

METHOD AND COMPUTING DEVICE FOR INFORMING ABOUT MALICIOUS WEB RESOURCES

This technique relates to the field of information security, in particular to a method and
5 computing device for informing about malicious web resources.

BACKGROUND

In order place a web resource on the Internet, it is necessary to upload its files to a web
server of the hosting provider, which is constantly connected to the Internet and which runs
10 special software necessary for processing requests to a web resource. When a hosting provider
is contacted, the owner of the web resource receives a personal account, and the web resource
receives an IP address issued by this hosting provider, and the issued IP address is assigned to
the account issued to the owner of the web resource. Thus, based on the IP address of the web
resource, it is possible at least to determine the hosting provider that issued the account using
15 this IP address. It shall be noted that hosting providers usually provide their services under
certain conditions, wherein the hosting provider may, among other things, suspend provision of
its services if a web resource with malicious and / or illegal content is hosted on its web server,
which implies blocking such a malicious web resource by the hosting provider by its IP address,
as a result of which this web resource is no longer accessible to the Internet users.

20 For the convenience of storing the address space of a web resource and enabling the
transition from one hosting provider to another hosting provider without the need to change a
single web resource pointer ("URL"), by entering which in the address line of the web browser,
the user can access specified web resource, the web resource owner can use the capabilities of
the domain name system, wherein such a web resource can be assigned a domain name that is
25 registered with a domain name registrar, at the same time, any combination of letters and
numbers that does not violate the rules of the selected domain zone can be chosen as a
registered domain name. To automatically convert the registered domain name of a web
resource to its IP address, usually specified when registering a domain name, the DNS servers
are used that store information about the compliance of certain domain names with IP addresses
30 of web resources issued by hosting providers. It shall be noted that domain name registrars,
similarly to hosting providers, also usually provide their services under certain conditions,
wherein the domain name registrar may, among other things, block a domain name registered
by this domain name registrar, if, for example, it will learn that this domain name belongs to a
web resource with malicious and / or illegal content. Thus, if the registrar blocks the domain
35 names of a specific domain name after a certain period of time, the domain name of the web
resource entered by the user in the address line of the browser will not be converted to an IP
address, as a result of which connection to the requested web resource will not occur (i.e., the

user will not be able to access the web resource), and the browser will give the user an error message, such as, for example, the message “Could not find the IP address of the server”.

Thus, one of the most significant reasons for suspending provision of the above services by a hosting provider and / or domain name registrar is that they will receive information that their associated web resource is malicious, that is, comprises malicious and / or illegal content.

In order to identify malicious web resources and to send notifications to authorized entities about detected malicious web resources, various intelligent systems are used for their subsequent blocking.

One of the illustrative examples of such an intelligent system is described in KR 101514984 B1 (publ. on 24.04.2015; G06F 21/56).

In particular, the patent KR 101514984 discloses a system for detecting malicious code distributed by web pages. The system under KR 101514984 is configured to connect to web pages of various web resources for the implementation of various user actions, to identify any behavioral model associated with the spread of malicious code, and to send a notification to the hosting server that hosts this malicious code, for ensuring that it can take the necessary measures before the distribution of this malicious code in accordance with the identified behavioral model.

In another patent KR20070049514 (publ. on 11.05.2007; G06F 11/00) a system claimed for detecting malicious code, comprising a block for obtaining references to many web resources; database for storing information about known malicious code; a search unit to search for malicious code among the received references by identifying whether the suspicious code matches the malicious code, information about which is stored in the database; and a notification block for sending a notification about the presence of malicious code to a web resource on which this malicious code was found by the search block, for later removal of the source code for generation of html documents, a program, an image, a pop-up window, etc., embedded in suspicious code, or blocking the domain through which the malicious code is distributed.

It shall be noted that the known information systems only allow to send a separate notification about one malicious web resource detected during a sequential check of the analyzed web resources for maliciousness to one authorized entity associated with this malicious web resource; however, there is a possibility that such notification will be ignored by an authorized entity, with the result that such a web resource will continue to work for the abusers, distributing malicious and / or illegal content on the web. It shall be noted that the known information systems do not use means and mechanisms that allow simultaneously informing a wide range of authorized entities that may influence the decision to block a web resource with malicious and / or illegal content or which may make such a decision about

malicious web-resources with similar signs of suspicion, having similar malicious activity and / or belonging to the same abuser or the same group of abusers.

Thus, there is an evident need to further improve the means for informing about malicious web resources, in particular, to improve the effectiveness of informing authorized entities about the identified web resources with malicious and / or illegal content.

Consequently, technical problem solved by this technique is the creation of improved means for informing about malicious web resources, in which the above-mentioned disadvantage of known information tools, is at least partially eliminated, consisting of the low efficiency of informing authorized entities in the identified web resources with malicious and / or illegal content.

DISCLOSURE

The said technical problem is solved in one of the aspects of this technique, wherein a method of informing about the malicious nature of web resources according to this technique is claimed, performed on a computing device, wherein according to this method: references to many web resources are obtained; malicious web resources in a specified set of web resources are identified; web resources associated with each of the identified malicious web resources are determined; malicious web resources among a set of identified related web resources are established; at least one authorized entity associated with each of the identified malicious web resources is determined; at least one report for at least one of the established authorized entities based on information about the detected malicious web resources associated with this authorized entity is generated; each generated report is sent to the appropriate authorized entity on the basis of the contact details of this authorized entity.

In one of embodiments of this technique, in order to obtain references to a set of web resources, at least one of the following operations are performed, wherein: a request is sent to at least one reference source in order to obtain at least one reference to a web resource from it; messages are received from at least one computing device, ensuring their processing to retrieve at least one reference to a web resource; messages are received from at least one mobile device ensuring their processing to retrieve at least one reference to a web resource; and search queries are entered into at least one search engine using a specific list of keywords to identify contextual advertising in search results received in response to each search query in each of these search engines, ensuring that at least one reference to a web resource is retrieved from the identified contextual advertising.

In another embodiment of this technique, in order to establish related web resources, at least one of the following is determined: whether the domain names of web resources have a similar spelling; whether the domain names are registered to the same person; whether the same registrant personal data is specified for registered domain names of web resources; whether the

domain names of web resources are located at the same IP address; and whether the references corresponding to the web resources have the same or similar single pointer to the web resource “URL”.

In another embodiment of this technique, in order to establish communication of web resources, at least the following operations is performed, wherein: a mathematical model in the form of a graph is created, wherein the vertices of the created graph correspond to at least the first web resource and to at least the second web resource, and the graph edges represent the references between at least the first web resource and at least the second web resource by at least one web resource parameter that is common for at least the first web resource and for at least the second web resource, wherein the number of references per parameter of the web resource between one first web resource and the second web resources is limited by a specified threshold value; by means of a known machine learning algorithm, the weights are assigned to the references between at least the first web resource and the second web resource based on the parameter of the first web resource and the second web resource; the link coefficient is determined as the ratio of the number of links one parameter of a web resource between one first web resource and second web resources and the weight of each link under one parameter of a web resource between the first web resource and the second web resources; and the links between at least the first web resource and at least the second web resource are deleted if the value of the determined communication coefficient is less than the predetermined threshold value.

In some embodiments of this technique, in order to identify malicious web resources, it is established whether each resulting reference is at least partially related to one of the known malicious references.

In other embodiments of this technique, in order to identify malicious web resources, in addition to the operation, wherein it is established whether each received reference at least partially coincides with one of the known malicious references, at least one of the following operations is performed, wherein: the domain name of the web resource is analyzed for maliciousness using at least one method of the domain names analysis; from a web resource at least one file is obtained for its analysis for maliciousness using at least one file analysis method; and the html-code of the web resource is obtained for its analysis for maliciousness using at least one html-code analysis method.

In some other embodiments of this technique, when analyzing the domain name of a web resource for maliciousness, it is further established whether this analyzed domain name matches one of the known malicious domain names.

In other embodiments of this technique, when analyzing a file received from a web resource, the hash sum of the analyzed file received from the web resource is additionally

calculated and it is established whether the calculated hash sum of the analyzed file matches the hash sum of one of the known malicious files.

In other embodiments of this technique, when analyzing the obtained html-code of a web resource, search is done in the specified html-code for specific keywords indicating the malicious nature of the web resource.

According to one of the embodiments of this technique, when establishing authorized entities associated with each of the identified malicious web resources, the owner, administrator, hosting provider, and / or domain name registrar associated with this malicious web resource is determined.

According to another embodiment of the this technique, the claimed method may include an additional step, wherein a threat type is set from a predetermined set of threat types for each detected malicious web resource, and when generating each report, a template from a predetermined set of report templates is used, with each template corresponding to one of the identified types of threats and one of the established authorized entities.

In another embodiment of this technique, the number of reports generated for each authorized entity may correspond to the number of identified types of threats.

In another embodiment of this technique, evidence of the maliciousness of each web resource, the details of which are comprised in this report, may be additionally added to each generated report.

BRIEF DESCRIPTION OF THE DRAWINGS

Accompanying drawings, which are provided for a better understanding of the essence of this technique, are included in this document to illustrate the following embodiments of this technique. Accompanying drawings, in conjunction with the description below, serve to explain the essence of this technique.

In the drawings:

Fig. 1 schematically shows a system for informing about malicious web resources;

Fig. 2 shows one of the options for implementation of a device for informing about malicious web resources;

Fig. 3 shows a flowchart of a method for informing about malicious web resources.

DETAILED DESCRIPTION

Some examples of possible embodiments of this technique are described below, and it shall not be assumed that the following description defines or limits the scope of this technique.

System for informing about malicious web resources

Fig. 1 schematically shows a system 300 for informing about malicious web resources, comprising a computing device 200 for informing about malicious web resources, a source of 120 references to web resources comprising references to potentially harmful web resources, a source of 130 references to web resources, comprising references to potentially harmful web resources, as well as computer device 140, mobile device 150 and Internet network 110.

In one of the embodiments of this technique, a source of 120 references to web resources may be an antiphishing.org site with references to known malicious web resources, and source of 130 references to web resources may be an antifraud.org site with the references to known malicious web resources. In such embodiment of this technique, all data streams transmitted from the source of 120 references, and all data streams transmitted from the source 130 shall be associated respectively with a unique identifier assigned to the source of 120 references and a unique identifier assigned to the source of 130 references, wherein the wherein computing device 200 described below shall be pre-programmed or configured to identify data streams from such sources of references, in particular from sources of 120, 130 references, based on their unique identifiers comprised in these data streams and previously known to computing device 200.

The source of 120 references comprises, among other things, a control block 122, an API-interface 124 that provides ability to interact with the control block 122, and a database of 126 references to web resources, which stores, for example, references to web resources collected from third-party sources that comprise potentially malicious and / or illegal content, and supporting information that attributes these references.

The source of 130 references also comprises, among other things, a control block 132, an API-interface 134 that provides ability to interact with the control block 132, and a database of 136 references to web resources, which stores, for example, references to web resources collected from third-party sources with potentially harmful and / or illegal content, and supporting information that attributes these references.

