US 20210273961A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2021/0273961 A1**

Humphrey et al. (43) **Pub. Date: Sep. 2, 2021**

(54) **APPARATUS AND METHOD FOR A CYBER-THREAT DEFENSE SYSTEM**

(71) Applicant: **Darktrace Limited**, Cambridge (GB)

(72) Inventors: **Dickon Murray Humphrey**, Cambridge (GB); **Timothy Owen Bazalgette**, Knebworth (GB)

(73) Assignee: **Darktrace Limited**

(21) Appl. No.: **17/187,385**

(22) Filed: **Feb. 26, 2021**

Related U.S. Application Data

(60) Provisional application No. 62/983,307, filed on Feb. 28, 2020, provisional application No. 63/078,092, filed on Sep. 14, 2020.

Publication Classification

(51) **Int. Cl.**

H04L 29/06 (2006.01)

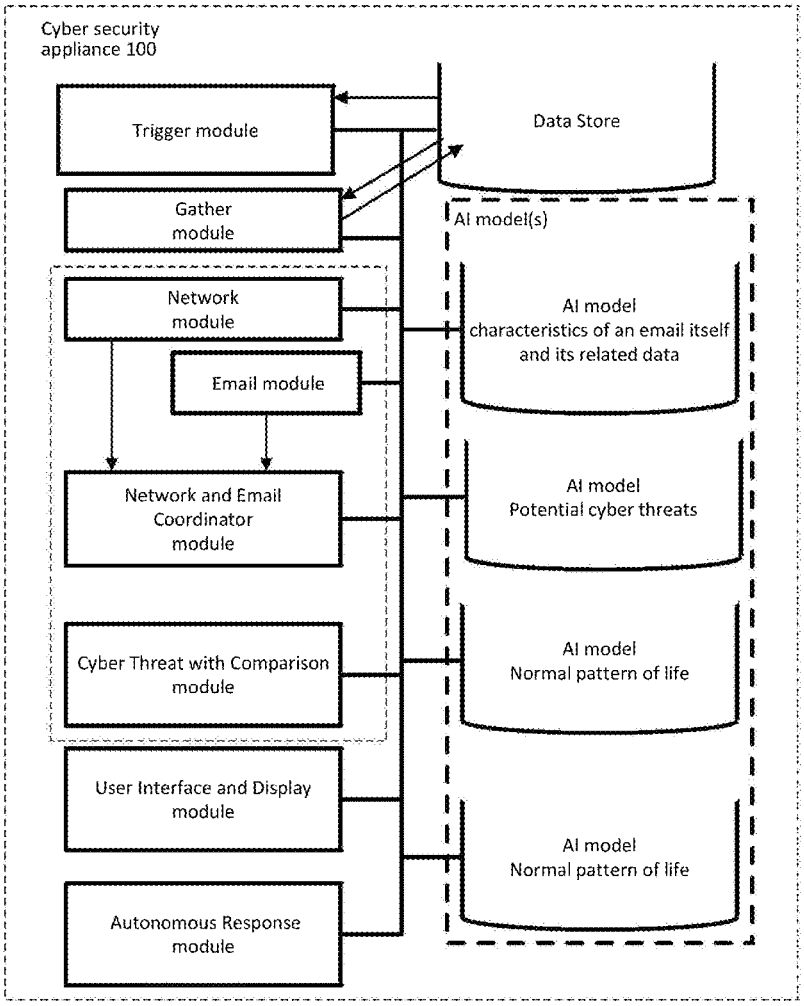
G06N 20/00 (2006.01)

(52) **U.S. Cl.**

CPC ***H04L 63/1425*** (2013.01); ***G06N 20/00*** (2019.01)

(57) **ABSTRACT**

An apparatus comprising: one or more machine learning modules that are trained on a normal behavior of entities associated with a network and interactions between the entities; an interface configured to receive a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation; where the interface is configured to work with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, in order to determine whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat, and thus, trigger an autonomous response from an autonomous response module to mitigate the cyber-threat.



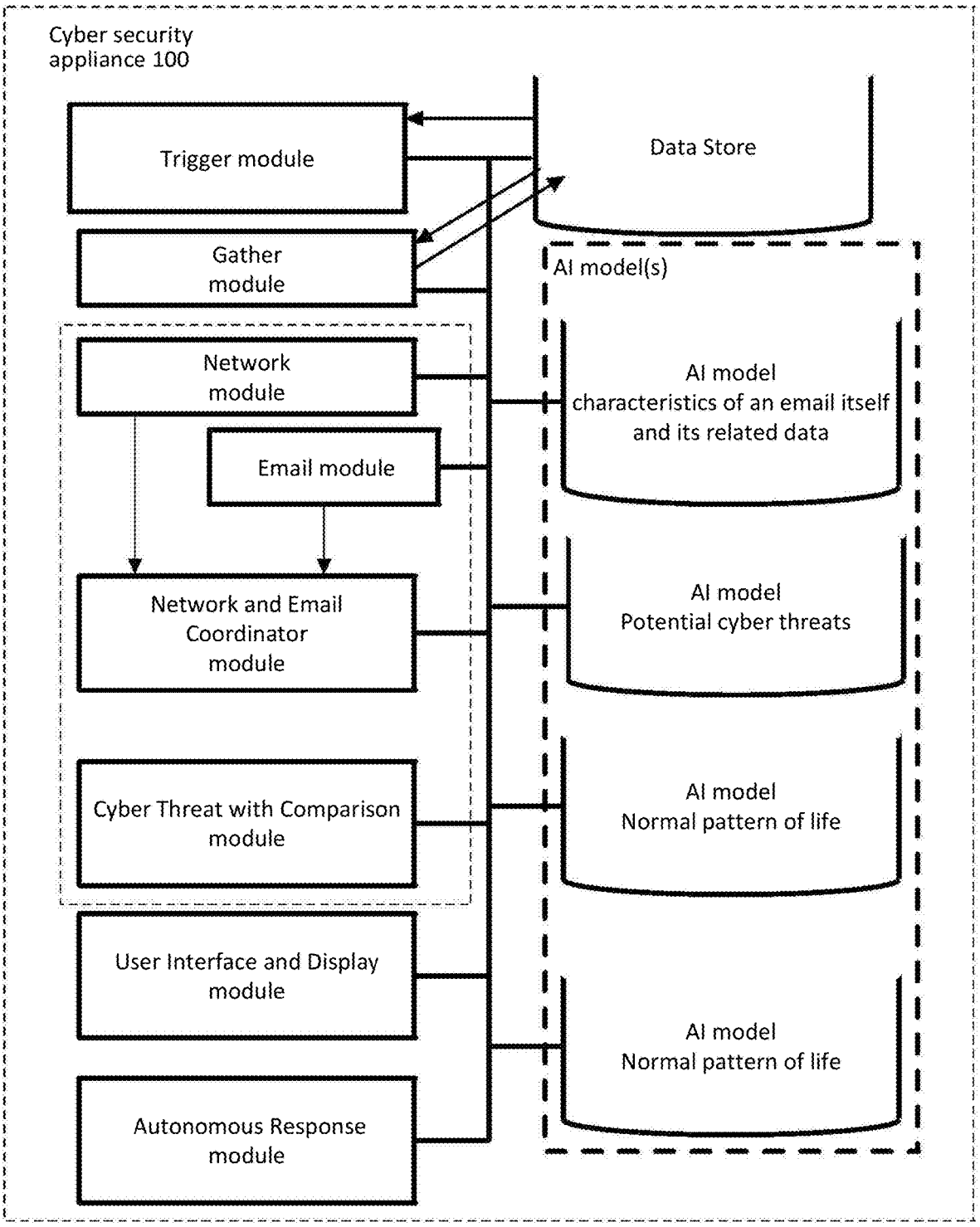

Cyber security appliance 100
Trigger module
Gather module
Network module
Email module
Network and Email Coordinator module
Cyber Threat with Comparison module
User Interface and Display module
Autonomous Response module
Data Store
AI model(s)
AI model characteristics of an email itself and its related data
AI model Potential cyber threats
AI model Normal pattern of life
AI model Normal pattern of life

Figure 1

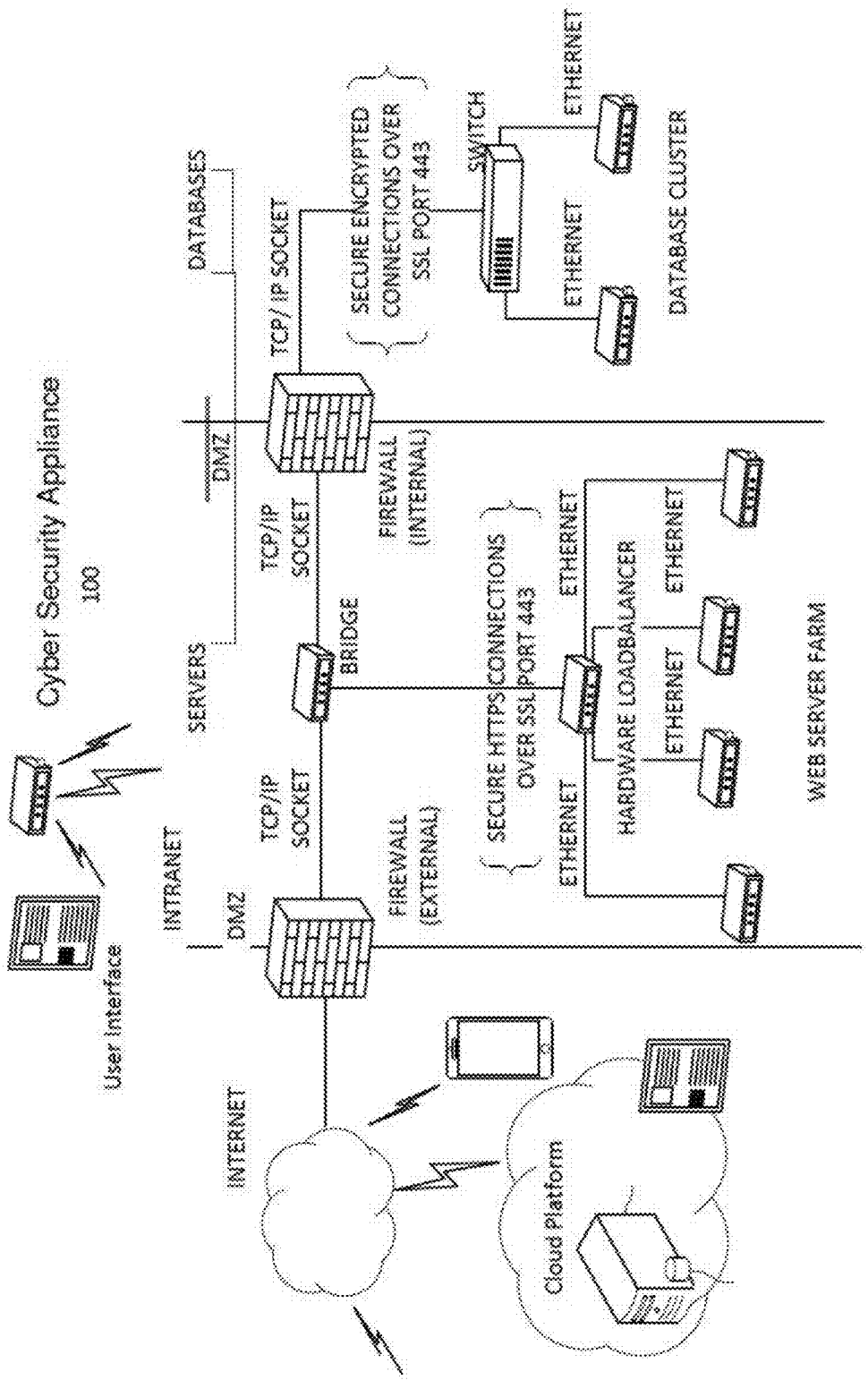

Cyber Security Appliance
100
User Interface
INTRANET
DMZ
TCP/IP SOCKET
FIREWALL (EXTERNAL)
SERVERS
BRIDGE
TCP/IP SOCKET
DMZ
FIREWALL (INTERNAL)
DATABASES
TCP/ IP SOCKET
SECURE ENCRYPTED CONNECTIONS OVER SSL PORT 443
SWITCH
ETHERNET
ETHERNET
DATABASE CLUSTER
SECURE HTTPS CONNECTIONS OVER SSL PORT 443
ETHERNET
ETHERNET
HARDWARE LOADBALANCER
ETHERNET
ETHERNET
WEB SERVER FARM
INTERNET
Cloud Platform

Figure 2

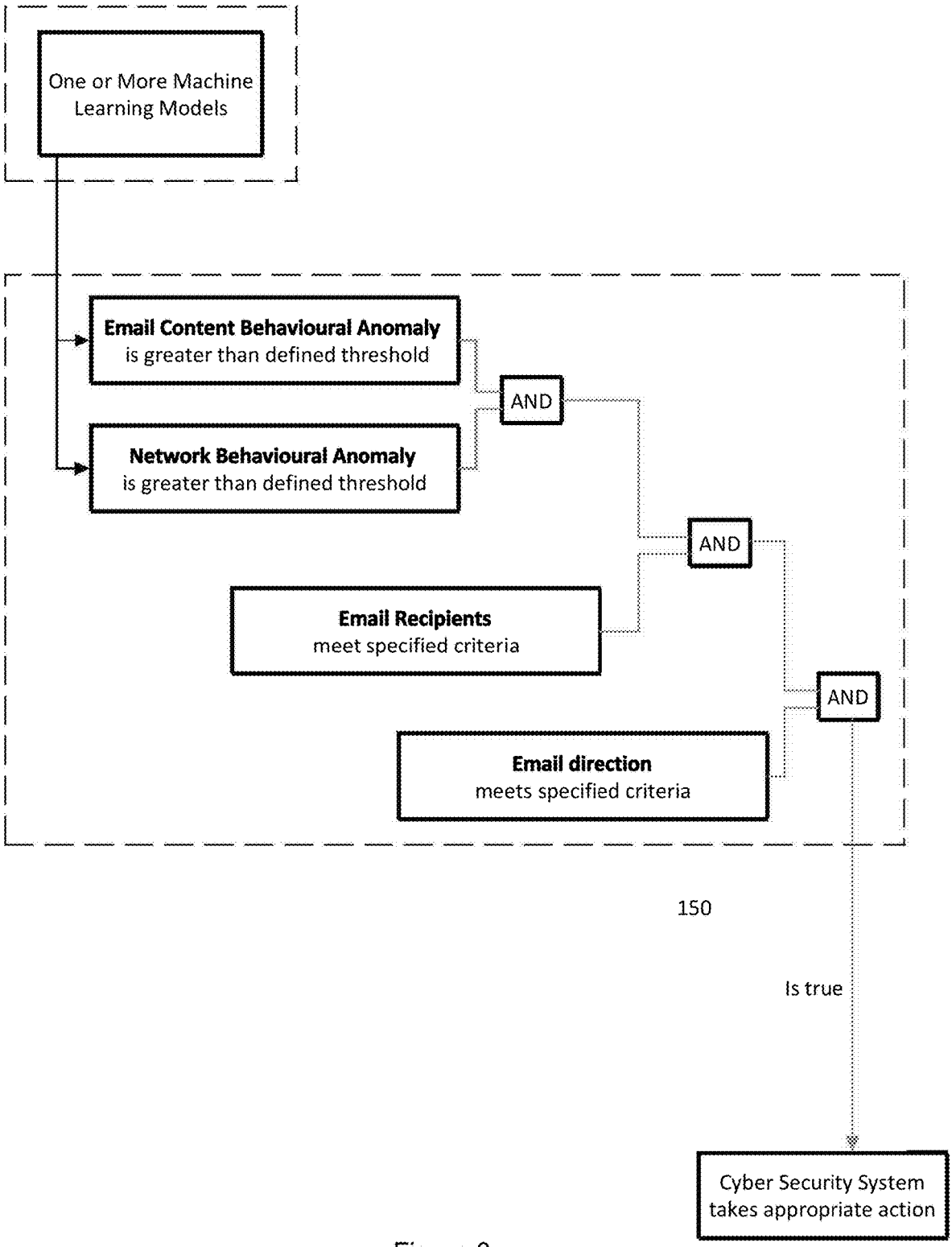

One or More Machine Learning Models
Email Content Behavioural Anomaly is greater than defined threshold
Network Behavioural Anomaly is greater than defined threshold
AND
AND
AND
Email Recipients meet specified criteria
Email direction meets specified criteria
150
Is true
Cyber Security System takes appropriate action

Figure 3

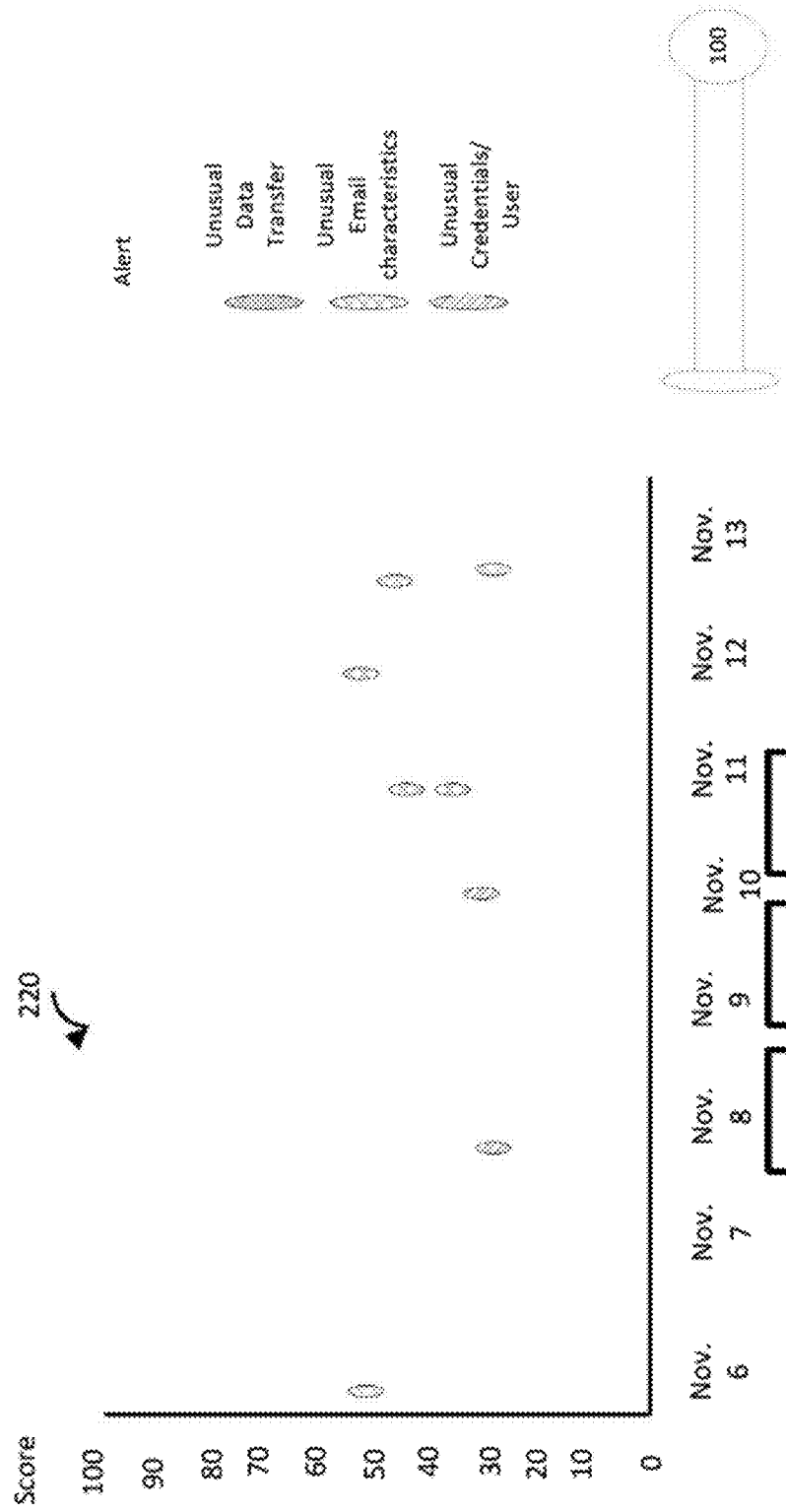

Alert
Unusual Data Transfer
Unusual Email characteristics
Unusual Credentials/ User
100
220
Score
100
90
80
70
60
50
40
30
20
10
0
Nov. 6
Nov. 7
Nov. 8
Nov. 9
Nov. 10
Nov. 11
Nov. 12
Nov. 13
Cluster type
Filters
Window of time

