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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 October 2012 (26.10.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/145066 A1

(51) International Patent Classification:
G06F 11/00 (2006.01)

(21) International Application Number:
PCT/US2012/026402

(22) International Filing Date:
23 February 2012 (23.02.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/089,191 18 April 2011 (18.04.2011) US

(71) Applicant *(for all designated States except US)*: **FIREEYE, INC.** [US/US]; 1390 McCarthy Blvd., Milpitas, California 95035 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **AZIZ, Ashar** [US/US]; FireEye, Inc., 1390 McCarthy Blvd., Milpitas, California 95035 (US). **UYENO, Henry** [US/US]; FireEye, Inc., 1390 McCarthy Blvd., Milpitas, California 95035 (US). **MANNI, Jay** [IN/US]; FireEye, Inc., 1390 McCarthy Blvd., Milpitas, California 95035 (US). **SUKHERA, Amin** [US/US]; FireEye, Inc., 1390 McCarthy Blvd., Milpitas, California 95035 (US). **STANIFORD, Stuart** [US/US]; FireEye, Inc., 1390 McCarthy Blvd., Milpitas, California 95035 (US).

(74) Agents: **DRAPINSKI, James, W.** et al.; Carr & Ferrell, LLP, 120 Constitution Dr., Menlo Park, California 94025 (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, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

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

(54) **Title**: ELECTRONIC MESSAGE ANALYSIS FOR MALWARE DETECTION

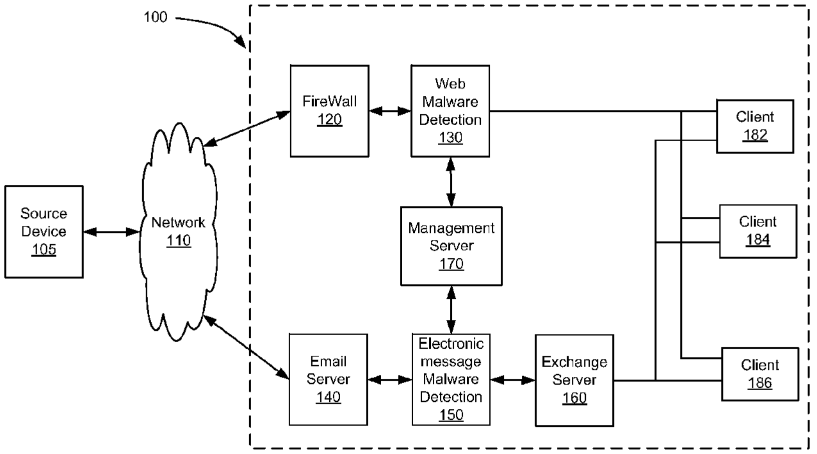

FIGURE 1

(57) **Abstract**: An electronic message is analyzed for malware contained in the message. Text of an electronic message may be analyzed to detect and process malware content in the electronic message itself. The present technology may analyze an electronic message and attachments to electronic messages to detect a uniform resource location (URL), identify whether the URL is suspicious, and analyze all suspicious URLs to determine if they are malware. The analysis may include re-playing the suspicious URL in a virtual environment which simulates the intended computing device to receive the electronic message. If the re-played URL is determined to be malicious, the malicious URL is added to a black list which is updated throughout the computer system.

Electronic Message Analysis for Malware Detection

Inventors:

Ashar Aziz
Henry Uyeno
Jay Manni
Amin Sukhera
Stuart Staniford

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part of U.S. patent application no. 11/717,474, filed March 12, 2007, entitled "Systems and Methods for Malware Attack Prevention", which is a continuation-in-part of U.S. patent application No. 11/494,990, filed July 28, 2006, entitled "Dynamic Signature Creation and Enforcement", which is a continuation-in-part of U.S. patent application No. 11/471,072, filed June 19, 2006, entitled "Virtual Machine with Dynamic Data Flow Analysis", which is a continuation-in-part of U.S. patent application No. 11/409,355, filed April 20, 2006, entitled "Heuristic Based Capture with Replay to Virtual Machine", which is a continuation-in-part of U.S. patent application No. 11/096,287, filed March 31, 2005, entitled "System and Method of Detecting Computer Worms", and is a continuation-in-part of U.S. patent application No. 11/151,812, filed June 13, 2005, entitled "System and Method of Containing Computer Worms," and is a continuation-in-part of U.S. patent application No. 11/152,286, June 13, 2005, entitled "Computer Worm Defense System and Method", U.S. patent application No. 11/096,287 claims the benefit of U.S. Provisional Application No. 60/559,198 filed on April 1, 2004, U.S. patent application No. 11/151,812 claims the benefit of U.S. Provisional Application No. 60/579,953 filed on June 14, 2004, and U.S. patent application No. 11/152,286 claims the benefit of U.S. Provisional Application No. 60/579,910 filed on June 14, 2004, all of which are incorporated by reference herein.

BACKGROUND

[0002] Presently, malicious network content (e.g., malicious software or malware) can attack various devices via a communication network. For example, malware may include any program or file that is harmful to a computer user, such as bots, computer viruses, worms, Trojan horses, adware, spyware, or any programming that gathers information about a computer user or otherwise operates without permission.

[0003] Various processes and devices have been employed to prevent the problems that malicious network content can cause. For example, computers often include antivirus scanning software that scans a particular client device for viruses. Computers may also include spyware and/or adware scanning software. The scanning may be performed manually or based on a schedule specified by a user associated with the particular computer, a system administrator, and so forth. Unfortunately, by the time a virus or spyware is detected by the scanning software, some damage on the particular computer or loss of privacy may have already occurred. Additionally, it can take days or weeks for new Anti-Virus signatures to be manually created and for an anti-virus application to be updated, by which time malware authors will have already created new versions that evade the signatures. Moreover, polymorphic exploits are also an issue that limits the effectiveness of some anti-virus applications.

[0004] Malicious network content may be distributed over a network via web sites, e.g., servers operating on a network according to an HTTP standard. Malicious network content distributed in this manner may be actively downloaded and installed on a user's computer, without the approval or knowledge of the user, simply by accessing the web site hosting the malicious network content. The web site hosting the malicious network content may be referred to as a malicious web site. The malicious network content may be embedded within data associated with web pages hosted by the malicious web site. For example, a web page may include JavaScript code, and malicious network content may be embedded within the JavaScript code. In this example, the malicious network content embedded within the JavaScript code may be obfuscated such that it is not apparent until the JavaScript code is executed that the JavaScript code contains malicious network content. Therefore, the malicious network content may attack or infect a user's computer before detection by antivirus software, firewalls, intrusion detection systems, or the like.

[0005] Additionally, malicious network content may be distributed by electronic messages, including email, using such protocols as POP, SMTP, IMAP, and various forms of web-based email. Malicious content may be directly attached to the message (for example as a document capable of exploiting a document reading application, such as a malicious Microsoft Excel document). Alternatively, electronic messages may contain URL links to malicious content hosted on web servers elsewhere on the network. When target users click on such links, they may be infected from the web in the manner described above. These techniques for infecting user computers via electronic messages are often used to make targeted attacks on particular "high-value" users at organizations, such as executives or key technical or operational staff.

[0006] What is needed is an improved system for detecting malicious content propagated in electronic messages.

