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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 October 2016 (20.10.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/168531 A1

(51) International Patent Classification:
H04L 29/06 (2006.01) *H04L 12/22* (2006.01)

(21) International Application Number:
PCT/US2016/027659

(22) International Filing Date:
15 April 2016 (15.04.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/148,232 16 April 2015 (16.04.2015) US
15/098,861 14 April 2016 (14.04.2016) US

(71) Applicant: **NEC LABORATORIES AMERICA, INC.** [US/US]; 4 Independence Way, Suite 200, Princeton, New Jersey 08540 (US).

(72) Inventors: **TANG, Lu-An**; 10 Manley Road, Pennington, New Jersey 08534 (US). **CHEN, Zhengzhang**; 44 York Road, Princeton Junction, New Jersey 08550 (US). **CHEN, Ting**; 90 Maple Street, Malden, Massachusetts 02148 (US). **JIANG, Guofei**; 5 Danby Court, Princeton, New Jersey 08540 (US). **XU, Fengyuan**; 2008 Housell Court, Franklin Park, New Jersey 08823 (US). **CHEN, Haifeng**; 11 Blackhawk Court, West Windsor, New Jersey 08550 (US).

(74) Agent: **KOLODKA, Joseph**; NEC Laboratories America, Inc., 4 Independence Way, Suite 200, Princeton, New Jersey 08540 (US).

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

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

Published:

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

(54) Title: INTEGRATED COMMUNITY AND ROLE DISCOVERY IN ENTERPRISE NETWORKS

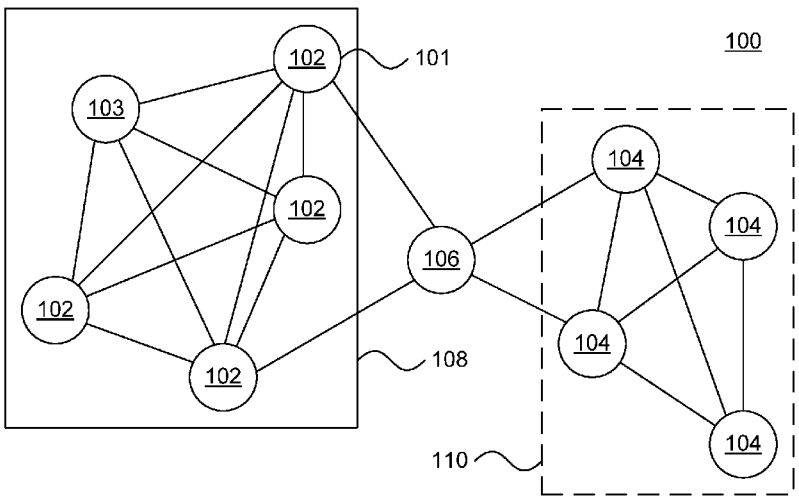

FIG. 1

(57) Abstract: Methods and systems for detecting anomalous communications include simulating a network graph based on community and role labels of each node in the network graph based on one or more linking rules. The community and role labels of each node are adjusted based on differences between the simulated network graph and a true network graph. The simulation and adjustment are repeated until the simulated network graph converges to the true network graph to determine a final set of community and role labels. It is determined whether a network communication is anomalous based on the final set of community and role labels.

WO 2016/168531 A1

INTEGRATED COMMUNITY AND ROLE DISCOVERY IN ENTERPRISE NETWORKS

RELATED APPLICATION INFORMATION

[0001] This application claims priority to 62/148,232, filed on April 16, 2015, incorporated herein by reference in its entirety.

BACKGROUND

Technical Field

[0002] The present invention relates to computer and network security and, more particularly, to integrated discovery of node community and role in such networks.

Description of the Related Art

[0003] Enterprise networks are key systems in corporations and they carry the vast majority of mission-critical information. As a result of their importance, these networks are often the targets of attack. Communications on enterprise networks are therefore frequently monitored and analyzed to detect anomalous network communication as a step toward detecting attacks.

[0004] However, accurate and effective detection is difficult if the system lacks knowledge of community and roles. Community represents the working group that a machine belongs to, while role represents the function of the machine (e.g., as an email server, as a data server, as a personal desktop, etc.). It often isn't possible for users to provide an accurate picture of community and role for an entire network.

[0005] Existing approaches to community and role detection treat the questions separately, for example detecting roles without taking community structures into account and detecting a node's community while ignoring its role, when in fact communities and roles are tightly coupled and cannot be separated in real networks.

SUMMARY

[0006] A method for detecting anomalous communications includes simulating a network graph based on community and role labels of each node in the network graph based on one or more linking rules. The community and role labels of each node are adjusted based on differences between the simulated network graph and a true network graph. The simulation and adjustment are repeated until the simulated network graph converges to the true network graph to determine a final set of community and role labels. It is determined whether a network communication is anomalous based on the final set of community and role labels.

[0007] A system for detecting anomalous communications includes a community and role detection module having a processor configured to simulate a network graph based on community and role labels of each node in the network graph based on one or more linking rules, to adjust the community and role labels of each node based on differences between the simulated network graph and a true network graph, and to repeat said simulation and adjustment until the simulated network graph converges to the true network graph to determine a final set of community and role labels. An anomaly detection module is configured to determine whether a network communication is anomalous based on the final set of community and role labels.

[0008] These and other features and advantages will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

[0009] The disclosure will provide details in the following description of preferred embodiments with reference to the following figures wherein:

[0010] FIG. 1 is directed to a network graph representing communities and roles of nodes in accordance with the present principles.

[0011] FIG. 2 is a block/flow diagram of a method of discovering community and role memberships and detecting anomalies in accordance with the present principles.