The computing device 200 according to this technique is connected to the source of 120 references and the source of 130 references, respectively, through the parser 160, configured to connect to the API-interface 124 of the source of 120 references and pre-configured to work with it, and the parser 170, configured to connect to the API-interface 134 of the source of 130 references and pre-configured to work with it, wherein the parser 160 is connected configured to communicate with the computing device 200 using the Internet network 110, and parser 170 is wired to communicate directly with the computing device 200 itself. It shall be noted that each of the API-interface 124 and API-interface 134 may have its own command syntax, so the parser 160 working with the API-interface 124 shall be pre-programmed to understand the command syntax of this API-interface 124, and the parser 170 , working with API-interface

134, shall be pre-programmed in a similar way to understand the command syntax of the API-interface 134, wherein setting the parser 160 and the parser 170 to work, respectively, with the API-interface 124 and the API-interface 134 occurs during the initial switching of the computing device 200 to the sources of 120, 130 references. It shall be noted that the parsers 160, 170 can each be implemented as a separate server or other known computing device.

The computing device 200 according to this technique is configured to send requests to each of the sources of 120, 130 references, for example, requests to send at least some references to potentially malicious web resources to the computing device 200, all references to potentially malicious web sites. resources or only references to potentially malicious web resources, stored respectively in the base of 126 references or base of 136 references for a given period of time. Due to the use of parsers 160, 170 preconfigured to work respectively with API-interfaces 124, 134, requests directed by the computing device 200 to the sources of 120, 130 references will comprise commands that are understandable respectively for the control blocks 122, 132, with the result that these blocks 122, 132 of management will be able to properly process and respond to these requests, in particular, transfer the requested references to potentially malicious web resources to the computing device 200 from which these requests were received.

In particular, in response to received requests, the control blocks 122, 132 receive access respectively to the base of 126 references and the base of 136 references, ensuring that they retrieve the requested references to potentially malicious web resources (also referred to in this document as potentially malicious references) and transmission, by means of API-interfaces 124, 134, extracted potentially malicious references, respectively, to parsers 160, 170, wherein the parser 160 provides configured to process the output data stream from API-interface 124 to extract from it potentially malicious references requested from reference source of 120 references and the parser 170 provides configured to process the output data stream from the API-interface 134 to extract from it potentially malicious references requested from the source of 130 references. It shall be noted that to extract the necessary references to web resources from the processed data stream, the parsers 160, 170 use each corresponding regular expression from a given set of regular expressions known to it. In particular, the output data stream from any of the API-interfaces 124, 134 comprises both potentially malicious references themselves and identification data describing the potentially malicious references being transmitted, for example, the date and time of the references entering the database references to web resources, reference source identification data and / or other necessary attributes of these transmitted references. The output data stream from any of the API-interfaces 124, 134 is typically a collection of characters as a string of characters with a specified description format, which is divided into structural elements using some predefined character, such as the “#” (grid) character, wherein the writing format of such a string of characters is known to parsers 160,

170, due to the fact that they are pre-programmed or configured to work with one of the corresponding API-interfaces 124, 134. In particular, parsers 160, 170 must know a keyword, a key symbol or a key label indicating the presence of the reference following it, and other key words / symbols / labels commonly used in the resulting character strings to indicate the presence of some other identifying information following such keywords. When these character strings are received from API-interfaces 124, 134, parsers 160, 170, respectively, extract from these received strings, each divided into a known set of structural elements, potentially malicious references to web resources, and at least some of the identification data, describing these potentially malicious references, ensuring that the retrieved potentially malicious references to web resources are transferred to computational device 200 for their subsequent analysis, features of which will be described below.

In the case of sending to one of the sources of 120, 130 references a request to transfer to the computing device 200 potentially malicious references saved respectively in the base of 126 references or in the base of 136 references for a given period of time, for example, all recently saved potentially malicious references starting from a certain point in time, for example, last few minutes, hours, days, weeks, months, etc. depending on the tasks, wherein such requested potentially malicious references are installed, for example, using the readings of the system clock of the corresponding source of references, in accordance with which, among other things, each of the saved potentially malicious references is set.

In one of the embodiments of this technique, the computing device 200 may be configured to connect directly to each of the sources of 120, 130 references with provision of direct access, respectively, to their bases of 126, 136 references to extract from them potentially malicious references for their subsequent processing by the computational device 20, the features of which are described below.

In another embodiment of this technique, the parsers 160, 170 may be both connected with a possibility of data exchange with computing device 200 using the Internet 110 network.

In another embodiment of this technique, the parsers 160, 170 can both be wired to communicate directly with the computing device 200 itself.

In some embodiments of this technique, sources of 120, 130 references can each be configured to exchange data with parsers 160, 170 using the Internet 110 network, and the parsers 160, 170 themselves can be both wired directly to a computing device 200.

The computer device 140, which may be represented, among other things, by a desktop computer, laptop, server, etc., is configured to communicate with the computing device 200 via the parser 180, wherein the computing device 140 is wire-connected to the parser 180 with a possibility of sending e-mails to it, for example, to an e-mail address associated with this parser 180, wherein the transmitted electronic messages have the specified description format, similarly to the above output streams of API-interfaces 124, 134. The parser 180 is pre-

programmed or configured to work with the computing device 140, so that the parser 180 receives electronic messages from the computing device 140, and the parser 180 knows the recording format of the received electronic messages. Similarly to the working process of the parser 160 or parser 170 described above, the parser 180 processes each received e-mail and extracts from its text the necessary references to web resources (these references have their own specific recording format) and at least some of the identification data, describing these extracted references, with the provision of transmission, through the Internet 110 network, of the extracted references, set in accordance with the extracted identification data, into the computing device 200 for their subsequent analysis, features of which will be described below. It shall be noted that the parser 180 uses the corresponding regular expression from a given set of regular expressions known to it to extract the necessary references to web resources from the text of the processed electronic messages. The parser 180 may be implemented as a separate server or other known computing device.

In one of the embodiments of this technique, the computing device 140 may be configured to transmit messages to the parser 180 using the Internet 110 network, and the parser 180 may be wired directly to the computing device 200.

The mobile device 150, which may be represented, among other things, by smartphone, cell phone, tablet, etc., is configured to communicate with the computing device 200 using two communication channels. In particular, for data exchange between the mobile device 150 and the computing device 200 through one of these communication channels, the mobile device 150 is wire-connected to the parser 190 with the possibility of sending electronic messages to it, comprising, among other things, references to potentially malicious web resources, at the e-mail address associated with this parser 190, wherein the transmitted e-mails have a set description format similar to the above output streams of API-interfaces 124, 134. The parser 190 is pre-programmed or configured to work with the mobile device 150, so that this parser 190 receives electronic messages from the mobile device 150, wherein the parser 190 knows recording format of the received electronic messages. Similarly to the working process of the parser 160 or parser 170 described above, the parser 190 extracts references to web resources from each received e-mail (these references have their own specific record format) and at least some of the identification data describing these extracted references, with the provision of transmission, via the Internet 110 network of the extracted references to web resources, set in accordance with some of the extracted identification data, to the computing device 200 for their subsequent analysis, the characteristics of which will be described below. It shall be noted that the parser 190 may be implemented as a separate server or other known computing device.

In addition, for data exchange between the mobile device 150 and the computing device 200 via another communication channel, the mobile device 150 is connected, via the cellular network 115, to the parser 195 configured to transfer, for example, SMS messages and / or

MMS messages comprising, among other things, references to web resources, by the contact number associated with this parser 195, wherein the transmitted SMS messages and / or MMS messages have the specified description format, similar to the above described output streams of API-interfaces 124, 134. Parser 195 is pre-programmed or configured to work with mobile device 150, due to which this parser 195 receive SMS messages and / or MMS messages from mobile device 150, wherein parser 195 knows the recording format of the received SMS messages and / or MMS messages . To receive SMS and MMS messages sent from mobile device 150 to parser 195 via cellular network 115, parser 195 is connected to an external modem equipped with a SIM card. Similarly to the working process of the parser 160 or parser 170 described above, the parser 195 extracts references to web resources from each received SMS or MMS message (and these references have their own specific recording format) and at least some of the identification data, describing these extracted references, such as the sender's contact number, with the transfer of these extracted references, set in accordance with some extracted identification data, into the computing device 200 wire-connected to the parser 195, for their subsequent analysis, the features of which will be described below. It shall be noted that the parser 195 uses the corresponding regular expression from a given set of regular expressions known to it to extract the necessary references to web resources from the text of the received electronic messages. The parser 195 may be implemented as a separate server or other known computing device.

In some embodiments of this technique, the converting module connected to the parser 180 and the converting module connected to the parser 190 can be implemented as a single converting module wire-connected an / or wirelessly configured to exchange data with the computing device 140 and the mobile device 150, and having functions similar to the functions of these connected converting modules.

In one of the embodiments of this technique, the computing device 200 may be configured to connect directly to each of the computing device 140 and the mobile device 150 with provision of direct access to their internal databases located in the memory of these devices, to receive messages from them, e.g. SMS, MMS, email, etc. (wherein on each of the computing device 140 and the mobile device 150, for example, a special client program shall be installed). Computing device 200 can process each received message to extract references from it for further processing by computing device 200, the features of which are described below.

In another embodiment of this technique, the parsers 180, 190, 195 may be each connected with a possibility of data exchange with computing device 200 using the Internet 110 network.

In another embodiment of this technique, the parsers 180, 190, 195 can each be wired to communicate directly with the computing device 200 itself.

In some embodiments of this technique, the computing device 140 and the mobile device 150 can each be configured to exchange data with the parsers 180, 190 using the Internet 110 network, and the parsers 180, 190 themselves can be wired both directly to the computing device 200.

5 In other embodiments of this technique, the parser 195 may be connected configured to communicate with a computing device 200 using the Internet 110 network.

It shall be noted that the source of 120 references, the source of 130 references, the computing device 140, and the mobile device 150 are shown on Fig. 1 solely as an example, that is, it shall not be considered that a possible embodiment of the system 300 for informing
10 about malicious web resources is limited to the example shown on Fig. 1, it shall be clear to those skilled in the art that system 300 may comprise two or more reference sources, each similar to the above referenced source 120, two or more reference sources, each similar to the above described reference source 130, two or more computing devices, each similar to the above described computing device 140, and / or two or more mobile devices, each similar to the
15 above described mobile device 150.

In one of the embodiments of this technique, each of the reference sources, each similar to the above-described reference source 120, can be connected to the computing device 200 by means of a separate parser with functionality similar to the above-described parser 160, and each such separate parser will be pre-programmed or configured to work with the appropriate
20 reference source to understand the syntax of the API-interface commands of this reference source.

In another embodiment of this technique, all reference sources in system 300, each similar to the above described reference source 120, can be connected to computing device 200 by means of a single parser with functionality similar to the above described parser 160, and
25 such a common parser shall be preprogrammed or configured to work with each of these connected reference sources to understand the syntax of the commands of its API-interface.