SUMMARY

[0007] The present technology analyzes an electronic message for malware contained in the message. Systems that analyze electronic messages typically analyze attached files for malware. The content of an electronic message itself may contain text, which is usually not examined by malware systems. The present technology analyzes text of an electronic message to detect and process malware content in the electronic message itself. In some embodiments, the present technology may analyze an electronic message to detect a uniform resource location (URL), identify whether the URL is suspicious, and analyze all suspicious URLs to determine if they are malware. The analysis may include re-playing the suspicious URL in a virtual environment which simulates the intended computing device to receive the electronic message. If the re-played URL is determined to be malicious, the malicious URL is added to a black list which is updated throughout the computer system.

[0008] In an embodiment, malicious network content may be detected by a network content processing system by receiving an electronic message. The electronic message may be determined to include content determined to be suspicious. The suspicious electronic message content may be executed in a virtual environment. The suspicious electronic message content may be identified as malicious based on execution of the suspicious electronic message content in the virtual environment.

BRIEF DESCRIPTION OF FIGURES

[0009] Figure 1 is a block diagram of an exemplary system for detecting malicious electronic messages.

[0010] Figure 2 is a block diagram of an exemplary e-mail malware detection module.

[0011] Figure 3 is a block diagram of an exemplary management server.

[0012] Figure 4 is a flowchart of an exemplary method for detecting malicious electronic messages.

[0013] Figures 5 is a flowchart of an exemplary method for identifying a suspicious URL.

[0014] Figure 6 is a flowchart of an exemplary method for identifying suspicious URLs.

[0015] Figure 7 is a flowchart of an exemplary method for updating a malware detection system.

[0016] Figure 8 is a block diagram of an exemplary computing device.

DETAILED DESCRIPTION

[0017] The present technology analyzes electronic messages for malware contained in the message. Systems that analyze electronic messages typically analyze attached files for malware in synthetic environments such as a virtual environment. Unlike prior systems, the present technology may analyze the content of an electronic message to detect malware in the message content. For example, the content may include a uniform resource locator (URL) address. The URL address may be analyzed to determine if the URL address is associated with malware. Additionally, the present technology may analyze attachments in a real operating system running in an instrumented virtual environment. In addition to analyzing the content within an email itself, the present technology may process attachments for emails that provide a location associated with malware. The attachments may include one or more files compatible with common applications, including Word, Excel and Powerpoint applications by Microsoft Corporation, of Redmond, Washington, and Adobe Reader application, by Adobe Systems Inc., of San Jose, California.

[0018] In some embodiments, the present technology may analyze an electronic message to detect a URL, identify whether the URL is suspicious, and analyze the suspicious URL to determine if it describes a location associated with malware. Determining if the URL is suspicious may include if comparing the URL to one or more lists of URLs. For example, the URL may be compared to a white list of acceptable URLS, a black list of malware URLs, and/or a list having a combination of URLs. If the URL is not found on any list, the URL is not determined to be malware and not determined to be acceptable, and therefore may be determined to be suspicious.

[0019] Analysis of a suspicious URL may include re-playing the suspicious URL in a virtual environment which simulates the intended computing device to receive the electronic message. Re-playing a URL may include executing the URL by a virtual component in the virtual environment to request content located from the URL address. Content is received by the virtual environment in a URL request response, the received content is loaded into the virtual environment, and executed while the virtual environment is monitored. If the re-played URL is determined to be malicious, the malicious URL is added to a black list which is updated throughout the computer system.

[0020] The electronic message content, for example a URL, may be identified as malicious by a first device or module that processes electronic messages to detect malware. Other first devices or modules in the system may process network traffic to detect malware. A central device or module may communicate with both the network traffic malware module and the electronic message malware module. In some embodiments, the central module may receive URLs detected to be malicious, may update a central URL blacklist based on the received URLs, and may transmit the updated URL blacklist to both the network traffic malware module and the electronic message malware module. This may cause a network malware module to examine more closely web traffic returning from requests to URLs passed in email, for example making it more likely that such web traffic was replayed in a virtual environment.

[0021] Figure 1 is block diagram of an exemplary system for detecting malicious electronic messages. The system of Figure 1 includes source device 105, network 110 and malware detection system 100. Malware detection system 100 includes firewall 120, web malware detection device 130, electronic message server 140, electronic message malware detection device 150, exchange server 160, management server 170, client device 182, client device 84 and client device 186. Though blocks within system 100 may be discussed herein as different devices, such as web malware detection 130 and electronic message malware detection 150, blocks of system 100 may be implemented as modules within a single device or combination of devices.

[0022] Source device 105 may transmit electronic messages and content page content, such as web page content, to malware detection system 100 over network 110. System 100 may receive network traffic content through firewall 120 and may receive electronic message content through electronic message server 140 via network 110.

[0023] Network 110 may transmit electronic message, content page, and other content between devices connected to network 110, including web malware detection system 130, electronic message malware detection system 150, and source device 105. Network 110 may include one or more private networks, public networks, LANs, WANs, intranets, the Internet, and a combination of these networks.

[0024] Firewall 120 may be a device that consists of hardware and/or software that detects and prevents unauthorized network traffic from being received by or sent by client devices 182, 184 and 186. Firewall 120 may communicate with network 110 and web malware detection system 130.

[0025] Web malware detection 130 may communicate with management server 170 and client devices 182-186. Web malware detection 130 may operate to intercept network traffic and analyze intercepted traffic to determine whether the traffic is malware. The intercepted traffic may be copied by web malware detection 130 and analyzed using heuristics and other techniques. The heuristics may be used to identify portions of the network traffic as suspicious. Portions of traffic not identified as suspicious are ignored and passed through web malware detection 130. The suspicious network traffic portions may be analyzed by replaying the traffic in a virtual environment. The replay may be monitored and used to identify malware content by web malware detection 130. A system for re-playing intercepted traffic in a virtual environment using virtual components is described in U.S. patent application no. 12/359,252, entitled "Detecting Malicious Network Content Using Virtual Environment Components", filed January 23, 2009, the disclosure of which is incorporated herein by reference.

[0026] Electronic message server 140 may receive and send electronic messages between network 110 and electronic message malware detection 150.

[0027] Electronic message malware detection 150 may communicate with exchange server 160, management server 170, and email server 140, and may be implemented on one or more devices such a mail transfer agents (MTAs). Electronic message malware detection 150 may intercept electronic message traffic directed towards client devices 182-186. Electronic message malware detection 150 may include logic which analyzes electronic messages transmitted to and from electronic message 140 to identify malicious content within the electronic message. Identifying malware may include identifying an electronic message as suspicious, analyzing suspicious electronic messages to identify a malicious message, and communicating the malicious content to management server 170 to inform the remainder of system 100. Analyzing the suspicious electronic message may include replaying a portion of the electronic message in a virtual environment and monitoring the replay of the content. In some embodiments, content examined by electronic message malware detection 150 may include a URL detected within the body or header of an electronic message received by system 100.

[0028] Exchange server 160 may transfer mail between client devices 182-186 and electronic message malware detection 150. Management server 170 may receive malicious URL notifications, aggregate the received URLs, and update a black list maintained at management server 170. The malicious URL notification may be received from system 150 or system 130. Management server 170 may also transmit the black list of URLs to web malware detection systems and electronic message malware detection systems throughout system 100.

[0029] Clients 182, 184 and 186 may be any kind of device within a system 100 on which one or more users may execute programs to access network content such as a web page and transmit electronic messages such as an electronic message, instant message, or other electronic message.

[0030] Figure 2 is a block diagram of an exemplary electronic message malware detection system. The system of Figure 2 includes network tap 210, URL analyzer 220, scheduler 230, virtual environment component pool 240, virtual environment 250 and URL database 260. Network tap 210 may intercept electronic messages such as electronic message and instant messages transmitted between electronic message server 140 and exchange server 160. Network tap 210 may make a copy of the electronic message to analyze within electronic message malware detection system 150. Though electronic messages may include email as well as other types of messages, email will be discussed herein as merely an example.