[0012] FIG. 3 is a block/flow diagram of a method of detecting anomalies in accordance with the present principles.

[0013] FIG. 4 is a block diagram of a system for discovering community and role memberships and detecting anomalies in accordance with the present principles.

[0014] FIG. 5 is a block diagram of a processing system in accordance with the present principles.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0015] In accordance with the present principles, the present embodiments detect communities and roles in a network in an integrated manner. In particular, every node in a network is associated not only with community membership, but also with role membership, so that the system can capture both community and role structures

simultaneously. When two nodes attempt to interact (e.g., when forming an edge between two nodes on the graph representing the network), both community and role memberships are considered when determining how probable the link is and, thus, whether the link can be considered anomalous. The community and role of each node is determined, in one embodiment, according to Gibbs sampling-based learning.

[0016] Referring now in detail to the figures in which like numerals represent the same or similar elements and initially to FIG. 1, an exemplary computer network 100 is illustratively depicted in accordance with one embodiment of the present principles. The network 100 is formed from a set of nodes 101, each of which has a role and a community. In the embodiment of FIG. 1, the nodes marked 102 have a community 108, while the nodes marked 104 have a community 110. It should be noted that the network graph 100 does not represent a physical network, but instead represents communications between the nodes 101, with each edge of the graph representing a communications link. There is nothing in principle stopping a node 102 from community 108 from forming a link with a node 104 in community 110. However, the present embodiments will consider the communities and roles of the nodes 101 in determining whether that link is anomalous. The nodes 101 are described herein as representing individual devices, but it should be understood that in some embodiments a single node 101 may incorporate multiple devices and, conversely, a single device may host multiple nodes 101. Similarly, a single node 101 may occupy multiple roles.

[0017] It should be understood that nodes 101 in different communities will have a low likelihood of interaction with one another (e.g., a low probability of forming a link). However, one exception is in the case of a node 106 that has a specific role, such as a

router or bridge. In this case, the node 106 may belong to one, both, or neither of the communities 108 and 110, and its role as an intermediary between those two communities will strongly influence its likelihood of forming connections with other nodes 101. This may be referred to as a background role-based connection. Note though that communities need not be identified with physical network segments—a community may instead simply represent for example a department or other organizational structure that communicates frequently within itself and relatively rarely with other departments.

[0018] Similarly, when two nodes are in the same community they will interact with a higher probability, but roles are also a strong factor. For example, a file server 103 within community 108 may interact more frequently with user terminals 102 than those nodes 102 interact with one another. This may be referred to as a within-community role-based connection.

[0019] Referring now to FIG. 2, a method for detecting anomalous links is shown. Block 202 generates an adjacency matrix representation of a blueprint graph, which is a heterogeneous graph constructed from a historical dataset of communications in the network 100, with nodes 101 representing physical devices on an enterprise network and edges reflecting the normal communication patterns among the nodes 101. For each pair of nodes in the adjacency matrix, block 204 generates community and role labels. The initial labels generated by block 204 may be random or may be generated according to any initial information that is available (e.g., based on known software installed on respective nodes 101 or based on an existing network map).

[0020] Block 206 then simulates the interactions of node pairs between different communities and roles. The simulation is based on a set of rules for known interactions

between community members and according to roles. For example, the nodes 104 marked by the labels as being members of community 110 will have a simulated link between them. In another example, server/client role relationships can be represented as links. This simulation is used to generate a simulated graph blueprint. Block 207 uses the simulated graph blueprint to form a synthetic adjacency matrix for the simulated graph.

[0021] If there are discrepancies between the adjacency matrix and the synthetic adjacency matrix, block 208 adjusts the community and role labels to bring the simulated links closer to the actual links in the blueprint graph. Block 210 then determines whether the synthetic matrix has converged with the real adjacency matrix, such that the links in the simulated graph match those of the blueprint graph. Convergence may be satisfied when the synthetic adjacency matrix is identical to the real adjacency matrix or may alternatively be based on a similarity metric for the matrices, where convergence is reached when the similarity metric is below a threshold. If so, block 212 uses the detected community and role labels to determine whether there is an anomaly. If not, processing returns to block 206 until the synthetic matrix does converge.

[0022] In one example of anomaly detection, consider a first node n_1 that has the role label of, "database server," and a community label of, "system team." A second node n_2 has the role label of, "email server," and the community label of, "operational team." If a new network connection between n_1 and n_2 is detected, the system can determine that the database server of one team will rarely have legitimate need to communicate with the email server of another team (with such information being set by the domain user). Block 212 may then determine that an intrusion has occurred.

[0023] The assignment of labels in block 204 may be performed as a respective community membership vector π_i and a respective role membership vector θ_i for each node i. When a pair of nodes (i,j) attempts to form a link, their community and role membership assignments $Z_{ij}^c, Z_{ji}^c, Z_{ij}^r, Z_{ji}^r$ are drawn according to a multinomial distribution parameterized by their membership distribution vectors, with Z_{ij}^c being the community assignment of node i for the pair of nodes (i,j) and Z_{ij}^r being the role assignment of node i for the pair of nodes (i,j). The question of whether a link is formed is represented as a Bernoulli event based on the community and role assignments of the two nodes and an interaction parameter B that characterizes the interaction probability between two community and role assignment tuples, for example (Z_{ij}^c, Z_{ij}^r).

[0024] The parameters π, θ, and B are treated as random variables, with Beta prior on each entry of B. The term $B_{\delta pq}$ is a Bernoulli distribution, and π_i and θ_i have a multinomial distribution with Dirichlet priors. The present model can then be summarized as follows:

[0025] For each entry (δ, p, q) in B:

[0026] draw $B_{\delta pq} \sim \text{Beta}(\xi_{\delta pq}^1, \xi_{\delta pq}^2)$.

