

COMPUTING DEVICE AND METHOD FOR DETECTING MALICIOUS DOMAIN NAMES IN A NETWORK TRAFFIC

Field

5 The present technique relates to the field of information security in computer systems and data networks, in particular to a computing device and method for detecting malicious domain names in network traffic.

Background

10 Nowadays, due to the continuous scientific and technological progress in the field of information technologies, information security is becoming increasingly important, in particular, protection against various malicious programs that allow attackers to remotely control infected computers (each computer individually, part of computers belonging to the network and/or the entire network).

15 A computer network consisting of a number of computers running autonomous malicious software (bots) running on them is called a botnet. Bots, as such, as part of a botnet, are not viruses, but are essentially a set of software that is secretly installed on a victim's computer and allows an attacker to perform some kind of malicious activity using the resources of an infected computer; firewalls, programs for remotely controlling a computer, as well as tools to hide from
20 the operating system. One of the possible sources of infection can be the Internet, local area network, flash drives, etc.

 Botnets have powerful computing resources, are a common cyber attack tool and a good way to make money for intruders. At the same time, the owner of a botnet can control anywhere infected from infected computers on the network: from another city, country or even from
25 another continent, and the organization of the Internet allows you to do this anonymously. In particular, botnets can be used by intruders to accomplish the following tasks and/or implement the following malicious actions:

1. Spamming.

 Spamming is the most common and one of the easiest botnet exploitation options.
30 According to expert estimates, more than 80% of spam is now sent from infected computers. Spam from botnets is not necessarily sent by network owners. For a fee, "spammers" can rent a botnet. Thousands of botnets allow spammers to send millions of mailings from infected machines in a short period of time. Addresses from which spam is actively sent are often blacklisted by mail servers, and emails coming from them are blocked or automatically marked
35 as spam, but the ability to send spam from a huge number of infected computers prevents you from using the same addresses for sending. In addition, botnets also allow you to collect email

addresses on infected computers, which can be sold to spammers or used when sending spam by the botnet owners themselves.

2. Cyber blackmail.

Botnets are also widely used for Distributed Denial of Service (DDoS) attacks. In the course of such distributed attacks from computers infected by a bot, a stream of false requests to the attacked server on the network is created. As a result, the server becomes unavailable to users due to overload. As a rule, attackers demand a "ransom" for stopping the attack. In today's world, many companies work only through the Internet, so for them the inaccessibility of servers means a complete stop of business, which can lead to significant financial losses.

3. Anonymous access to the network.

Attackers can access servers on the network using infected computers and on behalf of infected machines can commit cybercrime, for example, can hack websites or transfer stolen money.

4. Sale and rental of botnets.

One of the options for illegal earnings with the help of botnets is to rent or sell a botnet. Creating botnets for sale is a separate area of cybercrime business.

5. Phishing.

The addresses of phishing pages can be quite quickly blacklisted, but the botnet allows phishers to quickly change the address of the phishing page using infected computers as proxies, which allows them to hide the real address of the phisher's web server.

6. Theft of confidential data.

Botnets can also be used to steal various passwords (for example, access to e-mail, Skype, social networks, FTP resources, web services such as remote banking, etc.) and other confidential user data. A bot with which computers are infected in a botnet can download another malicious program, for example, a Trojan that steals passwords. In this case, all the computers included in this botnet will be infected with a Trojan program, and attackers will be able to get passwords from all infected computers. Stolen passwords can be resold or used, in particular, for mass infection of web pages (for example, passwords for all found FTP accounts) for further distribution of malicious bot software and botnet expansion.

Managing a computer that is infected with a bot can be direct and indirect. In the case of direct control, an attacker can establish communication with the infected computer and control it using the commands built into the body of the bot program. In the case of indirect control, the bots themselves establish contact with special command and control (C&C) servers or other computers on the network, send a request and execute the received command. One of the main commands received by bots from C&C servers are the following control commands:

1. "Update" type command.

The essence of this command is to download and run a given executable file or module from a given server. This command is basic, since it is implemented in the first place. It allows you to update the bot's executable file on the orders of the botnet owner in the case in which the owner wants to install an upgraded version of the bot. The same command allows you to infect your computer with other malware (viruses, worms), as well as install other bots on your computer. Using this command, Trojans can be simultaneously installed on all computers that search for all passwords ever entered on this computer and stored in its memory and send them to a server on the Internet.

2. "Flood" type command.

The essence of this command is to begin the process of creating a stream of false requests to a given server on the network to disable this server or to overload the Internet channel of a given global network segment. Creating such a stream can cause serious server problems leading to its inaccessibility for ordinary users.

3. "Spam" type command.

The essence of this command is to load a spam message template and start sending spam to specified addresses (there is a separate portion of addresses for each bot).

4. "Proxy" type command.

The essence of this command is to use a computer from a botnet as a proxy server to hide the real address of the attacker who controls the botnet. This functionality is often not separated into a separate command, but immediately included in the general functionality of the bot.

5. Other commands.

There are other less popular commands that are implemented in individual bots. For example, these additional commands allow to receive copies of the image from the user's screen, monitor the input of passwords from the keyboard, request a file with the user's network communication protocol (used to steal accounts and confidential data), send the specified file from the user's computer, request serial numbers, receive detailed information about the user's system and its environment, request a list of computers included in the botnet, etc.

To keep control and management of infected computers, botnets use a variety of methods, such as peer-to-peer networks, mail protocols, social networks or anonymous networks, such as TOR and i2p, but currently the most common domain generation (DGA) algorithms used to generate a large number of pseudo-random domain names, allowing them to establish a connection with a C&C computing device, and allowing to protect the C&C server from a single disconnection or adding its address blacklisted.

Infected computers that are included, for example, in the corporate network can be detected using antivirus programs installed on such computers. In addition, the task of detecting infected computers on a network can be solved centrally by analyzing network traffic, which

provides proactive protection and increases the transparency of the distribution of malicious code in a specified network.

One of the approaches to analyzing traffic is based on a predefined set of rules characterizing the activity of malicious software (software). This approach works well for detecting known malware, but it is almost useless against new threats. In this case, the situation can be corrected by adding learning algorithms.

When an infected computer establishes a communication channel with a C&C computing device, malware settings can be intercepted, which may include a list of C&C servers with which communication is planned, allowing you to track the intruder and identify other infected computers.

When using the DGA algorithm in malware, bots select one of the active domains based on some changing parameters, and in this case, each time the malware can select new domains for communication, and a static domain block will not neutralize the malware. To detect an infected computer, such malware using the DGA algorithm, in the prior art, uses a machine learning (ML) approach to look for anomalies in application data transfer protocols, in which the following rules are used to detect malware with the DGA algorithm:

- domain names include unnatural combinations of letters that are not found in natural language (for example, numbers in the middle of a word);
- most domains do not exist in reality;
- search of a large number of domains is carried out in a very short period of time.

It should be noted that the goal of any machine learning algorithm used to solve the above problem is to train the classifier to distinguish malicious activity that meets the above criteria.

One of the illustrative examples of detection of infected computers on the network is described in US patent No. 8402543 (publ. on March 19, 2013). In particular, US 8402543 discloses a system and method for detecting bots in network traffic by using a classification model trained during the training phase using machine learning algorithms on features extracted from network data packets associated with a known malicious client or a known non-malicious client, and applying a trained classification model to features extracted in real-time from current network data, , indicates the presence of an active communication channel between the malicious client and the controlling computing device: the number of data streams transmitted over a specified period of time, the number of data packets transmitted over a specified period of time, and the number of bits transmitted over a specified period of time.

Another illustrative example of detecting infected computers on a network is described in US Patent No. 8555388 (publ. on October 8, 2013). In particular, it is disclosed in US 8555388 the system and method for heuristic detection of botnet network by tracking all network traffic to identify it malicious network traffic and by detecting the bot using heuristic analysis of the

behavior of malicious network traffic using the processor, which includes command and control traffic associated with the control bot, while the monitored network traffic is assigned an estimate corresponding to the probability of attributing this monitored network traffic to the botnet network, the assigned score is increased if additional indicators of suspicious behavior associated with the monitored network traffic are met, and it is determined whether the suspicious behavior is associated with the botnet network based on the received score.

It should be noted that the detection mechanisms of infected devices in the data network, described in US 8402543 or US 8555388 patents, require significant computational resources and provide an insufficient probability of detecting infected devices, resulting in the implementation of a protection system at the level of end workstations or servers based on these detection mechanisms are often either impossible or inefficient.

Thus, there is an obvious need for further improvement of systems and methods of protection against malicious actions, for the organization of which the attackers use infected devices included in the botnet, in particular for the effective distribution of computing resources spent on providing such automated protection and increasing the probability of detecting such infected devices in the data network.

Consequently, the technical problem solved by the present technique consists in creating a system and a method for detecting infected devices in the data network, in which at least partially the above-mentioned lack of known systems and detection methods is eliminated, which consists in inefficient use of computing resources and a low probability of detecting malicious devices in the data network while providing automated protection.

Disclosure

The above-mentioned technical problem is solved in one aspect of the present technique, according to which a computing device for detecting malicious domain names in network traffic is proposed. The device comprises: a communication module configured to obtaining network traffic from a data network; a filtering module connected to the communication module to receive captured network traffic from it and perform at least the following operations: (i) extracting a plurality of data packets from the received network traffic, (ii) analyzing said extracted data packets to ensure that at least one domain name is extracted from them; an analyzing module configured to connect to the filtering module to obtain the specified domain names from it and to analyze each of the obtained domain names using a given set of analysis techniques and to assign a given numerical value to each of the specified set of features of a domain name suspiciousness corresponding to one of the given set of analysis methods for each analyzed domain name, depending on the results of its analysis using the specified methods of analysis; a processing module configured to connect to the analyzing module to obtain from it the specified features of suspiciousness with the assigned numerical values for

each domain name and with the possibility of their analysis using a given set of analysis rules to ensure that each domain name is classified as malicious domain names, if the results of the analysis of features of suspiciousness are characteristic of malicious domain names.

In one embodiment of the technique, the communication module in the computing device may be further configured to connect to at least one of the devices to capture network traffic included in the data network.

In another embodiment of the technique, the filtering module in the computing device may be further configured to determine if there is an analytical report for each of the extracted domain names.

In another embodiment of the technique, when detecting the presence of an analytical report, the filtering module in the computing device can perform at least the following operations: (i) obtaining data on analytical reports, (ii) using character-by-character comparison, establishing each of the extracted domain names with domain names, identifying analytical reports in the indicated obtained data, the correspondence of each of the extracted domain names to one of the available analytical reports.