[0031] URL analyzer 220 may detect URLs within a detected electronic message. Detecting a URL may include parsing the header and the body of an electronic message to identify a URL within the electronic message. Upon detecting a URL within a message, URL analyzer determines if the URL is suspicious and initiates an analysis of any suspicious URL. A URL may be suspicious if it does not appear in a list of acceptable URLs (a white list) and does not appear in a list of malware URLs (black list).

[0032] Upon detecting a suspicious URL, URL analyzer 220 provides the URL to scheduler 230. Scheduler 230 receives suspicious URLs and retrieves virtual environment components from virtual environment component pool 240. The virtual environment components may include components intended to replicate the actual environment at a client device intended to receive the electronic message message. For example, the virtual environments may include a virtual operating system, virtual applications, and a virtual network intended to replicate those associated with a particular client device intended to receive the message. Scheduler 230 then provides the URL and the retrieved virtual environment components to a virtual environment 250 in order to replay the URL within a virtual environment.

[0033] Virtual environment 250 receives the suspicious URL and virtual environment components and replays the URL within a virtual environment having the virtual components. Replaying the URL may be similar to performing a "click" operation on the suspicious URL. Upon performing a click on the URL, a request is sent to the URL for content, and the network server associated with the URL provides content and a response to the request. The content received in response to the request is then processed by the virtual environment and the environment is monitored to determine if any undesirable behavior occurs. If any undesirable behavior occurs in response to loading content associated with the URL, the URL is determined to be malware and added to a local black list by electronic message malware detection system 150. Undesirable behavior may unauthorized requests for data, sending or receiving data over a network, processing and/or storing data, changing a registry value, installing a file, executing a file, or other operations. The internal malware black list is transmitted to management server 170.

[0034] URL database 260 includes black URL list 262 and white URL list 264. URL analyzer may compare URLs detected in electronic messages to black URL list 262 to determine if there is a match. If there is a match, the URL is detected to be malware, and the electronic message may be blocked or the URL may be removed from the electronic message. If the URL is removed from the electronic message, an alert may be generated (e.g., within the message) indicating the URL has been removed and an administrator may be notified. If a detected URL matches a URL on the white URL list 264, the URL is determined to be acceptable and no further action is taken. If a detected URL does not match a URL on black URL list 262 or white URL list 264, the URL is identified as being suspicious and is processed in a virtual environment.

[0035] Figure 3 is a block diagram of an exemplary management server. Management server 180 of Figure 3 includes URL aggregator 310, URL black list 320, and communication manager 330. URL aggregator 310 aggregates received URLs and updates and stores URL black list 320. URL black list 320 is a list of confirmed malicious URLs maintained by management server 180. Communication manager 330 may receive URLs from electronic message malware detection systems and web malware detection systems within system 100. Communication manager 330 may provide the URLs to URL aggregator 310 to aggregate the URLs and update URL black list 320 maintained on management server 180. Communication manager 330 may also send the current URL black list to malware detection systems within system 100.

[0036] Figure 4 is a flowchart of an exemplary method for detecting malicious electronic messages. Though Figure 4 will be discussed in terms of an electronic message, other electronic messages, such as an instant message or other forms of communication, may be processed by the present technology.

[0037] An electronic message is received at step 405. The electronic message may be received by electronic message malware detection system 150 via electronic message server 140. The electronic message and/or an attachment to the message may be scanned to detect a URL at step 410. The electronic message may be scanned by a URL analyzer module to detect a URL in the electronic message header, body or other portion of the electronic message. The attachment may be scanned to detect a URL within the attachment. For example, if the attachment is a word processor or spreadsheet document, the attachment may be scanned to detect a URL in text of the word processor document or within a cell of the spreadsheet.

[0038] Detected URLs may be transmitted to a malware detection system at step 415. The malware detection system may be contained locally on electronic message malware detection system 150 or outside detection module 150. For example, electronic message malware detection system 150 may transmit detected URLs to web malware detection system 130 to process the URL to determine if the URL is malicious. In some embodiments, a URL is simply stored locally at electronic message malware detection system 150 at step 415 for further processing.

[0039] A suspicious URL may be identified from the detected URLs at step 420. A URL may be identified as suspicious if the URL does not match a black list of URLs or a white list of URLs maintained at electronic message malware detection system 150 (or accessible by detection module 150). Identifying suspicious URLs is discussed in more detail below with respect to the method of Figure 5.

[0040] Suspicious URLs are analyzed using virtual environment components to detect a malicious URL at step 425. Analyzing a suspicious URL may include selecting virtual components such as a virtual operating system, virtual applications, and virtual network, populating and configuring a virtual environment with the virtual components, and processing the URL within the virtual environment. Processing the URL within the environment may include replaying the URL within the virtual environment by performing a "click" operation on the URL. The URL may be identified as malicious if content received in response to the click operation on the URL results in an undesirable behavior within the virtual environment. An undesirable behavior may include attempts to change an operating system setting or configuration, execute an executable file within the virtual environment, transmit undesirable data, or other actions. In some embodiments, an undesirable behavior may include an unexpected behavior. If no undesirable behavior occurs in response to clicking the URL, the URL is determined to be acceptable and is added to a white list.

[0041] A malware detection system may be updated based on the detected malware URL at step 430. Updating may include communicating the malicious URL to other parts of a system. For example, electronic message malware detection system 150 may communicate one or more malicious URLs to management server 170, and server 170 may communicate the URL via an updated black list to web malware detection systems and electronic message malware detection systems within system 100. Updating a malware detection system is described in more detail below with respect to the method of Figure 7.

[0042] One or more factors may affect how a URL is determined to be suspicious and/or processed to determine if it is associated with malware. In an embodiment, any URL detected in an email may be transmitted by electronic message malware detection 150 to web malware detection 130. Upon detecting that content is being requested from the URL, for example in response to a user selection or "click" on the URL, the web malware detection 130 may increase the priority of the detected URL such that the URL is analyzed to determine if is suspicious and/or associated with malware. In this embodiment, the URL may not be processed by the web malware detection 130 until it is determined that content is actually being requested from the URL.

[0043] A large number of URLs may be detected by web malware detection 130 in network traffic travelling through firewall 120. One or more detected URLs detected by web malware detection 130 may be assigned a priority for analysis. Higher prioritized URLs are analyzed to determine if they are suspicious or associated with malware before lower priority URLs. In some embodiments, URLs detected in email are provided a lower priority than those detected as part of network traffic by web malware detection 130. The priority of a URL may be increased once it is determined to be present in both an email and network traffic (i.e., detected by both electronic message malware detection 150 and web malware detection 130, in any order). The level of priority increase may depend on the resources available to process URLs. For example, the level of priority increase may be less if there are a small number of virtual environments or components available to process a suspicious URL. If there is a large number of virtual environments and/or virtual components available to process a URL, there may be a large level of priority increase. Hence, the priority of URLS to be processed by may adjusted in such a way to avoid degradation of the normal functioning of web malware detection 130 under heavy load, while allowing thorough examination of all email URLs where load permits

[0044] Figure 5 is a flowchart of an exemplary method for identifying a suspicious URL. In some embodiments, the method of Figure 5 provides more detail for step 420 in the method of Figure 4. Each detected URL in an electronic message is compared to a URL white list at step 505. The URL white list may be maintained on electronic message malware detection system 150 and may include a list of acceptable URLs or URL domains. URLs that match the URL white list are ignored at step 510. The URLs that match the white list are determined to not be malicious and therefore are allowed to pass through to their intended client device.

