



- (51) **International Patent Classification:**
H04L 29/06 (2006.01) *H04L 29/08* (2006.01)
- (21) **International Application Number:**
PCT/EP2017/053363
- (22) **International Filing Date:**
15 February 2017 (15.02.2017)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
20165113 16 February 2016 (16.02.2016) FI
- (71) **Applicant:** NOKIA SOLUTIONS AND NETWORKS OY [FI/FI]; Karaportti 3, 02610 Espoo (FI).
- (72) **Inventors:** KURZ, Bernhard; Klara-Ziegler-Bogen 51, 81739 München (DE). LUEKEN, Joachim; Stroblstraße 13, 80686 Munich (DE). ABENDROTH, Joerg; Weiden-er Str. 3, 81737 Munich (DE).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

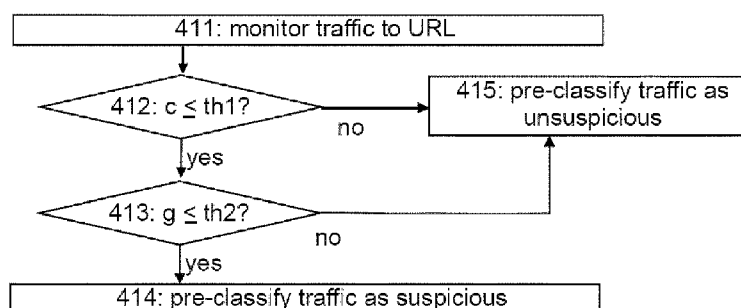
(54) **Title:** DETECTION OF MALWARE IN COMMUNICATIONS

Fig. 4B

(57) **Abstract:** A method for malware infection detection comprises monitoring (411), in a network node, network connections of communication devices to a network resource identifier in a communication system. Based on the monitoring, the network node pre-classifies (414) the network resource identifier as a suspicious network resource identifier if (412) the number of communication devices connected to the network resource identifier at least once within a pre-defined time period is not greater than a pre-defined connection threshold value, and, if (413) the increase rate of the number of communication devices connected to the network resource identifier at least once within a pre-defined time period does not exceed a pre-defined rate threshold value.



DESCRIPTION

TITLE

5

DETECTION OF MALWARE IN COMMUNICATIONS

TECHNICAL FIELD

The invention relates to communications.

BACKGROUND

10

Cybercrime is increasingly targeting smart phones and other mobile devices. Mobile malware may aim at stealing private data, committing fraud and extortion. Network-based malware detection may be used to protect the security and privacy of the mobile devices and their users in the mobile network and preserve the reputation of a network operator.

BRIEF DESCRIPTION

15

According to an aspect, there is provided the subject matter of the independent claims. Embodiments are defined in the dependent claims.

One or more examples of implementations are set forth in more detail in the accompanying drawings and the description below. Other features will be apparent from the description and drawings, and from the claims.

20 BRIEF DESCRIPTION OF DRAWINGS

In the following the invention will be described in greater detail by means of preferred embodiments with reference to the attached [accompanying] drawings, in which

In the following, the invention will be described in greater detail by means of preferred embodiments with reference to the accompanying drawings, in which

25

Figure 1 illustrates a wireless communication system to which embodiments of the invention may be applied;

Figure 2 illustrates a signalling diagram of a procedure for malware detection according to an embodiment of the invention;

30

Figure 3 is a blocks diagram illustrating an embodiment of the invention;

Figures 4A to 4C illustrate pre-classification according to an embodiment of the invention;

Figure 5 illustrates a process for malware detection according to an embodiment of the invention;

35

Figure 6 illustrates a communication behaviour analysis according to an embodiment of the invention;

Figure 7 is a blocks diagram illustrating an embodiment of the invention;

Figure 8 illustrates a blocks diagram of an apparatus according to an embodiment of the invention.

DETAILED DESCRIPTION OF SOME EMBODIMENTS

5 The following embodiments are exemplary. Although the specification may refer to “an”, “one”, or “some” embodiment(s) in several locations, this does not necessarily mean that each such reference is to the same embodiment(s), or that the feature only applies to a single embodiment. Single features of different embodiments may also be combined to provide other embodiments. Further-more, words “comprising” and “including” should be understood
10 as not limiting the described embodiments to consist of only those features that have been mentioned and such embodiments may contain also features/structures that have not been specifically mentioned.

Embodiments described may be implemented in a radio system, such as in at least one of the following: universal mobile telecommunication system (UMTS, 3G) based on basic
15 wideband-code division multiple access (W-CDMA), high-speed packet access (HSPA), long term evolution (LTE), LTE-advanced, and/or 5G system. The present embodiments are not, however, limited to these systems.

The embodiments are not, however, restricted to the system given as an example but a person skilled in the art may apply the solution to other communication systems provided
20 with necessary properties. One example of a suitable communications system is the 5G system, as listed above. It is assumed that network architecture in 5G will be quite similar to that of the LTE-advanced. 5G is likely to use multiple input – multiple output (MIMO) antennas, many more base stations or nodes than the current network deployments of LTE (a so-called small cell concept), including macro sites operating in co-operation with smaller local area access
25 nodes and perhaps also employing a variety of radio technologies for better coverage and enhanced data rates. 5G will likely be comprised of more than one radio access technology (RAT), each optimized for certain use cases and/or spectrum.

It should be appreciated that future networks will most probably utilize network functions virtualization (NFV) which is a network architecture concept that proposes virtual-
30 izing network node functions into “building blocks” or entities that may be operationally connected or linked together to provide services. A virtualized network function (VNF) may comprise one or more virtual machines running computer program codes using standard or general type servers instead of customized hardware. Cloud computing or cloud data storage may also be utilized. In radio communications this may mean node operations to be carried out, at
35 least partly, in a server, host or node operationally coupled to a remote radio head. It is also possible that node operations will be distributed among a plurality of servers, nodes or hosts. It should also be understood that the distribution of labour between core network operations

and base station operations may differ from that of the LTE or even be non-existent. Some other technology advancements probably to be used are software-defined networking (SDN), big data, and all-IP, which may change the way networks are being constructed and managed.

Figure 1 illustrates an example of a cellular communication system to which embodiments of the invention may be applied. Cellular radio communication networks, such as the long term evolution (LTE), the LTE-advanced (LTE-A) of the 3rd generation partnership project (3GPP), or the predicted future 5G solutions, are typically composed of at least one network element, such as a network element 110, providing a cell 100. Each cell may be, e.g. a macro cell, a micro cell, femto, or a pico cell, for example. The network element 110 may be an evolved node-B (eNB) as in the LTE and LTE-A, or any other apparatus capable of controlling radio communication and managing radio resources within a cell. For 5G solutions, the implementation may be similar to LTE-A, as described above. The network element 110 may be called a base station or an access node. The cellular communication system may be composed of a radio access network of network elements 110, 112, 114, e.g. eNBs, each controlling a respective cell or cells 100, 102, 104. The network elements 110 to 114 may each control a macro cell 100 to 104 providing wide area coverage for terminal devices 120. The network elements 110 to 114 may also be called access nodes because they provide the terminal devices 120 with wireless access to other networks such as the internet. Additionally, one or more local area access nodes 116 may be arranged within a control area of a network element 110, 112, 114 controlling a macro cell, 100 to 104. The local area access node 116 may provide wireless access within a sub-cell 106 that may be comprised within a macro cell 100. Examples of the sub-cell may include a micro, pico and/or femto cell. Typically, the sub-cell provides a hot spot within a macro cell. The operation of the local area access node 116 may be controlled by a network element 110 under whose control area the sub-cell is provided. The network element 110 and the other network elements 112 to 116 may support dual connectivity (DC) in which the terminal device 120 has established multiple connections with cells associated with a master eNB network element and a secondary eNB network element.

The network element 110 may employ carrier aggregation in which the terminal device 112 is allocated with resources from a plurality of component carriers that may be on contiguous frequency bands or on non-contiguous frequency bands. One network element 110 may provide one component carrier, e.g. a primary component carrier, while another network element 116 may provide another component carrier, e.g. a secondary component carrier. The network element 110 operating the primary component carrier may carry out scheduling of resources on all component carriers, or each network element 110, 116 may control scheduling of the component carrier it operates. Alternatively network element 110 may provide one component carrier, e.g. a primary component carrier, as well as another component carrier, e.g. a secondary component carrier.

