(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 February 2024 (15.02.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/035746 A1

(51) International Patent Classification:
G06F 16/901 (2019.01) *G06F 21/55* (2013.01)
G06F 21/50 (2013.01) *G06F 21/60* (2013.01)

(21) International Application Number:
PCT/US2023/029785

(22) International Filing Date:
08 August 2023 (08.08.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

63/396,105	08 August 2022 (08.08.2022)	US
17/991,173	21 November 2022 (21.11.2022)	US

(71) Applicant: **DARKTRACE HOLDINGS LIMITED** [GB/GB]; Maurice Wilkes Building, St John's Innovation Park, Cambridge, CB4 0DS (GB).

(72) Inventors; and
(71) Applicants: **BAZALGETTE, Timothy, Owen** [GB/GB]; 12 New Road, Woolmer Green, Knebworth SG3 6JX (GB). **HUMPHREY, Dickon, Murray** [GB/GB]; 2 Lents Way, Cambridge CB4 1UA (GB).

(72) Inventors: **FELLOWS, Simon, David Lincoln**; 19 Broad Street, Great Cambourne, Cambridge CB23 6EL (GB). **MARSENIC, Marko**; 10 Knoll Road, London SW18 2DF (GB). **SELLARS, Phillip**; Maurice Wikes Building, St John's Innovation Park, Cowley Road, Cambridge CB4 0DS (GB).

(74) Agent: **FERRILL, Thomas, S.**; Rutan & Tucker, LLP, 18575 Jamboree Road, 9th Floor, Irvine, CA 92612 (US).

(54) **Title:** A CYBER SECURITY RESTORATION ENGINE

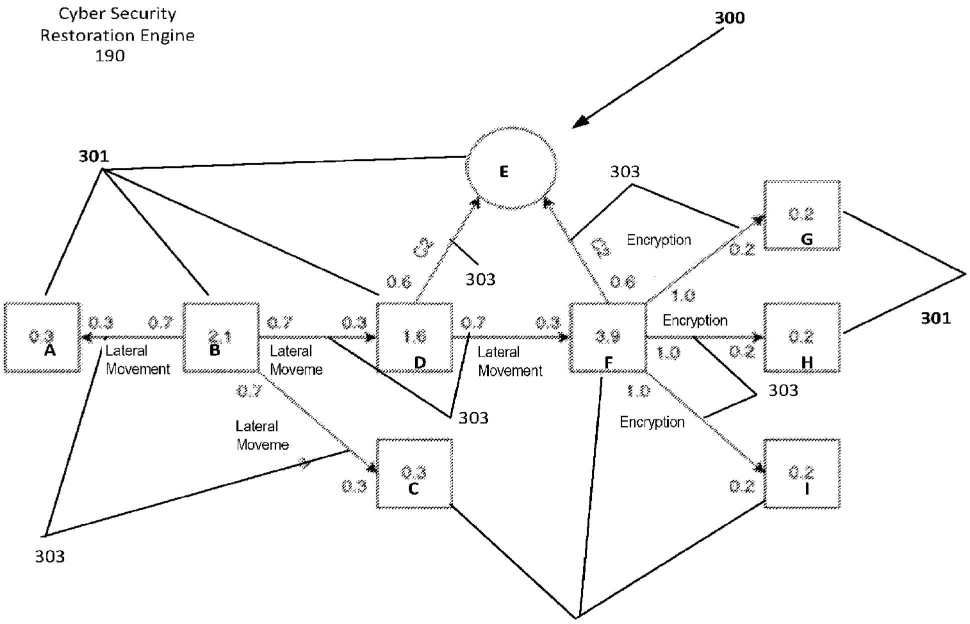


FIG 3B

(57) **Abstract:** A cyber security restoration engine prioritizes nodes in a graph of nodes in a computer network or system that are involved in a cyber attack for remediation actions. The cyber security restoration engine performs this prioritization by, for each node, determining one or more edges linking the node to other nodes in the graph, the edges representing interactions between two nodes; obtaining metadata indicative of a type of interaction between two nodes connected by the edge and the roles of the two nodes in that interaction; determining how severe the interaction represented by that edge is within the context of the cyber attack, based on the metadata of that edge; and determining a severity score for the node by combining the severity score for each of the one or more edges connected to the node. The cyber security restoration engine prioritizes nodes for remediation action based on the severity scores for the nodes.

WO 2024/035746 A1

[Continued on next page]

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

A CYBER SECURITY RESTORATION ENGINE

NOTICE OF COPYRIGHT

[001] 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.

RELATED APPLICATION

[002] This application claims priority under 35 USC 119 to U.S. provisional patent application No. 63/396105, titled "A CYBER THREAT PROTECTION SYSTEM" filed August 8, 2022. This application also claims priority under 35 USC 120 as a continuation in part to U.S. patent application No. 17/991173, titled "A CYBER SECURITY RESTORATION ENGINE" filed Nov. 21, 2022, which claims priority under 35 USC 119 to U.S. provisional patent application No. 63/281,978, titled "CYBER SECURITY TOOLS TO PROTECT A SYSTEM" filed November 22, 2021. All of these disclosures are incorporated herein by reference in their entirety.

FIELD

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

BACKGROUND

[004] 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

[005] Methods, systems, and apparatus are disclosed for an Artificial Intelligence-based cyber security system. The Artificial Intelligence-based cyber security system can include multiple Artificial Intelligence-based engines that cooperate to identify a cyber threat, mitigate that cyber threat and other cyber threats, restore from that cyber threat and other cyber threats, and factor in simulations of cyber threats.

[006] In an embodiment, a cyber security restoration engine is configured with software code and electronic hardware to prioritize nodes in a graph of nodes in a computer network or system that are involved in a cyber attack for remediation actions by, for each node: determining one or more edges linking the node to other nodes in the graph, the edges representing interactions between two nodes; obtaining metadata for each of the one or more edges, the metadata being indicative of a type of interaction between two nodes connected by the edge and the roles of the two nodes in that interaction; determining, for each of the one or more edges, a severity score representing how severe the interaction represented by that edge is within the context of the cyber attack, based on the metadata of that edge; and determining a severity score for the node by combining the severity score for each of the one or more edges connected to the node. The cyber security restoration engine is further configured to prioritize nodes for remediation action based on the severity scores for the nodes, with nodes having a higher severity score being prioritized above nodes having a lower severity score.

[007] 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

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

[009] Figure 1 illustrates a block diagram of an embodiment of the AI-based cyber security appliance with example components making up a detection engine that protects a system, including but not limited to a network/domain, from cyber threats.

[010] Figure 2 illustrates a block diagram of an embodiment of the AI-based cyber security appliance with example components making up a cyber security restoration (e.g. self-healing) engine that takes one or more autonomous remediation actions to recover from a cyberattack from a cyber threat.

[011] Figure 3A illustrates a diagram of an embodiment of the cyber security restoration engine configured to take one or more autonomous remediation actions based on Artificial Intelligence assistance to remediate one or more nodes in the graph of the system being protected back to the trusted operational state before a detected compromise by a cyber threat occurred in the protected system in order to assist in a recovery from the cyber threat.

[012] Figure 3B illustrates a graph of nodes in a computer network or system that are involved in a cyber attack and the metadata stored in the edges connecting the nodes, used by an embodiment of a cyber security restoration engine to prioritize nodes for remediation.

[013] Figure 4 illustrates a block diagram of an embodiment of an intelligent orchestration component configured to facilitate an Artificial Intelligence augmented and adaptive interactive response loop between multiple Artificial Intelligence-based engines.

[014] Figure 5 illustrates a diagram of an embodiment of i) the cyber threat detection engine using Artificial Intelligence algorithms configured and trained to perform a first machine-learned task of detecting the cyber threat, ii) an autonomous

response engine using 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, iii) a cyber-security restoration engine using Artificial Intelligence algorithms configured and trained to perform a third machine-learned task of remediating the system being protected back to a trusted operational state, and iv) a prediction engine using 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, working in tandem.

[015] 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.

[016] 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.

[017] Figure 8 illustrates a block diagram of an embodiment of the AI-based cyber security appliance with the cyber security restoration engine and other Artificial Intelligence-based engines plugging in as an appliance platform to protect a system.

[018] Figure 9 to Figure 14 illustrate diagrams of an embodiment of an intelligent orchestration component facilitating an example an Artificial Intelligence augmented and adaptive interactive response loop between the example four Artificial Intelligence-based engines.

[019] Figure 15 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 including the multiple Artificial Intelligence-based engines discussed herein.

[020] Figure 16 illustrates a graph of an embodiment of an example chain of unusual behavior for, in this example, the email activities and IT network activities deviating from a normal pattern of life in connection with the rest of the system/network under analysis.

[021] Figure 17 illustrates a diagram of an embodiment of the cyber security restoration engine cooperating with the other Artificial Intelligence-based engines of in a cyber security appliance to track and understand the cyber threat identified by the other Artificial Intelligence-based engines as well as track the one or more mitigation actions taken to mitigate the cyber threat during the cyberattack by the other Artificial Intelligence-based engines in order to assist in intelligently restoring the protected system, while still mitigating the cyber threat attack, back to a trusted operational state.

[022] 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

[023] 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.

[024] Again, a cyber security restoration engine is configured with software code and electronic hardware to take one or more autonomous remediation actions to remediate one or more nodes in a graph of a system being protected back to a trusted operational state in order to assist in a recovery from the cyber threat. The cyber security restoration engine has a tracking component for the operational state of each node in the graph of the protected system. The communication module also cooperates with the cyber security restoration engine to communicate with at least one of an external backup system and a recovery service to invoke backup remediation actions and/or recovery remediation actions to remediate one or more nodes potentially compromised by the cyber threat back to a trusted operational state, for example the state before the detected compromise by the cyber threat occurred in the protected system.

[025] Figure 1 illustrates a block diagram of an embodiment of the AI-based cyber security appliance with example components making up a detection engine 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 a system, such as 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 interface to an autonomous response engine, an interface to a restoration engine, an interface to a prediction engine, a 1st domain module, a 2nd domain module, and a coordinator module, and other similar components.

[026] The cyber threat detection engine includes a set of modules cooperating with one or more Artificial Intelligence models configured to perform a machine-learned task of detecting a cyber threat incident. The detection engine uses the set of modules cooperating with the one or more Artificial Intelligence models to detect anomalous behavior of one or more nodes, including at least user accounts, devices, and versions of source code files, in a graph of a system being protected. The detection engine uses the set of modules cooperating with the one or more Artificial Intelligence models to prevent a cyber threat from compromising the nodes and/or spreading through the nodes of the system.

[027] The cyber security appliance 100 with the Artificial Intelligence (AI)-based cyber security system may protect a network/domain from a cyber threat (insider attack, malicious files, malicious emails, etc.). In an embodiment, the cyber security appliance 100 can protect all of the 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.

[028] 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.

[029] 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.

[030] 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.

[031] 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.

[032] 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 patten 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.

[033] 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.

[034] 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.

[035] 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).

[036] 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.

[037] 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.

[038] 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.

[039] 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.

[040] The multiple Artificial Intelligence-based engines each have an interface to communicate with the other separate Artificial Intelligence-based engines. Each Intelligence-based engine has an interface to communicate with another separate Artificial Intelligence-based engine, which is configured to understand a type of information and communication that this other separate Artificial Intelligence-based engine needs to make determinations on an ongoing cyberattack from that other Artificial Intelligence-based engine's perspective. The autonomous response engine works with the assessment module in the detection engine when the cyber threat is detected and autonomously takes one or more actions to mitigate the cyber threat. Figure 1 shows the example components making up the detection engine to include interfaces to the prediction engine, the autonomous response engine, and the restoration engine.

[041] The cyber threat detection engine can also have an anomaly alert system in a formatting module configured to report out anomalous incidents and events as well as the cyber threat detected to a display screen viewable by a human cyber-security professional. Each Artificial Intelligence-based engine has a rapid messaging system to communicate with a human cyber-security team to keep the human cyber-security team informed on actions autonomously taken and actions needing human approval to be taken.

[042] Figure 2 illustrates a block diagram of an embodiment of the AI-based cyber security appliance with example components making up a cyber security restoration (e.g. self-healing) engine that takes one or more autonomous remediation actions to recover from a cyberattack from a cyber threat. Note, similarly named

components in the cyber security restoration engine can operate and function similar to as described for the detection engine.

[043] The cyber security restoration engine is configured to take one or more remediation actions based on configured and/or Artificial Intelligence assistance to remediate the one or more nodes in the graph of the system being protected back to a trusted operational state in a recovery from the cyber threat. These actions might be fully automatic, or require a specific human confirmation decision before they begin.

[044] The cyber security restoration engine is configured to cooperate with the other AI-based engines of the cyber security system, via the interfaces and/or direct integrations, to track and understand the cyber threat identified by the other components as well as track the one or more mitigation actions taken to mitigate the cyber threat during the cyberattack by the other components in order to assist in intelligently restoring the protected system while still mitigating the cyber threat attack back to a trusted operational state; and thus, as a situation develops with an ongoing cyberattack, the cyber security restoration engine is configured to take one or more remediation actions to remediate (e.g. restore) at least one of the nodes in the graph of the protected system back to a trusted operational state while the cyberattack is still ongoing.

[045] The cyber security restoration engine has a tracking component that includes at least one of i) a database to keep a record and track an operational state of each node in the graph of the protected system, ii) an Artificial Intelligence model trained to track the operational state of each node in the graph of the protected system, iii) a query to another artificial intelligence based engine that tracks the operational state of each node in the graph of the protected system from a different perspective, and iv) a combination of any of these, so that the cyber security restoration engine can then take the one or more autonomous remediation actions to remediate each particular node (e.g. user account and/or device) back to a trusted operational state for that node.

[046] The cyber security restoration engine can cooperate with the other Artificial Intelligence-based engines of the cyber security system to track and understand the cyber threat identified by the other Artificial Intelligence-based engines (detection engine and/or the prediction engine) as well as track the one or more

mitigation actions taken to mitigate the cyber threat during the cyberattack by an autonomous response engine and/or human cyber security team members in order to assist in intelligently restoring the protected system while still mitigating the cyber threat attack back to a trusted operational state. Thus, as a situation develops with an ongoing cyberattack, the cyber security restoration engine is configured to take the one or more remediation actions to remediate at least one of the nodes in the graph of the protected system back to a trusted operational state to restore portions of the protected system while the cyberattack is still ongoing. The cyber security restoration engine restores the affected nodes in the protected system by using incident modelling in the cyber threat analyst module (e.g. AI Analyst) to map and identify an entire lifecycle of attack, work with the AI models trained on cyber security threats in the detection engine to identify a source of the cyberattack, and recommend restore points (e.g. where in the protected system remediation action is needed).

[047] The communication module can cooperate with the cyber security restoration engine to communicate with the other Artificial Intelligence-based engines of the cyber security system. Again, the machine-learned tasks of the other Artificial Intelligence-based engines can include i) identifying the cyber threat itself and ii) taking one or more mitigation actions to mitigate the cyber threat during a cyberattack by the cyber threat. The communication module also communicates with one or more third party external backup and/or recovery services and systems to invoke backup remediation actions and recovery remediation actions to remediate the nodes from the cyber threat back to a trusted operational state, for example but not limited to the state before the detected compromise by the cyber threat occurred in the protected system. For example, the cyber security restoration engine can send a command to third party back up providers to invoke a full backup of a complete copy of all the files, folders, and operational settings for a device in the system. The cyber security restoration engine can use one or more Application Programming Interfaces (APIs) to translate desired remediation actions for the particular nodes in the system (e.g. its devices from potentially multiple different vendors, user accounts, etc.) being protected devices into a specific language and syntax utilized by the third party external backup and/or recovery services and systems to invoke the backup remediation actions and recovery

remediation actions to remediate the nodes. In addition, the cyber security restoration engine can send a request to the human cyber security team to take similar actions where it has no direct capability to do so itself but can recommend the remediation and recovery steps. The cyber security restoration engine can also prioritize the recommended remediation and recovery steps to enable the most efficient recovery of the system. In another example, the external 3rd party backup and/or recovery services and systems can include, for example, cloud data recovery, desktop and server backups to take disk images of hardware to restore all of the settings and data prior to an attack, and other forms of salvaging deleted, inaccessible, lost, corrupted, damaged, or formatted data and operational settings from these recovery services, switching to back up systems when the main system has been disrupted, etc.

[048] All of the Artificial Intelligence-based engines are configured to have bi-directional communications with the other Artificial Intelligence-based engines as well as with agents and sensors within the protected system under analysis. The communication module can use an instant messaging application between the cyber security restoration engine and members of a human cyber security team to report autonomous remediation actions taken by the cyber security restoration engine to restore the one or more nodes as well as proposed remediation actions needing the human cyber security team's authorization to remediate the one or more nodes in the protected system back to a trusted operational state. It may also use similar messaging applications to inform IT teams or other relevant but non cyber security teams that they need to take actions.

[049] The cyber security restoration engine can reference both i) a database of restoration response scenarios stored in the database and ii) a prediction engine configured to run Artificial Intelligence-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.

[050] The cyber security restoration engine can prioritize among the one or more nodes to restore, which nodes to remediate and an order of the nodes to

remediate. This can be done using a graph based representation of a cyber attack and using information stored in the graph as metadata that is derived by the restoration engine or other artificial intelligence based modules of the cyber security system.

[051] Figure 3A illustrates a block diagram of an embodiment of the cyber security restoration engine configured to take one or more autonomous remediation actions based on Artificial Intelligence assistance to remediate one or more nodes in the graph of the system being protected back to the trusted operational state before a detected compromise by a cyber threat occurred in the protected system in order to assist in a recovery from the cyber threat.

[052] As discussed, the communication module can be configured to communicate also with one or more external 3rd party backup and/or recovery services and systems to invoke backup remediation actions and recovery remediation actions to remediate the nodes from the cyber threat and restore likely affected users and devices in the protected system back to a trusted operational state. The external 3rd party backup and/or recovery services and systems receive instructions from the cyber security restoration engine to invoke specific and tailored backup remediation actions and recovery remediation actions to remediate the nodes from the cyber threat while still containing the cyber threat itself. As shown, the cyber security restoration engine can generate and maintain a graph of nodes making up the system under analysis.

[053] Figure 3B illustrates a graph 300 of nodes 301 in a computer network or system that are involved in a cyber attack. The nodes 301 that are involved in a cyber attack may be determined by Artificial Intelligence models by looking at deviations from a normal pattern of life, as described elsewhere herein. The cyber security restoration engine may communicate with these Artificial Intelligence modules to generate and maintain graph 300 when it is determined that a cyber attack is occurring as well as throughout the evolution of a cyber attack, for example as more devices on a network become infected by malware or are targeted during a cyber attack.

[054] Nodes 301 may represent devices within a computer network or system, user accounts, virtual computers, or any other discrete entities of a network. Nodes 301 are connected by edges 303, which represent and interaction between two nodes 301. The interaction may involve a lateral movement of data or software (especially malware)

between the two nodes, an encryption of one node by the other, or a command and control interaction of one node by the other, for example. Data need not necessarily be transferred in an interaction. It will also be appreciated that an edge may link a node within the computer network being protected with an external endpoint, such as when data is being sent from a device within the computer network being protected and an external endpoint.

[055] The interactions represented by an edge are the output of an analysis of the node behavior by an artificial intelligence module of the cyber security system and, in effect, is a summary or overview of what that interaction is. The edge does not necessarily just represent a single communication between two nodes. That is, not every communication between two nodes is represented by an individual edge. Indeed, some communications, not deemed reportable or abnormal by a cyber threat detection engine (i.e., they are within the normal pattern of life) will not have any corresponding edge at all. Generally, each edge will represent a plurality of individual communications between two nodes considered reportable or abnormal that are related to give an overall interaction, such as an "encryption" or "lateral movement (of data or in particular malware)" interaction. In some cases, an edge may represent an activity (or series of activities) that have already occurred. However, an edge may also represent an ongoing interaction, such as a node that is in the process of encrypting another node.

[056] Metadata about the interaction represented by an edge is stored within the graph, and this metadata is used by the restoration engine to prioritize nodes for restoration (also called healing). In one simple example, three bits of metadata are stored with each edge: the type of interaction, the source node, and the target node. This metadata is represented in Figure 3B by the arrows (representing which of the two nodes connected by the edge is the source node and which is the target node, with the arrow pointing from the source node to the target node) and the labels (specifying the interaction type).

[057] From this metadata, the restoration engine can construct, for each node, a severity score for each of the edges connected to that node. For example, in graph 300, node B has three edges connected to it, edge BA (connecting node B with node A), edge BC (connecting node B with node C), and edge BD (connecting node B with node

D). The restoration engine can determine a severity score for each edge BA, BC, and BD based on the metadata associated with each of these edges.

[058] Considering edge BA, the metadata indicates that this interaction is a “lateral movement” type of interaction, with node B as the source and node A as the target (i.e., destination of the data or malware). The restoration engine can use this information to assign a severity score of 0.7 to edge BA. Edges BD and BC are both the same type of interaction (“lateral movement” types with node B as the source), and so these edges are also assigned scores of 0.7. Once all of the edges connected to a node have had a severity score assigned to them by the restoration engine, the restoration engine can determine an overall severity score for the node. For node B, this can be determined by summing up the severity scores for edges BA, BC, and BD, giving 0.7 + 0.7 + 0.7 = 2.1. Therefore, the severity score of node B is 2.1. It should be noted that in this example the severity scores of the edges are summed, in general, they may be combined in some other way, such as an exponential decay of severity scores moving away from a given node.