[0045] Detected URLs which are not on the white list are then compared to the URL black list at step 515. URLs on the black list are known to be malicious and should not be passed through to a user associated with a client device. If a detected URL matches a URL on the black list, the URL is blocked and reported at step 520, and thereby prevented from being provided to the recipient client device. A URL may be prevented from delivery by either blocking transmission of the entire electronic message, removing the URL from the electronic message message, or in some other manner. URLs that do not match a URL on the white list or a URL on the black list are identified as suspicious URLs at step 526. The remaining URLs are characterized as suspicious because it is unknown whether they are acceptable or malicious.

[0046] Figure 6 is a flowchart of an exemplary method for identifying malicious URLs. The method of Figure 6 provides more detail for step 525 of the method of Figure 5. First, a suspicious URL is selected to analyze in a virtual environment at step 605.

Some URLs may be weighted with a higher priority to analyze. The higher priority URLs may be placed in a higher priority position in an analysis queue as opposed to lower priority URLs. A priority may be associated with a URL by a user, based on learning performed by the present system, or in some other manner. The priority may be associated with the URL domain, keywords in the URL, positioning within the electronic message for the URL, or other factors.

[0047] The present system may configure a virtual environment application, operating system, and network components at step 610. These virtual components may be retrieved from a component pool by a scheduler. A URL may be analyzed in the virtual environment configured with the virtual components at step 615. Analyzing the URL may include replaying the URL by performing a "click" operation on the URL within the virtual environment. Upon performing the click operation, an application may send a content request message to the URL and receive a response message in response to the URL request. For example, a network browser may be executed to provide the content received in response to the URL response received by the application. Actions performed within the virtual environment in response to receiving the URL content may be recorded and analyzed to determine if the URL is malicious.

[0048] A malicious URL may be identified at step 620. An identification as a malicious URL may be based on actions or changes that occur when a suspicious URL is replayed in the virtual environment. Actions that may indicate a malicious URL include changing an operating system configuration, performing requests or trying to install or execute file, or other actions performed in response to retrieving content from the URL location.

[0049] Figure 7 is a flowchart of an exemplary method for updating a malware detection system. The method of Figure 7 provides more detail for step 430 of the method of Figure 4. First, a management server receives a malicious URL detected by electronic message malware detection system at step 705. The malicious URLs are then aggregated by the management server at step 710. A URL black list is updated with the aggregated malicious URLs at step 715. The management server may then transmit the updated URL black list to electronic message malware detection systems and web malware detection systems at step 720. The transmission of the updated URL black list may be performed upon request, periodically, or upon occurrence of a particular event, such as when a URL black list has undergone a threshold number of changes.

[0050] In some applications of this technology, it may not be desired to fetch content from every URL seen in incoming electronic messages where such "clicks" may have undesired side effects on applications using the web (HTTP) as a communication protocol. Therefore, an alternative method can be used in such cases, in which all URLs received in electronic messages are forwarded to a web malware detection system, and are used to raise the probability of examining any particular piece of web content if it has previously been seen in electronic messages (e.g., email). Thus "targeted spear phishing" attacks in which malicious URLs are sent to particular email addresses in an effort to induce the recipient to click on the link will be examined by the malware detection system only in the event that the recipient does actually so click.

[0051] Since many URLs seen in electronic messages are also accessed via the web, the present invention also includes a dynamic method for setting the "email priority boost" used to enhance the priority of inspecting web content by noting the fraction of all the efforts of the web malware detection system devoted to examining URLs previously seen by the electronic message malware detection system. This "email priority boost" can be regulated to target a particular fraction of the virtual execution environments available on the web malware detection system, to avoid overloading the latter and causing loss of other web detection functionality, while still allowing complete examination of urls seen in electronic messages where system load allows.

[0052] Figure 8 is a block diagram of an exemplary computing device. The computing device of Figure 8 may be used to implement one or more devices in the system 100 of Figure 1, including but not limited to firewall 120, web malware detection 130, e-mail server 140, e-mail malware detection 150, management server 170, exchange server 160, or clients 182-186Figure 8 is a block diagram of an exemplary malicious network content detection device. In some embodiments, the method of Figure 8 provides more detail for malicious network content detection system 125 of Figure 1. Malicious network content detection system 125 comprises at least one or more processors 805, memory systems 810, and storage systems 815, each of which can be communicatively coupled with data bus 820. In some embodiments, data bus 820 may be implemented as one or more data buses. Malicious network content detection system 125 may also comprise communication network interface 825, input/output (I/O) interface 830, and display interface 835. Communication network interface 825 may be communicatively coupled with network 120 via communication medium 840. In some embodiments, malicious network content detection system 125 may be communicatively coupled with a network tap, such as network tap 115, which in turn may be communicatively coupled with network 120. Bus 920 provides communications between communications network interface 825, processor 805, memory system 810, storage system 815, I/O interface 830, and display interface 835.

[0053] Communications network interface 825 may communicate with other digital devices (not shown) via communications medium 840. Processor 905 executes instructions which may be stored on a processor-readable storage medium. Memory system 810 may store data permanently or temporarily. Some examples of memory system 810 include RAM and ROM. Storage system 815 also permanently or temporarily stores data. Some examples of storage system 815 are hard discs and disc drives. I/O interface 830 may include any device that can receive input and provide output to a user. I/O interface 830 may include, but is not limited to, a keyboard, a mouse, a touch screen, a keypad, a biosensor, a compact disc (CD) drive, a digital video disc (DVD) drive, an optical disk drive, or a floppy disk drive. Display interface 835 may include an interface configured to support a display, monitor, or screen. In some embodiments, malicious network content detection system 125 comprises a graphical user interface to be displayed to a user over a monitor in order to allow the user to control malicious network content detection system 125.

[0054] The foregoing detailed description of the technology herein has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the technology to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. The described embodiments were chosen in order to best explain the principles of the technology and its practical application to thereby enable others skilled in the art to best utilize the technology in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the technology be defined by the claims appended hereto.

CLAIMS

What is claimed is:

1. A method for detecting malicious network content by a network content processing system, comprising:

receiving an electronic message;

determining that the electronic message includes content determined to be suspicious;

executing the suspicious electronic message content in a virtual environment; and

identifying the suspicious electronic message content as malicious based on execution of the suspicious electronic message content in the virtual environment.

2. The method of claim 1, wherein the electronic message is an email.

3. The method of claim 1, wherein the content includes a uniform resource locator (URL) address.

4. The method of claim 3, further comprising:

comparing the URL to a first list of URLs; and

identifying the URL as suspicious if the URL is not in the first list of URLs.

5. The method of claim 4, wherein the first list includes URLs associated with malware.

6. The method of claim 4, wherein the first list includes URLs known to not be associated with malware.

7. The method of claim 5, wherein executing the URL includes sending a content request to the URL address by a virtual application component in the virtual environment.

8. The method of claim 7, wherein the virtual application component is a virtual network browser application.

9. The method of claim 8, wherein the electronic message is determined to be suspicious by an electronic message malware device, the suspicious electronic message content identified as malicious by a web malware device.

10. The method of claim 1, further comprising:

configuring a virtual environment component within a virtual environment to mimic a real application configured to process the suspicious network content, the virtual environment configured within the network content processing system;

processing the suspicious network content using the virtual environment component within the virtual environment; and

identifying the suspicious network content as malicious network content based on a behavior of the virtual environment component.

11. The method of claim 10, where the suspicious network content includes a file attached to an electronic message, the virtual environment component including an application configured to process the file.