In some embodiments of the technique, in response to the fact that there is no analytical report for the extracted domain name, the filtering module in the computing device may be further configured to transfer the specified domain name to the analyzing module.

In other embodiments of the present technique, when analyzing the obtained domain names, the analyzing module in the computing device may perform at least the following operations: (i) determining the number of characters in each of the obtained domain names, (ii) determining whether a certain number of characters of each analyzed domain name corresponds to a given threshold value of the length of the domain name, (iii) assigning a numerical value to the feature of a domain name suspiciousness representing the length of the analyzed domain name, depending on the result of the conformity assessment.

In some embodiments of the present technique, when analyzing the obtained domain names, the analyzing module in the computing device may additionally perform at least the following operations: (i) obtaining a set of known domain names, (ii) using character-by-character comparison of each of the extracted domain names with domain names, establishing the names in the obtained set of domain names, the coincidence of each of the extracted domain names with one of the known domain names, (iii) assignment of a numerical value to another feature of a domain name suspiciousness, representing the popularity of the analyzed domain name, depending on the result of establishing a match.

According to one embodiment of the present technique, when analyzing the obtained domain names, the analyzing module in the computing device may additionally perform at least the following operations: (i) determining the appearance frequencies of each of the plurality of N-grams in each of the obtained domain names, each N-gram corresponds to a combination of

N consecutive characters, (ii) determining the entropy of the N-gram of the specified analyzed domain name as a function of the specific frequency of occurrence of each N-gram in this the analyzed domain name, (iii) a comparison of the determined entropy of the N-gram of the analyzed domain name with the specified threshold value of the entropy of the N-gram, (iv)
5 assignment of a numerical value to another feature of a domain name suspiciousness representing the entropy of the domain name under analysis, depending on the result of the comparison.

According to another embodiment of the present technique, when analyzing the obtained domain names, the analyzing module in the computing device may additionally perform at least
10 the following operations: (i) obtaining data on language dictionaries, (ii) extracting at least one word from each of the received domain names names, (iii) establishing the coincidence of each of the specified extracted words with one of the words in the language dictionaries of the indicated obtained data, (iv) assignment of a numerical value to another feature of a domain name suspiciousness, representing the analyzed meaningful domain name, depending on the
15 result of establishing a match.

According to another embodiment of the present technique, when analyzing the obtained domain names, the analyzing module in the computing device may additionally perform at least the following operations: (i) obtaining data on language dictionaries, (ii) extracting at least one word from each of the obtained domain names , (iii) determining the Levenshtein distance
20 between each of the specified extracted words and the corresponding one of the words in the language dictionaries of the specified obtained data, (iv) comparing determined Levenshtein distance with a predetermined threshold, (v) assigning a numerical value to another feature of a domain name suspiciousness, representing the spelling of the analyzed domain name, depending on the result of the comparison.

25 According to some embodiments of the present technique, an analyzing module in a computing device might be additionally configured to change a given set of analysis techniques.

According to other embodiments of the present technique, in response to assigning a domain name to malicious domain names, the processing module in the computing device may further issue a warning message, block network traffic from infected devices and/or create an
30 analytical report for the malicious domain name.

According to another embodiments of the present technique, the processing module in the computing device can be configured to analyze the obtained set of features of suspiciousness of the analyzed domain name using at least one pre-trained machine learning algorithm.

35 In other embodiments of the present technique, the analyzing module can be further configured to receive, using a communication module, at least one or a plurality of domain names from at least one source of domain names.

The above-mentioned technical problem is also solved in another aspect of the present technique, according to which a method for detecting malicious domain names in network traffic is proposed. The method, implemented by a computing device, comprises: (i) obtaining, with a communication module, network traffic from a data network; (ii) extracting, using a filtering module, a set of data packets from the received network traffic; (iii) analyzing, using a filtering module, the specified extracted data packets and extracting at least one domain name from them; (iv) analysis, using an analysis module, of each of the domain names obtained using a given set of analysis techniques, ensuring that a given numerical value is assigned to each of the given set of domain name suspicion characteristics corresponding to one of the given set of analysis techniques, for each domain name analyzed, depending on the results of its analysis; (v) analysis, using a processing module, of the obtained suspicious features with numerical values assigned to them, using a specified set of analysis rules, ensuring that each domain name is classified as a malicious domain name, if the obtained results of the features of suspiciousness analysis are characteristic of malicious domain names.

In one embodiment, the step of receiving network traffic may further comprise connecting, using a communication module, to at least one of the network traffic capturing devices included in the data network.

In yet another embodiment of the present technique, for the extracted domain names, it is further determined by the filtering module whether there is an analytical report for each of these extracted domain names.

In another embodiment of the present technique, the step of detecting the presence of an analytical report may include performing, by a filtering module, at least the following sub-steps: (i) obtaining data on the analytical reports; (ii) using character-by-character comparison of each of the extracted domain names with domain names that identify analytical reports in the indicated obtained data, establish that each extracted domain name matches one of the available analytical reports.

In some embodiments of the present technique, in response to the fact that there is no analytical report for the extracted domain name, according to the proposed method, the filtering module transfers the specified domain name to the analyzing module.

In other embodiments of the present technique, the step of analyzing the obtained domain names may include performing, using the analyzing module, at least the following sub-steps: (i) determining the number of characters in each of the obtained domain names; (ii) determining whether a certain number of characters correspond to each analyzed domain name given the threshold value of the length of the domain name; (iii) assigning a numerical value to a feature of a domain name suspiciousness representing the length of the domain name being analyzed, depending on the result of establishing compliance.

In other embodiments of the present technique, the step of analyzing the obtained domain names may include performing, using the analyzing module, at least the following sub-steps: (i) obtaining a set of known domain names; (ii) using character-by-character comparison of each of the extracted domain names with domain names in the obtained set of domain names, establishing the match of each of the extracted domain names with one of the known domain names; (iii) the assignment of a numerical value to another feature of a domain name suspiciousness, representing the fame of the analyzed domain name, depending on the result of establishing a match.

According to one embodiment of the present technique, the step of analyzing the obtained domain names may include performing, using the analyzing module, at least the following sub-steps: (i) determining the frequencies of occurrence of each of the many N-grams in each of the obtained domain names, each N-gram corresponds to a combination of N consecutive characters; (ii) determining the entropy of the N-gram of the specified analyzed domain name as a function of the specific frequency of occurrence of each N-gram in this analyzed domain name; (iii) comparing the determined entropy of the N-gram of the analyzed domain name with the specified threshold value of the entropy of the N-gram; (iv) assigning a numerical value to another feature of a domain name suspiciousness representing the entropy of the domain name under analysis, depending on the result of the comparison.

According to another embodiment of the present technique, the step of analyzing the obtained domain names may include performing, using an analyzing module, at least the following sub-steps: (i) obtaining data on language dictionaries; (ii) extracting at least one word from each of the obtained domain names; (iii) establishing the match of each of the specified extracted words with one of the words in the language dictionaries of the specified obtained data; (iv) assigning a numerical value to another feature of a domain name suspiciousness, representing the meaningfulness of the domain name being analyzed, depending on the result of establishing a match.

According to another embodiment of the present technique, the step of analyzing the obtained domain names may include performing, using the analyzing module, at least the following sub-steps: (i) obtaining data on language dictionaries; (ii) extracting at least one word from each of the obtained domain names; (iii) determining the Levenshtein distance between each of the specified extracted words and the corresponding one of the words in the language dictionaries of the specified obtained data; (iv) comparing a certain Levenshtein distance with a given threshold value; (v) assignment of a numerical value to another feature of a domain name suspiciousness, representing the spelling of the analyzed domain name, depending on the result of the comparison.

According to some embodiments of the present technique, according to the proposed method, the specified set of analysis rules is further modified by the analyzing module.

According to other embodiments of the present technique, in response to assigning a domain name to malicious domain names, also provided, using a processing module, a warning message, blocking network traffic from infected devices and/or creating an analytical report for the malicious domain name.

5 According to other embodiments of the present technique, the step of analyzing the features of suspiciousness includes analyzing, using the processing module, the obtained set of features of suspiciousness of the analyzed domain name using at least one previously trained machine-learning algorithm.

10 According to another embodiment of the present technique, according to the proposed method, in the analyzing module, a plurality of domain names from at least one source of domain names is additionally obtained, using a communication module.

The above technical problem is also solved in another aspect of the present technique, according to which a computer readable medium for long-term data storage storing computer readable instructions which, when executed by a computing device, cause the steps of the
15 method according to the second aspect of the present technique to be performed.

The above technical problem is also solved in another aspect of the present technique, according to which a method for analyzing domain names is implemented by a computing device, the method comprises: obtaining, using a communication module, at least one domain name, analysis, using an analyzing module, each of the obtained domain names using a given
20 set of analysis techniques to ensure that a given numerical value is assigned to each of a given set of features of a domain name suspiciousness corresponding to a given set of analysis techniques for each analyzed domain name, depending on the results of its analysis using these methods of analysis, the analysis, using the processing module, obtained features of suspiciousness with assigned numerical values using a given set of analysis rules to ensure
25 that each domain name is classified as malicious domain names, if the results of the analysis of features of suspiciousness are characteristic of malicious domain names.

The above-mentioned technical problem is also solved in another aspect of the present technique, according to which a computing device for domain name analysis is proposed, the computing device comprises: a communication module is configured to obtain at least one
30 domain name from at least one domain name source, analyzing module is configured to connect to the communication module to obtain the specified domain names from it and configured to analyze each of the received domain names using a given set of analysis techniques with the provision of assigning a given numerical value to each of the specified set of features of a domain name suspiciousness corresponding to one of the given set of analysis techniques for
35 each analyzed domain name, depending on the results of its analysis using the specified analysis techniques, the processing module is configured to connect to the analyzing module to obtain the specified feature of suspiciousness from it with the assigned numerical values for

each domain name and configured to analyse them using a given set of analysis techniques to ensure that each domain name is classified as malicious domain names, if the results of the analysis of feature of suspiciousness are characteristic of malicious domain names.

Brief description of the drawings

The accompanying drawings, which are incorporated herein and constitute a part of it, illustrate various embodiments and aspects of the present technique, and together with the following description, explain the essence of the present technique. The drawings show:

Fig. 1 shows a block diagram of a system for detecting malicious domain names in network traffic according to the present technique;

Fig. 2 shows a block diagram of a method of operating a computing device used in the system shown in fig. 1.

Detailed description

To ensure information security in computer systems and data networks, system 10 can be used to detect malicious domain names in network traffic, as shown in fig. 1, which allows detecting suspicious domain names in network traffic, which are used to establish a communication channel of an infected workstation or an infected computer running malicious software (bot) running on it with a managing computing device to receive control commands from it.

