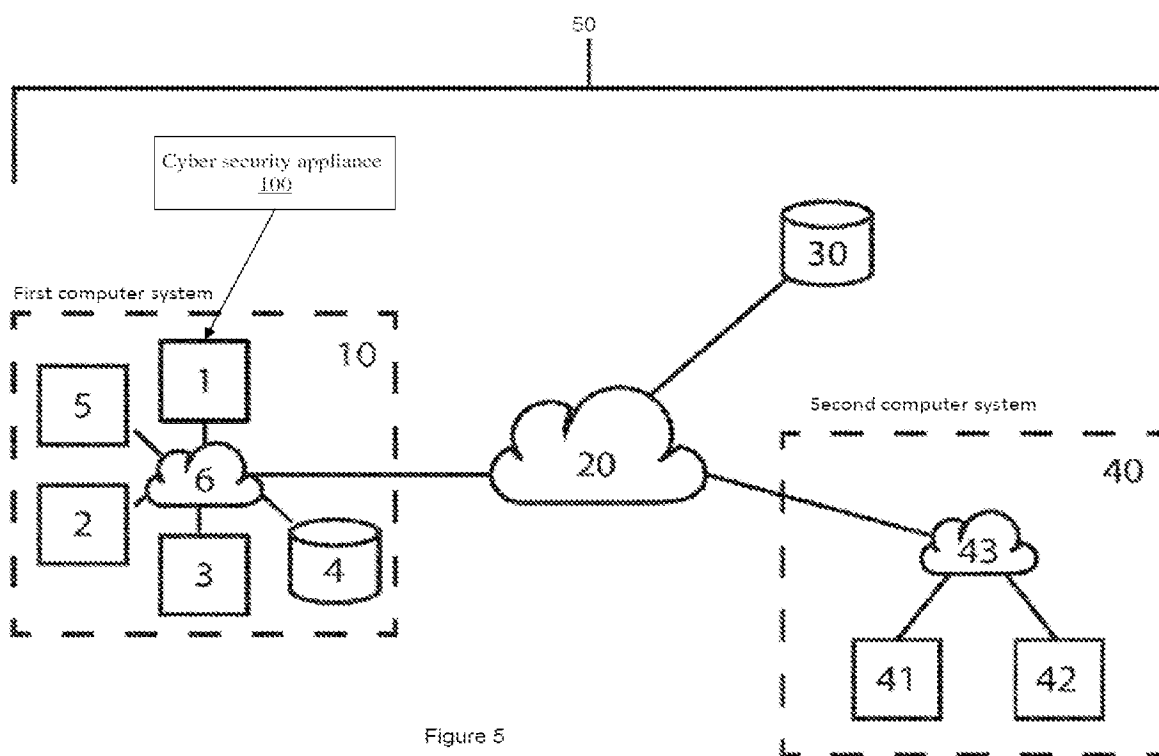


Figure 5

[0310] Figure 5 illustrates an example Artificial Intelligence based cyber security appliance 100 using a cyber threat analyst module 104 to protect an example network. The example network of computer systems 50 uses a cyber security appliance 100. The system depicted is a simplified illustration, which is provided for ease of explanation.

[0311] Referring to Figure 5, an exemplary and generalized embodiment of a system (e.g., enterprise network) 50 featuring computer systems 10 and 40, where one or more of these computer systems 10 and/or 40 may deploy the AI-based, cyber security appliance 100 of Figure 1 to protect the enterprise, is shown. Herein, the system 50 comprises a first computer system 10 within a building, which uses the threat detection system to detect and thereby attempt to prevent threats to computing devices within its bounds. The first computer system 10 comprises three computing devices 1, 2, 3, a local server 4, and a multifunctional device (MFD) 5 that provides printing, scanning and facsimile functionalities to each of the computers 1, 2, 3. All of the devices within the first computer system 10 are communicatively coupled via a first Local Area Network (LAN) 6. Consequently, all of the computing devices 1, 2, 3 are able to access the local server 4 via the first LAN 6 and use the functionalities of the MFD 5 via the LAN 6. The first LAN 6 of the first computer system 10 is connected to the Internet 20, which in turn provides computing devices 1, 2, 3 with access to a multitude of other computing devices, including a server 30 and a second computer system 40. The second computer system 40 also includes two computing devices 41, 42, connected by a second LAN 43.