[059] This process is repeated for each node of the network being protected within graph 300 (i.e., excluding external endpoints etc.) to give each node a severity score. It is noted that an edge may be assigned different severity scores when determining the severity score of each of the two nodes it connects. For example, as noted above, when determining the severity score for node B then edge BA was assigned a severity score of 0.7. However, when determining a severity score for node A, the edge connecting node A with node B (which can be labelled edge AB, for convenience, to distinguish from the discussion in relation to the same edge but when determining the severity score of node B above), has a severity score of 0.3 assigned to it because the metadata indicates that node A is the target of the “lateral movement” interaction. Therefore, it can be seen how the edge connecting nodes A and B may be assigned different severity scores depending upon which node is currently having it’s severity score being determined. Usually, the source of an interaction will be assigned a higher score (from that interaction) than the target of an interaction. In Figure 3B, the severity scores of the edges are placed proximal to the node to which they correspond for clarity.

[060] It should be noted that two nodes may be connected by two distinct edges if two types of interaction are ongoing or have occurred between them. For example, a first edge corresponding to a “lateral movement” interaction may represent movement of data from a first node to a second node, while a distinct, second edge corresponding to an “encryption” interaction may represent encryption of the first node by the second node.

[061] Returning to Figure 3B, while a “lateral movement” interaction is given a severity score of 0.7 or 0.3 depending upon whether a node is the source of the target of the interaction, different types of interaction may be assigned different severity scores. For example, in graph 300, the source of an “encryption” interaction (e.g., node F) may have 1.0 contributed to its overall severity score by that edge, while the target of an “encryption” interaction (e.g., nodes G, H and I) may have 0.2 contributed to its overall severity score by that edge.

[062] The different weightings given to different types of interaction will depend upon whether a node is the source or target of that interaction and can be preset, but may also be adaptable. In particular, artificial intelligence modules of the restoration engine may learn from past cyber attacks how effective different actions and sequences of actions were to remediate and restore compromised nodes and adjust the weights assigned to different types of interaction accordingly.

[063] Once every node has a severity score assigned to it, the nodes can then be ranked in order of their severity scores to give a prioritization order for remediation. For example, the nodes of graph 300 would be ranked in the following order: F (score 3.9), B (score 2.1), D (score 1.6), A (score 0.3), C (score 0.3), G (score 0.2), H (score 0.2), I (score 0.2). In this way, remediation actions can be performed in the most effective and efficient manner, beginning with remediation actions targeting node F.

[064] For instance, in the example graph of Figure 3B, the following course of action may be performed based on the priority order determined by the restoration engine.

1. Heal the node F. This node has the highest score because it is encrypting multiple other nodes (G, H and I). It is clearly infected and actively doing damage.

2. Heal the node B. This node has the second highest score because it is moving data laterally to multiple other nodes (A, C and D). It is actively spreading the infection through the network.
3. Heal the node D. This node has the third highest score because it is making command and control connections and performing some lateral movement to node F. It is likely infected and may continue to spread the infection further.
4. Heal nodes A and C. These nodes have the fourth highest scores because they have been targets of lateral movement but are not exhibiting signs of compromise themselves (no edges represent interactions with these nodes as the source). However, these nodes may be infected and could spread the infection further.
5. Heal the nodes G, H and I. These nodes have the lowest severity score because they have been targets of encryption. However, they are likely not compromised themselves, and so do not need to be addressed quickly, but still ultimately need to be restored.

[065] As illustrated by the above example, the prioritization order output by the restoration engine intrinsically accounts for dependencies. For example, in order to heal an encrypted node, the node encrypting it should be healed first, otherwise it may simply re-encrypt the recently healed node. By assigning the severity scores based on the metadata of the edges, taking into account the type of interaction and whether a node is the source or the target of the interaction, these dependencies are accounted for. As can be seen in the example above, the nodes being encrypted (G, H and I) are a lower priority than the node doing the encryption (node F). This is because the restoration engine assigns a higher severity score to the node doing the encryption (determined from the metadata of the edge) than to the node being encrypted (again, determined from the metadata of the edge).

[066] Other information may also be stored in the metadata associated with an edge or taken into account by the restoration engine. In particular, the metadata may comprise information about a node connected to an edge or about the edge itself from one of the other AI engines in the cyber security system.

[067] The cyber security restoration engine may be in communication with the autonomous response engine and receive information regarding any actions that have been automatically taken by the autonomous response engine. This information may be received by the restoration engine and incorporated into the metadata of the graph 300. For example, an autonomous response engine may have quarantined a compromised node. This information can be included in the metadata of the edges connected to this node and may lead to a lower severity score being assigned to those edges (because the node is quarantined and so cannot cause further damage, compared to other, unquarantined nodes).

[068] Similarly, the cyber security restoration engine may be in communication with a user interface module capable of receiving input from a human cyber security operative. This allows the human cyber security operative to input details about any response actions they have taken that the cyber security system may not be aware of otherwise (for example, physically removing a computer from a suspected inside attacker). This information can then be incorporated into the metadata in a similar way to information about any autonomous response actions as described above.

[069] The cyber security restoration engine may also be in communication with the cyber security prediction engine, which is configured to predict and simulate cyber attacks as described herein. From the prediction engine, the restoration engine can receive information regarding the vulnerability of nodes to cyber attacks and the importance of a node in terms of the potential damage or harm an attacker could cause if that node were compromised. This information can, again, be incorporated into the metadata associated with the edges of graph 300. For example, metadata indicating a node is vulnerable to an attack may lead to an increase in the severity score for edges connecting to that node, particularly for interactions having the vulnerable node as the target of the interaction. Similarly, metadata indicating that a node is important because a large amount of harm could be caused if that node were compromised may also lead to an increase in the severity score for edges connecting to that node, again, particularly for interactions having the important node as a target of the interaction.

[070] As discussed elsewhere, the restoration engine is configured to be able to autonomously perform restoration actions as well as determine restoration and

remediation actions that require a human operator to perform or approve the actions. Based on the analysis of the cyber attack performed by the detect and response engines, as well as its own analysis of the cyber attack, the restoration engine can determine remediation actions and estimate a time required to complete each of the remediation actions. The time required to complete a remediation action may be factored into the prioritization order.

[071] Returning to Figure 3B as an example, nodes A and C both have the same severity score (0.3) and so would be prioritized equally. However, based on the restoration engine's knowledge of these nodes (e.g., a type of node, a size, architecture, network connectivity, bandwidth etc.) the restoration engine may know that it would take less time to heal node A than node C. Therefore, the restoration engine will prioritize healing node A over node C, as this will lead to fewer nodes being compromised for so long.

[072] As previously discussed, the restoration engine may have a user interface module (see Figure 2). This enables the restoration engine to output the prioritization information to a human cyber security operative. For example, the user interface may enable an ordered list of nodes to be presented to a user based on their prioritization. This will enable the cyber security operative to take the appropriate action to restore the computer system under attack in an efficient and effective manner.

[073] The user interface module may further be configured to present to the user the severity scores for each node, as well as an indication of the factors contributing to the severity score. This can be particularly beneficial in instances where a human has taken actions that the cyber security system is not aware of. This is because, by transparently indicating how the restoration engine arrived at the severity scores and the prioritizations it determined, it allows the human cyber security operative to obtain an unbiased and informed view of the restoration engines understanding of the situation with enough information that the human cyber security operative can determine whether that understanding is accurate in view of any action they themselves have taken. If it is not accurate, because the human cyber security operative is aware of additional actions that have been performed, the user interface may be configured to enable the human cyber security operative to input information about the actions that

have been taken, which the restoration engine can then use to update the prioritization accordingly, as discussed above.

[074] In particular examples, the user interface module presents, for each node, the severity scores that have been combined to give rise to the overall severity score of that node. Alternatively or in addition, the user interface module presents, for each node, a description of the interaction represented by each of the nodes connected to that node or an overall natural language description of why that node has a high or low prioritization. For example, the user may be presented with text such as "Node F is encrypting multiple other nodes (G, H and I). There is a high confidence that it is infected and actively doing damage." This can enable human operatives to transparently understand the recommendations being provided to them by the restoration engine.

[075] Alongside the prioritization of the nodes, the user interface can present remediation actions for a human to perform or authorize. The remediation actions may be based on the prioritization order determined by the restoration engine, the metadata within the graph of the cyber incident, and/or a high level, artificial intelligence classification of the cyber incident by the restoration engine, as discussed subsequently. These may be in the form of one or more “playbooks”. Each playbook is a series of steps and actions to be performed by a human cyber security operative and/or the restoration engine in order to fully restore a node and/or an entire system. A playbook comprises of a combination of “playbooklets”, which are similar to playbooks (i.e. defined series of steps and actions) but address a specific independent problem. Each edge on graph 300, corresponding to a type of interaction, may have one or more playbooklets associated with that type of interaction that would typically be required to recover from the effects of the interaction. To generate a playbook for repairing a node, the restoration engine may further combine the playbooklets corresponding to each edge connected to that node, based on the metadata associated with the edge (e.g., taking into account the incident type, whether the node in question is the source or target, etc.).

[076] Some incidents may look identical from the point of view of the cyber security system on the network, but may require different remediation actions

depending upon details not available to the cyber security system. In this case, the user interface may request further input from a human cyber security operative or may present different options for the human cyber security operative to select from, depending upon the additional knowledge they have. For instance, a cyber attack may look the same but require different remediation actions to be performed depending upon whether it is an insider threat by a malicious party within the organization, an accidental threat by an ignorant or careless party within the organization, or an external threat. The user interface module may, therefore, present three playbooks and ask a human cyber security operative to select the appropriate playbook.

[077] Thus, the cyber-security restoration engine in the cyber security system determines the order of what files/devices/things should be healed first, leveraging knowledge from AI Analyst incidents, and performing a scoring algorithm that takes the type of action that happened to the object and its degree in determining its score. The cyber-security restoration engine has Artificial Intelligence/machine learning trained on how to – prioritize and confirm restoration of network devices/users/things in the network in response to a cyber attack while still mitigating the cyber threat. The cyber-security restoration engine determines the order of what files/devices/things should be healed first, leveraging knowledge from AI Analyst incidents, and performing a scoring algorithm that takes the type of action that happened to the object and its degree in determining its score. The cyber-security restoration engine in the cyber security system can use intelligence on prioritization and confirmation the order of what files/devices/things should be healed first in this network after a cyber attack is detected and mitigated in order to restore the network to a desired condition before the cyber attack occurred.

[078] The cyber-security restoration engine can then use time series prediction machine learning models applied to network activity to detect whether it has been healed successfully. After a device has been restored from backup this is done to check that it is behaving as expected. This could be done by computing the difference between the network activity previously seen on the device and the expected network activity. In the cyber-security restoration engine, the artificial intelligence can use various methodologies to understand what should be healed first when a cyber attack/ a

compromise has been detected on one or more devices in the network. An example way that the system prioritizes is utilizing scores of importance and damage, etc. derived from the prevent component running simulations of the cyber attack and subsequent recovery and, where the compromised devices fall in the overall compromise. For example, was this network device, the one that connected to all of the others network devices and spread ransomware to them during the simulations of the cyber attack; and thus, isolating that device and limiting types of communication from network devices connected downstream of the original compromised network device will determine the prioritization. The cyber-security restoration engine can pull that information, from the AI analyst incident information about how the cyber attack did or is predicted to unfold and then factoring in knowing an importance and damage and exposure metrics from a prevent component pull to create a method of scoring. The cyber-security restoration engine determines things like what should be prioritized for healing (e.g. remediation) whether that's restoring a backup save, rolling back (e.g. removing) or restoring network users' permissions back to a point zero in order to recover from the cyber attack while still mitigating the detected cyber attack to minimize times that a portion of the network is down due to the cyber attack. The cyber-security restoration engine can use time series prediction machine learning models. The cyber-security restoration engine can look at the predicted network behavior of the device(s) on a normal timeframe, for example, a time window of seven days of what we'd expect each device to do. The cyber-security restoration engine can then compare that predicted network behavior of the device(s) on a normal timeframe to the state of devices as a result of the cyber attack and obtain prioritization data. The cyber-security restoration engine determines essentially what devices/users/things should be healed (either fully or mostly restored) first and any required order of the healing and restoration, and the extent the restoration of permissions and connections need to be restored during the restoration process while still mitigating the threat from the cyber attack. The cyber-security restoration engine assists in determining a restoration process and automatically performing the restoration and/or at least presenting the restoration process to a user on a display screen, saving an actual human hours of precious time trying to figure out how to prioritize the devices/users/things from this particular cyber attack.

[079] The cyber-security restoration engine performs higher level incident classification and analysis by using a graph neural networks. An AI Cyber Analyst can identify complex cross-entity by using graph-based analysis. 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. 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, the model can classify subgraphs associated with types of incidents, identifying an overall description, highlighting the kinds of overall activity involved, and aiding in the suggestion of remediation steps. The model can perform graph-based meta analysis of events and then graph a neural network on that. The system trains machine learning algorithms on graphs to try and classify those graphs into associated categories of graphs, so when an incident in a network with the AI analyst deployed occurs, it can analyze the series of nodes in the graph, such as boundary network devices and/or endpoint computing devices connected together by linking factors. The neural network is trained on those structures to try and derive an overall shape what it thinks that incident is. For example, linking factors on each of the nodes and the labels on the edges are analyzed, the model thinks this is ransomware, which was confirmed through supervised learning and then derive an overall type of incident and then suggest appropriate remediation actions. The system creates the shape of the incident through a graph neural network in order to correlate that to a type and/or category of that incident. The system then takes the next step to suggest the appropriate actions to take and present a title to explain it to an end user.

[080] Thus, the cyber-security restoration engine creates a graph based analysis / representation of a cyber security incident in order to determine/create a priority of restoration/ recovery of assets (e.g. devices and user privileges) during and after the cyber security incident. The cyber-security restoration engine uses a pre

existing graphical representation of the cyber incident to recommend recovery from the cyber incident utilizing machine learning approaches to both, summarize that incident structure in an understandable format, and to recommend 1) based upon machine learning approaches and/or 2) based upon a kind of weighted system, i) the most appropriate way to recover from that in that cyber incident, ii) the highest priority nodes within that cyber incident to which to address and iii) also to recommend a course of action. If the cyber-security restoration engine is not able to, with which both the cyber-security restoration engine may take and the user may take and taking also into account input from other systems that may be part of the cyber security platform (e.g. the prediction engine, the autonomous response engine, etc.), including whether then any of the assets involved in this, the ongoing side of instant under any kind of respond control, or may choose to the instant may spread further, based upon sort of known aspects from these other parts of the cyber security platform (e.g. the prediction engine, the autonomous response engine, etc.). The cyber-security restoration engine from a cyber threat attack can have a prioritization module/component to help to derive from a high level AI graph representation of an incident the tailored recommended steps to recover the network assets (e.g. devices and privileges associated with a user) involved, and an unbiased view of the most effective priority order to restore those assets in light of the occurring cyber attack. The cyber-security restoration engine can, via its user interface present a graph based analysis / representation of a cyber security incident in order to determine/create a priority of restoration/ recovery of devices and user privileges during and after the cyber security incident. that provides guidance to a human (and/or actually take autonomous restoration steps itself) for the types of steps the system might want to take knowing the exact details of this particular network being protected by a cyber security appliance, helping to inform what types of steps the system might want to take to restore and recover from that particular cyber attack and then be able to classify on top of that what is actually happened so that it can be understood. The cyber-security restoration engine produces a high level representation of an incident that allows a user to then do further analysis both using more AI and in other ways that will help you assess that ongoing cyber attack incident, and ultimately respond to the cyber attack to restore assets during the cyber attack to restore components and abilities while still mitigating the cyber threat causing the cyber attack itself, and heal the assets in the network being protected by the cyber security appliance. The cyber-security restoration engine uses the information that different components in the cyber security system derived from the cyber attack and information already known about that particular network under attack and using AI in the restoration engine to generate an abstraction graph of that cyber incident on the nodes making up the assets of the particular network under attack, and then taking in that information and using that information to determine what kind of prioritizing (priority order and extent of privileges and/or capabilities) restored is applied that particular asset in light of the current state of the ongoing cyber attack (both what type of cyber attack incident is occurring and what assets -users & devices that are affected) to continue to mitigate the effects and spread of the cyber attack while restoring as much normal operations or at least limited operations back to the assets of that network being protected.

[081] In an embodiment, a cyber security restoration engine is configured with software code and electronic hardware to take one or more autonomous remediation actions to remediate one or more nodes in a graph of a system being protected back to a trusted operational state before a detected compromise by a cyber threat occurred in the protected system in order to assist in a recovery from the cyber threat, while a cyber attack is ongoing as well as after the cyber attack has been mitigated/eliminated. The cyber security restoration engine has a tracking component that includes at least one of i) a database to keep a record and track an operational state of each node in the graph of the protected system, ii) an Artificial Intelligence model trained to track the operational state of each node in the graph of the protected system, iii) a query to another Artificial Intelligence based engine that tracks the operational state of each node in the graph of nodes of the system being protected, and iv) a combination of any of these, so that the cyber security restoration engine can then take the one or more autonomous remediation actions to remediate one or more nodes back to a trusted operational state for that node. In addition, (add some claim language on) The cyber security restoration engine cooperates with AI model(s) 160 trained with machine learning on the normal behavior of the devices and users in the system. This way , the

cyber security restoration engine knows what is best to do based on knowing the normal behavior of that particular user and/or device.

[082] Figure 4 illustrates a block diagram of an embodiment of an intelligent orchestration component configured to facilitate an Artificial Intelligence augmented and adaptive interactive response loop between the multiple Artificial Intelligence-based engines. The example multiple Artificial Intelligence-based engines cooperating with each other can include i) the cyber threat detection engine, ii) an autonomous response engine, iii) a cyber-security restoration engine, and iv) a prediction engine. i) The cyber threat detection engine can be configured to use Artificial Intelligence algorithms trained to perform a machine-learned task of detecting the cyber threat. (See for example Figure 1) ii) The autonomous response engine can be configured to use Artificial Intelligence algorithms trained to perform a machine-learned task of taking one or more mitigation actions to mitigate the cyber threat. iii) The cyber-security restoration engine can be configured to use Artificial Intelligence algorithms trained to perform a machine-learned task of remediating the system being protected back to a trusted operational state, as well as recommending and prioritizing remediation actions to be performed or authorized by human cyber security operators. (See for example Figure 2) iv) The prediction engine can be configured to use Artificial Intelligence algorithms trained to perform a 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. (See, for example, Figure 6)

[083] The multiple Artificial Intelligence-based engines have communication hooks in between them to exchange a significant amount of behavioral metrics including data between the multiple Artificial Intelligence-based engines to work in together to provide an overall cyber threat response.

[084] The intelligent orchestration component can be configured as a discreet intelligent orchestration component that exists on top of the multiple Artificial Intelligence-based engines to orchestrate the overall cyber threat response and an interaction between the multiple Artificial Intelligence-based engines, each configured to

perform its own machine-learned task. Alternatively, the intelligent orchestration component can be configured as a distributed collaboration with a portion of the intelligent orchestration component implemented in each of the multiple Artificial Intelligence-based engines to orchestrate the overall cyber threat response and an interaction between the multiple Artificial Intelligence-based engines. In an embodiment, whether implemented as a distributed portion on each AI engine or a discrete AI engine itself, the intelligent orchestration component can use self-learning algorithms to learn how to best assist the orchestration of the interaction between itself and the other AI engines, which also implement self-learning algorithms themselves to perform their individual machine-learned tasks better.

[085] The multiple Artificial Intelligence-based engines can be configured to cooperate to combine an understanding of normal operations of the nodes, 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 Artificial Intelligence-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.

[086] One or more Artificial Intelligence models in the detection engine can be configured to maintain what is considered to be normal behavior for that node, which is constructed on a per node basis, on the system being protected from historical data of that specific node over an operation of the system being protected.

[087] The multiple Artificial Intelligence-based engines each have an interface to communicate with the other separate Artificial Intelligence-based engines configured to understand a type of information and communication that the other separate Artificial Intelligence-based engine needs to make determinations on an ongoing cyberattack from that other Artificial Intelligence-based engine’s perspective. Each Artificial Intelligence-based engine has an instant messaging system to communicate with a human cyber-security team to keep the human cyber-security team informed on actions

autonomously taken and actions needing human approval as well as generate reports for the human cyber-security team.