System

A system 10 for detecting malicious domain names in network traffic, shown as a simplified block diagram in fig. 1, comprises a computing device 1 in the form of, for example, a server and a traffic capture device 2 connected to a data transfer network 3, the traffic capture device 2 is configured to extract network traffic from the data transfer network 3. Computing device 1 is also connected with the ability to exchange data with traffic capturing devices 2 to receive network traffic from them, extracted from data transfer network 3, and configured to further process the received network traffic to determine the malicious domain names used to establish the infected a work station or an infected computer with a control command center and then receive control commands from it to execute a set of programmed malware waist actions, such as sending spam, phishing, anonymous access to the network, stealing confidential data, etc. depending on the objectives of the attackers, as described above in the section "Prior art".

In system 10 shown in fig. 1, the traffic capturing devices 2 are connected to the computing device 1 by a wired method, for example, using a network cable, and represent at least one of the following known network devices for intercepting and transmitting network

traffic: L2-level network switches operating using network traffic mirroring technology of necessary network segments, such as, for example, SPAN-technology for mirroring network traffic in CISCO equipment, network transparency tools, also called to the Security Delivery Platform or Network Packet Brokers (NPB), and Test Access Points of various types, as well as proxy servers with ICAP-Protocol support, working within the established TCP-connection, mail servers with SMTP-Protocol support, etc.

In one of the embodiments of the present technique, the computing device 1 can be connected directly to the data transfer network 3 or connected directly to the data transfer network 3 with the ability to extract or capture network traffic from it for further analysis and processing in order to detect malicious domain names in network traffic. In other words, in such an implementation, computing device 1 may have the functionality of traffic capturing devices 2 described above, while computing device 1 may, if necessary, use a secure data transmission/reception channel to receive network traffic extracted from data transfer network 3.

In another embodiment of the present technique, the above-described network traffic capturing devices 2 may be embedded or integrated into the computing device 1.

In some embodiments of the present technique, computing device 1 may be connected to traffic capturing devices 2 wirelessly.

Computing device

In system 10 shown in fig. 1, the computing device 1 is essentially a hardware-software complex and comprises a communication module 1.1, a filtering module 1.2, an analyzing module 1.3, a processing module 1.4 and a local data storage 1.5, each of which is connected to a communication bus 1.6, while each of the communication module 1.1, a filtering module 1.2, an analyzing module 1.3 and a processing module 1.4 is configured to exchange data, using a communication bus 1.6, with a local data storage 1.5, a communication module 1.1 is configured to exchange data with the filter module 1.2, which in turn is configured to exchange data, using the bus 1.6 connection, with the analyzing module 1.3, and the analyzing module 1.3 is configured to exchange data, using the bus 1.6 connection, with the processing module 1.4.

Local storage database

Local data storage 1.5 is designed to store executable software instructions that can control the operation of the communication module 1.1, the filtering module 1.2, the analyzing module 1.3 and the processing module 1.4. In addition, local data storage 1.5 is used in the computing device 1 for storing various data used in the operation of the computing device 1.

In particular, data on the available analytical reports are stored in the local data repository 1.5, each of which is assigned to a specific domain name and each of which contains data on the numerical values assigned by the analyzing module 1.3 to a specified set of features of suspiciousness characterizing the specified specific domain name as well as data related to the decision made by the processing module 1.4 in relation to the domain name being analyzed, in particular, data on the set of analysis rules used processing module 1.4 to analyze the specified set of features of suspiciousness, and data on the final decision regarding the harmfulness or power of attorney of the analyzed domain name, as will be described in more detail below.

The local data storage 1.5 also stores data about known domain names and data about language dictionaries. It should be noted that the well-known domain names in the data stored in the local data storage 1.5 can be used, for example, the 10,000 most popular sites from lists of popular sites hosted, for example, on rating sites such as alexa.com, webomer.com, trends.google.com, etc. As language dictionaries stored in the local data storage 1.5, various language dictionaries can be used, including dictionaries of abbreviations, abbreviations, slang words, various professional slang, etc., while dictionaries of the Russian language and other languages other than English language stored in Latin transliterated form.

In one embodiment of the present technique, the computing device 1, in particular its analyzing module 1.3, may periodically, after a specified period of time establish a connection with at least one of the above rating sites or access them using the module 1.6 connection with the extraction from the specified rating sites of at least one list of popular sites for updating data on known domain names stored in the local storage 1.5 data, on the basis of the specified updated list of popular sites; at that, a list of popular sites from one of these rating sites can be used for updating, as well as a summary list of popular sites compiled by the analyzing module 1.3 on the basis of separate lists of popular domain names obtained from each of these rating sites, by, for example, a comparative assessment of popularity indicators assigned to sites in these separate lists, in accordance with a given set of evaluation rules to identify a given number of the most popular sites.

In some embodiments of the present technique, at least some of the above data stored in a local storage 1.5 data can be stored in a cloud data storage (not shown), while these data can be located both on a single remote file server included in a network 3 data transmission or in a different data network than a network 3 data transmission, and on a set of remote file servers distributed in a network 3 data transmission or in a different data network 3 data transmission network, in this case, the computing device 1 can be configured to establish a connection for data exchange, in particular a secure communication channel for receiving/transmitting data, with such a cloud data storage using a wired and/or wireless connection method. In the embodiment, in which the computing device 1 connected to the data network 3 is connected to

the cloud data storage via at least one other data network other than the data network 3, it can be used to connect these different data networks at least one special switching device, such as network hubs, also known as network hubs, network routers, and other known switching devices.

5 In one of the embodiments of the present technique, the computing device 1 may comprise a separate local storage of analytical reports (not shown), designed to store data about existing analytical reports. Such a separate local repository of analytical reports can be connected via the communication bus 1.6 to the filtering module 1.2, allowing the filtering module 1.2 to access such local repository of analytical reports to retrieve data about the
10 available analytical reports from it, as well as connected via the communication bus 1.6 with the processing module 1.4, the processing module 1.4 is configured to get access to such a local repository of analytical reports for recording in it new analytical report, which is described in more detail below.

In another embodiment of the present technique, the data on the available analytical
15 reports can be stored in a separate remote analytical report storage (not shown), while the filtering module 1.2 can be configured to connect or access, using the communication module 1.1, with which the filtering module 1.2 is connected via the communication bus 1.6, to such a remote analytical report store to obtain data on the available analytical reports from it for the purpose of their subsequent use to identify domain names, previously analyzed with the help of
20 the computing device 1, and the processing module 1.4 can be configured to connect or access, using the communication module 1.1, with which the processing module 1.4 is connected using the communication bus 1.6, to such a remote storage of analytical reports for recording in it a new analytical report generated by the processing module 1.4 in relation to the analyzed domain name, which is described in more detail below.

25 In one of the embodiments of the present technique, the computing device 1 may comprise a separate local storage of domain names (not shown), designed to store data about known domain names. Such local storage of domain names can be connected via the communication bus 1.6 to the analyzing module 1.3, allowing the analyzing module 1.3 to access such local storage of domain names to retrieve data about known domain names from
30 it for later use to determine the numerical value appropriate feature of suspiciousness for the analyzed domain name, as described in more detail below.

In another embodiment of the present technique, there may be a separate remote domain name storage (not shown) for storing data about known domain names, and the analyzing module 1.3 can be configured to connect or access using the communication module
35 1.1, with which the analyzing module 1.3 is connected via a communication bus 1.6 to such a remote storage of domain names in order to obtain data about known domain names from it with a view to their subsequent use for determining the numerical value of the corresponding

feature of suspiciousness for the analyzed domain name, which is described in more detail below.

In one of the embodiments of the present technique, the computing device 1 may comprise a separate local storage of dictionaries (not shown) for storing data about existing language dictionaries. Such local storage of dictionaries can be connected via the communication bus 1.6 to the analyzing module 1.3 with the possibility for the analyzing module 1.3 to access such local storage of dictionaries to obtain data from it on existing language dictionaries for the purpose of their subsequent use for determining the numerical value of the corresponding attribute suspicion for the analyzed domain name, as described in more detail below.

In another embodiment of the present technique, there may be a separate remote dictionary storage (not shown) for storing data about existing language dictionaries, and the analyzing module 1.3 can be configured to connect or access using the communication module 1.1 with which the analyzing module 1.3 is connected via the communication bus 1.6 to such a remote storage of dictionaries to retrieve data from it about the existing language dictionaries for the purpose of their subsequent use for determining the numerical value of the corresponding feature of suspiciousness for the analyzed domain name, as described in more detail below.

The filtering module 1.2, the analyzing module 1.3, and the processing module 1.4 can be implemented as a single processor, such as a general-purpose processor or a special-purpose processor (for example, processors for digital signal processing, specialized integrated circuits, etc.). Thus, the processor implementing the filtering module 1.2, analyzing module 1.3, and processing module 1.4 in the computing device 1 can be configured to execute program instructions stored in the local storage 1.5 data with the implementation of the functionality of the filtering module 1.2 for extracting domain names from data packets in the received network traffic, the functionality of the analyzing module 1.3 to determine the numerical values of features of suspiciousness for the analyzed domain name and the functionality of the processing module 1.4 to make a final decision regarding the harmfulness of the analyzed domain name using the resulting set of feature of suspiciousness in the numerical values assigned to them.

In one of the embodiments of the present technique, each of the filtering module 1.2, the analyzing module 1.3, and the processing module 1.4 can be implemented as at least one separate processor. In this embodiment, the first processor used in the computing device 1 for implementing the filtering module 1.2 can be configured to execute software instructions stored in the local data storage 1.5 and ensure the functionality of the filtering module 1.2 to extract domain names from the data packets in the received network traffic. The second processor implementing the analyzing module 1.3 can be configured to execute software instructions stored in the local data storage 1.5 and ensuring the implementation of the functionality of the analyzing module 1.3 for determining the numerical values of feature of suspiciousness for the

analyzed domain name. The third processor used in the computing device 1 to implement the processing module 1.4 can be configured to execute software instructions stored in the local data storage 1.5 and ensuring the functionality of the processing module 1.4 to make a final decision regarding the harmfulness of the analyzed domain name using the obtained a set of features of suspiciousness with the numerical values assigned to them.

Local data storage 1.5 may be implemented, for example, in the form of one or more computer-readable media for long-term data storage. In some embodiments of the present technique, local data storage 1.5 may be implemented using a single physical device (for example, one optical storage device, magnetic storage device, organic storage device or another type of storage device, or storage on disks), and in other embodiments Implementing local data storage 1.5 can be implemented using two or more physical devices.

