US009270689B1

(12) **United States Patent**
Wang et al.

(10) **Patent No.: US 9,270,689 B1**
(45) **Date of Patent: Feb. 23, 2016**

(54) **DYNAMIC AND ADAPTIVE TRAFFIC SCANNING**

(75) Inventors: **Jisheng Wang**, Belmont, CA (US); **Daniel Quinlan**, San Bruno, CA (US); **Lee Jones**, Hayard, CA (US)

(73) Assignee: **Cisco Technology, Inc.**, San Jose, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 611 days.

(21) Appl. No.: **13/529,134**

(22) Filed: **Jun. 21, 2012**

(51) **Int. Cl.**
H04L 29/06 (2006.01)

(52) **U.S. Cl.**
CPC ***H04L 63/1408*** (2013.01)

(58) **Field of Classification Search**
CPC H04L 63/1408
USPC 726/24
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,851,058	B1 *	2/2005	Gartside	726/24
7,882,560	B2	2/2011	Kraemer et al.	
8,370,943	B1 *	2/2013	Dongre et al.	726/24
2012/0110667	A1 *	5/2012	Zubrilin et al.	726/24
2013/0291109	A1 *	10/2013	Staniford et al.	726/23

OTHER PUBLICATIONS

Anti-Virus Comparative, Mar. 2012, Retrieved from the Internet <URL: .av-comparatives.org/images/stories/test/ondret/avc_fd_mar2012_intl_en.pdf>, pp. 1-12 as printed.*

Oberheide, CloudAV: N-version Antivirus in the Network Cloud, 2008, Retrieved from the Internet <URL: jon.oberheide.org/files/usenix08-clouday.pdf>, pp. 1-16 as printed.*

Grover, "A fast quantum mechanical algorithm for database search", Proceedings, STOC, 1996, pp. 1-8.

Press, "Strong profiling is not mathematically optimal for discovering rare malfeasors", Los Alamos National Laboratory, 2006, 18 pages.

Montanaro, "Quantum search with advice", Department of Computer Science, University of Bristol, Sep. 14, 2009, pp. 1-14.

Hoogstrate et al., "Minimizing the average number of inspections for detecting rare items in finite populations", 2011 IEEE Computer Society, pp. 203-208.

* cited by examiner

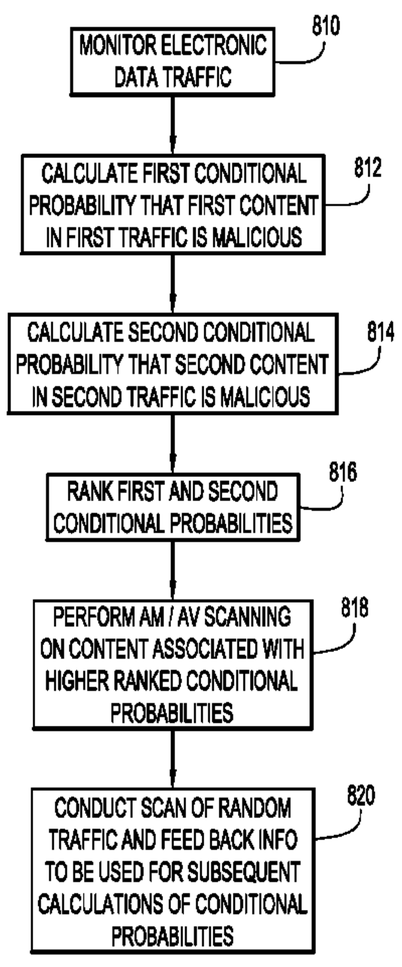
Primary Examiner — Michael Chao

(74) *Attorney, Agent, or Firm* — Edell, Shapiro & Finnan, LLC

(57) **ABSTRACT**

Systems and methods are provided that enable probabilistic application of data traffic scanning in an effort to catch malicious software or code being carried by the data traffic. The methodology and systems operate by monitoring data traffic in an data network via an interface with the data network, calculating a first conditional probability that content in first given data traffic being monitored is malicious, calculating a second conditional probability that content in second given data traffic being monitored is malicious, ranking the first and second conditional probabilities resulting in ranked conditional probabilities, and performing at least one of anti-virus (AV) or anti-malware (AM) scanning of the content of the first or second given data traffic depending on whose conditional probability is ranked higher in the ranked conditional probabilities.

13 Claims, 8 Drawing Sheets



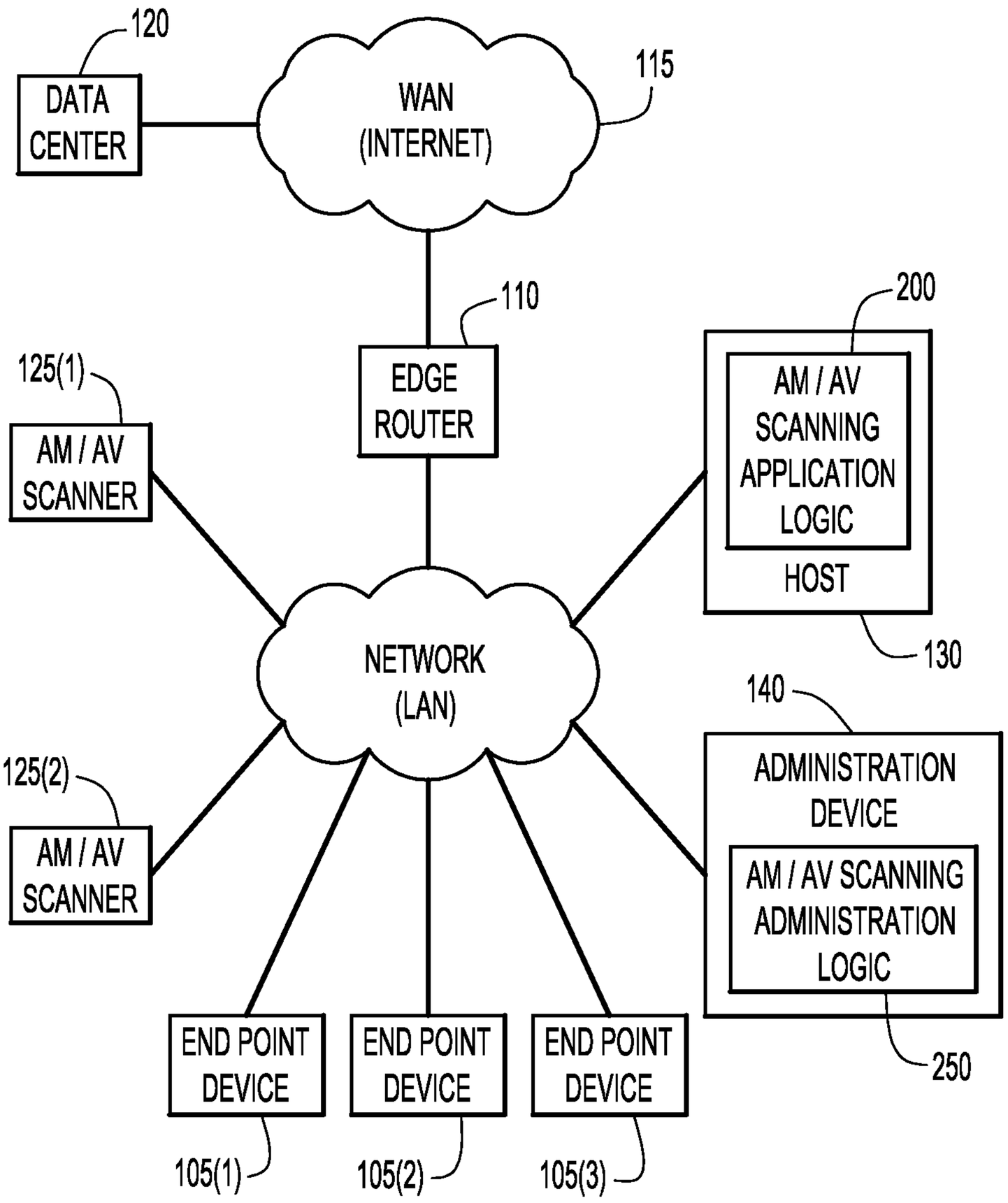

120
DATA CENTER
WAN (INTERNET)
115
110
EDGE ROUTER
125(1)
AM / AV SCANNER
200
AM / AV SCANNING APPLICATION LOGIC
HOST
130
NETWORK (LAN)
125(2)
AM / AV SCANNER
140
ADMINISTRATION DEVICE
AM / AV SCANNING ADMINISTRATION LOGIC
250
END POINT DEVICE
END POINT DEVICE
END POINT DEVICE
105(1)
105(2)
105(3)