[088] Figure 5 illustrates a diagram of an embodiment of i) the cyber threat detection engine using Artificial Intelligence algorithms trained to perform a first machine-learned task of detecting the cyber threat, ii) an autonomous response engine using Artificial Intelligence algorithms trained to perform a second machine-learned task of taking one or more mitigation actions to mitigate the cyber threat, iii) a cyber-security restoration engine using Artificial Intelligence algorithms trained to perform a third machine-learned task of remediating the system being protected back to a trusted operational state, and iv) a prediction engine using Artificial Intelligence algorithms 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, in order for these four Artificial Intelligence-based engines to work together. In addition, the intelligent orchestration component can use Artificial Intelligence algorithms trained to perform a fifth machine-learned task of adaptive interactive response between the multiple Artificial Intelligence-based engines to provide information each Artificial Intelligence engine needs to work cohesively to provide an overall incidence response that mitigates different types of cyber threats while still minimizing an impact tailored to this particular system being protected. For example, when a conversation occurs between the AI-based engines such as a system that can be positively affected by both proposed mitigation actions and proposed restoration actions, any of which might be attempted but fail or only partially succeed, then the intelligent orchestration component can arbitrate and evolve the best result for this particular system being protected. The intelligent orchestration component can help anticipate i) the needs of and ii) cohesive response of each Artificial Intelligence-based engine based on a current detected cyber threat.

[089] The cyber-security restoration engine receives and sends inputs through communication hooks (e.g.) interfaces to all of these Artificial Intelligence-based engines each configured with self-learning AI machine learning algorithms to,

respectively, i) to detect the cyber threat, ii) to respond to mitigate that cyber threat, and iii) to predict how that cyber threat might occur and likely progress through simulations. Each of these Artificial Intelligence-based engines has bi-directional communications, including the exchange of raw data, with each other as well as with software agents resident in physical and /or virtual devices making up the system being protected as well as bi-directional communications with sensors within the system being protected. Note, the system under protection can be, for example, an IT network, an OT network, a Cloud network, an email network, a source code database, an endpoint device, etc.

[090] In an example, the autonomous response engine uses its intelligence to cooperate with a cyber threat prediction engine and its Artificial Intelligence-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. The autonomous response engine, rather than the human taking the action, is configured to autonomously cause the one or more mitigation actions to be taken to contain the cyber threat when a threat risk parameter from an assessment module in the detection engine is equal to or above an actionable threshold. Example mitigation actions can include 1) the autonomous response engine monitoring and sending signals to a potentially compromised node to restrict communications of the potentially compromised node to merely normal recipients and types of communications according to the Artificial Intelligence model trained to model the normal pattern of life for each node in the protected system, 2) the autonomous response engine trained on how to isolate a compromised node as well as to take mitigation acts with other nodes that have a direct nexus to the compromised node.

[091] In another example, the cyber threat prediction engine and its Artificial Intelligence-based simulations use intelligence to cooperate with the cyber-security restoration engine 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.

[092] In another example, the cyber security restoration engine restores the one or more nodes in the protected system by cooperating with at least two or more of 1) an Artificial Intelligence model trained to model a normal pattern of life for each node in the protected system, 2) an Artificial Intelligence 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 engine.

[093] 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 virtualized instances 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 virtualized instance of the network to be tested. As shown in Figure 6, the various cooperating modules residing in the prediction engine 105 may include, but are not limited to, a collections module, a cyberattack generator (e.g. phishing email generator with a paraphrasing engine), an email module, a network module, an analyzer module, a payloads module with first and second payloads, a communication module, a training module, a simulated attack module, a cleanup module, a scenario module, a user interface, a reporting module, a formatting module, an orchestration module, an AI classifier with a list of specified classifiers.

[094] The prediction engine 105 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 pentest one or more defenses provided by 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. The prediction engine 105 may include and cooperate with one or more AI models trained with machine learning on the contextual knowledge of the organization. These trained AI models 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 105 may use the trained AI models to cooperate with one or more AI classifier(s) by producing a list of specific organization-based classifiers for the AI classifier. The prevent engine 105 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 determine 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 prior to an autonomous response action is taken in response to the actual cyberattack of the cyber incident.

[095] Figure 7 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. The prediction engine 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 notes to certain key servers.

[096] Again, similarly named components in each Artificial Intelligence-based engine can 1) perform similar functions and/or 2) have a communication link from that component located in one of the Artificial Intelligence-based engines and then information is needed from that component is communicated to another Artificial

Intelligence-based engine that through the interface to that Artificial Intelligence-based engine.

[097] Training of AI pre-deployment and then during deployment

[098] In step 1, an initial training of the Artificial Intelligence 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 for 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. 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.

[099] 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.

[0100] Thus, a deployed Artificial Intelligence model 160 trained on a normal behavior 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 Artificial Intelligence models 160 trained with machine learning on a normal 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. Alternatively, pre-deployment machine learning training of one or more Artificial Intelligence models trained on a normal pattern of life of entities in the system can occur. Initial training of one or more Artificial Intelligence models 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. 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.

[0101] 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.

[0102] Clustering Methods

[0103] 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.

[0104] 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.

[0105] 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 threat analysis. 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. More on the training of the AI models that are trained to create one or more possible hypotheses and perform cyber threat investigations will be discussed later.

[0106] Next, the various Artificial Intelligence 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 cyber threat attacks. The Artificial Intelligence models and classifiers combine use of unsupervised and supervised machine learning constantly revises assumptions about behavior, using probabilistic mathematics, that is always up to date on what a current normal behavior is, and not solely reliant on human input. The Artificial Intelligence 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.

[0107] 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.

[0108] Ranking the Cyber Threat

[0109] 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 and 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.

[0110] More on the operation of the cyber security appliance 100

[0111] As discussed in more detail below, 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.

[0112] All of the above AI models 160 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.

[0113] Anomaly detection/ deviations

[0114] 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 cyber-attack, 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.

[0115] 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.

[0116] Again, the AI models trained on a normal pattern of life of entities in a network (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.

[0117] Next, as discussed further below, 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.

[0118] 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.

[0119] 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

[0120] **<u>Additional module interactions</u>**

[0121] 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.

[0122] 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.

[0123] 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.

[0124] 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.

[0125] 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.

[0126] 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.

[0127] 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.

[0128] 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.

[0129] **<u>Determination of whether something is likely malicious</u>**

[0130] 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.

[0131] 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 of 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.

[0132] 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.

[0133] 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.

[0134] 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.

[0135] 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.

[0136] 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.

[0137] 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.

[0138] 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.

[0139] 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.

[0140] 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.

[0141] 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?

[0142] 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.

[0143] The cyber threat analyst module 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.

[0144] In an example, a behavioural pattern analysis of what are the unusual behaviours of the network/system/device/user under analysis by the machine learning models may be as follows. The coordinator module 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. Figure 16 illustrates a graph 220 of an embodiment of an example chain of unusual behaviour for, in this example, the email activities and IT network activities deviating from a normal pattern of life in connection with the rest of the system/network under analysis. The cyber threat analyst module and/or analyzer module can cooperate with one or more machine learning models. The one or more 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 behaviour of a malicious actor – human, program, or other threat. The cyber threat analyst module 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. An example 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 detects a chain of anomalous behavior of unusual data transfers three times, unusual characteristics in emails 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 emails 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 behaviour of a malicious actor. The analyser module can cooperate with one or more models trained on cyber threats and their behaviour to try to determine if a potential cyber threat is causing these unusual behaviours. The cyber threat analyst module 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) a 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 behaviours 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 behaviours under analysis, that are equal to or above a configurable level of threat posed by this malicious actor.

[0145] The autonomous response engine of the cyber security system 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 engine is configured to reference an Artificial Intelligence model trained to track a normal pattern of life for each node of the protected system to perform an autonomous act of 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. Similarly named components in the cyber security restoration engine can operate and function similar to as described for the detection engine.

[0146] 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.

[0147] An assessment of the cyber threat in order to determine appropriate autonomous actions, for example, those by the autonomous response engine

[0148] In the next step, the analyzer module and/or cyber threat analyst module generates one or more supported possible cyber threat hypotheses from the possible set of cyber threat hypotheses. The analyzer module generates the supporting data and details of why each individual hypothesis is supported or not. The analyzer module can also generate one or more possible cyber threat hypotheses and the supporting data and details of why they were refuted.

[0149] In general, the analyzer module cooperates with the following three sources. The analyzer module cooperates with the AI models 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 cooperates with the AI models trained on a normal behavior of entities in the network under analysis. The analyzer module 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.

[0150] In the next step, an assessment module with the AI classifiers is configured to cooperate with the analyzer module. The analyzer module supplies the identity of the supported possible cyber threat hypotheses from the possible set of cyber threat hypotheses to the assessment module. The assessment module with the AI classifiers cooperates with the AI model 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 with the AI classifiers cooperates with the one or more AI models trained on possible cyber threats in order to assign a numerical assessment of a given cyber threat hypothesis that was found likely to be supported by the analyzer module 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 with the AI classifiers output can be a score (ranked number system, probability, etc.) that a given identified process is likely a malicious process.

[0151] The assessment module 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, 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.

[0152] The cyber threat analyst module in the AI-based cyber security appliance 100 component provides an advantage over competitors' products as it reduces the time taken for cybersecurity investigations, provides an alternative to manpower for small organizations and improves detection (and remediation) capabilities within the cyber security platform.

[0153] The AI-based cyber threat analyst module performs its own computation of threat and identifies interesting network events with one or more processers. These methods of detection and identification of threat all add to the above capabilities that make the AI-based cyber threat analyst module a desirable part of the cyber security appliance 100. The AI-based cyber threat analyst module offers a method of prioritizing which is not just a summary or highest score alert of an event evaluated by itself equals the most bad, and prevents more complex attacks being missed because their composite parts/individual threats only produced low-level alerts.

[0154] The AI classifiers can be part of the assessment component, which scores the outputs of the analyzer module. 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.

[0155] In the next step, the formatting module can generate an output such as a printed or electronic report with the relevant data. The formatting module can cooperate with both the analyzer module and the assessment module depending on what the user wants to be reported.

[0156] The formatting module 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 one or more report templates populated with the data for that incident.

[0157] The formatting module 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 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.

[0158] 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 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 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 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.

[0159] In the next step, the assessment module 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, then cooperates with the autonomous response engine to take an autonomous action such as i) deny access in or out of the device or the network and/or ii) shutdown activities involving a detected malicious agent.

[0160] The autonomous response engine, rather than a human taking an action, can be configured to cause one or more rapid autonomous mitigation actions to be taken to counter the cyber threat. A user interface for the response module can program the autonomous response engine 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 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 engine 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, and other autonomous actions against the devices and/or users. The autonomous response engine uses one or more Artificial Intelligence models that are configured to intelligently work with other third-party defense systems in that customer's network against threats. The autonomous response engine can send its own protocol commands to devices and/or take actions on its own. In addition, the autonomous response engine uses the one or more Artificial Intelligence models 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 engine can be an autonomous self-learning 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 module can cooperate with the autonomous response engine 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.

[0161] The trigger module, analyzer module, assessment module, and formatting module 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.

[0162] Figure 9 to Figure 14 illustrate diagrams of an embodiment of an intelligent orchestration component facilitating an example of an Artificial Intelligence augmented and adaptive interactive response loop between the four Artificial Intelligence-based engines. Again, the multiple (e.g. four) Artificial Intelligence-based engines have communication hooks in between them to exchange a significant amount of behavioral metrics including data between the multiple Artificial Intelligence-based engines to work in together to provide an overall cyber threat response. The AI adaptive incident response loop has interaction and orchestration between the multiple (four) self-learning AI components, each trained and focused on their individual machine-learned tasks of i) detecting a cyber threat, ii) how to conduct a simulation and make the prediction about a cyberattack, iii) how to make and what types of autonomous mitigation responses can be made in response to a cyberattack and iv) what level of restrictions are needed and how to invoke restoration actions to restore nodes in the system being protected while still mitigating effects of the cyberattack. The Artificial Intelligence in each of the engines trained and focused on performing their corresponding machine-learned tasks as well as the orchestration between the Artificial Intelligence-based engines drive the exchange to make them work in together against a cyberattack by the cyber threat (e.g. malicious actor). The intelligent orchestration component facilitates the multiple example stages of the Artificial Intelligence augmented and adaptive interactive response loop between these four Artificial Intelligence-based engines.

[0163] Figure 9 illustrates a diagram of an example first stage of the interactive Artificial Intelligence-based response loop between the multiple Artificial Intelligence-based engines working in tandem to provide an overall cyber threat response.

[0164] The human security team is just working through these top cyber threat risks identified by the prediction engine simulations, while the AI detection engine looks for cyber threats.

[0165] The detection engine monitoring the example network being protected detects a potential cyber threat. The detection engine informs all the other engines of the new detection and details (e.g. the symptoms detected and any devices possibly compromised). The detection engine also sends a report to the human security team to review.

[0166] The detection engine detects a cyberattack through AI analysis and communicates the cyber threat and one or more nodes (user accounts, devices, etc.) showing anomalous behavior indicative of being compromised by the cyber threat to the autonomous response engine. The detection engine communicates to the autonomous response engine about all the cyber events and indicated nodes that need containing, even the ones that potentially individually seem normal.

[0167] The autonomous response engine reviews the information from the detection engine to take one or more autonomous actions to mitigate the cyberattack. The artificial intelligence in the autonomous response engine determines what additional information the autonomous response engine needs to take the one or more autonomous actions to mitigate the cyberattack while still minimizing an effect on the rest of the non-compromised nodes in the system and other mitigation actions needed to contain the cyber threat. The autonomous response engine reasons and acts, e.g. blocking specific network connections, enforcing a pattern of life on users/devices, and locking accounts. The autonomous response engine is trained to mitigate the cyber threat and compromised nodes with the minimum number of changes to achieve that goal, and with a minimum disruption to legitimate on-going operations

[0168] The detection engine detects a cyberattack through AI analysis and also communicates the cyber threat and one or more nodes (user accounts, devices, etc.) showing anomalous behavior indicative of being compromised by the cyber threat to the prediction engine. The prediction engine reviews the information and runs one or more Artificial Intelligence-based simulations of cyberattacks to assist in determining 1) how this cyberattack might have occurred in the system being protected, and 2) using the cyberattack information from the detection engine to run simulations to preempt possible escalations of this ongoing actual cyberattack. The prediction engine reviews the information from the detection engine and the artificial intelligence in the prediction engine determines what additional information the prediction engine needs to run one or more simulations of this cyberattack. The prediction engine reasons, based on the one or more possible cyberattacks detected from the events and indicators detected, the likely next set of actions that might take place in the attack – especially the actions that would constitute the most dangerous escalations.

[0169] The prediction engine communicates with the autonomous response engine to supplement the actions and events it can take to mitigate the cyberattack based on the simulations. Again, the autonomous response engine reasons and acts.

[0170] The autonomous response engine communicates with the restoration engine what autonomous actions have been taken to mitigate the cyber threat. The autonomous response engine communicates e.g. that a specific device and account are being contained now, and restoration is therefore less urgent. The restoration engine pulls information from the detection engine to determine an operational state for each node in the system being protected. The artificial intelligence in the restoration engine analyzes the operational state of the nodes in the network/system being protected. The artificial intelligence in the restoration engine analyzes the operational state of the nodes and what level of mitigation might be needed to contain the cyber threat during this ongoing cyberattack. The artificial intelligence in the restoration engine also analyzes the information from the Artificial Intelligence-based simulations of cyberattacks to assist in determining what level of restrictions are needed on the potentially compromised nodes and their neighboring nodes to still mitigate against an ongoing cyberattack but minimize an effect on the rest of the system.

[0171] Figure 10 illustrates a diagram of an example second stage of the interactive Artificial Intelligence-based response loop between the multiple Artificial Intelligence-based engines working in tandem to provide an overall cyber threat response. Again, the autonomous response engine communicates with the human cyber security team and restoration engine about the things it has successfully contained. The human cyber security team does not need to have been involved at all to this point. The restoration engine also communicates to the human cyber security team a report. The restoration engine is configured to be able to take autonomous recovery actions in some cases such as this one, and can autonomously e.g. revert one or more of the changes made to an account and give sole ownership back to the intended user.

[0172] The cyber security restoration engine is trained to heal the protected system back to a healthy state with the minimum number of changes to achieve that goal, and with a minimum disruption to legitimate on-going operations. The cyber security restoration engine can take various example remediation actions in real time. The cyber security restoration engine to restore the protected system can use historic SaaS/IaaS resource access, privilege information and clustered similar users to reset user access and permissions after a privilege escalation incident.

[0173] Figure 11 illustrates a diagram of an example third stage of the interactive Artificial Intelligence-based response loop between the multiple Artificial Intelligence-based engines working in tandem to provide an overall cyber threat response. The restoration engine sends another report after its analysis to the human cyber security team. The report communicates that it cannot be sure whether an initially compromised device can be trusted now and it recommends rebuilding from an effective date prior to the initial indications of the cyberattack. This a major operation and the device and its user will lose some data. The cyber-security restoration engine checks with the human team to obtain approval to perform this restoration action. The human cyber security team decide whether to confirm this rebuild action, potentially including in that decision information that cannot be known by the cyber security system (e.g. calling a device's user to discuss the impact of the rebuild). The restoration engine enables human security teams to make quick and confident decisions with the goal of keeping the business up and running. The report can identify assets affected by a cyberattack, their condition, and how best to restore them during and after an attack.

[0174] Next, a period of time has occurred between the initial detection of the cyber threat.

[0175] Figure 12 illustrates a diagram of an example fourth stage of the interactive Artificial Intelligence-based response loop between the multiple Artificial Intelligence-based engines working in tandem to provide an overall cyber threat response. The human security team confirms the recommended device rebuild. The restoration engine communicates back to the detection engine and the human cyber security team after it has automatically recovered an affected account. The detection engine reviews the information from the restoration engine and can choose to enhance the monitoring of this account temporarily, since its recent attack makes it more likely to be attacked again or for new activity to reveal the restoration did not fully resolve the problem. The artificial intelligence in the detection engine determines what additional information the detection engine needs to set the appropriate revised thresholds and types of behavior to look for. The detection engine communicates to both the restoration engine and the human cyber security team that it will monitor the restored user account and associated devices with a heightened scrutiny and increase the threat score of potential unusual events including abnormal behavior from the user and the devices for a set period of time. The detection engine then sets detection of cyber threats and abnormal behavior with the lower thresholds on the specific user account and the associated devices for that set period of time. The prediction engine through its simulations can inform the detection engine and restoration engine whether a recovered device is still extremely weak to other types of cyberattack.

[0176] Figure 13 illustrates a diagram of an example fifth stage of the interactive Artificial Intelligence-based response loop between the multiple Artificial Intelligence-based engines working in tandem to provide an overall cyber threat response. The detection engine sends a new communication to the autonomous response engine and the human security team that it has detected and found some new related abnormal activity; thus, this attacker is probably still active. The autonomous response engine immediately takes actions as before.

[0177] The detection engine communicates the new and overall cyber threat activity to the prediction engine. The Artificial Intelligence in the prediction engine reviews the communicated cyber threat activity and gathers additional information to run simulations. The Artificial Intelligence in the prediction engine reasons the most likely and/or most dangerous next moves by the attacker, and their most likely ultimate goals based on the results of the cyberattack simulations.

[0178] The human cyber security team reviews the updated report from the detection engine. The human team is active and occasionally needs to confirm recommendations or make decisions, but a large part of the overall incident response is assisted by the AI engines working together without human input.

[0179] Figure 14 illustrates a diagram of an example sixth stage of the interactive Artificial Intelligence-based response loop between the multiple Artificial Intelligence-based engines working in tandem to provide an overall cyber threat response. The prediction engine after running the Artificial Intelligence-based simulations communicates to the autonomous response engine the locations where it could block the likely and/dangerous next moves by the attacker. The Artificial Intelligence in the autonomous response engine analyzes the simulation results and grabs any additional information needed to decide what nodes need autonomous actions and what mitigation actions to take to each node that is compromised and potentially its neighboring nodes.

The Artificial Intelligence in the autonomous response engine reasons and takes action. The AI engines also update the report visible to the human cyber security team.

[0180] This interactive Artificial Intelligence-based response loop between the multiple Artificial Intelligence-based engines working together continues on. The intelligent orchestration component uses unsupervised machine learning algorithms to self-learn from previous cyber threat incidents (and their aftermath) on tasks such as how the response went, what worked, what did not, how long things took and how this compared to previous occasions and to expectations, and then uses this information to adjust future incident response expectations and priorities. The intelligent orchestration component can use action success/completion and time taken as measures of improvement. Likewise, the restoration engine can use unsupervised machine learning algorithms to self-learn from previous cyber threat incidents to get better at healing the system being protected to mitigate the cyber threat while minimizing an impact on the system being protected, in particular by classifying a graph of the nodes involved in a cyber attack. Referring back to Figure 3B, the restoration engine can comprise a cyber threat analysis module having one or more neural networks configured to analyze the graph generated by the restoration engine during a cyber attack, such as graph 300 in Figure 3B. The neural network(s) are trained on data of graphs from previous attacks, which may be labelled, or preferably unlabeled data, or some combination of both, to learn how to group and identify different types of cyber attacks. This classification is at a high level, and whilst based on all of the interactions within the graph is typically defined by certain key interactions. For example, cyber attacks may be classified, based on their graphs, as ransomware attacks, data theft attacks, etc. etc. Due to the non-Euclidean structure of cyber incidents, message passing neural networks have been found to be particularly suited to this type of task. Additionally, due to the time-varying nature of the graphs being classified by the neural network (which vary as a cyber attack progresses, moving through a network, and as response and remediation actions are performed), message passing neural networks have been used in combination with sequential models such as recurrent neural networks or transformers to analyze incidents over time.