Communications module

In the computing device 1 of the system 10 shown in fig. 1, the communication module 1.1 is wired, for example, using a coaxial cable, twisted pair, fiber optic cable or other physical connection, to traffic capturing devices 2 with the possibility of receiving network traffic from them. Thus, the communication module 1.1 is configured to connect to one or more of the above-described traffic capture devices 2 included in the data network 3, and to receive network traffic extracted by said devices 2 to capture traffic from the data network 3 from this one or more traffic capturing devices 2.

In another embodiment of the present technique, the communication module 1.1 may be connected to the traffic capturing devices 2 wirelessly, for example using a communication line based on WiFi technology, a communication line based on 3G technology, a communication line based on LTE technology etc.

In an embodiment of the present technique in which the above-described cloud data storage is used to store various data used in the operation of the computing device 1, the communication module 1.1 in the computing device 1 is connected to this cloud data storage wirelessly, for example, using a communication line based on WiFi technology, communication lines based on 3G technology, communication lines based on LTE technology and/or communication lines based on other known wireless communication technology.

The communication module 1.1 in the computing device 1 according to the present technique can be implemented as a network adapter equipped with the necessary connectors for connecting the necessary types of physical cables to them depending on the types of physical connections used to communicate with the traffic capturing devices 2. In one embodiment of the present technique, the communication module 1.1 in the computing device 1 may be implemented as a network adapter in the form of a WiFi adapter, a 3G adapter, an LTE adapter, or another wireless communication adapter depending on the type of wireless link

used to provide communication any external devices and remote data storages that can establish a wireless channel of communication with the computing device to exchange necessary data with it.

Thus, the communication module 1.1 in the computing device 1 can essentially be configured to receive input data from one or more devices in a wired and/or wireless manner, as well as with the possibility of sending output data to other devices in a wired manner and/or wirelessly.

The communication module 1.1 may also be a known communication device, such as a transmitter, receiver, transceiver, modem and/or network interface card for exchanging data with external devices of any type via a wired or wireless communication network, for example using an Ethernet network connection, digital subscriber line (DSL), telephone line, coaxial cable, cellular telephone system, etc.

In some embodiments of the present technique, network traffic received by communication module 1.1 may be at least temporarily stored in local data storage 1.5. In other embodiments of the present technique, the traffic received by the communication module 1.1 may be at least temporarily stored in a separate network traffic network storage (not shown), different from the local data storage 1.5 and connected, via the communication bus 1.6, to the communication module 1.1. In other embodiments of the present technique, the network traffic received by the communication module 1.1 may be at least temporarily stored in a separate remote network traffic storage (not shown) connected to the data exchange capability with the communication module 1.1 via a wired and/or wireless method.

Filtering module

In the computing device 1 of the system 10 shown in fig. 1, the filtering module 1.2 is configured to connect or communicate, via the communication bus 1.6, to the communication module 1.1 so as to enable it to receive captured network traffic. The filtering module 1.2 extracts many individual data packets from the received network traffic using, for example, known technologies for extracting data packets based on open source software, such as Suricata, Bro, Kuko, etc. technologies, including using advanced technologies based on at least one of the above technologies. The filtering module 1.2 is also configured to analyze each of the plurality of extracted data packets, ensuring the extraction of at least one domain name from them.

In an embodiment of the present technique in which the received network traffic is stored in the local data storage 1.5, the filtering module 1.2 can be configured to access the local data storage 1.5 or to communicate with it using the bus 1.6 to retrieve the stored network traffic for the subsequent extraction of multiple data packets from it, as described above. In an embodiment of the present technique, in which the received network traffic is stored in a separate local storage of network traffic (not shown), the filtering module 1.2 can be configured

to access or connect to such local storage of network traffic using bus 1.6 while ensuring retrieval of saved network traffic from it for subsequent retrieval of multiple data packets from it, as described above. In an embodiment of the present technique, in which the received network traffic is stored in a separate remote network traffic repository (not shown), the filtering module 1.2 can be configured to access or connect to such remote network traffic storage using the communication module 1.1 connected to the filtering module 1.2 via the communication bus 1.6, ensuring that the saved network traffic is extracted from it for subsequent retrieval of multiple data packets from it, as described above.

In one embodiment of the present technique, the data packets extracted by the filtering module 1.2 from the received network traffic can be stored in the local data storage 1.5, while the filtering module 1.2 is configured to access the local data storage 1.5 or to establish a connection with it using bus 1.6 with the provision of extracting from it the stored data packets extracted from the received network traffic, for their subsequent analysis with the provision of extracting at least one domain name from them. In another embodiment of the present technique, the data packets extracted by the filtering module 1.2 from the received network traffic can be stored in a separate local data storage (not shown) other than the local data storage 1.5, while the filtering module 1.2 can be configured to receive access to such local data storage or to establish a connection with it using the communication bus 1.6, ensuring the retrieval from it of the stored data packets extracted from the received network traffic, for their subsequent analysis with the provision of extracting at least one domain name from them. In another embodiment of the present technique, the data packets extracted by the filtering module 1.2 from the received network traffic can be stored in a separate remote data storage (not shown), while the filtering module 1.2 can be configured to access such a remote data store or establish a connection with it using the communication module 1.1 connected to the filtering module 1.2 using the communication bus 1.6, with the provision of extracting from it the stored data packets extracted from the received network traffic for their further analysis with the provision of extraction of at least one domain name from them.

In one embodiment of the present technique, the domain names extracted by the filtering module 1.2 from a plurality of data packets extracted from the received network traffic can be stored in the local data storage 1.5, and the filtering module 1.2 can be configured to access the local data storage 1.5 or to establish a connection with it using the communication bus 1.6 ensuring the extraction of stored domain names from it for subsequent transfer to the analyzing module 1.3. In another embodiment of the present technique, the domain names extracted by the filtering module 1.2 from a plurality of data packets extracted from the received network traffic can be stored in a separate local data store (not shown) other than the local data storage 1.5, while the filtering module 1.2 may be configured to access such local data storage or to communicate with it using the communication bus 1.6 to ensure retrieval of stored domain

names from it for their subsequent transfer to the analyzing module 1.3. In another embodiment of the present technique, the domain names extracted by the filtering module 1.2 from a plurality of data packets extracted from the received network traffic can be stored in a separate remote data repository (not shown), while the filtering module 1.2 can be configured to access such remote data storage or to establish a connection with it using the communication module 1.1 connected to the filtering module 1.2 using the communication bus 1.6, ensuring the extraction of domain names stored in it for their subsequent transmission to the analyzing module 1.3.

Further, after extracting domain names from the data packets being analyzed, the filtering module 1.2 additionally determines whether there is an analytical report for each of these extracted domain names.

To detect the presence of an analytical report, the filtering module 1.2 performs at least the following operations: (i) gets access to the local data storage 1.5 or communicates with it using the communication bus 1.6 ensuring that it receives data about the available analytical reports, each of which is supplied in accordance with the specific domain name for which this analytical report was created, (ii) establishes, by character-by-character comparison of each of the extracted domain names with domain names identifying analytical reports in the specified data, the correspondence of each of the extracted domain names to one of the available analytical reports. Thus, the filtering module 1.2 compares each extracted domain name with domain names that identify behavioral reports in the obtained analytical reports, and the coincidence of the extracted domain name with the domain name that identifies one of the available analytical reports means that for the extracted domain name there is already an analytical report stored in the local data storage 1.5, which indicates that the extracted domain name has already been analyzed in computing ohm device 1, in particular in its processing module 1.4.

In an embodiment in which data on analytical reports are stored in a separate local storage of analytical reports (not shown), the filtering module 1.2 gets access to such local repository of analytical reports or communicates with it using the communication bus 1.6 to ensure that data on the available analytical reports is received from it, each of which reports is assigned to a specific domain name for which this analytical report was created to further establish the compliance of the extracted domain name with one of the available analytical reports. In an embodiment in which data on analytical reports are stored in a separate remote storage of analytical reports (not shown), the filtering module 1.2 gets access to such remote repository of analytical reports or communicates with it using the communication module 1.1 connected to the filtering module 1.2 using the communication bus 1.6, ensuring that data about the available analytical reports are obtained from it, each of which is aligned with a specific domain name for which this analysis report was created, in order to further establish the compliance of the extracted domain name with one of the available analysis reports.

Further, the filtering module 1.2 transfers to the analyzing module 1.3 only those extracted domain names for which there is no analytical report in the local data storage 1.5, that is, only those domain names that have not been previously analyzed in computing device 1, in particular in its processing module 1.4.

Thus, according to the above description of the functionality of the filtering module 1.2, the filtering module 1.2 provides the ability to filter at least part of the domain names extracted from the data packets of the received network traffic using the minimum computing resources of the computing device 1, which ultimately significantly speeds up the process identify malicious domain names in network traffic.

Analyzing module

In the computing device 1 of the system 10 shown in fig. 1, the analyzing module 1.3 is configured to connect to the filtering module 1.2 or to establish communication with it using the communication bus 1.6, ensuring that it can obtain domain names from it, for which the filtering module 1.2 revealed the absence of an analytical report, as described above.

In an embodiment of the present technique in which the domain names extracted by the filtering module 1.2 are stored in the local data storage 1.5, the analyzing module 1.3 accesses or communicates with the local data storage 1.5 using the communication bus 1.6 to ensure that data is received from it the extracted domain names to be further analyzed in this analyzing module 1.3, as will be described below.

In another embodiment of the present technique, in which the domain names extracted by the filtering module 1.2 are stored in a separate local data storage, the analyzing module 1.3 gets access to such local data storage or communicates with it using the communication bus 1.6 so that it can receive from it information about extracted domain names to be further analyzed in this analyzing module 1.3, as will be described below.

In another embodiment of the present technique, in which the domain names extracted by the filtering module 1.2 are stored in a separate remote data storage, the analyzing module 1.3 accesses or communicates with such local data storage using the communication module 1.1 connected to the analysis module 1.3 using the communication bus 1.6, ensuring that data about the extracted domain names are retrieved from it, which are subject to further analysis in this analyzing module 1.3, as will be described below.

Analyzing module 1.3 analyzes each of the obtained domain names using a given set of methods or methods of analyzing suspiciousness, ensuring that a given numerical value is assigned to each of a given set of feature of a domain name suspiciousness corresponding to one of a given set of suspiciousness analysis techniques, depending on the results of its analysis using the specified methods of analysis.

In one embodiment of the present technique, the analyzing module 1.3 can be equipped with a special framework (software platform) that allows, for example, the operator or the user of a computing device 1 to change the specified set of techniques for analyzing suspiciousness used by the analyzing module 1.3 to assign a given numerical value to each of a given set of features of a domain name suspiciousness corresponding to one of a given set of methods of analysis, for example, to set or add new methods for analyzing suspiciousness, allowing to determine the numerical values of the new feature of suspiciousness, which will be used in the future in the work of the processing module 1.4, which provides an extension of the functionality of both the analyzing module 1.3 separately and the computing device 1 as a whole.