In some embodiments of this technique, each of the reference sources, each similar to the above-described reference source 130, may be connected to the computing device 200 via a separate parser with functionality similar to the above-described parser 170, wherein each such
30 individual parser will be pre-programmed or configured to work with the appropriate source of references to understand the syntax of the API-interface commands of this reference source.

In other embodiments of this technique, all reference sources in system 300, each similar to the above described reference source 130, may be connected to computing device 200 via a single parser with functionality similar to the above described parser 170, and such a
35 common parser shall be preprogrammed or configured to work with each of these connected sources of links to understand the syntax of the commands of its API-interface.

In other embodiments of this technique, each of the computing devices each similar to the above described computing device 140 may be connected to the computing device 200 via a separate parser with functionality similar to the above described parser 180, wherein each such separate parser will be pre-programmed or configured to work with appropriate computer device
5 to understand the format of recording electronic messages received from this computing device.

In some other embodiments of this technique, all computing devices in system 300, each similar to the above-described computing device 140, may be connected to computing device 200 via a single parser with functionality similar to the above-described parser 170, and such a common parser shall be preprogrammed or configured to work with each of these
10 connected computing devices to understand the format for recording electronic messages received from this computing device.

In some embodiment of this technique, each of the mobile devices, each similar to the above described mobile device 150, can be connected to the computing device 200 via a separate parser with functionality similar to the above described parsers 190, 195, wherein each
15 such parser will be pre-programmed or configured to work with an appropriate mobile device to understand the format of recording messages received from this mobile device, in particular electronic messages, SMS-messages and / or MMS-messages.

In another embodiment of this technique, all mobile devices in system 300, each similar to the mobile device 150 described above, can be connected to the computing device 200 by
20 means of a single parser with functionality similar to the parsers 190, 195, wherein such a common parser is first programmed or configured to work with each of these connected mobile devices to understand the format for recording messages received from this mobile device, in particular e-mails, SMS-messages and / or MMS-messages.

According to one of the embodiments of this technique, at least a part of reference
25 sources, each similar to reference source 120, reference sources, each similar to reference source 130, computing devices each similar to the computing device 140, and mobile devices each similar to the mobile device 150 can be connected to the computing device 200 by means of one parser with functionalities similar to the above described parsers 160, 170, 180, 190 and 195, wherein such a common parser shall be properly preprogrammed or configured to work with
30 each of the connected reference sources to understand the syntax of the commands of its API-interface, each of the connected computing devices to understand the recording format of electronic messages received from this computing device, and each of the connected mobile devices to understand the message recording format of the types described above, received from this mobile device.

According to another embodiment of this technique, a computing device may be
35 subscribed to an RSS casting under at least one of the reference sources, each similar to the above described source of 120 references, and / or an RSS casting of at least one of the

reference sources, each similar to the source of 130 references, to receive at least one report from specified reference sources, indicating, for example, the appearance of at least one new reference to a web resource in the corresponding reference source.

In accordance with some embodiments of this technique, system 300 may additionally comprise a separate reference base that is external or remote with respect to computing device 200, wherein the parsers 160, 170, 180, 190 and 195 each can be executed with the possibility to gain access to this external reference bases configured to record references in it, extracted properly in accordance with the description, as the result of which this external reference base comprises many references to potentially malicious web resources each put in accordance with the auxiliary identification data describing the reference, such as the date and time of archiving and / or at least one other identifier. Computing device 200 is configured to obtain access to such a reference base with the possibility of extracting from it the necessary references for their subsequent processing, the features of which are described below. As an addition or alternative in this embodiment, the above external base of the links may also comprise many links to known malicious web resources.

According to other embodiments of this technique, the system 300 may comprise only a computing device 200 and a structured reference base that is external or remote with respect to the computing device 200. In this embodiment, the external reference base comprises references to potentially malicious web resources archived from any different sources, with each reference in this external reference base being associated with auxiliary identification data describing this reference, for example, date and time of storage and / or at least one other identifier. Computing device 200 is configured to gain access to such an external reference base with the possibility of extracting necessary references from it for subsequent processing, the features of which are described below. As an addition or alternative in this embodiment, the above external base of the links may also comprise many links to known malicious web resources.

Computing device for informing about malicious web resources

The computing device 200 shown on Fig. 2, is designed to inform authorized entities about the identified malicious web resources and is essentially a hardware-software complex implemented as a general-purpose computer, having the structure described below, which is well known to those skilled in the art.

It shall be noted that in this document, an authorized entity is an individual who can block the operation of a web resource or influence the decision to block a malicious web resource or suspend its operation, for example, the administrator of a web resource, the owner of a web resource, etc., or a legal entity that can block the operation of a web resource or influence the decision to block or suspend a malicious web resource, such as a domain name registrar, a hosting provider etc.

In particular, a general-purpose computer usually comprises a central processor, system memory, and a system bus, which in turn comprises various system components, including memory associated with the central processor. A system bus in such a general-purpose computer comprises a memory bus and a memory bus controller, a peripheral bus and a local bus, configured with a possibility of interaction with any other bus architecture. System memory comprises read-only memory (ROM) and random access memory (RAM). The Basic Input / Output System (BIOS) comprises the basic procedures that ensure the transfer of information between the elements of such a general-purpose computer, for example, when the operating system boots using the ROM. In addition, a general purpose computer comprises a hard disk for reading and recording data, a magnetic disk drive for reading and recording to removable magnetic disks, and an optical drive for reading and recording on removable optical disks such as CD-ROM, DVD-ROM and other optical storage media, but other types of computer storage media can be used to store data in machine-readable form, such as solid-state drives, flash cards, digital disks, etc., and connected to the system bus via controller. At a general-purpose computer, a hard disk, a magnetic disk drive and an optical drive are connected to the system bus via a hard disk interface, a magnetic disk interface and an optical drive interface, respectively. Drives and associated computer storage media are non-volatile means of storing computer instructions, data structures, program modules and other general-purpose computer data. A general purpose computer has a file system that stores a recorded operating system, as well as additional software applications, other software modules and program data. The user can enter commands and information into a general-purpose computer using known input devices, such as a keyboard, mouse, microphone, joystick, game console, scanner, etc., wherein these input devices are usually connected to a general-purpose computer via a serial port, which is in turn connected to the system bus, but they can also be connected in some other way, for example, using a parallel port, a game port, or a universal serial USB bus. A monitor or other type of display device is also connected to the system bus via an interface, such as a video adapter. In addition to the monitor, a personal computer can be equipped with other peripheral output devices, such as speakers, a printer, etc. A general purpose computer can work in a network environment, and a network connection can be used to connect to one or more remote computers. Network connections can form a local area network (LAN) and wide area network (WAN). Such networks are usually used in corporate computer networks and internal networks of companies, wherein they have access to the Internet. In a LAN or WAN network, a general purpose computer is connected to the local network via a network adapter or network interface. When using networks, a general purpose computer may use a modem, network card, adapter or other means of providing connection with a global computer network, such as the Internet, and

these means of communication are connected to the system bus via a serial port. It shall be noted that in the ROM of the general purpose computer or at least in any of the above computer-readable media that can be used in a general-purpose computer, the computer-readable instructions can be stored which can be accessed by the CPU of the general-purpose computer, wherein execution of these machine-readable instructions on a general-purpose computer may cause the central processor to execute various procedures or operations described later in this document.

In one of the embodiments of this technique, the computing device 200 may be implemented as a single computer server, such as a Dell TM PowerEdge TM server using the Ubuntu Server 18.04 operating system. Besides, in other embodiments of this technique, the computing device 200 may be presented in the form of a desktop personal computer, laptop, netbook, smartphone, tablet, and other electronic computing device suitable for solving the set tasks.

In other embodiments, the computing device 200 may be manufactured in the form of any other combination of hardware, software or software and hardware complex, suitable for solving tasks.

In some embodiments of this technique, the system 300 may comprise at least two computing devices, each similar to computing device 200, and the functionality described below of the computing device 200 may be divided in any appropriate way between at least two computing devices, wherein each of them for example, can be manufactured as a separate computer server.

The computing device 200 shown on Fig. 2 comprises a communication module 10, an analyzing module 100 and a local data storage 20, each connected to a communication bus 30, wherein each communication module 10 and the analyzing module 100 being able to exchange data via the communication bus 30 with a local storage 20 data, and the communication module 10 is also configured to exchange data with the analyzing module 100.

In one of the embodiments of this technique, the above-described parsers 160, 170, 180, 190 and 195 can each be implemented as a separate data preprocessing module embedded in the computing device 200 (i.e., included in this computing device 200) and having the above-described functionality of one of the corresponding parsers 160, 170, 180, 190 and 195, in particular the functionality for providing interaction or data exchange between the computing device 200 and one of the corresponding reference source 120, reference source 130, computing device 140 and mobile device 150 (i.e., each of these separate data preprocessing modules shall be pre-programmed to work with one of the corresponding reference source 120, reference source 130, computing device 140 and mobile device 150) and on processing input data streams from one of the corresponding reference source 120, reference source 130, computing device 140 and mobile device 150. In one form of this embodiment, the communication module 10 of

the computing device 200 may be made multi-channel, for example, four-channel, with each of the communication channels in such a communication module 10 being pre-configured to exchange data via the communication bus 30 with one of the above described modules for data processing and data exchange with one of the corresponding reference source 120, reference source 130, computing device 140 and mobile device 150. In another version of this embodiment, the computing device 200 may be equipped with four communication modules, each similar to the communication module 10, each of these communication modules being pre-configured to exchange data via the communication bus 30 with one of the above described separate data processing modules and data exchange with one of the corresponding reference source 120, reference source 130, computing device 140 and mobile device 150. In this embodiment, separate data-processing modules (not shown) are also each configured to interact, via the communication bus 30, with the analyzing module 100 to process requests for receiving references that can be generated by this analyzing module 100, and then sending them from the computing device 200 to one the corresponding above-described reference source 120, reference source 130, computing device 140 and mobile device 150. It shall also be noted that when processing the input data streams received from one of the corresponding reference source 120, reference source 130, the computing device 140 and the mobile device 150, each of these separate preprocessing modules (not shown) can, among other things, identify or recognize the format of the description of the received input data stream. If the identified data description format does not conform to the unified data description format appropriate for the computing device 200, then each of the separate data preprocessing modules can be further configured to convert this received input data stream into the specified unified format, wherein it can be further implemented with the possibility of communication, via the communication bus 30, with local data storage 20, ensuring that data about the unified data description format (as described below), understood by the computing device 200, are obtained, and with the possibility of comparing the identified and unified data formats indicated to decide whether they are appropriate or inconsistent with each other. Thus, if any of the above described separate data preprocessing modules reveals that among the input data streams received from one of the corresponding reference source 120, the reference source 130, computing device 140 and mobile device 150, there are, for example, voice messages or video messages, then such a separate data preprocessing module converts such messages into text, that is, into such data description format which is understandable to the computing device 200, followed by extracting from it of the references to potentially malicious web resources.