[0181] The cyber security restoration engine can also use one or more unsupervised machine learning algorithms, as a self-learning entity, to have an ability to learn how to restore the one or more nodes in the graph of the protected system back to the trusted operational state while still mitigating against the cyber threat so the cyber security restoration engine gets better over time of a deployment of the cyber security restoration engine by learning from previous restoration attempts (e.g. action success/completion and time taken as measures, action effectiveness as a measure, etc., as well as including or adapting changes to previous recommendations made by the human security team.

[0182] The cyber threat detection engine, the autonomous response engine, the prediction engine all perform their machine-learned task and send inputs to each other to assist in determining what nodes are impacted, what cyber threat is causing the problems, and how the cyberattack likely occurred and will progress based upon possible mitigation and restoration actions taken so that the restoration engine can rely on the determinations by the Artificial Intelligence in those AI-based engines to give the restoration engine a fantastic starting point for figuring out what is the system being protected is trying to recover from and then a best way to restore the nodes in the system.

[0183] There are four discrete AI-based engines working to achieve aims with their own machine learning approaches. Each separate AI contributes data that has been processed intelligently through machine learning approaches and then hands over the processed behavioral metrics to another AI engine which then performs its own individualized machine-learned task.

[0184] The prevent engine in conducting simulations can use AI Analyst with external data input (e.g., crowdstrike) and cooperate with the detection engine to identify an infected patient zero and additional devices actually compromised and/or directly linked to devices actually compromised in need of remediation. The linked devices or the activity may not be directly visible to the detection engine alone and the external data input fills in the big picture. The cyber security restoration engine to restore the protected system can potentially use the external data input that the system is receiving from third party integrations (e.g., from host-based agents from 3rd party vendors, antivirus and-based testing antivirus, etc.) to identify patient zero of the attack, identify, where the attack has happened and is happening, identify devices that the system reasonably believes are linked to the compromised entity, and recommend remediation or perform remediation via AI alone, and/or AI in combination with human assistance. The cyber security restoration engine can restore the protected system back to a state before a compromise (e.g. abnormalities started) by a cyber threat occurred to the protected system. The cyber security restoration engine restores nodes in the protected system to cyberattacks in progress – so heals in real time, as an attack happens, as well as can assist in healing after an attack has taken place.

[0185] The trusted operational state of a node can be an operational state for a date and time before the earliest detection of a possible compromise of a node in the graph (device and/or user account) plus a threshold buffer amount of time.

[0186] In an example, the detection engine can use historic IaaS data on virtual resource usage to identify errant virtual resources and the autonomous response engine to spin down those resources or disable overactive microservices like lambdas. In another example, the detection engine can use historic IaaS data on virtual resource usage to understand when a client is undergoing some kind of DDOS and the autonomous response engine acts to do scaling to handle the load until the overload is over. The restoration engine can recommend controlling the scaling when the system understands deliberate overloading of traffic is occurring and then bringing that scaling back down again to assist their service architectures to deal with situations when some cyber threat is trying to overload those systems to bring that customer down.

[0187] In another example, the cyber security restoration engine to restore the protected system can use historic source codebase information and modelling from the AI models in the detection engine for development to revert commits and code changes that potentially introduce bad or compromised code. The cyber security restoration engine to restore the protected system can also use historic records of a source code database information to find out when during the development of a product that the cyber-attack occurred on the source code in order to restore the source code back to the state before the compromise occurred, as well as use historic code base analysis and understanding to identify supply chain and products vulnerable to bad code /

compromised code and sending an update package / at least a notice to revert those products and further prevent the source code vulnerabilities from trickling down the supply chains from the vendor to the end user. Once file data of a cyber threat is identified, then that file data and its characteristics are captured in an inoculation package and then cascade that file information to each cyber security appliance in the fleet of cyber security appliances, and quarantine the identical and very similar files in order to remove them from all of the environments before anything can spread even more than it has via immediate remediation and also using the system's own inoculation data.

[0188] In an example, the autonomous response engine can stop a device that is infected from connecting to other nodes. In addition, the autonomous response engine can restrict reading and writing traffic and/or types of data/information being communicated in that traffic to restrict traffic movement and process activity to nodes close to an entity that the system thinks is performing erroneously or infected.

[0189] Referring to Figure 4, the autonomous response engine is configured to use one or more Application Programming Interfaces to translate desired mitigation actions for nodes (devices, user accounts, etc.) into a specific language and syntax utilized by that device, user account, etc. from potentially multiple different vendors being protected in order to send the commands and other information to cause the desired mitigation actions to change, for example, a behavior of a detected threat of a user and/or a device acting abnormal to the normal pattern of life. The selected mitigation actions on the selected nodes minimize an impact on other parts of the system being protected (e.g. devices and users) that are i) currently active in the system being protected and ii) that are not in breach of being outside the normal behavior benchmark. The autonomous response engine can have a discovery module to i) discover capabilities of each node being protected device and the other cyber security devices (e.g. firewalls) in the system being protected and ii) discover mitigation actions they can take to counter and/or contain the detected threat to the system being protected, as well as iii) discover the communications needed to initiate those mitigation actions.

[0190] For example, the autonomous response engine cooperates and coordinates with an example set of network capabilities of various network devices. The network devices may have various capabilities such as identity management including setting user permissions, network security controls, firewalls denying or granting access to various ports, encryption capabilities, centralize logging, antivirus anti-malware software quarantine and immunization, patch management, etc., and also freeze any similar, for example, network activity, etc. triggering the harmful activity on the system being protected.

[0191] Accordingly, the autonomous response engine will take an autonomous mitigation action to, for example, shutdown the device or user account, block login failures, perform file modifications, block network connections, restrict the transmission of certain types of data, restrict a data transmission rate, remove or restrict user permissions, etc. The autonomous response engine for an email system could initiate example mitigation actions to either remedy or neutralize the tracking link, when determined to be the suspicious covert tracking link, while not stopping every email entering the email domain with a tracking link, or hold the email communication entirely if the covert tracking link is highly suspicious, and also freeze any similar, for example, email activity triggering the harmful activity on the system being protected.

[0192] The autonomous response engine has a default set of autonomous mitigation actions shown on its user interface that it knows how to perform when the different types of cyber threats are equal to or above a user configurable threshold posed by this type of cyber threat. The autonomous response engine is also configurable in its user interface to allow the user to augment and change what type of automatic mitigation actions, if any, the autonomous response engine may take when different types of cyber threats that are equal to or above the configurable level of threat posed by a cyber threat.

[0193] The autonomous response engine can also reference its artificial intelligence trained to perform mitigation actions. Again, the autonomous response engine has an administrative tool in its user interface to program/set what autonomous mitigation actions the autonomous response engine can take, including types of mitigation actions and specific mitigation actions the autonomous response engine is capable of, when the cyber-threat module in the detection engine indicates the threat risk parameter is equal to or above the actionable threshold, selectable by the cyber professional. The cyber professional can also indicate what types of mitigation actions can be performed for different users and parts of the system as well as what actions need the cyber professional to approve. Again, the autonomous response engine can also reference a default library of mitigation actions, types of mitigation actions and specific mitigation actions the autonomous response engine is capable of on a particular node.

[0194] Referring to Figure 6, the prediction engine 105 using Artificial Intelligence-based simulations is communicatively coupled to a cyber security appliance 100, an open source (OS) database server 122, an email system 115, one or more endpoint computing devices 101A-B, and an IT network system 112 with one or more entities 130-142, over one or more networks 110/112 in the system being protected.

[0195] The prediction engine 105 with Artificial Intelligence-based simulations is configured to integrate with the cyber security appliance 100 and cooperate with components within the cyber security appliance 100 installed and protecting the network from cyber threats by making use of outputs, data collected, and functionality from two or more of a data store, other modules, and one or more AI models already existing in the cyber security appliance 100.

[0196] The prediction engine 105 may include a cyber threat generator module to generate many different types of cyber threats with the past historical attack patterns to attack the simulated system to be generated by the simulated attack module that will digitally/virtually replicate the system being protected, such as a phishing email generator configured to generate one or more automated phishing emails to pentest the email defenses and/or the network defenses provided by the cyber security appliance 100. For example, the system being protected can be an email system and then the phishing email generator may be configured to cooperate with the trained AI models to customize the automated phishing emails based on the identified data points of the organization and its entities.

[0197] The email module and network module may use a vulnerability tracking module to track and profile, for example, versions of software and a state of patches and/or updates compared to a latest patch and/or update of the software resident on devices in the system/network. The vulnerability tracking module can supply results of the comparison of the version of software as an actual detected vulnerability for each particular node in the system being protected, which is utilized by the node exposure score generator and the prediction engine 105 with Artificial Intelligence-based simulations in calculating 1) the spread of a cyber threat and 2) a prioritization of remediation actions on a particular node compared to the other network nodes with actual detected vulnerabilities. The node exposure score generator is configured to also factor in whether the particular node is exposed to direct contact by an entity generating the cyber threat (when the threat is controlled from a location external to the system e.g. network) or the particular node is downstream of a node exposed to direct contact by the entity generating the cyber threat external to the network.

[0198] The node exposure score generator and the simulated attack module in the prediction engine 105 cooperate to run the one or more hypothetical simulations of an actual detected cyber threat incident and/or a hypothetical cyberattack incident to calculate the node paths of least resistance in the virtualized instance/modeled instance of the system being protected. The progress through the node path(s) of least resistance through the system being protected are plotted through the various simulated instances of components of the graph of the system being protected until reaching a suspected end goal of the cyber-attack scenario, all based on historic knowledge of connectivity and behavior patterns of users and devices within the system under analysis. The simulated attack module, via a simulator and/or a virtual network clone creator, can be programmed to model and work out the key paths and devices in the system (e.g. a network, with its nets and subnets,) via initially mapping out the system being protected and querying the cyber security appliance on specific's known about the system being protected by the cyber security appliance 100. The simulated attack module is configured to search and query, two or more of i) a data store, ii) modules in the detection engine, and iii) the one or more Artificial Intelligence (AI) models making up the cyber security appliance 100 protecting the actual network under analysis from cyber threats, on what, i) the data store, ii) the modules, and iii) the one or more AI models in the cyber security appliance 100, already know about the nodes of the system, under analysis to create the graph of nodes of the system being protected. Thus, the prediction engine 105 with Artificial Intelligence-based simulations is configured to construct the graph of the virtualized version of the system from knowledge known and stored by modules, a data store, and one or more AI models of a cyber security appliance 100 protecting an actual network under analysis. The knowledge known and stored is obtained at least from ingested traffic from the actual system under analysis. Thus, the virtualized system, and its node components/accounts connecting to the network, being tested during the simulation are up to date and accurate for the time the actual system under analysis is being tested and simulated because the prediction engine 105 with Artificial Intelligence-based simulations is configured to obtain actual network data collected by two or more of 1) modules, 2) a data store, and 3) one or more AI models of a cyber security appliance protecting the actual network under analysis from cyber threats. The simulated attack module will make a model incorporating the actual data of the system through the simulated versions of the nodes making up that system for running simulations on the simulator. Again, a similar approach is taken when the simulated attack module uses a clone creator to spin up and create a virtual clone of the system being protected with virtual machines in the cloud.

[0199] The prediction engine 105 with Artificial Intelligence-based simulations is configured to simulate the compromise of a spread of the cyber threat being simulated in the simulated cyber-attack scenario, based on historical and/or similar cyber threat attack patterns, between the devices connected to the virtualized network, via a calculation on an ease of transmission of the cyber threat algorithm, from 1) an originally compromised node by the cyber threat, 2) through to other virtualized/ simulated instances of components of the virtualized network, 3) until reaching a suspected end goal of the cyber-attack scenario, including key network devices. The prediction engine 105 with Artificial Intelligence-based simulations also calculates how likely it would be for the cyber-attack to spread to achieve either of 1) a programmable end goal of that cyber-attack scenario set by a user, or 2) set by default an end goal scripted into the selected cyber-attack scenario.

[0200] The email module and the network module can include a profile manager module. The profile manager module is configured to maintain a profile tag on all of the devices connecting to the actual system/network under analysis based on their behavior and security characteristics and then supply the profile tag for the devices connecting to the virtualized instance of the system/network when the construction of the graph occurs. The profile manager module is configured to maintain a profile tag for each device before the simulation is carried out; and thus, eliminates a need to search and query for known data about each device being simulated during the simulation. This also assists in running multiple simulations of the cyberattack in parallel.

[0201] The prediction engine 105 with Artificial Intelligence-based simulations module is configured to construct the graph of the virtualized system, e.g. a network with its nets and subnets, where two or more of the devices connecting to the virtualized network are assigned with different weighting resistances to malicious compromise from the cyber-attack being simulated in the simulated cyber-attack scenario based on the actual cyber-attack on the virtualized instance of the network and their node vulnerability score. In addition to a weighting resistance to the cyberattack, the calculations in the model for the simulated attack module factor in the knowledge of a layout and connection pattern of each particular network device in a network, an amount of connections and/or hops to other network devices in the network, how important a particular device (a key importance) determined by the function of that network device, the user(s) associated with that network device, and the location of the device within the network. Note, multiple simulations can be conducted in parallel by the orchestration module. The simulations can occur on a periodic regular basis to pentest the cyber security of the system and/or in response to a detected ongoing cyberattack in order to get ahead of the ongoing cyberattack and predict its likely future moves. Again, the graph of the virtualize instance of the system is created with two or more of 1) known characteristics of the network itself, 2) pathway connections between devices on that network, 3) security features and credentials of devices and/or their associated users, and 4) behavioral characteristics of the devices and/or their associated users connecting to that network, which all of this information is obtained from what was already know about the network from the cyber security appliance.

[0202] During an ongoing cyberattack, the simulated attack module is configured to run the one or more hypothetical simulations of the detected cyber threat incident and feed details of a detected incident by a cyber threat module in the detection engine into the collections module of the prediction engine 105 using Artificial Intelligence-based simulations. The simulated attack module is configured to run one or more hypothetical simulations of that detected incident in order to predict and assist in the triggering an autonomous response by the autonomous response engine and then restoration by the restoration engine to the detected incident.

[0203] The simulated attack module ingests the information for the purposes of modeling and simulating a potential cyberattacks against the network and routes that an attacker would take through the network. The simulated attack module can construct the graph of nodes 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 cyber-attack would use during the cyber-attack, via modeling the cyber-attack on at least one of 1) a simulated device version and 2) a virtual device version of the system being protected under analysis. Correspondingly, the calculated likelihood of the compromise and timeframes for the spread of the cyberattack is tailored and accurate to each actual device/user account (e.g. node) being simulated in the system because the cyber-attack scenario is based upon security credentials and behavior characteristics from actual traffic data fed to the modules, data store, and AI models of the cyber security appliance.

[0204] The prediction engine 105 with its Artificial Intelligence trained on how to conduct and perform cyberattack in a simulation in either a simulator or in a clone creator spinning up virtual instances on virtual machines will take a sequence of actions and then evaluate the actual impact after each action in the sequence, in order to yield a best possible result to contain/mitigate the detected threat while minimizing the impact on other network devices and users that are i) currently active and ii) not in breach, from different possible actions to take. Again, multiple simulations can be run in parallel so that the different sequences of mitigation actions and restoration actions can be evaluated essentially simultaneously. The prediction engine 105 with Artificial

Intelligence-based simulations in the prediction engine 105 is configured to use one or more mathematical functions to generate a score and/or likelihood for each of the possible actions and/or sequence of multiple possible actions that can be taken in order to determine which set of actions to choose among many possible actions to initiate. The one or more possible actions to take and their calculated scores can be stacked against each other to factor 1) a likelihood of containing the detected threat acting abnormal with each possible set of actions, 2) a severity level of the detected threat to the network, and 3) the impact of taking each possible set of actions i) on users and ii) on devices currently active in the network not acting abnormal to the normal behavior of the network, and then communicate with the cyber threat detection engine, the autonomous response engine, and the cyber-security restoration engine, respectively, to initiate the chosen set of actions to cause a best targeted change of the behavior of the detected threat acting abnormal to the normal pattern of life on the network while minimizing the impact on other network devices and users that are i) currently active and ii) not in breach of being outside the normal behavior benchmark. The prediction engine cooperates with the AI models modelling a normal pattern of life for entities/nodes in the system being protected.

[0205] The simulated attack module is programmed itself and can cooperate with the artificial intelligence in the restoration engine to factor an intelligent prioritization of remediation actions and which nodes (e.g. devices and user accounts) in the simulated instance of the system being protected should have a priority compared to other nodes. This can also be reported out to assist in allocating human security team personnel resources that need human or human approval to restore the nodes based on results of the one or more hypothetical simulations of the detected incident.

[0206] Note, the prediction engine does not need to not calculate every theoretically possible path from the virtualized instance of the source device to the end goal of the cyber-attack scenario but rather a set of the most likely paths, each time a hop is made from one node in the virtualized network to another device in the virtualized network, in order to reduce an amount of computing cycles needed by the one or more processing units as well as an amount of memory storage needed in the one or more non-transitory storage mediums.

[0207] Figure 8 illustrates a block diagram of an embodiment of the AI-based cyber security appliance 100 with the cyber security restoration engine and other Artificial Intelligence-based engines plugging in as an appliance platform to protect a system. The probes and detectors monitor, in this example, email activity and IT network activity to feed this data to determine what is occurring in each domain individually to their respective modules configured and trained to understand that domain's information as well as correlate causal links between these activities in these domains to supply this input into the modules of the cyber security appliance 100. The network can include various computing devices such as desktop units, laptop units, smart phones, firewalls, network switches, routers, servers, databases, Internet gateways, etc.

[0208] Figure 17 illustrates a diagram of an embodiment of the cyber security restoration engine cooperating with the other Artificial Intelligence-based engines of the cyber security system to track and understand the cyber threat identified by the other Artificial Intelligence-based engines as well as track the one or more mitigation actions taken to mitigate the cyber threat during the cyberattack by the other Artificial Intelligence-based engines in order to assist in intelligently restoring the protected system, while still mitigating the cyber threat attack, back to the trusted operational state before indications of when the compromised started by the cyber threat attack. The Artificial Intelligence-based cyber security system can also use 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. 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.

[0209] The first computer system 10 comprises three computers 1, 2, 3, a local server 4, and a multifunctional device 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 Local Area Network 6. Consequently, all of the computers 1, 2, 3 are able to access the local server 4 via the LAN 6 and use the functionalities of the MFD 5 via the LAN 6.

[0210] The LAN 6 of the first computer system 10 is connected to the Internet 20, which in turn provides computers 1, 2, 3 with access to a multitude of other computing devices 18 including server 30 and second computer system 40. The second computer system 40 also includes two computers 41, 42, connected by a second LAN 43.

[0211] In this exemplary embodiment of the cyber security appliance 100 with the multiple Artificial Intelligence-based engines is implemented on computer 1. Computer 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. As such, the computer system 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 a network interface for collecting the required information for the probes and other sensors collecting data from the network under analysis.

[0212] The cyber security appliance 100 in computer 1 builds and maintains a dynamic, ever-changing model of the 'normal behavior' of each user and machine within the system 10. The approach is based on Bayesian mathematics, and monitors all interactions, events, and communications within the system 10 - which computer is talking to which, files that have been created, networks that are being accessed.

[0213] For example, computer 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 second computer system 40 between 9.30 AM and midday, and is active from about 8:30 AM until 6 PM.

[0214] 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 security appliance 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 network under analysis is used as a moving benchmark, allowing the cyber security appliance 100 to spot behavior on a system that seems to fall outside of this normal pattern of life, and flags this behavior as anomalous, requiring further investigation and/or autonomous action.

[0215] 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 Artificial Intelligence model(s) 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.

[0216] 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 network by analyzing data on the activity on the network 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.

[0217] The cyber security appliance 100 can use unsupervised machine learning to works things out without pre-defined labels. In the case of sorting a series of different entities, such as animals, the system analyzes the information and works out the different classes of animals. This allows the system 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.

[0218] The cyber security appliance 100's 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.

[0219] 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.

[0220] Using RBE, the cyber security appliance 100's AI models are able to constantly adapt themselves, in a computationally efficient manner, as new information becomes available to the system. The cyber security appliance 100's AI models continually recalculate threat levels in the light of new evidence, identifying changing attack behaviors where conventional signature-based methods fall down.

[0221] Training a 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 model can be to find a set of weights and biases that have low loss, on average, across all examples.

[0222] 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 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 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.

[0223] The methods and systems shown in the Figures and discussed in the text herein can be coded to be performed, at least in part, 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. 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.

[0224] A computing system can be, wholly or partially, part of one or more of the server or client computing devices in accordance with some embodiments. Components of the computing system can include, but are not limited to, a processing unit having one or more processing cores, a system memory, and a system bus that couples various system components including the system memory to the processing unit.

[0225] Computing devices

[0226] Figure 15 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 including the multiple Artificial Intelligence-based engines discussed herein.

[0227] The computing device may include one or more processors or 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.