One of the suspiciousness analysis techniques allows, using the analyzing module 1.3, to determine the numerical value of such a feature of suspiciousness of the analyzed domain name as the length of the domain name. When analyzing the obtained domain names using this method of suspicion analysis, the analyzing module 1.3 performs at least the following operations: (i) determines the number of characters in each of the received domain names, (ii) determines whether a certain number of characters of each analyzed domain name corresponds to a specified threshold value of the domain name length, which can be set manually by the expert or by the operator/user of the computing device 1, as well as automatically using the specified rules for determining the threshold value for the length of the domain name, (iii) assigns a specified numerical value to the feature of a domain name suspiciousness, which represents the length of the analyzed domain name, depending on the result of establishing compliance.

In one embodiment of the technique, if a domain name suspiciousness analysis method is used based on its length, the threshold value of the domain name length can be set automatically, for example as follows: the analyzing module 1.3 can get data about known domain names that can be stored in local data storage 1.5, a detached local domain name storage, or a separate remote domain name storage, and then calculate, for example, the average length of the known domain name, followed by using this calculated value as a threshold value of length of a domain name. In another embodiment of the present technique, if a domain name suspiciousness analysis method is used based on its length, a given set of rules can be used to process the length values of known domain names in the data obtained by the analyzing module 1.3 to calculate the threshold length of the domain name based on them that is used in the work of the analyzing module 1.3.

Another suspiciousness analysis technique allows, using the analyzing module 1.3, to determine the numerical value of such a feature of suspiciousness of the analyzed domain name as the popularity of the analyzed domain name. When analyzing the obtained domain names using this method of suspicion analysis, the analyzing module 1.3 performs at least the following operations: (i) establishes a connection with the local data store 1.5 (a separate local or remote

domain name store depending on the implementation option, as described above) to obtain from it a set of known domain names, (ii) establishes, by symbolizing comparison of each of the extracted domain names with the domain names in the resulting set of domain names, the coincidence of each of the extracted domain names with one of the known domain names and
5 (iii) assigns a given numerical value to another feature of a domain name suspiciousness, which represents the popularity of the analyzed domain name, depending on the result of establishing a match.

Another suspiciousness analysis technique used by the analyzing module 1.3 in its work allows determining the numerical value of such a feature of suspiciousness of the analyzed
10 domain name as the entropy of the analyzed domain name. When analyzing the obtained domain names using this method of suspicion analysis, the analyzing module 1.3 performs at least the following operations: (i) determines the frequency of occurrence of each of the set of N-grams in each of the obtained domain names, each n-gram corresponds to a combination of N consecutive characters, (ii) determines the entropy of the N-gram of the specified analyzed
15 domain name as a function of a certain frequency of occurrence of each N-gram in this analyzed domain name, (iii) compares a certain entropy of the N-gram of the domain name under analysis with a given threshold of the N-gram entropy, and (iv) assigns a given numerical value to another feature of a domain name suspiciousness representing the entropy of the analyzed domain name, depending on the result of the comparison. Thus, this method of analyzing
20 suspiciousness allows to obtain, for example, the numerical values of three feature of suspiciousness of the analyzed domain name, representing each entropy of the analyzed domain name, respectively, for example, for 2-gram, 3-gram and 4-gram (the most appropriate lengths of combinations from consecutive characters).

In one embodiment of the present technique, the number of consecutive characters in a
25 combination corresponding to an N-gram, when determining entropy in a suspiciousness analysis technique related to determining the entropy of an analyzed domain name, can be set manually by an expert or user of computing device 1 or can be set automatically, for example, depending on the above described threshold for the length of a domain name or the average length of known domain names, which can be determined the analyzing module 1.3 using, for
30 example, data on known domain names from the local data storage 1.5, as described in more detail above, or in accordance with any other specified rule.

Another suspiciousness analysis technique used by the analyzing module 1.3 in its work allows determining the numerical value of such a feature of suspiciousness of the analyzed domain name as the meaningfulness of the analyzed domain name. When analyzing the
35 resulting domain names using this suspiciousness analysis technique, the analysis module 1.3 performs at least the following operations: (i) establishes a link with the local data store 1.5 (a separate local or remote domain name store depending on the implementation option, as

described above) to obtain data on the language dictionaries from it, (ii) extracts at least one word from each, (iii) establishes the coincidence of each of the specified extracted words with one of the words in the language dictionaries from the specified obtained data, (iii) assigns a given numerical value to another feature of a domain name suspiciousness, representing the
5 meaningfulness of the analyzed domain name, depending on the result of establishing a match.

In one embodiment of the present technique, in the case of using a method of analyzing the suspiciousness of a domain name on the basis of its meaningfulness, when determining the coincidence of a word extracted from the analyzed domain name with one of the words in the resulting language dictionaries, the analyzing module 1.3 additionally checks whether the
10 extracted word coincides with one of the words in the resulting language dictionaries when writing this extracted word in reverse order and/or whether the meaningful part of this extracted word coincides (for example, common abbreviation of such word) with one of the words in the resulting language dictionaries, while further analyzing module 1.3 also assigns a given numerical value to another feature of a domain name suspiciousness, representing the
15 meaningfulness of the analyzed domain name, depending on the results of such additional verification.

In another embodiment of the present technique, if a domain name suspiciousness analysis technique is used based on its meaningfulness, after extracting at least one word from each of the obtained domain names, the analyzing module 1.3 can additionally identify the
20 language of the specified extracted word, and if it's established that this extracted word is written in a language other than English (for example, in Russian) to transliterate it in Latin in accordance with the set of the transliteration rules. In another embodiment of the present technique, in the case of using a technique for analyzing the suspiciousness of a domain name on the basis of its meaningfulness, the analyzing module 1.3 can be additionally configured to
25 establish a connection, using the communication module 1.1, with a source of language dictionaries to obtain from it data on a new dictionary, followed by updating the existing language dictionaries stored in the local data storage 1.5, on the basis of this new dictionary from the data obtained, at the same time, the analyzing module 1.3 can also identify the language of words of this new word, and if the words in this new dictionary are written in a language other than
30 English (for example, in Russian), it can transliterate all the words from this new dictionary in Latin in accordance with a given set of transliteration rules.

Another suspiciousness analysis technique used by the analyzing module 1.3 in its work allows determining the numerical value of such a feature of suspiciousness of the analyzed domain name as the correctness of the spelling of the analyzed domain name. When analyzing
35 the resulting domain names using this suspect analysis technique, the analysis module 1.3 performs at least the following operations: (i) establishes a link with the local data storage 1.5 (a separate local or remote domain name store depending on the implementation option, as

described above) to obtain data on the language dictionaries from it, (ii) extracts at least one word from each, (iii) determine the Levenshtein distance between each of the specified extracted words and the corresponding word in the language dictionaries from the specified received data, (iv) compare a certain Levenshtein distance with a given threshold value, which, for example, can be used constant equal to two (2), (v) assigns a given numerical value to another feature of a domain name suspiciousness, representing the correctness of the spelling of the analyzed domain name, depending on the result of the comparison (in particular, depending on the types of errors when spelling the analyzed domain name).

As a result of applying the above-described suspiciousness analysis techniques to each domain name analyzed in analyzing module 1.3, the analyzing module 1.3 creates or prepares a separate log file with analysis results of a specific domain name in the analyzing module 1.3, which contains at least numerical values assigned analyzing module 1.3 to the above described features of suspiciousness, characterizing the specified specific domain name and corresponding to each one of the applied techniques for the analysis of suspiciousness.

In one embodiment of the present technique, the communication module 1.1 may be further configured to establish communication with at least one local or remote source of domain names, or with the ability to connect to such a source of domain names to ensure that it receives at least one domain name or sets of domain names for their subsequent transfer to the analyzing module for the analysis of each of these obtained domain names using a given set of analysis techniques, as it described in detail above. Such sources of domain names can include at least one of the following local or remote sources: a remote computing device, server, host, domain, database, antivirus system or program, website, etc.

Processing module

In the computing device 1 of the system 10 shown in fig. 1, the processing module 1.4 is configured to connect to the analyzing module 1.3 or to establish communication with it using the communication bus 1.6, ensuring that it can receive log files prepared or created by the analyzing module 1.3, as described above.

The processing module 1.4 analyzes the features of a domain name suspiciousness with assigned numerical values contained in each of the received log files using a specified set of analysis rules ensuring that each domain name for which the corresponding log file was created is classified as malicious domain names, if the results of the analysis of the features of suspiciousness are characteristic of malicious domain names, or of trusted domain names, if the results of the analysis of the features of suspiciousness are characteristic for trusted domain names. Thus, the processing module 1.4 makes the final decision regarding the harmfulness or harmlessness of the domain name using the log file created by the analyzing module 1.3 for each extracted domain name.

In particular, the processing module 1.4 is based on a previously trained classifier, which makes its decision regarding the harmfulness or harmlessness of a specific domain name using a set of feature of this domain name suspiciousness with its assigned numerical values, which the processing module 1.4 extracts using known algorithms or scripts from the log file obtained from the analyzing module 1.3, each of the used set of feature of suspiciousness essentially reflects one of the characteristics of a particular domain name that the analyzing module 1.3 analyzes, and is indicative of the potential malignancy of the specified specific domain name, and the numerical value assigned to this domain name reflects the degree of its suspiciousness. The pre-trained classifier used in the processing module 1.4 is a classifier based on one of the known machine learning algorithms (MLA), for example, based on the machine learning algorithm "Random Forest", based on the machine learning algorithm "gBoost", based on the machine learning algorithm "AdaBoost", etc. The classifier on the basis of one of the MLA algorithms makes its decision on the harmfulness or harmlessness of each domain name based on the extracted weighted features of suspiciousness obtained as a result of assigning a given weight to each of the set of features of a domain name suspiciousness extracted from the resulting log file, and other static (not having a given weight) features using an algorithm that implements, for example, the most effective decision tree selected at the algorithm training stage by experts using known malicious domain names.