Figure 4

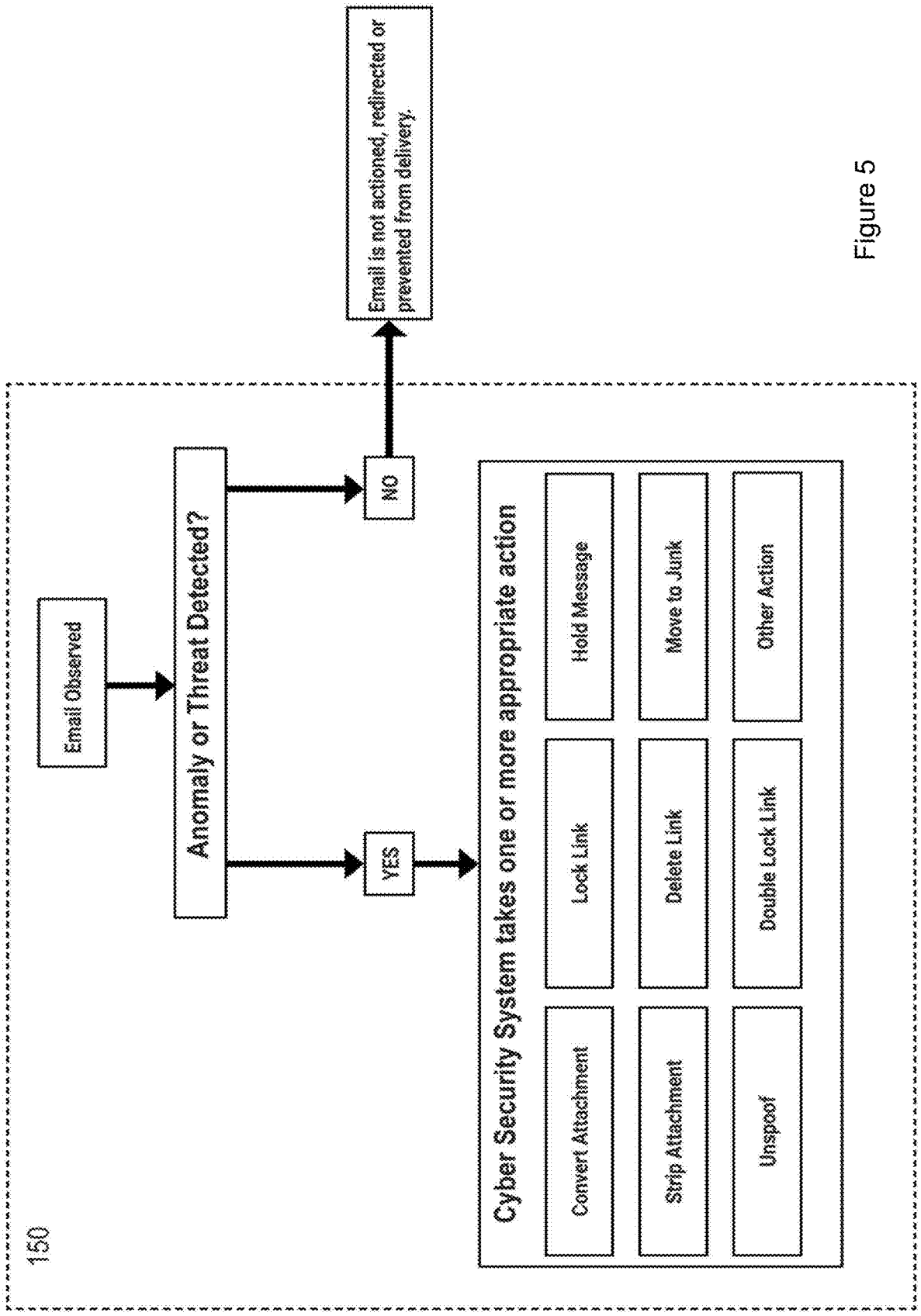

150
Email Observed
Anomaly or Threat Detected?
NO
Email is not actioned, redirected or prevented from delivery.
YES
Cyber Security System takes one or more appropriate action
Convert Attachment
Lock Link
Hold Message
Strip Attachment
Delete Link
Move to Junk
Unspoof
Double Lock Link
Other Action

Figure 5

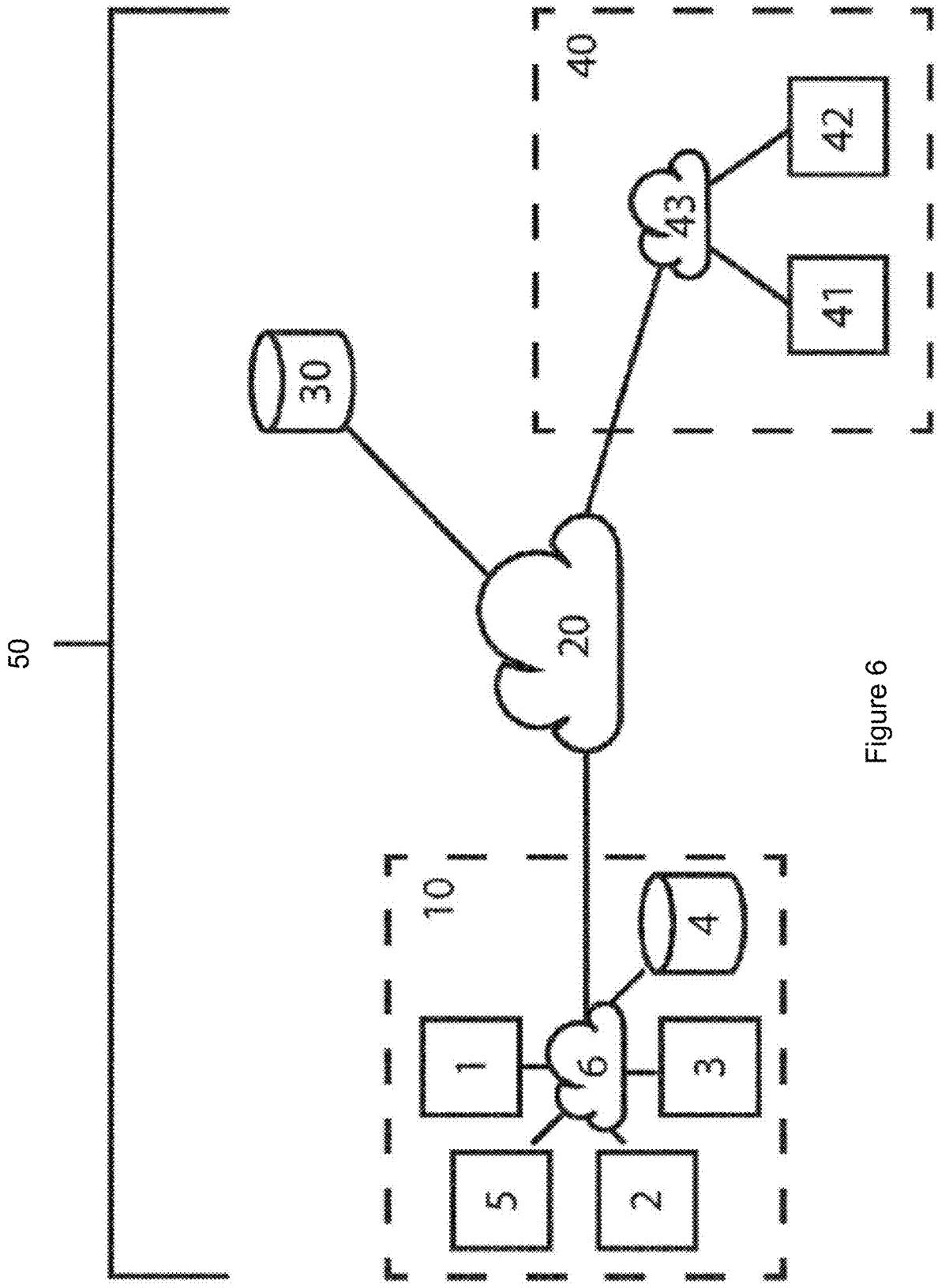


Figure 6

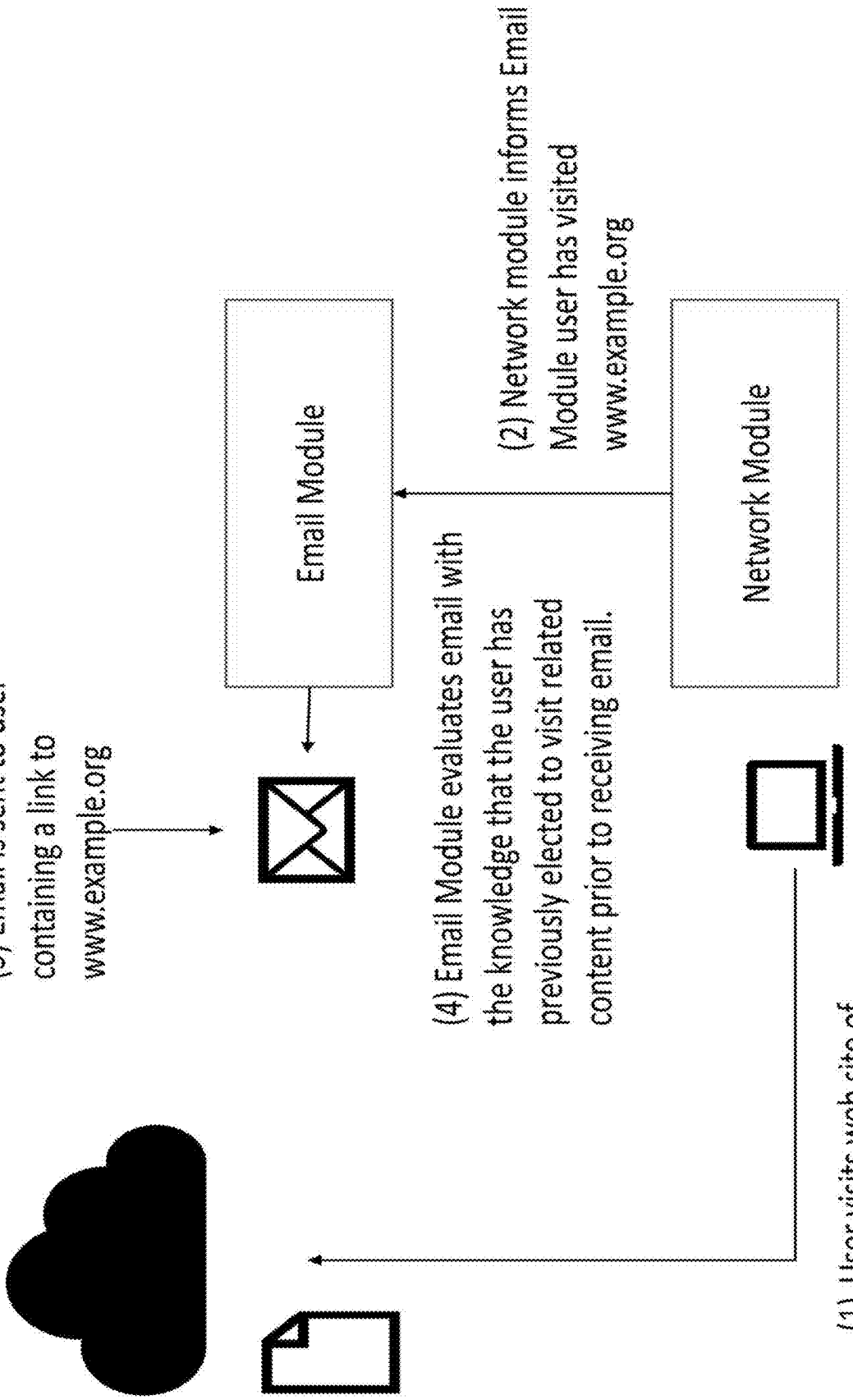

(1) User visits web site of www.example.org
(2) Network module informs Email Module user has visited www.example.org
(3) Email is sent to user containing a link to www.example.org
(4) Email Module evaluates email with the knowledge that the user has previously elected to visit related content prior to receiving email.
Email Module
Network Module

Figure 7

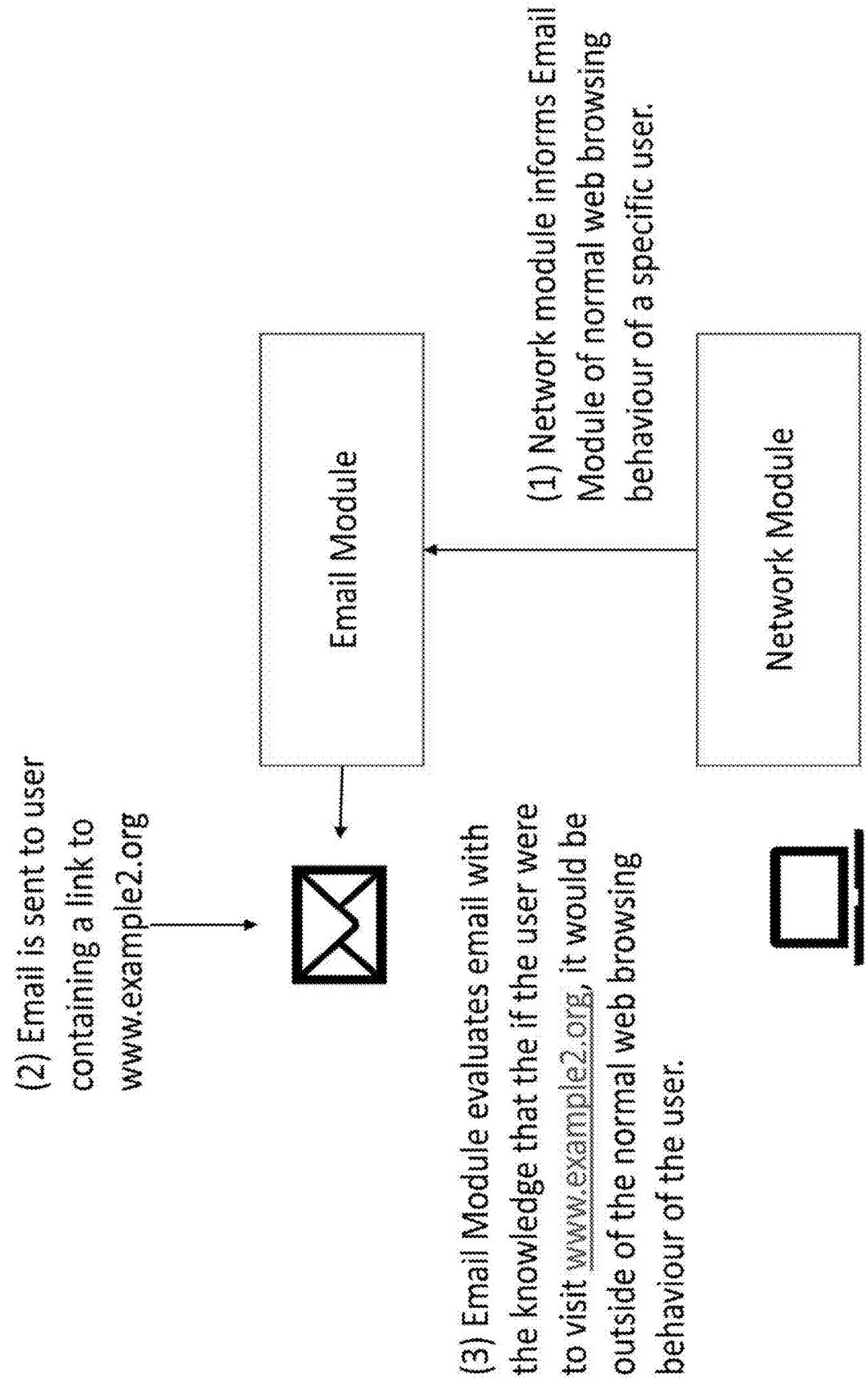

(2) Email is sent to user containing a link to www.example2.org
Email Module
(1) Network module informs Email Module of normal web browsing behaviour of a specific user.
Network Module
(3) Email Module evaluates email with the knowledge that the if the user were to visit www.example2.org, it would be outside of the normal web browsing behaviour of the user.

Figure 8

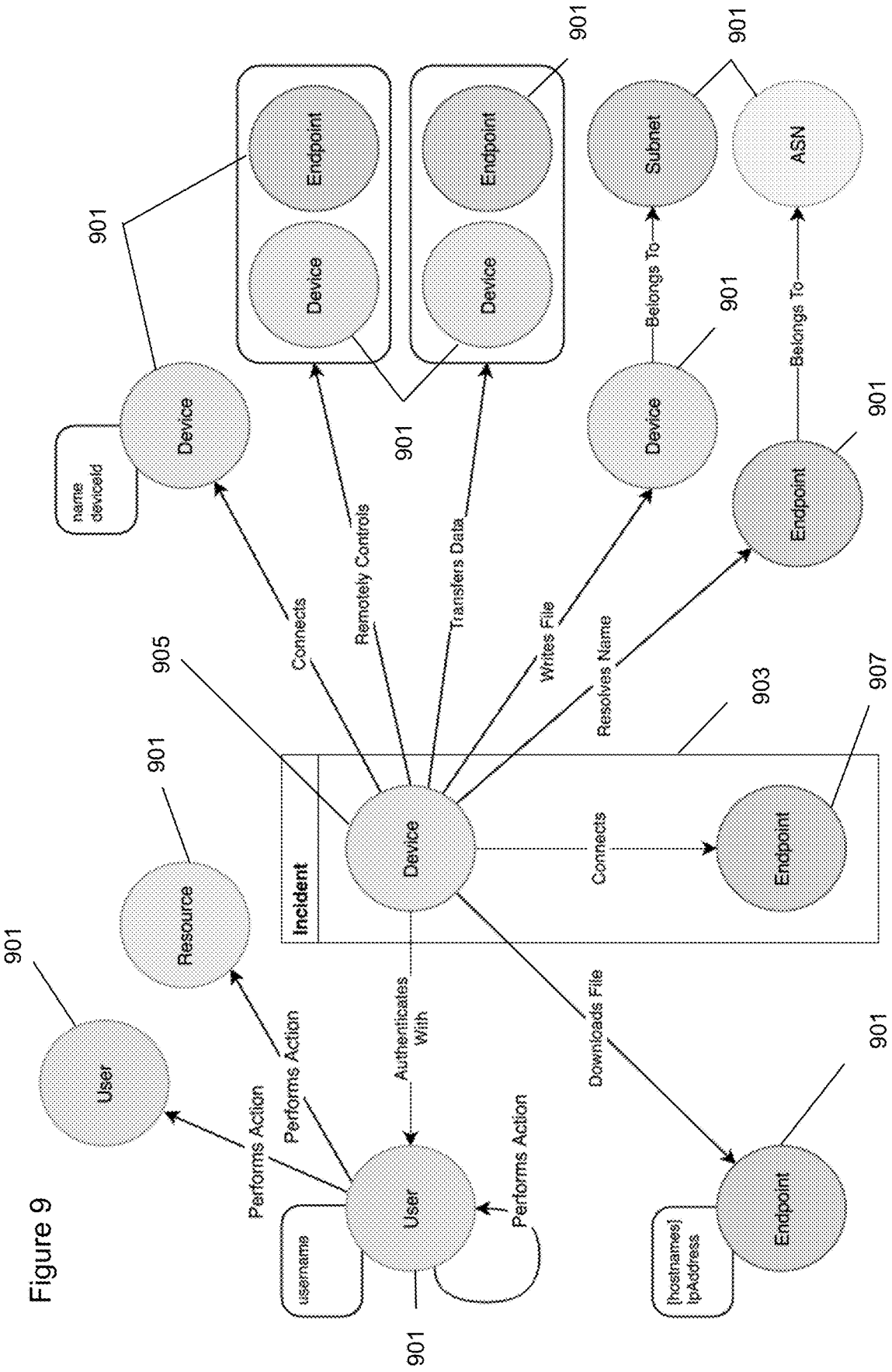

901
User
Resource
Device
Endpoint
Subnet
ASN
Incident
Performs Action
Authenticates With
Connects
Remotely Controls
Transfers Data
Writes File
Resolves Name
Downloads File
Belongs To
905
903
907
username
name
deviceId
[hostnames]
ipAddress