12. The method of claim 11, where the file is a Microsoft Word type document, the virtual environment component including a Microsoft Word program.

13. The method of claim 11, where the file is a Microsoft Excel type document, the virtual environment component including a Microsoft Excel program.

14. The method of claim 11, where the file is a Microsoft Powerpoint type document, the virtual environment component including a Microsoft Powerpoint program.

15. The method of claim 11, where the file is a Portable Document Format (PDF) document, the virtual environment component including an Adobe PDF Reader program.

16. The method of claim 1, further comprising monitoring changes to the virtual environment operating system by an agent, the suspicious network content identified as malicious network content based on detected improper changes to the virtual environment operating system.

17. The method of claim 4, further comprising transmitting one or more malicious URLS to a remote device, the remote device configured receive the one or more malicious URLs, consolidating malicious URLs, and transmitting an updated list of URLs associated with URLs.

18. The method of claim 17, further comprising:

transmitting one or more URLs from an electronic message malware detection system to a web malware detection system; and

raising a priority associated with examining one or more URLs received from a network by the web malware detection system, the priority raised based on the transmitted URLs.

19. The method of claim 18, further comprising:

dynamically adjusting a priority for processing URLs detected within an email based on the web malware detection system load.

20. A computer readable storage medium having stored thereon instructions executable by a processor for performing a method for detecting malicious network content, the method comprising:

receiving an electronic message;

determining that the electronic message includes content determined to be suspicious;

executing the suspicious electronic message content in a virtual environment; and

identifying the suspicious electronic message content as malicious based on execution of the suspicious electronic message content in the virtual environment.

21. The computer readable storage medium of claim 20, wherein the electronic message is an email.

22. The computer readable storage medium of claim 20, wherein the content includes a uniform resource locator (URL) address.

23. The computer readable storage medium of claim 22, the method further comprising:

comparing the URL to a first list of URLs;

identifying the URL as suspicious if the URL is not in the first list of URLs.

24. The computer readable storage medium of claim 22, wherein the first list includes URLs associated with malware.

25. The computer readable storage medium of claim 22, wherein the first list includes URLs known to not be associated with malware.

26. The computer readable storage medium of claim 22, wherein executing the URL includes sending a content request to the URL address by a virtual application component in the virtual environment.

27. The computer readable storage medium of claim 20, wherein the virtual application component is a virtual network browser application.

28 The computer readable storage medium of claim 20, wherein the electronic message is determined to be suspicious by an electronic message malware device, the suspicious electronic message content identified as malicious by a web malware device.

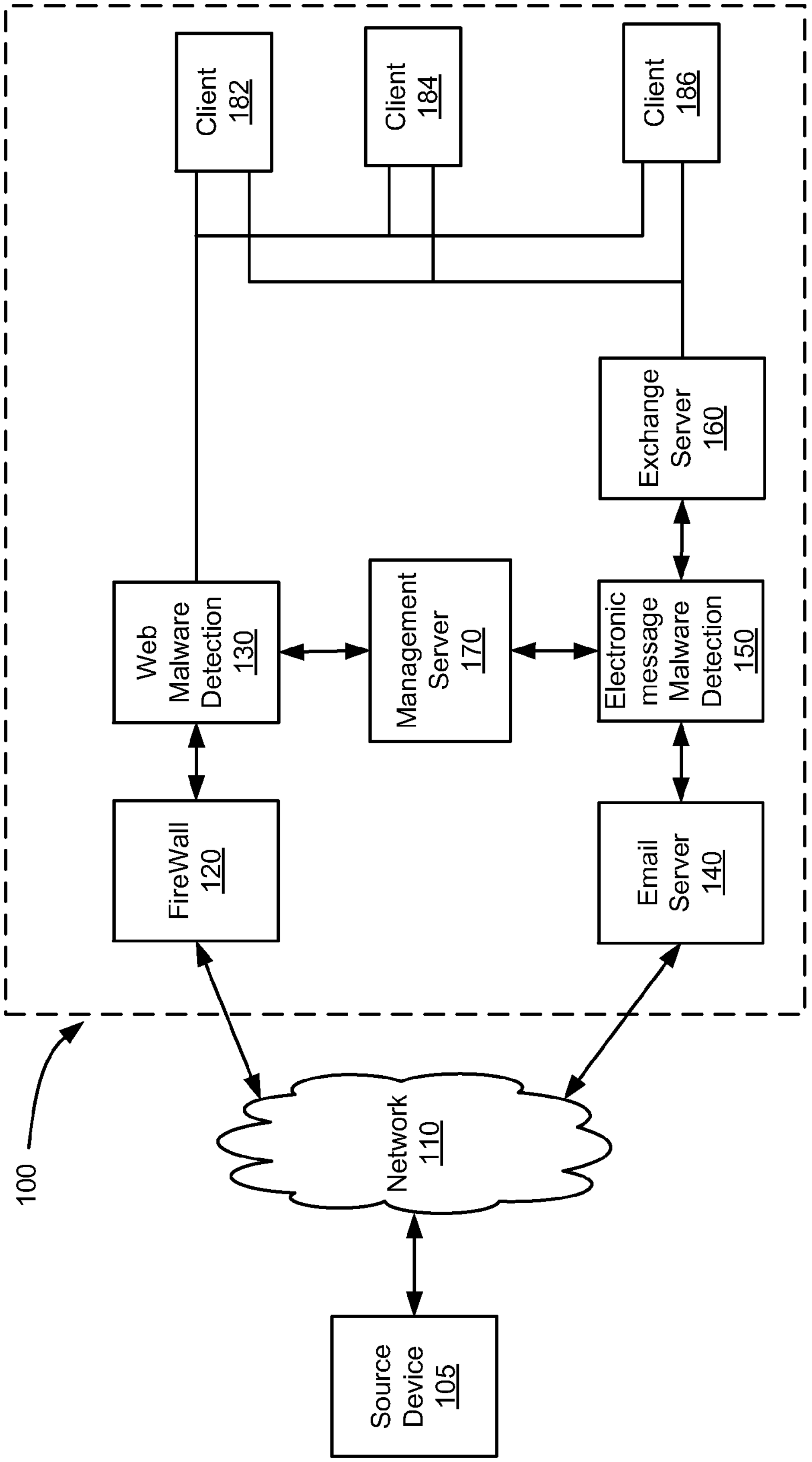
100
Client 182
Client 184
Client 186
FireWall 120
Web Malware Detection 130
Management Server 170
Electronic message Malware Detection 150
Exchange Server 160
Email Server 140
Network 110
Source Device 105