[0027] For each node i:

[0028] Draw a community membership vector $\pi_i \sim \text{Dirichlet}(\alpha^c)$

[0029] Draw a role membership distribution vector $\theta_i \sim \text{Dirichlet}(\alpha^r)$

[0030] For each node pair (i,j):

[0031] Draw node i's community $Z_{ij}^c \sim \text{Multinomial}(\pi_i)$

[0032] Draw node j's community $Z_{ji}^c \sim \text{Multinomial}(\pi_j)$

[0033] Draw node i's role $Z_{ij}^r \sim \text{Multinomial}(\theta_i)$

[0034] Draw node j's role $Z_{ji}^{r} \sim \text{Multinomial}(\theta_j)$

[0035] Draw link $E_{ij} \sim \text{Bernoulli}\left(B_{\delta\left(z_{ij}^{c}, z_{ji}^{c}\right), z_{ij}^{r}, z_{ji}^{r}}\right)$

[0036] Under the above generative model, when the adjacency matrix E_{ij} is observed, the posterior distribution of hidden variables, such as membership vectors, can be inferred. Given the network communications data, the posterior distribution and, in particular, the posterior mean, of the variables in the model are inferred. Due to the complicated integrals over hidden states in the posterior inference, exact inference is intractable. The present embodiments therefore employ Gibbs sampling inference, though it should be understood that other types of inference may be used instead.

[0037] In Gibbs sampling, a Markov chain is maintained. The chain sequentially reaches its next state by sampling a variable from its distribution when conditioned on current values of all of the other variables. When the Markov chain approaches an equilibrium distribution, the subsequent samples are generated from the target distribution. Using collapsed Gibbs sampling, direct samples of the Dirichlet membership variables π and θ are avoided by integrating those variable out. Thus, only the membership assignments of a pair of nodes (i,j) are sampled at a time according to the pair's conditional distribution. The conditional distribution P is therefore computed, representing the community and role assignments of the pair of nodes (i,j) given the adjacency matrix E_{ij} and current assignments of the other node pairs. The conditional distribution P is defined as:

$$P \propto \frac{\left(n^{-ij}_{\delta(a,b)pq+} + \xi^1\right)^{E_{ij}} \left(n^{-ij}_{\delta(a,b)pq-} + \xi^2\right)^{1-E_{ij}}}{n^{-ij}_{\delta(a,b)pq+} + n^{-ij}_{\delta(a,b)pq-} + \xi^1 + \xi^2} \left(h^{-ij}_{ia} + \alpha^c\right)\left(h^{-ij}_{jb} + \alpha^c\right)\left(m^{-ij}_{ip} + \alpha^r\right)\left(m^{-ij}_{ip} + \alpha^r\right)$$

where $a = Z^c_{ij}$, $b = Z^c_{ji}$, $p = Z^r_{ij}$, $q = Z^r_{ji}$, h_{ia} is the count of the node i assigned to community a, m_{ip} is the count of the node i assigned to role b, $n^{-ij}_{\delta(a,b)pq+}$ is a count of linked node pairs with community assignments a and b and role assignments p and q, $n^{-ij}_{\delta(a,b)pq-}$ is a count of unlinked node pairs with community assignments a and b and role assignments p and q, ξ^1 and ξ^2 are scalar Beta hyperparameters for (k, p, q) in the interaction tensor B.

[0038] It is worth noting that the conditional distribution P is proportional to two parts: the rate of link/non-link given the community and role assignments of the two nodes, and the ratio (after normalization) of community and role membership assignments of both nodes. Both parts are calculated by excluding their current assignments.

[0039] The Markov chain can then be initialized by a given community and role membership assignments for all node pairs. The chain can be run by sequentially re-sampling assignments of each pair of nodes conditioned on the rest. Once the assignments of a pair of nodes are updated, the counters n, m and h are also updated. After enough iterations, the Markov chain approaches the equilibrium distribution. The subsequent samples of the community and role assignments can be collected to estimate the posterior distribution of the variables.

[0040] The community membership of node i is Dirichlet distributed, and its mean at a^{th} dimension is:

$$\pi_{ia} = \frac{(h_{ia} + \alpha^c)}{\sum_{a=1}^{K^c} h_{ia} + K^c \alpha^c}$$

where K^c is the number of communities and α^c is the Dirichlet hyperparameter for π_i. The role membership of the node *i* is also Dirichlet distributed, and its mean at the p^{th} dimension is given by:

$$\theta_{ip} = \frac{(m_{ip} + \alpha^r)}{\sum_{p=1}^{K^r} m_{ip} + K^r \alpha^r}$$

where K^r is the number of roles and α^r is the Dirichlet hyperparameter for θ_i. The interaction tensor *B* is Beta distributed, with the mean of each entry being estimated by:

$$B_{kpq} = \frac{n_{kpq+} + \xi^1}{n_{kpq+} + n_{kpq-} + \xi^1 + \xi^2}$$

[0041] Blocks 206 and 207 therefore compute the conditional distribution for each pair of nodes (i, j) and block 208 determines π_{ia}, θ_{ip}, and B_{kpq}.

[0042] Embodiments described herein may be entirely hardware, entirely software or including both hardware and software elements. In a preferred embodiment, the present invention is implemented in software, which includes but is not limited to firmware, resident software, microcode, etc.

[0043] Embodiments may include a computer program product accessible from a computer-usable or computer-readable medium providing program code for use by or in connection with a computer or any instruction execution system. A computer-usable or computer readable medium may include any apparatus that stores, communicates, propagates, or transports the program for use by or in connection with the instruction execution system, apparatus, or device. The medium can be magnetic, optical, electronic, electromagnetic, infrared, or semiconductor system (or apparatus or device) or a

propagation medium. The medium may include a computer-readable storage medium such as a semiconductor or solid state memory, magnetic tape, a removable computer diskette, a random access memory (RAM), a read-only memory (ROM), a rigid magnetic disk and an optical disk, etc.