In the case of multiple eNBs in the communication network, the eNBs may be connected to each other with an X2 interface as specified in LTE. Other communication methods between the network elements may also be possible. The network elements 110 to 116 may be further connected via an S1 interface to an evolved packet core (EPC) 130, more specifically to a mobility management entity (MME) 132 and to a system architecture evolution gateway (SAE-GW) 134.

The radio system of Figure 1 may support machine type communication (MTC). MTC may enable providing service for a large amount of MTC capable devices, such as the at least one terminal device 120. The at least one terminal device 120 may comprise mobile phones, smart phones, tablet computers, laptops and other devices used for user communication with the radio communication network, such as a MTC network. These devices may provide further functionality compared to the MTC scheme, such as communication link for voice, video and/or data transfer. However, in MTC perspective the at least one terminal device 120 may be understood as a MTC device. The at least one terminal device 120 may also comprise another MTC capable device, such as a sensor device providing position, acceleration and/or temperature information to name a few examples.

Network-based detection of mobile malware typically relies on communication patterns which have been extracted in advance. The extraction may be carried out by a static and/or dynamic analysis of code samples in an emulation environment which allows to simulate user interaction with the software application and to monitor communication activities of the software application, such as retrieving content from URL. Such extracted communication patterns may then be detected by a traffic sensor in the mobile network, indicating that the device which was sending or receiving this communication pattern is infected with a respective malware.

However, variants of a specific malware may easily be created using dedicated frameworks and generators. Therefore, lots of variants exist which have not been analysed - and probably will never be - and which thus cannot be detected in the network. A new malware type is also not detectable until a sample of the new malware type has been caught, identified as malware and analysed in an emulation environment. In order to be able to detect every malware, it may be required to intensively analyse every existing sample in advance, in order to be able to classify it as malicious, and then to extract communication patterns from the network traffic for the malware detection.

In reality, the dynamic malware analysis is typically restricted to emulating only fractions of the possible execution parameters, as the interaction with the software application is simulated by scripts, and the available execution time for the emulation is limited to a few minutes or less. Malicious activities may therefore not be carried out during the analysis period, yielding a false classification of a benign software application. Besides, modern malware is often able to recognize an ongoing analysis procedure, for instance, by detecting that the

malware is running in an emulation environment. Such malware does not carry out malicious activities while being analysed, causing the analysis to classify it wrongly as a benign software application. As a consequence, the detection rate of mobile malware in the mobile networks tends to be low, and the detection mechanism bears a high risk of missing new types of malware, variants of existing malware and more sophisticated malware.

The effectiveness of a malware detection mechanism based on pre-analysed communication patterns depends on the frequency and quality of the updates which extend the communication pattern list with communication patterns of new malware types and new variants of already known malware. This requires the availability of corresponding malware code samples for the analysis and efficient analysis methods. In practice, typically only subsets of the malware variants and malware types are covered. Besides that, it may take a considerable time for the new malware before it gets analysed. Machine learning methods are theoretically usable for malware detection, but the amount of data which needs to be continuously processed in the mobile network causes the machine learning approach to be unfeasible in practice.

Figure 2 illustrates an embodiment for malware infection detection between a communication device 120 and network element of a cellular communication system, e.g. a network element 110.

Referring to Figure 2, the network element, such as network element 110, monitors (block 201) network connections (block 202) of communication devices (such as terminal device 120) to network resource identifiers. Based on the monitoring, the network element 110 pre-classifies (block 203) a network resource identifier as a suspicious network resource identifier or unsuspicious network resource identifier. Based on the pre-classifying, the network element, performs (block 204) a communication behaviour analysis on communication activities regarding network connections of an individual communication device to a suspicious network resource identifier, wherein the result of the communication behaviour analysis is utilized, in the network element, to detect (i.e. classify, block 205) whether or not the network resource identifier is indicative of a malware infection of the communication devices connecting to the network resource identifier.

Thus, in an embodiment, a network apparatus is configured to perform malware infection detection by evaluating communication data of communication devices. The evaluating may comprise pre-classification and a communication behaviour analysis (see Figure 3). The behaviour analysis comprises monitoring and analysing the communication behaviour of individual communication devices (e.g. terminal devices) in the (cellular) communication network. The pre-classification comprises classifying network resource identifiers (e.g. URLs or domain names) as suspicious or unsuspicious based on the connections to these network resource identifiers from each communication device in the mobile network. Network resource

identifiers classified as unsuspicious in the pre-classification, are not considered in the communication behaviour analysis of individual communication devices, thus reducing the processing effort of the malware infection detection.

In an embodiment, if a device is communicating with URL which has not yet been
5 classified as unsuspicious, the communication behaviour of said device with said URL is monitored and analysed by behaviour analyser functionality. The results of behaviour analyses of multiple devices for a specific potentially suspicious URL are aggregated and evaluated. If the aggregated results show adequate indications of maliciousness, the URL in question is classified as indicative of a malware infection, and each contact of the communication devices to that
10 URL is considered as an evidence of the malware infection.

In an embodiment, malware infections caused by malware including the complete range of variants, may be detected.

In an embodiment, malware infections caused by new types of malware which have not been a subject to a static and/or dynamic code analysis, may be detected.

15 In an embodiment, malware infections caused by malware for which no code samples are identifiable, may be detected.

In an embodiment, malware infections caused by malware for which the static and/or dynamic analysis of an identified code sample is not able to determine any maliciousness at all, and/or for which the static and/or dynamic analysis or is not able to extract communication patterns for malware detection, may be detected.
20

In an embodiment, malware infections caused by malware which is able to hide its malicious activities while being analysed in an emulation environment, may be detected.

In an embodiment, malware infections caused by malware which uses encrypted communication, may be detected (in this case the granularity of the malware detection may be
25 reduced to a network domain name level, if full URL information is not available, e.g. in https).

In an embodiment, instead of relying on a pre-defined fixed set of communication patterns, the malware detection functionality itself is capable of deriving communication patterns for identified malware infections. The functionality may be synergistically usable together with conventional malware detection based on pre-analysed communication patterns.

30 In an embodiment, the malware detection mechanism is able to cope with a large volume of traffic under real-time conditions and is thus superior compared to machine learning anomaly detection algorithms.

During the pre-classification, the number of devices which are connected to a specific URL at least once during a pre-defined time period (e.g. within 24 hours), is tracked. While
35 the traffic is monitored (block 401), URL and device information from the traffic is got (block 402) to check (block 403), whether or not a history list H_i of the device already contains the

URL. The history list, also called URL list, contains those URLs to which the device i in the mobile network has connected during a pre-defined time period (such as during the current day, 48 hours, etc.; i.e. it is a connection history of the device).

If this is the first time during the pre-defined time period the device connects to the URL (block 403: no), the URL is added in block 404 to the history list H_i , associated with time information. Further, for each URL gotten in block 402 and added in block 404 a counter indicating the number of devices in the mobile network which have connected at least once to the URL u during the pre-defined time period is maintained, and the counter is incremented in block 405 by one, the increment being associated with time. Then the process continues the monitoring (block 401). In other words, if the device-specific history list is updated, the URL-specific counter is also incremented.

If within the pre-defined time period the device has earlier connected to the URL (block 403: yes), the process continues the monitoring (block 401).

The information thus collected may be used in pre-classification as will be described with Figures 4B and 4C to detect new mobile malware. The pre-classification may be based on a following assumption on mobile malware which is also backed by the results of practical evaluations in real operator networks. It is assumed that new mobile malware does not have high penetration rates or high increase rates, as there is no active spreading taking place, and that drive-by malware infections are not common with software applications in mobile devices.

Referring to 4B, traffic monitoring includes URL-specific monitoring (block 411) for pre-classifying. If the counter value c for the URL, the counter value c indicating the number of devices in the mobile network which have connected at least once to the URL u during a pre-defined time period, does not exceed a threshold $th1$ (block 412: yes), and the increase rate g of the connected devices during the pre-defined time period does not exceed a corresponding threshold $th2$ (block 413: yes), the traffic will be pre-classified (block 414) as suspicious traffic. The threshold value $th1$ may be equal to known malware infected devices in the network. In other words, a threshold for unsuspicious high connection rates, i.e. $th1$, may be determined by using the number of known malware infected devices in the network, as initially determined e.g. by the communication pattern-based malware detection. For example, the threshold value $th1$ may be the number of known malware infected devices in the network to which number a safety margin (Δ) is added. The Δ , i.e. the safety margin, ensures that URLs generated by new malware leading to increased infection rates will not be classified as benign. The same principles may be used to determine the threshold for increase, i.e. $th2$.