Figure 9

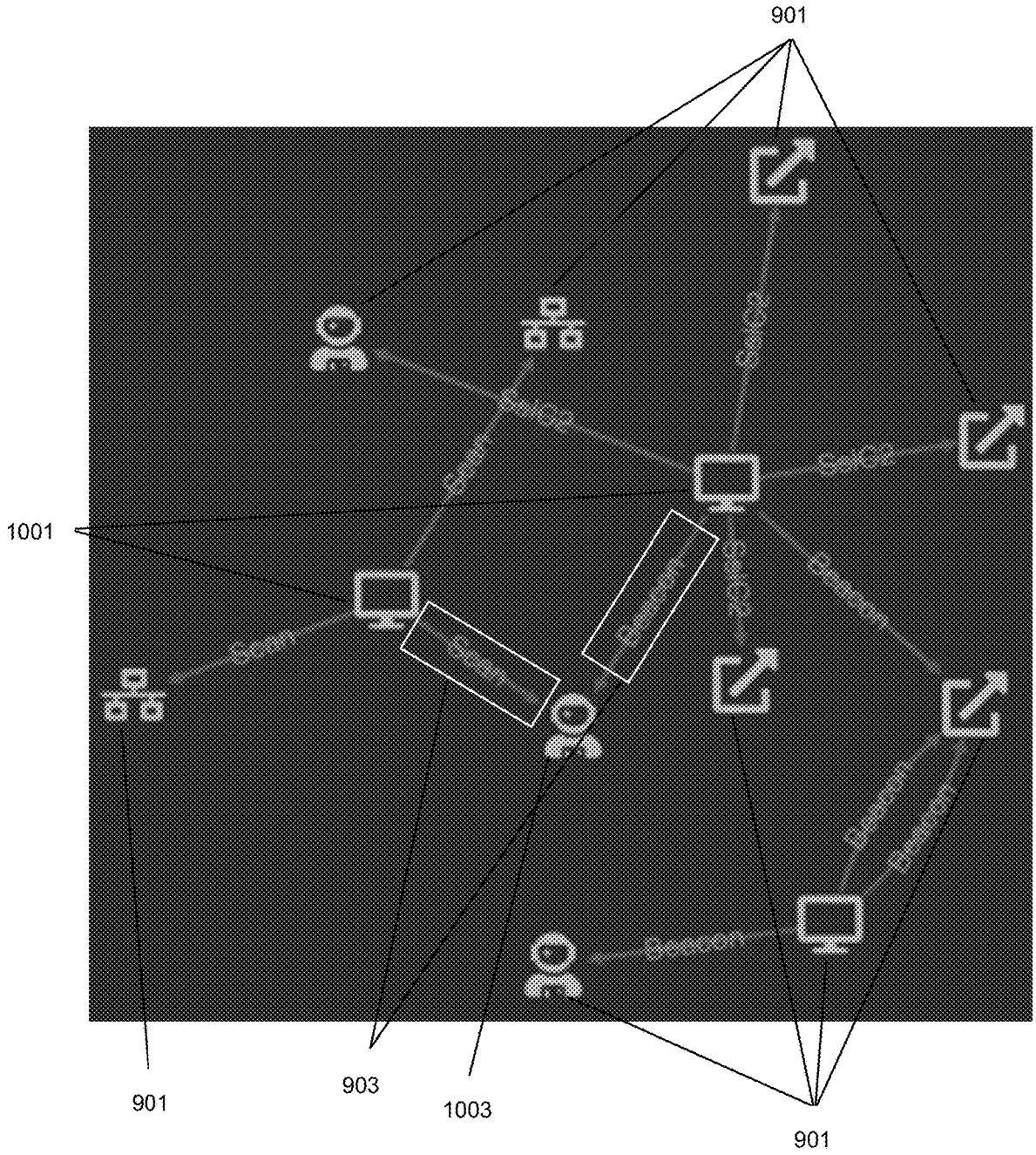


Figure 10

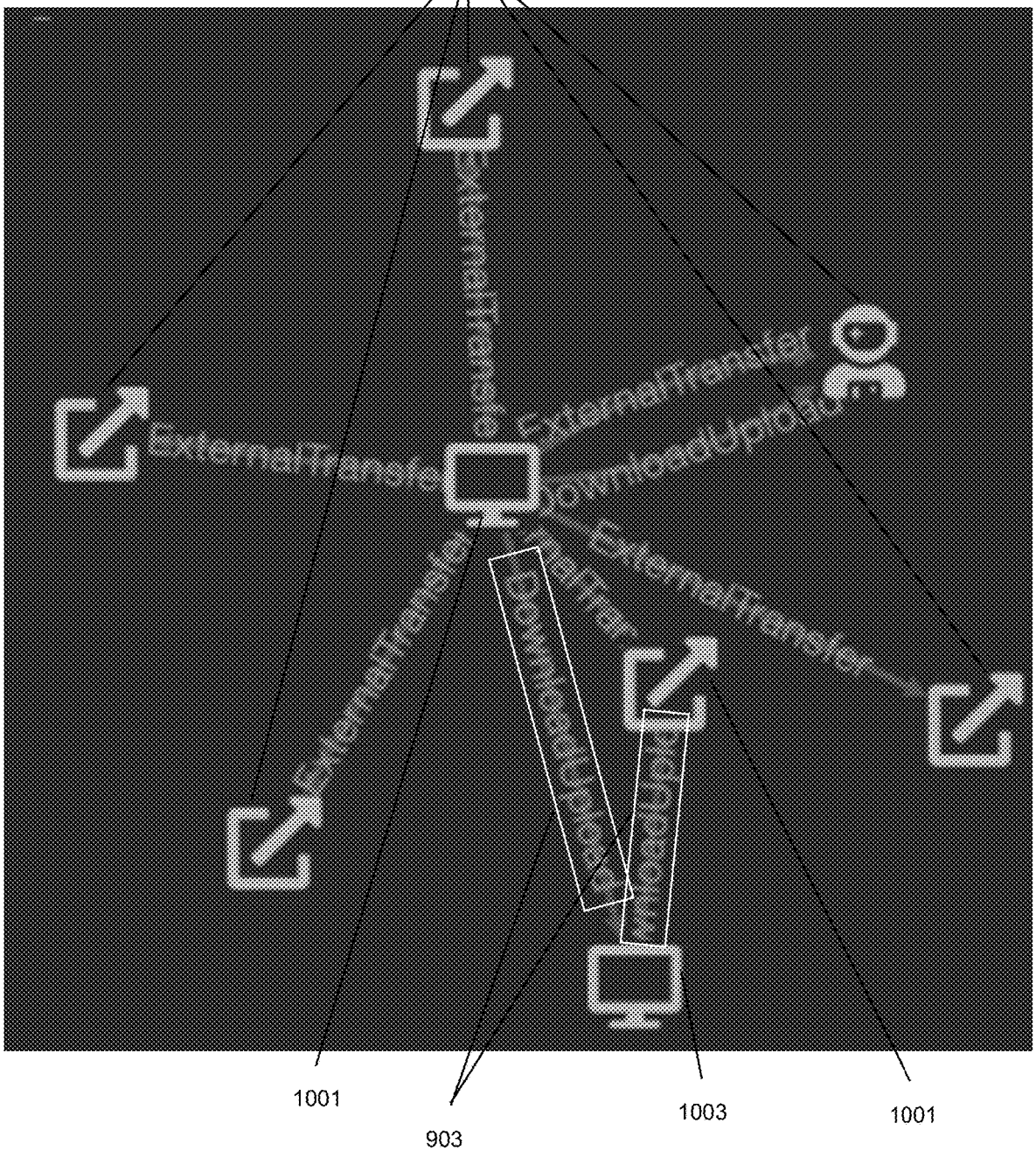


Figure 11

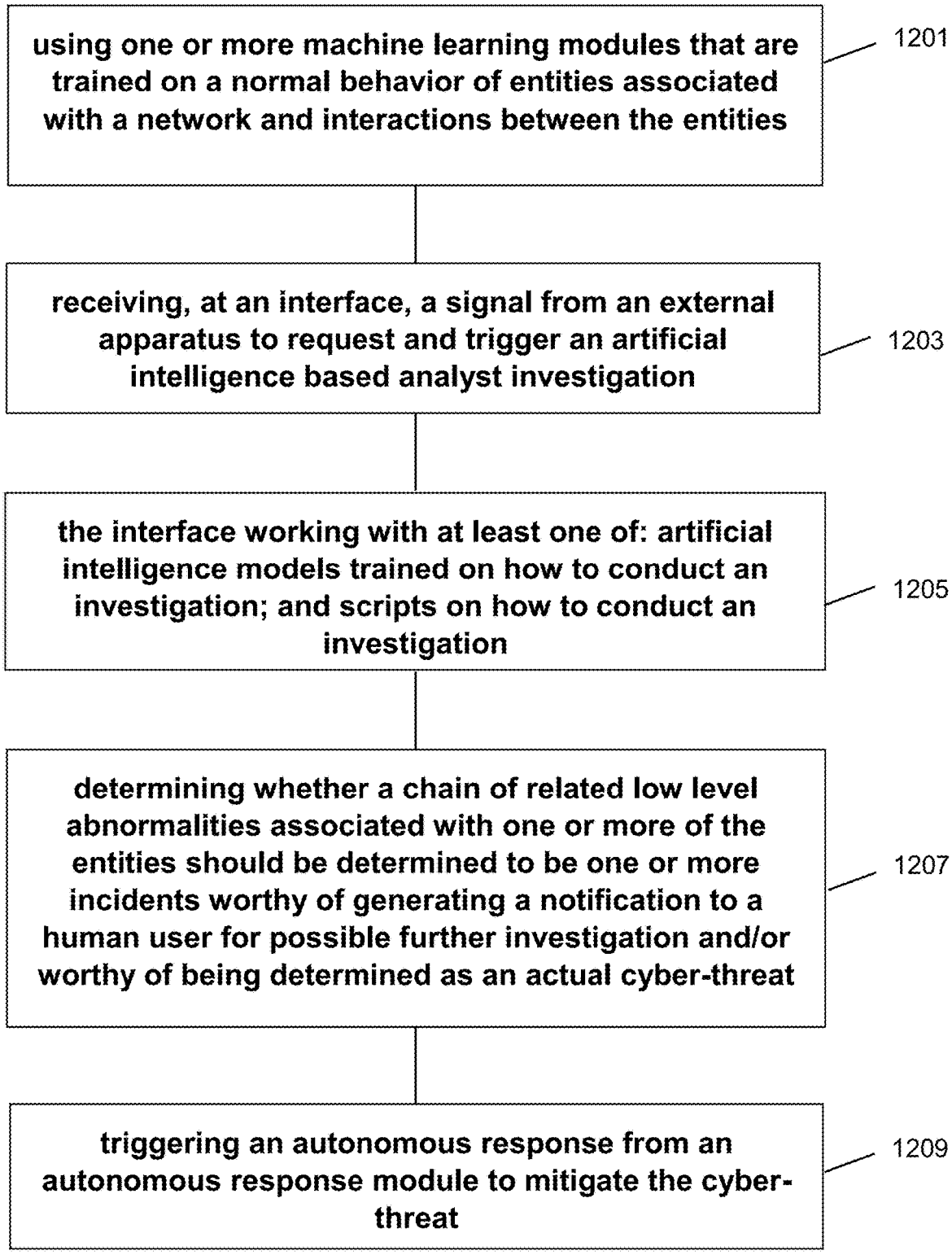

using one or more machine learning modules that are trained on a normal behavior of entities associated with a network and interactions between the entities
1201
receiving, at an interface, a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation
1203
the interface working with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation
1205
determining whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat
1207
triggering an autonomous response from an autonomous response module to mitigate the cyber-threat
1209

Figure 12

APPARATUS AND METHOD FOR A CYBER-THREAT DEFENSE SYSTEM

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] This application claims priority to U.S. Provisional Application No. 62/983307 filed 28 Feb. 2020, entitled AN ARTIFICIAL INTELLIGENCE BASED CYBER SECURITY SYSTEM, and US Provisional Application No. 63/078092 filed 14 Sep. 2020, entitled AN INTELLIGENT CYBER SECURITY SYSTEM, the disclosure of each of which is hereby expressly incorporated by reference herein in its entirety.

NOTICE OF COPYRIGHT

FIELD

[0003] Embodiments of the design provided herein generally relate to a cyber threat defense system. In an embodiment, Artificial Intelligence is applied to analyzing Cyber Security threats.

BACKGROUND

[0004] Existing methods such as vulnerability scanning performed by humans are less targeted and may lead to security resource allocation in the wrong places. Also, some vulnerability scanners actually test and compromise the actual network devices themselves, which may adversely affect the network during this testing and scanning.

[0005] Cyber threat protection systems generally ingest network data to detect cyber threats but not to assess how a cyber threat might spread through a network. A human Red team of cyber security professionals typically is hired to test a network's vulnerability to cyber-attacks.

[0006] In the cyber security environment, firewalls, endpoint security methods and other tools such as SIEMs and sandboxes are deployed to enforce specific policies, and provide protection against certain threats. These tools currently form an important part of an organization's cyber defense strategy, but they are insufficient in the new age of cyber threat. Legacy tools are failing to deal with new cyber threats because the traditional approach relies on being able to pre-define the cyber threat in advance, by writing rules or producing signatures. In today's environment, this approach to defend against cyber threats is fundamentally flawed:

[0007] Threats are constantly evolving—novel attacks do not match historical-attack "signatures", and even subtle changes to previously understood attacks can result in them going undetected by legacy defenses;

[0008] Rules and policies defined by organizations are continually insufficient—security teams simply can't imagine every possible thing that may go wrong in future; and

[0009] Employee 'insider' threat is a growing trend—it is difficult to spot malicious employees behaving inappropriately as they are a legitimate presence on the business network.

[0010] The reality is that modern threats bypass the traditional legacy defense tools on a daily basis. These tools need a new tool based on a new approach that can complement them and mitigate their deficiencies at scale across the entirety of digital organizations. In the complex modern world it is advantageous that the approach is fully automated as it is virtually impossible for humans to sift through the vast amount of security information gathered each minute within a digital business.

[0011] Cyber threat including email threats can be subtle and rapidly cause harm to a network. Having an automated response can allow a system to rapidly counter these threats.

SUMMARY

[0012] The inventors of the present invention have appreciated the need for additional triggers to trigger an artificial intelligence investigation of one or more entities such as devices, users, or ports. Provision of additional triggers, rather than the triggers only being based on anomaly alerts, is advantageous. Existing hypotheses relating to the cyber-threat may be supplemented by additionally triggered investigations, or new hypotheses may be formed, but significantly, investigation may be carried out on one or more entities being associated with behavior which would not have reached an alert threshold for investigation, and therefore would not have been otherwise investigated.

[0013] In addition, the inventors of the present invention have appreciated the need to be able to assess cyber-attacks at a high level. That is, it is advantageous to view a network compromise as a whole, rather than viewing each affected device individually at a lower level.

[0014] The invention is defined by the independent claims to which reference should now be made. Optional features are set forth in the dependent claims.

[0015] According to an aspect of the present invention, there is provided an apparatus comprising: one or more machine learning modules that are trained on a normal behavior of entities associated with a network and interactions between the entities; an interface configured to receive a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation, where the interface is configured to work with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, in order to determine whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat, and thus, trigger an autonomous response from an autonomous response module to mitigate the cyber-threat.

[0016] The signal from the external apparatus may be provided by a manual user input. The manual user input may comprise investigation instructions which may comprise one or more of: a time period to be investigated, designation of the one or more entities to be investigated, and a particular investigation environment.

[0017] The signal from the external apparatus may be provided by a third party threat intelligence component. The

third party threat intelligence component may provide additional data relating to behavior of the one or more entities

[0018] The apparatus may further comprise an analyser module configured to, in response to the determination of one or more incidents, generate a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially indicative of cyber threats.

[0019] The directed graph may comprise a plurality of nodes, each node of the plurality of nodes may correspond to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

[0020] The analyser module may be further configured to group the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

[0021] The apparatus may further comprise a formatting module configured to generate a visual representation of the directed graph for display.

[0022] The autonomous response module may be configurable to know when the response module should take the autonomous actions to mitigate the cyber-threat when one or more incidents are worth of being determined as a cyber-threat, where the autonomous response module has an administrative tool, configurable through the interface, to set what autonomous actions the autonomous response module can take, including types of actions and specific actions the autonomous response module is capable of.

[0023] According to another aspect of the present invention, there is provided a method for a cyber-threat defense system, the method comprising: using one or more machine learning models that are trained on a normal behavior of entities associated with a network and interactions between the entities; receiving, at an interface, a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation; the interface working with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, and determining whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat; and triggering an autonomous response from an autonomous response module to mitigate the cyber-threat.

[0024] The signal from the external apparatus may be provided by a manual user input at the interface. The manual user input may comprise investigation instructions which may comprise one or more of: a time period to be investigated, designation of the one or more entities to be investigated, and a particular investigation environment.

[0025] The signal from the external apparatus may be provided by a third party threat intelligence component. The third party threat intelligence component may provide additional data relating to behavior of the one or more entities.

[0026] The method may further comprise, in response to the determination of one or more incidents, generating a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially indicative of cyber-threats.

[0027] The directed graph may comprise a plurality of nodes, each node of the plurality of nodes corresponding to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

[0028] The method may further comprise grouping the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

[0029] The method may further comprise generating a visual representation of the directed graph.

[0030] According to another aspect of the present invention, there is provided a non-transitory computer-readable medium including executable instructions that, when executed with one or more processors, cause the cyber-threat defense system to perform any of the methods as described above.

DRAWINGS

[0031] The drawings refer to some embodiments of the design provided herein in. The invention will be described, by way of example, with reference to the drawings in which:

[0032] FIG. **1** illustrates a block diagram of an embodiment of a cyber-threat defense system embodying an aspect of the present invention with a cyber-threat module that references machine learning models that are trained on the normal behavior of email activity and user activity associated with at least the email system, where the cyber-threat module determines a threat risk parameter that factors in 'the likelihood that a chain of one or more unusual behaviors of the email activity and user activity under analysis fall outside of derived normal benign behavior'; and thus, are likely malicious behavior;

[0033] FIG. **2** illustrates a block diagram of an embodiment of the cyber-threat defense system embodying an aspect of the present invention, the cyber-threat defense system monitoring email activity and network activity to feed this data to correlate causal links between these activities to supply this input into the cyber-threat analysis;

[0034] FIG. **3** illustrates a block diagram of an embodiment of the cyber-threat module embodying an aspect of the present invention, the cyber-threat module determining a threat risk parameter that factors in how the chain of unusual behaviors correlate to potential cyber threats and 'the likelihood that this chain of one or more unusual behaviors of the email activity and user activity under analysis fall outside of derived normal benign behavior;' and thus, is malicious behavior;

[0035] FIG. **4** illustrates a block diagram of an embodiment of an example chain of unusual behavior for the email(s) in connection with the rest of the network under analysis;

[0036] FIG. **5** illustrates a block diagram of an embodiment of example autonomous actions that the autonomous rapid response module can be configured to take without a human initiating that action;

[0037] FIG. **6** illustrates an example cyber-threat defense system embodying an aspect of the present invention protecting an example network;

[0038] FIG. **7** illustrates an example of the network module informing the email module of a computer's network activity prior to the user of that computer receiving an email containing content relevant to that network activity;

[0039] FIG. **8** illustrates an example of the network module informing the email module of the deduced pattern of life information on the web browsing activity of a computer

prior to the user of that computer receiving an email which contains content which is not in keeping with that pattern of life;

[0040] FIG. **9** is a schematic diagram illustrating directed graph analysis embodying an aspect of the present invention;

[0041] FIG. **10** is a schematic diagram illustrating directed graph analysis embodying an aspect of the present invention;

[0042] FIG. **11** is a schematic diagram illustrating directed graph analysis embodying an aspect of the present invention; and

[0043] FIG. **12** is a flow diagram illustrating a method for a cyber-defense system embodying an aspect of the present invention.