[0044] Each computer program may be tangibly stored in a machine-readable storage media or device (e.g., program memory or magnetic disk) readable by a general or special purpose programmable computer, for configuring and controlling operation of a computer when the storage media or device is read by the computer to perform the procedures described herein. The inventive system may also be considered to be embodied in a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

[0045] A data processing system suitable for storing and/or executing program code may include at least one processor coupled directly or indirectly to memory elements through a system bus. The memory elements can include local memory employed during actual execution of the program code, bulk storage, and cache memories which provide temporary storage of at least some program code to reduce the number of times code is retrieved from bulk storage during execution. Input/output or I/O devices (including but not limited to keyboards, displays, pointing devices, etc.) may be coupled to the system either directly or through intervening I/O controllers.

[0046] Network adapters may also be coupled to the system to enable the data processing system to become coupled to other data processing systems or remote printers or storage devices through intervening private or public networks. Modems, cable

modem and Ethernet cards are just a few of the currently available types of network adapters.

[0047] Referring now to FIG. 3, a method of performing intrusion detection based on an integrated network-level analysis that includes both community and role information is shown. Block 302 collects data from agents installed on each of the nodes 101. The agents collect information regarding each node's activities, including for example host-level activities (e.g., user-to-process events, process-to-file events, user-to-registry events, etc.) and network-level activities (e.g., TCP and UDP connections with other nodes 101 on the network 100).

[0048] Block 304 performs a network-level analysis using the collected information. The network-level analysis is described in greater detail above and integrates both node community membership and node role membership to detect anomalous communications. Block 306 performs a host-level analysis based on the collected information to determine whether anomalous behavior has occurred locally within a single node 101.

[0049] Block 308 integrates the network-level and host-level anomalies to provide intrusion detection events. This may include further contextual analysis to detect interactions between network-level and host-level anomalies, for example noting that certain host-level and network-level anomalies may have greater import when occurring together. Block 310 then presents the detected intrusion events to a user for review and for further action. In some embodiments, block 312 may automatically respond to the intrusion detection event. The response may include, for example, blocking certain network-level communications, restricting access on the level of an individual host, changing security policies, and providing alerts to interested parties, such as a system

administrator. Block 312 may consider the specific intrusion information determined by block 308 to determine a best course of action.

[0050] Referring now to FIG. 4, a network-level anomaly detection system 400 is shown. The detection system 400 includes a hardware processor 402 and a memory 404, as well as a network interface 405. The system 400 further includes certain functional modules that may, in some embodiments, be implemented as software that is stored in the memory 404 and executed by processor 402. In other embodiments, the functional modules may be implemented as one or more discrete hardware components, for example in the form of an application-specific integrated chip or field programmable gate array.

[0051] The system 400 collects historical data 406 regarding the network 100 via the network interface 405 and stores the historical data 406 in the memory 404. This historical data 406 includes information that reflects communications between nodes 101 on the network 100 and is provided by agents at the individual nodes 101 that report what each respective node 101 is doing. The historical data 406 is used to construct a blueprint graph 410 of the network 100, with nodes 101 of the blueprint graph representing individual hosts on the network 100 and edges representing normal communications between the nodes 101.

[0052] A community and role detection module 408 automatically discovers the community and role memberships of each node 101 in the network 100 as described in detail above. The community and role detection module 408 uses the processor 402 to analyze the blueprint graph 410 and provides membership vectors θ and π. Anomaly detection module 412 uses the membership vectors and the blueprint graph to review incoming information about current network communications and to determine whether a

given communication is anomalous. The anomaly detection module 412 furthermore uses the incoming network communications to make adjustments to the blueprint graph 410, which in turn may lead to adjustments in the community and role memberships.

[0053] Referring now to FIG. 5, an exemplary processing system 500 is shown which may represent the network-level anomaly detection system 400. The processing system 500 includes at least one processor (CPU) 504 operatively coupled to other components via a system bus 502. A cache 506, a Read Only Memory (ROM) 508, a Random Access Memory (RAM) 510, an input/output (I/O) adapter 520, a sound adapter 530, a network adapter 540, a user interface adapter 550, and a display adapter 560, are operatively coupled to the system bus 502.

[0054] A first storage device 522 and a second storage device 524 are operatively coupled to system bus 502 by the I/O adapter 520. The storage devices 522 and 524 can be any of a disk storage device (e.g., a magnetic or optical disk storage device), a solid state magnetic device, and so forth. The storage devices 522 and 524 can be the same type of storage device or different types of storage devices.

[0055] A speaker 532 is operatively coupled to system bus 502 by the sound adapter 530. A transceiver 542 is operatively coupled to system bus 502 by network adapter 540. A display device 562 is operatively coupled to system bus 502 by display adapter 560.

[0056] A first user input device 552, a second user input device 554, and a third user input device 556 are operatively coupled to system bus 502 by user interface adapter 550. The user input devices 552, 554, and 556 can be any of a keyboard, a mouse, a keypad, an image capture device, a motion sensing device, a microphone, a device incorporating the functionality of at least two of the preceding devices, and so forth. Of course, other

types of input devices can also be used, while maintaining the spirit of the present principles. The user input devices 552, 554, and 556 can be the same type of user input device or different types of user input devices. The user input devices 552, 554, and 556 are used to input and output information to and from system 500.