[0228] Computing device 602 typically includes a variety of computing machine-readable media. 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, computing machine-readable media use includes storage of information, such as computer-readable instructions, data structures, other executable software, or other data. Computer-storage 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. 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.

[0229] 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.

[0230] 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 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.

[0231] 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.

[0232] 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 most effectively convey the substance of their work 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. A module may be implemented in hardware electronic components, software components, and a combination of both. A software engine is a core component of a complex system consisting of hardware and 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.

[0233] 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.

[0234] 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.

Claims

1. An apparatus, comprising:

a cyber security restoration engine to prioritize nodes in a graph of nodes in a computer network or system that are involved in a cyber attack for remediation actions, for each node, by being configured to perform operations as follows:

determine one or more edges linking the node to other nodes in the graph, the edges representing interactions between two nodes;

obtain metadata for each of the one or more edges, the metadata being indicative of a type of interaction between two nodes connected by the edge and the roles of the two nodes in that interaction;

determine, for each of the one or more edges, a severity score representing how severe the interaction represented by that edge is within the context of the cyber attack, based on the metadata of that edge; and

determine a severity score for the node by combining the severity score for each of the one or more edges connected to the node;

wherein the cyber security restoration engine is further configured to prioritize nodes for remediation action based on the severity scores for the nodes, with nodes having a higher severity score being prioritized above nodes having a lower severity score; and

where any instructions for the cyber security restoration engine are stored in an executable format on one or more non-transitory computer readable mediums, which are executable by one or more processors.

2. The apparatus of claim 1, wherein the type of interaction between two nodes may be one or more of a lateral movement of data, an encryption interaction, and a command and control interaction.

3. The apparatus of claim 1, wherein the cyber security restoration engine is further configured to determine, for each node, one or more remediation actions to remediate

the node back to a trusted operational state based on the metadata of the edges connected to that node.

4. The apparatus of claim 3, wherein at least one of the one or more remediation actions to remediate the node back to a trusted operational state are configured to be performed while a cyber attack is ongoing.

5. The apparatus of claim 3, wherein the cyber security restoration engine is configured to prioritize nodes for remediation action based on an expected time to complete the determined remediation actions for each node, with nodes having a lower time to complete the determined remediation actions being prioritized above nodes having a higher time to complete the determined remediation actions.

6. The apparatus of claim 5, wherein nodes are prioritized for remediation action first based on the severity scores for the nodes and secondly, for any nodes that have equal severity scores, based on the time to complete the determined remediation actions.

7. The apparatus of claim 1, further comprising an autonomous response engine configured to take one or more autonomous actions to respond to a cyber threat;

wherein the metadata for a given edge comprises information regarding any autonomous actions taken by the autonomous response engine in relation to the interaction represented by that edge, such that the severity score for the edge is at least in part based on any autonomous actions taken by the autonomous response engine.

8. The apparatus of claim 1, further comprising a user interface module configured to receive input from a human cyber security operative;

wherein the metadata for a given edge comprises information regarding input from the human cyber security operative about any actions taken by a human in relation to the interaction represented by that edge, such that the severity score for the edge is at least in part based on any actions taken by a human.

9. The apparatus of claim 1, further comprising a cyber security prediction engine configured to simulate cyber attacks to determine how vulnerable a node is to cyber attack, wherein the metadata for a given edge further comprises information regarding the vulnerability of one or both of the nodes connected to the edge, such that the severity score for the edge is at least in part based on the vulnerability of one or both of the nodes.

10. The apparatus of claim 1, further comprising a cyber security prediction engine configured to simulate cyber attacks to determine an importance of a node based on potential damage that could be caused by a node if that node were compromised in a cyber attack;

wherein the metadata for a given edge further comprises information regarding the importance of one or both of the nodes connected to the edge, such that the severity score for the edge is at least in part based on the importance of one or both of the nodes.

11. The apparatus of claim 1, further comprising a user interface module, the user interface module configured to present the nodes to the user with an indication of their priority as determined based on the severity scores.

12. The apparatus of claim 11, wherein the user interface module is further configured to present, for each node, the severity scores for the edges combined to give the severity score for the node.

13. The apparatus of claim 11, wherein the user interface module is further configured to present, for each node, a description of the interaction represented by each edge connected to that node.

14. The apparatus of claim 1, further comprising a cyber threat detection engine comprising one or more artificial intelligence based analysis modules configured to

analyze the computer network or system to determine nodes that may be involved in a cyber attack by comparing the activity of nodes to a normal pattern of life for each node and to analyze any differences between a present activity of a node and the normal pattern of life for that node;

wherein the cyber security restoration engine is configured to communicate with the one or more artificial intelligence based analysis modules of the cyber threat detection engine to determine, for each node, one or more remediation actions to restore the node back to a trusted operational state based on the normal pattern of life for that node and/or a present activity of that node.

15. The apparatus of claim 1, further comprising a cyber threat analysis module having a neural network configured to analyze the graph of the computer network or system as a cyber attack is occurring to classify the type of cyber attack.

16. The apparatus of claim 15, wherein the cyber security restoration engine is further configured to determine one or more remediation actions to mitigate the cyber attack based on the classification of type of cyber attack.

17. The apparatus of claim 15, wherein the neural network is a message passing neural network.

18. The apparatus of claim 15, wherein the neural network is trained on unlabelled data comprising graphs of computer networks or systems during cyber attacks.

19. The apparatus of claim 1, wherein the cyber security restoration engine is configured to continually or at intervals update the prioritization of nodes for remediation actions as the graph of the computer network or system changes throughout a cyber attack.

20. A method for a cyber security restoration engine to prioritize nodes in a graph of nodes of a computer network or system that are involved in a cyber attack for remediation actions by, for each node, comprising:

determining one or more edges linking the node to other nodes in the graph, the edges representing interactions between two nodes;

obtaining metadata for each of the one or more edges, the metadata being indicative of a type of interaction between two nodes connected by the edge and the roles of the two nodes in that interaction;

determining, for each of the one or more edges, a severity score representing how severe the interaction represented by that edge is within the context of the cyber attack, based on the metadata of that edge;

determining a severity score for the node by combining the severity score for each of the one or more edges connected to the node; and

wherein the method further comprises prioritizing nodes for remediation action based on the severity scores for the nodes, with nodes having a higher severity score being prioritized above nodes having a lower severity score.

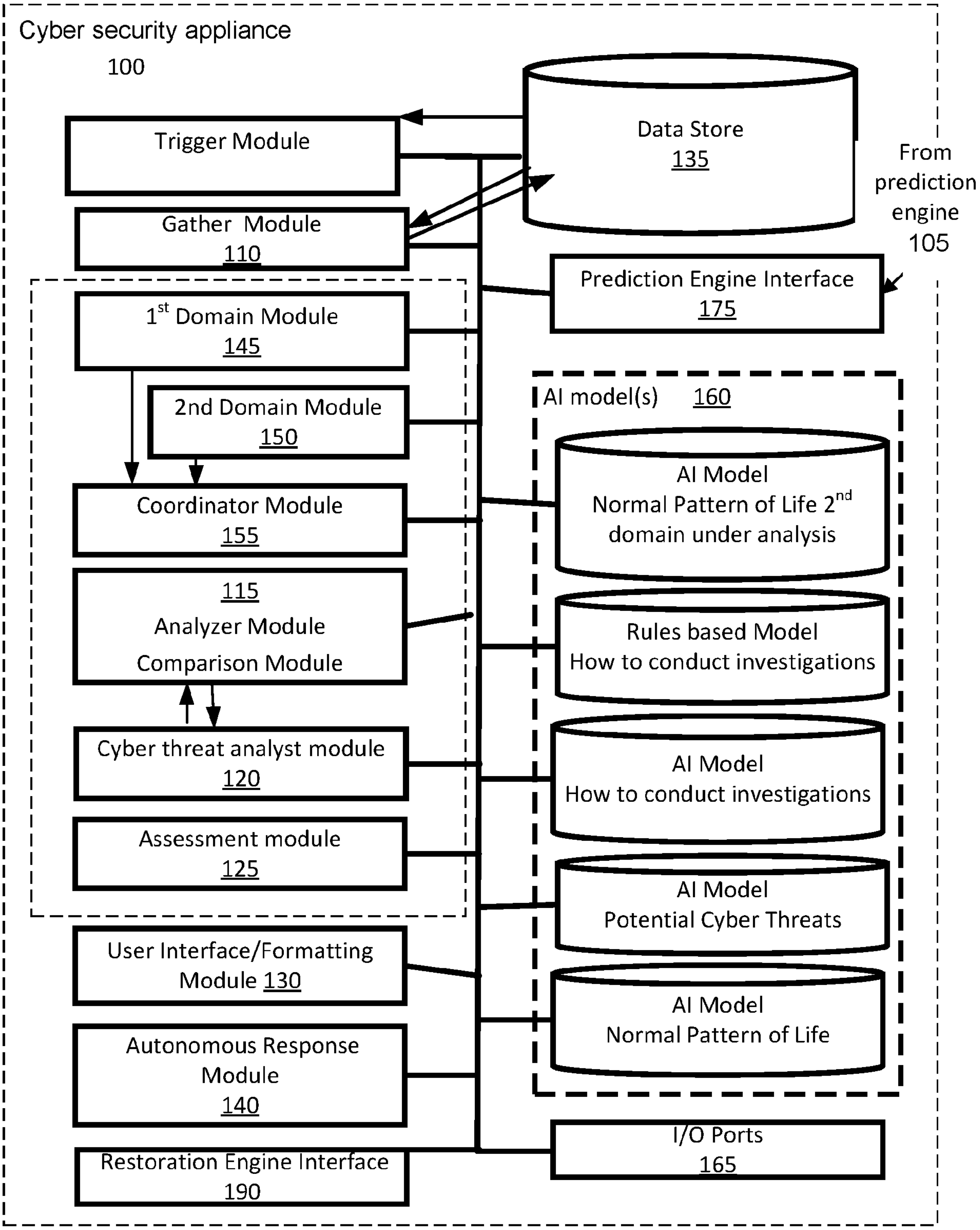

Cyber security appliance
100
Trigger Module
Data Store
135
From prediction engine
105
Gather Module
110
Prediction Engine Interface
175
1st Domain Module
145
2nd Domain Module
150
AI model(s) 160
Coordinator Module
155
AI Model
Normal Pattern of Life 2nd domain under analysis
115
Analyzer Module
Comparison Module
Rules based Model
How to conduct investigations
Cyber threat analyst module
120
AI Model
How to conduct investigations
Assessment module
125
AI Model
Potential Cyber Threats
User Interface/Formatting Module 130
AI Model
Normal Pattern of Life
Autonomous Response Module
140
I/O Ports
165
Restoration Engine Interface
190

Fig. 1

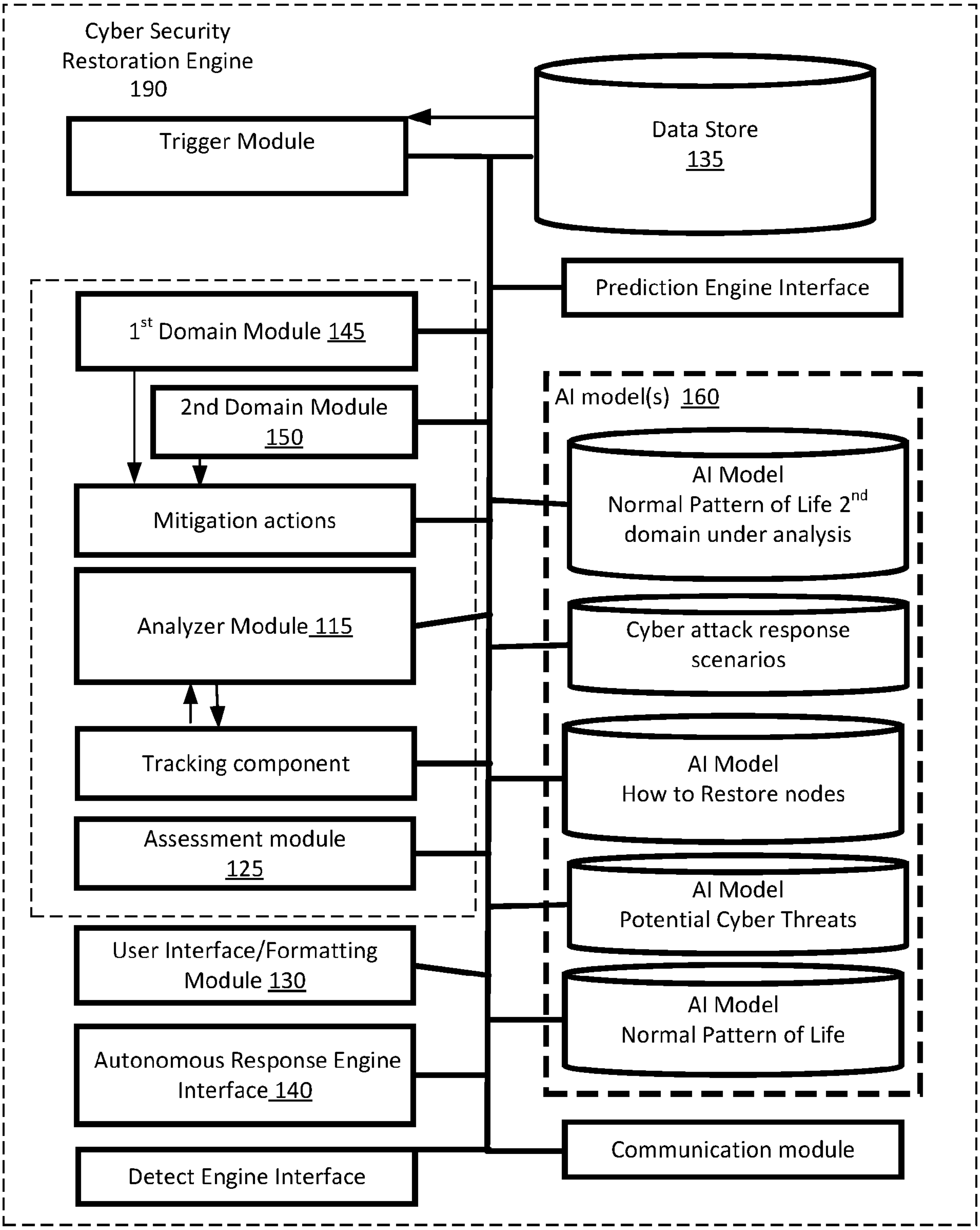
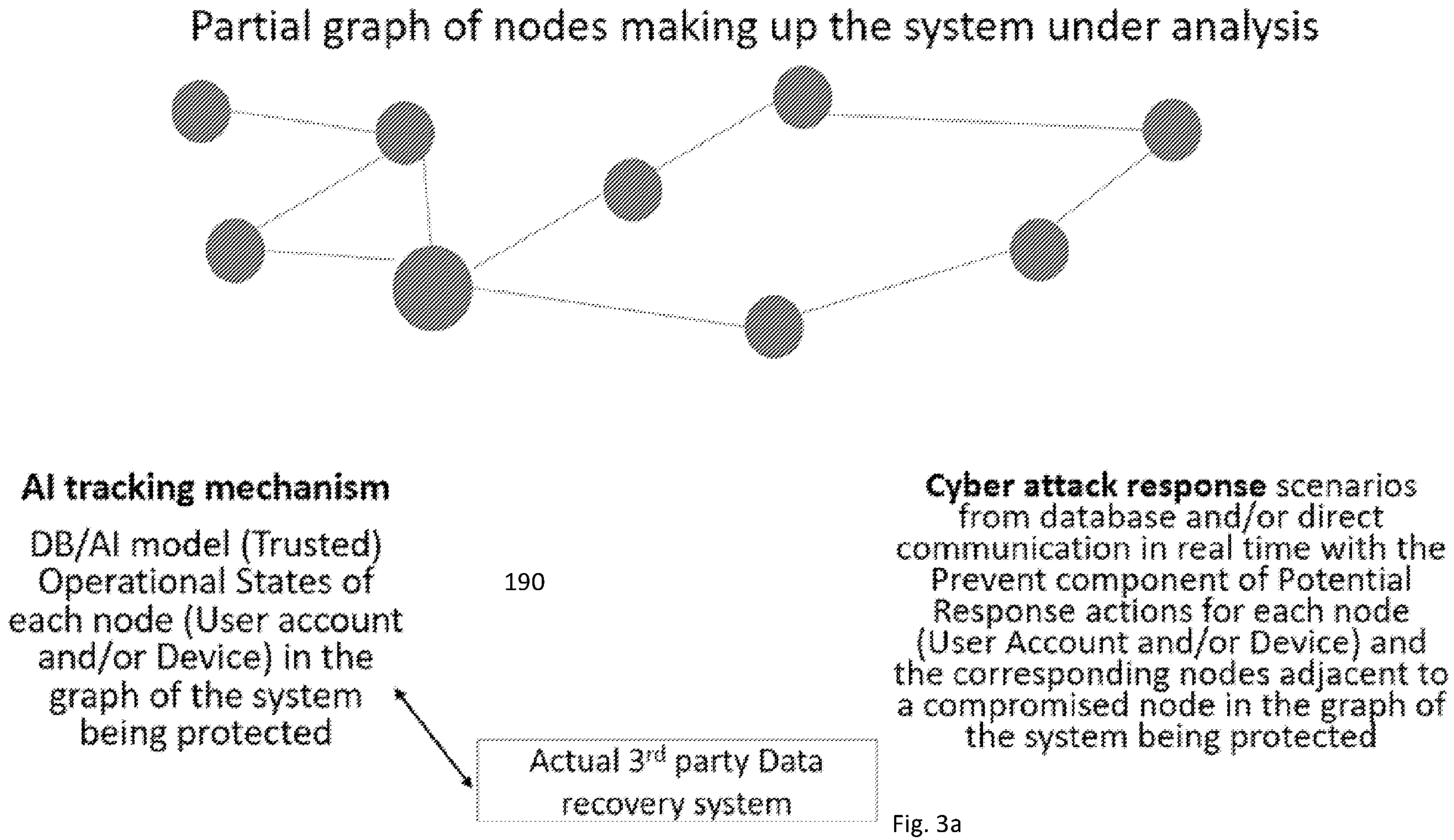

Cyber Security
Restoration Engine
190
Trigger Module
Data Store
135
Prediction Engine Interface
1st Domain Module 145
2nd Domain Module
150
Mitigation actions
Analyzer Module 115
Tracking component
Assessment module
125
User Interface/Formatting
Module 130
Autonomous Response Engine
Interface 140
Detect Engine Interface
AI model(s) 160
AI Model
Normal Pattern of Life 2nd
domain under analysis
Cyber attack response
scenarios
AI Model
How to Restore nodes
AI Model
Potential Cyber Threats
AI Model
Normal Pattern of Life
Communication module

Fig. 2



Partial graph of nodes making up the system under analysis

AI tracking mechanism

DB/AI model (Trusted) Operational States of each node (User account and/or Device) in the graph of the system being protected

190

Actual 3rd party Data recovery system

Cyber attack response scenarios from database and/or direct communication in real time with the Prevent component of Potential Response actions for each node (User Account and/or Device) and the corresponding nodes adjacent to a compromised node in the graph of the system being protected