FIGURE 1

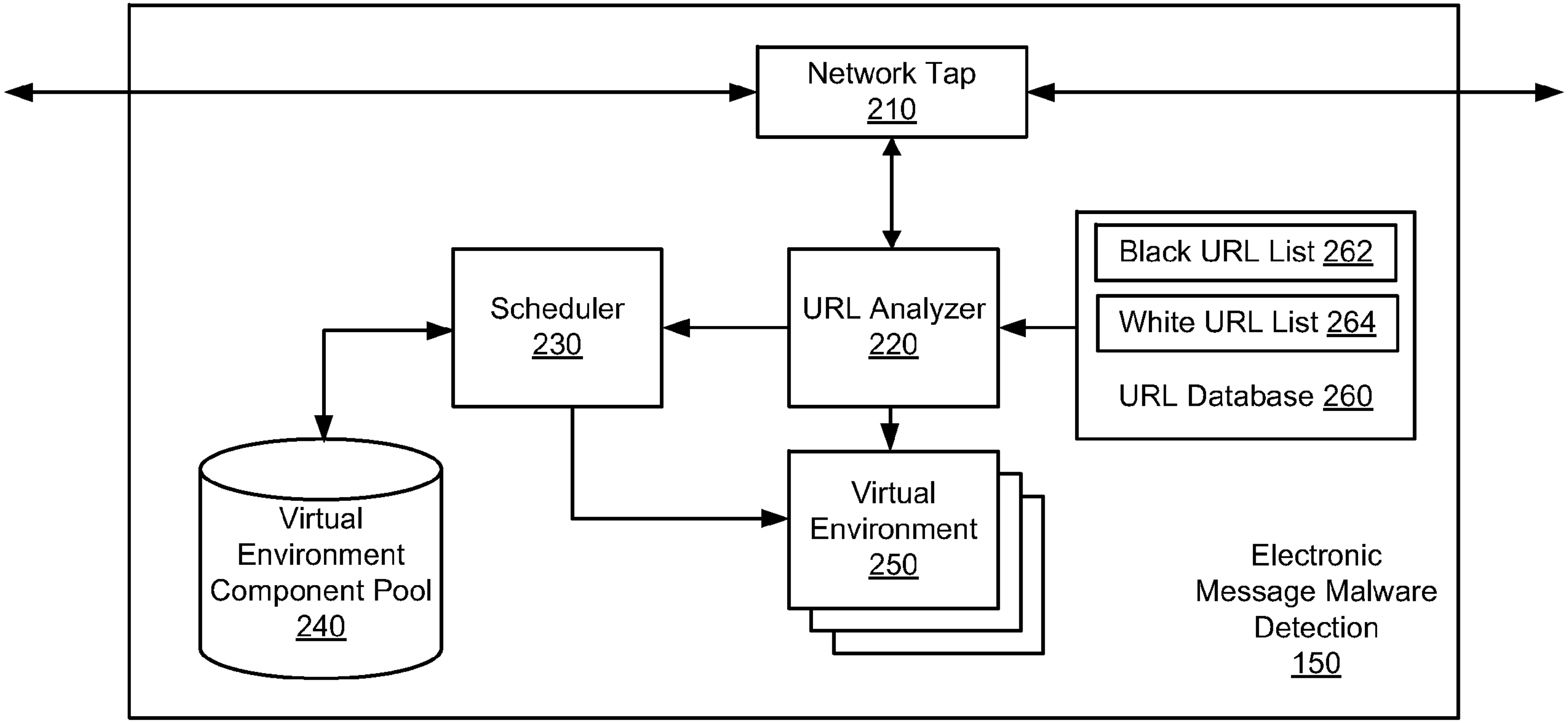
Network Tap
210
Black URL List 262
White URL List 264
URL Database 260
Scheduler
230
URL Analyzer
220
Virtual
Environment
Component Pool
240
Virtual
Environment
250
Electronic
Message Malware
Detection
150

FIGURE 2

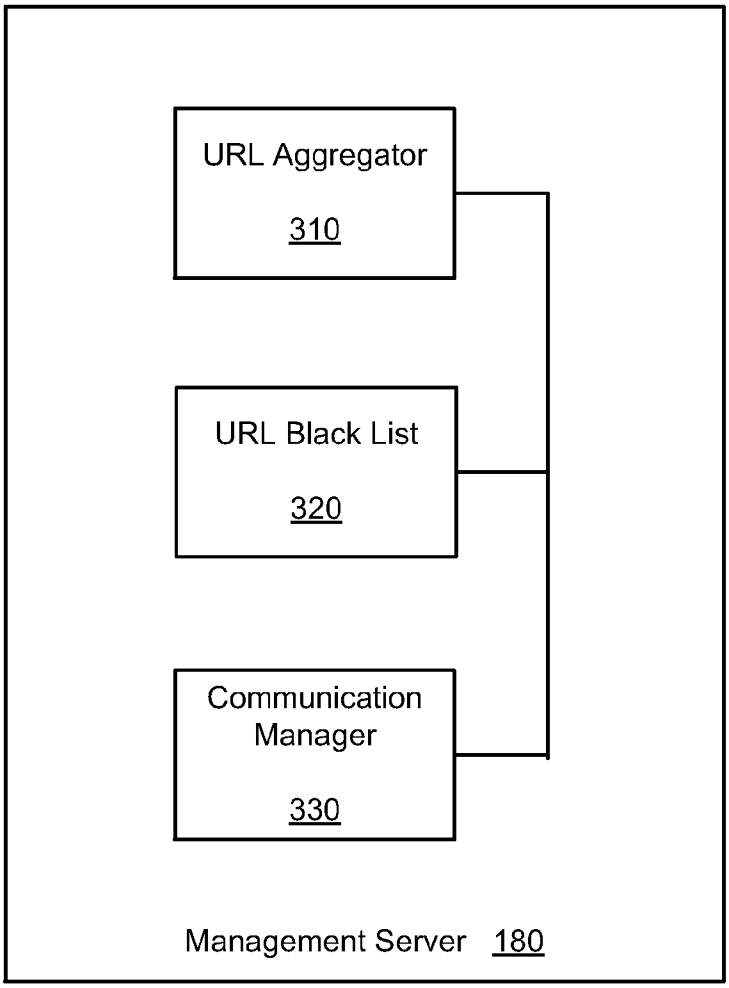
URL Aggregator
310
URL Black List
320
Communication Manager
330
Management Server 180

FIGURE 3

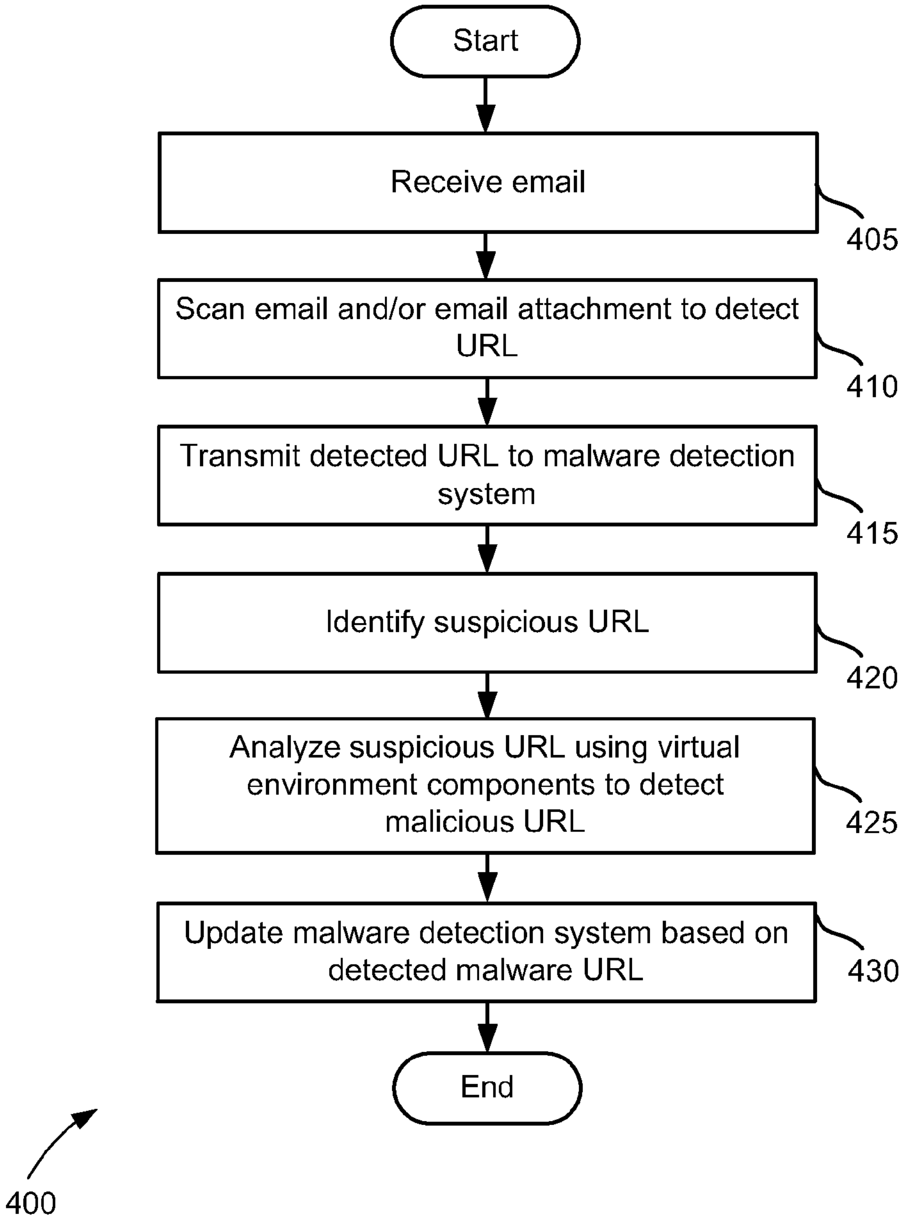
Start
Receive email
405
Scan email and/or email attachment to detect URL
410
Transmit detected URL to malware detection system
415
Identify suspicious URL
420
Analyze suspicious URL using virtual environment components to detect malicious URL
425
Update malware detection system based on detected malware URL
430
End
400