[0057] Of course, the processing system 500 may also include other elements (not shown), as readily contemplated by one of skill in the art, as well as omit certain elements. For example, various other input devices and/or output devices can be included in processing system 500, depending upon the particular implementation of the same, as readily understood by one of ordinary skill in the art. For example, various types of wireless and/or wired input and/or output devices can be used. Moreover, additional processors, controllers, memories, and so forth, in various configurations can also be utilized as readily appreciated by one of ordinary skill in the art. These and other variations of the processing system 500 are readily contemplated by one of ordinary skill in the art given the teachings of the present principles provided herein.

[0058] The foregoing is to be understood as being in every respect illustrative and exemplary, but not restrictive, and the scope of the invention disclosed herein is not to be determined from the Detailed Description, but rather from the claims as interpreted according to the full breadth permitted by the patent laws. It is to be understood that the embodiments shown and described herein are only illustrative of the principles of the present invention and that those skilled in the art may implement various modifications without departing from the scope and spirit of the invention. Those skilled in the art could implement various other feature combinations without departing from the scope and spirit of the invention. Having thus described aspects of the invention, with the details and

particularity required by the patent laws, what is claimed and desired protected by Letters Patent is set forth in the appended claims.

WHAT IS CLAIMED IS:

1. A method for detecting anomalous communications, comprising:

simulating a network graph based on community and role labels of each node in the network graph based on one or more linking rules;

adjusting the community and role labels of each node based on differences between the simulated network graph and a true network graph;

repeating said simulating and adjusting until the simulated network graph converges to the true network graph to determine a final set of community and role labels; and

determining whether a network communication is anomalous based on the final set of community and role labels.

2. The method of claim 1, wherein adjusting the community and role labels of each node comprises determining a conditional distribution for each pair of nodes in a network graph based on a rate of linking for a community and role label of each node in the pair of nodes and a ratio of community and role labels of both nodes.

3. The method of claim 1, further comprising determining initial community and role labels for each of a plurality of nodes.

4. The method of claim 3, wherein determining initial community and role labels comprises randomly assigning a community and role label to each node.

5. The method of claim 1, wherein the true network graph is based on historical communications between the nodes.

6. The method of claim 1, wherein repeating said simulating and adjusting comprises determining a true adjacency matrix based on the true network graph and a synthetic adjacency matrix based on the simulated network graph.

7. The method of claim 6, wherein repeating said simulating and adjusting further comprises determining whether the simulated network graph has converged to the true network graph by determining a similarity of the synthetic adjacency matrix to the true adjacency matrix.

8. The method of claim 1, wherein determining whether a network communication is anomalous comprises determining a probability of the network communication taking place between an associated first node and second node based on the community and role labels of the respective first and second nodes.

9. The method of claim 1, further comprising automatically responding to a detected intrusion event, said response comprising one or more of blocking the network communication, restricting access, changing security policies, and alerting a system administrator.

10. A system for detecting anomalous communications, comprising:

a community and role detection module comprising a processor configured to simulate a network graph based on community and role labels of each node in the network graph based on one or more linking rules, to adjust the community and role labels of each node based on differences between the simulated network graph and a true network graph, and to repeat said simulation and adjustment until the simulated network graph converges to the true network graph to determine a final set of community and role labels; and

an anomaly detection module configured to determine whether a network communication is anomalous based on the final set of community and role labels.

11. The system of claim 10, wherein the community and role detection module is further configured to determine a conditional distribution for each pair of nodes in a network graph based on a rate of linking for a community and role label of each node in the pair of nodes and a ratio of community and role labels of both nodes.

12. The system of claim 10, wherein the community and role detection module is further configured to determine initial community and role labels for each of a plurality of nodes.

13. The system of claim 12, wherein the community and role detection module is further configured to randomly assign a community and role label to each node.

14. The system of claim 10, wherein the true network graph is based on historical communications between the nodes.

15. The system of claim 10, wherein the community and role detection module is further configured to determine a true adjacency matrix based on the true network graph and a synthetic adjacency matrix based on the simulated network graph.

16. The system of claim 15, wherein the community and role detection module is further configured to determine whether the simulated network graph has converged to the true network graph by determining a similarity of the synthetic adjacency matrix to the true adjacency matrix.

17. The system of claim 10, wherein the anomaly detection module is further configured to determine a probability of the network communication taking place between an associated first node and second node based on the community and role labels of the respective first and second nodes.

18. The system of claim 10, wherein the anomaly detection module is further configured to automatically responding to a detected intrusion event, said response comprising one or more of blocking the network communication, restricting access, changing security policies, and alerting a system administrator.