If the number of devices is bigger than the threshold $th1$ (block 412: no), or the increase rate is bigger than the threshold $th2$ (block 413: no), the traffic will be pre-classified (block 415) as unsuspicious. In other words, high enough connection rates and or rapid enough increase indicating unsuspicious traffic can be determined.

The above described pre-classification is contrary to known solutions in which traffic is assumed to be malicious if the amount and/or increase rate exceeds a threshold, i.e. is high and/or rapid enough, unless specifically indicated as benign.

It should be appreciated that the pre-defined time periods used above may be the same, or different time periods may be used for different purposes. For example, the pre-defined time period used in block 412 may be different from the pre-defined time period in block 413.

Figure 4C illustrates an exemplary tracking curve for 30 days, illustrating the number of devices connecting to a specific URL (URL A, URL B, URL C) at least once a day. Using the principle disclosed above and the process of Figure 4B, URLs with constant high connection rates, for example, established sites (e.g. URL C in Figure 4C) and URLs with significant increase rates in their connection curve (identifying increasingly popular sites such as new game or community sites, e.g. URL B in Figure 4C) may be pre-classified as unsuspecting. URLs (URL A) for which the assorted curves are low and slowly increasing, are pre-classified as potentially suspicious. In other words, if the number of communication devices connected to the individual network resource identifier at least once within a pre-defined time period is not greater than a pre-defined connection threshold value, and, if the increase rate of the number of communication devices connected to an individual network resource identifier at least once within a pre-defined time period does not exceed a pre-defined rate threshold value, the individual network resource identifier is pre-classified as potentially suspicious. Communication of devices with these potentially suspicious URLs may be monitored and analysed on a per-device basis by using the behaviour analysis.

In an embodiment, additional characteristics of URLs may be used to assist the pre-classification of URLs, e.g. known benign URLs showing specifically steep or otherwise characteristic communications patterns may be used to further increase the precision of the pre-classifying of unknown URLs.

The behaviour analysis may be performed based on one or more of the following data structures. The URL list PCBEN contains each URL which has been pre-classified as unsuspecting, causing that these URLs are treated as benign. The URL list BABEN contains each URL which has been classified as benign during the behaviour analysis. The URL list BAMAL contains each URL which has been classified as indicative of a malware infection during the behaviour analysis. The blacklist BL contains URLs which have been identified as malware communication typically by off-line analysis of code samples. The URL list H_i containing each URL to which device i in the mobile network has connected during a pre-defined time period (such as during the current day, 48 hours, etc.; connection history of the device). $c_{u,t}$ indicates the number of devices in the mobile network which have connected at least once to URL u during day t . thr_c indicates the threshold for c_u . An average gradient of curve (graph) c_u represents the values for $c_{u,t}$ for a further pre-defined time period (such as several days, a week, 2 weeks, etc.),

i.e. the average gradient of the curve of the number of devices which have connected at least once a day to URL u over a period of days. thr_g indicates the threshold for the gradient of curve c_u .

Figure 5 illustrates an embodiment for malware detection. Referring to Figure 5, when a network apparatus notices (block 501) that a communication device connects to URL u , it may be checked (block 502), if URL u is contained on blacklist BL (which is an optional list containing URLs which have been identified as related to malware, typically by (conventional) off-line analysis of malware code) or on BAMAL (i.e. URL u is already classified as indicative of a malware infection by the malware detection mechanism). If based on the checking it is detected that URL u is contained on BL and/or BAMAL, no further processing is needed, and the device is considered (block 503) as malware infected.

It may also be checked (block 504), if URL u is contained on PCBEN (i.e. based on the pre-classification URL u is considered to be unsuspicious) or on BABEN (i.e. the malware detection mechanism has classified URL u as benign). If based on the checking it is detected that URL u is contained on PCBEN and/or BABEN, no further processing is needed, and the connection between the device and URL u is considered (block 505) as benign.

If these simple cases are ruled out (i.e. it is detected that URL u is not contained on BL, BAMAL, PCBEN and/or BABEN), URL u is considered as a candidate URL for communication behaviour analysis.

The pre-classification may comprise checking (block 506), if URL u is contained in URL list H_i , i.e. if device i has already connected to URL u , for example, during the current day, as any connection from device i to URL u only once per day is counted. If URL u is not contained in URL list H_i , counter $c_{u,t}$ (with t representing the current day, for example) is incremented (block 508) by 1, URL u is added (block 507) to URL list H_i in order to prevent further counting for that device during that day, and the aggregated connection counts for URL u are evaluated in order to determine whether the pre-classification may be carried out. For that purpose, it is checked (block 509) if counter $c_{u,t}$ is greater than threshold thr_c , i.e. if the connection count for URL u for the current day has exceeded a pre-defined threshold value. An upper bound for this threshold thr_c may be assumed to be the actual number of known malware infected devices in the network. If $c_{u,t}$ exceeds thr_c , then URL u may be pre-classified (block 511) as benign (u is added to PCBEN (block 510)) according to the assumption described above, and no further processing is needed.

The pre-classification may further comprise checking (block 512), if the average gradient of the curve of c_u exceeds a pre-defined threshold thr_g . In order URL u to be pre-classified as benign, the curve of c_u , i.e. the curve of the connection count for URL u over a period of several days, is to be significantly increasing: the average gradient of this curve is therefore to exceed a pre-defined threshold thr_g . The average gradient may be computed based on the values of the past m days, where m is as a value chosen, for example, in the range of 3 to 10,

which efficiently limits the amount of required data. If the average gradient of the curve is greater than threshold thr_g , URL u may be pre-classified (block 511) as benign (u is added to PCBEN (block 510)) according to the assumption described above, and no further processing is needed.

5 If neither of the criteria for the pre-classification as benign are fulfilled (i.e. if $c_{u,t}$ does not exceed thr_c /if the average gradient of the curve does not exceed thr_g), the processing continues the same way as if URL u were already contained in URL list H_i while starting the behaviour analysis.

The behaviour analysis may be carried out as one or more dedicated computational modules (block 515) which are run in parallel and implement a specific behaviour analysis procedure, e.g. a periodicity detector. It is checked (block 516), if the aggregated data for URL u are sufficient for a classification decision on whether URL u is indicative of a malware infection or not. If the aggregated data for URL u are sufficient for a classification decision on whether URL u is indicative of the malware infection or not, u is respectively moved via a classifier (block 517) either to BAMAL (block 521, 522) or BABEN (block 519, 520) depending on the classification decision (block 518). As long as the data are not yet sufficient for making the classification decision, the behaviour analysis is continued (block 523) upon detecting the next connection from any device in the network to URL u .

The behaviour analysis typically requires a minimal set of collected data in order to run the parallel computational modules 515. During an initial period where not enough data has yet been collected, optional load reducer functionality enables reducing the computational load by skipping computations related to the search for traces of malicious behaviour. Traces of malicious behaviour are searched in the collected communication data by the respective behaviour analyser module. The load reducer may check (block 513), if enough data has been collected for the behaviour analysis. If enough data has been collected for the behaviour analysis, the process continues to the behaviour analysis (block 515). If not enough data has been collected for the behaviour analysis yet, the process continues by collecting observations (block 514).

Figure 6 illustrates the functioning of the behaviour analyser. For the behaviour analysis, each of the parallel computational modules may monitor (block 601) specific communication activities regarding the connection of the device to URL u . This includes collecting (block 602) and storing relevant data, e.g. for a behaviour analysis module which aims at detecting periodicities in network communication, timestamps of the monitored communication activities are stored. The collected data are processed (block 603) and checked (block 604) for traces of malicious behaviour at every execution of the module. If traces of malicious behaviour are found, the module returns (block 605) a state "malicious" and terminates its execution. If no traces of malicious behaviour are found, the behaviour analysis module has an internal logic

to decide (block 606) if continuation of the behaviour analysis looks promising. For periodicities detection, for instance, the behaviour analysis is continued for a pre-defined period of time. If after that time still no traces of malicious behaviour are found, it may be assumed that the continuation of the behaviour analysis is not promising. Other types of behaviour analysis modules may implement different and/or more complex decision logic.