Fig. 3a

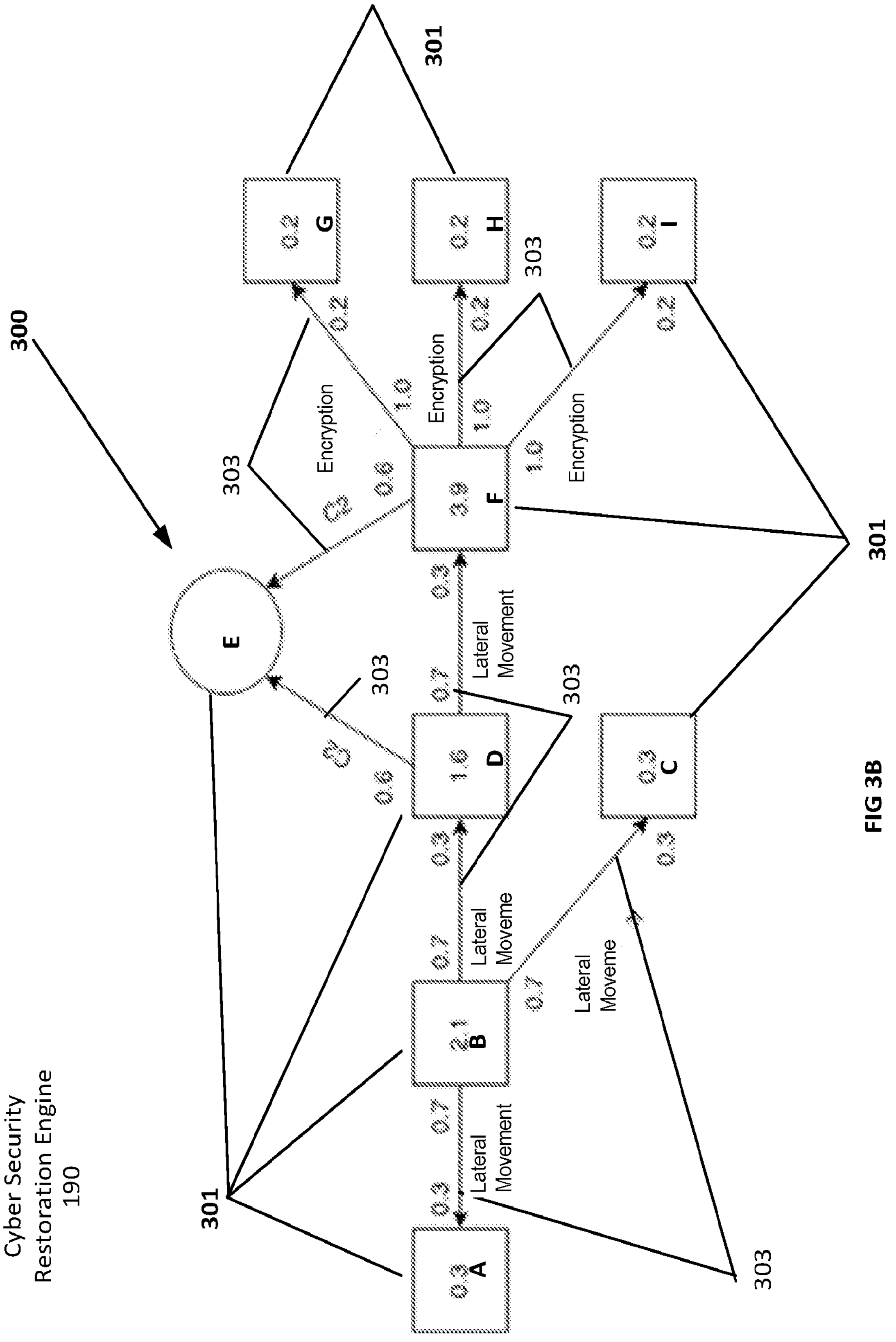

300
Cyber Security
Restoration Engine
190
301
303
A
0.3
B
2.1
C
0.3
D
1.6
E
F
3.9
G
0.2
H
0.2
I
0.2
0.3
0.7
Lateral
Movement
0.7
Lateral
Moveme
0.3
0.7
Lateral
Moveme
0.3
0.6
0.7
Lateral
Movement
0.3
0.6
1.0
Encryption
1.0
Encryption
1.0
Encryption
0.2
0.2
0.2
301
303
303
303
303
301
303

FIG 3B

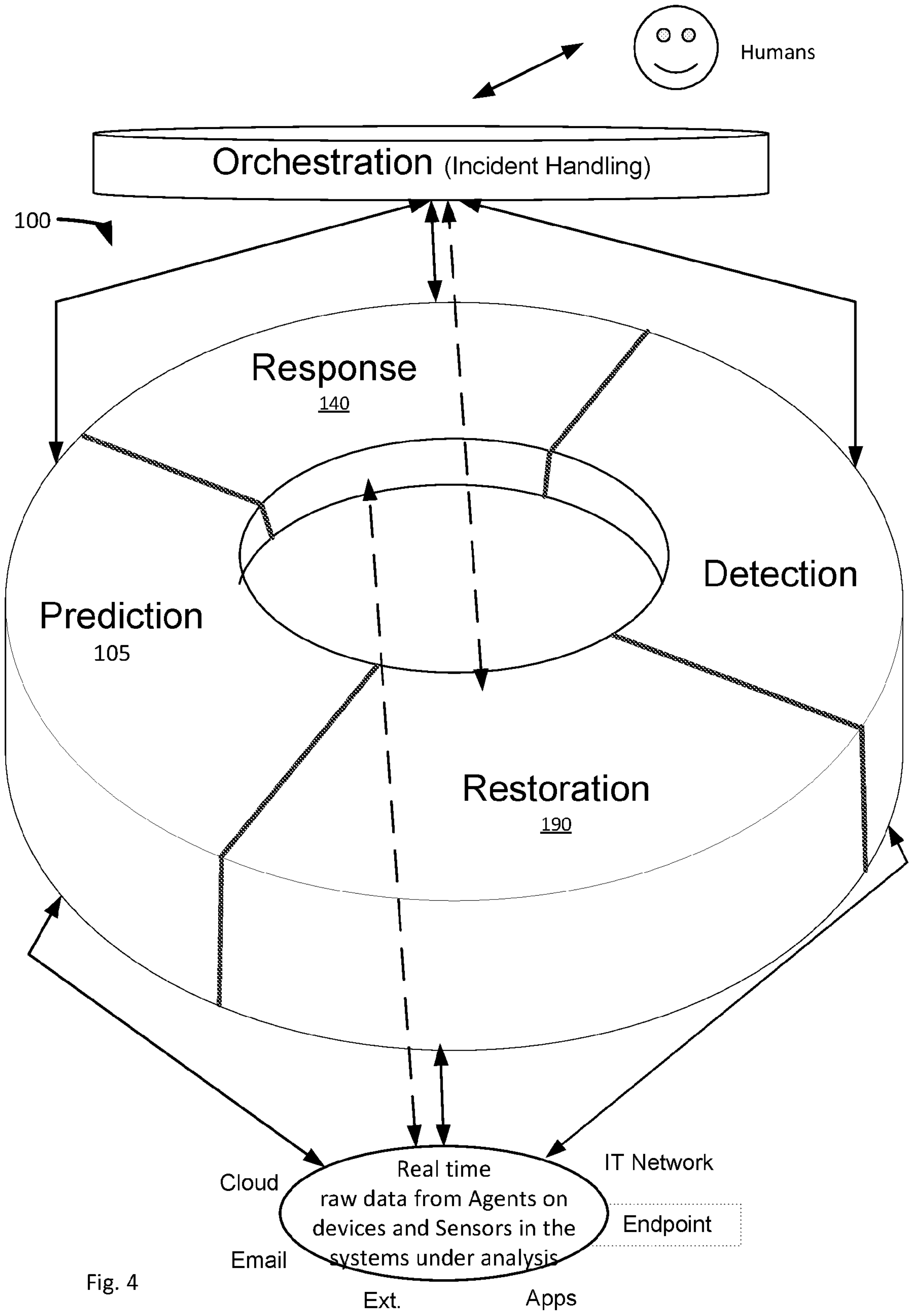
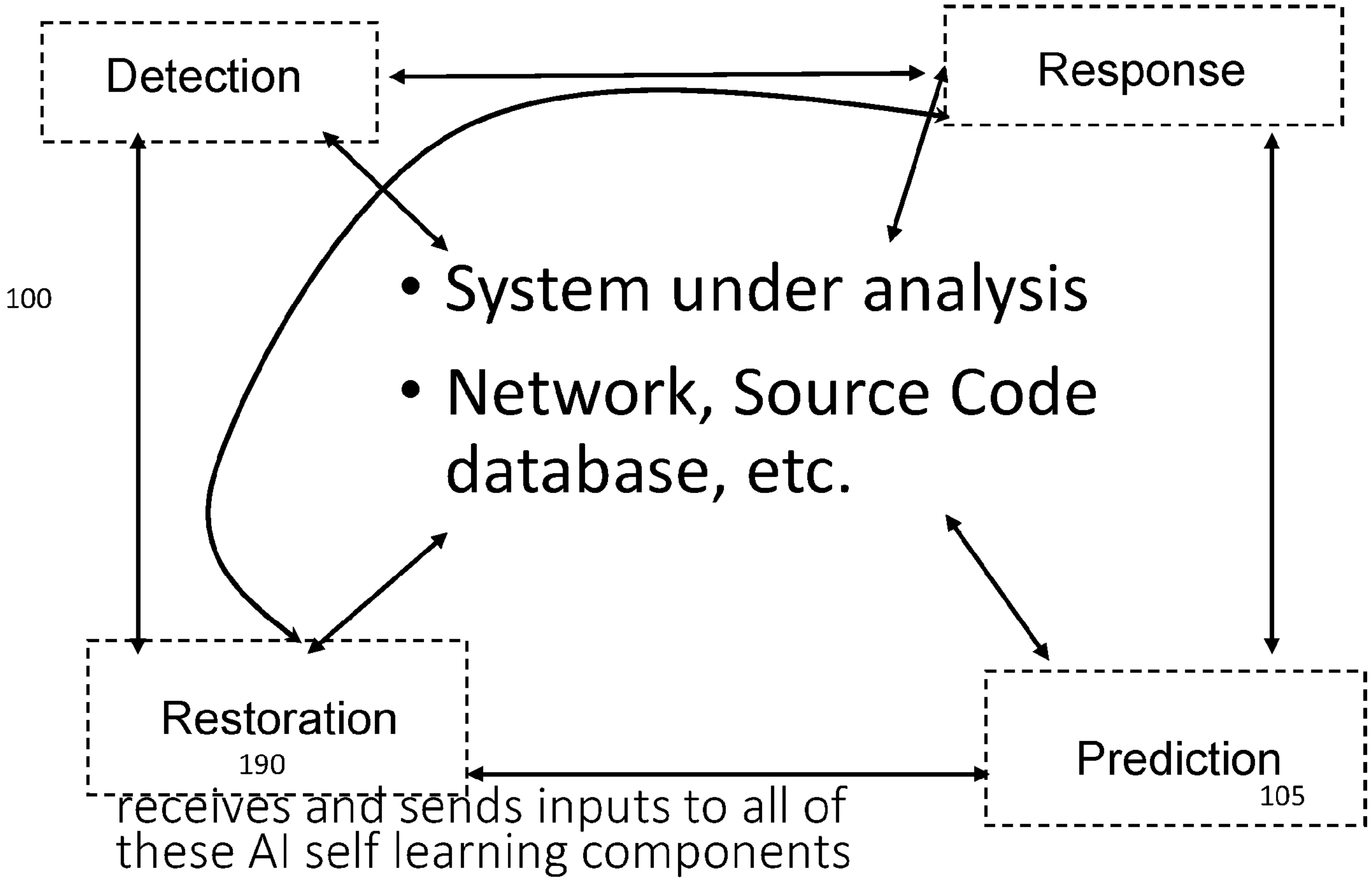

Humans
Orchestration (Incident Handling)
100
Response
140
Detection
Prediction
105
Restoration
190
Real time
raw data from Agents on
devices and Sensors in the
systems under analysis
Cloud
IT Network
Endpoint
Email
Ext.
Apps

Fig. 4


Detection
Response
100
• System under analysis
• Network, Source Code database, etc.
Restoration
190
Prediction
105
receives and sends inputs to all of these AI self learning components

The AI components have bi-directional communications with each other as well as with Agents and Sensors within the system under analysis

Fig. 5

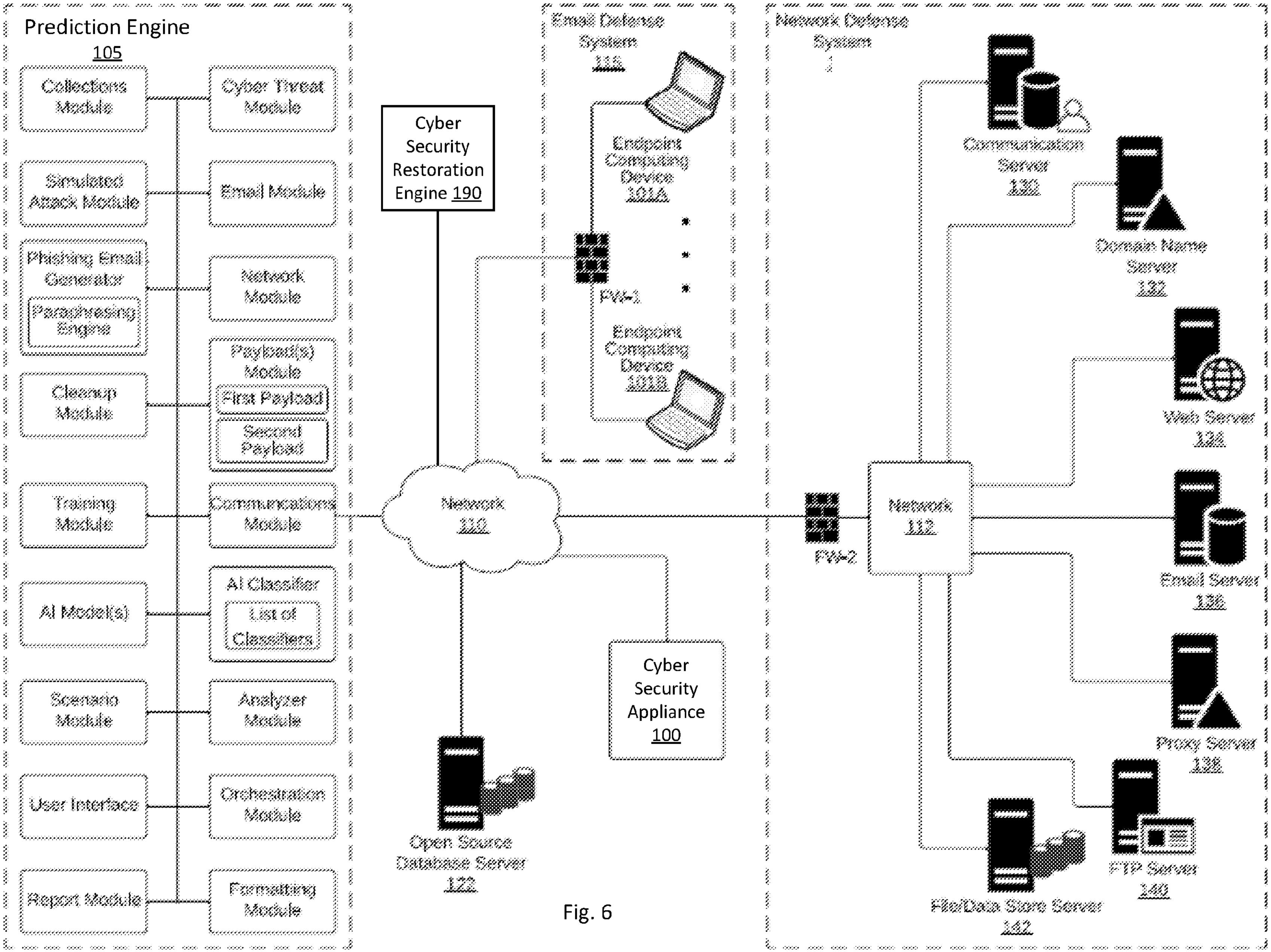

Prediction Engine
105
Collections Module
Cyber Threat Module
Simulated Attack Module
Email Module
Phishing Email Generator
Paraphrasing Engine
Network Module
Cleanup Module
Payload(s) Module
First Payload
Second Payload
Training Module
Communications Module
AI Model(s)
AI Classifier
List of Classifiers
Scenario Module
Analyzer Module
User Interface
Orchestration Module
Report Module
Formatting Module
Cyber Security Restoration Engine 190
Network
110
Open Source Database Server
122
Cyber Security Appliance
100
Email Defense System
115
Endpoint Computing Device
101A
FW-1
Endpoint Computing Device
101B
Network Defense System
1
FW-2
Network
112
Communication Server
130
Domain Name Server
132
Web Server
134
Email Server
136
Proxy Server
138
FTP Server
140
File/Data Store Server
142

Fig. 6

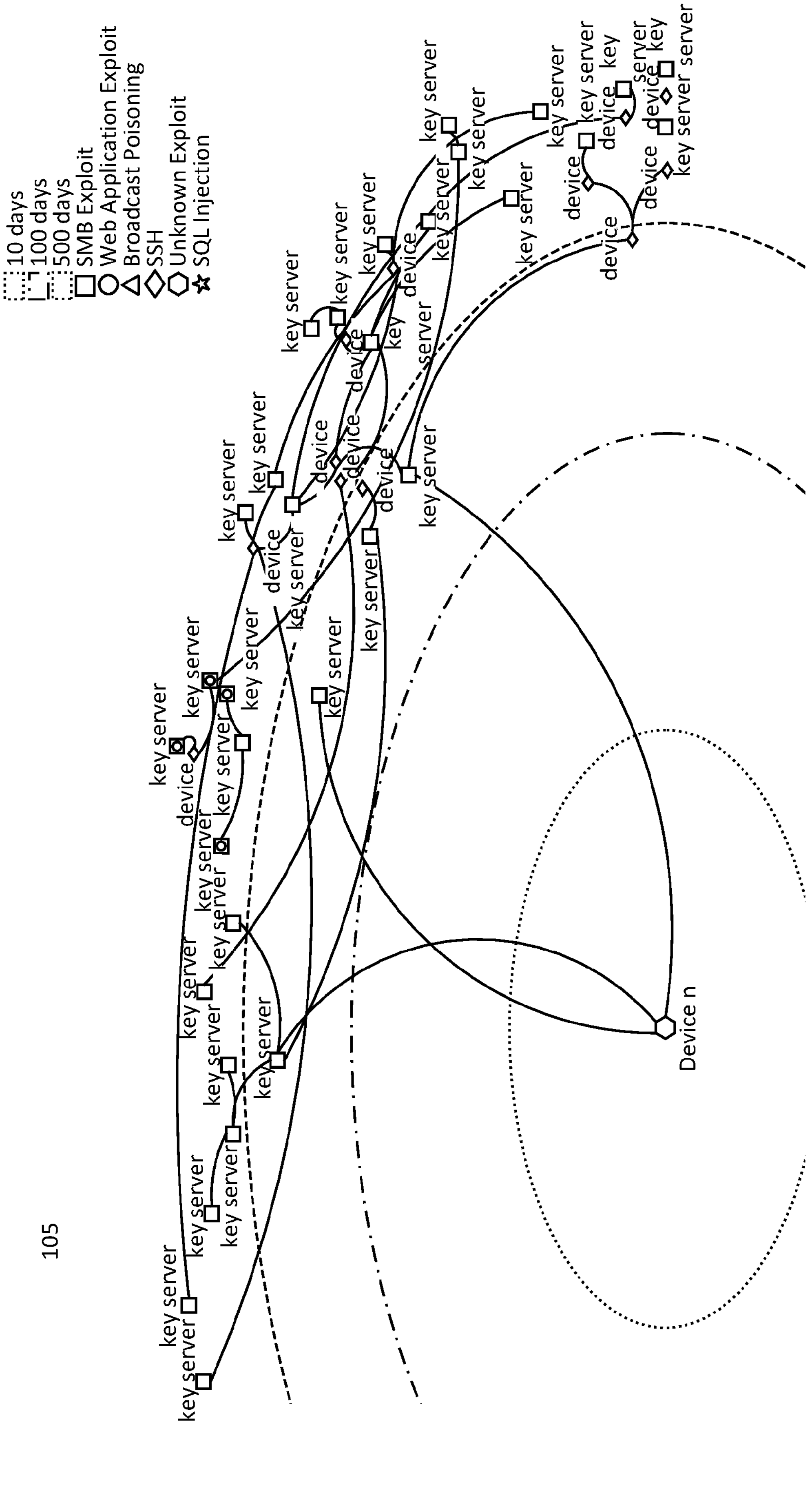

10 days
100 days
500 days
SMB Exploit
Web Application Exploit
Broadcast Poisoning
SSH
Unknown Exploit
SQL Injection
key server
device
key
server
Device n
105