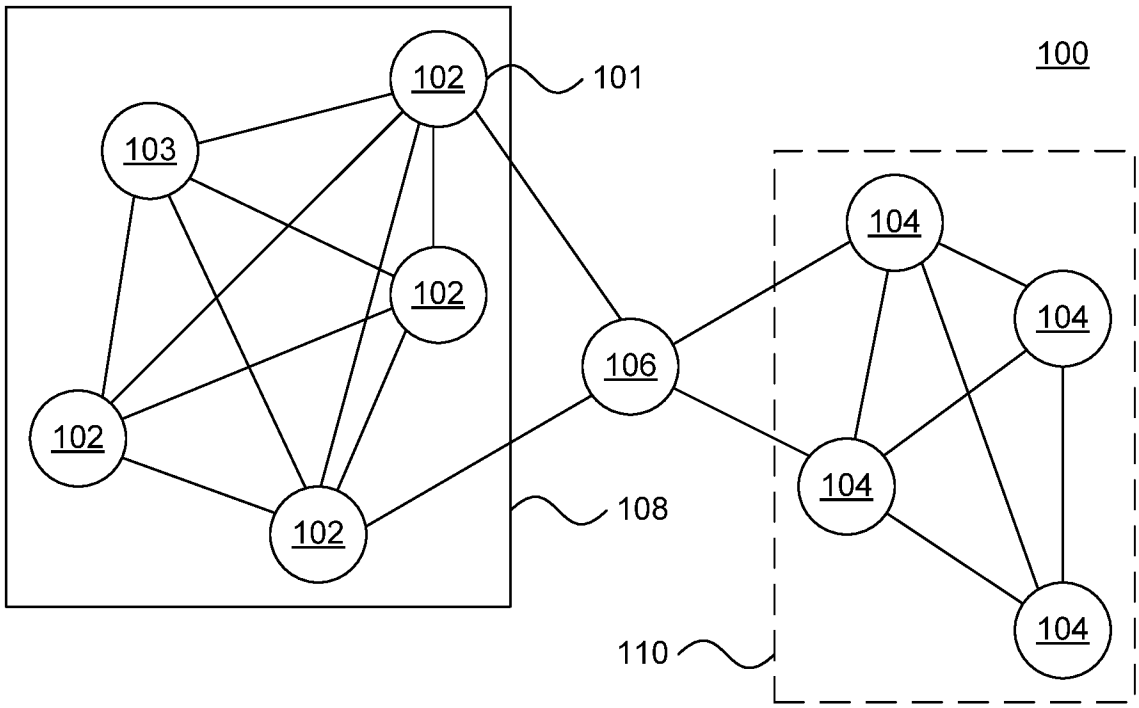

FIG. 1

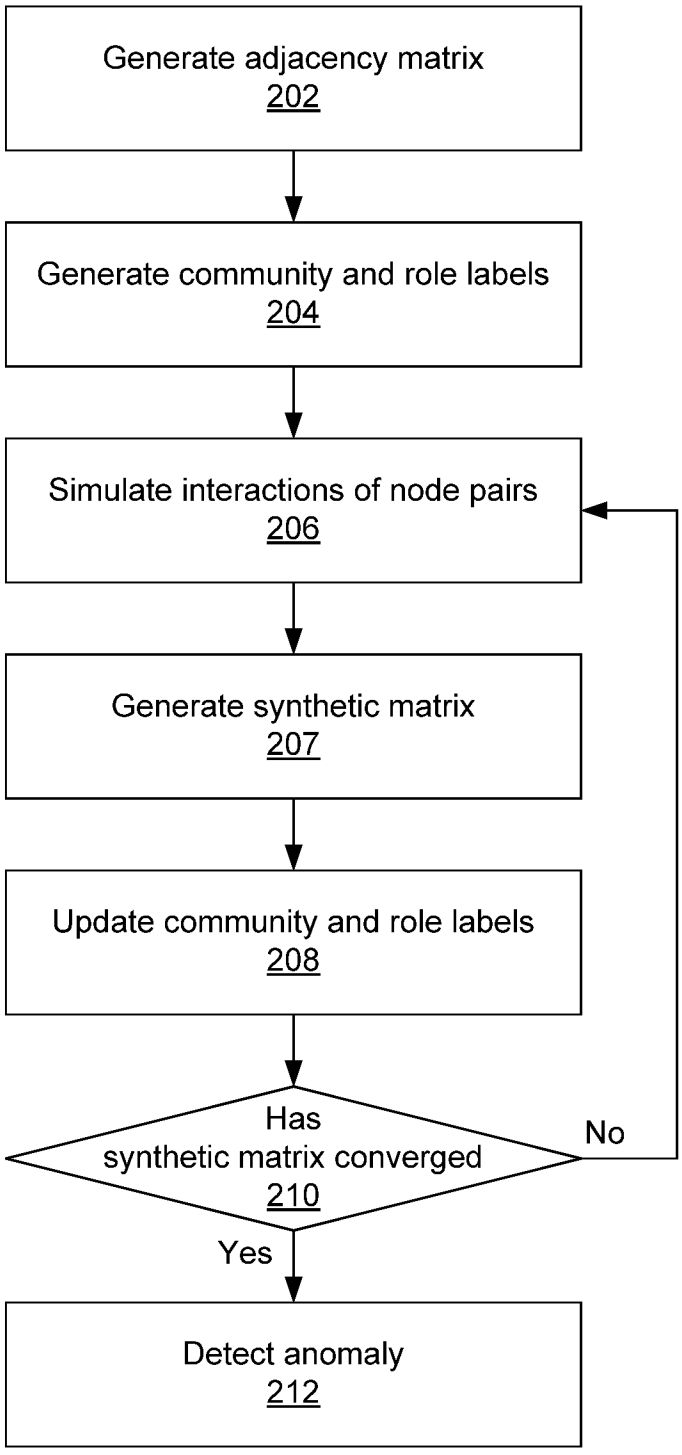
Generate adjacency matrix
202
Generate community and role labels
204
Simulate interactions of node pairs
206
Generate synthetic matrix
207
Update community and role labels
208
Has
synthetic matrix converged
210
No
Yes
Detect anomaly
212