FIG.1

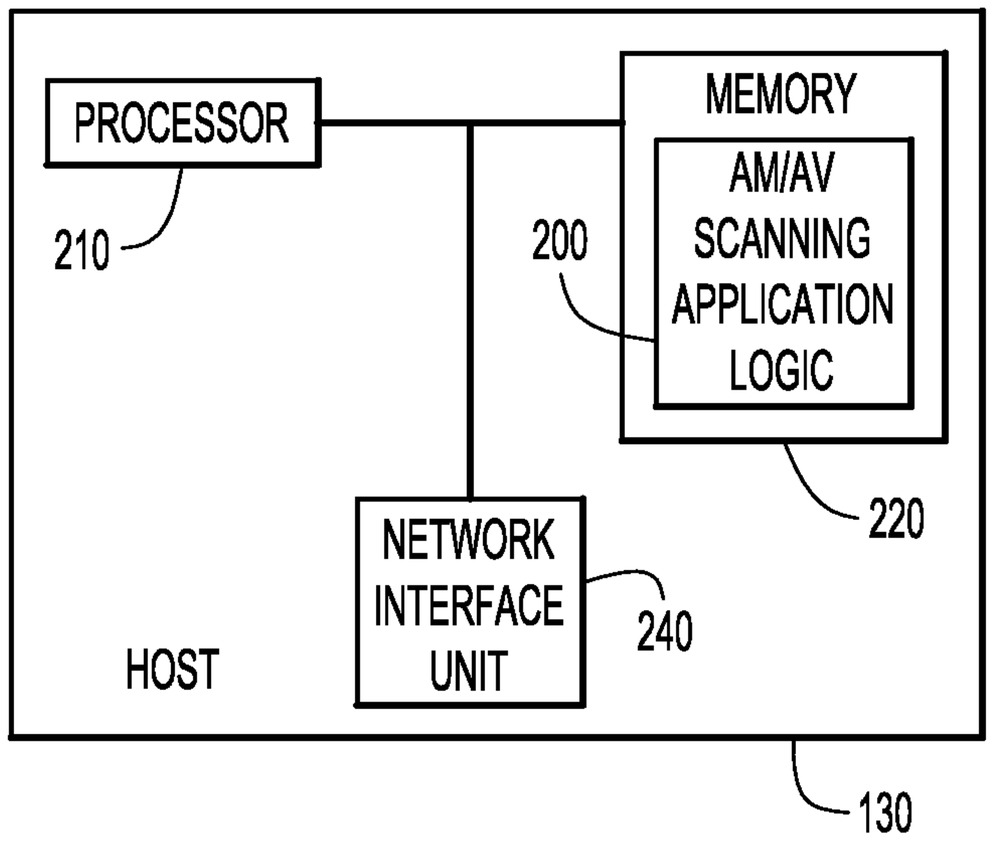

PROCESSOR
210
MEMORY
AM/AV
SCANNING
APPLICATION
LOGIC
200
220
NETWORK
INTERFACE
UNIT
240
HOST
130

FIG.2

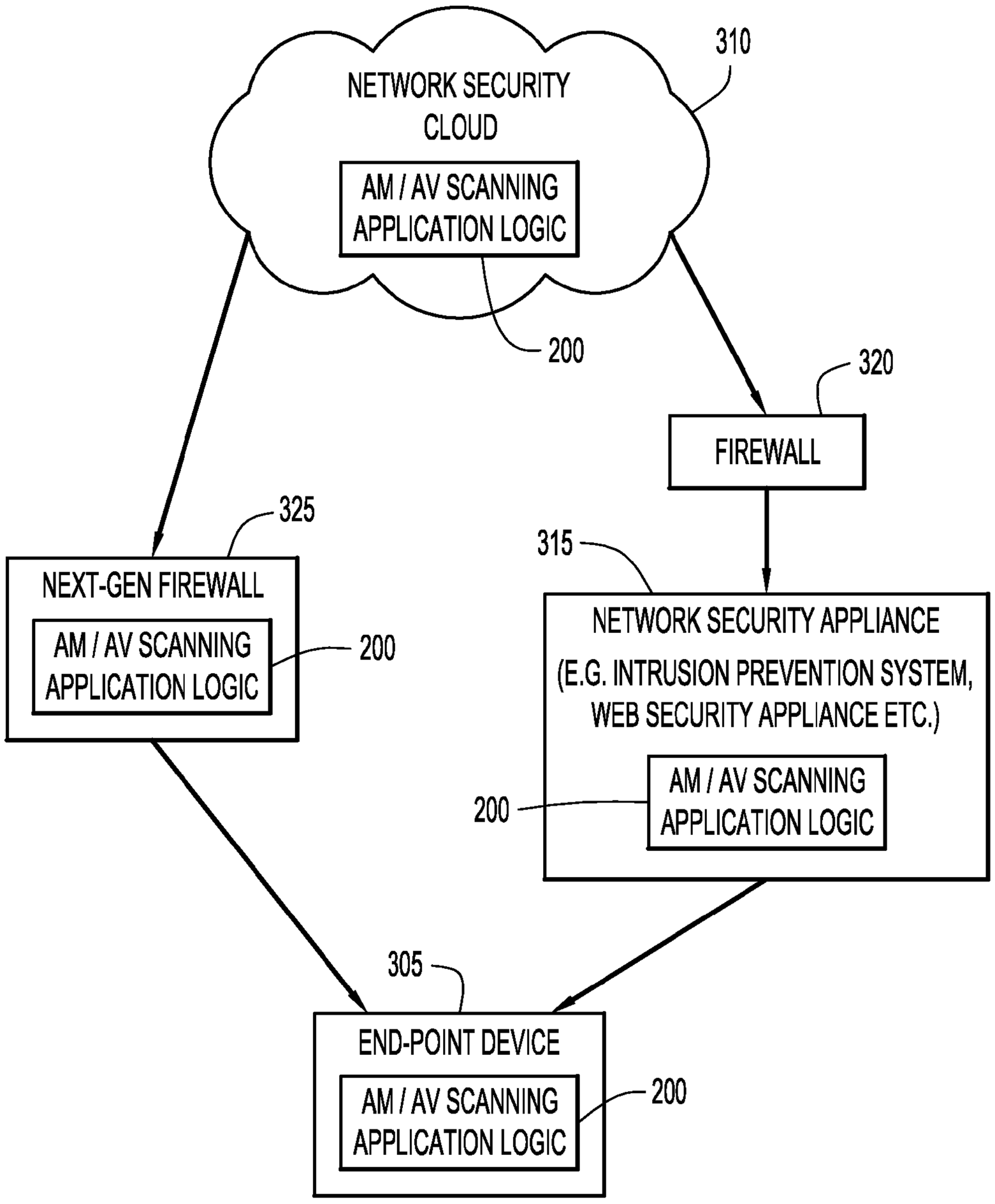

NETWORK SECURITY CLOUD
310
AM / AV SCANNING APPLICATION LOGIC
200
320
FIREWALL
325
NEXT-GEN FIREWALL
AM / AV SCANNING APPLICATION LOGIC
200
315
NETWORK SECURITY APPLIANCE
(E.G. INTRUSION PREVENTION SYSTEM, WEB SECURITY APPLIANCE ETC.)
200
AM / AV SCANNING APPLICATION LOGIC
305
END-POINT DEVICE
AM / AV SCANNING APPLICATION LOGIC
200