If the module decides to continue the behaviour analysis, a state "not yet decided" is returned (block 607). If the module does not find traces of malicious behaviour and decides not to continue the analysis, a state "benign" is returned (block 608) and the execution of the module is terminated. In any case, the logic of the modules ensures that one of the terminating states is reached within an appropriate period of time.

Information for the behaviour analysis is collected and analysed on a per-device basis. An individual instance of each type of behaviour analysis module is created for each device which connects to a monitored URL, and subsequent connections of this device to this URL are analysed until the module is taking a decision on whether, based on the individual data for this device, the module considers the URL to be indicative of a malware infection or not (see Figure 7). For the final classification decision, the decision states ("indicative of malware infection" or "benign") which the module instances (each representing the analysis result for a specific device and a specific analysis type) are reporting to the final classification which is a single global instance per analysed URL, are aggregated and evaluated. The final classification depends on the characteristics and the number of the different used types of behaviour analysis modules. For example, if only a single type of behaviour analysis is used (for instance the periodicities detection), the final classification may consider URL as indicative of a malware infection, if in at least half of the devices which had connected to this URL, malicious behaviour has been detected.

The behaviour analyser incorporates knowledge on the communication characteristics of malware, which is implemented by a set of behaviour analysis module types. For example, a periodicities behaviour analysis module may be used to detect malicious communication, as malware is known to regularly communicate with its command and control infrastructure to obtain commands or to update its settings. The inherent periodicities are detected by analysing the recorded time parameters of communication of the device to the specific URL. For example, it may be detected if a malware tries to connect to URL representing its command and control center every 30 minutes.

In an embodiment, the combination of the pre-classifier and behaviour analyser is used, however, implementations which use only the pre-classifier or only the behaviour analyser, are also possible.

An embodiment provides an apparatus comprising at least one processor and at least one memory including a computer program code, wherein the at least one memory and

the computer program code are configured, with the at least one processor, to cause the apparatus to carry out the procedures of the above-described network element or network node. The at least one processor, the at least one memory, and the computer program code may thus be considered as an embodiment of means for executing the above-described procedures of the network element or the network node. Figure 8 illustrates a block diagram of a structure of such an apparatus. The apparatus may be comprised in the network element or in the network node, e.g. the apparatus may form a chipset or a circuitry in the network element or in the network node. In some embodiments, the apparatus is the network element or the network node. The apparatus comprises a processing circuitry 10 comprising the at least one processor. The processing circuitry 10 may comprise a monitor 12 configured to monitor network connections of communication devices to network resource identifiers. The processing circuitry 10 may comprise a pre-classifier 14 configured to pre-classify a network resource identifier as a suspicious network resource identifier or unsuspicious network resource identifier. The pre-classifier 14 may be configured to pre-classify the network resource identifier as a suspicious network resource identifier or unsuspicious network resource identifier, as described above, and output information on the suspicious network resource identifier to a behaviour analyser 16 configured to perform a communication behaviour analysis on communication activities regarding network connections of an individual communication device to the suspicious network resource identifier. The behaviour analyser 16 may be configured to perform the communication behaviour analysis, as described above, and output information on the result of the communication behaviour analysis to a classifier 18 configured to detect whether or not the network resource identifier is indicative of a malware infection of the communication devices connecting to the network resource identifier.

The processing circuitry 10 may comprise the circuitries 12, 14, 16 and 18 as sub-circuitries, or they may be considered as computer program modules executed by the same physical processing circuitry. The memory 20 may store one or more computer program products 24 comprising program instructions that specify the operation of the circuitries 12 to 18. The memory 20 may further store a database 26 comprising definitions for the behaviour analysis and pre-classification, for example. The apparatus may further comprise a radio interface 22 providing the apparatus with radio communication capability with the terminal devices. The radio interface may comprise a radio communication circuitry enabling wireless communications and comprise a radio frequency signal processing circuitry and a baseband signal processing circuitry. The baseband signal processing circuitry may be configured to carry out the functions of a transmitter and/or a receiver. In some embodiments, the radio interface may be connected to a remote radio head comprising at least an antenna and, in some embodiments, radio frequency signal processing in a remote location with respect to the base station. In such embodiments, the radio interface may carry out only some of radio frequency signal processing or no radio frequency signal processing at all. The connection between the radio interface and

the remote radio head may be an analogue connection or a digital connection. In some embodiments, the radio interface may comprise a fixed communication circuitry enabling wired communications.

As used in this application, the term 'circuitry' refers to all of the following: (a) hardware-only circuit implementations such as implementations in only analog and/or digital circuitry; (b) combinations of circuits and software and/or firmware, such as (as applicable): (i) a combination of processor(s) or processor cores; or (ii) portions of processor(s)/software including digital signal processor(s), software, and at least one memory that work together to cause an apparatus to perform specific functions; and (c) circuits, such as a microprocessor(s) or a portion of a microprocessor(s), that require software or firmware for operation, even if the software or firmware is not physically present.

This definition of 'circuitry' applies to all uses of this term in this application. As a further example, as used in this application, the term "circuitry" would also cover an implementation of merely a processor (or multiple processors) or portion of a processor, e.g. one core of a multi-core processor, and its (or their) accompanying software and/or firmware. The term "circuitry" would also cover, for example and if applicable to the particular element, a baseband integrated circuit, an application-specific integrated circuit (ASIC), and/or a field-programmable grid array (FPGA) circuit for the apparatus according to an embodiment of the invention.

The processes or methods described above in connection with Figures 1 to 8 may also be carried out in the form of one or more computer process defined by one or more computer programs. The computer program shall be considered to encompass also a module of a computer programs, e.g. the above-described processes may be carried out as a program module of a larger algorithm or a computer process. The computer program(s) may be in source code form, object code form, or in some intermediate form, and it may be stored in a carrier, which may be any entity or device capable of carrying the program. Such carriers include transitory and/or non-transitory computer media, e.g. a record medium, computer memory, read-only memory, electrical carrier signal, telecommunications signal, and software distribution package. Depending on the processing power needed, the computer program may be executed in a single electronic digital processing unit or it may be distributed amongst a number of processing units.

The present invention is applicable to cellular or mobile communication systems defined above but also to other suitable communication systems. The protocols used, the specifications of cellular communication systems, their network elements, and terminal devices develop rapidly. Such development may require extra changes to the described embodiments. Therefore, all words and expressions should be interpreted broadly and they are intended to illustrate, not to restrict, the embodiment.

Thus, a computer program product is disclosed, embodied on a non-transitory distribution medium readable by a computer and comprising program instructions which, when loaded into the computer, execute a computer process comprising causing a network node to perform the method.

5 Further, an apparatus is disclosed comprising means for monitoring network connections of communication devices to a network resource identifier in a communication system; means for pre-classifying the network resource identifier as a suspicious network resource identifier or unsuspicious network resource identifier, based on the monitoring; and means for performing, based on the pre-classifying, a communication behaviour analysis on
10 communication activities regarding network connections of an individual communication device to a suspicious network resource identifier, wherein the result of the communication behaviour analysis is utilized to detect whether an individual communication device is malware infected or not.

Yet further, an apparatus is disclosed comprising a monitor for monitoring network connections of communication devices to a network resource identifier in a communication system; a pre-classifier for pre-classifying the network resource identifier as a suspicious network resource identifier or unsuspicious network resource identifier, based on the monitoring; and a behaviour analyser for performing, based on the pre-classifying, a communication behaviour analysis on communication activities regarding network connections of an individual communication device to a suspicious network resource identifier, wherein the result of the communication behaviour analysis is utilized to detect whether an individual communication device is malware infected or not.

It will be obvious to a person skilled in the art that, as the technology advances, the inventive concept can be implemented in various ways. The invention and its embodiments are
25 not limited to the examples described above but may vary within the scope of the claims.