[0044] 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

[0045] 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.

[0046] FIG. **1** shows some modules of an example cyber security appliance **100**, or cyber-threat defense system. Various Artificial Intelligence models and modules of a cyber security appliance **100** cooperate to protect a system, including but not limited to an email network, from cyber threats. The cyber security appliance **100** may include a trigger module, a gatherer module, an analyser module, an assessment module, a formatting module, an autonomous report composer, a data store, one or more Artificial Intelligence models trained on potential cyber threats and their characteristics, symptoms, remediations, etc., one or more Artificial Intelligence models trained with machine learning on a normal pattern of life for entities in the network, one or more Artificial Intelligence models trained with machine learning on threat report generation, and multiple libraries of text and visual representations to cooperate the library of page templates to populate visual representations, such as graphs, and text on the pages of the threat report. An example network of an email system will be used to illustrate portions of a cyber security appliance **100**.

[0047] It will be appreciated that an email system and network is used for exemplary purposes only, and embodiments of the present invention are applicable to any aspect of the network. For example, embodiments below describe AI models trained with machine learning on a normal email pattern of life for entities in an email network. However, it will be appreciated that arrangements of the present invention are equally applicable to AI models trained on entities associated with a network and interactions between the entities, the entities including at least one of users, ports, and devices.

[0048] Referring to FIG. **1**, the trigger module may detect time stamped data indicating an event is occurring and then triggers that something unusual is happening. In some embodiments, the gatherer module may be triggered by specific events or alerts of i) an abnormal behaviour, ii) a suspicious activity, and iii) any combination of both. The trigger module may identify, with one or more AI models trained with machine learning on a normal email pattern of life for entities in the email network, at least one of i) an abnormal behaviour, ii) a suspicious activity, and iii) any combination of both, from one or more entities in the system.

[0049] The inline data may be gathered on the deployment when the traffic is observed. The gatherer module may initiate a collection of data to support or refute each of the one or more possible cyber threat hypotheses that could include this abnormal behaviour or suspicious activity by the one or more AI models trained on possible cyber threats. The gatherer module cooperates with a data store. The data store 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.

[0050] The data store can store the metrics and previous threat alerts associated with network traffic for a period of time, which may be, 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 meta data associated with the network traffic in the data store. FIG. **2** illustrates an example cyber security appliance **100** using an intelligent-adversary simulator cooperating with a network module and network probes ingesting traffic data for network devices and network users in the network under analysis.

[0051] Referring back to FIG. **1**, the gatherer module 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 gatherer module for each possible hypothesis.

[0052] The gatherer module may further extract data, at the request of the analyser module, on each possible hypothetical threat that would include the abnormal behaviour or suspicious activity; and then, filter that collection of data down to relevant points of data to either 1) support or 2) refute each particular hypothesis of what the potential cyber threat, e.g. the suspicious activity and/or abnormal behaviour, relates to. The gatherer module and the data store can

cooperate to store an inbound and outbound email flow received over a period of time as well as autonomous actions performed by the autonomous response module on that email flow. The gatherer module may send the filtered down relevant points of data to either 1) support or 2) refute each particular hypothesis to the analyser module, comprised of one or more algorithms used by the AI models trained with machine learning on possible cyber threats to make a determination on a probable likelihood of whether that particular hypothesis is supported or refuted.

[0053] A feedback loop of cooperation between the gatherer module and the analyser module may be used to apply one or more models trained on different aspects of this process.

[0054] The analyser module can form one or more hypotheses on what are a possible set of activities including cyber threats that could include the identified abnormal behaviour and/or suspicious activity from the trigger module with one or more AI models trained with machine learning on possible cyber threats. The analyser module may request further data from the gatherer module to perform this analysis. The analyser module can cooperate with the one or more Artificial Intelligence models trained with machine learning on the normal email pattern of life for entities in the email network to detect anomalous email which is detected as outside the usual pattern of life for each entity, such as a user, of the email network. It will be appreciated again that, here, a normal email pattern of life and email network are used as examples only, and a pattern of life can be established for any appropriate aspect of the network. The analyser module can cooperate with the Artificial Intelligence models trained on potential cyber threats to detect suspicious emails that exhibit traits that may suggest a malicious intent, such as phishing links, scam language, sent from suspicious domains, etc. In addition, the gatherer module and the analyser module may use a set of scripts to extract data on each possible hypothetical threat to supply to the analyser module. The gatherer module and analyser module may use a plurality of scripts to walk through a step-by-step process of what to collect to filter down to the relevant data points (from the potentially millions of data points occurring in the network) to make a decision what is required by the analyser module.

[0055] The analyser module may further analyse a collection of system data, including metrics data, to support or refute each of the one or more possible cyber threat hypotheses that could include the identified abnormal behaviour and/or suspicious activity data with the one or more AI models trained with machine learning on possible cyber threats. The analyser module then generates at least one or more supported possible cyber threat hypotheses from the possible set of cyber threat hypotheses as well as could include some hypotheses that were not supported/refuted.

[0056] The analyser module may get threat information from Open Source APIs as well as from databases as well as information trained into AI models. Also, probes collect the user activity and the email activity and then feed that activity to the network module to draw an understanding of the email activity and user activity in the email system.

[0057] The analyser module learns how expert humans tackle investigations into specific cyber threats. The analyser module may use i) one or more AI models and/or ii) rules-based models and iii) combinations of both that are hosted within the plug-in appliance connecting to the network.

[0058] The AI models use data sources, such as simulations, database records, and actual monitoring of different human exemplar cases, as input to train the AI model on how to make a decision. The analyser module also may utilize repetitive feedback, as time goes on, for the AI models trained with machine learning on possible cyber threats via reviewing a subsequent resulting analysis of the supported possible cyber threat hypothesis and supply that information to the training of the AI models trained with machine learning on possible cyber threats in order to reinforce the model's finding as correct or inaccurate.

[0059] Each hypothesis has various supporting points of data and other metrics associated with that possible threat, and a machine learning algorithm will look at the relevant points of data to support or refute that particular hypothesis of what the suspicious activity and/or abnormal behaviour relates to.

[0060] The analyser module may perform analysis of internal and external data including readout from machine learning models, which output a likelihood of the suspicious activity and/or abnormal behaviour related for each hypothesis on what the suspicious activity and/or abnormal behaviour relates to with other supporting data to support or refute that hypothesis.

[0061] The assessment module may assign a probability, or confidence level, of a given cyber threat hypothesis that is supported, and a threat level posed by that cyber threat hypothesis, which includes this abnormal behaviour or suspicious activity, with the one or more AI models trained on possible cyber threats. The assessment module can cooperate with the autonomous response module to determine an appropriate response to mitigate various cyber-attacks that could be occurring.

[0062] The analyser module can reference machine learning models that are trained on the normal behaviour of email activity and user activity associated with at least the email system, where the analyser module cooperates with the assessment module to determine a threat risk parameter that factors in 'the likelihood that a chain of one or more unusual behaviours of the email activity and user activity under analysis fall outside of derived normal benign behaviour;' and thus, are likely malicious behaviour.

[0063] FIG. **7** illustrates a block diagram of an embodiment of the cyber-threat module comparing the analyzed metrics on the user network and computer activity and email activity compared to their respective moving benchmark of parameters that correspond to the normal pattern of life for the computing system used by the self-learning machine learning models and the corresponding potential cyber threats. The cyber-threat module can then determine, in accordance with the analyzed metrics and the moving benchmark of what is considered normal behavior, a cyber-threat risk parameter indicative of a likelihood of a cyber-threat.

[0064] The cyber-threat defense system **100** may also include one or more machine learning models trained on gaining an understanding of a plurality of characteristics on an email itself and its related data including classifying the properties of the email and its meta data.

[0065] The cyber-threat module can also reference the machine learning models trained on an email itself and its

related data to determine if an email or a set of emails under analysis have potentially malicious characteristics. The cyber-threat module can also factor this email characteristics analysis into its determination of the threat risk parameter.

[0066] The network module may have one or more machine learning models trained on a normal behavior of users, devices, and interactions between them, on a network, which is tied to the email system. A user interface has one or more windows to display network data and one or more windows to display emails and cyber security details about those emails through the same user interface on a display screen, which allows a cyber professional to pivot between network data and email cyber security details within one platform, and consider them as an interconnected whole rather than separate realms on the same display screen.

[0067] The cyber-threat module can also factor this network analysis into its determination of the threat risk parameter.

[0068] FIG. **8** illustrates an example of the network module informing the email module of the deduced pattern of life information on the web browsing activity of a computer prior to the user of that computer receiving an email which contains content which is not in keeping with that pattern of life.

[0069] The user interface can display a graph **220** (illustrated in FIG. **4**) of an example chain of unusual behavior for the email(s) in connection with the rest of the network under analysis.

[0070] The network & email module can tie the alerts and events from the email realm to the alerts and events from the network realm.

[0071] The cyber-threat module cooperates 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 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.

[0072] This is 'a behavioral pattern analysis' of what are the unusual behaviors of the network/system/device/user/email under analysis by the cyber-threat module and the machine learning models. The cyber defense system uses unusual behavior deviating from the normal behavior and then builds a chain of unusual behavior and the causal links between the chain of unusual behavior to detect cyber threats. An example behavioral pattern analysis of what are the unusual behaviors may be as follows. The unusual pattern may be determined by filtering out what activities/events/alerts that fall within the window of what is the normal pattern of life for that network/system/device/user/email under analysis, and then the pattern of the behavior of the activities/events/alerts that are left, after the filtering, can be analyzed to determine whether that pattern is indicative of a behavior of a malicious actor—human, program, email, or other threat. The defense system can go back and pull in some of the filtered out 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 defense system 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. When the behavioral pattern analysis of any individual behavior or of the chain as a group is believed to be indicative of a malicious threat, then a score of how confident is the defense 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 parameter (e.g. score or probability) indicative of what level of threat does this malicious actor pose to the system. Lastly, the cyber-threat defense system is configurable in its user interface of the defense system on what type of automatic response actions, if any, the defense system may take when for different types of cyber threats that are equal to or above a configurable level of threat posed by this malicious actor.

[0073] The cyber-threat module may chain the individual alerts and events that form the unusual pattern into a distinct item for cyber-threat analysis of that chain of distinct alerts and/or events. The cyber-threat module may reference the one or more machine learning models trained on 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.

[0074] One or more machine learning models may also be trained on characteristics and aspects of all manner of types of cyber threats to analyze the threat risk associated with the chain/cluster of alerts and/or events forming the unusual pattern. The machine learning technology, using advanced mathematics, can detect previously unidentified threats, without rules, and automatically defend networks.

[0075] The models may perform by the threat detection through a probabilistic change in normal behavior through the application of an unsupervised Bayesian mathematical model to detect behavioral change in computers and computer networks. The core threat detection system is termed the 'Bayesian probabilistic'. 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. From the email and network raw sources of data, a large number of metrics can be derived each producing time series data for the given metric.

[0076] The detectors in the cyber-threat module including its network module and email module components can be discrete mathematical models that implement a specific mathematical method against different sets of variables with the target. Thus, each model is specifically targeted on the pattern of life of alerts and/or events coming from, for example, i) that cyber security analysis tool, ii) analyzing various aspects of the emails, iii) coming from specific devices and/or users within a system, etc.

[0077] At its core, the cyber-threat defense system mathematically characterizes what constitutes 'normal' behavior based on the analysis of a large number/set of different measures of a device's network behavior. The cyber-threat defense system can build a sophisticated 'pattern of life'—that understands what represents normality for every person, device, email activity, and network activity in the system being protected by the cyber-threat defense system.

[0078] As discussed, each machine learning model may be trained on specific aspects of the normal pattern of life for the system such as devices, users, network traffic flow, outputs from one or more cyber security analysis tools analyzing the system, email contact associations for each user, email characteristics, etc. The one or more machine learning models may use at least unsupervised learning algorithms to establish what is the normal pattern of life for the system. The machine learning models can train on both i) the historical normal distribution of alerts and events for that system as well as ii) factored in is a normal distribution information from similar peer systems to establish the normal pattern of life of the behavior of alerts and/or events for that system. Another set of machine learning models train on characteristics of emails and the activities and behavior of its email users to establish a normal for these.

[0079] Note, when the models leverage at least two different approaches to detecting anomalies: e.g. comparing each system's behavior to its own history, and comparing that system to its peers' history and/or e.g. comparing an email to both characteristics of emails and the activities and behavior of its email users, this multiple source comparison allows the models to avoid learning existing bad behavior as 'a normal' because compromised devices/users/components/emails will exhibit behavior different to their immediate peers.

[0080] In addition, the one or more machine learning models can use the comparison of i) the normal pattern of life for that system corresponding to the historical normal distribution of alerts and events for that system mapped out in the same multiple dimension space to ii) the current chain of individual alerts and events behavior under analysis. This comparison can yield detection of the one or more unusual patterns of behavior within the plotted individual alerts and/or events, which allows the detection of previously unidentified cyber threats compared to finding cyber threats with merely predefined descriptive objects and/or signatures. Thus, increasingly intelligent malicious cyber threats that try to pick and choose when they take their actions in order to generate low level alerts and event will still be detected, even though they have not yet been identified by other methods of cyber analysis. These intelligent malicious cyber threats can include malware, spyware, key loggers, malicious links in an email, malicious attachments in an email, etc. as well as nefarious internal information technology staff who know intimately how to not set off any high level alerts or events.

[0081] In essence, the plotting and comparison is way to filter out what is normal for that system and then be able to focus the analysis on what is abnormal or unusual for that system. Then, for each hypothesis of what could be happening with the chain of unusual events and/or alerts, the gatherer module may gather additional metrics from the data store including the pool of metrics originally considered 'normal behavior' to support or refute each possible hypothesis of what could be happening with this chain of unusual behavior under analysis.

[0082] Note, each of the individual alerts and/or events in a chain of alerts and/or events that form the unusual pattern can indicate subtle abnormal behavior; and thus, each alert and/or event can have a low threat risk associated with that individual alert and/or event. However, when analyzed as a distinct chain/grouping of alerts and/or events behavior forming the chain of unusual pattern by the one or more machine learning models, then that distinct chain of alerts and/or events can be determine to now have a much higher threat risk than any of the individual alerts and/or events in the chain.

[0083] Note, in addition, today's cyberattacks can be of such severity and speed that a human response cannot happen quickly enough. Thanks to these self-learning advances, it is now possible for a machine to uncover these emerging threats and deploy appropriate, real-time responses to fight back against the most serious cyber threats.

[0084] The threat detection system has the ability to self-learn and detect normality in order to spot true anomalies, allowing organizations of all sizes to understand the behavior of users and machines on their networks at both an individual and group level. Monitoring behaviors, rather than using predefined descriptive objects and/or signatures, means that more attacks can be spotted ahead of time and extremely subtle indicators of wrongdoing can be detected. Unlike traditional legacy defenses, a specific attack type or new malware does not have to have been seen first before it can be detected. A behavioral defense approach mathematically models both machine, email, and human activity behaviorally, at and after the point of compromise, in order to predict and catch today's increasingly sophisticated cyber-attack vectors. It is thus possible to computationally establish what is normal, in order to then detect what is abnormal. In addition, the machine learning constantly revisits assumptions about behavior, using probabilistic mathematics. The cyber-threat defense system'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.

[0085] The user interface and output module may also project the individual alerts and/or events forming the chain of behavior onto the user interface with at least three-dimensions of i) a horizontal axis of a window of time, ii) a vertical axis of a scale indicative of the threat risk assigned for each alert and/or event in the chain and a third dimension of iii) a different color (e.g. red, blue, yellow, etc., and if gray scale—different shades of gray black and white with potentially different hashing patterns) for the similar characteristics shared among the individual alerts and events forming the distinct item of the chain. These similarities of events and/or alerts in the chain may be, for example, alerts or events are coming from same device, same user credentials, same group, same source ID, same destination IP address, same types of data transfers, same type of unusual activity, same type of alerts, same rare connection being made, same type of events, etc., so that a human can visually see what spatially and content-wise is making up a particular chain rather than merely viewing a textual log of data. Note, once the human mind visually sees the projected pattern and corresponding data, then the human can ultimately decide if a cyber-threat is posed. Again, the at least three-dimensional projection helps a human synthesize this information more easily. The visualization onto the User Interface allows a human to see data that supports or refutes why the cyber-threat defense system thinks these aggregated alerts and/or events could be potentially malicious. Also, instead of generating the simple binary outputs 'malicious' or 'benign,'

the cyber-threat defense system's mathematical algorithms produce outputs that indicate differing degrees of potential compromise.

[0086] The cyber-threat defense system **100** may use at least three separate machine learning models. Each machine learning model may be trained on specific aspects of the normal pattern of life for the system such as devices, users, network traffic flow, outputs from one or more cyber security analysis tools analyzing the system, etc. One or more machine learning models may also be trained on characteristics and aspects of all manner of types of cyber threats. One or more machine learning models may also be trained on characteristics of emails themselves.

[0087] The email module monitoring email activity and the network module monitoring network activity may both feed their data to a network & email coordinator module to correlate causal links between these activities to supply this input into the cyber-threat module. The application of these causal links is demonstrated in the block diagrams of FIG. **7** and FIG. **8**.

[0088] The cyber-threat module can also factor this network activity link to a particular email causal link analysis into its determination of the threat risk parameter (see FIG. **8**).

[0089] The cyber-threat defense system **100** uses various probes to collect the user activity and the email activity and then feed that activity to the data store and as needed to the cyber-threat module and the machine learning models. The cyber-threat module uses the collected data to draw an understanding of the email activity and user activity in the email system as well as updates a training for the one or more machine learning models trained on this email system and its users. For example, email traffic can be collected by putting hooks into the e-mail application, such as Outlook or Gmail, and/or monitoring the internet gateway from which the e-mails are routed through. Additionally, probes may collect network data and metrics via one of the following methods: port spanning the organizations existing network equipment; inserting or re-using an in-line network tap, and/or accessing any existing repositories of network data. (e.g. See FIG. **2**)

[0090] The cyber-threat defense system **100** may use multiple user interfaces. A first user interface may be constructed to present an inbox-style view of all of the emails coming in/out of the email system and any cyber security characteristics known about one or more emails under analysis. The user interface with the inbox-style view of emails has a first window/column that displays the one or more emails under analysis and a second window/column with all of the relevant security characteristics known about that email or set of emails under analysis. The complex machine learning techniques determine anomaly scores which describe any deviation from normal that the email represents, these are rendered graphically in a familiar way that users and cyber professionals can recognize and understand.

[0091] The cyber-threat defense system **100** can then take actions to counter detected potential cyber threats.

[0092] The autonomous response module, rather than a human taking an action, can be configured to cause one or more rapid autonomous actions to be taken to contain the cyber-threat when the threat risk parameter from the cyber-threat module is equal to or above an actionable threshold. The cyber-threat module's configured cooperation with the autonomous response module, to cause one or more autonomous actions to be taken to contain the cyber threat, improves computing devices in the email 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.

[0093] The cyber-threat defense system **100** may be hosted on a device, on one or more servers, and/or in its own cyber-threat appliance platform. (e.g. see FIG. **2**)

[0094] FIG. **2** illustrates a block diagram of an embodiment of the cyber-threat defense system monitoring email activity and network activity to feed this data to correlate causal links between these activities to supply this input into the cyber-threat analysis. The network can include various computing devices such as desktop units, laptop units, smart phones, firewalls, network switches, routers, servers, databases, Internet gateways, the cyber-threat defense system **100**, etc.

[0095] The network module uses the probes to monitor network activity and can reference the machine learning models trained on a normal behavior of users, devices, and interactions between them or the internet which is subsequently tied to the email system.

[0096] The user interface has both i) one or more windows to present/display network data, alerts, and events, and ii) one or more windows to display email data, alerts, events, and cyber security details about those emails through the same user interface on a display screen. These two sets of information shown on the same user interface on the display screen allows a cyber professional to pivot between network data and email cyber security details within one platform, and consider them as an interconnected whole rather than separate realms.