FIG.3

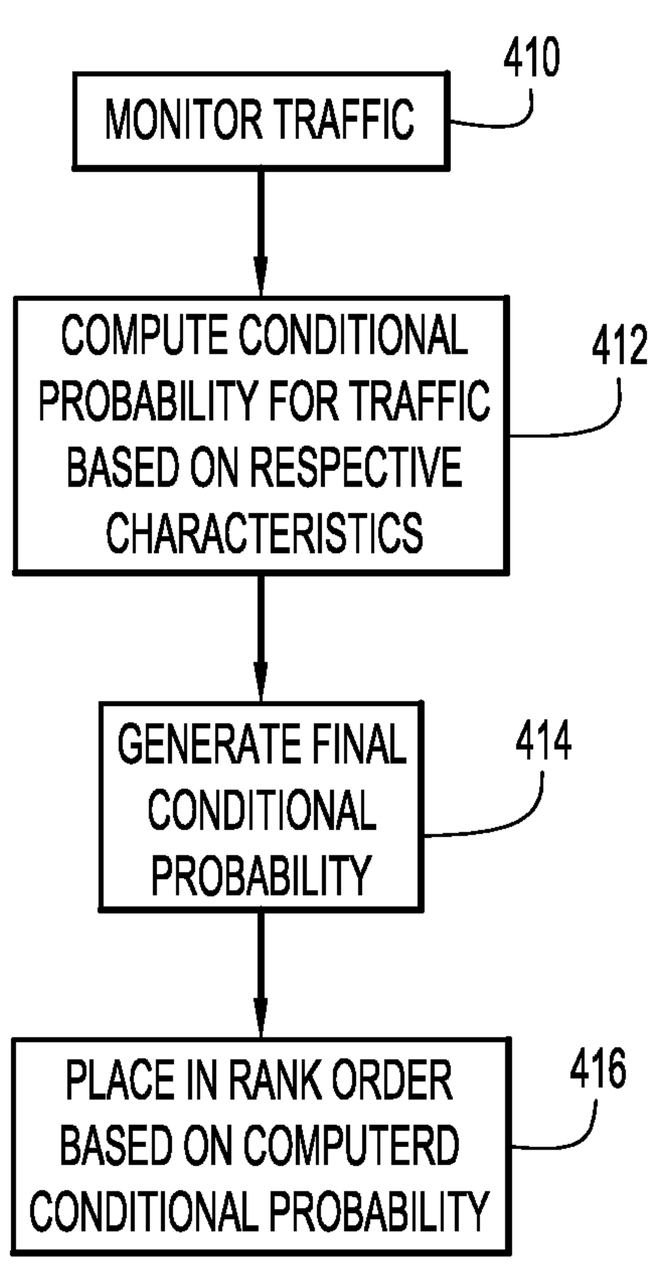

MONITOR TRAFFIC
410
COMPUTE CONDITIONAL PROBABILITY FOR TRAFFIC BASED ON RESPECTIVE CHARACTERISTICS
412
GENERATE FINAL CONDITIONAL PROBABILITY
414
PLACE IN RANK ORDER BASED ON COMPUTERD CONDITIONAL PROBABILITY
416