List of abbreviations

	BA	behaviour analyser
	H	history
30	BL	blacklist
	BAMAL	behavioural analyser malicious (list of URLs)
	BABEN	behavioural analyser benign (list of URLs)
	PCBEN	pre-classifier benign (list of URLs)

CLAIMS

1. A method for mobile malware infection detection, the method comprising:
monitoring, in a network node, network connections of communication devices to
network resource identifiers in a communication system;
5 based on the monitoring, pre-classifying, in the network node, a network resource
identifier as a suspicious network resource identifier, if the number of communication devices
connected to the network resource identifier at least once within a pre-defined time period is
not greater than a pre-defined connection threshold value, and, if the increase rate of the num-
ber of communication devices connected to the network resource identifier at least once within
10 a pre-defined time period does not exceed a pre-defined rate threshold value.
2. A method according to claim 1, the method comprising performing, in the net-
work node, based on the pre-classifying, a communication behaviour analysis on communica-
tion activities regarding network connections of an individual communication device to a sus-
picious network resource identifier, wherein the result of the communication behaviour anal-
15 ysis is utilized, in the network node, to detect whether or not the network resource identifier
is indicative of a malware infection of the communication devices connecting to the network
resource identifier.
3. A method according to claim 1 or 2, the method comprising
aggregating, in the network node, the results of the communication behaviour
20 analyses on communication activities regarding network connections of multiple communica-
tion devices to an individual suspicious network resource identifier; and
if the aggregated results indicate an adequate level of maliciousness, classifying, in
the network node, said individual suspicious network resource identifier as indicative of the
malware infection, wherein each connection of the communication devices to said individual
25 suspicious network resource identifier is considered as an evidence of the malware infection.
4. A method according to claim 1, 2 or 3, wherein
the pre-classifying comprises tracking, in the network node, the number of com-
munication devices connected to an individual network resource identifier at least once within
a pre-defined time period.
- 30 5. A method according to any of claims 1 to 4, wherein the method comprises pre-
classifying, in the network node, a network resource identifier as an unsuspicious network re-
source identifier, if the number of communication devices connected to the network resource
identifier at least once within the pre-defined time period has exceeded the pre-defined con-
nection threshold value.
- 35 6. A method according to any of claims 1 to 5, wherein the method comprises stor-
ing, in the network node, one or more of the following data:

a first network resource identifier list containing information on each network resource identifier pre-classified as unsuspicious, wherein the network resource identifiers on first network resource identifier list are treated as benign network resource identifiers,

5 a second network resource identifier list containing information on each network resource identifier classified as benign during the communication behaviour analysis,

a third network resource identifier list containing information on each network resource identifier classified during the communication behaviour analysis as indicative of the malware infection,

10 a fourth network resource identifier list containing information on network resource identifiers identified as related to malware communication and possibly generated by an off-line analysis of code samples, and

a fifth network resource identifier list containing information on each network resource identifier to which an individual communication device has connected during a pre-defined time period.

15 7. A method according to claim 6, wherein the method comprises, in the network node,

checking, in response to a communication device connecting to a network resource identifier, if said network resource identifier is contained on at least one of the third and fourth network resource identifier list,

20 wherein, if based on the checking it is detected that the network resource identifier is contained on at least one of the third and fourth network resource identifier list, the network connection between said communication device and said network resource identifier is considered as indicative of the malware infection.

25 8. A method according to claim 6 or 7, wherein the method comprises, in the network node,

checking, in response to a communication device connecting to a network resource identifier, if said network resource identifier is contained on at least one of the first and second network resource identifier list,

30 wherein, if based on the checking it is detected that the network resource identifier is contained on at least one of the first and second network resource identifier list, the network connection between said communication device and said network resource identifier is considered as benign in the sense of not being related to malware activities.

9. A method according to claim 6, 7 or 8, wherein the method comprises, in the network node,

35 checking, in response to a communication device connecting to a network resource identifier, if said network resource identifier is contained on at least one of the first, second, third, and fourth network resource identifier list,

wherein, if based on the checking it is detected that the network resource identifier is not contained on at least one of the first, second, third, and fourth network resource identifier list, said network resource identifier is considered as a candidate network resource identifier for communication behaviour analysis.

5 10. A method according to any of claims 6 to 9, wherein the method comprises, if a network resource identifier is considered as a candidate network resource identifier for communication behaviour analysis,

checking, in the network node, in response to a communication device connecting to said network resource identifier, if said network resource identifier is contained on the fifth
10 network resource identifier list,

wherein, if based on the checking it is detected that the network resource identifier is not contained on the fifth network resource identifier list, the method comprises
incrementing a network resource identifier-specific connection counter by 1; and
adding said network resource identifier on the fifth network resource identifier
15 list.

11. A method according to any of claims 6 to 10, wherein the method comprises, if it is detected that the network resource identifier is contained on the fifth network resource identifier list,

checking if the number of communication devices connected to an individual network resource identifier at least once within a pre-defined time period is greater than a pre-defined connection threshold value,

wherein if the number of the communication devices connected to the individual network resource identifier at least once within the pre-defined time period is greater than the pre-defined connection threshold value, the method comprises
25 pre-classifying said network resource identifier as benign.

12. A method according to any of claims 6 to 11, wherein the method comprises, if it is detected that the network resource identifier is contained on the fifth network resource identifier list,

checking if the increase rate of the number of communication devices connected to an individual network resource identifier at least once within a pre-defined time period exceeds a pre-defined rate threshold value,

wherein if the increase rate of the number of communication devices connected to an individual network resource identifier at least once within the pre-defined time period exceeds the pre-defined rate threshold value, the method comprises
35 pre-classifying said network resource identifier as benign.

13. A method according to any of claims 1 to 12, wherein the method comprises, if the number of communication devices connected to an individual network resource identifier at least once within a pre-defined time period is not greater than a pre-defined connection

threshold value, and if the increase rate of the number of communication devices connected to an individual network resource identifier at least once within the pre-defined time period does not exceed a pre-defined rate threshold value,

performing the communication behaviour analysis for said network resource identifier by monitoring the communication activities regarding each individual communication device connecting to said network resource identifier.

14. An apparatus comprising
at least one processor; and
at least one memory including a computer program code, wherein the at least one
memory and the computer program code are configured, with the at least one processor, to
cause the apparatus to

monitor network connections of communication devices to a network resource identifier in a communication system;

based on the monitoring, pre-classify the network resource identifier as a suspicious network resource identifier, if the number of communication devices connected to the network resource identifier at least once within a pre-defined time period is not greater than a pre-defined connection threshold value, and, if the increase rate of the number of communication devices connected to the network resource identifier at least once within a pre-defined time period does not exceed a pre-defined rate threshold value.

15. An apparatus according to claim 14, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus to, based on the pre-classifying, perform a communication behaviour analysis on communication activities regarding a network connection of an individual communication device to a suspicious network resource identifier, wherein the result of the communication behaviour analysis is utilized to detect whether the individual communication device is malware infected or not.

16. An apparatus according to claim 14 or 15, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus to perform the communication behaviour analysis for said network resource identifier pre-classified as suspicious by monitoring the communication activities regarding each individual communication device connecting to said network resource identifier.

17. An apparatus according to claim 14, 15 or 16, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus to perform any of the method steps of claims 2 to 12.

18. An apparatus comprising means for performing a method according to any of claims 1 to 13.

19. A non-transitory computer readable media having stored thereon instructions that, when executed by a computing device, cause the computing device to:

monitor network connections of communication devices to a network resource identifier in a communication system;

5 based on the monitoring, pre-classify the network resource identifier as a suspicious network resource identifier, if the number of communication devices connected to the network resource identifier at least once within a pre-defined time period is not greater than a pre-defined connection threshold value, and, if the increase rate of the number of communication devices connected to the network resource identifier at least once within a pre-defined
10 time period does not exceed a pre-defined rate threshold value.

20. A non-transitory computer readable media as claimed in claim 19, having stored thereon further instructions that, when executed by a computing device, cause the computing device further to, based on the pre-classifying, perform a communication behaviour analysis on communication activities regarding a network connection of an individual communication device to a suspicious network resource identifier, wherein the result of the communication behaviour analysis is utilized to detect whether the individual communication device
15 is malware infected or not.

21. A non-transitory computer readable media as claimed in claim 19 or 20, having stored thereon further instructions that, when executed by a computing device, cause the computing device further to to perform the communication behaviour analysis for said network resource identifier pre-classified as suspicious by monitoring the communication activities regarding each individual communication device connecting to said network resource identifier.
20

22. A computer program product comprising program instructions configuring an apparatus to perform any of the steps of a method as claimed in any of claims 1 to 13 when the
25 computer program is run.