[0097] The network module and its machine learning models are utilized to determine potentially unusual network activity in order to provide an additional input of information into the cyber-threat module in order to determine the threat risk parameter (e.g. a score or probability) indicative of the level of threat.

[0098] A particular user's network activity can be tied to their email activity because the network module observes network activity and the network & email coordinator module receives the network module observations to draw that into an understanding of this particular user's email activity to make an appraisal of potential email threats with a resulting threat risk parameter tailored for different users in the e-mail system. The network module tracks each user's network activity and sends that to the network & email coordinator component to interconnect the network activity and email activity to closely inform one-another's behavior and appraisal of potential email threats.

[0099] The cyber-threat defense system **100** can now track possible malicious activity observed by the network module on an organization's network back to an event such as a specific email event observed by the e-mail module, and use the autonomous rapid response module to shut down any potentially harmful activity on the network itself, and also freeze any similar email activity triggering the harmful activity on the network.

[0100] The probes collect the user activity as well as the email activity. The collected activity is supplied to the data store and evaluated for unusual or suspicious behavioral activity, e.g. alerts, events, etc., which is evaluated by the

cyber-threat module to draw an understanding of the email activity and user activity in the email system. The collected data can also be used to potentially update the training for the one or more machine learning models trained on the normal pattern of life for this email system, its users and the network and its entities.

[0101] An example probe for the email system may be configured to work directly with an organization's email application, such as an Office 365 Exchange domain and receive a Blind Carbon Copy (BCC) of all ingoing and outgoing communications. The email module will inspect the emails to provide a comprehensive awareness of the pattern of life of an organization's email usage.

[0102] FIG. **3** illustrates a block diagram of an embodiment of the cyber-threat module determining a threat risk parameter that factors in how the chain of unusual behaviors correlate to potential cyber threats and 'the likelihood that this chain of one or more unusual behaviors of the email activity and user activity under analysis fall outside of derived normal benign behavior;' and thus, is malicious behavior.

[0103] The user interface **150** can graphically display logic, data, and other details that the cyber-threat module goes through.

[0104] The user interface **150** displays an example email that when undergoing analysis exhibits characteristics, such as header, address, subject line, sender, recipient, domain, etc. that are not statistically consistent with the normal emails similar to this one.

[0105] Thus, the user interface **150** displays an example email's unusual activity that has it classified as a behavioral anomaly.

[0106] During the analysis, the email module can reference the one or more machine learning models that are self-learning models trained on a normal behavior of email activity and user activity associated with an email system. This can include various e-mail policies and rules that are set for this email system. The cyber-threat module may also reference the models that are trained on the normal characteristics of the email itself. The cyber-threat module can apply these various trained machine learning models to data including metrics, alerts, events, meta data from the network module and the email module. In addition, a set of AI models may be responsible for learning the normal 'pattern of life' for internal and external address identities in connection with the rest of the network, for each email user. This allows the system to neutralize malicious emails which deviate from the normal 'pattern of life' for a given address identity for that user in relation to its past, its peer group, and the wider organization.

[0107] Next, the email module has at least a first email probe to inspect an email at the point it transits through the email application, such as Office 365, and extracts hundreds of data points from the raw email content and historical email behavior of the sender and the recipient. These metrics are combined with pattern of life data of the intended recipient, or sender, sourced from the data store. The combined set of the metrics are passed through machine learning algorithms to produce a single anomaly score of the email, and various combinations of metrics will attempt to generate notifications which will help define the 'type' of email.

[0108] Email threat alerts, including the type notifications, triggered by anomalies and/or unusual behavior of 'emails and any associated properties of those emails' are used by the cyber-threat module to better identify any network events which may have resulted from an email borne attack.

[0109] In conjunction with the specific threat alerts and the anomaly score, the system may provoke actions upon the email designed to prevent delivery of the email or to neutralize potentially malicious content.

[0110] Next, the data store stores the metrics and previous threat alerts associated with each email for a period of time, which is, by default, at least **27** days. This corpus of data is fully searchable from within the user interface **150** and presents an invaluable insight into mail flow for email administrators and security professionals.

[0111] Next, the cyber-threat module can issue an anomaly rating even when an unusual email does not closely relate to any identifiable malicious email. This value indicates how unusual the cyber-threat module considers this email to be in comparison to the normal pattern of life for the organization and the specific internal user (either inbound recipient or outbound sender). The cyber-threat module considers over **750** metrics and the organizational pattern of life for unusual behavior for a window of time. For example, the cyber-threat module considers metrics and the organizational pattern of life for unusual behavior and other supporting metrics for the past 7 days when computing the anomaly score, which is also factored into the final threat risk parameter.

[0112] 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 cyber security appliance uses unusual behaviour deviating from the normal behaviour and then builds a chain of unusual behaviour and the causal links between the chain of unusual behaviour to detect cyber threats (for example see FIG. **4**). FIG. **4** illustrates a block diagram in the form of a graph of an embodiment of an example chain of unusual behaviour for the email(s) deviating from a normal pattern of life in connection with the rest of the network under analysis. The unusual pattern can be determined by filtering out what activities/events/alerts that fall within the window of what is the normal pattern of life for that network/system/device/user under analysis, and then the pattern of the behaviour of the activities/events/alerts that are left, after the filtering, can be analysed to determine whether that pattern is indicative of a behaviour of a malicious actor—human, program, or other threat. Next, the cyber security appliance can go back and pull in some of the filtered out 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. 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 is configurable in a user interface, by a user, enabling what type of automatic response actions, if any, the cyber security appliance 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.

[0113] The assessment module may rank supported candidate cyber threat hypotheses by a combination of likelihood that this candidate cyber threat hypothesis is supported as well as a severity threat level of this incident type.

[0114] The formatting module can be coded to generate the report with the identified critical devices connecting to the virtualized instance of the network that should have the priority to allocate security resources to them, along with one or more portions of the constructed graph (See FIG. **2**). The formatting module can have an autonomous email-report composer that cooperates with the various Al models and modules of the cyber security appliance **100** as well as at least a set of one or more libraries of sets of prewritten text and visual representations to populate on templates of pages in the email threat report. The autonomous email-report composer can compose an email threat report on cyber threats that is composed in a human-readable format with natural language prose, terminology, and level of detail on the cyber threats aimed at a target audience being able to understand the terminology and the detail. The modules and AI models cooperate with the autonomous email-report composer to indicate in the email threat report, for example, an email attack's 1) purpose and/or 2) targeted group (such as members of the finance team, or high-level employees).

[0115] The formatting module may format, present a rank for, and output the current email threat report, from a template of a plurality of report templates, that is outputted for a human user's consumption in a medium of, any of 1) a printable report, 2) presented digitally on a user interface, 3) in a machine readable format for further use in machine-learning reinforcement and refinement, and 4) any combination of the three.

[0116] The system may use at least three separate machine learning models. For example, a machine learning model may be trained on specific aspects of the normal pattern of life for entities in the system, such as devices, users, network traffic flow, outputs from one or more cyber security analysis tools analysing the system, etc. One or more machine learning models may also be trained on characteristics and aspects of all manner of types of cyber threats. One or more machine learning models may also be trained on composing email threat reports.

[0117] The various modules cooperate with each other, the AI models, and the data store to carry out the operations discussed herein. The trigger module, the Al models, the gatherer module, the analyser module, the assessment module, the formatting module, and the data store cooperate to improve the analysis and formalized report generation with less repetition to consume less CPU cycles, as well as doing this more efficiently and effectively than humans. For example, the modules can repetitively go through these steps and re-duplicate steps to filter and rank the one or more supported possible cyber threat hypotheses from the possible set of cyber threat hypotheses and/or compose the detailed information to populate into the email threat report.

[0118] One or more processing units are configured to execute software instructions associated with the intelligent-adversary simulator, the formatting module, other modules, and models in the cyber security appliance **100**.

[0119] One or more non-transitory storage mediums are configured to store at least software associated with the intelligent-adversary simulator, the other modules, and AI models.

[0120] FIG. **5** illustrates a block diagram of an embodiment of example autonomous actions that the autonomous rapid response module can be configured to take without a human initiating that action.

[0121] The autonomous rapid response module is configurable, via the user interface **150**, to know when it should take the autonomous actions to contain the cyber-threat when i) a known malicious email or ii) at least highly likely malicious email is determined by the cyber-threat module. The autonomous rapid response module has an administrative tool, configurable through the user interface, to program/set what autonomous actions the autonomous rapid response module can take, including types of actions and specific actions the autonomous rapid response module is capable of, when the cyber-threat module indicates the threat risk parameter is equal to or above the actionable threshold, selectable by the cyber professional, that the one or more emails under analysis are at least highly likely to be malicious.

[0122] The types of actions and specific actions the autonomous rapid response module customizable for different users and parts of the system; and thus, configurable for the cyber professional to approve/set for the autonomous rapid response module to automatically take those actions and when to automatically take those actions.

[0123] The autonomous rapid response module has a library of response actions types of actions and specific actions the autonomous rapid response module is capable of, including focused response actions selectable through the user interface **150** that are contextualized to autonomously act on specific email elements of a malicious email, rather than a blanket quarantine or block approach on that email, to avoid business disruption to a particular user of the email system. The autonomous rapid response module is able to take measured, varied actions towards those email communications to minimize business disruption in a reactive, contextualized manner.

[0124] The autonomous response module works hand-in-hand with the AI models to neutralize malicious emails, and deliver preemptive protection against targeted, email-borne attack campaigns in real time.

[0125] The cyber-threat module cooperating with the autonomous response module can detect and contain, for example, an infection in the network, recognize that the infection had an email as its source, and identify and neutralize that malicious email by either removing that from the corporate email account inboxes, or simply stripping the malicious portion of that before the email reaches its intended user. The autonomous actions range from flattening attachments or stripping suspect links, through to holding emails back entirely if they pose a sufficient risk.

[0126] The cyber-threat module can identify the source of the compromise and then invoke an autonomous miss response action by sending a request to the autonomous response model. This autonomous response action will rapidly stop the spread of an emerging attack campaign, and give human responders the crucial time needed to catch up.

[0127] In an embodiment, initially, the autonomous response module can be run in human confirmation mode—all autonomous, intelligent interventions must be confirmed initially by a human operator. As the autonomous response module refines and nuances its understanding of an organization's email behavior, the level of autonomous action can be increased until no human supervision is required for each

autonomous response action. Most security teams will spend very little time in the user interface **150** once this level is reached. At this time, the autonomous response module response action neutralizes malicious emails without the need for any active management. The autonomous response module may take one or more proactive or reactive action against email messages, which are observed as potentially malicious. Actions are triggered by threat alerts or by a level of anomalous behavior as defined and detected by the cyber-security system and offer highly customizable, targeted response actions to email threat that allows the end user to remain safe without interruption. Suspect email content can be held in full, autonomously with selected users exempted from this policy, for further inspection or authorization for release. User behavior and notable incidents can be mapped, and detailed, comprehensive email logs can be filtered by a vast range of metrics compared to the model of normal behavior to release or strip potentially malicious content from the email.

Example Possible Actions

[0128] The following selection of example actions, categorized into delivery actions, attachment actions, link actions, header and body actions, etc., appear on the dashboard and can be taken by or at least suggested to be taken by the autonomous response module when the threat risk parameter is equal to or above a configurable set point set by a cyber security professional:

[0129] Hold Message: The autonomous response module has held the message before delivery due to suspicious content or attachments. Held emails can be reprocessed and released by an operator after investigation. The email will be prevented from delivery, or if delivery has already been performed, removed from the recipient's inbox. The original mail will be maintained in a buffered cache by the data store and can be recovered, or sent to an alternative mailbox, using the 'release' button in the user interface **150**.

[0130] Lock Links: The autonomous response module replaces the URL of a link such that a click of that link will first divert the user via an alternative destination. The alternative destination may optionally request confirmation from the user before proceeding. The original link destination and original source will be subject to additional checks before the user is permitted to access the source.

[0131] Convert Attachments: The autonomous response module converts one or more attachments of this email to a safe format, flattening the file typically by converting into a PDF through initial image conversion. This delivers the content of the attachment to the intended recipient, but with vastly reduced risk. For attachments which are visual in nature, such as images, pdfs and Microsoft Office formats, the attachments will be processed into an image format and subsequently rendered into a PDF (in the case of Microsoft Office formats and PDFs) or into an image of the original file format (if an image). In some email systems, the email attachment may be initially removed and replaced with a notification informing the user that the attachment is undergoing processing. When processing is complete the converted attachment will be inserted back into the email.

[0132] Double Lock Links: The autonomous response module replaces the URL with a redirected Email link. If the link is clicked, the user will be presented with a notification to that user that they are not permitted to access the original destination of the link. The user will be unable to follow the link to the original source, but their intent to follow the link will be recorded by the data store via the autonomous response module.

[0133] Strip Attachments: The autonomous response module strips one or more attachments of this email. Most file formats are delivered as converted attachments; file formats which do not convert to visible documents (e.g. executables, compressed types) are stripped to reduce risk. The 'Strip attachment' action will cause the system to remove the attachment from the email, and replace it with a file informing the user that the original attachment was removed.

[0134] Junk action: The autonomous response module will ensure the email classified as junk or other malicious email is diverted to the recipient's junk folder, or other nominated destination such as 'quarantine'.

[0135] Redirect: The autonomous response module will ensure the email is not delivered to the intended recipient but is instead diverted to a specified email address.

[0136] Copy: The autonomous response module will ensure the email is delivered to the original recipient, but a copy is sent to another specified email address.

[0137] Do not hold or alter: Can be set on a particular user basis. The autonomous response module will ensure the email(s) are never held, and never altered in any way by the system, regardless of actions performed by other models or triggered by the general anomaly threat level.

[0138] Take no action on attachments: Can be set on a particular user basis. This action will override any attachment actions that would be otherwise taken by the autonomous response module whether in response to a particular threat alert or overall detected anomaly level.

[0139] Header and body action: The autonomous response module will insert specific, custom text into the email Body or Subject Line to add to or substitute existing text, images, or other content in a header and/or body of the email.

[0140] Unspoof: The autonomous response module will identify standard email header address fields (e.g. rfc822 type) and replace the Personal Name and the header email address with an alternative name or email address which might reveal more about the true sender of the email. This mechanism significantly reduces the psychological impact of spoof attempts.

[0141] The paragraphs above set out autonomous actions on emails and those initiated by an email component. Autonomous actions may, however, be taken by a wider cybersecurity system. For example, the AI Analyst may be implemented, in other areas of autonomous actions, such as in a network, in SaaS, and in host-based agents.

[0142] For example, a cyber-threat coordinator-component may be provided that identifies devices and/or users that are in a breach state of a benchmark of parameters, utilized by AI models, that correspond to the normal pattern of life for the network. The cyber-threat coordinator-component sends an external communication to selected network devices in order to initiate actions with that network device in order to counter a behavior of a detected threat of at least one a user and/or a device acting abnormal to the normal pattern of life on the network. The initiated actions are also targeted to minimize an impact on other network devices and users that are i) currently active in the network and ii) that are not in breach of being outside the normal behavior benchmark. The cyber-threat coordinator-component can be an autonomous self-learning digital response coordinator that is trained specifically to control and reconfigure the

actions of traditional legacy computer defenses (e.g. firewalls, switches, proxy servers, etc.) to contain threats propagated by, or enabled by, networks and the internet. The cyber-threat coordinator-component can then take actions to counter detected potential cyber threats. The autonomous response module, rather than a human taking an action, can be configured to cause one or more rapid autonomous actions to be taken to contain the cyber threat when the threat risk parameter from the cyber threat module is equal to or above an actionable threshold. The autonomous response module can communicate to hosts/IPs to one or more third-party networking devices such as Firewalls so that they can end current/prevent future connections.

[0143] Some example types of capabilities of network devices; and thus, actions that can be directed from the cyber-threat coordinator-component:

[0144] open or block access to a port and IP address combination;

[0145] open or block access for a limited amount of time;

[0146] redirect communication to another URL or redirect the user to a specific URL;

[0147] prevent the device from reading a file share on server;

[0148] block access to specific IP addresses and/or to specific types of devices;

[0149] open or block outbound web access for a specific user and/or device;

[0150] open or block outbound DNS access for a specific user and/or device;

[0151] open or block all outbound access for a specific user and/or device;

[0152] altering the user's permissions, restricting login in for that person; blocking a connection based on a source or destination address;

[0153] look at type of traffic and then reroute the network traffic of a specific type via security device;

[0154] slow down a transfer rate by allocating a lowest bandwidth to that port, user, or device;

[0155] quarantine files and/or send shutdown signal to a device;

[0156] close authentication to different parts of the network; and

[0157] do all of the above actions until a specified window of time is elapsed, such as an hour, or until reset by a human network administrator. A cyber threat defense system can incorporate data from a Software-as-a-Service (SaaS) application administrated by a third-party operator to identify cyber threats related to that SaaS application. The cyber threat defense module can have a SaaS module to collect third-party event data from the third-party operator. The cyber threat defense system can have a comparison module to compare third-party event data for a network entity to at least one machine-learning model of a network entity using a normal behavior benchmark to spot behavior deviating from normal benign behavior. The comparison module can identify whether the network entity is in a breach state. The cyber threat defense system can have a cyber threat module to identify whether the breach state and a chain of relevant behavioral parameters correspond to a cyber threat. An autonomous response module can execute an autonomous response in response to the cyber threat. The autonomous response module, rather than a human taking an action, can be configured to cause one or more rapid autonomous actions to be taken to contain the cyber threat when the threat risk parameter from the cyber threat module is equal to or above an actionable threshold.

[0158] The cyber threat defense system can decide based upon the cyber threat detected that disabling the user account or device for a fixed period is the most appropriate and least disruptive method to suppress the action. Note, the cyber threat defense system is not limited to applying this method just on the specific platform where the threat was detected; the user may be disabled on all SaaS platforms, all distributed platforms (such as Cloud, Email, etc.) or disabled on the SaaS platform due to potentially malicious actions on the physical network. The connector may send an application programming interface (API) request to the SaaS product to modify the affected user and remove their permission or alternatively send a reset (RST) packet. The autonomous response module thus can log users out of their accounts, disable their accounts for a period of time, etc.

[0159] A method may be provided for detection of a cyber-threat to an end point computing device, such as a mobile phone, tablet, laptop, desktop, Internet of Things appliance, etc., via using a software End Point Monitoring Agent resident on the computing device and that reports Pattern of Life metadata, events and alerts to a gateway hosting a cyber threat response system at periodic intervals whenever connected to the internet, where the End Point Monitoring agent exchanges communications with the cyber threat response system on the network and/or on a backend platform, and then based on an identified cyber threat and a threat score, an autonomous response module is configured to take an action preapproved by a human user to autonomously attempt to counter a malicious threat.

[0160] Again, the autonomous response module, rather than a human taking an action, is configured to cause one or more autonomous actions to be taken to contain the cyber threat when a potential cyber threat is detected. i) the cyber security appliance can have the autonomous response module, or ii) a portion of the autonomous response module can exist on the endpoint agent while the majority remains on the cyber security appliance due to greater processing power. A user programmable interface hosted on the cyber security appliance having any of i) fields, ii) menus, and iii) icons is scripted to allow a user to preauthorize the autonomous response module to take actions to contain the cyber threat. The user programmable fields/menus to allow a user to preauthorize the module to take actions such as killing individual processes, revoking specific privileges, preventing the download of specific files, allowing only processes observed in the pattern of life for peer devices to be active for a set period, asking other EPPs to quarantine suspicious files, etc. while not disturbing operations of other processes going on inside that device. The user interface has the granularity in options available to the user to program the autonomous response module to take very specific actions such as killing individual processes, revoking specific privileges while still permitting other permissions for that user, getting live terminal access, preventing the download of specific files, allowing only processes observed in the pattern of life for peer devices to be active for a set period, asking other EPPs to quarantine suspicious files, etc. while not shutting down an entire device, or blocking all outside communications, or revoking one or more but not all of that user's privileges. Actions such as revoking only some user privileges or enforcing the peer pattern of life allow the user to continue working but just not perform certain connections

or run certain processes, which most likely a malicious piece of software was initiating, such as accessing and downloading sensitive files while the user, completely unaware of the malicious software using their credentials, is doing a normal activity for that user such as typing out a document or entering data into a program.