FIG.4

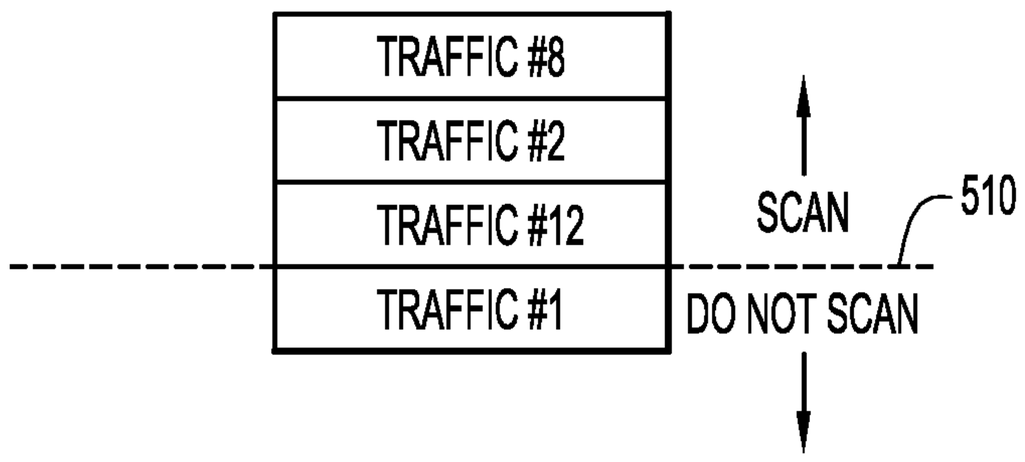


FIG.5

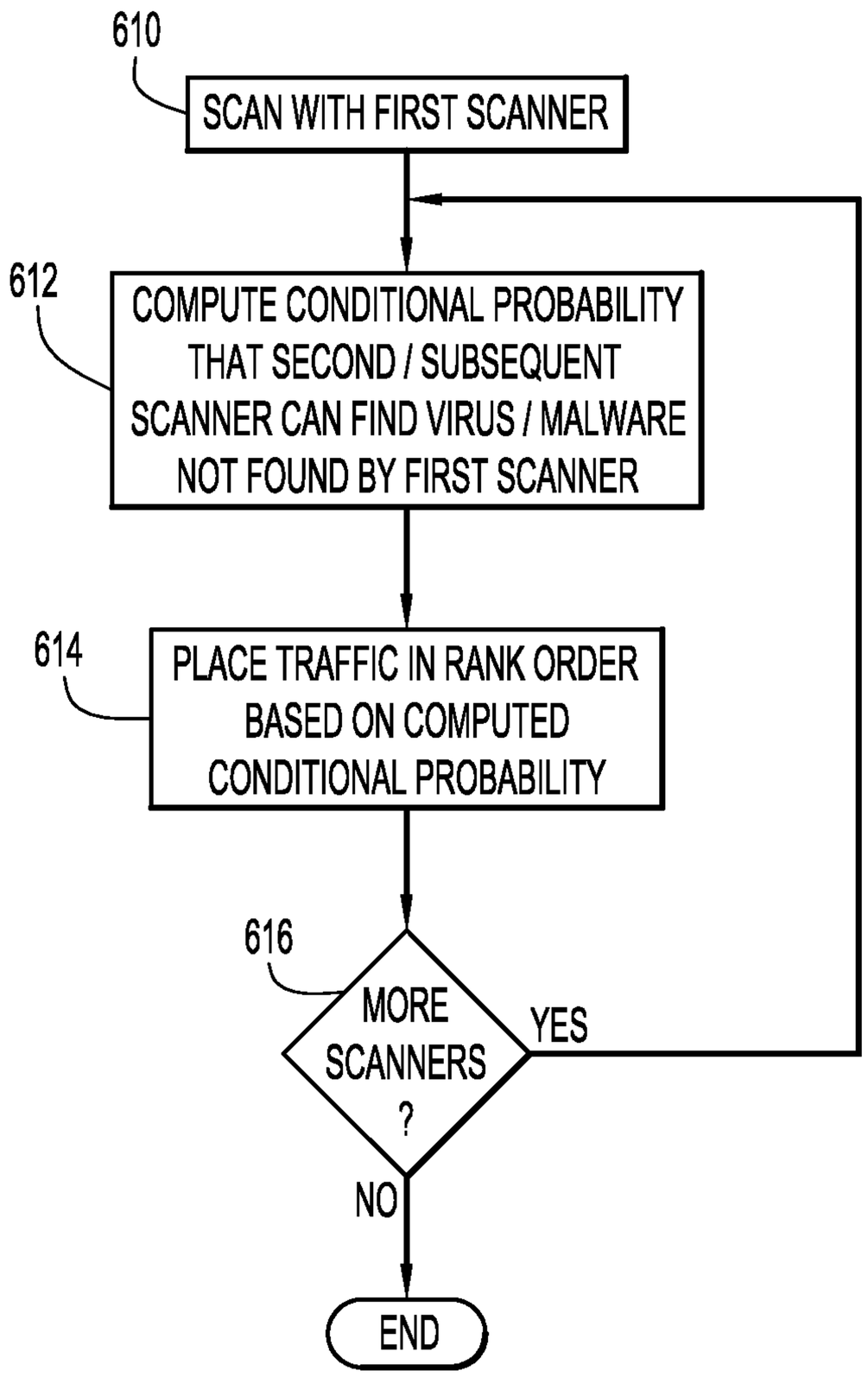

610
SCAN WITH FIRST SCANNER
612
COMPUTE CONDITIONAL PROBABILITY THAT SECOND / SUBSEQUENT SCANNER CAN FIND VIRUS / MALWARE NOT FOUND BY FIRST SCANNER
614
PLACE TRAFFIC IN RANK ORDER BASED ON COMPUTED CONDITIONAL PROBABILITY
616
MORE SCANNERS ?
YES
NO
END

FIG.6

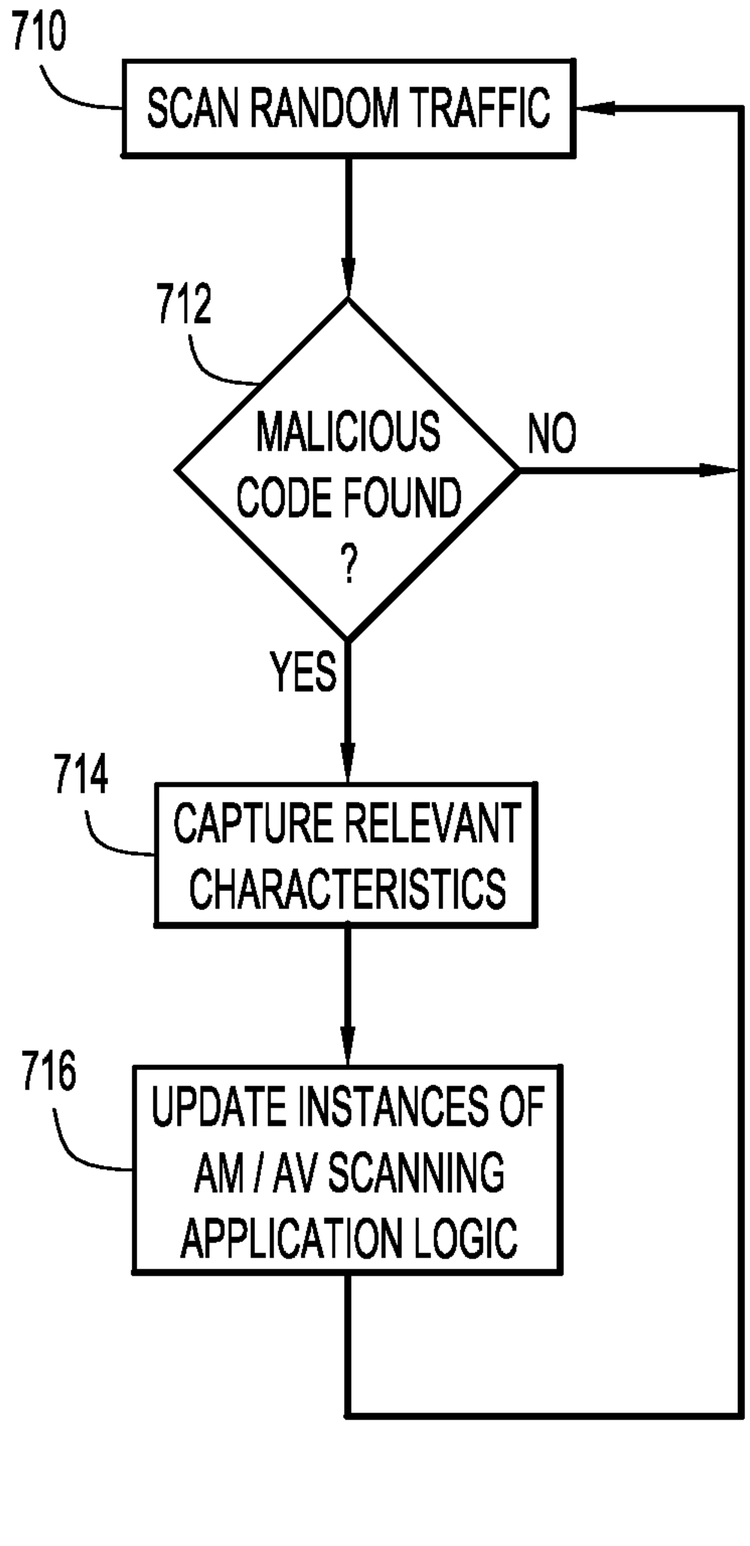

710
SCAN RANDOM TRAFFIC
712
MALICIOUS CODE FOUND ?
NO
YES
714
CAPTURE RELEVANT CHARACTERISTICS
716
UPDATE INSTANCES OF AM / AV SCANNING APPLICATION LOGIC