In another embodiments of this technique, the above-described parsers 150, 160, 170, 180, 190 and 195 can be implemented as a single data preprocessing module (not shown) embedded in the computing device 200 (i.e., included in this computing device 200) and having the above-described functionality of all parsers 150, 160, 170, 180, 190 and 195, in particular

the functionality to provide interaction or data exchange between the computing device 200 and each of the reference source 120, the reference source 130, a computing device 140 and a mobile device 150 (i.e., such a single preprocessing module shall be pre-programmed to work with each of the reference source 120, the reference source 130, computing device 140 and mobile device 150) and processing input data streams from each of the reference source 120, reference source 130, computing device 140 and mobile device 150. In this embodiment, a single data processing module (not shown) shall also be connected in the computing device 200 to the communication bus 30 with the possibility of data exchange with the communication module 10 providing interaction between the computing device 200 and the reference source 120, the reference source 130, computing device 140 and mobile device 150, wherein the communication module 10 of the computing device 200 can then be performed, for example, as multichannel, and each of the communication channels in such connection module 10 can be pre-set to communicate with one of the corresponding reference source 120, reference source 130, computing device 140 and mobile device 150. In this embodiment, a single data preprocessing module (not shown) is also configured to interact, via the communication bus 30, with the analyzing module 100 to process requests for receiving references which can be generated by this analyzing module 100, with their subsequent forwarding from the computing device 200 to the above-described reference source 120, reference source 130, computing device 140, and mobile device 150. It shall also be noted that when processing input data streams received from reference source 120, reference source 130, computing device 140 and mobile device 150, a single preprocessing module (not shown) can, among other things, identify or recognize the format for describing this input data streams, and if the identified data description format does not conform to a unified data description format suitable for the computing device 200, then it can additionally be implemented configured to convert these received input data streams into the specified unified format, wherein this single preprocessing module can be additionally configured to communicate, via the communication bus 30, with the local data storage 20, ensuring receipt of data about the unified data description format (as described below), understandable to computing device 200, and configured to compare specified identified and unified data formats for decision-making on their compliance or non-compliance with each other. Thus, if the above described single data preprocessing module reveals that among the input data streams it received from the reference source 120, the reference source 130, the computing device 140 and the mobile device 150, there are, for example, voice messages or video messages, then such single data preprocessing module converts such messages into text, that is, into such data description format that is understandable to computing device 200, with the subsequent extraction of references to potentially malicious web resources from it.

In some embodiments of this technique, the functionality of the above-described parsers 160, 170, 180, 190, 195 can be implemented as additional functionality of the analyzing module 100, in particular, each of the parsers 160, 170, 180, 190, 195 or all of these parsers can be implemented as a separate software module embedded in the computing device 200 and executed by the analyzing module 100.

In one embodiment of this technique, the computing device 200 may further comprise an auxiliary module for collecting contextual advertising (not shown), configured to automatically collect contextual advertising shown or demonstrated to users in known search engines, such as, for example, Bing, Google, Yandex etc., with the provision of extraction from contextual advertising, collected at least in one of these well-known search engines, of at least one reference to a web resource. The contextual advertising collection module is connected to the communication bus 30 and is configured to exchange data via the communication bus 30 with the communication module 10, the local data storage 20 and the analyzing module 100. It shall be noted that recently abusers often resort to distributing references to malicious web resources by placing these links in contextual advertising of well-known search engines, and this malicious advertisement is usually targeted to the most frequent search queries of users in each of these search engines, since such lists of the most popular keywords by users are freely available on the sites of these search engines. In this embodiment of this technique, the data storage 20 further has a separate database of search query keywords comprising several sections, each with the stored keywords of the most frequent search queries of the corresponding one of the well known search engines, to the work with which the module for collection of contextual advertising is pre-configured or programmed, so that all keywords in each specific section of this base are aligned with one of the well known search systems. The module for collecting contextual advertising is also configured to at least periodically update (for example, daily) the database of search query keywords placed in the local data storage 20 for at least one of the search engines known to it, for example, by periodically automatically obtaining an up-to-date list of key words that are most popular with users in a particular search engine, using a specific link to the web page of the site of this search engine stored in the local data storage 20 and retrieved from there by the specified module of collection of contextual advertising when updating a specific section of the database of search queries corresponding to the specified search engine, followed by updating the existing list of keywords of search queries in the base section of the keyword of search queries corresponding to the specified search system, based on the obtained current list of keywords. The contextual advertising collection module is also configured to form at least one search query for at least one of the search engines known to it using at least a part of the keywords comprised in one of the sections of the search query keyword database corresponding to this search engine, and configured to automatically transfer this generated search query to this search engine. The contextual advertising collection

module is also configured to retrieve search results issued by a search engine in response to a transmitted request, and configured to filter the search results to detect contextual advertising in the form of advertisements among them, based on, for example, the “advertising” tag, which is provided for such advertisements, wherein each such advertisement has, among other things, at least one reference to the web resources. The contextual advertising collection module is additionally configured to retrieve, for example, a regular expression known to it, such as, for example, **(https?|ftp)://(-\.)?([^\s/?#-]+\.\.?)+/([^\s]*)?@\$@iS**, at least one reference to web resources from each detected advertisement with provision of transmission by communication bus 30 of each of these references to a web resource to the analyzing module 100 for its subsequent analysis of harmfulness to identify or to determine whether a web resource located in this reference pertains to the malicious web resources, as described in more details hereinafter. Thus, the contextual advertising collection module, for example, can sequentially generate search queries for each particular search engine using some combination of keywords formed from at least part of the keywords in the existing keyword list corresponding to that search engine until the end of this keyword list is reached. It shall be noted that the above described method of obtaining references to web resources by computing device 200 may be an alternative or addition to the above methods of obtaining references to web resources used in system 300. In the described embodiment, an auxiliary module for collecting contextual advertising can be implemented, for example, as a separate processor embedded in the computing device 200.

In one of the embodiments of this technique, the functionality of the context advertising collection module described above can be implemented as additional functionality of the analyzing module 100, in particular, the context advertising collection module can be implemented as a separate software module included in the computing device 200 and executed, for example, by analyzing module 100. In another embodiment of this technique, the contextual advertising collection module may be one of the functional submodules of the analyzing module 100.

In yet another embodiment of this technique, the context advertisement collection module described above may be a separate source of references, for example, a separate server that is external to the computing device 200 and is wire-connected to it and / or wirelessly, ensuring that it can send references to the web -resources, wherein the references to web resources transmitted from such an external source of references can be received by the communication module 10 of the computing device 200.

Local data storage

Local data storage 20 is also designed to store executable software instructions that allow to control the operation of functional modules embedded in computing device 200, in

particular communication module 10 and analyzing module 100, and allow these functional modules to implement their functionality when executing these software instructions. Executable software instructions stored in the local data store 20 also allow to control the operation of any submodules, which in some disclosed embodiments are included in some of the functional modules, for example, the analyzing module 100, and allow these submodules to implement their functionality when executing these software instructions.

Local data storage 20 can also store executable software instructions that allow to control the operation of any additional functional modules embedded in computing device 200 and their submodules, and which allow these additional functional modules and their submodules to implement their functionality when executing these software instructions.

In addition, the local data storage 20 is designed for storing various data used in the operation of the computing device 200, in particular, data on a unified data description format understandable to the computing device 200, data on known malicious references, data on known malicious domain names, data on hash -sum of known malicious files, data on keywords indicating the harmful nature of a web resource, data on hosting provider, data on domain name registrar, a list of known authorized entities, a set of known types of malicious threats to web resources, a set of report templates, etc. The local data storage 20 may also store other data used in the operation of the various functional modules embedded in the computing device 200 and the operation of at least some of the sub-modules included in some of these functional modules.

In addition, auxiliary data used in the work of the analyzing module 100 can also be stored in the local data storage 20, for example, data on language dictionaries and a predetermined threshold value used in the method of analyzing domain names based on the correctness of their spelling; virtual machine image files and a set of rules for analyzing changes in virtual machine state parameters used in suspicious file analysis methods based on changes in virtual machine state parameters, a set of regular expressions used to extract references to web resources from input data streams analyzed in the analyzing module 100, and other auxiliary data.

In the computing device 200 shown on Fig. 2, the communication module 10 is configured to receive the extracted references to web resources transmitted by the parsers 160, 170, 180, 190 and 195 to the computing device 200, and then to save the received references to the web resources in the local data storage 20 in which these received data can be transmitted via communication bus 30. Thus, the local data storage 20 can store the references to web resources extracted from data streams from a references source 120, references to web resources extracted from data streams from a references source 130, references to web resources extracted from messages from computing device 140, and / or references to web resources extracted from messages from mobile device 150, and at least some of the extracted identification data describing such stored references.

In some embodiments of this technique, the local data storage 20 in the computing device 200 may comprise one or more databases, each configured to store at least one separate group of the above groups of data used in the operation of the computing device 200, and / or at least some of the accepted references to web resources.

5 In other embodiments of the computing device 200, at least one separate remote data storage (not shown) can be used, to which the analyzing module 100 of the computing device 200 can gain access using the communication module 10, to store therein at least some of the above described groups data and / or at least part of the accepted references to web resources.

10 In some other embodiments of this technique, computing device 200 may comprise at least one local data storage and at least one remote data storage (not shown), each designed to store at least one of the data groups described above and / or at least parts of the received references to the web resources; in addition, the local data storages are each connected to the analyzing module 100 via the communication bus 30, and the indicated remote data storages are each connected with analyzing module 100 via communication module 10. Thus, for example,
15 an embodiment of this technique is possible in which the computing device 200 comprises a single local data storage 20 that stores, for example, only received references to web resources, and comprises several remote data storages, each storing at least some of the above groups of data used in the operation of the computing device 200.

20 In one embodiment of this technique at least one of the above groups of data and / or received references to web resources can be stored in the corresponding separate local data store (not shown), different from the local data storage 20 and connected via connection bus 30, with the analyzing module 100, which in turn is designed to connect to any of these separate local data storages with the provision of extracting from them of the necessary references to web resources.

25 The analyzing module 100 may 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.), and configured to execute software instructions stored in local data storage 20, with the implementation of the following functionality of the analyzing module 100.

30 Local data storage 20 may be implemented, for example, in the form of one or more known physical computer-readable media for long-term data storage. In some embodiments of this technique, local data storage 20 may be implemented using a single physical device (for example, a single optical storage device, a magnetic storage device, an organic storage device, a storage device on disks, or a different type of storage device), and in other embodiments, local
35 data storage 20 may be implemented using two or more known storage devices.

Communication module

The communication module 10 used in the computing device 200 shown on Fig. 1 and 2, has a wireless connection with the above-described parsers 160, 180, 190 configured to exchange data with them, and also has a wired connection with the above-described parsers 170, 195 configured to exchange data with them.