[0161] Example autonomous actions available to be pre-approved by a human user for the autonomous response module can include a general prompt to the user on the display screen of the end-point computing-device along with the action of: Prevent or slow down activity related to the threat; Quarantine or semi-quarantine people, processes, devices; Feed threat intelligence to EPP and EDR processes and devices to take third party or vendor specific actions such as quarantine or firewall blocks; ending anomalous processes on the client device, etc.; and in most cases without disrupting the normal day to day activity of users or other processes on the end-point computing-device.

[0162] FIG. **6** illustrates an example cyber security appliance 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.

[0163] 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**.

[0164] 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**.

[0165] In this exemplary embodiment of the cyber security appliance **100**, computer **1** on the first computer system **10** has the hardware and software 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 from the lightweight probes.

[0166] The cyber security appliance **100** in computer **1** builds and maintains a dynamic, ever-changing model of the 'normal behaviour' 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.

[0167] 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.

[0168] 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 threat detection system 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 'normal' of the model of the normal pattern of life is used as a moving benchmark, allowing the system to spot behaviour on a system that seems to fall outside of this normal pattern of life, and flags this behaviour as anomalous, requiring further investigation.

[0169] 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, using normal software protocol. Any attack process thus stops or 'backs off' automatically if the mouse or keyboard is used. However, yet more sophisticated attacks try the opposite, hiding in memory under the guise of a normal process and stealing CPU cycles only when the machine is active, in an attempt to defeat a relatively-simple policing process. These sophisticated attackers look for activity that is not directly associated with the user's input. As an APT (Advanced Persistent Threat) attack typically has very long mission windows of weeks, months or years, such processor cycles can be stolen so infrequently that they do not impact machine performance. But, however cloaked and sophisticated the attack is, there will always be a measurable delta, even if extremely slight, in typical machine behavior, between pre and post compromise. This behavioral delta can be observed and acted on with the form of Bayesian mathematical analysis used by the threat detection system installed on the computer **1**.

[0170] The cyber defense self-learning platform uses machine-learning technology. The machine learning technology, using advanced mathematics, can detect previously unidentified threats, without rules, and automatically defend networks. Note, today's attacks can be of such severity and speed that a human response cannot happen quickly enough. Thanks to these self-learning advances, it is now possible for a machine to uncover emerging threats and deploy appropriate, real-time responses to fight back against the most serious cyber threats.

[0171] The cyber security appliance 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**.

[0172] The threat detection system has the ability to self-learn and detect normality in order to spot true anomalies, allowing organizations of all sizes to understand the behavior of users and machines on their networks at both an individual and group level. Monitoring behaviors, rather than using predefined descriptive objects and/or signatures, means that more attacks can be spotted ahead of time and extremely subtle indicators of wrongdoing can be detected. Unlike traditional legacy defenses, a specific attack type or new malware does not have to have been seen first before it can be detected. A behavioral defense approach mathematically models both machine and human activity behaviorally, at and after the point of compromise, in order to predict and

catch today's increasingly sophisticated cyber-attack vectors. It is thus possible to computationally establish what is normal, in order to then detect what is abnormal.

[0173] This intelligent system is capable of making value judgments and carrying out higher value, more thoughtful tasks. Machine learning requires complex algorithms to be devised and an overarching framework to interpret the results produced. However, when applied correctly these approaches can facilitate machines to make logical, probability-based decisions and undertake thoughtful tasks.

[0174] 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 analyses 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 system does not always know what it is looking for, but can independently classify data and detect compelling patterns.

[0175] Advanced machine learning is at the forefront of the fight against automated and human-driven cyber-threats, overcoming the limitations of rules and signature-based approaches:

[0176] The machine learning learns what is normal within a network—it does not depend upon knowledge of previous attacks.

[0177] The machine learning thrives on the scale, complexity and diversity of modern businesses, where every device and person is slightly different.

[0178] The machine learning turns the innovation of attackers against them—any unusual activity is visible.

[0179] The machine learning constantly revisits assumptions about behavior, using probabilistic mathematics.

[0180] The machine learning is always up to date and not reliant on human input. Utilizing machine learning in cyber security technology is difficult, but when correctly implemented it is extremely powerful. The machine learning means that previously unidentified threats can be detected, even when their manifestations fail to trigger any rule set or signature. Instead, machine learning allows the system to analyze large sets of data and learn a 'pattern of life' for what it sees

[0181] Machine learning can approximate some human capabilities to machines, such as:

[0182] Thought: it uses past information and insights to form its judgments;

[0183] Real time: the system processes information as it goes; and

[0184] Self-improving: the model's machine learning understanding is constantly being challenged and adapted, based on new information

[0185] Unsupervised learning works things out without pre-defined labels. In the case of sorting the series of different 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. The system does not always know what it is looking for, but can independently classify data and detect compelling patterns.

[0186] 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.

[0187] The cyber threat defense system uses unique implementations of unsupervised machine learning algorithms to analyze network data at scale, intelligently handle the unexpected, and embrace uncertainty. Instead of relying on knowledge of past threats to be able to know what to look for, it is able to independently classify data and detect compelling patterns that define what may be considered to be normal behavior. Any new behaviors that deviate from those, which constitute this notion of 'normality,' may indicate threat or compromise. The impact of the cyber threat defense system's unsupervised machine learning on cyber security is transformative.

[0188] Threats from within, which would otherwise go undetected, can be spotted, highlighted, contextually prioritized and isolated using these algorithms.

[0189] The application of machine learning has the potential to provide total network visibility and far greater detection levels, ensuring that networks have an internal defense mechanism.

[0190] Machine learning has the capability to learn when to action automatic responses against the most serious cyber threats, disrupting in progress attacks before they become a crisis for the organization.

[0191] This new mathematics not only identifies meaningful relationships within data, but also quantifies the uncertainty associated with such inference. By knowing and understanding this uncertainty, it becomes possible to bring together many results within a consistent framework—the basis of Bayesian probabilistic analysis. The mathematics behind machine learning is extremely complex and difficult to get right. Robust, dependable algorithms are developed, with a scalability that enables their successful application to real-world environments.

[0192] 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 behaviour 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.

[0193] The cyber security appliance **100** can use a Recursive Bayesian Estimation. To combine these multiple analyses of different measures of network behaviour 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.

[0194] 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 behaviours where conventional signature-based methods fall down.

[0195] Training a model can be accomplished by having the model learn good values for all of the weights and the

bias for labelled 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.

[0196] An anomaly detection technique that can be used is supervised anomaly detection that requires a data set that has been labelled 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 unlabelled 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 behaviour 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.

[0197] In an embodiment, a closer look at the cyber threat defense system's machine learning algorithms and approaches is as follows.

[0198] The cyber threat defense system's probabilistic approach to cyber security is based on a Bayesian framework. This allows it to integrate a huge number of weak indicators of potentially anomalous network behavior to produce a single clear measure of 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.

[0199] Ranking Threats

[0200] Crucially, the cyber threat defense system's approach accounts for the inevitable ambiguities that exist in data, and distinguishes between the subtly differing levels of evidence that different pieces of data may contain. Instead of generating the simple binary outputs 'malicious' or 'benign,' the cyber threat defense system's mathematical algorithms produce outputs that indicate differing degrees of potential compromise. This output enables users of the system to rank different alerts in a rigorous manner and prioritize those that most urgently require action, simultaneously removing the problem of numerous false positives associated with a rule-based approach.

[0201] At its core, the cyber threat defense system mathematically characterizes what constitutes 'normal' behavior based on the analysis of a large number/set of different measures of a devices network behavior, examples include:

- **[0202]** Server access;
- **[0203]** Data access;
- **[0204]** Timings of events;
- **[0205]** Credential use;
- **[0206]** DNS requests; and
- **[0207]** other similar parameters.

[0208] Each measure of network behavior is then monitored in real time to detect anomalous behaviors.

[0209] Clustering

[0210] To be able to properly model what should be considered as normal for a device, its behavior must be analyzed in the context of other similar devices on the network. To accomplish this, the cyber threat defense system leverages the power of unsupervised learning to algorithmically identify naturally occurring groupings of devices, a task which is impossible to do manually on even modestly sized networks.

[0211] In order to achieve as holistic a view of the relationships within the network as possible, the cyber threat defense system simultaneously employs a number of different clustering methods including matrix based clustering, density based clustering and hierarchical clustering techniques. The resulting clusters are then used to inform the modeling of the normative behaviors of individual devices.

[0212] Clustering: At a glance:

- **[0213]** Analyzes behavior in the context of other similar devices on the network;
- **[0214]** Algorithms identify naturally occurring groupings of devices—impossible to do manually; and
- **[0215]** Simultaneously runs a number of different clustering methods to inform the models.

[0216] Network Topology

[0217] Any cyber threat detection system must also recognize that a network is far more than the sum of its individual parts, with much of its meaning contained in the relationships among its different entities, and that complex threats can often induce subtle changes in this network structure. To capture such threats, the cyber threat defense system employs several different mathematical methods in order to be able to model multiple facets of a networks topology.

[0218] One approach is based on iterative matrix methods that reveal important connectivity structures within the network. In tandem with these, the cyber threat defense system has developed innovative applications of models from the field of statistical physics, which allow the modeling of a network's 'energy landscape' to reveal anomalous substructures that may be concealed within.

[0219] Network Structure

[0220] A further important challenge in modeling the behaviors of network devices, as well as of networks themselves, is the high-dimensional structure of the problem with the existence of a huge number of potential predictor variables.

[0221] Observing packet traffic and host activity within an enterprise LAN, WAN and Cloud is difficult because both input and output can contain many inter-related features (protocols, source and destination machines, log changes and rule triggers, etc.). Learning a sparse and consistent structured predictive function is crucial to avoid the curse of over fitting.

[0222] In this context, the cyber threat defense system has employed a cutting edge large-scale computational approach to learn sparse structure in models of network behavior and connectivity based on applying L1-regularization techniques (e.g. a lasso method). This allows for the discovery of true associations between different network components and events that can be cast as efficiently solvable convex optimization problems and yield parsimonious models.

[0223] Recursive Bayesian Estimation

[0224] To combine these multiple analyses of different measures of network behavior to generate a single comprehensive picture of the state of each device, the cyber threat defense system takes advantage of the power of Recursive Bayesian Estimation (RBE) via an implementation of the Bayes filter.

[0225] Using RBE, the cyber threat defense system's mathematical models are able to constantly adapt them-

selves, in a computationally efficient manner, as new information becomes available to the system. They continually recalculate threat levels in the light of new evidence, identifying changing attack behaviors where conventional signature-based methods fall down.

[0226] The cyber threat defense system's innovative approach to cyber security has pioneered the use of Bayesian methods for tracking changing device behaviors and computer network structures. The core of the cyber threat defense system's mathematical modeling is the determination of normative behavior, enabled by a sophisticated software platform that allows for its mathematical models to be applied to new network data in real time. The result is a system that is able to identify subtle variations in machine events within a computer networks behavioral history that may indicate cyber-threat or compromise.

[0227] The cyber threat defense system uses mathematical analysis and machine learning to detect potential threats, allowing the system to stay ahead of evolving risks. The cyber threat defense system approach means that detection no longer depends on an archive of previous attacks. Instead, attacks can be spotted against the background understanding of what represents normality within a network. No predefinitions are needed, which allows for the best possible insight and defense against today's threats. On top of the detection capability, the cyber threat defense system can create digital antibodies automatically, as an immediate response to the most threatening cyber breaches. The cyber threat defense system approach both detects and defends against cyber threat. Genuine unsupervised machine learning eliminates the dependence on signature-based approaches to cyber security, which are not working. The cyber threat defense system's technology can become a vital tool for security teams attempting to understand the scale of their network, observe levels of activity, and detect areas of potential weakness. These no longer need to be manually sought out, but are flagged by the automated system and ranked in terms of their significance.

[0228] Machine learning technology is the fundamental ally in the defense of systems from the hackers and insider threats of today, and in formulating response to unknown methods of cyber-attack. It is a momentous step change in cyber security. Defense must start within.

[0229] An Example Method

[0230] The threat detection system shall now be described in further detail with reference to a flow of the process carried out by the threat detection system for automatic detection of cyber threats through probabilistic change in normal behavior through the application of an unsupervised Bayesian mathematical model to detect behavioral change in computers and computer networks.

[0231] The core threat detection system is termed the 'Bayesian probabilistic'. The Bayesian probabilistic is a Bayesian system of automatically determining periodicity in multiple time series data and identifying changes across single and multiple time series data for the purpose of anomalous behavior detection.

[0232] Human, machine or other activity is modeled by initially ingesting data from a number of sources at step S**1** and deriving second order metrics at step S**2** from that raw data.

[0233] The raw data sources include, but are not limited to: raw network IP traffic captured from an IP or other network TAP or SPAN port; machine generated log files; building access ("swipe card") systems; IP or non IP data flowing over an Industrial Control System (ICS) distributed network; individual machine, peripheral or component power usage; telecommunication signal strength; and/or machine level performance data taken from on-host sources (CPU usage/memory usage/disk usage/disk free space/network usage/etc.)

[0234] From these raw sources of data, a large number of metrics can be derived each producing time series data for the given metric. The data are bucketed into individual time slices (for example, the number observed could be counted per **1** second, per 10 seconds or per 60 seconds), which can be combined at a later stage where required to provide longer range values for any multiple of the chosen internal size. For example, if the underlying time slice chosen is 60 seconds long, and thus each metric time series stores a single value for the metric every 60 seconds, then any new time series data of a fixed multiple of 60 seconds (120 seconds, 180 seconds, 600 seconds etc.) can be computed with no loss of accuracy. Metrics are chosen directly and fed to the Bayesian probabilistic by a lower order model which reflects some unique underlying part of the data, and which can be derived from the raw data with particular domain knowledge. The metrics that are obtained depends on the threats that the system is looking for. In order to provide a secure system, it is common for a large number of metrics relating to a wide range of potential threats to be obtained. Communications from components in the network contacting known suspect domains.

[0235] The actual metrics used are largely irrelevant to the Bayesian probabilistic system, which is described here, but some examples are provided below.

[0236] Metrics derived from network traffic could include data such as: the number of bytes of data entering or leaving a networked device per time interval; file access; the commonality/rarity of a communications process; invalid SSL certification; failed authorization attempt; email access patterns.

[0237] In the case where TCP, UDP or other Transport Layer IP protocols are used over the IP network, and in cases where alternative Internet Layer protocols are used (e.g. ICMP, IGMP), knowledge of the structure of the protocol in use and basic packet header analysis can be utilized to generate further metrics, such as: the number of multicasts per time interval originating from a networked device and intended to reach publicly addressable IP ranges; the number of internal link-local IP Broadcast requests originating from a networked device; the size of the packet payload data; and the number of individual TCP connections made by a device, or data transferred by a device, either as a combined total across all destinations or to any definable target network range, (e.g. a single target machine, or a specific network range).

[0238] In the case of IP traffic, in the case where the Application Layer protocol can be determined and analyzed, further types of time series metric can be defined, for example: the number of DNS requests a networked device generates per time interval, again either to any definable target network range or in total; the number of SMTP, POP or IMAP logins or login failures a machine generates per time interval; the number of LDAP logins or login failures a generated; data transferred via file sharing protocols such as SMB, SMB**2**, FTP, etc; and logins to Microsoft Windows

Active Directory, SSH or Local Logins to Linux or Unix Like systems, or other authenticated systems such as Kerberos.

[0239] The raw data required to obtain these metrics may be collected via a passive fiber or copper connection to the networks internal switch gear, from virtual switching implementations, from cloud based systems, or from communicating devices themselves. Ideally, the system receives a copy of every communications packet to provide full coverage of an organization.

[0240] For other sources, a number of domain specific time series data are derived, each chosen to reflect a distinct and identifiable facet of the underlying source of the data, which in some way reflects the usage or behavior of that system over time.

[0241] Many of these time series data are extremely sparse, and have the vast majority of data points equal to **0**. Examples would be employee's using swipe cards to access a building or part of a building, or user's logging into their workstation, authenticated by Microsoft Windows Active Directory Server, which is typically performed a small number of times per day. Other time series data are much more populated, for example the size of data moving to or from an always-on Web Server, the Web Servers CPU utilization, or the power usage of a photocopier.

[0242] Regardless of the type of data, it is extremely common for such time series data, whether originally produced as the result of explicit human behavior or an automated computer or other system to exhibit periodicity, and have the tendency for various patterns within the data to recur at approximately regular intervals. Furthermore, it is also common for such data to have many distinct but independent regular time periods apparent within the time series.

[0243] At step S**3**, detectors carry out analysis of the second order metrics. Detectors are discrete mathematical models that implement a specific mathematical method against different sets of variables with the target network. For example, HMM may look specifically at the size and transmission time of packets between nodes. The detectors are provided in a hierarchy that is a loosely arranged pyramid of models. Each detector model effectively acts as a filter and passes its output to another model higher up the pyramid. At the top of the pyramid is the Bayesian probabilistic that is the ultimate threat decision making model. Lower order detectors each monitor different global attributes or 'features' of the underlying network and/or computers. These attributes consist of value over time for all internal computational features such as packet velocity and morphology, endpoint file system values, and TCP/IP protocol timing and events. Each detector is specialized to record and make decisions on different environmental factors based on the detectors own internal mathematical model such as an HMM.

[0244] While the threat detection system may be arranged to look for any possible threat, in practice the system may keep watch for one or more specific threats depending on the network in which the threat detection system is being used. For example, the threat detection system provides a way for known features of the network such as desired compliance and Human Resource policies to be encapsulated in explicitly defined heuristics or detectors that can trigger when in concert with set or moving thresholds of probability abnormality coming from the probability determination output. The heuristics are constructed using complex chains of weighted logical expressions manifested as regular expressions with atomic objects that are derived at run time from the output of data measuring/tokenizing detectors and local contextual information. These chains of logical expression are then stored in and/or on online libraries and parsed in real-time against output from the measures/tokenizing detectors. An example policy could take the form of "alert me if any employee subject to HR disciplinary circumstances (contextual information) is accessing sensitive information (heuristic definition) in a manner that is anomalous when compared to previous behavior (Bayesian probabilistic output)". In other words, different arrays of pyramids of detectors are provided for detecting particular types of threats.

[0245] The analysis performed by the detectors on the second order metrics then outputs data in a form suitable for use with the model of normal behavior. As will be seen, the data is in a form suitable for comparing with the model of normal behavior and for updating the model of normal behavior.

[0246] At step S**4**, the threat detection system computes a threat risk parameter indicative of a likelihood of there being a threat using automated adaptive periodicity detection mapped onto observed behavioral pattern-of-life analysis. This deduces that a threat over time exists from a collected set of attributes that themselves have shown deviation from normative collective or individual behavior.

[0247] The automated adaptive periodicity detection uses the period of time the Bayesian probabilistic has computed to be most relevant within the observed network and/or machines. Furthermore, the pattern of life analysis identifies how a human and/or machine behaves over time, i.e. when they typically start and stop work. Since these models are continually adapting themselves automatically, they are inherently harder to defeat than known systems. The threat risk parameter is a probability of there being a threat in certain arrangements. Alternatively, the threat risk parameter is a value representative of there being a threat, which is compared against one or more thresholds indicative of the likelihood of a threat.

[0248] In practice, the step of computing the threat involves comparing current data collected in relation to the user with the model of normal behavior of the user and system being analyzed. The current data collected relates to a period in time, this could be in relation to a certain influx of new data or a specified period of time from a number of seconds to a number of days. In some arrangements, the system is arranged to predict the expected behavior of the system. The expected behavior is then compared with actual behavior in order to determine whether there is a threat.

[0249] The system uses machine learning/Artificial Intelligence to understand what is normal inside a company's network, and when something's not normal. The system then invokes automatic responses to disrupt the cyber-attack until the human team can catch up. This could include interrupting connections, preventing the sending of malicious emails, preventing file access, preventing communications outside of the organization, etc. The approach begins in as surgical and directed way as possible to interrupt the attack without affecting the normal behavior of say a laptop, but if the attack escalates, it may ultimately become necessary to quarantine a device to prevent wider harm to an organization.