Fig. 7

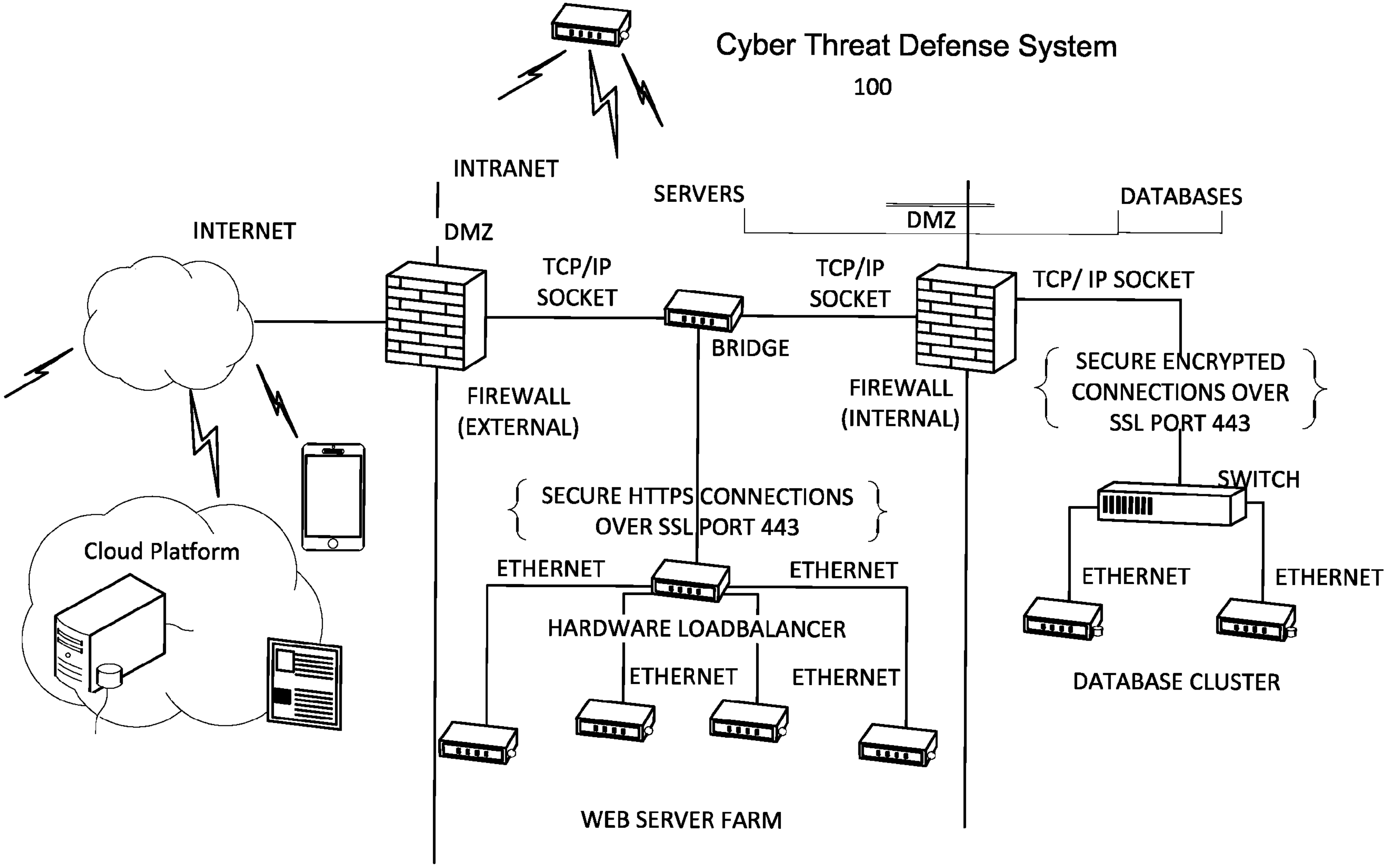
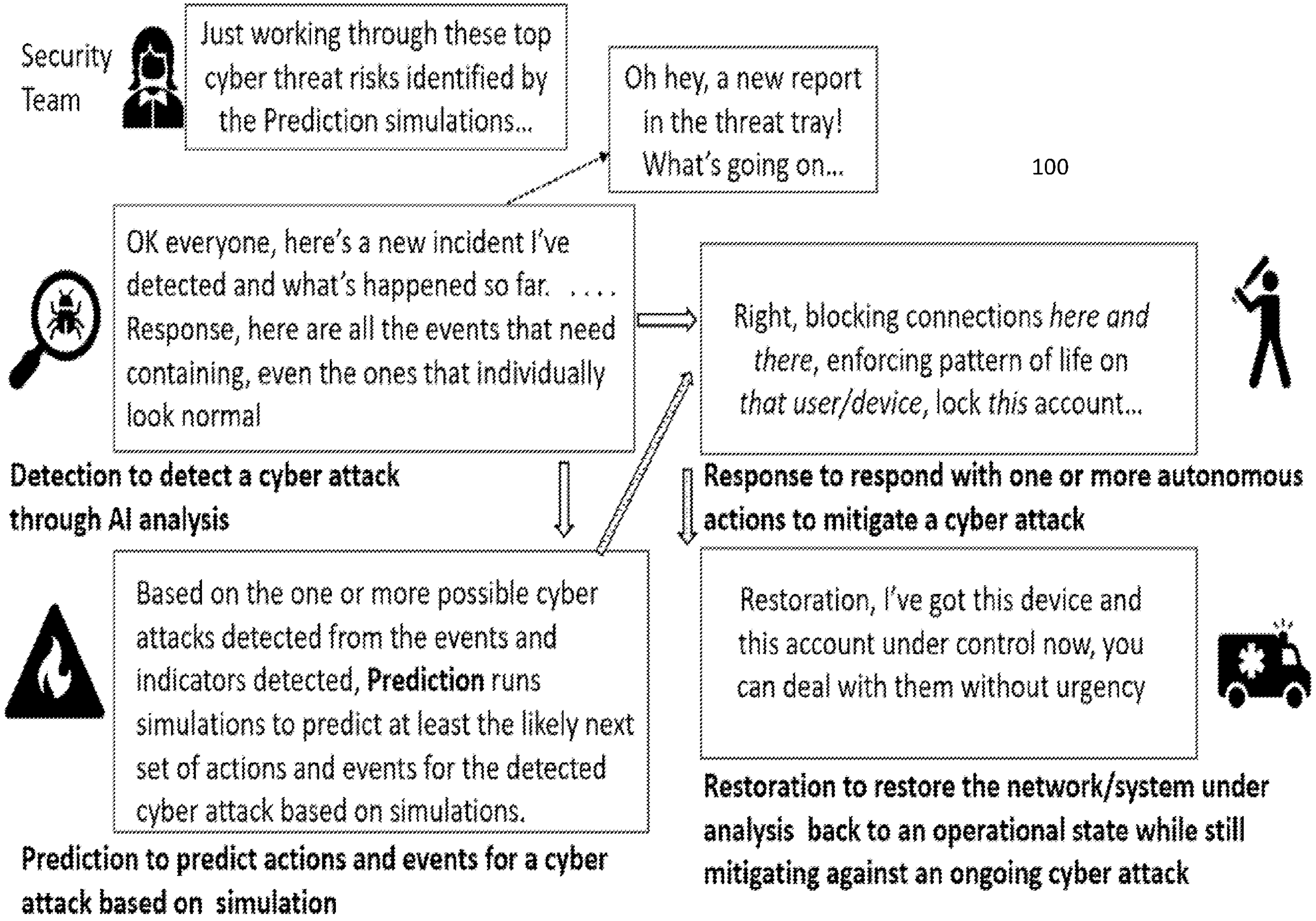


FIG. 8 Network



Just working through these top cyber threat risks identified by the Prediction simulations...

Oh hey, a new report in the threat tray! What's going on...

OK everyone, here's a new incident I've detected and what's happened so far. .... Response, here are all the events that need containing, even the ones that individually look normal

Detection to detect a cyber attack through AI analysis

Right, blocking connections *here and there*, enforcing pattern of life on *that user/device*, lock *this* account...

Response to respond with one or more autonomous actions to mitigate a cyber attack

Based on the one or more possible cyber attacks detected from the events and indicators detected, **Prediction** runs simulations to predict at least the likely next set of actions and events for the detected cyber attack based on simulations.

Prediction to predict actions and events for a cyber attack based on simulation

Restoration, I've got this device and this account under control now, you can deal with them without urgency

Restoration to restore the network/system under analysis back to an operational state while still mitigating against an ongoing cyber attack

Fig. 9

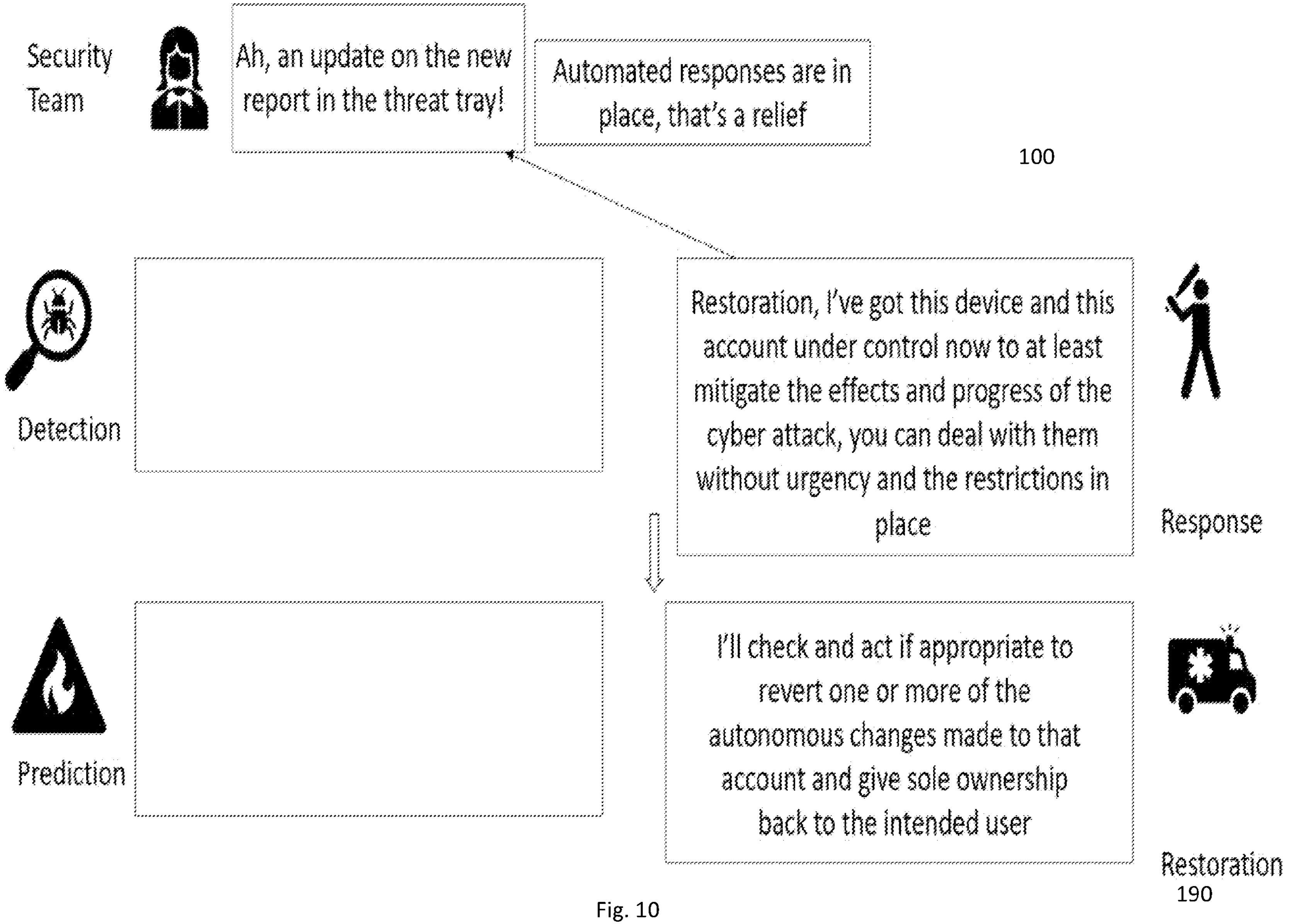
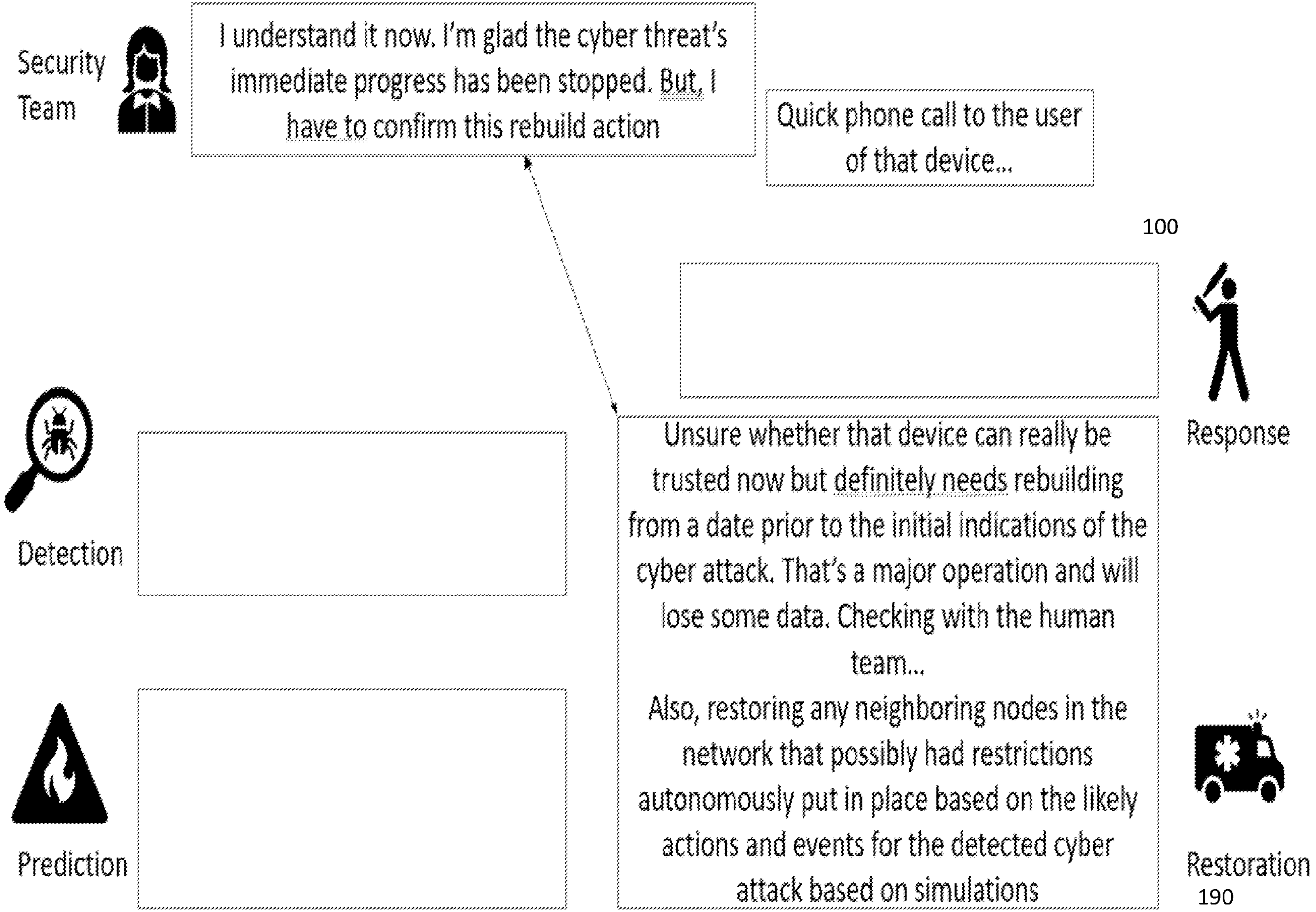

100
Security Team
Ah, an update on the new report in the threat tray!
Automated responses are in place, that's a relief
Detection
Restoration, I've got this device and this account under control now to at least mitigate the effects and progress of the cyber attack, you can deal with them without urgency and the restrictions in place
Response
Prediction
I'll check and act if appropriate to revert one or more of the autonomous changes made to that account and give sole ownership back to the intended user
Restoration
190

Fig. 10



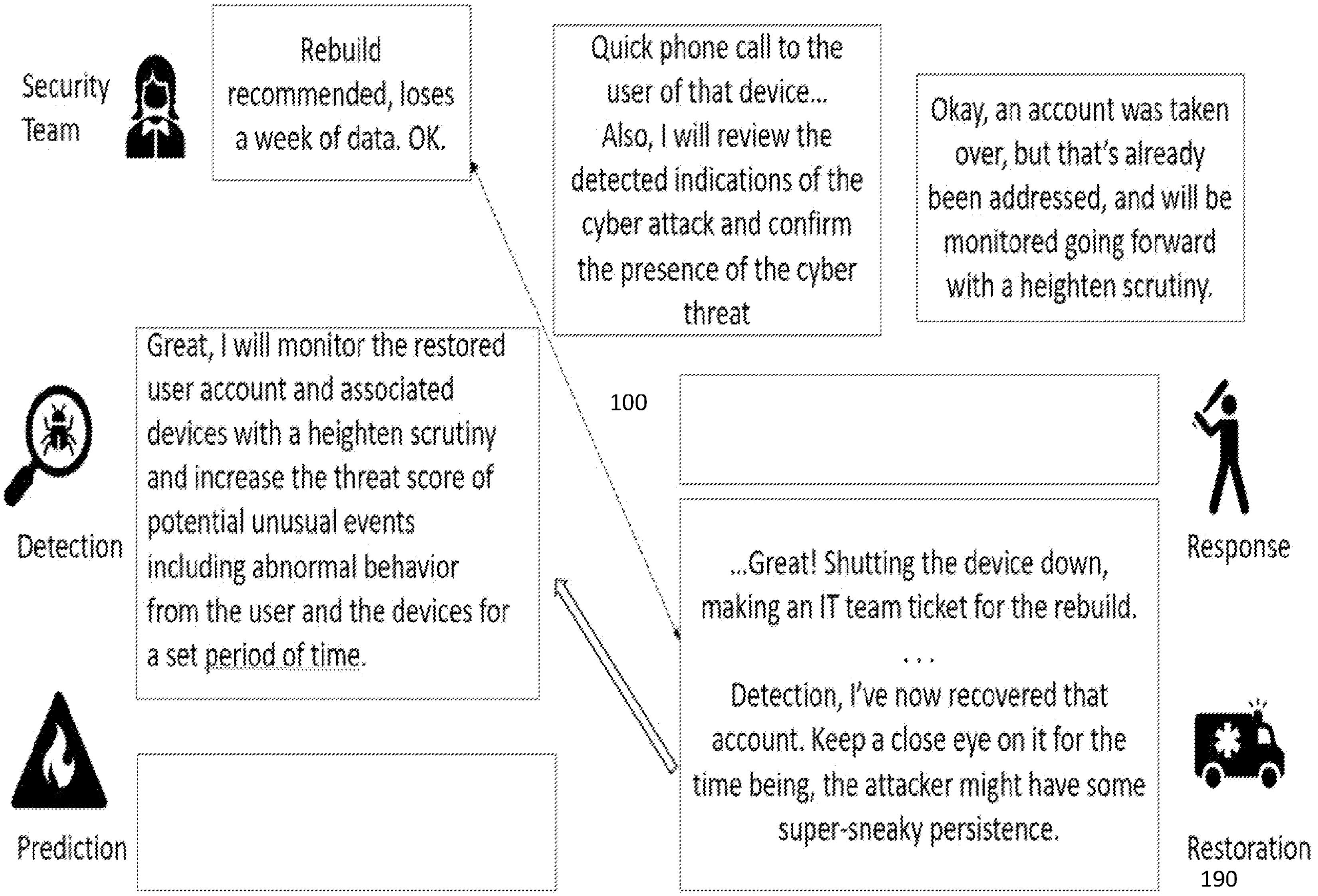
I understand it now. I'm glad the cyber threat's immediate progress has been stopped. But, I have to confirm this rebuild action

Quick phone call to the user of that device...

Unsure whether that device can really be trusted now but definitely needs rebuilding from a date prior to the initial indications of the cyber attack. That's a major operation and will lose some data. Checking with the human team...

Also, restoring any neighboring nodes in the network that possibly had restrictions autonomously put in place based on the likely actions and events for the detected cyber attack based on simulations

Fig. 11



Rebuild recommended, loses a week of data. OK.

Quick phone call to the user of that device...
Also, I will review the detected indications of the cyber attack and confirm the presence of the cyber threat

Okay, an account was taken over, but that's already been addressed, and will be monitored going forward with a heighten scrutiny.

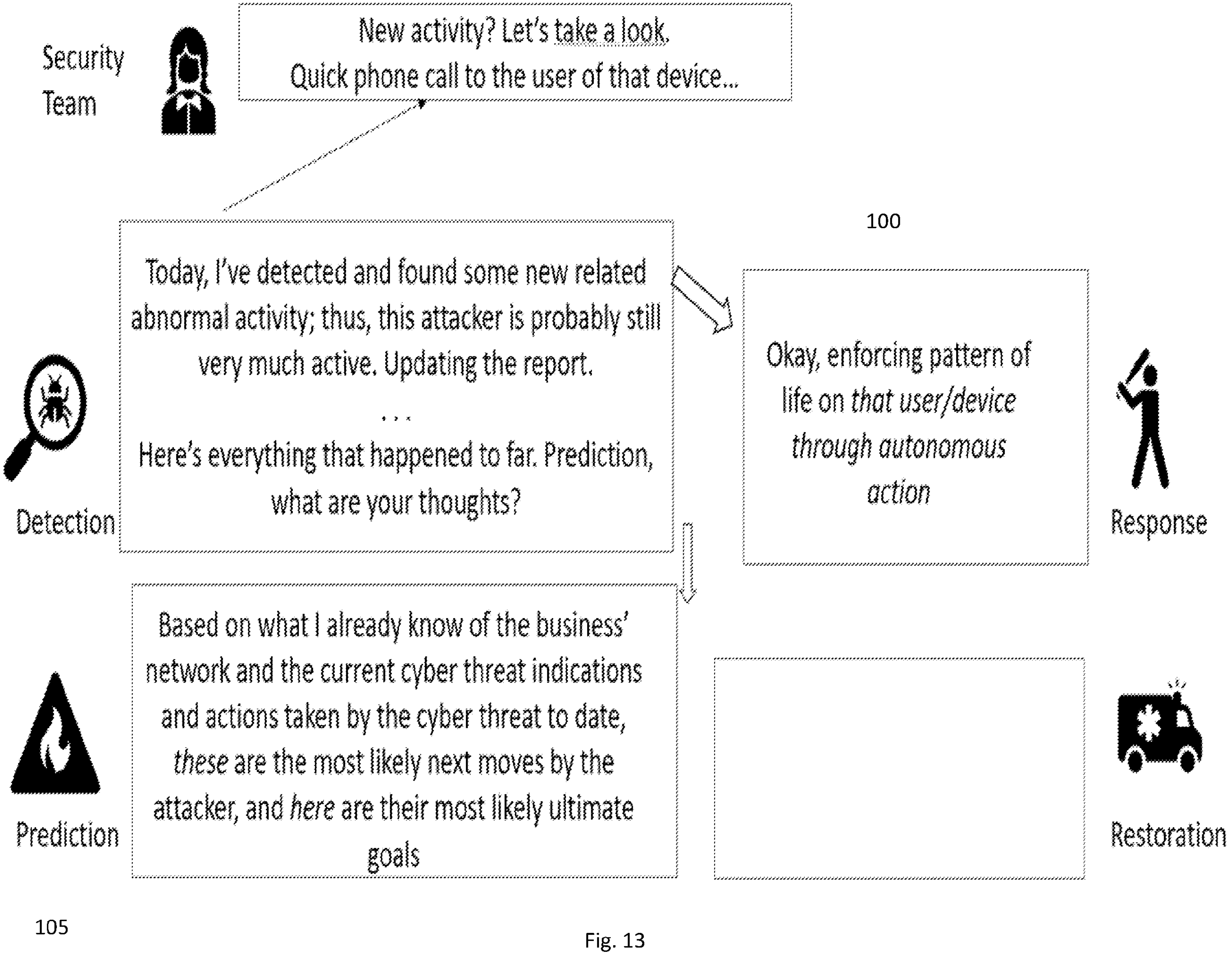
Great, I will monitor the restored user account and associated devices with a heighten scrutiny and increase the threat score of potential unusual events including abnormal behavior from the user and the devices for a set period of time.