In one of the embodiments of this technique, the communication module 10 may be connected to all parsers 160, 170, 180, 190, 195 in a wired manner configured to exchange data with them, for example using a coaxial cable, twisted pair, fiber optic cable or other physical connection. In this embodiment, the communication module 10 may be implemented, for example, in the form of 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 provide communication with the parsers 160, 170, 180, 190, 195.

In another embodiment of this technique, the communication module 10 can be connected to all parsers 160, 170, 180, 190, 195 wirelessly configured to exchange data with them, for example using a communication line based on "WiFi" technology, a communication line based on 3G technology, LTE-based communication lines and / or the like. In this embodiment, the communication module 10 may be implemented, for example, 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 communication link used to provide connection with parsers 160, 170, 180, 190, 195.

In other embodiments of this technique, the communication module 10 may use any suitable combination of wired and wireless communication lines to exchange data with at least some of the parsers 160, 170, 180, 190, 195 included in the system 300.

Communication module 10 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, the computing device 200 may additionally be equipped with a SIM card modem for receiving SMS messages and / or MMS messages from mobile devices, such as mobile device 150.

Analyzing module

The analyzing module 100 included in the computing device 200 shown on Fig. 2, may 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.), for example, as a central processor of the above-described general-purpose computer, in the form of which computing device 200 may be implemented.

The analyzing module 100 is configured to access local data storage 20 (separate local data storage or remote data storage, depending on the embodiment, as described above in this document) or to communicate with it using communication bus 30 to ensure extraction from it of the references to web resources for their subsequent analysis, as it will be described below.

In one of the embodiments of this technique, the analyzing module 100 may be configured to communicate, via the communication bus 30, with the communication module 10, ensuring that it can receive links to web resources for their subsequent analysis, as it will be described in more details below. Thus, in this embodiment, the analyzing module 100 may receive links to web resources directly from the communication module 10 immediately after receiving these links by the communication module 10.

In the embodiments of this technique in which the obtained references to web resources are stored in a separate local storage other than the local data storage 20, or in a remote data storage, the analyzing module 100 may be configured to access such a separate or remote data storage or configured to communicate with it using the communication bus 30, ensuring that stored web resource references are extracted from it for subsequent analysis, as will be described in more details below.

The analyzing module 100 is configured to analyze each of the obtained or extracted references to web resources in order to identify or establish web resources with malicious and / or illegal content, also called malicious web resources, among web resources that are located under the analyzed references as will be described in more details below.

In particular, to detect malicious web resources, when analyzing references to web resources, the analyzing module 100 (i) gains access to the local data storage 20 (a separate local data storage or a remote data storage, depending on the embodiment, as described earlier in this document) or establishes communication with it using the communication bus 30, ensuring that data about known malicious references is obtained from it; and (ii) establishes, by character-by-character comparison of each analyzed references with known malicious references from the indicated obtained data, the fact of at least partial coincidence of the analyzed reference with at least one of the known malicious references.

Thus, if the analyzing module 100 has established or discovered that a specific references has at least partial concurrence with at least one of the known malicious references, then this indicates that this reference refers to malicious references and, accordingly, the web resource located under this reference refers to malicious web resources.

If, however, the analyzing module 100 has established or discovered that the analyzed reference does not at least partially coincide with any of the known malicious references, then it additionally performs at least one of the following operations, wherein it: 1) analyzes the

domain name for the analyzed reference for harmfulness using at least one domain name analysis method known to it; 2) obtains or downloads at least one file located under the analyzed references, followed by its analysis for maliciousness using at least one file analysis method known to it; and 3) obtains the html-code of the web resource located under the analyzed reference, followed by its analysis for harmfulness using at least one html-code analysis method known to it.

When analyzing a domain name of any analyzed reference for maliciousness, the analyzing module 100 (i) gets access to local data storage 20 (separate local data storage or remote data storage, depending on the embodiment, as described earlier in this document) or communicates with it using the communication bus 30 to ensure that data about known malicious domain names is obtained from it, (ii) establishes or detects, by character-by-character comparison of each analyzed domain name with known malicious domain names from the obtained data, the fact that this analyzed domain name is at least partially concurrent with one of the known malicious domain names. If the analyzing module 100 found or discovered that the analyzed domain name does not at least partially coincide with any of the known malicious domain names, then it can additionally apply to such an analyzed domain name at least one of the methods of domain name analysis for suspiciousness known to it, for example, a domain name analysis method based on its length (the longer a domain name is, the more suspicious it is), a domain name analysis method based on its entropy (wherein, the higher information entropy calculated for a particular domain name of the well-known Shannon formula, the more suspicious the domain name is), a method for analyzing a domain name based on its meaningfulness and / or analyzing technique of the domain name based on the correctness of its spelling. As an example, when the analyzing module 100 analyzes a domain name for maliciousness using a domain name analysis method based on correctness of its spelling, it performs at least the following operations, wherein it: (i) communicates with the local data storage 20 (by a separate local or remote data storage depending on the embodiment, as described earlier in this document) to retrieve language dictionary data from it, (ii) extracts at least one word from each of the obtained domain names, (iii) determines the Levenshtein distance between each of the specified extracted words and one of the corresponding words in the language dictionaries of the specified obtained data, and (iv) compares a certain Levenshtein distance with a specified threshold value, for which a constant equal to two (2) may be used, ensuring that the analyzed domain name is classified as the malicious domain names, if a certain Levenshtein distance exceeds a specified threshold value, equal to two (2), for example.

Thus, if the analyzing module 100 has established or discovered, through at least one of the above described analysis methods, that the domain name for a particular analyzed reference belongs to malicious domain names, this indicates that this reference refers to malicious

references and, accordingly, a web resource located under this reference is a malicious web resource.

When analyzing a file located under the analyzed reference for maliciousness, the analyzing module 100 performs at least the following operations, wherein it: (i) obtains a file located under the analyzed reference; (ii) calculates the hash sum of the resulting file; (iii) gains access to local data storage 20 (separate local data storage or remote data storage, depending on the embodiment, as described earlier in this document) or establishes communication with it using communication bus 30 to ensure that data is received from it on hash-sums of known malicious files; (iv) establishes, by comparing the calculated hash-sum of the file with the hash-sums of known malicious files from the specified data, the fact that the calculated hash-sum of the file matches one of the hash-sums of known malicious files.

Thus, if the analyzing module 100 established or discovered that the hash sum of a particular file matches one of the hash sums of known malicious files, then this file belongs to malicious files, which indicates that this reference belongs to malicious links and, accordingly, the web resource located under this reference belongs to malicious web resources.

If the analyzing module 100 found or discovered that the hash sum of the received file does not match any of the hash sums of known malicious files, then it can additionally apply to such a received file at least one of the methods of file analysis for suspiciousness known to it, for example, a method of file analysis for suspiciousness based on a change in the state parameters of virtual machines, wherein the analyzing module 100 performs at least the following operations, wherein it: (i) launches every file received on at least one virtual machine characterized by a given set of state parameters, (ii) records changes in a given set of state parameters of at least one specified virtual machine for a given period of time, (iii) analyzes the obtained state change parameters using a specified set of analysis rules to ensure that the specified launched file is classified as malicious files, if the analyzed changes in the state parameters are typical for malicious files.

Thus, if the analyzing module 100 has established or discovered, using at least one of the above-described analysis methods, that the file located under a specific reference belongs to malicious files, this indicates that this reference belongs to malicious links and, accordingly, a web resource located under this reference belongs to malicious web resources.

When analyzing the html-code of a web resource located under the analyzed reference for harmfulness, the analyzing module 100 performs at least following operations, wherein it: (i) loads the html-code of the web resource located under this reference; (ii) analyzes downloaded html-code for maliciousness using at least one of html-code analysis methods known to it, for example, methods for analyzing html-code based on keywords indicating the harmful nature of a web resource. In addition, when analyzing downloaded html-code for maliciousness, the analyzing module 100 can also download all images and / or other files associated with a web

resource, for example, graphic design elements (*.JPG, *.PNG, etc.), style sheets (*.css), JS scripts, etc., based on the lists of such images and / or other files obtained by the analyzing module 100 from the extracted html-code, ensuring verification of the so-called screen-signatures, i.e. search for similar images and analysis of related web resources, wherein the search for similar images may be performed, e.g., using techniques of similar image search on the basis of a well-known method of search for the nearest neighbors. During such a search, the analyzing module 100 determines whether, for example, the images placed on the analyzed web resource correspond to the domain name and registration data of the web resource, wherein the analyzing module 100 can also additionally calculate the hash sums of all the images present on the analyzed web resource, and determine whether each calculated image hash sum matches one of the hash sums of known malicious elements that can be stored, for example, in the local data storage 20. In addition, the analyzing module 100 can additionally check the so-called resource signatures, for which it can calculate the hash sums of all previously loaded resources of the analyzed web resource, such as images, cascading style sheets (CSS), JS files, fonts, etc. and to establish or determine whether each calculated hash sum of a resource matches one of the hash sums of known malicious resources that can be stored, for example, in the local data storage 20.

Thus, if the analyzing module 100 has established or discovered, by means of at least one of the html-code analysis methods described above, that the web resource located under a specific reference comprises malicious content, this indicates that this reference belongs to malicious references and, accordingly, the web resource located under this reference, belongs to malicious web resources.

The analyzing module 100 is also configured to save information about each malicious web resource detected or installed using at least one of the above-described methods for analyzing web resources for maliciousness in a database of malicious web resources stored in the local data storage 20 (a separate local data storage of interconnected malicious web resources to which the analyzing module 100 can access or with which it can communicate using the communication bus 30 , or in the isolated remote data storage of interconnected malicious web resources, which the analyzing module 100 can access or communicate with using the communication module 10 connected to the analyzing module 100 via the communication bus 30, depending on the embodiment of this technique).

The analyzing module 100 is also configured to establish or identify web resources associated with each of the malicious web resources detected in the analyzing module 100 using at least one of the above-described methods for analyzing web resources for maliciousness.

In order to identify the web resources associated with each of the identified malicious web resources, the analyzing module 100 (i) gains access to the local data storage 20 (separate local data storage or remote data storage, depending on the embodiment, as described earlier in this document) or establishes communication with it using the communication bus 30, ensuring

that all other saved references to web resources are obtained from it; (ii) establishes a possible reference between each malicious link that comprises the corresponding identified malicious web resource and each of the received reference; and (iii) in the case of establishing this connection between the references, combines the web resources located under these connected references into a group of interrelated web resources. It shall be noted that each such group of interconnected web resources is formed from one malicious web resource and at least one associated web resource, considered as a potentially harmful web resource.