[0312] In this exemplary embodiment of the cyber security appliance 100, a first computing device 1 on the first computer system 10 has the electronic hardware, modules, models, and various software processes of the cyber security appliance 100; and therefore, runs threat detection for detecting threats to the first computer system 10. As such, the first computing device 1 includes one or more processors arranged to run the steps of the process described herein, memory storage components required to store information related to the running of the process, as well as one or more network interfaces for collecting information from various security controls (e.g., sensors, probes, etc.) collecting data associated with the system (network) 50 under analysis.

[0313] The cyber security appliance 100 in the first computing device 1 builds and maintains a dynamic, ever-changing model of the 'normal behavior' of each user and machine within the first computer system 10. The approach is based on Bayesian mathematics, and monitors all interactions, events and communications within the first computer system 10 - which computing device is talking to which, files that have been created, networks that are being accessed. For example, a second computing device 2

is based in a company's San Francisco office and operated by a marketing employee who regularly accesses the marketing network, usually communicates with machines in the company's U.K. office in the second computer system 40 between 9.30 AM and midday, and is active from about 8:30 AM until 6 PM. The same employee virtually never accesses the employee time sheets, very rarely connects to the company's Atlanta network and has no dealings in South-East Asia. The cyber security appliance 100 takes all the information that is available relating to this employee and establishes a 'pattern of life' for that person and the devices used by that person in that system, which is dynamically updated as more information is gathered. The model of the normal pattern of life for an entity in the system 50 under analysis is used as a moving benchmark, allowing the cyber security appliance 100 to spot behavior on the system 50 seems to fall outside of this normal pattern of life, and flags this behavior as anomalous, requiring further investigation and/or autonomous action.

[0314] The cyber security appliance 100 is built to deal with the fact that today's attackers are getting stealthier, and an attacker/malicious agent may be 'hiding' in a system to ensure that they avoid raising suspicion in an end user, such as by slowing their machine down. The AI model(s) 160 in the cyber security appliance 100 builds a sophisticated 'pattern of life' – that understands what represents normality for every person, device, and network activity in the system being protected by the cyber security appliance 100. The self-learning algorithms in the AI can, for example, understand each node's (user account, device, etc.) in an organization's normal patterns of life in about a week, and grows more bespoke with every passing minute. Conventional AI typically relies solely on identifying threats based on historical attack data and reported techniques, requiring data to be cleansed, labelled, and moved to a centralized repository. The detection engine self-learning AI can learn "on the job" from real-world data occurring in the system and constantly evolves its understanding as the system's environment changes. The Artificial Intelligence can use machine learning algorithms to analyze patterns and 'learn' what is the 'normal behavior' of the system (network) 50 by analyzing data on the activity on the system 50 at the device and employee level. The unsupervised machine learning does not need humans to supervise the learning in the model but rather discovers hidden patterns or data groupings without the need for human intervention. The unsupervised

machine learning discovers the patterns and related information using the unlabeled data monitored in the system itself. Unsupervised learning algorithms can include clustering, anomaly detection, neural networks, etc. Unsupervised learning can break down features of what it is analyzing (e.g., a network node of a device or user account), which can be useful for categorization, and then identify what else has similar or overlapping feature sets matching to what it is analyzing.

[0315] The cyber security appliance 100 can use unsupervised machine learning to work things out without pre-defined labels. In the case of sorting a series of different entities, such as different devices, the system analyzes the information and works out the different classes of devices. This allows the system 50 to handle the unexpected and embrace uncertainty when new entities and classes are examined. The modules and models of the cyber security appliance 100 do not always know what they are looking for but can independently classify data and detect compelling patterns. The cyber security appliance's 100 unsupervised machine learning methods do not require training data with pre-defined labels. Instead, they are able to identify key patterns and trends in the data, without the need for human input. The advantage of unsupervised learning in this system is that it allows computers to go beyond what their programmers already know and discover previously unknown relationships. The unsupervised machine learning methods can use a probabilistic approach based on a Bayesian framework. The machine learning allows the cyber security appliance 100 to integrate a huge number of weak indicators/low threat values by themselves of potentially anomalous network behavior to produce a single clear overall measure of these correlated anomalies to determine how likely a network device is to be compromised. This probabilistic mathematical approach provides an ability to understand important information, amid the noise of the network – even when it does not know what it is looking for.

[0316] The cyber security appliance 100 can use a Recursive Bayesian Estimation to combine these multiple analyzes of different measures of network behavior to generate a single overall/comprehensive picture of the state of each device, the cyber security appliance 100 takes advantage of the power of Recursive Bayesian Estimation (RBE) via an implementation of the Bayes filter. Using RBE, the cyber security appliance's 100 AI models are able to constantly adapt themselves, in a computationally efficient manner,

as new information becomes available to the system. The AI model(s) of the cyber security appliance 100 may be configured to continually recalculate threat levels in the light of new evidence, identifying changing attack behaviors where conventional signature-based methods fall down.