FIG.7

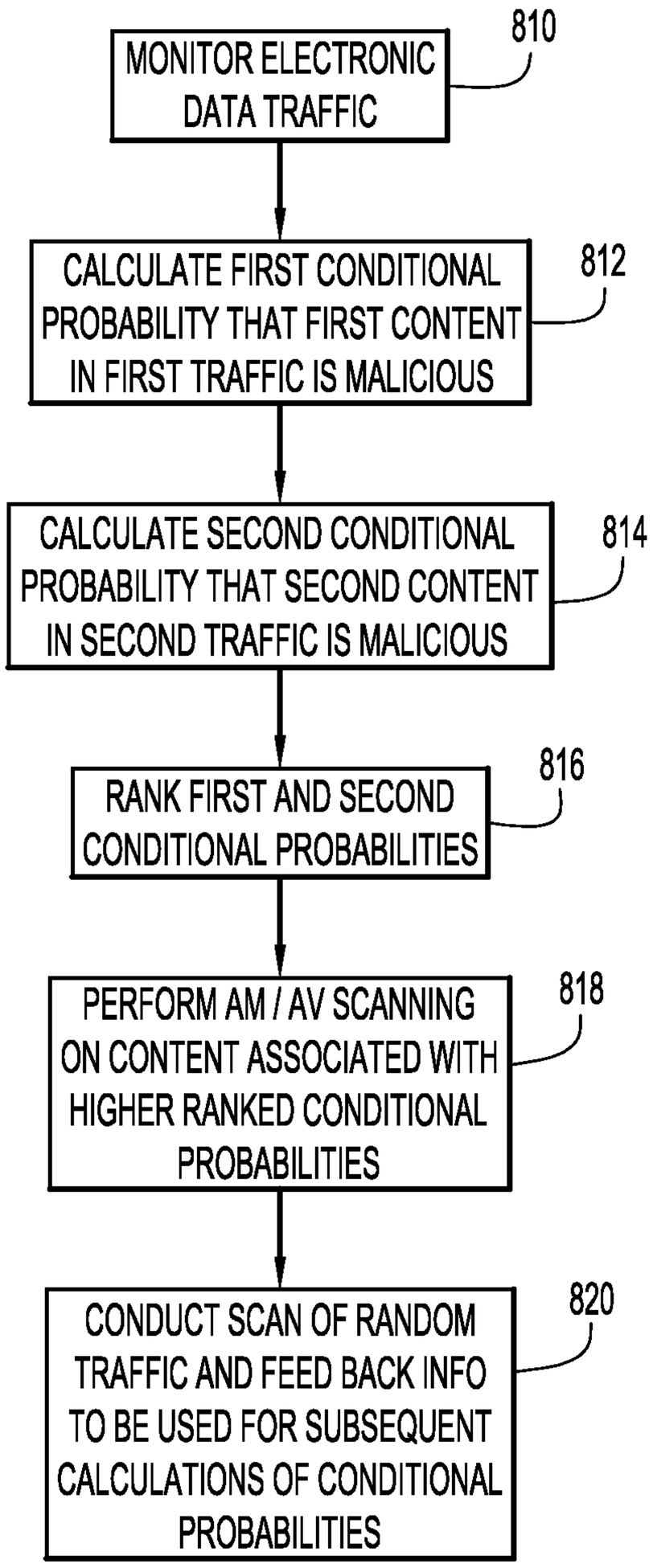

MONITOR ELECTRONIC DATA TRAFFIC
810
CALCULATE FIRST CONDITIONAL PROBABILITY THAT FIRST CONTENT IN FIRST TRAFFIC IS MALICIOUS
812
CALCULATE SECOND CONDITIONAL PROBABILITY THAT SECOND CONTENT IN SECOND TRAFFIC IS MALICIOUS
814
RANK FIRST AND SECOND CONDITIONAL PROBABILITIES
816
PERFORM AM / AV SCANNING ON CONTENT ASSOCIATED WITH HIGHER RANKED CONDITIONAL PROBABILITIES
818
CONDUCT SCAN OF RANDOM TRAFFIC AND FEED BACK INFO TO BE USED FOR SUBSEQUENT CALCULATIONS OF CONDITIONAL PROBABILITIES
820

FIG.8

DYNAMIC AND ADAPTIVE TRAFFIC SCANNING

TECHNICAL FIELD

The present disclosure relates to electronic data network security.

BACKGROUND

Computer systems, networks and data centers are exposed to a constant and differing variety of attacks that expose vulnerabilities of such systems in order to compromise their security and/or operation. As an example, various forms of malicious software program attacks include viruses, worms, Trojan horses and the like that computer systems can obtain over a network such as the Internet. Quite often, users of such computer systems are not even aware that such malicious programs have been obtained within the computer system. Once resident within a given computer, a malicious program that executes might disrupt operation of that computer to a point of inoperability and/or might spread itself to other computers within a network or data center by exploiting vulnerabilities of the computer's operating system or resident application programs. Other malicious programs, such as "Spyware," might operate within a computer to secretly extract and transmit information within the computer to remote computer systems for various suspect purposes.

To combat the proliferation of malicious software program attacks, Anti-Malware (AM) and Anti-Virus (AV) scanners have been widely deployed in different security appliances and Intrusion Prevention Systems (IPSs) to detect known malicious software, and to thereafter block or filter such threats. Compared with, e.g., universal resource locator (URL) and reputation-based malware filtering, which rely broadly on the source of content by monitoring, e.g., an IP address or a domain name, AM/AV scanning may enjoy a lower false positive rate by employing, e.g., well-defined signatures on the actual content that is responsible for malicious behavior. On the other hand, AM/AV scanning can be expensive in terms of central processing unit (CPU) and memory usage which, in many instances, especially as electronic data network traffic grows at increasing rates, makes it impractical to apply AM/AV scanning to all of the content being carried by network traffic, or even destined for a single given computer within that network.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** depicts an example architecture including the deployment of AM/AV scanning application logic and AM/AV scanners.

FIG. **2** depicts an example host device on which AM/AV scanning application logic may be deployed.

FIG. **3** depicts a schematic diagram of the locations at which AM/AV scanning application logic can be deployed.

FIG. **4** is a flow chart depicting example operations for placing electronic data traffic in a ranked order according to computed conditional probabilities.

FIG. **5** illustrates an example rank list based on final conditional probabilities calculated for multiple segments or portions of electronic data traffic

FIG. **6** is a flow chart depicting example operations for selecting a second or subsequent scanner to scan electronic data traffic.

FIG. **7** is a flow chart depicting example operations for updating instances of AM/AV scanning application logic based on randomly sampled and scanned electronic data traffic.

FIG. **8** is flow chart depicting example operations for implementing a probabilistic methodology for AM/AV scanning.

DESCRIPTION OF EXAMPLE EMBODIMENTS

Overview

Systems and methods are described that enable probabilistic application of data traffic scanning in an effort to catch malicious software or code being carried by the data traffic. The methodology and systems operate by monitoring data traffic in a data network via an interface with the electronic data network, calculating a first conditional probability that content in first given data traffic being monitored is malicious, calculating a second conditional probability that content in second given data traffic being monitored is malicious, ranking the first and second conditional probabilities resulting in ranked conditional probabilities, and performing at least one of anti-virus (AV) or anti-malware (AM) scanning of the content of the first or second given data traffic depending on whose conditional probability is ranked higher in the ranked conditional probabilities. In one embodiment, the functionality for computing or determining the conditional probabilities is embodied in computer executable instructions or code that is hosted by a network device or appliance, which can be deployed in one of several different locations in a network topology.

Example Embodiments

Embodiments described herein include a computer or other processing system configured to execute a probabilistic application of Anti-Malware/Anti-Virus (AM/AV) scanning. The number of security threats introduced by electronic network (e.g., world wide web) traffic is reaching epidemic proportions. Traditional gateway defenses are proving inadequate against a variety of malware, leaving corporate electronic data networks exposed. The speed, variety and damage potential of malware attacks highlight the utility of a robust, secure platform to protect the enterprise network perimeter or, at the very least, individual end point or end user devices within that perimeter.

FIG. **1** illustrates an example computer networking environment **100** suitable for use in explaining example embodiments disclosed herein. The computer networking environment **100** includes a computer network **101** such as a local area network (LAN) that interconnects multiple end point devices **105**(**1**), **105**(**2**), **105**(**3**), etc. End point devices **105** may be wired or wirelessly connected desktop computers or laptop computers, mobile devices including mobile telephones and tablets, or like devices that provide some sort of user experience to an end user as a result of receiving data over the network.

Network **101** is also connected to an edge router **110** that couples network **101** to a wide area network (WAN) **115** such as the Internet and that allows communication between the end point devices **105** and other computers, servers and other devices worldwide. Such other computers, servers, etc. (not shown) may be disposed, for example, inside a data center **120** that may be a physical data center or a virtual data center that functions, from the perspective of an end point device **105**, as a dedicated physical data center.