...Great! Shutting the device down, making an IT team ticket for the rebuild.

. . .

Detection, I've now recovered that account. Keep a close eye on it for the time being, the attacker might have some super-sneaky persistence.

Fig. 12



New activity? Let's take a look.
Quick phone call to the user of that device...

Today, I've detected and found some new related abnormal activity; thus, this attacker is probably still very much active. Updating the report.

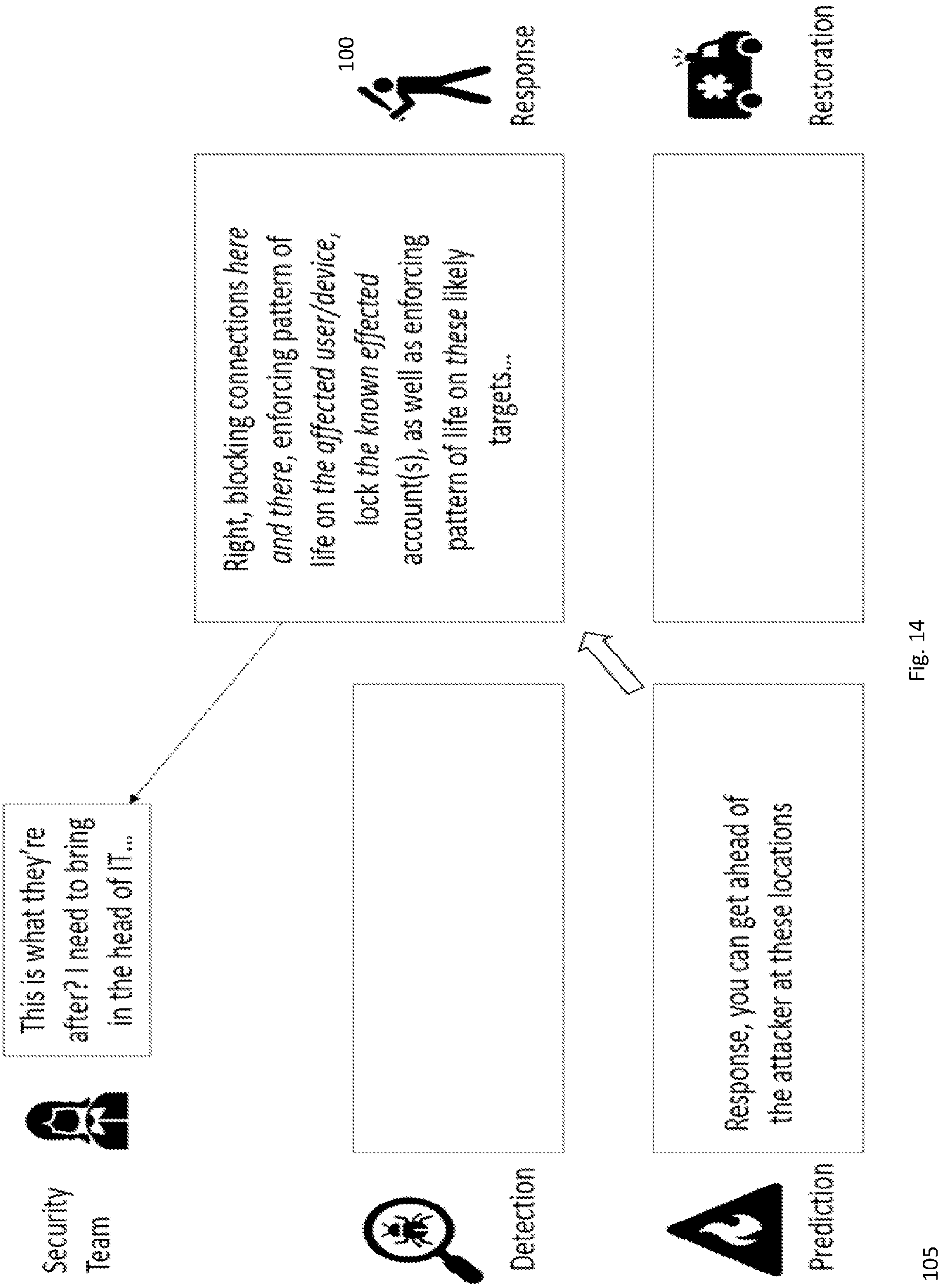
. . .

Here's everything that happened to far. Prediction, what are your thoughts?

Okay, enforcing pattern of life on *that user/device through autonomous action*

Based on what I already know of the business' network and the current cyber threat indications and actions taken by the cyber threat to date, *these* are the most likely next moves by the attacker, and *here* are their most likely ultimate goals

Fig. 13



This is what they're after? I need to bring in the head of IT...

Right, blocking connections *here and there*, enforcing pattern of life on *the affected user/device*, lock *the known effected* account(s), as well as enforcing pattern of life on *these* likely targets...

Response, you can get ahead of the attacker at these locations

Fig. 14

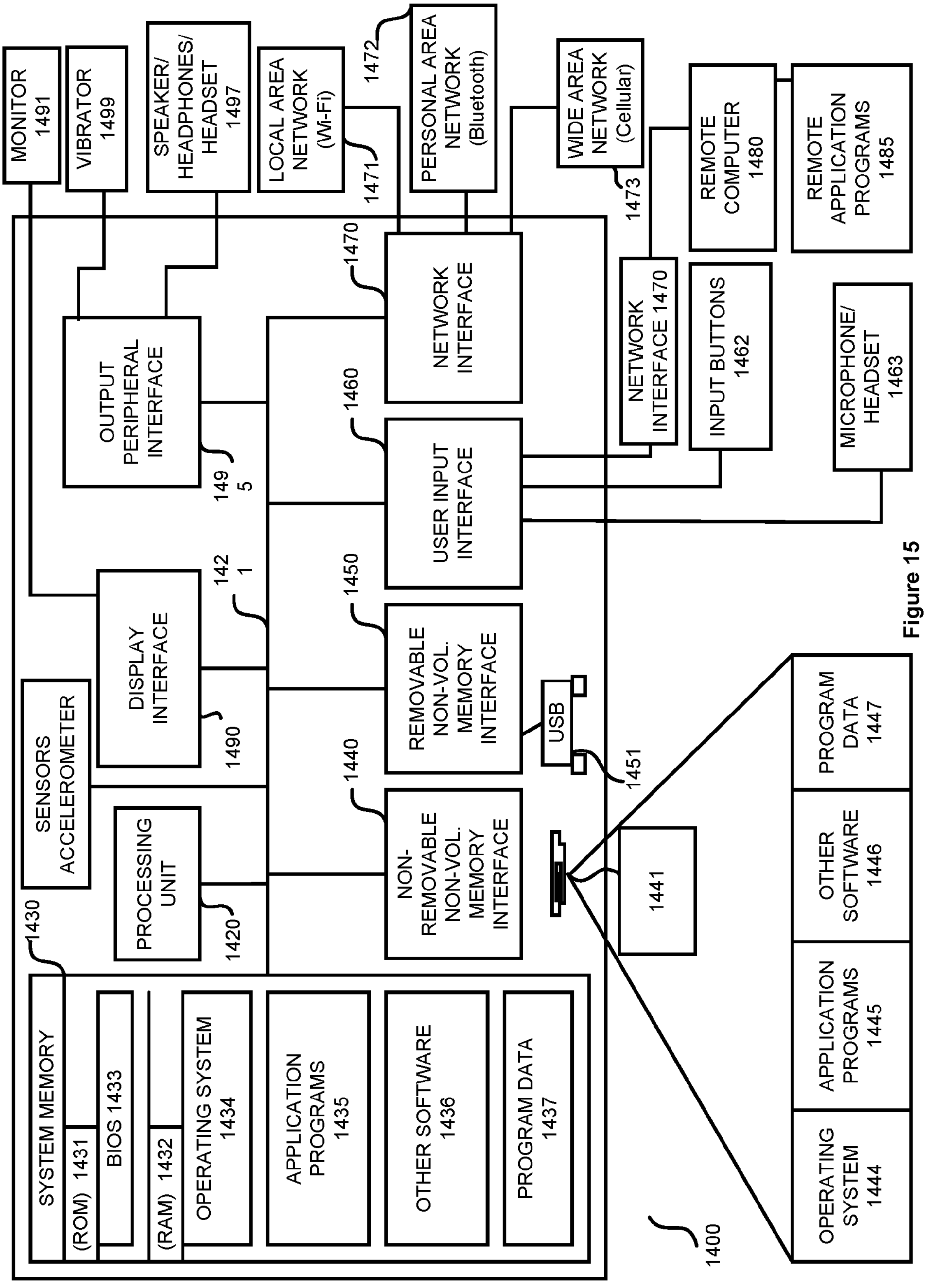

SYSTEM MEMORY
(ROM) 1431
BIOS 1433
(RAM) 1432
OPERATING SYSTEM 1434
APPLICATION PROGRAMS 1435
OTHER SOFTWARE 1436
PROGRAM DATA 1437
1430
PROCESSING UNIT
1420
SENSORS ACCELEROMETER
DISPLAY INTERFACE
1490
142 1
OUTPUT PERIPHERAL INTERFACE
149 5
NON-REMOVABLE NON-VOL. MEMORY INTERFACE
1440
REMOVABLE NON-VOL. MEMORY INTERFACE
1450
USER INPUT INTERFACE
1460
NETWORK INTERFACE
1470
USB
1451
1441
OPERATING SYSTEM 1444
APPLICATION PROGRAMS 1445
OTHER SOFTWARE 1446
PROGRAM DATA 1447
MONITOR 1491
VIBRATOR 1499
SPEAKER/ HEADPHONES/ HEADSET 1497
LOCAL AREA NETWORK (W-Fi)
1471
PERSONAL AREA NETWORK (Bluetooth)
1472
WIDE AREA NETWORK (Cellular)
1473
REMOTE COMPUTER 1480
REMOTE APPLICATION PROGRAMS 1485
NETWORK INTERFACE 1470
INPUT BUTTONS 1462
MICROPHONE/ HEADSET 1463
1400

Figure 15

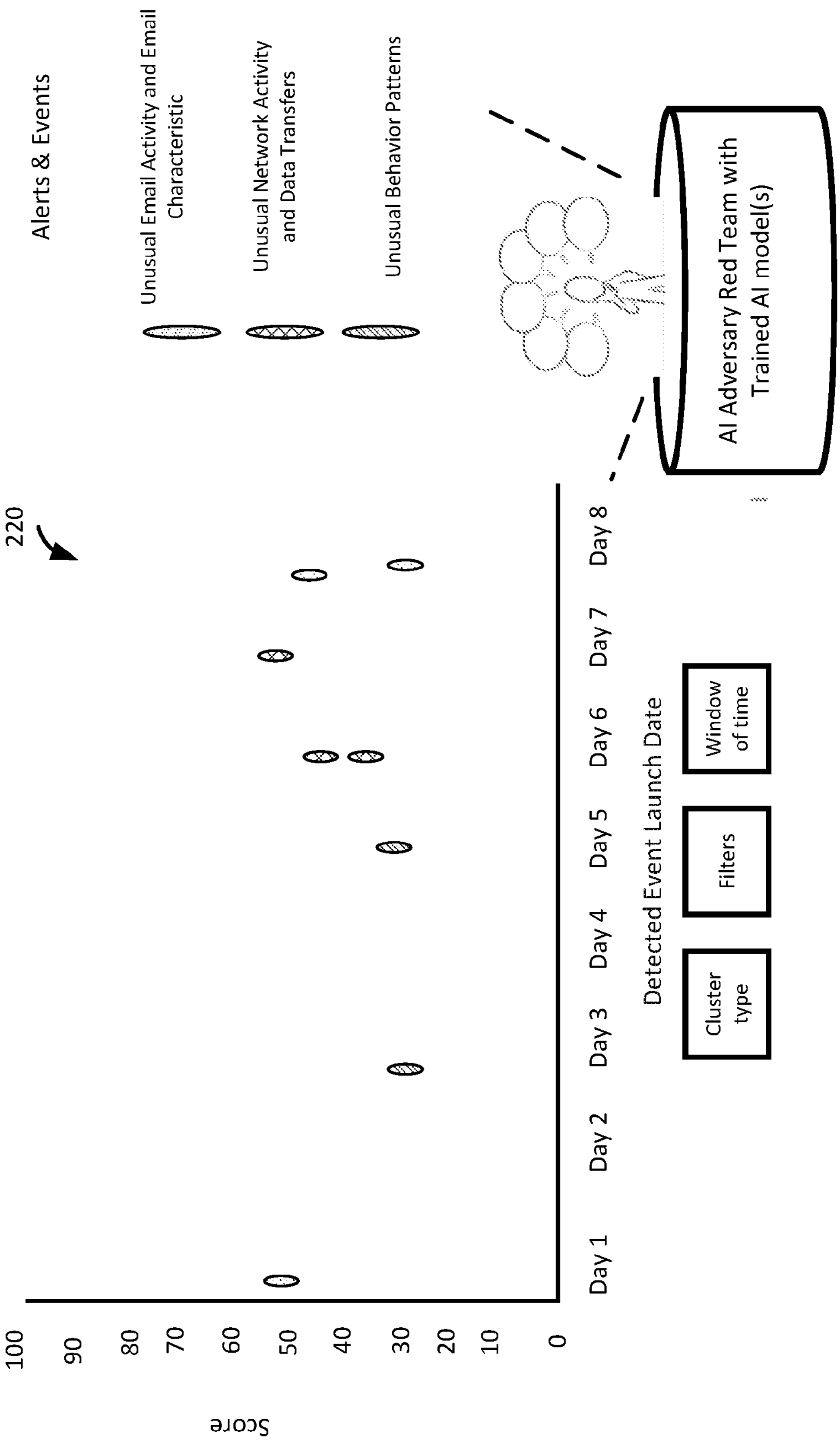

220
Score
100
90
80
70
60
50
40
30
20
10
0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Detected Event Launch Date
Cluster type
Filters
Window of time
Alerts & Events
Unusual Email Activity and Email Characteristic
Unusual Network Activity and Data Transfers
Unusual Behavior Patterns
AI Adversary Red Team with Trained AI model(s)

FIG. 16

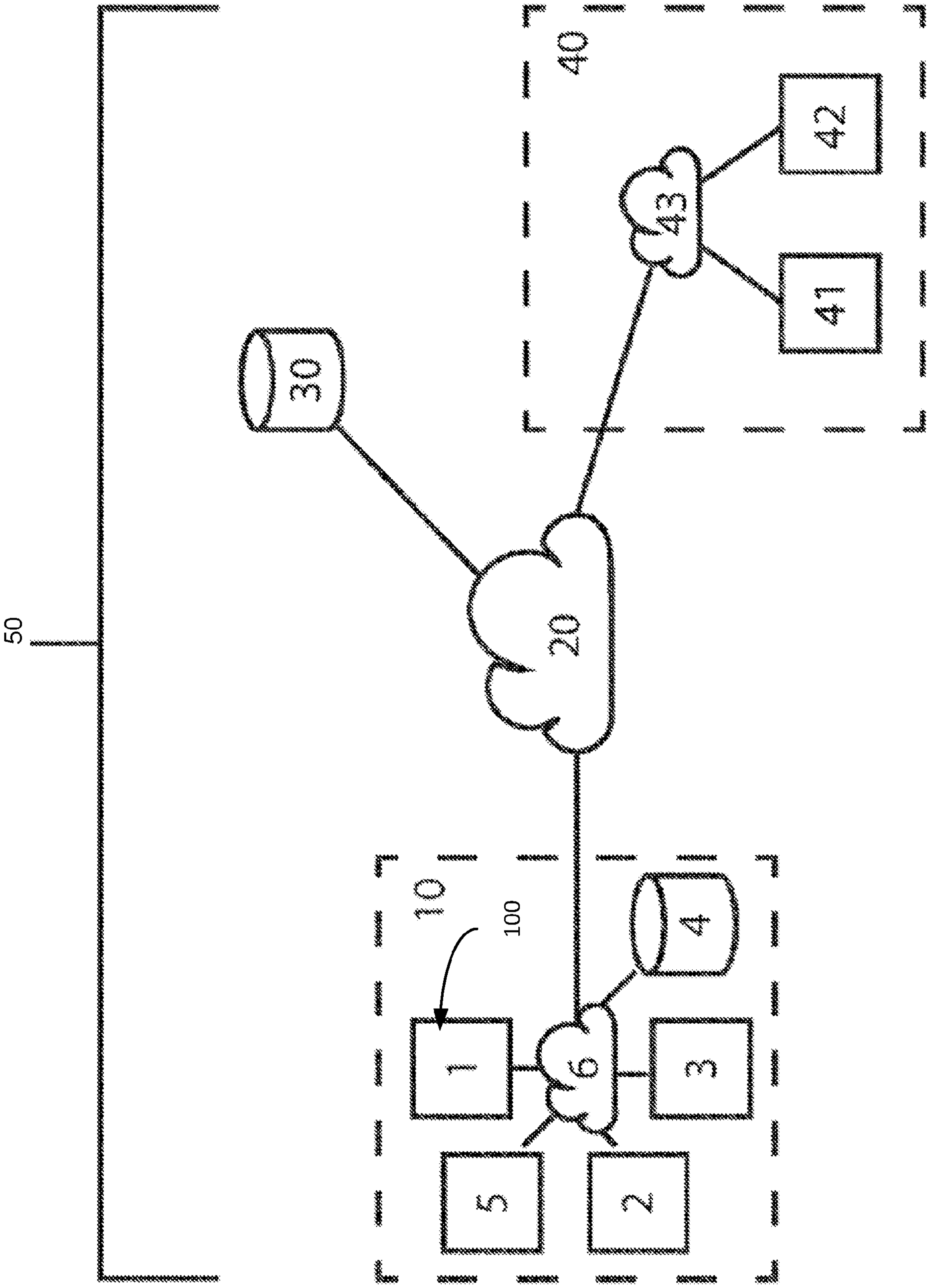


FIG. 17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 23/29785

A. CLASSIFICATION OF SUBJECT MATTER

IPC - INV. G06F 16/901, G06F 21/50, G06F 21/55, G06F 21/60 (2023.01)
ADD.

CPC - INV. G06F 16/9024, G06F 21/50, G06F 21/55, G06F 21/60

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2021/0273961 A1 (DARKTRACE LIMITED) 02 September 2021 (02.09.2021), entire document, especially Fig. 1, 2, 11; para [0019], [0022], [0032], [0046], [0083], [0084], [0090], [0112] - [0114], [0118], [0131], [0196], [0200], [0245], [0255], [0260], [0262], [0265], [0270]	1-14, 19, 20 ---------------- 15-18
Y	US 11,010,472 B1 (ARCHITECTURE TECHNOLOGY CORPORATION) 18 May 2021 (18.05.2021), entire document, especially Fig. 1; col 4, ln 35-47; col 11, ln 16-29	15-18
A	US 2019/0362219 A1 (SCHEFFLER) 28 November 2019 (28.11.2019), entire document	1-20
A	US 2020/0358799 A1 (DARKTRACE LIMITED) 12 November 2020 (12.11.2020), entire document	1-20
A	US 2017/0230392 A1 (DARKTRACE LIMITED) 10 August 2017 (10.08.2017), entire document	1-20

□ Further documents are listed in the continuation of Box C. □ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"D"	document cited by the applicant in the international application	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent but published on or after the international filing date		
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search 30 September 2023	Date of mailing of the international search report NOV 0 8 2023
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Kari Rodriquez Telephone No. PCT Helpdesk: 571-272-4300