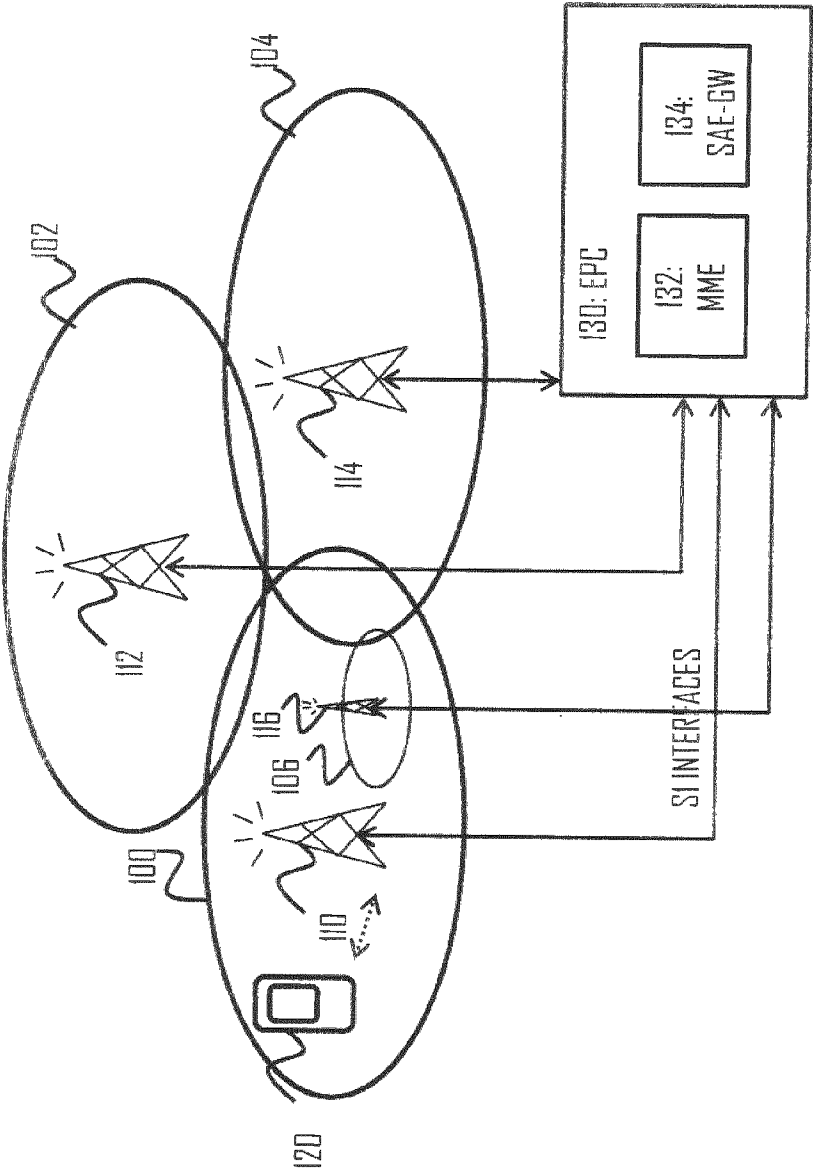
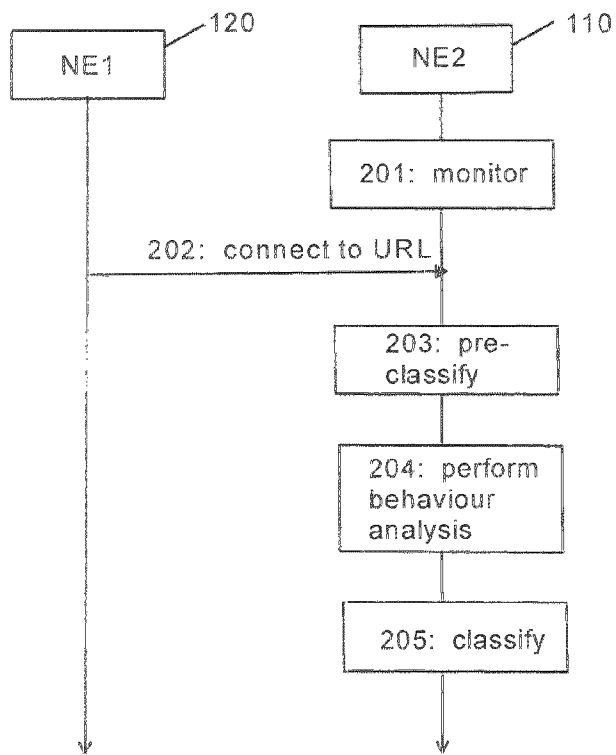