Also shown in FIG. **1** and interconnected by network **101** are AM/AV scanners **125**(**1**), **125**(**2**), the function of which will be explained in more detail later herein. A host **130** is also shown connected to network **101**. Host **130** could be a stand alone computer or other network device, or could be any one of the network connected devices **105**, **110**, **120**, **125** among others. Host **130** hosts, stores or otherwise incorporates AM/AV scanning application logic **200** in the form of, e.g., computer executable instructions or code. As will be explained in detail below, AM/AV scanning application logic **200** is configured to select which electronic data traffic traversing network paths within computer architecture **100** is to be scanned for viruses/malware. Logic **200** may be further configured to select which one of a plurality of AM/AV scanners is to be employed as well as an order by which selected AM/AV scanners may be applied to given electronic data traffic.

In addition, yet another device shown in computer architecture **100** is an administration device **140**, which hosts AM/AV scanning administration logic **250**. Administration device **140** may be a stand alone computer or device as shown or may be incorporated into any one of the network connected devices shown in FIG. **1**. As will be explained below, AM/AV scanning administration logic **250** may be used to control or configure the operation of AM/AV scanning application logic **200** to change the way in which electronic data traffic traversing networks paths within computer architecture **100** may be processed by AM/AV scanners **125**.

FIG. **2** depicts an example host **130** on which AM/AV scanning application logic **200** may be deployed. Host **130** may comprise a processor **210**, associated memory **220** (a tangible medium), which may include AM/AV scanning application logic **200**, and a network interface unit **240** such as a network interface card, which enables host **130** to communicate externally with other devices via network **101**. Although not shown, host **130** may also include input/output devices such as a keyboard, mouse and display (not shown) to enable direct control of host **130** by a user. Those skilled in the art will appreciate that host **130** may be a rack mounted device, such as a blade in a server rack, and that may not have dedicated respective input/output devices. Instead, such a rack mounted device might be accessible via a centralized console, or some other (even remote) arrangement by which host **130** can be accessed, controlled or configured by, e.g., a user, perhaps using AM/AV administration logic **250**. It is noted that administration device **140** may be configured similarly to host **130** and store in memory thereof AM/AV administration logic **250**.

FIG. **3** depicts a schematic diagram of several locations at which instances of AM/AV scanning application logic **200** can be deployed. As computer architecture continues to mature, there may be advantages in terms of cost, efficiency, processing capacity, among other considerations, to host AM/AV scanning application logic **200** in different locations. For example, AM/AV scanning application logic **200** may be deployed, in accordance with one possible embodiment, within what is referred to as the "cloud" or, in this case, network security cloud **310**. This cloud may physically map, for example, to data center **120** shown in FIG. **1**. AM/AV scanning application logic **200** may also be hosted by a network security appliance **315** that is logically located between an end point device **305** (like **105** in FIG. **1**) and a firewall **320**. Network security appliance **315** may be configured as an intrusion detection system, web security appliance, or other like computing component. In still another embodiment, AM/AV scanning application logic **200** may be deployed on what may be referred to as a "next-generation" firewall **325**. Such a component might perform the functions of both network security appliance **315** and firewall **320**. Finally, in yet another embodiment, AM/AV scanning application logic **200** may be deployed directly on end point devices **305** (**105**). Again, precisely where in a given computer network architecture AM/AV scanning application logic **200** is deployed may be a function of one or more factors.

In the same way, although AM/AV scanners **125** shown in FIG. **1** are depicted as stand alone components, the functionality of AM/AV scanners **125** may be incorporated into any of the components depicted in FIG. **3**. That is, AM/AV scanners **125** could be incorporated into network security cloud **310**, next-generation firewall **325**, network security appliance **315** or end point device **305**. Thus, those skilled in the art will appreciate that AV/AM scanning application logic **200** and AM/AV scanners **125** may be hosted by a same physical component or be hosted by different network elements within network architecture **100**. Likewise, AM/AV scanning administration logic **200** could also be hosted separately as shown in FIG. **1** or be hosted on any of the components shown in FIG. **3**.

The functionality of AM/AV scanning application logic **200** will now be described. At a high level, a main function of AM/AV scanning application logic **200** is to identify, based on probabilistic methodologies, which electronic data traffic is to be scanned by a given one of the AM/AV scanners **125**, and where appropriate, to determine, in a probabilistic manner, which additional AM/AV scanner or scanners **125** may also be applied to the electronic data traffic after a given first one of the AM/AV scanners **125** has completed its scan operations.

A typical AM/AV scanner can accommodate approximately 100 scanning requests per second, i.e., a request to scan content carried by given electronic data traffic. In today's network environment, however, in order to effectively scan substantially all electronic data traffic, a scanner would have to support potentially thousands of requests per second. To address this problem, some scanning implementations scan only higher risk but lower volume file types, such as executable (.exe.) and zipped (.zip) file types. However, the trend is that malware and viruses are increasingly embedded in many other file types including hyper text markup language (HTML) files, JavaScript files, portable document format (PDF) documents and other image format files

To address the forgoing, AM/AV scanning application logic **200** provides a dynamic electronic data traffic scanning approach that determines which high-risk traffic to scan based on a probabilistic prioritization, selects which AM/AV scanner(s) **125** to user if there are two or more AM/AV scanners available for use, and monitors overall system load to dynamically scan more content when a given one or more of AM/AV scanners **125** may be idle. AM/AV scanner application logic **200** may also periodically cause AM/AV scanners **125** to scan random electronic data traffic, determine whether that traffic includes malicious software programs, and then based on results of the determination, provide a feedback mechanism by which future probabilistic prioritization techniques can be tuned or optimized.

More specifically, probabilistic prioritization, in accordance with principles of embodiments described herein, is based on conditional probability. In the case of identifying malicious software in electronic data traffic, the conditional probability of an event B (i.e., existence of malicious software programs) is the probability that the event will occur given the knowledge that an event A (i.e., selected characteristics of the electronic data traffic are present) has already occurred. This probability is written as P(B|A) and referred to as "the prob-

ability of B given A." In the instant context, this would be referred to as the probability of finding malicious code or software in given electronic data traffic given the attributes or characteristics of that electronic data traffic.

If events A and B are not independent, then the probability of the intersection of A and B (the probability that both events occur) is defined by P(A and B)=P(A)P(B|A). From this definition, the conditional probability P(B|A) is obtained by dividing by P(A):

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Event "A" in the context of characteristics of the electronic data traffic might include, for example, the URL from which the electronic data traffic is being received, and/or a corresponding reputation thereof, a category of the content being received (e.g., gambling pages, or cooking recipes), or type of file be carried by the electronic data traffic (e.g., .exe, .pdf, etc.), a user-agent type (e.g., an application used to request the content), geographical of either or both a server from which the electronic data traffic is being received or a user or end point device that is to receive the electronic data traffic, among many other possibilities.

In one implementation, AM/AV scanning application logic **200** has a set of a priori knowledge about the likelihood (probability) of given electronic data traffic carrying malicious software given certain characteristics of that given electronic data traffic. As the traffic passes through host **130**, AM/AV scanning application logic **200** operates on or processes the traffic and generates a conditional probability that the traffic is carrying malicious software. It is noted that there may be multiple characteristics that AM/AV scanning application logic **200** considers in determining the conditional probability. As such, a final conditional probability that given electronic data traffic carries malicious software may be a combination of multiple conditional probabilities.