[0317] Training an AI model can be accomplished by having the model learn good values for all of the weights and the bias for labeled examples created by the system, and in this case; starting with no labels initially. A goal of the training of the AI model can be to find a set of weights and biases that have low loss, on average, across all examples.

[0318] The AI classifier can receive supervised machine learning with a labeled data set to learn to perform their task as discussed herein. An anomaly detection technique that can be used is supervised anomaly detection that requires a data set that has been labeled as "normal" and "abnormal" and involves training a classifier. Another anomaly detection technique that can be used is an unsupervised anomaly detection that detects anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal, by looking for instances that seem to fit least to the remainder of the data set. The AI model representing normal behavior from a given normal training data set can detect anomalies by establishing the normal pattern and then test the likelihood of a test instance under analysis to be generated by the AI model. Anomaly detection can identify rare items, events or observations which raise suspicions by differing significantly from the majority of the data, which includes rare objects as well as things like unexpected bursts in activity.

[0319] The method and system are arranged to be performed by one or more processing components with any portions of software stored in an executable format on a computer readable medium. Thus, any portions of the method, apparatus and system implemented as software can be stored in one or more non-transitory memory storage devices in an executable format to be executed by one or more processors. The computer readable medium may be non-transitory and does not include radio or other carrier waves. The computer readable medium could be, for example, a physical computer readable medium such as semiconductor memory or solid state memory, magnetic tape, a removable

computer diskette, a random access memory (RAM), a read-only memory (ROM), a rigid magnetic disc, and an optical disk, such as a CD-ROM, CD-R/W or DVD.

[0320] The various methods described above may also be implemented by a computer program product. The computer program product may include computer code arranged to instruct a computer to perform the functions of one or more of the various methods described above. The computer program and/or the code for performing such methods may be provided to an apparatus, such as a computer, on a computer readable medium or computer program product. For the computer program product, a transitory computer readable medium may include radio or other carrier waves.