In one of the embodiment options, the processing module 1.4 can be based on a set of pre-trained classifiers, each of which makes its decision regarding the harmfulness or harmlessness of a specific domain name using the same set of feature of this domain name suspiciousness and their numerical values, while, for example, the above classifiers can be used as such classifiers. In one form of this embodiment of the present technique, the final decision regarding the safety of a domain name can be made by the processing module 1.4, for example, in the case in which at least one of the specified set of classifiers or at least one of the specified set of classifiers has passed against the safety of the domain name, otherwise, in which none of the specified set of classifiers produced a personal decision regarding the safety of the domain name processing module 1.4 may reach a final decision regarding the harmfulness of the domain name. In another variation of this embodiment, the processing module 1.4 can render its final decision regarding the harmfulness or harmlessness of the domain name by comparing the total weight of personal decisions received by a given set of classifiers in the processing module 1.4, with a given threshold value, which is essentially constant, while any personal decision rendered by each classifier from a given set of classifiers is automatically assigned its own prior given weighting. Thus, the processing module 1.4 accepts one or another final decision, taking into account all personal decisions (intermediate decisions) of the used classifiers that have one or another weight in the total weight taken into account by the processing module 1.4 for rendering the final decision.

After the final decision, the processing module 1.4 prepares or creates a separate analytical report for each domain name analyzed in this processing module 1.4, and this analytical report at least contains all the information from the corresponding log file created by the analyzing module 1.3 for the specified domain name, as well as data on the final decision rendered by the processing module 1.4 with respect to the domain name, as well as data on the classifier used by the processing module 1.4 for the final decision. In an embodiment of the present technique, in which several classifiers are used in the operation of the processing module 1.4, an analytical report prepared or created by the processing module 1.4 may also contain data on each of the classifiers used by the processing module 1.4 to render the final decision, and personal decisions in relation to the harmfulness or harmlessness of the domain name taken out by each of these classifiers separately.

In some embodiments of the present technique, the processing module 1.4 can be equipped with a special framework (software platform), allowing, for example, the operator of the computing device 1 or experts to make changes to a given set of classifiers used by the processing module 1.4 to make a final decision on the harmfulness or harmlessness domain name as described in more detail above.

In addition, in the case when the processing module 1.4 makes a final decision about the harmfulness of a domain name, that is, assigns a specific domain name based on its numerical values assigned in the analyzing module 1.3 to its features of suspiciousness, the processing module 1.4 can also perform at least one of the following actions: issue a warning message that a malicious domain name is detected in network traffic and block network traffic from an infected device or infected devices, which can be identified by their IP-address as defined by the revealed a malicious domain name.

To save a prepared or created analytical report in the local data storage 1.5, the processing module 1.4 accesses or communicates with the local data storage 1.5 using the communication bus 1.6 to ensure the transfer of the specified analytical report to the local data storage 1.5 to save it in this local data storage 1.5.

In an embodiment of the present technique, in which a prepared or created analytical report is stored in a separate local storage of analytical reports (not shown), the processing module 1.4 accesses or communicates with such a local storage of analytical reports with it using the communication bus 1.6 analytical report in the local repository of analytical reports for its storage in this local repository of analytical reports.

In an embodiment of the present technique, in which the prepared or created analytical report is stored in a separate remote repository of analytical reports (not shown), the processing module 1.4 accesses or establishes a connection to such remote repository of analytical reports with the use of the communication module 1.6, which is connected with the processing module 1.4 using the communication bus 1.6, ensuring the transfer of the specified analytical report to

the remote analytical report storage by wired and/or wireless one way to save it in this remote repository of analytical reports.

It should be noted that communication module 1.1, filtering module 1.2, analyzing module 1.3, and processing module 1.4 implement their above functions essentially in real time, that is, communication module 1.1 continues to receive network traffic, filtering module 1.2 performs its filtering operations at least part of the domain names extracted from the data packets in the received network traffic, the analyzing module 1.3 performs its operations to analyze domain names and assign the specified numerical values to the feature of suspiciousness, followed by the creation of a log file for each domain name separately, and the processing module 1.4 performs its operations to analyze at least part of the log files obtained from the analyzing module 1.3, and save at least part of the created or prepared analytical reports in the local data storage 1.5.

In one of the embodiments of the present technique, the analyzing module 1.3 and the processing module 1.4 can be combined, for example, into a single module for detecting malicious domain names, having all the above-described functionalities inherent in the analyzing module 1.3 and the processing module 1.4. In another embodiment of the present technique, the filtering module 1.2 and the analyzing module 1.3 can be combined, for example, into a single preprocessing module having all the above-described functionalities inherent in the filtering module 1.2 and the analyzing module 1.3. In other embodiments of the present technique, the filtering module 1.2, the analyzing module 1.3 and the processing module 1.4 can be combined, for example, into a single module for identifying malicious domain names that has all the above-described functionality inherent in the filtering module 1.2, the analyzing module 1.3 and the processing module 1.4.

In another embodiment of the present technique, communication module 1.1 may be divided into several separate communication modules, each of which provides at least one of the known methods of wired and/or wireless communication in computing device 1.

In some embodiments of the present technique, the analyzing module 1.3 can be divided into several independent modules, each of which can perform at least one of the above-described functionalities inherent in the analyzing module 1.3, and which are adapted to communicate with each other and the rest of device 1 via communication bus 1.6. In this case, one of such independent modules can be configured to create a log file, as described above, or there may be, for example, an additional module for creating log files, configured to communicate with these independent modules using, for example, bus 1.6 ensuring that each of them receives, for example, data on a specific feature of suspiciousness with a numerical value assigned to it for the same domain name.

In other embodiments of the present technique, the processing module 1.4 can be divided into several independent modules, each of which can, for example, implement one of

the above-described pre-trained classifiers that underlie the processing module 1.4, while these independent modules can be configured to communicate with each other and with the rest of the design modules of the computing device 1, in particular with the analyzing module 1.3, using the communication bus 1.6. In this case, one of such independent modules can be configured to create an analytical report, as described above, or there can be, for example, an additional module for creating analytical reports, configured to communicate with these independent modules using, for example, communication bus 1.6 with provision obtaining data from them about the results of the analysis of the received log file using the specified set of analysis rules for creating an analytical report for a specific domain name.

In another embodiment of the present technique, the filtering module 1.2 can be divided into several other independent modules, each of which can perform at least one of the above-described functionalities inherent in the filtering module 1.2, and which are configured to communicate with each other and the other structural modules computing device 1, in particular with the communication module 1.1 and the analyzing module 1.3, using the communication bus 1.6.

In other embodiments of the present technique, each classifier from a predetermined set of classifiers used in the processing module 1.4 can be implemented as a separate classifying module, issuing its personal decision regarding the harmfulness or harmlessness of a particular domain name, with the specified individual classifying modules can be configured to communicate with the processing module 1.4 using, for example, the communication bus 1.6, and the processing module 1.4 in this case can make a final decision on the harmfulness or harmlessness of a domain name using a given set of rules for making a final decision and personal decisions made by the classifying modules.

According to one embodiment of the present technique, at least some of the above-described functionality inherent in the filtering module 1.2, the analyzing module 1.3 and/or the processing module 1.4 can be implemented as a separate functional submodule or functional block included in the corresponding one of the modules 1.2, 1.3 and 1.4. Thus, the filtering module 1.2 can contain several of its submodules, each of which implements at least one of the above-described functional capabilities inherent in the filtering module 1.2, the analyzing module 1.3 can contain several of its submodules, each of which implements at least one of the above-described functional capabilities inherent in the analyzing module 1.3, and the processing module 1.4 may contain several of its submodules, each of which implements at least one of the above functional capabilities inherent in the processing unit 1.4.

Thus, the above-described functionality of the filtering module 1.2 provides the ability to effectively filter at least part of domain names extracted from network traffic data packets, resulting in the subsequent analysis of domain names in the analyzing module 1.3 and the processing module requiring significantly less computational resources allocated computing

device 1. In addition, the above-described functionality of the filtering analyzing module 1.3 makes it possible to efficiently allocate a set of features of suspiciousness with assigned numerical values for each domain name, and the above-described functionality of the processing module 1.4 provides the ability to make an accurate and reliable decision regarding the harmful or harmless domain name.

According to another aspect, a method for detecting malicious domain names in network traffic is proposed, shown as a simplified flowchart in fig. 2.

The method shown in fig. 2 begins in block 2, according to which the presence of a computing device 1 for detecting malicious domain names in network traffic is provided, the structural units and functionality of which are described above. In block 2.1 of the method, network traffic from data network 3 is received via communication module 1.1. In block 2.2 of the method, a plurality of data packets are extracted from the received network traffic using a filtering module 1.2. In block 2.3 of the method, the extracted data packets are analyzed using the filtering module 1.2, ensuring the extraction of at least one domain name from the specified set of data packets. In block 2.4, each of the obtained domain names is analyzed using an analyzing module 1.3 using a given set of analysis techniques, ensuring that a given numerical value is assigned to each of a given set of features of a domain name suspiciousness corresponding to one of a given set of analysis methods for each analyzed domain name in depending on the results of its analysis using the specified methods of analysis. In block 2.5 of the method, the obtained features of a domain name suspiciousness with numerical values assigned to them are analyzed using the processing module 1.4 using a specified set of analysis rules. In block 2.6 of the method, it is established, using the processing module 1.4, whether the analyzed features of suspiciousness are characteristic with the assigned numerical values for malicious domain names. If in block 2.6 of the method it was determined that the analyzed features of suspiciousness with the assigned numerical values are characteristic of malicious domain names (the “Yes” line in fig. 2), the method proceeds to block 2.7 of the method according to which the domain name is assigned to malicious domain names using the processing module 1.4, followed by the completion of the method in block 2.9 of the method. Otherwise, the method proceeds to block 2.8 (the “No” line in fig. 2), according to which, using the processing module 1.4, the domain name is assigned to trusted domain names, followed by completion of the method in block 2.9 of the method.

The operations in block 2.1 of the method include connecting, using the communication module 1.1, to at least one of the network traffic capturing devices included in the data network.

For the domain names extracted in block 2.3 of the method, it is further determined using the filtering module 1.2 whether there is an analytical report for each of these extracted domain names. At the same time, the above operation of detecting the availability of an analytical report includes at least the following sub-steps: (i) obtaining data on analytical reports, (ii) establishing

compliance of each of the extracted domain names to one of the available analytical reports using character-by-character comparison of each of the extracted domain names with domain names identifying the analytical reports in the indicated data.

In response to the fact that there is no analytical report for the extracted domain name, according to the proposed method, the filtering module additionally transfers the specified domain name to the analyzing module.

According to the proposed method, the operations in block 2.4 include the execution, using the analyzing module 1.3, of at least the following sub-steps: (i) determining the number of characters in each of the obtained domain names; (ii) determining whether a certain number of characters correspond to each analyzed domain name given the threshold value of the length of the domain name; (iii) the assignment of a given numerical value to a feature of a domain name suspiciousness, representing the length of the analyzed domain name, depending on the result of establishing compliance.