[0250] In order to improve the accuracy of the system, a check can be carried out in order to compare current behavior of a user with associated users, i.e. users within a single office. For example, if there is an unexpectedly low level of activity from a user, this may not be due to unusual activity from the user, but could be due to a factor affecting the office as a whole. Various other factors can be taken into account in order to assess whether or not abnormal behavior is actually indicative of a threat.

[0251] Finally, at step **S5** a determination is made, based on the threat risk parameter, as to whether further action need be taken regarding the threat. This determination may be made by a human operator after being presented with a probability of there being a threat, or an algorithm may make the determination, e.g. by comparing the determined probability with a threshold.

[0252] In one arrangement, given the unique global input of the Bayesian probabilistic, a form of threat visualization is provided in which the user can view the threat landscape across all internal traffic and do so without needing to know how their internal network is structured or populated and in such a way as a 'universal' representation is presented in a single pane no matter how large the network. A topology of the network under scrutiny is projected automatically as a graph based on device communication relationships via an interactive **3**D user interface. The projection is able to scale linearly to any node scale without prior seeding or skeletal definition.

[0253] The threat detection system that has been discussed above therefore implements a propriety form of recursive Bayesian estimation to maintain a distribution over the probability state variable. This distribution is built from the complex set of low-level host, network and traffic observations or 'features'. These features are recorded iteratively and processed in real time on the platform. A plausible representation of the relational information among entities in dynamic systems in general, such as an enterprise network, a living cell or a social community, or indeed the entire internet, is a stochastic network, which is topological rewiring and semantically evolving over time. In many high-dimensional structured I/O problems, such as the observation of packet traffic and host activity within a distributed digital enterprise, where both input and output can contain tens of thousands, sometimes even millions of interrelated features (data transport, host-web-client dialogue, log change and rule trigger, etc.), learning a sparse and consistent structured predictive function is challenged by a lack of normal distribution. To overcome this, the threat detection system consists of a data structure that decides on a rolling continuum rather than a stepwise method in which recurring time cycles such as the working day, shift patterns and other routines are dynamically assigned. Thus, providing a non-frequentist architecture for inferring and testing causal links between explanatory variables, observations and feature sets. This permits an efficiently solvable convex optimization problem and yield parsimonious models. In such an arrangement, the threat detection processing may be triggered by the input of new data. Alternatively, the threat detection processing may be triggered by the absence of expected data. In some arrangements, the processing may be triggered by the presence of a particular actionable event.

[0254] Graph Anomaly Analysis

[0255] The cyber threat detection platform is configured to create a graph and map incidents onto that graph to detect anomalies; and thus, potential indicators of a cyber threat. The Graph Anomaly Analysis method uses graph theory (creating directed graphs), where AI Analyst incidents are rendered as edges of the graph and the metadata in those incidents connects nodes of devices to form the rest of the graph. AI Analyst may be a general term used to describe at least some of the apparatus of the present disclosure. That is, AI Analyst may refer to the cyber-security appliance, or at least part of the cyber-security appliance, used to identify and/or response to cyber-attacks in the manner previously described. Note, devices and ports involved are used to connect nodes. Each device may be a physical networking device, such as a laptop or desktop computer, but it does not need to be a physical networking device and may be, for example, a virtual device. Each device may form one of, for example, a client device, a virtualized entity such as a virtual machine (VM) or a container, or a user in a SaaS environment. Alternatives to ports, and in particular, shared ports, may be used to connect nodes. For example, an indication may be provided in the metadata (for example, the same server message block (SMB) credential, or the same external IP seen used in two rare SaaS administrative actions). Significantly, the graph anomaly analysis may be built out of higher level abstractions of incidents and other relations between devices, rather than just low level things like connections, which is likely to produce a lot more noise. Metadata may be more than shared endpoints or edges. The nodes in this case are devices and ports in the network. Joining these nodes with incidents allows for incidents to be grouped into an overview of a compromise, linking affected devices and tracing how an infection or malicious actor spread across the whole network. The Graph Anomaly Analysis breaks down individual events and provides an overall summary (AI analyst can use the direct graph to detect anomalies). The Graph Anomaly Analysis method graphs incidents to create a meta of incidents, which encompass a full compromise or network-wide event. The directed graph also uses time to retrace the path of a meta-incident and locate patient zero. This is highly desirable for an end-user as this linking would have to be done manually by a security team after an incident. Also, the Graph Anomaly Analysis method can be used on Model Breaches.

[0256] FIG. **9** is a schematic diagram illustrating directed graph analysis embodying an aspect of the present invention.

[0257] Analysis using graph theory can be used to investigate anomalies, which may be indicative of a cyber-threat, in further detail to the embodiments described above. In particular, producing a directed graph allows a high level analysis of incidents in a network. An incident may comprise one or more events; events may be a group of anomalies or network actions investigated by the cyber-threat security appliance that pose a likely cyber threat. Using, for example, an arrangement as described above, a particular incident may be identified and analysed at a low level. That incident may, for example, affect a particular entity. By producing a directed graph, an overview of a plurality of incidents may be provided to view multiple incidents affecting a network at a higher level.

[0258] The directed graph of FIG. **9** comprises a plurality of nodes **901**, **905**, **907** each node **901** corresponding to an entity. In this example, the entities may comprise devices, users, and ports, which may be in the form of external

endpoints or network ranges. The graph also comprises an edge **903** corresponding to an incident. The edge **903** joins two of the nodes **905**, **907**.

[0259] In the arrangement, one or more machine learning modules are trained on a normal behavior of entities associated with a network and interactions between those entities. A cyber-threat module has one or more machine learning models trained on cyber-threats in the network. The cyber-threat module is configured to reference the models that are trained on the normal behavior of the entities associated with the network. The cyber-threat module determines a threat risk parameter that factors in the likelihood that a chain of one or more unusual behaviors of the entities under analysis fall outside of derived normal behavior. These unusual behaviors are determined to be one or more incidents **903**. Based on this information, an analyser module generates the directed graph.

[0260] An incident **903** has been identified by the cyber-threat module in the manner described. The incident **903** comprises an entity **905** (a device) connecting to a suspicious or malicious entity **907** (an endpoint). The devices make various other connections to other entities which have been identified by the cyber-threat module. These entities, in turn, have further connections with yet more entities. The connections may include a direct connection (physical or a communication connection), a remote control connection, data transfer, writing of a file, name resolution, authentication, file downloading, or any such connection.

[0261] Incidents **903** may be grouped into a meta of incidents, representing a compromise which links entities in the network affected by the incidents. By providing this link, not only is a high level view of the compromise in the network provided, but also a significant reduction in noise is provided. In other words, when investigating a network for cyber-threats, a plurality of entities may be examined, only some of which are affected by an incident, or are compromised. In arrangements of the present invention, the directed graph produced comprises those entities or nodes **901**, **905**, **907** which are involved in, or connected to, an edge or incident **903**, and therefore are potentially affected by the incident or cyber-attack. In other words, unaffected entities are not included in analysis, and therefore the analysis is more refined and noise is reduced. Analysis in this way beneficially provides insights automatically, without the need to specify an investigation into specific entities; the affected or so-called interesting entities are already highlighted in the directed graph. In some examples, each entity in a network may be affected.

[0262] This analysis advantageously helps enable the identification of the origin of a particular incident. Additionally, the directed graph allows identification of which entities or incidents may be prioritized for analysis. For example, those entities which are the source of one or more incidents spreading in a network.

[0263] The links or edges **903** are based on relevant behavior of entities, and entities are connected based on such behavior. For example, two entities which are not necessarily connected to one another, both connect to a suspicious or malicious endpoint. These two entities may be linked, for example by an edge **903** in a directed graph because of this behavior. This is more efficient than linking or connecting entities purely based on their physical or network communication connections; entities connected to one another in these ways may not all be affected by incidents. The range of entities to be analysed in more detail is therefore substantially narrowed by connecting entities by their behavior, rather than their physical or communication connections. This is illustrated in FIG. **10** described below.

[0264] In addition, by connecting entities by their behavior, an incident or cyber-attack may be described most effectively. In doing this, an accurate response to the cyber-attack may be provided more efficiently. A behavioral chain detection is capable of identifying more subtle incidents or cyber-attacks, which may not be identified if, for example, a human aims to connect entities by their connections.

[0265] A visual representation of the directed graph may be provided to a user. This may be provided via the user interface. This provides the user with a convenient overview of a compromise as explained above. The user may then determine which entity or entities may be the source of an incident, and which entity or entities need further investigation.

[0266] FIG. **10** illustrates a similar directed graph to that illustrated in FIG. **9**, but at yet a higher level. In addition, FIG. **10** illustrates a directed graph including a plurality of incidents **903**, and shows an overview of a compromise in a network.

[0267] The cyber-threat module identifies a plurality of incidents **903**, involving entities **1001** connecting to a malicious or suspicious entity **1003**. The incidents **903** represented as edges of the directed graph, connect entities **1001** involved in the incidents **903**. These incidents are grouped into a meta of incidents representing a compromise on the network. The directed graph provides a useful overview of the compromise, showing which entities **1001** have been affected, and connections involving those entities. That is, the edges **903** represent the incidents and illustrate the entities involved, but the graph analysis also shows further connections between the affected entities **1001** and additional entities **901**. In this way, an overview of the compromise may be viewed, and affected and potentially affected entities can be analysed. Significantly, there is no direct connection between affected entities **1001**. In existing cyber-security systems, a connection between these entities would not be made. In arrangements of the present invention, a behavioral connection between the entities is highlighted, and analysed using the directed graph.

[0268] The visual representation of the directed graph illustrated in FIG. **10** may be provided to a user, for example via the user interface.

[0269] As explained above, in providing the overview of the compromise, further analysis may be carried out on affected entities **1001**, and an appropriate response to the cyber-attack may be efficiently implemented.

[0270] For example, the behavioral links between affected devices provides an effective description of the incident **903** or cyber-attack. Based on this description, an autonomous response module may, rather than a human taking an action, cause one or more autonomous actions to be taken to contain the cyber-threat when the threat risk parameter from the cyber-threat module is equal to or above an actionable threshold. The autonomous response module is configurable to know when the response module should take the autonomous actions to contain the cyber-threat when incidents are determined by the cyber-threat module. The autonomous response module has an administrative tool, configurable through the user interface, to set what autonomous actions the autonomous response module can take, including types

of actions and specific actions the autonomous response module is capable of, when the cyber-threat module indicates the threat risk parameter is equal to or above the actionable threshold. The autonomous response module has a library of response actions types of actions and specific actions the autonomous response module is capable of, including focused response actions selectable through the user interface that are contextualized to autonomously act on specific incidents of the one or more incidents.

[0271] FIG. **11** illustrates another representation of a directed graph similar to that illustrated in FIGS. **9** and **10**. Incidents **903** have been identified involving affected entities **1001** connecting to a suspicious or malicious entity **1003**, by the cyber-threat module. Entities **1001** are linked due to their behavior and their involvement in the incident **903**, in particular, due to their connection to the entity **1003**. Also shown are connections between other entities **901** which have connections with the affected entities **1001** involved in the incident. Not only does this provide an overview of the compromise, but also provides for the identification of other potentially affected entities **901** if this is desired.

[0272] An external interface for AI Analyst Triggered Investigations in an Enterprise Immune System

[0273] An Artificial Intelligence based Analyst may conduct an autonomous investigation of anomalous events/etc., assess any potential cyber threats, and generate a report indicating its findings. AI Analyst investigations of anomalous events/etc. may be based upon internal anomaly triggers from different parts of the cyber security appliance also housing the AI Analyst. With an external interface, an AI Analyst triggered investigation also maybe triggered on demand from a Third-Party Threat Intelligence component and/or manually initiated by a user. Multiple methods are proposed to trigger analysis manually (by user or by input from a third party Threat Intelligence component) so that it can aid an investigation or integrate with third-party tools. Some example methods include a manual trigger icon or menu option within a user interface, a trigger icon or menu option in the email component of the cyber threat defense system, an API trigger/control signal via third party inputs such as Syslog or other to manually trigger investigations which rely on third-party data. The manual request can specify a request for, for example, the specific period of time and/or particular environment e-mail, SaaS, etc

[0274] The AI Analyst can then trigger cyber threat investigations based upon internal anomaly alerts—the component would investigate potential anomalous behaviour and choose to investigate. In addition, investigations by AI Analyst can be manually triggered (via a user internally pushing an icon or selecting the command in the menu or externally receiving a signal by a third party) via one or more of the following methods:

[0275] a) A user interacts with a device which is behaving strangely, but has not triggered a full anomaly event. The user requests via—e.g. click a button to request—that AI analyst autonomously start to investigate potential cyber threats by evaluating anomalies at the current time the manual request is detected.

[0276] b) A user requests an AI analyst autonomous investigation on a given device during a given timeframe, potentially in response to later indications of compromise.

[0277] c) A user on another system (e.g. a cyber threat protection system for email, a console of a SaaS cyber threat protection system, etc.) manually requests an investigation of a network device known to be associated with a credential as the credential is showing signs of compromise.

[0278] d) A user of a cyber threat protection system for email manually requests the AI analyst autonomous investigation of any associated SaaS accounts as an email address is indicating signs of compromise.

[0279] e) An operator configures third-party alerts (such as those from a SIEM or a firewall) to be ingested by the cyber security appliance, where these alerts can trigger an AI Analyst autonomous investigation of a device based on their contents. Most likely implemented with rate limiting.

[0280] f) An operator configures third-party alerts which contain additional data which is not previously seen by the cyber security appliance. AI Analyst is triggered to investigate and uses the data from the third-party alert to perform the investigation, whether partially or fully.

[0281] g) An operator configures the cyber security appliance to have access to their third-party threat intelligence systems. The operator triggers an investigation on a specific device and requests that the third-party intelligence be used during the investigation.

[0282] In scenarios a, b, c, and g, AI Analyst will have to apply all hypotheses. In scenario d, it can apply just SaaS-based hypotheses which has less computational requirements. In scenarios e and f, the type of anomalous activity would be indicated by the third-party and then only related hypotheses would be applied.

[0283] Furthermore, information supplied in method g) could be used for standard investigation.

[0284] Embodiments described above comprise an investigation being triggered by an alert raised by the system to some suspicious or unusual behavior of one or more entities on a network. These lower level anomaly alerts trigger investigation similar to how a human may, albeit in a far more efficient way. That is, when something suspicious or unusual is noticed, an investigation into a suspicious or unusual entity takes place. In addition, or alternatively, investigation into one or more entities may be triggered by an interface receiving a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation, which may be used to supplement existing hypotheses relating to a potential cyber-threat based on the behavior of one or more entities, and/or to produce new hypotheses. This may be particularly advantageous as investigation into one or more entities may be triggered when there is not enough unusual behavior falling outside of the normal pattern of life for an investigation to be automatically triggered, so investigation would not otherwise take place. Without the signal from the external apparatus triggering the investigation, only entities which have already been alerted as being suspicious or unusual would be investigated. By using such triggers, the system is not only an alert generator, but is capable of investigating any entity, regardless of its behavior. Such arrangements provide a user friendly way of identifying all compromised entities quickly. This is far more efficient than a human analyst who may analyse un-useful data.

[0285] FIG. **12** is a flow diagram illustrating a method for a cyber-security defense system. The method may be considered a method for supplementing hypotheses for cyber-

attack detection. The further investigation triggered by the interface receiving the signal from an external apparatus provides further information which may form additional hypotheses, or supplement existing hypotheses.

[0286] At step **1201**, one or machine learning models trained on a normal behavior of entities associated with a network and interaction between the entities are used. The entities may be at least one of users, ports, and devices.

[0287] At step **1203**, an interface receives a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation. In this example, the signal from the external apparatus is provided by a manual user input via the interface. The interface may comprise a screen configured to enable user interaction, and in this example, the user selects an entity to be investigated on the screen. In addition, the manual user input may comprise specific details about the investigation to be carried out. This may non-exhaustively include: selection of one or more entities of a plurality of entities to be investigated; a period of time over which to carry out the investigation (such as investigating the behavior of a particular entity over the last day, or **7** days, for example); and a particular investigation environment to be investigated. An environment may comprise, for example, e-mail, SaaS, or any appropriate environment. The details may additionally or alternatively include any suitable details regarding an investigation.

[0288] It will be appreciated that, whilst the signal from the external apparatus may be provided by manual user input, the trigger input may additionally or alternatively comprise an input from a third party threat intelligence component. The third party threat intelligence component may be any appropriate third party capable of triggering an investigation, such as third party software or a third party system. The third party intelligence component may provide additional data relating to behavior of the one or more entities. This data may be provided to the interface. This may aid an investigation into the one or more entities, making the determination of, and response to, a cyber-threat more efficient. Further hypotheses may be determined based on the additional data provided by the third party system. This may be particularly advantageous, for example, when an entity is temporarily disconnected from a network running the cyber-threat security appliance. For example, an employee may take home a computing device, disconnecting it from a network. Whilst disconnected, the computing device (or entity) may be subject to, or perform, suspicious of unusual behavior. When the computing device is reconnected to the network, third party software running on the computing device may provide the signal to the interface requesting and triggering an investigation to be carried out on the computing device. The third party software may indicate to the cyber-threat security system that unusual behavior has taken place, and provide data comprising details of the behavior of the entity to the interface, thereby aiding the investigation of the entity. Such unusual behavior may not have been found without the trigger input being provided, and a cyber-threat may have been missed as the device was disconnected from the network.

[0289] The interface may request further information from the third party intelligence component that provided the signal, or any third party intelligence component. The further information may relate to further unusual behavior, and trigger yet further investigation into one or more entities.

[0290] At step **1205**, the interface works with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation. The investigation may be conducted in the manner of any embodiments described above. A cyber-threat module may comprise one or more of the interface, the artificial intelligence models, and the scripts. The cyber-threat security appliance, or the Cyber AI Analyst, may use a model breach as a starting point to investigate the device. The behavioral analysis it performs may discover anomalies or patterns of activity that were not the original trigger point for the model breach but are worthy of investigation. Consequently, the event period may not correspond with the breach time. Additionally, some model breaches require sustained behaviors such as repeated connections before breaching, so the final breach trigger may be later than the connection of interest. The interface or cyber-threat module contextualizes the behavior of the one or more entities being investigated against historic activity and connectivity for the entity and its peers. If a finished investigation has the result "No Incident Found", the cyber-threat security appliance, or AI Analyst did not locate any identifiable anomalous activity around the investigation time. If anomalous activity is detected, one or more incident events will be created.

[0291] At step **1207**, the interface, working with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, determines whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat. This determination may be performed in the manner of any of the embodiments described above. For example, the interface, or a cyber-threat module, may use one or more machine learning models trained on cyber-threats in the network, and reference the one or more machine learning models that are trained on the normal behavior of the entities in the network, in order to establish what is unusual behavior falling outside of normal behavior, or what is considered an abnormality.

[0292] Behavior identified as abnormal may be determined to be one or more incidents. In this case, a notification may be provided to a user that one or more incidents have been found. This notification may be provided to the user through the interface, for example, on the screen. The user may then have the opportunity to trigger further investigation into the one or more entities involved in the one or more incidents. In addition or alternatively, investigation may be triggered to investigate entities connected to entities involved in the incident, but not involved in the incident themselves. This may aid in preventing the spread of an incident.

[0293] In addition, the one or more incidents may be determined as an actual cyber-threat. In this case, at step **1209**, an autonomous response is triggered from an autonomous response module in order to mitigate the cyber threat. The autonomous response module may provide an autonomous response in accordance with any of the embodiments as set out above.

[0294] In response to the determination of one or more incidents, the method may further comprise generating a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially

indicative of cyber-threats. This may be done by an analyser module, and may be done in the manner of any of the embodiments described above.

[0295] An apparatus such as a computer may be configured in accordance with such code to perform one or more processes in accordance with the various methods discussed herein.

[0296] Web Site

[0297] A web site is configured as a browser-based tool or direct cooperating app tool for configuring, analyzing, and communicating with the cyber threat defense system.

[0298] Network

[0299] A number of electronic systems and devices can communicate with each other in a network environment. The network environment has a communications network. The network can include one or more networks selected from an optical network, a cellular network, the Internet, a Local Area Network ("LAN"), a Wide Area Network ("WAN"), a satellite network, a 3rd party 'cloud' environment; a fiber network, a cable network, and combinations thereof. In some embodiments, the communications network is the Internet. There may be many server computing systems and many client computing systems connected to each other via the communications network.