FIGURE 4

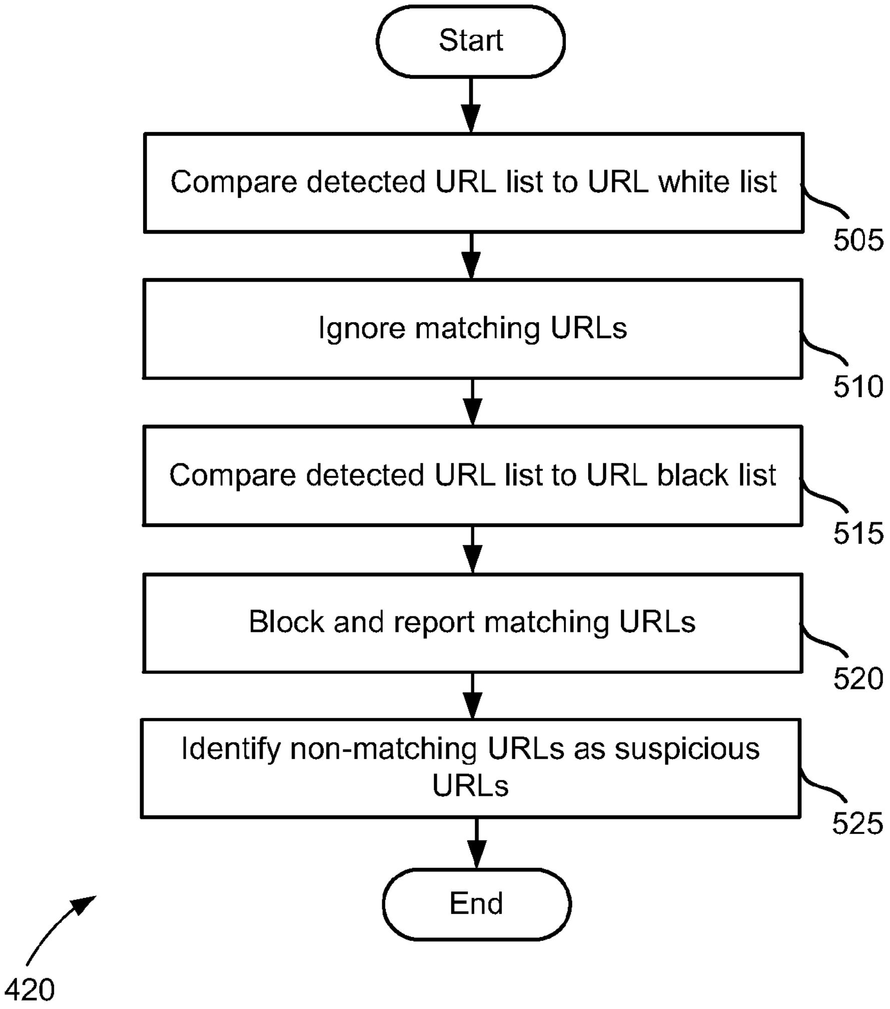
Start
Compare detected URL list to URL white list
505
Ignore matching URLs
510
Compare detected URL list to URL black list
515
Block and report matching URLs
520
Identify non-matching URLs as suspicious URLs
525
End
420

FIGURE 5

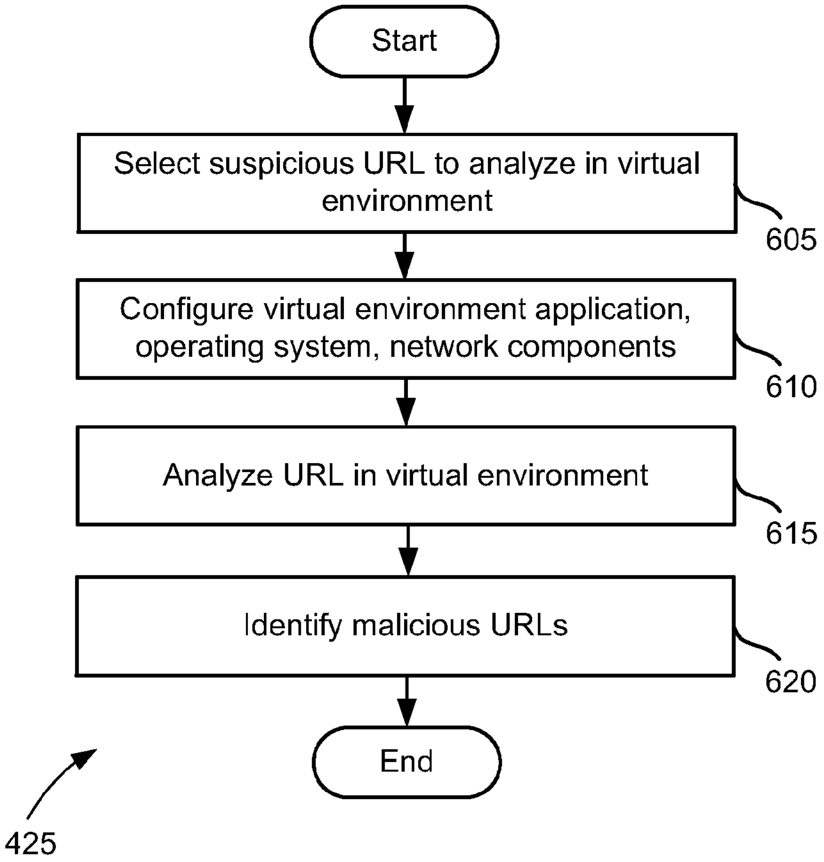
Start
Select suspicious URL to analyze in virtual environment
605
Configure virtual environment application, operating system, network components
610
Analyze URL in virtual environment
615
Identify malicious URLs
620
End
425