According to the proposed method, the operations in block 2.4 also include performing, using the analyzing module 1.3, at least the following sub-steps: (i) obtaining a set of known domain names; (ii) establishing the coincidence of each of the extracted domain names with one of the known domain names using character-by-character comparison of each of the extracted domain names with domain names in the obtained set of domain names; (iii) the assignment of a given numerical value to another feature of a domain name suspiciousness, representing the fame of the analyzed domain name, depending on the result of establishing a match.

According to the proposed method, the operations in block 2.4 also include the execution, using the analyzing module 1.3, of at least the following sub-steps: of N consecutive characters; (ii) determining the entropy of the N-gram of the specified analyzed domain name as a function of the specific frequency of occurrence of each N-gram in this analyzed domain name; (iii) comparing the determined entropy of the N-gram of the analyzed domain name with the specified threshold value of the entropy of the N-gram; (iv) the assignment of a given numerical value to another feature of a domain name suspiciousness, representing the entropy of the domain name under analysis, depending on the result of the comparison.

According to the proposed method, the operations in block 2.4 also include the execution, using the analyzing module 1.3, of at least the following sub-steps: (i) obtaining data about language dictionaries; extracting at least one word from each of the obtained domain names; (ii) establishing the coincidence of each of the specified extracted words with one of the words in the language dictionaries of the specified obtained data; (iii) the assignment of a given numerical value to another feature of a domain name suspiciousness, representing the meaningfulness of the analyzed domain name, depending on the result of establishing a match.

According to the proposed method, the operations in block 2.4 also include the execution, using the analyzing module 1.3, of at least the following sub-steps: (i) obtaining data about language dictionaries; (ii) extracting at least one word from each of the obtained domain names; (iii) determining the Levenshtein distance between each of the specified extracted words and the corresponding one of the words in the language dictionaries of the specified obtained data; (iv) comparing a certain Levenshtein distance with a given threshold value; (v) the assignment of a given numerical value to another feature of a domain name suspiciousness, representing the spelling of the analyzed domain name, depending on the result of the comparison.

According to the proposed method, the specified set of analysis rules is additionally modified using an analyzing module.

In response to assigning a domain name to malicious domain names, according to the proposed method, they also provide, using a processing module 1.4, a warning message, blocking network traffic from infected devices and/or creating an analytical report for the malicious domain name.

The operations in block 2.5 of the method include the analysis, using the processing module 1.4, of the obtained set of features of suspiciousness of the analyzed domain name using at least one previously trained machine learning algorithm.

In one embodiment, the implementation of the present technique according to the proposed method in the analyzing module 1.3 can additionally receive, using the communication module 1.1, a plurality of domain names from at least one source of domain names, such as a remote device, server, host, domain, database, anti-virus system or program, website, etc.

According to another aspect of the present technique, a computer-readable medium for long-term storage of data is proposed that stores computer-readable instructions that, when executed by a processor by a computing device, cause the steps of the method described in this document to be performed. Machine-readable instructions may comprise a machine-readable program code that can be transmitted using any suitable medium, including wireless means, wired facilities, fiber optic cable, radio frequency, and/or the like, and/or any the right combination. A computer-readable program code can be written in one of the programming languages or any combination of programming languages containing an object-oriented programming language, such as "Java", "Smalltalk", "C++", and/or the like, and conventional procedural programming languages such as the programming language "C". The computer-readable program code can be fully or partially executed on the computing device 1.

Thus, machine-readable software instructions stored on a machine-readable medium can control the computing device 1 in such a way that it will function as described above, so that computer-readable instructions stored in a computer-readable medium create an industrial

product containing software instructions that implement the functions/the actions indicated in the blocks of the flowchart of fig. 2, illustrating the operation of the computing device 1.

As a machine-readable media for long-term storage of data can be used one of the following material machine-readable media designed to store data for a long period of time: hard
5 disk drives, read-only memory (ROM), compact disk drives (CD), drives on digital versatile disks (DVDs), floppy drives, Blu-ray drives, etc.

CLAIMS

1. A computing device configured to detect malicious domain names in network traffic in real time, the computing device comprises:
 - 5 a communication module configured to receive network traffic from a data network,
 - a filtering module, configured to connect to the communication module to receive the network traffic from the communication module and to perform at least the following:
 - extracting a plurality of data packets from the received network traffic,
 - analyzing said extracted data packets, ensuring the extraction of at least one
 - 10 domain name from said plurality of data packets,
 - an analyzing module, configured to connect to the filtering module to receive the specified domain names from the filtering module and to analyze each of the obtained domain names using a given set of analysis methods, ensuring that a given numerical value is assigned to each of a given set of features of a domain name suspiciousness corresponding to one of the
 - 15 given set of analysis methods, for each analyzed domain name, depending on the results of its analysis using the specified analysis methods,
 - a processing module, configured to connect to the analyzing module to receive from the analyzing module the indicated features of suspiciousness with the assigned numerical values for each domain name and with the possibility of analyzing them using a specified set of analysis
 - 20 rules ensuring that each domain name is classified as malicious domain names, if the received results of the analysis of features of suspiciousness are characteristic of malicious domain names.
2. The computing device of claim 1, wherein the communication module is further
- 25 configured to connect to at least one of the network traffic capture devices connected to the data network.
3. The computing device of claim 1, wherein the filtering module is further configured to determine if there is an analytical report for each of the extracted domain names.
- 30 4. The computing device of claim 3, wherein, in response to a presence of the analytical report, the filtering module is configured to perform at least the following operations:
 - receiving data on analytical reports,
 - establishing the correspondence of each of the extracted domain names to one of the
 - 35 available analytical reports using character-by-character comparison of each of the extracted domain names with domain names that identify analytical reports in the indicated obtained data.

5. The computing device of claim 3 or 4, wherein, in response to an absence of the analytical report for the extracted domain name, the filtering module is additionally configured to transfer the specified domain name to the analysing module.
- 5 6. The computing device of claim 1, wherein, when analyzing the obtained domain names, the analyzing module is configured to perform at least the following operations:
determining the number of characters in each of the received domain names,
determining whether a certain number of characters correspond to each analyzed domain name given the threshold value of the length of the domain name,
10 assignment of a given numerical value to a suspiciousness attribute of a domain name, representing the length of the domain name being analyzed, depending on the result of establishing compliance.
7. The computing device of claim 6, wherein, when analyzing the obtained domain names,
15 the analyzing module is additionally configured to perform at least the following operations:
getting a set of known domain names,
establishing the coincidence of each of the extracted domain names with one of the known domain names using character-by-character comparison of each of the extracted domain names with domain names in the obtained set of domain names,
20 assignment of a given numerical value to another feature of a domain name suspiciousness, representing popularity of the analyzed domain name, depending on the result of establishing a match.
8. The computing device of claim 7, wherein, when analyzing the obtained domain names,
25 the analyzing module is additionally configured to perform at least the following operations:
determining the frequency of occurrence of each of the many N-grams in each of the obtained domain names, each N-gram corresponding to a combination of N consecutive characters,
determining the entropy of the N-gram of the specified analyzed domain name as a
30 function of the specific frequency of occurrence of each N-gram in this analyzed domain name,
comparison of the determined entropy of the N-gram of the analyzed domain name with the specified threshold value of the entropy of the N-gram,
assignment of a given numerical value to another feature of a domain name suspiciousness, representing entropy of the analyzed domain name, depending on the result of
35 the comparison.

9. The computing device of claim 8, wherein, when analyzing the obtained domain names, the analyzing module is additionally configured to perform at least the following operations:

getting data about language dictionaries,

extracting at least one word from each of the received domain names,

5 matching each of the specified extracted words with one of the words in the language dictionaries from the specified received data,

assignment of a given numerical value to another feature of a domain name suspiciousness, representing meaningfulness of the analyzed domain name, depending on the result of establishing a match.

10. The computing device of claim 9, wherein, when analyzing the obtained domain names, the analyzing module is additionally configured to perform at least the following operations:

getting data about language dictionaries,

extracting at least one word from each of the received domain names,

15 determining the Levenshtein distance between each of the specified extracted words and the corresponding one of the words in the language dictionaries of the specified obtained data,

comparison of a certain Levenshtein distance with a predetermined threshold value

20 assignment of a given numerical value to another feature of a domain name suspiciousness, representing correctness of spelling of the analyzed domain name, depending on the result of the comparison.

11. The computing device of claim 1, wherein the analyzing module is further configured to change the given set of analysis methods.

12. The computing device of claim 1, wherein, in response to assigning a domain name to malicious domain names, the processing module is further configured to issue a warning message, block network traffic from infected devices and/or create an analytical report for the malicious domain name.

13. The computing device of claim 10, wherein the processing module is configured to analyze the obtained set of features of suspiciousness of the analyzed domain name using at least one previously trained machine-learning algorithm.

14. The computing device of claim 1, wherein the analyzing module is further configured to receive, via the communication module, at least one domain name from at least one source of domain names.

15. A method to detect malicious domain names in network traffic in real time implemented by a computing device, the method comprising:

receiving, using a communication module of the computing device, network traffic from
5 a data network,

extracting, using a filtering module of the computing device, a plurality of data packets from the received network traffic,

analysing, using the filtering module, said extracted data packets, ensuring the extraction of at least one domain name from said plurality of data packets,

10 analysing, using an analyzing module of the computing device, obtained domain names, using a given set of analysis methods, ensuring that a given numerical value is assigned to each of a given set of features of a domain name suspiciousness corresponding to one of the given set of analysis methods for each analyzed domain name depending on the results of its analysis using the specified methods of analysis,

15 analysing, using a processing module of the computing device, obtained features of a domain name suspiciousness with the assigned numerical values using a specified set of analysis rules with ensuring that each domain name is classified as malicious domain names, if the results of the analysis of features of suspiciousness are characteristic of malicious domain names.

20 16. The method of claim 15, wherein the step of receiving network traffic further comprises connecting, using the communication module, to at least one of the network traffic capture devices included in the data network.

25 17. The method of claim 15, wherein, for the extracted domain names, it is further determined by the filtering module whether there is an analytical report for each of these extracted domain names.

18. The method of claim 17, wherein the step of detecting the presence of the analytical
30 report includes performing, by the filtering module, at least the following sub-steps:

receiving data on analytical reports,

establishing the correspondence of each of the extracted domain names to one of the available analytical reports using character-by-character comparison of each of the extracted domain names with domain names that identify analytical reports in the indicated obtained data.

19. The method of claim 17, wherein, in response to the absence of analytical report for the extracted domain name, the specified domain name is transmitted to the analyzing module using the filtering module.

20. The method of claim 15, wherein the step of analyzing the obtained domain names comprises performing, with the analyzing module, at least the following sub-steps:

determining the number of characters in each of the received domain names,

determining whether a certain number of characters correspond to each analyzed domain name given the threshold value of the length of the domain name,

assignment of a given numerical value to a suspiciousness attribute of a domain name, representing the length of the domain name being analyzed, depending on the result of establishing compliance.