[0300] The communications network can connect one or more server computing systems selected from at least a first server computing system and a second server computing system to each other and to at least one or more client computing systems as well. The server computing systems can each optionally include organized data structures such as databases. Each of the one or more server computing systems can have one or more virtual server computing systems, and multiple virtual server computing systems can be implemented by design. Each of the one or more server computing systems can have one or more firewalls and similar defenses to protect data integrity.

[0301] At least one or more client computing systems for example, a mobile computing device (e.g., smartphone with an Android-based operating system can communicate with the server(s). The client computing system can include, for example, the software application or the hardware-based system in which the client computing system may be able to exchange communications with the first electric personal transport vehicle, and/or the second electric personal transport vehicle. Each of the one or more client computing systems can have one or more firewalls and similar defenses to protect data integrity.

[0302] A cloud provider platform may include one or more of the server computing systems. A cloud provider can install and operate application software in a cloud (e.g., the network such as the Internet) and cloud users can access the application software from one or more of the client computing systems. Generally, cloud users that have a cloud-based site in the cloud cannot solely manage a cloud infrastructure or platform where the application software runs. Thus, the server computing systems and organized data structures thereof can be shared resources, where each cloud user is given a certain amount of dedicated use of the shared resources. Each cloud user's cloud-based site can be given a virtual amount of dedicated space and bandwidth in the cloud. Cloud applications can be different from other applications in their scalability, which can be achieved by cloning tasks onto multiple virtual machines at run-time to meet changing work demand. Load balancers distribute the work over the set of virtual machines. This process is transparent to the cloud user, who sees only a single access point.

[0303] Cloud-based remote access can be coded to utilize a protocol, such as Hypertext Transfer Protocol ("HTTP"), to engage in a request and response cycle with an application on a client computing system such as a web-browser application resident on the client computing system. The cloud-based remote access can be accessed by a smartphone, a desktop computer, a tablet, or any other client computing systems, anytime and/or anywhere. The cloud-based remote access is coded to engage in 1) the request and response cycle from all web browser based applications, 3) the request and response cycle from a dedicated on-line server, 4) the request and response cycle directly between a native application resident on a client device and the cloud-based remote access to another client computing system, and 5) combinations of these.

[0304] In an embodiment, the server computing system can include a server engine, a web page management component, a content management component, and a database management component. The server engine can perform basic processing and operating-system level tasks. The web page management component can handle creation and display or routing of web pages or screens associated with receiving and providing digital content and digital advertisements. Users (e.g., cloud users) can access one or more of the server computing systems by means of a Uniform Resource Locator ("URL") associated therewith. The content management component can handle most of the functions in the embodiments described herein. The database management component can include storage and retrieval tasks with respect to the database, queries to the database, and storage of data.

[0305] In some embodiments, a server computing system can be configured to display information in a window, a web page, or the like. An application including any program modules, applications, services, processes, and other similar software executable when executed on, for example, the server computing system, can cause the server computing system to display windows and user interface screens in a portion of a display screen space. With respect to a web page, for example, a user via a browser on the client computing system can interact with the web page, and then supply input to the query/fields and/or service presented by the user interface screens. The web page can be served by a web server, for example, the server computing system, on any Hypertext Markup Language ("HTML") or Wireless Access Protocol ("WAP") enabled client computing system (e.g., the client computing system **802**B) or any equivalent thereof. The client computing system can host a browser and/or a specific application to interact with the server computing system. Each application has a code scripted to perform the functions that the software component is coded to carry out such as presenting fields to take details of desired information. Algorithms, routines, and engines within, for example, the server computing system can take the information from the presenting fields and put that information into an appropriate storage medium such as a database (e.g., database). A comparison wizard can be scripted to refer to a database and make use of such data. The applications may be hosted on, for example, the server computing system and served to the specific application or

browser of, for example, the client computing system. The applications then serve windows or pages that allow entry of details.

[0306] Computing Systems

[0307] 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. The system bus may be any of several types of bus structures selected from a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures.

[0308] The computing system typically includes a variety of computing machine-readable media. Computing machine-readable media can be any available media that can be accessed by computing system 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 **900**. Transitory media, such as wireless channels, are not included in the machine-readable media. Communication media typically embody computer readable instructions, data structures, other executable software, or other transport mechanism and includes any information delivery media.

[0309] The system memory includes computer storage media in the form of volatile and/or nonvolatile memory such as read only memory (ROM) and random access memory (RAM). A basic input/output system (BIOS) containing the basic routines that help to transfer information between elements within the computing system, such as during start-up, is typically stored in ROM. RAM typically contains data and/or software that are immediately accessible to and/or presently being operated on by the processing unit. By way of example, and not limitation, the RAM can include a portion of the operating system, application programs, other executable software, and program data.

[0310] The drives and their associated computer storage media discussed above, provide storage of computer readable instructions, data structures, other executable software and other data for the computing system.

[0311] A user may enter commands and information into the computing system through input devices such as a keyboard, touchscreen, or software or hardware input buttons, a microphone, a pointing device and/or scrolling input component, such as a mouse, trackball or touch pad. The microphone can cooperate with speech recognition software. These and other input devices are often connected to the processing unit through a user input interface that is coupled to the system bus, but can be connected by other interface and bus structures, such as a parallel port, game port, or a universal serial bus (USB). A display monitor or other type of display screen device is also connected to the system bus via an interface, such as a display interface. In addition to the monitor, computing devices may also include other peripheral output devices such as speakers, a vibrator, lights, and other output devices, which may be connected through an output peripheral interface.

[0312] It should be noted that the present design can be carried out on a single computing system and/or on a distributed system in which different portions of the present design are carried out on different parts of the distributed computing system.

[0313] Note, an application described herein includes but is not limited to software applications, mobile apps, and programs 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++, or other similar languages. Also, an algorithm can be implemented with lines of code in software, configured logic gates in software, 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.

[0314] It should be borne in mind, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. 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.

[0315] Many functions performed by electronic hardware components can be duplicated by software emulation. Thus, a software program written to accomplish those same functions can emulate the functionality of the hardware components in input-output circuitry.

[0316] 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.

[0317] The method, apparatus and system are arranged to be performed by one or more processing components with any portions of software stored in an executable format on a computer readable medium. Thus, any portions of the method, apparatus and system implemented as software can be stored in one or more non-transitory memory storage devices in an executable format to be executed by one or more processors. The computer readable medium may be non-transitory and does not include radio or other carrier waves. The computer readable medium could be, for example, a physical computer readable medium such as semiconductor memory or solid state memory, magnetic tape, a removable computer diskette, a random access memory (RAM), a read-only memory (ROM), a rigid magnetic disc, and an optical disk, such as a CD-ROM, CD-R/W or DVD.

[0318] The various methods described above may 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.

[0319] Note, an application described herein includes but is not limited to software applications, mobile applications, and programs 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 C, C++, HTTP, Java, or other similar languages. Also, an algorithm can be implemented with lines of code in software, configured logic gates in software, 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.

[0320] Generally, an application includes programs, routines, objects, widgets, plug-ins, and other similar structures that perform particular tasks or implement particular abstract data types. Those skilled in the art can implement the description and/or figures herein as computer-executable instructions, which can be embodied on any form of computing machine-readable media discussed herein.

[0321] Many functions performed by electronic hardware components can be duplicated by software emulation. Thus, a software program written to accomplish those same functions can emulate the functionality of the hardware components in input-output circuitry.

[0322] 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.

[0323] Embodiments of the invention may be described by the following numbered clauses.

[0324] CLAUSES:

[0325] 1. An apparatus comprising:

[0326] one or more machine learning modules that are trained on a normal behavior of entities associated with a network and interactions between the entities;

[0327] an interface configured to receive a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation, where the interface is configured to work with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, in order to determine whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat, and thus, trigger an autonomous response from an autonomous response module to mitigate the cyber-threat.

[0328] 2. An apparatus according to clause 1 wherein the signal from the external apparatus is provided by a manual user input.

[0329] 3. An apparatus according to clause 2 wherein the manual user input comprises investigation instructions comprising one or more of: a time period to be investigated, designation of the one or more entities to be investigated, and a particular investigation environment.

[0330] 4. An apparatus according to any of clauses 1 to 3 wherein the signal from the external apparatus is provided by a third party threat intelligence component.

[0331] 5. An apparatus according to clause 4 wherein the third party threat intelligence component provides additional data relating to behavior of the one or more entities.

[0332] 6. An apparatus according to any of clauses 1 to 5 further comprising:

[0333] an analyser module configured to, in response to the determination of one or more incidents, generate a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially indicative of cyber threats.

[0334] 7. An apparatus according to clause 6 wherein the directed graph comprises a plurality of nodes, each node of the plurality of nodes corresponding to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

[0335] 8. An apparatus according to clause 6 or clause 7, wherein the analyser module is further configured to group the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

[0336] 9. An apparatus according to any of clauses 6 to 8 wherein the directed graph is configured to be used to investigate the compromise at a higher level than each incident of the one or more incidents by providing an overview of the plurality of nodes affected by the one or more incidents.

[0337] 10. An apparatus according to any of clauses 6 to 9 further comprising a formatting module configured to generate a visual representation of the directed graph for display.

[0338] 11. An apparatus according to any of clauses 1 to 10, wherein the autonomous response module is configurable to know when the response module should take the autonomous actions to mitigate the cyber-threat when one or more incidents are worth of being determined as a cyber-threat, where the autonomous response module has an administrative tool, configurable through the interface, to set what autonomous actions the autonomous response module can take, including types of actions and specific actions the autonomous response module is capable of.

[0339] 12. An apparatus according to any of clauses 1 to 11 wherein the autonomous response module has a library of response actions types of actions and specific actions the autonomous response module is capable of, including focused response actions selectable through the user interface that are contextualized to autonomously act on specific incidents of the one or more incidents.

[0340] 13. A method for a cyber-threat defense system, the method comprising:

[0341] using one or more machine learning models that are trained on a normal behavior of entities associated with a network and interactions between the entities;

[0342] receiving, at an interface, a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation;

[0343] the interface working with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, and determining whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat; and

[0344] triggering an autonomous response from an autonomous response module to mitigate the cyber-threat.

[0345] 14. A method according to clause 13 wherein the signal from the external apparatus is provided by a manual user input at the interface.

[0346] 15. A method according to clause 14 wherein the manual user input comprises investigation instructions comprising one or more of: a time period to be investigated, designation of the one or more entities to be investigated, and a particular investigation environment.

[0347] 16. A method according to any of clauses 13 to 15 wherein the signal from the external apparatus is provided by a third party threat intelligence component.

[0348] 17. A method according to clause 16 wherein the third party threat intelligence component provides additional data relating to behavior of the one or more entities.

[0349] 18. A method according to any of clauses 13 to 17 further comprising, in response to the determination of one or more incidents, generating a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially indicative of cyber-threats.

[0350] 19. A method according to clause 18 wherein the directed graph comprises a plurality of nodes, each node of the plurality of nodes corresponding to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

[0351] 20. A method according to clause 18 or clause 19 further comprising grouping the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

[0352] 21. A method according to any of clauses 18 to 20 further comprising using the directed graph to investigate the compromise at a higher level than each individual incident of the one or more incidents by providing an overview of the plurality of nodes affected by the one or more incidents.

[0353] 22. A method according to any of clauses 18 to 21 further comprising generating a visual representation of the directed graph.

[0354] 23. A method according to any of clauses 13 to 22 further comprising:

[0355] making the autonomous response module configurable to know when the response module should take the autonomous actions to mitigate the cyber-threat; and

[0356] taking the autonomous actions to mitigate the cyber-threat when one or more incidents are worth of being determined as a cyber-threat, wherein the autonomous response module has an administrative tool, configurable through the interface, to set what autonomous actions the autonomous response module can take, including types of actions and specific actions the autonomous response module is capable of.

[0357] 24. A method according to any of clauses 13 to 23 further comprising using a library of response actions types of actions and specific actions the autonomous response module is capable of, including focused response actions selectable through the user interface that are contextualized to autonomously act on specific incidents of the one or more incidents.

[0358] 25. A non-transitory computer-readable medium including executable instructions that, when executed with one or more processors, cause a cyber-threat defense system to perform the method of any of clauses **13** to **24**.

[0359] 26. An apparatus, comprising:

[0360] one or more machine learning modules that are trained on a normal behavior of entities associated with a network and interactions between the entities;

[0361] a cyber-threat module with one or more machine learning models trained on cyber threats in the network, where the cyber-threat module is configured to reference the models that are trained on the normal behavior of the entities associated with the network, where the cyber-threat module determines a threat risk parameter that factors in the likelihood that a chain of one or more unusual behaviors of the

entities under analysis fall outside of derived normal benign behavior, and determines this behavior as one or more incidents; and

[0362] an analyser module configured to generate a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially indicative of cyber-threats.

[0363] 27. An apparatus according to clause 26 wherein the directed graph comprises a plurality of nodes, each node of the plurality of nodes corresponding to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

[0364] 28. An apparatus according to clause 26 or clause 27 wherein the analyser module is further configured to group the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

[0365] 29. An apparatus according to any of clauses 26 to 28 wherein the directed graph is configured to be used to investigate the compromise at a higher level than each incident of the one or more incidents by providing an overview of the plurality of nodes affected by the one or more incidents.

[0366] 30. An apparatus according to any of clauses 26 to 29 further comprising a formatting module configured to generate a visual representation of the directed graph.

[0367] 31. An apparatus according to any of clauses 26 to 30 further comprising an autonomous response module configured to, rather than a human taking an action, cause one or more autonomous actions to be taken to contain the cyber-threat when the threat risk parameter from the cyber-threat module is equal to or above an actionable threshold.

[0368] 32. An apparatus according to any of clauses 26 to 31 wherein the autonomous response module is configurable to know when the response module should take the autonomous actions to contain the cyber-threat when one or more incidents are determined by the cyber-threat module, where the autonomous response module has an administrative tool, configurable through a user interface, to set what autonomous actions the autonomous response module can take, including types of actions and specific actions the autonomous response module is capable of, when the cyber-threat module indicates the threat risk parameter is equal to or above the actionable threshold.

[0369] 33. An apparatus according to clause 32 wherein the autonomous response module has a library of response actions types of actions and specific actions the autonomous response module is capable of, including focused response actions selectable through the user interface that are contextualized to autonomously act on specific incidents of the one or more incidents.

[0370] 34. A method for a cyber-threat defense system, the method comprising:

[0371] using one or more machine learning models that are trained on a normal behavior of entities associated with a network and interactions between the entities, the entities including at least one of users, ports, and devices;

[0372] using a cyber-threat module with one or more machine learning models trained on cyber threats in the network, where the cyber-threat module is configured to reference the models that are trained on the normal behavior of the entities associated with the network;

[0373] determining a threat risk parameter that factors in the likelihood that a chain of one or more unusual behaviors of the entities under analysis falls outside of a derived normal benign behavior and determining this behavior as one or more incidents; and

[0374] generating a directed graph, using graph theory, mapping one or more incidents onto the graph to detect anomalies potentially indicative of cyber-threats.

[0375] 35. A method according to clause 34 wherein generating the directed graph comprises generating a plurality of nodes, each node of the plurality of nodes corresponding to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

[0376] 36. A method according to clause 34 or clause 35 further comprising grouping the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

[0377] 37. A method according to any of clauses 34 to 36 further comprising using the directed graph to investigate the compromise at a higher level than each individual incident of the one or more incidents by providing an overview of the plurality of nodes affected by the one or more incidents.

[0378] 38. A method according to any of clauses 34 to 37 further comprising generating a visual representation of the directed graph.

[0379] 39. A method according to any of clauses 34 to 38 further comprising using an autonomous response module, rather than a human taking an action, to cause one or more autonomous actions to be taken to contain the cyber-threat when the threat risk parameter from the cyber-threat module is equal to or above an actionable threshold.

[0380] 40. A method according to clause 39 further comprising:

[0381] making the autonomous response module configurable to know when the response module should take the autonomous actions to contain the cyber-threat; and

[0382] taking the autonomous actions to contain the cyber-threat when the when one or more incidents are determined by the cyber-threat module, wherein the autonomous response module has an administrative tool, configurable through a user interface, to set what autonomous actions the autonomous response module can take, including types of actions and specific actions the autonomous response module is capable of, when the cyber-threat module indicates the threat risk parameter is equal to or above the actionable threshold.

[0383] 41. A method according to clause 39 or clause 40 further comprising using a library of response actions types of actions and specific actions the autonomous response module is capable of, including focused response actions selectable through the user interface that are contextualized to autonomously act on specific incidents of the one or more incidents.

[0384] 42. A non-transitory computer-readable medium including executable instructions that, when executed with one or more processors, cause the cyber-threat defense system to perform the method of any of clauses 34 to 41.

1. An apparatus comprising:
one or more machine learning modules that are trained on a normal behavior of entities associated with a network and interactions between the entities;
an interface configured to receive a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation, where the interface is configured to work with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, in order to determine whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat, and thus, trigger an autonomous response from an autonomous response module to mitigate the cyber-threat.

2. An apparatus according to claim 1 wherein the signal from the external apparatus is provided by a manual user input.

3. An apparatus according to claim 2 wherein the manual user input comprises investigation instructions comprising one or more of: a time period to be investigated, designation of the one or more entities to be investigated, and a particular investigation environment.

4. An apparatus according to claim 1 wherein the signal from the external apparatus is provided by a third party threat intelligence component.

5. An apparatus according to claim 4 wherein the third party threat intelligence component provides additional data relating to behavior of the one or more entities.

6. An apparatus according to claim 1 further comprising:
an analyser module configured to, in response to the determination of one or more incidents, generate a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially indicative of cyber threats.

7. An apparatus according to claim 6 wherein the directed graph comprises a plurality of nodes, each node of the plurality of nodes corresponding to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

8. An apparatus according to claim 6, wherein the analyser module is further configured to group the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

9. An apparatus according to claim 6 further comprising a formatting module configured to generate a visual representation of the directed graph for display.

10. An apparatus according to claim 1, wherein the autonomous response module is configurable to know when the response module should take the autonomous actions to mitigate the cyber-threat when one or more incidents are worth of being determined as a cyber-threat, where the autonomous response module has an administrative tool, configurable through the interface, to set what autonomous actions the autonomous response module can take, including types of actions and specific actions the autonomous response module is capable of.

11. A method for a cyber-threat defense system, the method comprising:
using one or more machine learning models that are trained on a normal behavior of entities associated with a network and interactions between the entities;
receiving, at an interface, a signal from an external apparatus to request and trigger an artificial intelligence based analyst investigation;
the interface working with at least one of: artificial intelligence models trained on how to conduct an investigation; and scripts on how to conduct an investigation, and determining whether a chain of related low level abnormalities associated with one or more of the entities should be determined to be one or more incidents worthy of generating a notification to a human user for possible further investigation and/or worthy of being determined as an actual cyber-threat; and
triggering an autonomous response from an autonomous response module to mitigate the cyber-threat.

12. A method according to claim 11 wherein the signal from the external apparatus is provided by a manual user input at the interface.

13. A method according to claim 12 wherein the manual user input comprises investigation instructions comprising one or more of: a time period to be investigated, designation of the one or more entities to be investigated, and a particular investigation environment.

14. A method according to claim 11 wherein the signal from the external apparatus is provided by a third party threat intelligence component.

15. A method according to claim 14 wherein the third party threat intelligence component provides additional data relating to behavior of the one or more entities.

16. A method according to claim 11 further comprising, in response to the determination of one or more incidents, generating a directed graph, using graph theory, to map the one or more incidents onto the graph to detect anomalies potentially indicative of cyber-threats.

17. A method according to claim 16 wherein the directed graph comprises a plurality of nodes, each node of the plurality of nodes corresponding to a respective entity of the entities, the plurality of nodes being connected by one or more edges corresponding to the one or more incidents.

18. A method according to claim 16 further comprising grouping the one or more incidents into a meta of incidents representing a compromise linking entities in the network affected by the one or more incidents.

19. A method according to claim 16 further comprising generating a visual representation of the directed graph.

20. A non-transitory computer-readable medium including executable instructions that, when executed with one or more processors, cause a cyber-threat defense system to perform the method of claim 11.

* * * * *