FIGURE 6

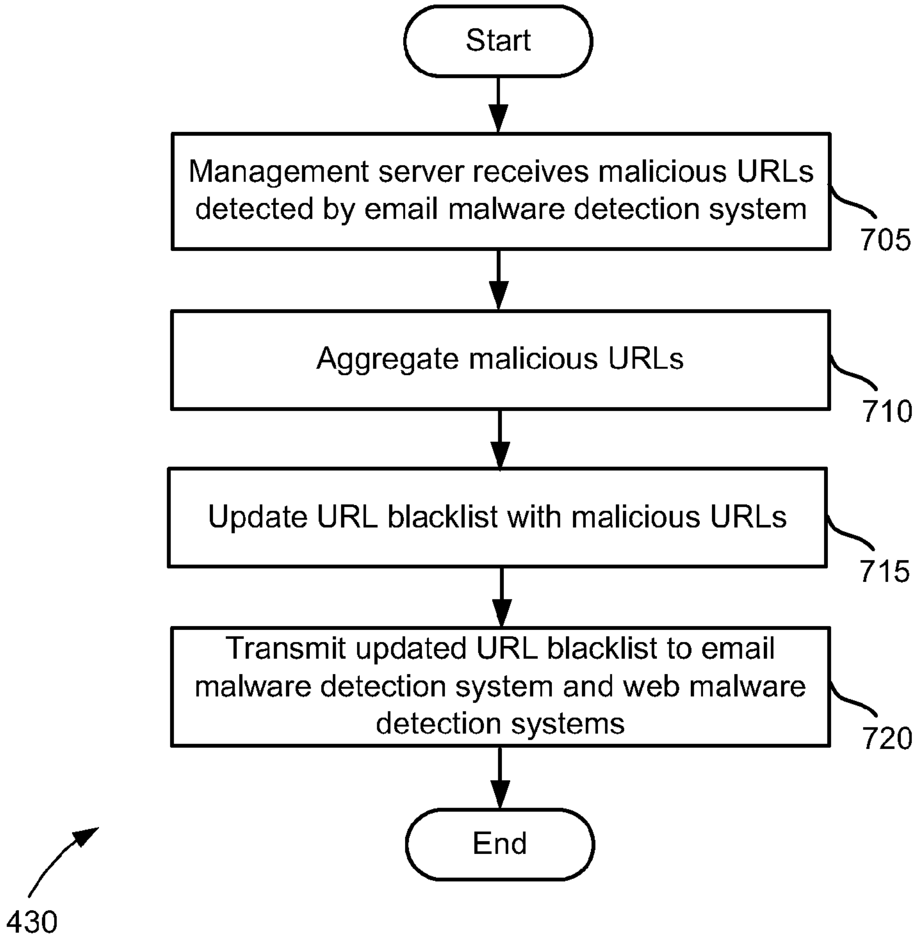
Start
Management server receives malicious URLs detected by email malware detection system
705
Aggregate malicious URLs
710
Update URL blacklist with malicious URLs
715
Transmit updated URL blacklist to email malware detection system and web malware detection systems
720
End
430

FIGURE 7

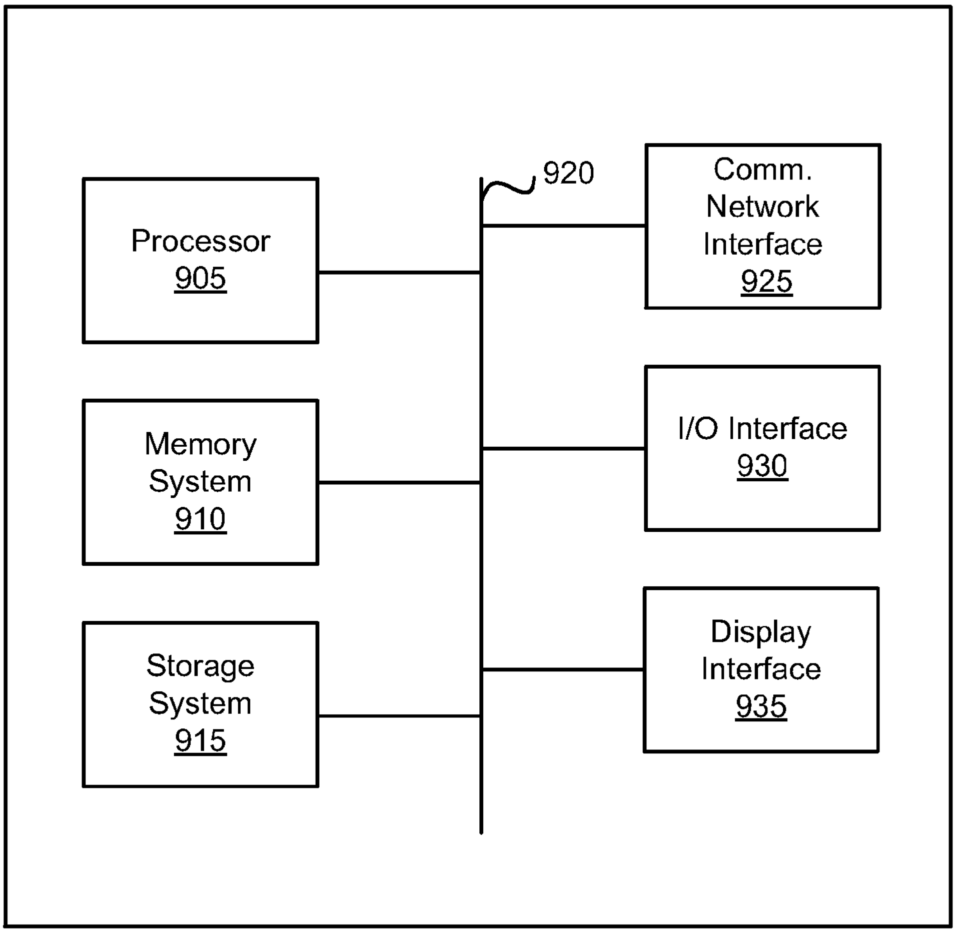
Processor 905
Memory System 910
Storage System 915
920
Comm. Network Interface 925
I/O Interface 930
Display Interface 935

FIGURE 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/26402

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 11/00 (2012.01)
USPC - 726/24
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)
USPC: 726/24; IPC: G06F 11/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 726/22, 24

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST: (PGPB, USPT, EPAB, JPAB); Google Scholar
Search Terms: detect malicious software, detect malicious content, malware, network, electronic message, email, electronic mail, suspicious, intrusive, threat, unauthorized, hostile, harmful, virtualize, VM, virtual machine, virtual environment, sandbox, emulate, mimic

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- Y	US 2007/0174915 A1 (GRIBBLE et al.) 26 July 2007 (26.07.2007), abstract and para [0014], [0016]-[0018], [0029]-[0042], [0044]-[0047], [0064], [0067], [0070]-[0071], [0076].	1, 3-11, 16-20, 22-28 ---------------------------------- 2, 12-15, 21
Y	US 2010/0077481 A1 (POLYAKOV et al.) 25 March 2010 (25.03.2010), abstract and para [0002], [0045].	2, 21
Y	US 2010/0281102 A1 (CHINTA et al.) 04 November 2010 (04.11.2010), abstract and para [0046], [0090], [0092], [0129], [0334], [0342], [0347].	12-15
A	US 2011/0047620 A1 (MAHAFFEY et al.) 24 February 2011 (24.02.2011), entire document.	1-28

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

* 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
07 May 2012 (07.05.2012)

Date of mailing of the international search report
25 MAY 2012

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-3201

Authorized officer:
Lee W. Young
PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774