FIG. 2

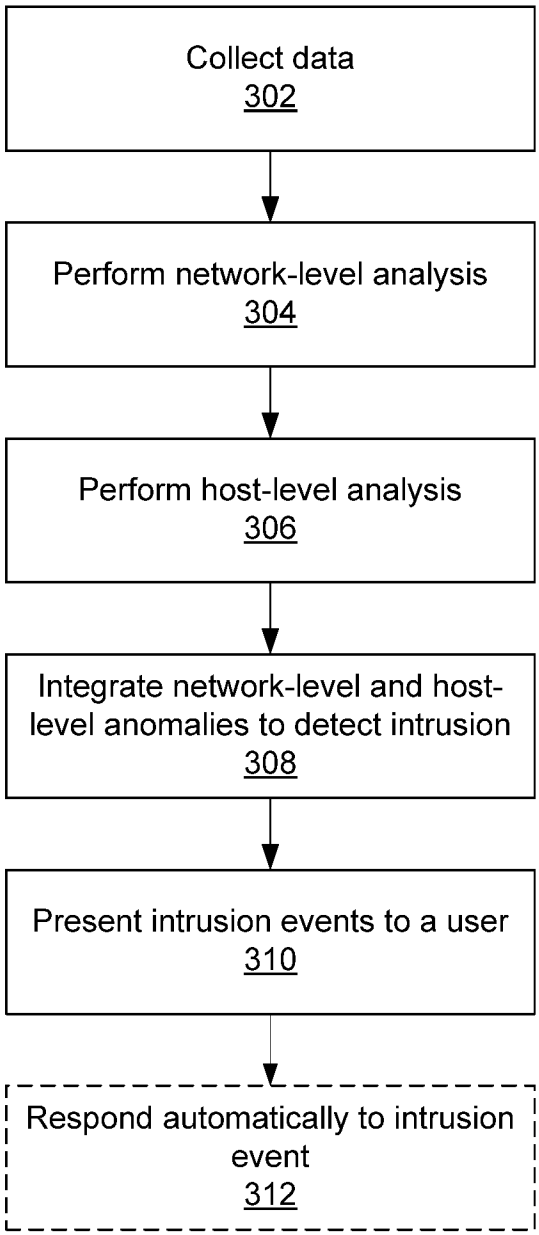
Collect data
302
Perform network-level analysis
304
Perform host-level analysis
306
Integrate network-level and host-level anomalies to detect intrusion
308
Present intrusion events to a user
310
Respond automatically to intrusion event
312

FIG. 3

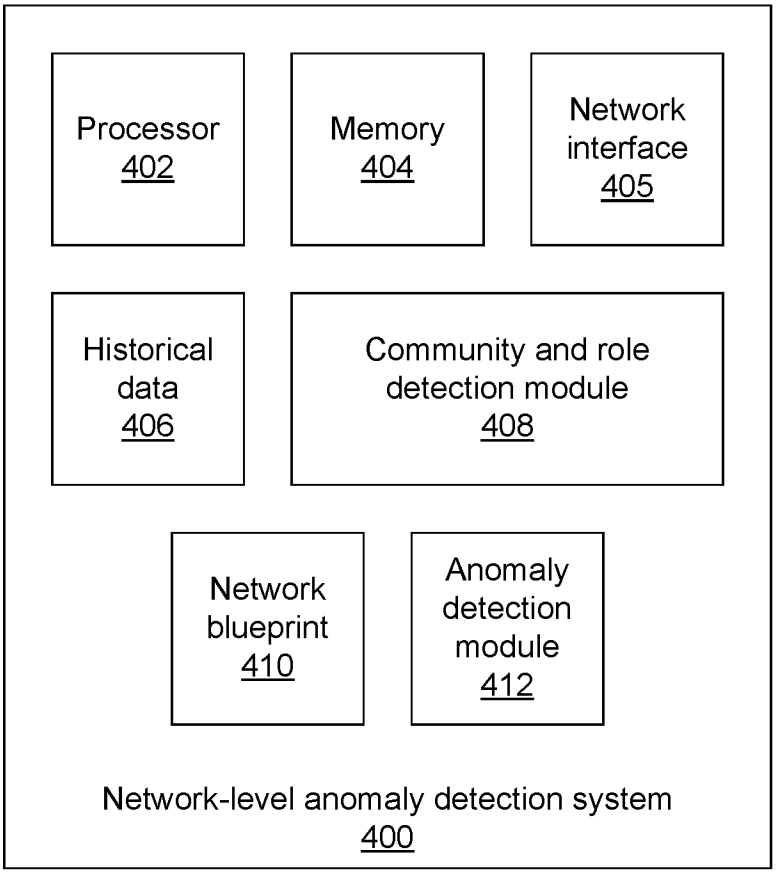
Processor
402
Memory
404
Network interface
405
Historical data
406
Community and role detection module
408
Network blueprint
410
Anomaly detection module
412
Network-level anomaly detection system
400