21. The method of claim 20, wherein the step of analyzing the obtained domain names comprises performing, with the analyzing module, at least the following sub-steps:

getting a set of known domain names,

establishing the coincidence of each of the extracted domain names with one of the known domain names using character-by-character comparison of each of the extracted domain names with domain names in the obtained set of domain names,

assignment of a given numerical value to another feature of a domain name suspiciousness, representing popularity of the analyzed domain name, depending on the result of establishing a match.

22. The method of claim 21, wherein the step of analyzing the obtained domain names comprises performing, with the analyzing module, at least the following sub-steps:

determining the frequency of occurrence of each of the many N-grams in each of the obtained domain names, each N-gram corresponding to a combination of N consecutive characters,

determining the entropy of the N-gram of the specified analyzed domain name as a function of the specific frequency of occurrence of each N-gram in this analyzed domain name,

comparison of the determined entropy of the N-gram of the analyzed domain name with the specified threshold value of the entropy of the N-gram,

assignment of a given numerical value to another feature of a domain name suspiciousness, representing entropy of the analyzed domain name, depending on the result of the comparison.

23. The method of claim 22, wherein the step of analyzing the obtained domain names comprises performing, with the analyzing module, at least the following sub-steps:

getting data about language dictionaries,

extracting at least one word from each of the received domain names,

5 matching each of the specified extracted words with one of the words in the language dictionaries from the specified received data,

assignment of a given numerical value to another feature of a domain name suspiciousness, representing meaningfulness of the analyzed domain name, depending on the result of establishing a match.

10 24. The method of claim 23, wherein the step of analyzing the obtained domain names comprises performing, with the analyzing module, at least the following sub-steps:

getting data about language dictionaries,

extracting at least one word from each of the received domain names,

15 determining the Levenshtein distance between each of the specified extracted words and the corresponding one of the words in the language dictionaries of the specified obtained data,

comparison of a certain Levenshtein distance with a predetermined threshold value

20 assignment of a given numerical value to another feature of a domain name suspiciousness, representing correctness of spelling of the analyzed domain name, depending on the result of the comparison.

25 25. The method of claim 15, wherein the given set of analysis methods is further modified by the analyzing module.

26. The method of claim 15, wherein, in response to assigning a domain name to malicious domain names, the method further comprises, using the processing module, issuing an alert message, blocking network traffic from infected devices and/or creating an analytical report for the malicious domain name.

30 27. The method of claim 24, wherein the step of analyzing the features of suspiciousness includes analyzing, with the processing module, the obtained set of features of suspiciousness of the analyzed domain name using at least one previously trained machine-learning algorithm.

35 28. The method of claim 15, wherein according to the analyzing module, a plurality of domain names from at least one source of domain names are additionally obtained via the communication module.

29. Machine-readable medium for long-term storage of data storing machine-readable instructions that, when executed by a computing device, cause the execution of the steps of the method of any of claims 15-28.

5

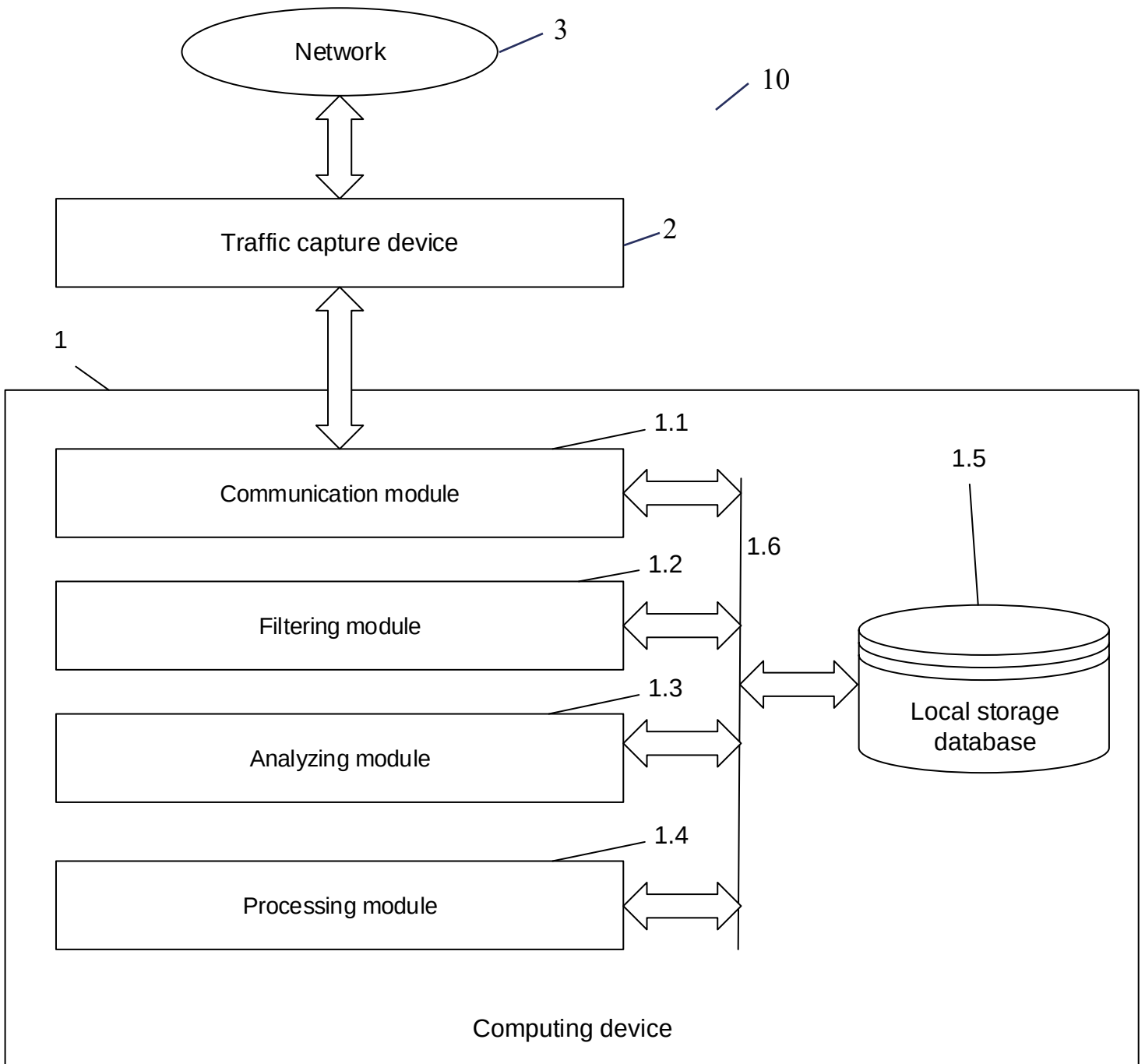


Fig. 1

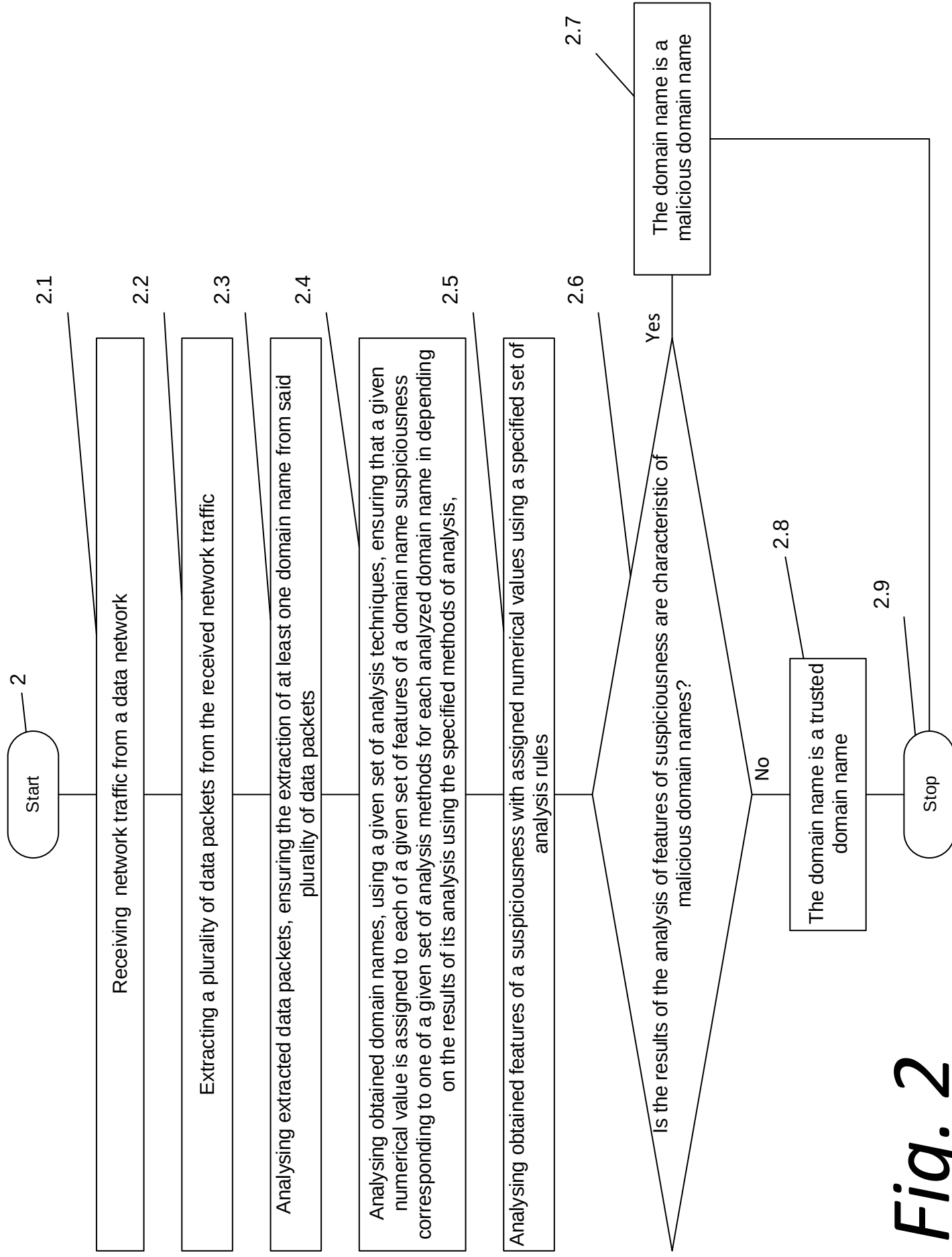


Fig. 2