Fig 1

2/8

*Fig. 2*

3/8

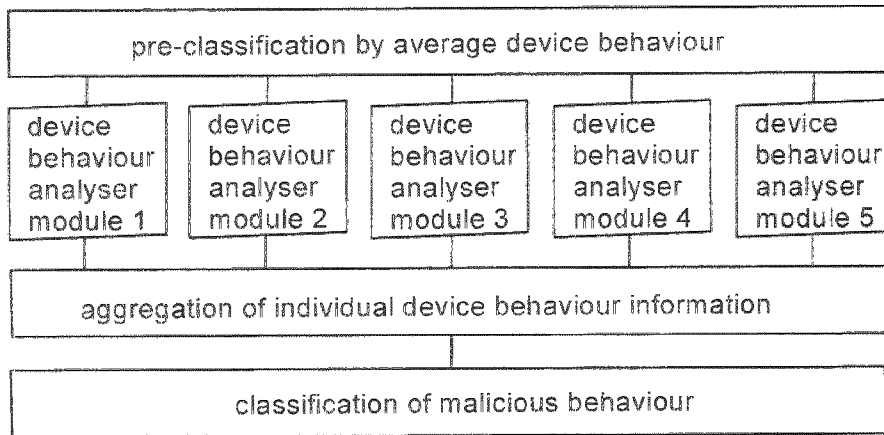


Fig. 3

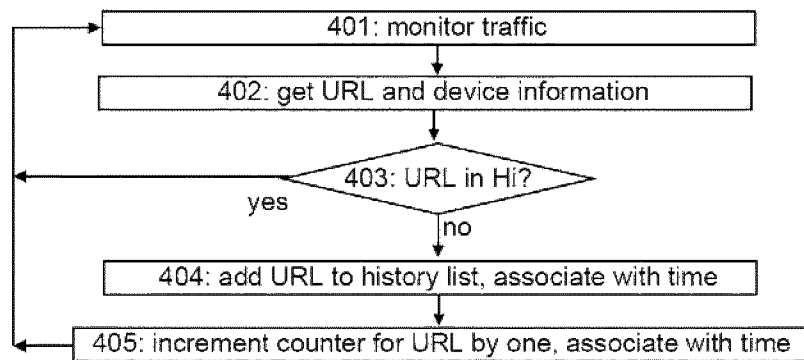


Fig. 4A

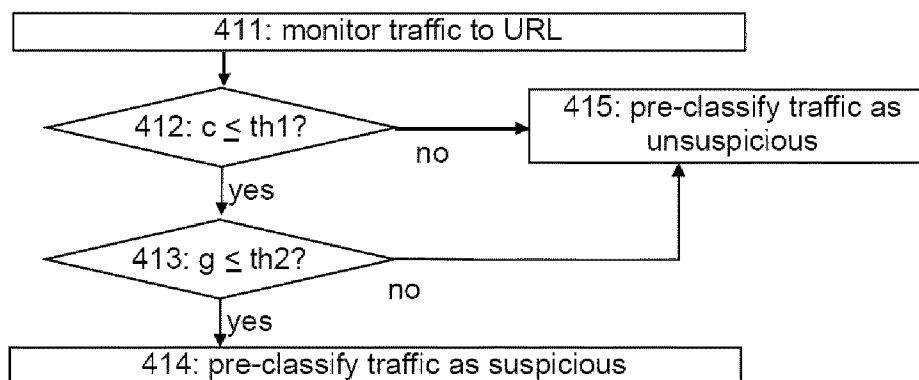


Fig. 4B

4/8

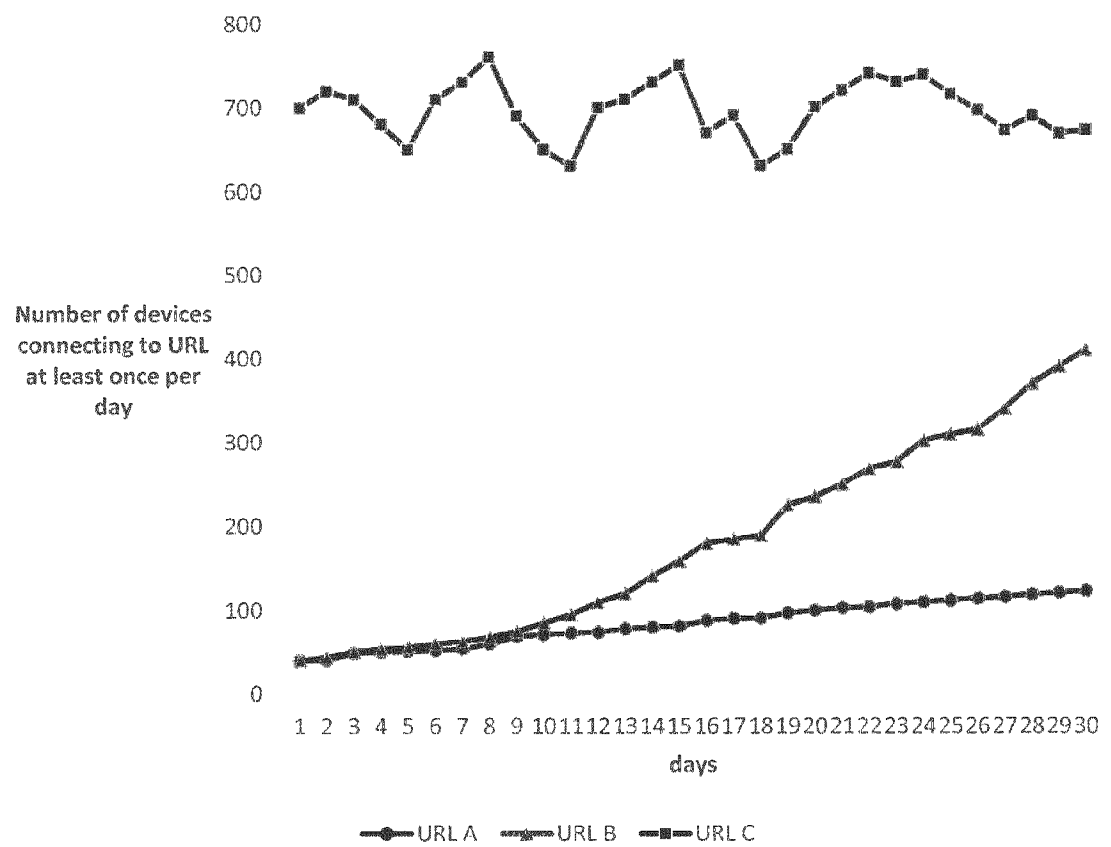
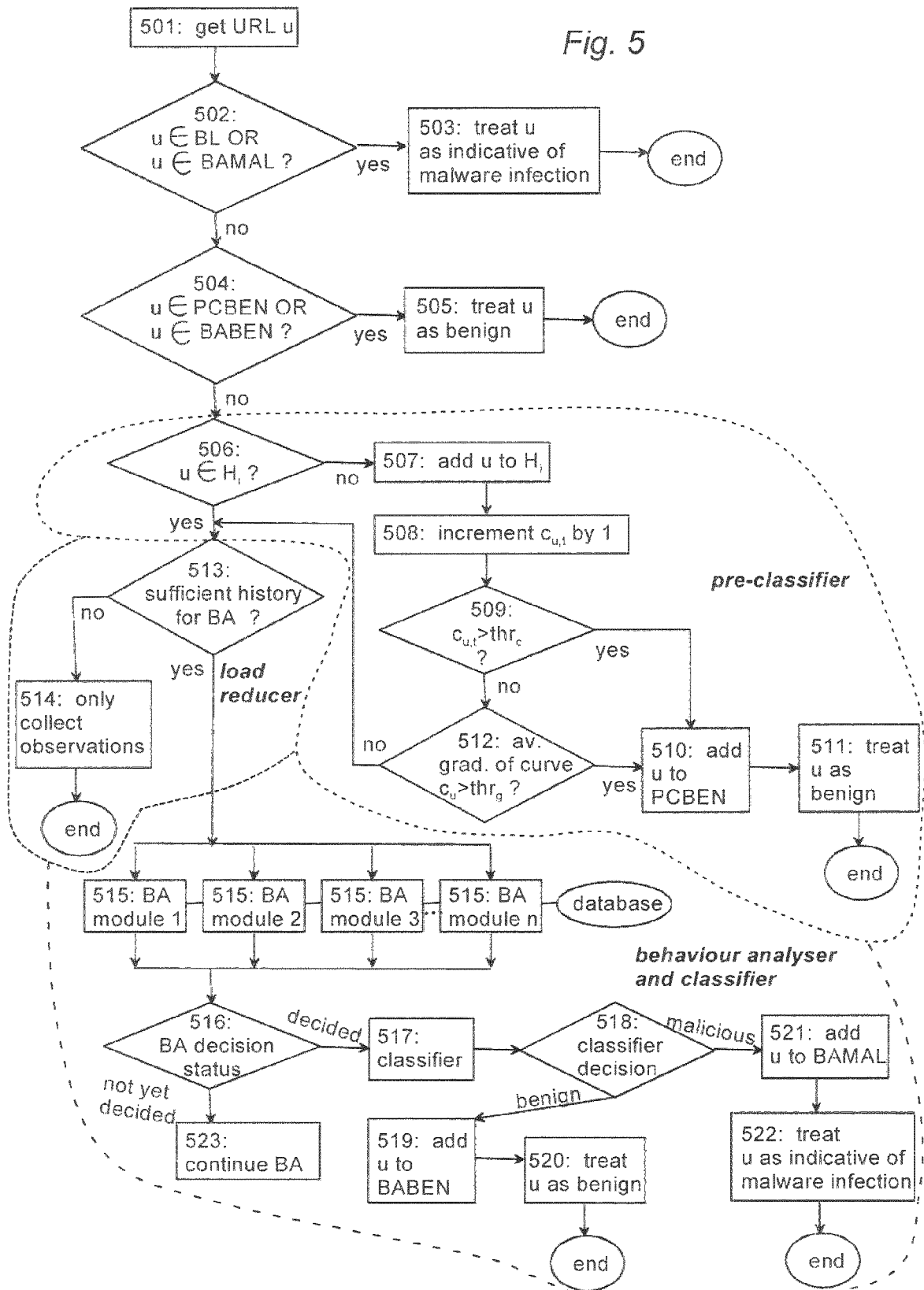