FIG. **4** is a flow chart depicting example operations for placing electronic data traffic in a ranked order according to computed conditional probability. At **410**, the electronic data traffic is monitored. At **412**, conditional probabilities that given electronic data traffic contains or carries malicious software are calculated in connection with respective characteristics of the electronic data traffic (e.g., URL/reputation, content category, file type, etc.). At **414**, the calculated conditional probabilities are added or otherwise aggregated to obtain a final conditional probability for the given electronic data traffic. At **416**, an entry is made in, e.g., a table or list stored in memory wherein the entry is placed in rank order in the list.

FIG. **5** illustrates an example ranked list of electronic data traffic based on final conditional probabilities calculated for respective segments or portions of electronic data traffic. In the example shown, the portions are ranked in order as #8, #2, #12 and #1. These portions may represent, e.g., individual sessions, individual packets, collections of packets, or any segmentation of the overall traffic that transits a host on which AM/AV scanning application logic **200** resides. As shown in the drawing, traffic portions #8, #2 and #12 are ranked near the top of the list. In this case, and as shown by broken line **510**, those three portions will be scanned by one or more scanners **125**, whereas portion #1 will not be scanned at all, as the conditional probability that portion #1 contains a virus or malware is relatively lower than the other three portions. By ranking segments of electronic data traffic in this way, it is possible to optimize which traffic is scanned since it may not be efficient (or even possible) to scan all such traffic.

That is, rather than setting hard thresholds or limiting scanning of content carried by electronic data traffic to certain types of files, embodiments described herein rank the electronic data traffic based on a plurality of characteristics and resulting conditional probabilities. Scanning can then take place on the electronic data traffic (i.e., the content therein) with the highest rankings. In this way, scanner **125** can be put to use in the a more optimized way by passing content through the scanners that is determined to more likely carry malicious software.

FIG. **6** is a flow chart depicting example operations for selecting a second or subsequent scanner to scan electronic data traffic. Preliminarily, it is noted that scanners can be configured or tuned in such a way to be better at catching or identifying malicious code in different types of files. For instance, one type of scanner might be better at catching malicious code in executable files, while another scanner might be better at catching malicious code in a PDF type file. Thus, as an initial matter, one embodiment of AM/AV scanning application logic **200** selects which scanner from among possibly multiple scanners should be used to scan any given electronic data traffic. Not only can the selection be based on the type of traffic that is to be analyzed or scanned, but the selection of a given scanner might also be based on its availability at that moment to process the traffic. It should be kept in mind that the methodologies being described herein may be operating in real-time or near-real-time contexts, such that end users might detect latency should scanning requests become backlogged. By selecting available scanners **125**, such latency can be reduced.

Thus, in accordance with an embodiment, AM/AV scanning application logic **200** first selects an available and appropriate scanner **125** to employ for a first pass of scanning. Then, if another scanner might be available and it is determined that such a scanner might catch malicious software that the first scanner did not detect, then AM/AV scanning application logic **200** may further cause the electronic data traffic being processed to be processed by a second or even third, etc. scanner.

Still with reference to FIG. **6**, using AM/AV scanning application logic **200**, at **610**, the electronic data traffic is scanned with a first scanner. At **612**, a conditional probability that a second or subsequent scan by another scanner can find a virus or malware beyond that already found by the first scan is computed. The particular traffic may then be placed in ranked order similar to the ranking shown in FIG. **5**. At **616**, it is determined whether more scanners might be available for subsequent scanning of the same electronic data traffic. If no, the operation ends. If yes, another conditional probability is computed with respect to any such subsequent scanner, as indicated by **612**.

With the additional scanner(s) so selected, the same electronic data traffic can be passed through the additional scanner(s) such that an overall catch rate of malicious code can be improved.

To keep the system and methodology relevant and updated over time, a periodic feedback mechanism is employed. FIG. **7** shows example operations for updating multiple instances of AM/AV scanning application logic **200** such that the conditional probabilities that are computed can be as accurate as practicable.

Referring to FIG. **7**, and at **710**, random electronic data traffic is scanned on a periodic basis. The period can be on the order of minutes, hours, days, or weeks, etc. AM/AV scanning administration logic **250** can be used to set the periodicity of

the random scanning of **710**. Likewise, the amount of traffic that is scanned can also be configured via AM/AV scanning administration logic **250**. Random single packets, groups of packets, parts or complete sessions may be selected for scanning.

At **712**, it is determined whether malicious software has been found via the scanning procedure. If not, operation **710** is repeated. If malicious code is found, then at **714**, relevant characteristics of the electronic data traffic are captured. These characteristics might include, e.g., a URL, the category of the content being carried or the type of file be carried, among many other possible characteristics. At **716** this information (namely that malicious software was (or was not) found in electronic data traffic having the captured characteristics) is broadcast or otherwise made available to other instances of AM/AV scanning application logic **200** that may be operating within other network architectures. In this way, different instances of AM/AV scanning application logic **200** can share information with one another, enabling individual instances to be as up to date as may be practicable so that the conditional probabilities that are calculated or as accurate as possible.

Reference has been made to portions or segments of electronic data traffic. In the context of the embodiments described herein, those segments or portions can be based on information gleaned from layers 3-7 of Open Systems Interconnection (OSI) model wherein the headers and payloads of various packets or frames of the electronic data traffic at these several layers are inspected for maliciousness and for the characteristics that are used to compute the conditional probabilities.

Thus, as explained, electronic data traffic transiting a security appliance or host is dynamically selected to be scanned based on a function of the probability that the traffic contains malicious software or code. In one embodiment, scanners are employed at their maximum throughput such that the greatest amount of electronic data traffic can be scanned without substantially impacting an expected quality of service by an end user. Scanners can also be used in sequence in an attempt to capture as much malicious code as possible. Periodically, a random sample of electronic data traffic is scanned by one or more scanners and information gleaned from such scans is distributed to network appliances that operate in accordance with the principles described herein.

Reference is now made to FIG. **8**, which depicts example operations for performing conditional probabilities-based AM/AV scanning according to the principles described herein.

Operation **810** includes monitoring electronic data traffic in an electronic data network via an electronic interface with the electronic data network. Operation **812** includes calculating a first conditional probability that content in first given electronic data traffic being monitored is malicious. Operation **814** includes calculating a second conditional probability that content in second given electronic data traffic being monitored is malicious.

Operation **816** includes ranking the first and second conditional probabilities resulting in ranked conditional probabilities, and operation **818** includes performing at least one of anti-virus (AV) or anti-malware (AM) scanning of the content of the first or second given electronic data traffic depending on whose conditional probability is ranked higher in the ranked conditional probabilities.

Operation **820** includes performing at least one of AV or AM scanning of content of third given electronic data traffic, determining, as a result of the AV or AM scanning, whether the content of the third given electronic data traffic is malicious, and when the content of the third given electronic data traffic is determined to be malicious, capturing characteristics of the third given electronic data traffic, wherein the characteristics are employed to calculate conditional probabilities that content of still other given electronic data traffic is malicious

Operations may further include selecting a first one of a plurality of scanners for performing the at least one of AV or AM scanning, wherein selecting is based on an efficacy value associated with respective ones of the plurality of scanners for a given type of content that is being carried by the first or second given electronic data traffic whose conditional probability is ranked higher in the ranked conditional probabilities.

Operations may still further comprise selecting a second one of the plurality of scanners for a performing a subsequent operation of performing at least one of AV or AM scanning after scanning by the first one of the plurality of scanners, wherein selecting of the second one of the plurality of scanners is based on a catch rate of malicious content not caught by the first one of the plurality of scanners.

In one embodiment, the functionality for calculating the conditional probabilities and controlling which scanners are to be employed for scanning may be deployed at a network security appliance that is logically disposed between an end-point device and a firewall that is in communication with the Internet. This functionality may also be deployed with a cloud computing environment.