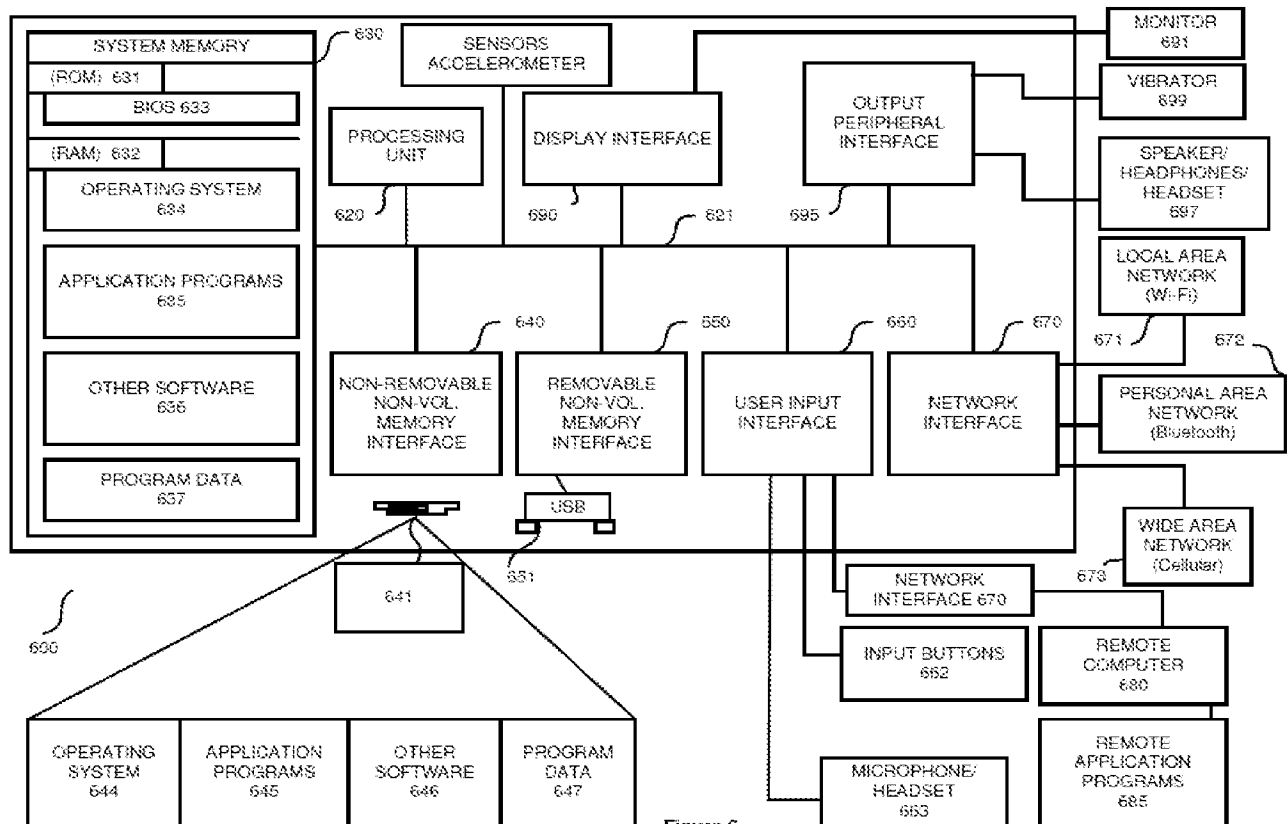


Figure 6

[0321] COMPUTING DEVICES

[0322] Figure 6 illustrates a block diagram of an embodiment of one or more computing devices that can be a part of an AI-based, cyber security system including the cyber security appliance 100, the restoration engine, the prediction engine, etc. for an embodiment of the current design discussed herein.

[0323] The computing device may include one or more processors (e.g. processing units) 620 to execute instructions, one or more memories 630-632 to store information, one or more data input components 660-663 to receive data input from a user of the computing device 600, one or more modules that include the management module, a network interface communication circuit 670 to establish a communication link to communicate with other computing devices external to the computing device, one or more sensors where an output from the sensors is used for sensing a specific triggering condition and then correspondingly generating one or more preprogrammed actions, a display screen 691 to display at least some of the information stored in the one or more memories 630-632 and other components. Note, portions of this design implemented in software 644, 645, 646 are stored in the one or more memories 630-632 and are executed by the one or more processors 620. The processing unit 620 may have one or more processing cores, which couples to a system bus 621 that couples various system components including the system memory 630. The system bus 621 may be any of several types of bus structures selected from a memory bus, an interconnect fabric, a peripheral bus, and a local bus using any of a variety of bus architectures.

[0324] Computing device 602 typically includes a variety of computing machine-readable media. Non-transitory machine-readable media can be any available media that can be accessed by computing device 602 and includes both volatile and nonvolatile media, and removable and non-removable media. By way of example, and not limitation, non-transitory machine-readable media use includes storage of information, such as computer-readable instructions, data structures, other executable software, or other data. Non-transitory machine-readable media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other tangible medium which can be used to store the desired information, and which can be accessed by the computing device 602. Transitory media such as wireless channels are not included in the machine-readable media. Machine-readable media typically embody computer readable instructions, data structures, and other executable software.

[0325] In an example, a volatile memory drive 641 is illustrated for storing portions of the operating system 644, application programs 645, other executable software 646, and program data 647.

[0326] A user may enter commands and information into the computing device 602 through input devices such as a keyboard, touchscreen, or software or hardware input buttons 662, a microphone 663, a pointing device and/or scrolling input component, such as a mouse, trackball or touch pad 661. The microphone 663 can cooperate with speech recognition software. These and other input devices are often connected to the processing unit 620 through a user input interface 660 that is coupled to the system bus 621, but can be connected by other interface and bus structures, such as a lighting port, game port, or a universal serial bus (USB). A display monitor 691 or other type of display screen device is also connected to the system bus 621 via an interface, such as a display interface 690. In addition to the monitor 691, computing devices may also include other peripheral output devices such as speakers 697, a vibration device 699, and other output devices, which may be connected through an output peripheral interface 695.

[0327] The computing device 602 can operate in a networked environment using logical connections to one or more remote computers/client devices, such as a remote computing system 680. The remote computing system 680 can be a personal computer, a mobile computing device, a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements described above relative to the computing device 602. The logical connections can include a personal area network (PAN) 672 (e.g., Bluetooth®), a local area network (LAN) 671 (e.g., Wi-Fi), and a wide area network (WAN) 673 (e.g., cellular network). Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet. A browser application and/or one or more local apps may be resident on the computing device and stored in the memory.