Fig. 4 C

5/8

Fig. 5



6/8

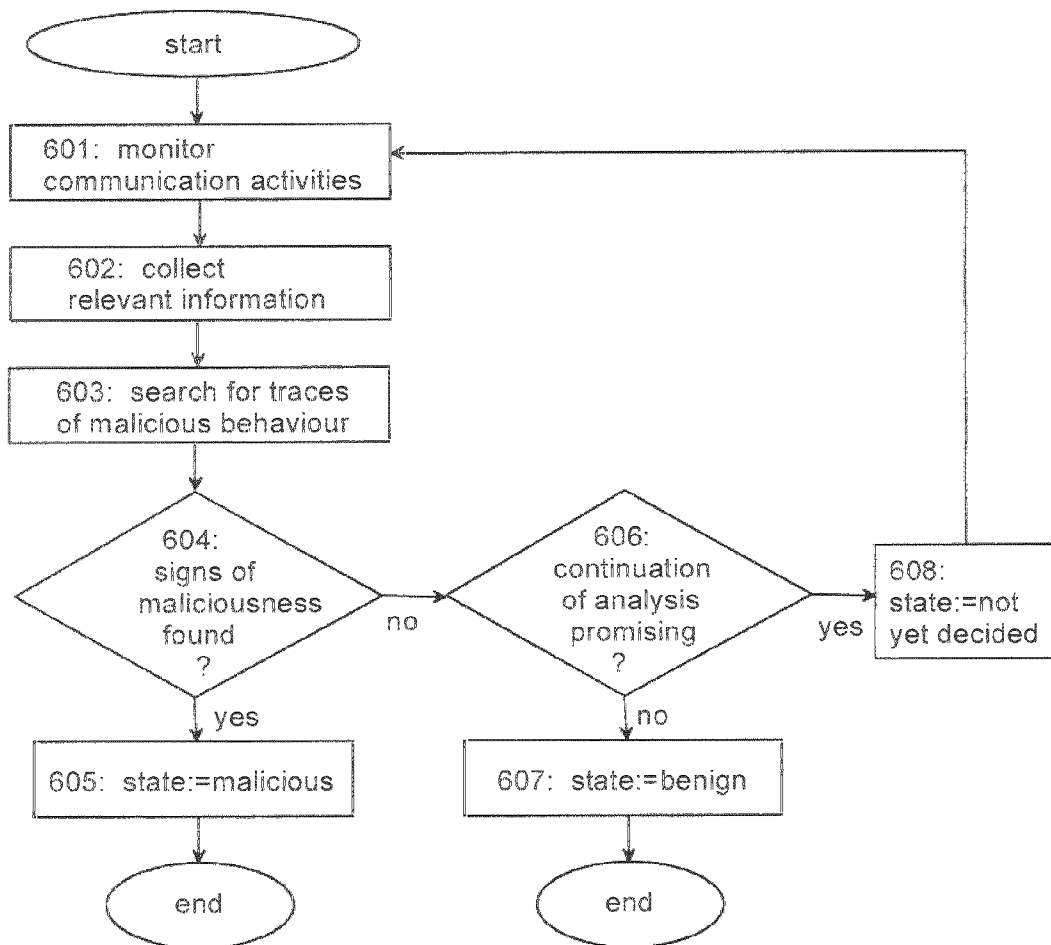


Fig. 6

7/8

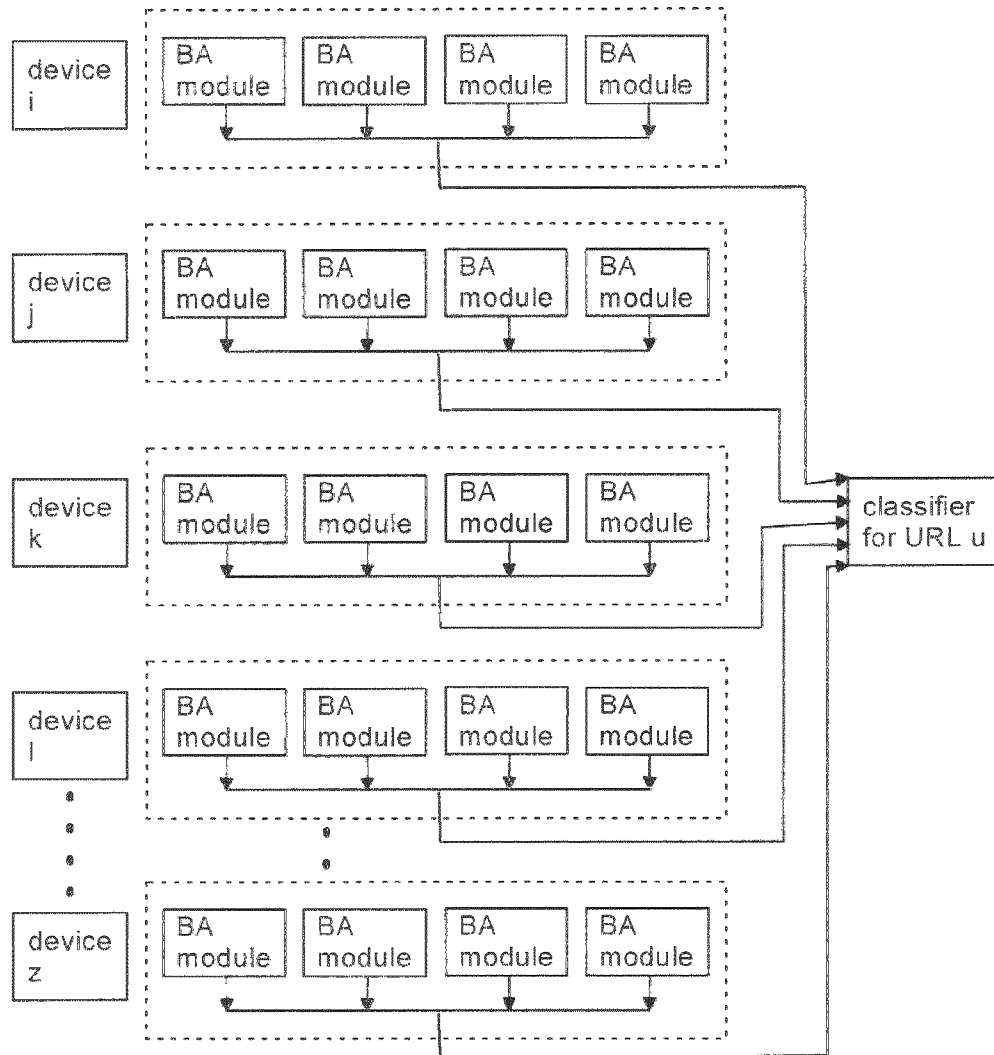
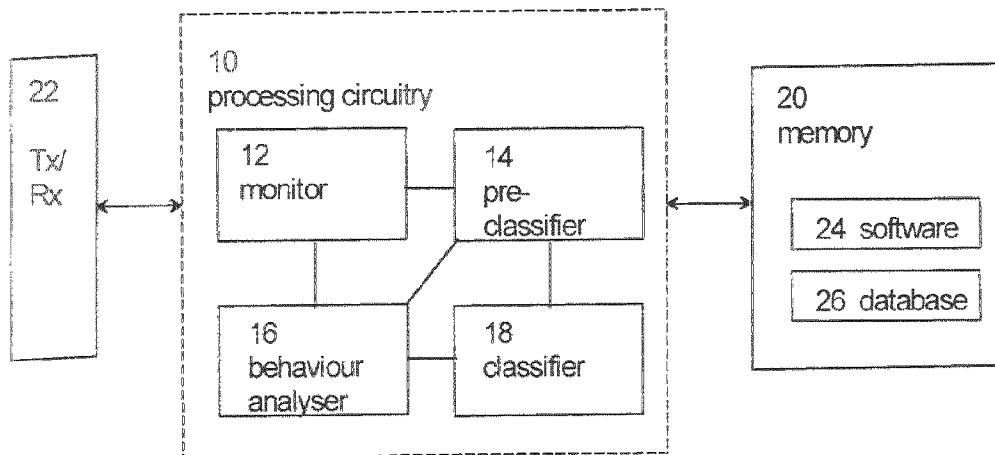


Fig. 7

8/8

*Fig. 8*

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/053363

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04L29/06 H04L29/08
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04L G06F H04W

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2015/141665 A1 (NIPPON TELEGRAPH & TELEPHONE [JP]) 24 September 2015 (2015-09-24) paragraph [0005] paragraph [0006] paragraph [0013] paragraph [0021] paragraph [0022] paragraph [0028] figures 3,6,9	1-22
X	US 9 049 221 B1 (YEN TING-FANG [US] ET AL) 2 June 2015 (2015-06-02) column 5, line 11 - line 43 column 2, line 58 - column 3, line 5 ----- -/-	1-22



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

4 April 2017

Date of mailing of the international search report

13/04/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Medvesan, Oana

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/053363

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 9 195 826 B1 (FANG CHUNSHENG [US] ET AL) 24 November 2015 (2015-11-24) column 7, line 11 - line 46 -----	1-22
A	US 2012/317642 A1 (ROYAL PAUL [US] ET AL) 13 December 2012 (2012-12-13) paragraph [0009] paragraph [0119] -----	1-22
A	US 2012/102545 A1 (CARTER III RUSSELL BERTRAND [US] ET AL) 26 April 2012 (2012-04-26) paragraph [0020] -----	1-22

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2017/053363

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015141665 A1	24-09-2015	EP 3101580 A1 JP 6030272 B2 US 2017070520 A1 WO 2015141665 A1	07-12-2016 24-11-2016 09-03-2017 24-09-2015
US 9049221 B1	02-06-2015	US 9049221 B1 US 9503468 B1	02-06-2015 22-11-2016
US 9195826 B1	24-11-2015	NONE	
US 2012317642 A1	13-12-2012	NONE	
US 2012102545 A1	26-04-2012	CN 103221959 A EP 2630611 A2 JP 5702470 B2 JP 5973017 B2 JP 2013541781 A JP 2015146188 A KR 20140051102 A US 2012102545 A1 WO 2012054449 A2	24-07-2013 28-08-2013 15-04-2015 17-08-2016 14-11-2013 13-08-2015 30-04-2014 26-04-2012 26-04-2012