To establish the said link between the references, the analyzing module 100 performs at least one of the following operations, wherein it establishes at least one of the following for each pair of compared references: (1) whether domain names have a similar spelling (for example, by comparing them character-by-character, calculating the Levenshtein distance between domain names, comparing their hash sums calculated by the analyzing module 100, and / or another well-known technique); (2) whether domain names are registered to the same person; (3) whether the same personal data of the registrant, that is, the individual or legal entity to which the domain names are registered, in particular, the telephone number, actual address and / or email address, are indicated for the registered domain names; (4) whether domain names are located at the same IP address; and (5) whether the references have the same or a similar single web resource pointer "URL" (for example, by comparing them character-by-character, calculating the Levenshtein distance between these "URLs", comparing their hash sums calculated by the analyzing module 100, and / or another well-known methodology), for example www.site.com and www.sile.com, with information about the persons to whom the domain names are registered, information about the registrant's personal data (included in the domain name registration data) specified for registered domain names and IP addresses at which the registered domain names are located, can be automatically retrieved by analyzing module 100 using, for example, the online service Whois, in particular by automatically sending a suitable search query to the online service Whois and extracting the necessary information from the response of the online service Whois or from a web page with the results of a search query by using, for example, a special parser embedded in the analyzing module 100 and analyzing, for example, the text of the response of the online service Whois or the html code of the specified web page.

According to one of the embodiments of this technique, as an addition or alternative, the above reference relationship can also be established by analyzing module 100 by comparing for each pair of compared references the history of changes in IP addresses, operating services, history of domain names, history of DNS servers, history of changes in DNS records, SSL keys, SSH prints, executable files and other parameters of web resources. It shall be noted that the existence of a connection between the compared references can be established or determined by the analyzing module 100 based on the concurrence of at least one of the above parameters of

the web resources. In particular, in one of the embodiments of this technique, the connection between web resources located under the analyzed references can be established by the analyzing module 100 by creating a well-known mathematical model in the form of a graph, wherein the vertices of the created graph correspond at least to the first web resource, resource
5 and at least to the second web resource, and the graph edges represent the links between at least the first web resource and at least the second web resource under at least one the parameters of the above parameters, common for at least the first web resource and for at least the second web resource. In such embodiment of this technique described above, the analyzing module 100 may be configured to assign, by, for example, a well-known machine learning algorithm, weights to
10 connections between at least the first web resource and the second web resource based on the parameter of the first web resource and second web resource, wherein the number of links under a single web resource parameter between one first web resource and second web resources can be limited by a threshold value. The analyzing module 100 is additionally configured to determine a link coefficient as a ratio of the number of links under one parameter between one
15 first web resource and second web resources and the weight of each link under one parameter between the first web resource and the second web resources and configured to delete links between at least the first web resource and at least the second web resource if the value of a certain communication coefficient is less than a predetermined threshold value.

The analyzing module 100 is further configured to analyze the maliciousness of each of
20 the potentially harmful web resources in each formed group of interconnected web resources to identify malicious web resources among these potentially harmful web resources by implementing at least one of the above described analysis methods of the web resources for maliciousness.

If the malicious nature of at least one of the above potentially harmful web resources in
25 a specific group of interconnected web resources is confirmed, the analyzing module 100 stores information about each of these interconnected malicious web resources into the above described malicious web resources database, wherein the data stored for each malicious web resource, comprise, among other things, data indicating that this malicious web resource is associated with at least one other malicious web resource.

It shall be noted that when each new reference to a web resource is received, the
30 analyzing module 100 additionally checks whether the web resource located under this received reference belongs to malicious web resources, for which this analyzing module 100 (i) gets access to the above described a database of malicious web resources to retrieve information about detected malicious web resources from it; (ii) searches for this analyzed web resource
35 among the detected malicious web resources of the obtained data by character-by-character comparison of the reference that comprises the web resource being analyzed with each of the references under which these identified malicious web resources are located to determine

whether they at least partially coincide. Thus, if for the received new reference it was established that it at least partially coincides with one of the references under which previously detected malicious web resources are located, then the analyzing module 100 classifies the web resource located under this new reference as pertaining to the malicious web resources.

5 Otherwise, that is, when the new received reference does not even have a partial match with any of the references, under which the previously detected malicious web resources are located, in relation to the web resource located under this new reference, the above analysis for maliciousness is carried out.

The analyzing module 100 is also designed to classify or identify the type of threat each
10 of the identified malicious web resources carries, depending on the malicious content of this malicious web resource detected using at least one of the above web resource analysis methods for maliciousness (each type of threat corresponds to one or another characteristic malicious element, for example, text inviting the user to perform an action, a file of a certain format , scripts, replaced logos, etc.). For example, the analyzing module 100 may identify that a
15 particular malicious web resource is related to threats like phishing, malicious code, fraud, bot-no, and / or the like. Thus, for each of the detected malicious web resources, the analyzing module 100 is additionally configured to store data on the type of threat that this malicious web resource carries in the above-described malicious web resources database, and this stored data on the type of threat will be associated with a specific malicious web resource.

20 In addition, for each detected malicious web resource, the analyzing module 100 is configured to store evidence or grounds obtained by using at least one of the above described methods for analyzing web resources for maliciousness in this web database. during its analysis and which allowed to classify this analyzed web resource as malicious web resources, and such stored evidence or grounds for web resource maliciousness will be associated with a specific
25 malicious web resource.

In addition, the analyzing module 100 is configured to establish or identify for each of the detected malicious web resources, information about which is stored in the above-described base of malicious web resources, at least one authorized entity associated with this malicious web resource. Authorized entities associated with each of the identified malicious web resources
30 can be the administrator of this malicious web resource, the owner of this malicious web resource, the domain name registrar, the hosting provider and / or other known individuals and entities that can block the operation of this malicious web resource or influence the decision to block or suspend the operation of this malicious web resource.

In order to identify authorized entities associated with each of the identified malicious
35 web resources, the analyzing module 100 is pre-configured or programmed to determine at least one of the owner, administrator, hosting provider and / or domain name registrar associated with

this malicious web resource, as well as their contact details, such as, for example, the actual address, contact telephone number, e-mail address, etc.

It shall be noted that the said authorized entities, established or detected by analyzing module 100, can be determined using any of the known online services, for example the online Whois service, and / or any of the known utilities, such as, for example, the utility “nslookup”, based on, for example, the domain name used to form a search query. It shall also be noted that the necessary contact details of at least some of the required authorized entities can also be obtained using any of these well-known online services and / or any of these well-known utilities, since they are included in the domain name registration data specified for registered domain names in these services and / or utilities. In particular, in any of the well-known online services and / or any of the known utilities, contact details of the owner of a particular web resource can be obtained, namely, his/ her contact phone number, the actual address of his/ her place of residence and / or his/ her email address, as well as (if available) the contact details of the administrator of this web resource, namely his/ her contact phone number, the actual address of his/ her place of residence and / or his/ her email address.

Thus, to determine the owner, administrator, hosting provider and / or registrar of domain names associated with a specific malicious web resource, and to obtain contact information of the owner and / or administrator of this malicious web resource, the analyzing module 100 is configured to automatically send, for example, to the online service “Whois” a suitable search query, formed on the basis of the domain name, extracted by the analyzing module 100 from the reference under which this malicious web resource is located, and with the possibility to automatically extract the necessary information from the response of this online Whois service or from a web page with the results of a search query by using, for example, a special parser embedded in the analyzing module 100 and analyzing, for example, the text of the response of the online Whois service or html-code of the specified web page. Thus, from the information received from any of the known online services and / or any of the known utilities, the analyzing module 100 can uniquely determine the owner and administrator of domain names for each of the detected malicious web resources, as well as contact information of each of them, and to establish the names of the domain name registrar and hosting provider associated with this malicious web resource.

In the local data storage 20, an updated database of authorized entities is pre-stored for storing information about known authorized entities, in particular, a list of known domain name registrars and their contact information, a list of known hosting providers and their contact information, and a list of the state institutions that can influence the decision to block or suspend functioning of the malicious web resources, etc., and their contact information, wherein the contact details in this database of authorized entities are set in accordance with the specific authorized entity to which they relate. The analyzing module 100 is configured to access the

local data storage 20 or to communicate with it using the communication bus 30, ensuring that at least one of the interested entities associated with a specific malicious web resource is retrieved from the database of authorized entities, based on the names of these entities of interest previously installed by the analyzing module 100 using any of the known online services and / or any of the known utilities, as described in more details earlier in this document. Thus, the analyzing module 100 retrieves from the database of authorized entities the contact information of the domain name registrar and / or hosting provider previously established by the analyzing module 100 for the detected malicious web resource using any of the known online services and / or any of the known utilities.

For each of the detected malicious web resources, the analyzing module 100 is additionally configured to store the names of authorized entities associated with this identified malicious web resource and the contact data of these authorized entities in the malicious web resources database described above. Thus, for each of the detected malicious web resources, the analyzing module 100 stores in the base of malicious web resources the name of the owner of this web resource and his/her contact data, the name of the administrator of this web resource and his/her contact details, the name of the domain name registrar for this web -resource and his/her contact details and / or the name of the hosting provider for this web resource and its contact information, wherein each contact details in the database of malicious web resources are associated with a specific authorized entity from the above authorized entities to which they relate, and with a specific malicious web resource, with which the authorized entities are associated.

In one of the embodiments of this technique, for each detected malicious web resource, the analyzing module 100 may be further configured to gain access to local data storage 20 (separate local data storage or remote data storage, depending on the embodiment, as described above in this document) or configured to communicate with it using the communication bus 30 to ensure that the database of malicious web resources comprises information about authorized entities associated with this malicious web resource, that is, for example, the name of the owner of this malicious web resource and his/her contact details, the administrator's name of the malicious web resource and his/her contact details, the name of the domain name registrar for this malicious web resource, and its contact details and / or the name of the hosting provider for this malicious web resource and its contact details. If the analyzing module 100 determines that the database of malicious web resources already comprises all necessary information about authorized entities associated with this malicious web resource, or at least a part of such necessary information, then the analyzing module 100 does not perform the above described operations related to the direction of search queries to well-known online services and / or well-known utilities and receiving an access to the database of authorized entities, and immediately begins the process described below of formation of at least one report of at least one of the

authorized entities associated with this malicious web resource on the basis of specified information on authorized entities from the database of malicious web-resources.

In another embodiment of this technique, in which the computing device 200 receives, via the communication module 10, the references to known malicious web resources from at least one reference source having a unique identifier, by which the analyzing module 100 determines that the received data streams from the specified at least one references source comprise references to web resources with malicious and / or illegal content, the analyzing module 100 may not perform the above analysis of such received references for maliciousness, and may immediately send a search query to the above malicious web resource database to determine whether this database comprises information about authorized entities associated with a malicious web resource located under a received reference, and then generating at least one report under at least one authorized entity associated with this malicious web resource, based on the specified information about authorized entities from the malicious web resource base, as described in more details below. Otherwise, that is, in the absence of information about authorized entities associated with a malicious web resource in the base of malicious web resources, the analyzing module 100 performs the above described operations related to sending search queries to well-known online services and / or known utilities and obtaining access to the database of authorized entities, followed by the formation of at least one report under at least one authorized entity associated with this malicious web resource, based on specified information about authorized entities from the base of malicious web resources, as described in more detail below.