FIG. 4

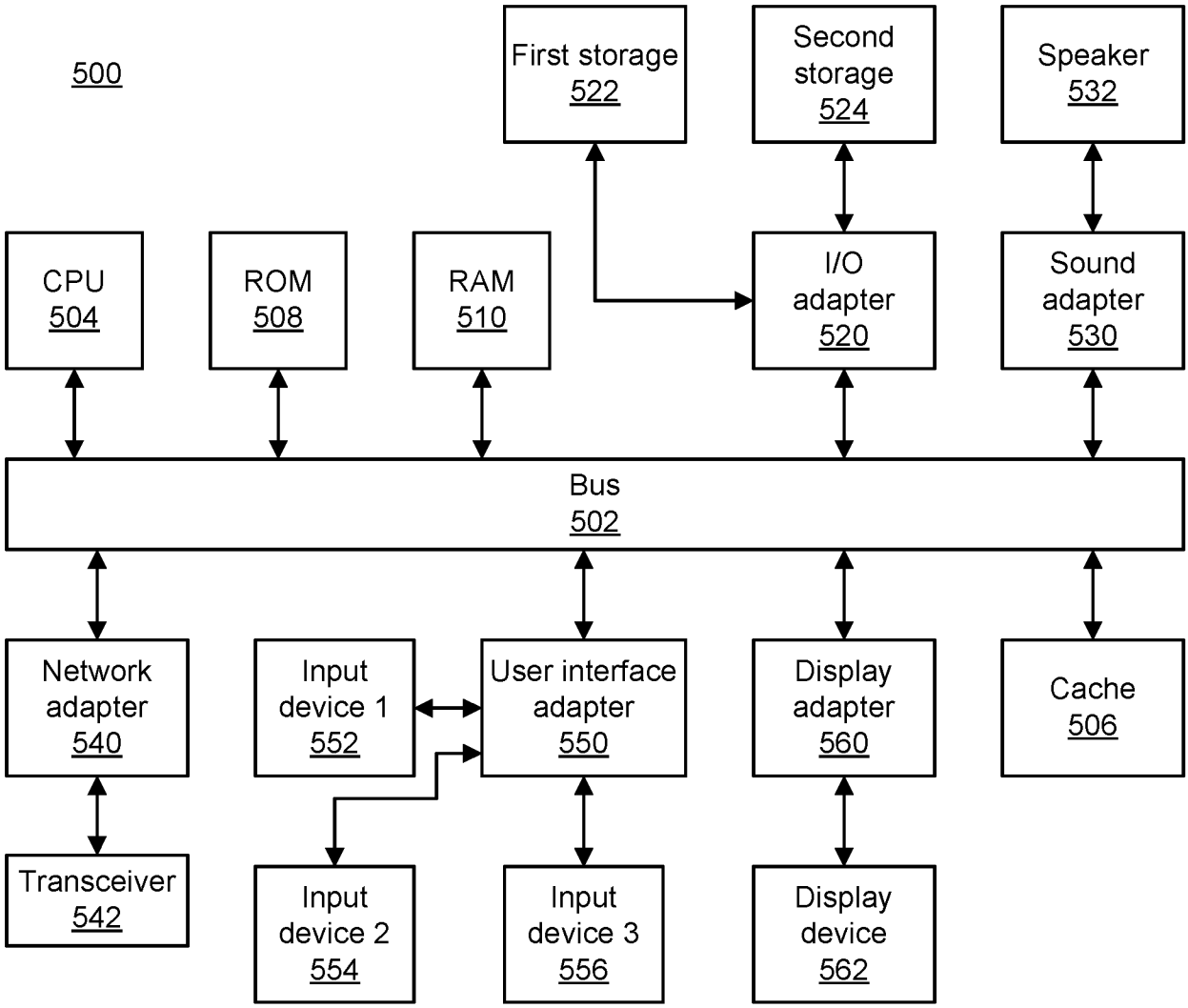
500
First storage 522
Second storage 524
Speaker 532
CPU 504
ROM 508
RAM 510
I/O adapter 520
Sound adapter 530
Bus 502
Network adapter 540
Input device 1 552
User interface adapter 550
Display adapter 560
Cache 506
Transceiver 542
Input device 2 554
Input device 3 556
Display device 562

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/027659

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06(2006.01)i, H04L 12/22(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04L 29/06; H04J 1/16; G06F 15/00; G06F 11/34; H04L 12/24; H04L 12/26; H04L 12/22

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & keywords: anomalous communication, simulation, adjustment, repeat, community and role lable

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2005-0169185 A1 (LILI QIU et al.) 04 August 2005 See paragraphs [0044], [0118]-[0126], claim 1 and figure 7.	1,3,9-10,12,18
A		2,4-8,11,13-17
A	KR 10-2009-0040017 A (KOREA INFORMATION SECURITY AGENCY) 23 April 2009 See paragraphs [0018]-[0040] and figures 1-3.	1-18
A	US 2009-0031176 A1 (TSUYOSHI IDE et al.) 29 January 2009 See paragraphs [0075]-[0085] and figures 3-6.	1-18
A	KR 10-2014-0035664 A (GROOS) 24 March 2014 See paragraphs [0041]-[0075] and figures 3-10.	1-18
A	US 7808916 B1 (KENICHI FUTAMURA et al.) 05 October 2010 See column 3, line 56 - column 9, line 30 and figure 2.	1-18

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

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

Date of the actual completion of the international search	Date of mailing of the international search report
08 July 2016 (08.07.2016)	**08 July 2016 (08.07.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
KIM, KI HO
Telephone No. +82-42-481-8691

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2016/027659

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2005-0169185 A1	04/08/2005	AT 350832 T	15/01/2007
		CN 1665205 A	07/09/2005
		CN 1665205 B	29/09/2010
		CN 1665205 C	07/09/2005
		DE 602005000383 D1	15/02/2007
		DE 602005000383 T2	26/04/2007
		DK 1560366 T3	14/05/2007
		EP 1560366 A1	03/08/2005
		EP 1560366 B1	03/01/2007
		ES 2279479 T3	16/08/2007
		JP 04786908 B2	05/10/2011
		JP 2005-223906 A	18/08/2005
		KR 10-1098744 B1	23/12/2011
		KR 10-2006-0042903 A	15/05/2006
		US 2005-0169186 A1	04/08/2005
		US 2005-0204028 A1	15/09/2005
		US 7583587 B2	01/09/2009
		US 7606165 B2	20/10/2009
		US 7613105 B2	03/11/2009
KR 10-2009-0040017 A	23/04/2009	KR 10-0951144 B1	07/04/2010
		US 2009-0106844 A1	23/04/2009
US 2009-0031176 A1	29/01/2009	JP 03922375 B2	30/05/2007
		JP 2005-216066 A	11/08/2005
		US 2005-0193281 A1	01/09/2005
		US 7346803 B2	18/03/2008
		US 7647524 B2	12/01/2010
KR 10-2014-0035664 A	24/03/2014	KR 10-1380768 B1	02/04/2014
US 7808916 B1	05/10/2010	None	