Although the apparatus, system and method are illustrated and described herein as embodied in one or more specific examples, it is nevertheless not intended to be limited to the details shown, since various modifications and structural changes may be made therein without departing from the scope of the apparatus, system, and method and within the scope and range of equivalents of the claims. A Data Center can represent any location supporting capabilities enabling service delivery that are advertised. A Provider Edge Routing Node represents any system configured to receive, store or distribute advertised information as well as any system configured to route based on the same information. Accordingly, it is appropriate that the appended claims be construed broadly and in a manner consistent with the scope of the apparatus, system, and method, as set forth in the following.

What is claimed is:

1. A method comprising:

monitoring data traffic in a data network via an electronic interface with the data network, wherein monitoring comprises monitoring data packets traversing the data network at an appliance that is logically disposed between an end-point device and the Internet;

calculating a first conditional probability that content in first given data traffic being monitored is malicious;

calculating a second conditional probability that content in second given data traffic being monitored is malicious;

ranking the first and second conditional probabilities with respect to each other resulting in ranked conditional probabilities; and

performing at least one of anti-virus or anti-malware scanning of the content of the first or second given data traffic depending on whose conditional probability is ranked higher in the ranked conditional probabilities to the exclusion of the other of the first or second given data traffic;

selecting a first one of a plurality of scanners for performing the at least one of anti-virus or anti-malware scanning, wherein selecting is based on an efficacy value associated with respective ones of the plurality of scan-

ners for a given type of content that is being carried by the first or second given data traffic whose conditional probability is ranked higher in the ranked conditional probabilities;

selecting a second one of the plurality of scanners for a performing a subsequent operation of performing at least one of anti-virus or anti-malware scanning after scanning by the first one of the plurality of scanners, wherein selecting of the second one of the plurality of scanners is based on a conditional probability that the second one of the plurality of scanners can catch malicious content not caught by the first one of the plurality of scanners,

wherein a same content of the first or second given data traffic is processed by the first one of a plurality of scanners and the second one of a plurality of scanners, and

wherein selecting the first one of a plurality of scanners is based on a data type of the content of the first or second given data traffic.

2. The method of claim **1**, wherein monitoring data traffic comprises monitoring data packets at a network security appliance that is logically disposed between an end-point device and a firewall that is in communication with the Internet.

3. The method of claim **1**, further comprising:

performing at least one of anti-virus or anti-malware scanning of content of third given data traffic;

determining, as a result of the anti-virus or anti-malware scanning, whether the content of the third given data traffic is malicious; and

when the content of the third given data traffic is determined to be malicious, capturing characteristics of the third given data traffic, wherein the characteristics are employed to calculate conditional probabilities that content of still other given data traffic is malicious.

4. The method of claim **3**, wherein capturing characteristics comprises capturing a universal resource locator from which the third given data traffic was received.

5. The method of claim **1**, further comprising selecting the first one of a plurality of scanners for performing the at least one of anti-virus or anti-malware scanning based on available throughput of respective ones of the plurality of scanners.

6. An apparatus comprising:

a network interface unit configured to communicate over a data network;

a memory; and

a processor configured to:

monitor electronic data traffic in the data network via the interface, by monitoring data packets traversing the data network when the appliance is logically disposed between an end-point device and the Internet;

calculate a first conditional probability that content in first given data traffic being monitored is malicious;

calculate a second conditional probability that content in second given data traffic being monitored is malicious;

rank the first and second conditional probabilities with respect to one another resulting in ranked conditional probabilities; and

cause at least one of anti-virus or anti-malware scanning of the content of the first or second given data traffic to be performed depending on whose conditional probability is ranked higher in the ranked conditional probabilities to the exclusion of the other of the first or second given data traffic,

wherein the processor is further configured to select a first one of a plurality of scanners for performing the at least one of anti-virus or anti-malware scanning based on an efficacy value associated with respective ones of the plurality of scanners for a given type of content that is being carried by the first or second given electronic data traffic whose conditional probability is ranked higher in the ranked conditional probabilities,

wherein the processor is further configured to select a second one of the plurality of scanners for performing a subsequent operation of performing at least one of anti-virus or anti-malware scanning after scanning by the first one of the plurality of scanners, wherein the processor is configured to so select the second one of the plurality of scanners based on a conditional probability that the second one of the plurality of scanners can catch rate malicious content not caught by the first one of the plurality of scanners,

wherein a same content of the first or second given data traffic is processed by the first one of a plurality of scanners and the second one of a plurality of scanners, and

wherein the first one of a plurality of scanners is selected based on a data type of the content of the first or second given data traffic.

7. The apparatus of claim **6**, wherein the processor is further configured to:

cause at least one of anti-virus or anti-malware scanning of content of third given data traffic to be performed;

determine, as a result of the anti-virus or anti-malware scanning, whether the content of the third given data traffic is malicious; and

when the content of the third given electronic data traffic is determined to be malicious, capture characteristics of the third given data traffic, wherein the characteristics are employed to calculate conditional probabilities that content of still other given data traffic is malicious.

8. The apparatus of claim **7**, wherein the capture of characteristics comprises capturing a universal resource locator from which the third given data traffic was received.

9. The apparatus of claim **6**, wherein the processor is further configured to select the first one of a plurality of scanners for performing the at least one of anti-virus or anti-malware scanning based on available throughput of respective ones of the plurality of scanners.

10. One or more non-transitory computer readable storage media encoded with software comprising computer executable instructions and when the software is executed operable to:

monitor data traffic in a data network via an interface;

calculate a first conditional probability that content in first given data traffic being monitored is malicious;

calculate a second conditional probability that content in second given data traffic being monitored is malicious;

rank the first and second conditional probabilities with respect to one another resulting in ranked conditional probabilities; and

cause at least one of anti-virus or anti-malware scanning of the content of the first or second given data traffic to be performed depending on whose conditional probability is ranked higher in the ranked conditional probabilities to the exclusion of the other of the first or second given data traffic,

wherein the instructions are further operable to select a first one of a plurality of scanners for performing the at least one of anti-virus or anti-malware scanning based on an efficacy value associated with respective ones of the plurality of scanners for a given type of content that is being carried by the first or second given electronic data

traffic whose conditional probability is ranked higher in the ranked conditional probabilities,

wherein the instructions are further operable to select a second one of the plurality of scanners for performing a subsequent operation of performing at least one of anti-virus or anti-malware scanning after scanning by the first one of the plurality of scanners, wherein the processor is configured to so select the second one of the plurality of scanners based on a conditional probability that the second one of the plurality of scanners can catch malicious content not caught by the first one of the plurality of scanners,

wherein a same content of the first or second given data traffic is processed by the first one of a plurality of scanners and the second one of a plurality of scanners, and

wherein the first one of a plurality of scanners is selected based on a data type of the content of the first or second given data traffic.

11. The computer readable storage media of claim **10**, wherein the instructions are further operable to:

cause at least one of anti-virus or anti-malware scanning of content of third given data traffic to be performed;

determine, as a result of the anti-virus or anti-malware scanning, whether the content of the third given data traffic is malicious; and

when the content of the third given electronic data traffic is determined to be malicious, capture characteristics of the third given data traffic, wherein the characteristics are employed to calculate conditional probabilities that content of still other given data traffic is malicious.

12. The computer readable storage media of claim **10**, wherein the instructions are further operable to capture a universal resource locator from which the third given data traffic was received.

13. The computer readable storage media of claim **10**, wherein the instructions are further operable to select the first one of a plurality of scanners for performing the at least one of anti-virus or anti-malware scanning based on available throughput of respective ones of the plurality of scanners.

* * * * *