It shall be noted that a predefined set of report templates is pre-stored in the local data storage 20, with each report template essentially being a pre-composed letter of appeal informing a specific authorized entity about the malicious nature of at least one specific web resource asking for a decision about blocking or suspending the operation of the specified at least one malicious web resource, or influencing an adoption of such a decision, wherein each template from this set of report templates is set up to comply with or associated with one of the known types of threats that may be carried by malicious web resources, and one of the authorized entities. Thus, for each known authorized entity, several report templates can be stored in the local data storage 20, each pre-compiled according to only one type of threat from known types of threats.

The analyzing module 100 is also configured with an possibility to generate at least one report for at least one authorized entity after a predetermined period of time (for example, every 10 minutes, once every half hour, every hour, every few hours, once a day, once a week, etc.) or essentially in real time based on the following information:

- data on at least one of the malicious web resources associated with one of the specified authorized entities and extracted by the analyzing module 100 from the above described base of malicious web resources, at least based on the name of this authorized entity, and

- a specific report template corresponding to one of the specified authorized entities and one of the types of threats identified by the analyzing module 100 for the specified malicious web resources, and extracted by it from the malicious web resources base, at least based on information about the specified malicious web resources, in particular, the unique identifier of each of these malicious web resources.

Thus, the analyzing module 100 can, for example, generate one report for one of the well-known hosting providers and one of the well-known domain name registrars, wherein each such report can include information about several specific malicious web resources at once (if these web resources are a threat of the same type, for example, a phishing threat, and are associated respectively with the same hosting provider or domain name registrar), and also specific information on one malicious web resource (if it carries a threat of a type other than other malicious web resources, and / or is associated respectively with a hosting provider or domain name registrar different from other malicious web resources). As an addition or alternative, the analyzing module 100 may, for example, generate one report for each of the web resource administrators associated with malicious web resources, information about which was included in the above-described report for the hosting provider and the report for the domain name registrar, at the same time, each report can include information about several specific malicious web resources at once (in case these web resources pose a threat of the same type, for example, a type of “fraud”, associated respectively with the same administrator) and information about only one specific malicious web resource (in case it carries a threat of a type different than other malicious web resources, and / or associated respectively with an administrator different from other malicious web resources).

It shall also be additionally noted that the number of reports generated by analyzing module 100 for each of the above authorized entities for malicious web resources associated with this authorized entity, in an amount received over a specified period of time, will correspond to the number of types of threats that carry these malicious web resources.

There is a possible embodiment in which the analyzing module 100 for each of the malicious web resources will generate reports for each of the authorized entities associated with this malicious web resource, in real time, immediately after establishing the fact that the web resource located under the accepted reference belongs to malicious web resources that carry a specific type of threat, as described in more detail earlier in this document.

In one of the embodiments of this technique, the analyzing module 100 may also add to at least one of the reports generated by the analyzing module 100 for authorized entities, evidence of the harmfulness of each web resource that was included in this report, wherein the

analyzing module 100 may obtain all the necessary evidence from the base of malicious web resources, in which they are set in accordance with a specific malicious web resource.

Analyzing module 100 is also configured to send each above-described generated report to the appropriate authorized entity on the basis of the contact information of this authorized entity, received by the analyzing module 100 from the malicious web resources base, to inform this authorized entity of at least one web resource with malicious and / or illegal content.

According to one of the embodiments of this technique, at least part of the above-described functionality of the analyzing module 100 can be implemented using at least one separate functional unit or module, which can be necessarily connected configured to exchange data with the analyzing module 100 and with each other.

As an example, in one of the embodiments of this technique, the above-described analyzing module 100 may be configured to perform exclusively the above-described operation of detecting malicious web resources in a plurality of web resources located under the received references. The computing device 200 may additionally comprise, for example, a separate module for identifying the interconnected web resources connected to the analyzing module 100 configured to exchange data and configured to perform the above described establishment of the web resources associated with each of the malicious web resources detected by the analyzing module 100, and a separate module for informing about malicious web resources, connected to the module for identifying interconnected web resources and analyzing module 100 with the possibility of exchange with them data and performed configured to perform the above operation to establish at least one authorized entity associated with each of the malicious web resources detected by the analyzing module 100 and / or the module for identifying the interconnected web resources, as well as the above described operation for generating at least one report for at least one of the established authorized entities based on information about the detected malicious web resources associated with this authorized entity and the above operation of sending each generated report to the appropriate authorized entity on the basis of the contact details of this authorized entity. It shall be noted that in such embodiment of this technique, the analyzing module 100 may be configured to exchange data with the communication module 10 and the local data storage 20 using the communication bus 30, the interconnected web resource detection module may be configured to exchange data with the local data storage 20 using the communication bus 30 and the module for informing about malicious web resources can be configured to exchange data with the local data storage 20 using communication bus 30.

As another example of another embodiment of this technique, the above-described analysis module 100 may be configured to perform exclusively the above-described operation of detecting malicious web resources in a plurality of web resources located under the received references. The computing device 200 may additionally comprise, for example, a separate module for identifying the interconnected web resources connected to the analyzing module 100

configured to exchange data and configured to perform the above described establishment of the web resources associated with each of the malicious web resources detected by the analyzing module 100, as well as a separate module for the establishment of authorized entities, connected to the module for identifying interconnected web resources and the analyzing module 100 exchanging data with them and configured to perform the above operation to establish at least one authorized entity associated with each of the malicious web resources detected by analyzing module 100 and / or module for identifying interconnected web resources and a separate report generation module connected to the module of establishment of authorized entities with an ability to exchange data and configured to perform the above operation of generating at least one report for at least one of the authorized entities based on data of identified malicious web resources associated with this authorized entity, and the above-described operation of sending each generated report to the relevant authorized entity on the basis of the contact information of such authorized entity. It shall be noted that in such embodiment of this technique, the analyzing module 100 may be configured to exchange data with the communication module 10 and the local data storage 20 using the communication bus 30, and each module for identifying related web resources, the module for establishing authorized entities and the reporting module may be configured to exchange data with the local data storage 20 using the communication bus 30.

As another example, in yet another embodiment of this technique, the above described analyzing module 100 may be configured to perform the above-described operation of detecting malicious web resources in a plurality of web resources located under the received references, as well as performing the above-described operation of establishing web resources associated with each of the identified malicious web resources. Computing device 200 may additionally comprise, for example, a separate module for establishing authorized entities, connected to the analyzing module 100 configured to exchange data and configured to perform the above-described operation of establishing at least one authorized entity associated with each of the malicious web resources detected by the analyzing module 100, and a separate reporting module connected to the module for establishing authorized entities configured to exchange data and configured to perform the above described operation of generating at least one report for at least one of the established authorized entities based on information about detected malicious web resources associated with this authorized entity, as well as the above operation of sending each generated report to the appropriate authorized entity based on the contact details of this authorized entity. It shall be noted that in such embodiment of this technique, the analyzing module 100 may be configured to exchange data with the communication module 10 and the local data storage 20 using the communication bus 30, and each of the authorized entities establishment module and the report generating module can be performed configured to exchange data with the local data storage 20 using the communication bus 30.

According to another embodiment of this technique, the analyzing module 100 may be formed from at least one submodule configured to implement at least part of the above described functionality of the analyzing module 100, wherein such functional submodules in the analyzing module 100 can be connected to each other as necessary configured to exchange data.

5 As an example, in one of the embodiments of this technique, the above-described analyzing module 100 may be formed from a sub-module for detecting malicious web resources made with the possibility of performing the above-described operation of detecting malicious web resources in a variety of web resources located under the received references, submodule for detection of interconnected web resources connected with a submodule for detection of
10 malicious web resources configured to exchange data and executed configured to perform the above operation of establishment of the web resources associated with each of the malicious web resources identified by the malicious web resources detection submodule, as well as the submodule for identifying authorized entities connected to the submodule for detecting interconnected web resources and data exchange with them and performed configured to
15 perform the above operation to establish at least one authorized entity associated with each of the malicious web resources detected by the submodule for detecting malicious web resources and / or submodule for detecting interconnected web resources, and a submodule for generating reports connected to the submodule for identifying authorized entities with possibility of exchanging data, and configured to perform the above described operation of generating at least
20 one report for at least one of the established authorized entities based on information about detected malicious web resources associated with this authorized entity, and the above operation of sending each generated report to the appropriate authorized entity on the basis of the contact information of the authorized entity. It shall be noted that in such embodiment of this technique, the submodule for detecting malicious web resources can be configured to exchange data with
25 communication module 10 and local data storage 20 using communication bus 30, and each sub-module for detecting interconnected web resources, sub-module establishing authorized entities and the sub-module generating reports can be configured to exchange data with the local data storage 20 using the communication bus 30.

As an example, in one of the embodiments of this technique, the above-described
30 analyzing module 100 may be formed from a sub-module for detecting malicious web resources made with the possibility of performing the above-described operation of detecting malicious web resources in a variety of web resources located under the received references, submodule of detection of interconnected web resources, connected to a submodule for detecting malicious web resources configured to exchange data and configured to perform the above operation of
35 detection of the web resources associated with each of the malicious web resources identified by the malicious web resources detection submodule, as well as the malicious web resources information submodule connected to the interconnected web resources submodule and the

malicious web resources submodule with the possibility of exchanging data with them and configured to perform the above operation to establish at least one authorized entity associated with each of the malicious web resources identified by the submodule for detecting malicious web resources and / or submodules for identifying interconnected web resources, as well as the above described operation of generating at least one report for at least one of the established authorized entities based on the detected malicious web resources associated with this authorized entity, and the above operation of sending each generated report to the appropriate authorized entity based on the contact details of this authorized entity. It shall be noted that in such embodiment of this technique, the submodule for detecting malicious web resources may be configured to exchange data with the communication module 10 and the local data storage 20 using the communication bus 30, and each of the submodules for detecting interconnected web resources and submodule for informing about malicious web resources can be configured to exchange data with the local data storage 20 using the communication bus 30.

As another example, in yet another variation of such embodiment of the this technique, the above-described analysis module 100 may be formed from a submodule for detecting malicious web resources, configured to perform the above-described operation to detect malicious web resources in a variety of web resources located under received references, as well as the implementation of the above operation of establishing web resources associated with each of the identified malicious web resources, as well as a submodule for detection of the authorized entities connected to a submodule for detection of malicious web resources configured to exchange data and configured to perform the above operation to establish at least one authorized entity associated with each of the malicious web resources detected by the submodule for detecting malicious web resources and a submodule for generating reports, connected with a submodule for establishing authorized entities configured to exchange data and configured to perform the above operation of generating at least one report for at least one of the established authorized entities based on information about detected malicious web resources associated with this authorized entity, as well as the above operation of sending each generated report to the appropriate authorized entity based on the contact details of this authorized entity. It shall be noted that in such embodiment of this technique, the submodule for detecting malicious web resources can be configured to exchange data with communication module 10 and local data storage 20 using communication bus 30, and each of the submodules establishing authorized entities and the reporting module can be configured to exchange data with the local data storage 20 using the communication bus 30.