[0328] When used in a LAN networking environment, the computing device 602 is connected to the LAN 671 through a network interface 670, which can be, for example, a Bluetooth® or Wi-Fi adapter. When used in a WAN networking environment (e.g., Internet), the computing device 602 typically includes some means for establishing

communications over the WAN 673. With respect to mobile telecommunication technologies, for example, a radio interface, which can be internal or external, can be connected to the system bus 621 via the network interface 670, or other appropriate mechanism. In a networked environment, other software depicted relative to the computing device 602, or portions thereof, may be stored in the remote memory storage device. By way of example, and not limitation, remote application programs 685 as reside on remote computing device 680. It will be appreciated that the network connections shown are examples and other means of establishing a communications link between the computing devices that may be used. It should be noted that the present design can be carried out on a single computing device or on a distributed system in which different portions of the present design are carried out on different parts of the distributed computing system.

[0329] In certain situations, each of the terms “engine,” “module” and “component” is representative of hardware, firmware, and/or software that is configured to perform one or more functions. As hardware, the engine (or module or component) may include circuitry having data processing and/or storage functionality. Examples of such circuitry may include, but are not limited or restricted to a processor, a programmable gate array, a microcontroller, an application specific integrated circuit, wireless receiver, transmitter and/or transceiver circuitry, semiconductor memory, or combinatorial logic. Alternatively, or in combination with the hardware circuitry described above, the engine (or module or component) may be software in the form of one or more software modules, which may be configured to operate as its counterpart circuitry. For instance, a software module may be a software instance that operates as or is executed by a processor, namely a virtual processor whose underlying operations is based on a physical processor such as virtual processor instances for Microsoft® Azure® or Google® Cloud Services platform or an EC2 instance within the Amazon® AWS infrastructure, for example. Illustrative examples of the software module may include an executable application, a daemon application, an application programming interface (API), a subroutine, a function, a procedure, an applet, a servlet, a routine, source code, a shared library/dynamic load library, or simply one or more instructions. A module may be implemented in hardware electronic components, software components, and a combination of both. A module is a core component of a

complex system consisting of hardware and/or software that is capable of performing its function discretely from other portions of the entire complex system but designed to interact with the other portions of the entire complex system. The term “computerized” generally represents that any corresponding operations are conducted by hardware in combination with software and/or firmware. The terms “computing device” or “device” should be generally construed as physical device with data processing capability, data storage capability, and/or a capability of connecting to any type of network, such as a public cloud network, a private cloud network, or any other network type. Examples of a computing device may include, but are not limited or restricted to, the following: a server, a router or other intermediary communication device, an endpoint (e.g., a laptop, a smartphone, a tablet, a desktop computer, a netbook, IoT device, networked wearable, etc.) Finally, the terms “or” and “and/or” as used herein are to be interpreted as inclusive or meaning any one or any combination. As an example, “A, B or C” or “A, B and/or C” mean “any of the following: A; B; C; A and B; A and C; B and C; A, B and C.” An exception to this definition will occur only when a combination of elements, functions, steps or acts are in some way inherently mutually exclusive.

[0330] Note, an application described herein includes but is not limited to software applications, mobile applications, and programs routines, objects, widgets, plug-ins that are part of an operating system application. Some portions of this description are presented in terms of algorithms and symbolic representations of operations on data bits within a computer memory. These algorithmic descriptions and representations are the means used by those skilled in the data processing arts to convey the substance of their work most effectively to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated. It has proven convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers, or the like. These algorithms can be written in a number of different software programming languages such as Python, C, C++, Java, HTTP, or other similar languages. Also, an algorithm can be

implemented with lines of code in software, configured logic gates in hardware, or a combination of both. In an embodiment, the logic consists of electronic circuits that follow the rules of Boolean Logic, software that contain patterns of instructions, or any combination of both. Note, many functions performed by electronic hardware components can be duplicated by software emulation. Thus, a software program written to accomplish those same functions can emulate the functionality of the hardware components in the electronic circuitry.

[0331] Unless specifically stated otherwise as apparent from the above discussions, it is appreciated that throughout the description, discussions utilizing terms such as "processing" or "computing" or "calculating" or "determining" or "displaying" or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system's registers and memories into other data similarly represented as physical quantities within the computer system memories or registers, or other such information storage, transmission or display devices.

[001] While the foregoing design and embodiments thereof have been provided in considerable detail, it is not the intention of the applicant(s) for the design and embodiments provided herein to be limiting. Additional adaptations and/or modifications are possible, and, in broader aspects, these adaptations and/or modifications are also encompassed. Accordingly, departures may be made from the foregoing design and embodiments without departing from the scope afforded by the following claims, which scope is only limited by the claims when appropriately construed.