Fig. 3 shows a flowchart of a method 400 for informing about the malicious nature of a web resource according to this technique. It shall be noted that the method 400 can be performed using the computing processor of any known computing device, in particular using

the above-described analyzing module 100 of the computing device 200 to inform about the malicious nature of the web resources shown in Fig. 2.

Method 400 shown on Fig. 3 begins with stage 410, under which the links to a plurality of web resources are obtained.

5 In one of embodiments of this technique, in order to obtain links to a plurality of web resources at step 410, at least one of the following operations shall be performed, wherein: (1) sending a request to at least one source of references to obtain from it at least one reference to a web resource; (2) receiving messages from at least one computing device ensuring their processing to retrieve at least one web resource; (3) receiving messages from at least one mobile
10 device ensuring their processing to retrieve at least one reference to a web resource; and (4) entering search queries in at least one search engine using a specific list of keywords to identify contextual advertising in the search results received in response to each search query in each of these search engines, ensuring the extraction of at least one reference to a web resource from the identified contextual advertising.

15 Subsequently, method 400 proceeds to execution of step 420, wherein malicious web resources are detected in the indicated set of web resources, and then to execution of step 430, wherein web resources associated with each of the malicious web resources identified in the above step 420 are determined.

In one of the embodiments of this technique, at least one of the following is determined
20 in order to establish related web resources at step 430: (i) whether the domain names of the web resources have a similar spelling; (ii) whether the domain names are registered to the same person; (iii) whether the same personal data of the registrant, that is, the individual or legal entity to which the domain names are registered, is indicated for the registered domain names of the web resources; (iv) whether the domain names of the web resources are located at the same
25 IP address; and (v) whether the links corresponding to the web resources have the same or similar single web resource index "URL" (for example, www.site.com and www.sile.com).

In another embodiment of this technique, to establish a link to web resources, at step 430, at least the following operations are performed, wherein: (i) creating a mathematical model in the form of a graph, wherein the vertices of the created graph correspond to at least the first
30 web resource and at least the second web resource, and the graph edges are links between at least the first web resource and the at least second web resource by at least one web resource parameter common at least for the first web resource and at least for the second web resource, wherein the number of links per parameter of the web resource between the first web resource and the second web resources is limited by a specified threshold value; (ii) assigning, by a
35 known machine learning algorithm, weights to the links between at least the first web resource and the second web resource based on the parameter of the first web resource and the second web resource; (iii) determining the link coefficient as the ratio of the number of links for one

parameter of a web resource between one first web resource and second web resources and the weights of each link for one parameter of a web resource between the first web resource and second web resources; and (iv) removing links between at least the first web resource and the at least second web resource in case the value of a certain communication coefficient is less than a predetermined threshold value.

Subsequently, method 400 proceeds to execution of step 440, wherein malicious web resources are detected in the associated web resources established in step 430.

In some embodiments of this technique, to identify malicious web resources, at step 420 or step 440, it is established whether each resulting reference is at least partially related to one of the known malicious references.

In other embodiments of this technique, in order to detect malicious web resources at step 420 or step 440, at least one of the following operations is performed in addition to the operation wherein each received link at least partially coincides with one of the known malicious links. wherein: (1) analyzing domain name of a web resource for maliciousness using at least one method of analyzing domain names; (2) receiving at least one file from a web resource for analyzing its maliciousness using at least one file analysis method; and (3) receiving html-code of the web resource for analyzing its maliciousness using at least one method of analyzing the html-code.

In some other embodiments of this technique, when analyzing the domain name of a web resource for maliciousness during the execution of operation (1) within the framework of execution of step 420 or step 440, it is further established whether this analyzed domain name matches one of the known malicious domain names.

In other embodiments of this technique, when analyzing a file received from a web resource, when performing operation (2) as part of executing step 420 or step 440, the hash sum of the analyzed file obtained from the web resource is additionally calculated and it is determined whether the calculated hash sum of the analyzed file coincides with the hash sum of one of the known malicious files.

In other embodiments of this technique, when analyzing the received html-code of a web resource, when performing operation (3) as part of executing step 420 or step 440, further search is conducted in the specified html-code for specific keywords indicating the malicious nature of the web resource.

Subsequently, method 400 proceeds to execution of step 450, wherein at least one authorized entity is established, associated with each of the malicious web resources detected at step 420 and / or step 440.

In one of the embodiments of this technique, when establishing authorized entities associated with each of the detected malicious web resources, at step 450, the owner, administrator, hosting provider and / or domain name registrar associated with this malicious

web resource is determined. In another embodiment of this technique, for the owner of a malicious web resource, determined at stage 450 when establishing authorized entities associated with each of the identified malicious web resources, a request is also sent to the hosting provider and / or domain name registrar, also determined at stage 450 under identification of the authorized entities associated with each of the identified malicious web resources and associated with this malicious web resource, ensuring receipt of that additional links to the web resources, associated with the specified owner.

Subsequently, method 400 proceeds to execution of step 460, wherein at least one report is generated for at least one of the authorized entities established at step 450, based on information about detected malicious web resources associated with this authorized entity.

In one of the embodiments of this technique, the method 400 may include an additional step, wherein a threat type is set from a predetermined set of threat types for each malicious web resource detected at step 420 and / or step 440, and when generating each report, a template is used from a specified set of report templates, with each template corresponding to one of the identified types of threats and one of the established authorized entities.

In another embodiment of this technique, the number of reports generated for each authorized entity at the stage 460 may correspond to the number of identified types of threats.

In another embodiment of this technique, evidence of maliciousness of each web resource, the details of which are comprised in this report, may be added to each report generated at step 460.

Subsequently, method 400 proceeds to perform the final step 470, wherein each report generated at step 460 is sent to the appropriate authorized entity on the basis of the contact details of this authorized entity.

It shall be noted that the claimed method 400 improves the efficiency of informing authorized entities about the identified web resources with malicious and / or illegal content both by expanding the circle of authorized entities receiving such reports and by improving the informational representativeness of each report that can immediately cover the entire group of malicious web resources that are involved by the abusers and carry the same type of threat.

The presented illustrative embodiments, examples and description are merely designed to provide an understanding of the proposed technical solution and are not restrictive. Other possible embodiments will be clear to the specialist from the above description. The scope of this technique is limited only by the attached claims.

CLAIMS

1. A method of informing about malicious web resources running on a computing device, the method comprising:

5 obtaining references to a set of web resources;
 identifying malicious web resources in the set of web resources;
 establishing web resources that are associated with each of the identified malicious web resources;

 identifying additional malicious web resources in the established associated web
10 resources;

 determining at least one authorized entity associated with each of the identified malicious web resources and the identified additional malicious web resources;

 generating at least one report for at least one of the determined authorized entities based on the data on the identified malicious web resources and the identified additional malicious
15 web resources associated with this authorized entity; and

 sending each generated report to the determined authorized entity on the basis of contact details of the authorized entity.

2. The method of claim 1, wherein determining the at least one authorized entity associated with
20 each of the identified malicious web resources and the identified additional malicious web resources comprises determining an owner, administrator, hosting provider and / or domain name registrar associated with the identified malicious web resource and the identified additional malicious web resources.

25 3. The method of claim 1, further comprising establishing a type of threat from a given set of types of threats for each of the identified malicious web resources and the identified additional malicious web resource; and

 generating the at least one report comprises using a template from a specified set of report templates, each template corresponding to one of the given set of types of threats and one
30 of the determined authorized entities.

4. The method of claim 3, wherein a number of reports generated for each authorized entity corresponds to a number of identified types of threats.

35 5. The method of any of claims 1-4, further comprising adding, to each report, evidence of maliciousness of the each web resource, information about which is comprised in this report.

6. The method of claim 1, wherein identifying the malicious web resources comprises establishing whether each obtained reference at least partially coincides with one of known malicious references.

5 7. The method of claim 6, wherein identifying the malicious web resources or the additional malicious web resources comprises carrying out at least one of the following operations:

- analyzing domain name of the web resource for harmfulness using at least one method for analyzing domain names;

- obtaining at least one file from a web resource and analyzing the at least one file for maliciousness using at least one method of analyzing files; and

- obtaining html-code of a web resource and analyzing the html-code for maliciousness using at least one method of analyzing html-code.

8. The method of claim 7, wherein analyzing the domain name of the web resource for maliciousness comprises establishing whether this analyzed domain name matches one of known malicious domain names.

9. The method of claim 7, wherein analyzing the at least one file received from the web resource for maliciousness comprises calculating a hash sum of the at least one file and determining whether the calculated hash sum of the analyzed file matches a hash sum of one of known malicious files.

10. The method of claim 7, wherein analyzing the obtained html-code of the web resource comprises conducting a search in the obtained html-code for specific keywords indicating a malicious nature of the web resource.

11. The method of claim 1, wherein establishing the associated web resources comprises determining at least one of the following:

- whether domain names of the web resources have a similar spelling;

- whether domain names are registered to a same person;

- whether same registrant personal data are specified for registered domain names of the web resources;

- whether domain names of the web resources are located at a same IP address; and

- whether references corresponding to the web resources have a same or similar uniform pointer of the web resource "URL".

12. The method of claim 1, wherein establishing the associated web resources comprises performing at least the following operations:

- creating a mathematical model in the form of a graph, wherein vertices of the graph correspond to at least a first web resource and at least a second web resource, and graph edges are links between at least the first web resource and at least the second web resource at least under one web resource parameter, common for at least the first web resource and at least the second web resource, wherein a number of the links under the one web resource parameter between one first web resource and the second web resources is limited by a predetermined threshold value;

- assigning weights to the links between at least the first web resource and the second web resource by means of a known machine learning algorithm, based on a parameter of the first web resource and the second web resource;

- determining a link coefficient as a ratio of the number of the links under the one parameter of a web resource between one first web resource and the second web resources and the weight of each link for a single web resource parameter between the first web resource and the second web resources; and

- removing the links between at least the first web resource and at least the second web resource in case a value of a certain communication coefficient is less than a predetermined threshold value.

13. The method of claim 1, wherein obtaining the references to the set of web resources comprises performing at least one of the following operations:

- sending a request to at least one source of references to obtain from it at least one reference to a web resource;

- receiving messages from at least one computing device with a provision of their processing to retrieve at least one reference to a web resource;

- receiving messages from at least one mobile device with a provision of their processing to retrieve at least one reference to a web resource; and

- entering search queries in at least one search engine using a specific list of keywords to identify contextual advertising in search results received in response to each request in each of these search engines, ensuring an extraction of at least one reference to a web resource from the identified contextual advertising.

14. A computing device for informing about malicious web resources, comprising a memory for storing machine-readable instructions and at least one computing processor, configured to execute computer-readable instructions while ensuring implementation of the method for informing about malicious web resources according to any one of claims 1-13.