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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 February 2016 (04.02.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/016885 A1

(51) International Patent Classification:
G06F 21/00 (2013.01)

(21) International Application Number:
PCT/IL2015/050769

(22) International Filing Date:
27 July 2015 (27.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/029,576 28 July 2014 (28.07.2014) US

(71) Applicant: **B. G. NEGEV TECHNOLOGIES AND APPLICATIONS LTD., AT BEN-GURION UNIVERSITY** [IL/IL]; P.o.b. 653, 8410501 Beer Sheva (IL).

(72) Inventors: **SHABTAI, Asaf**; Hulda 2, 76842 Hulda (IL). **ELOVICI, Yuval**; 9 Moshav Arugot, 7986400 D.N. Lachish (IL).

(74) Agents: **LUZZATTO, Kfir** et al.; Luzzatto & Luzzatto, P.O. Box 5352, 8415202 Beer Sheva (IL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

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

[Continued on next page]

(54) Title: MISUSEABILITY ANALYSIS FOR IT INFRASTRUCTURE

(57) Abstract: The present invention relates to a method and a framework that automatically and dynamically derives a misuseability score for every IT component (e.g., PC, laptop, server, router, smartphone, and user or any other element that can be connected to organization network or to the internet). The dynamic framework of the present invention supports the risk analysis process. The misuseability score encapsulates the potential damage that can be caused to the organization in case that an asset is compromised and misused, for example, as part of a cyber-attack.

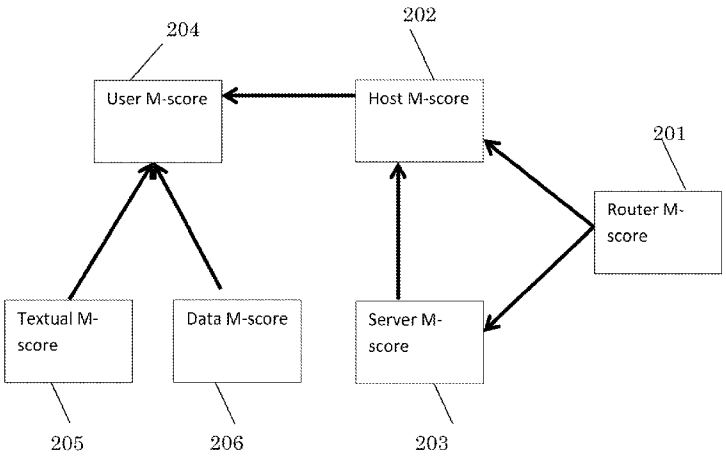

Fig. 2

WO 2016/016885 A1

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

MISUSEABILITY ANALYSIS FOR IT INFRASTRUCTURE

Field of the invention

The invention relates to the field of data security. More specifically the invention relates to misuseability analysis for IT infrastructure.

Background of the invention

Today, organizations are facing many cyber-attacks which make the detection of cyber-attacks a very difficult task. In addition, the organizations have limited resources that can be allocated to the detection of complex cyber-attacks. Some of the challenges are: the need to analyze massive amount of data that is collected from the information technology (IT) infrastructure; the highly advanced attacks that are in continuously becoming more and more sophisticated; the fact that attacks can be originated by an insider or by an external entity; the rapid introduction of new technologies that are integrated into the organization's infrastructure; and the variety and costs of security solutions.

Because of these challenges organizations must conduct a thorough risk analysis process in order to focus their efforts and resources on the protection of the highly critical assets. In the risk analysis process the organization identifies the most important assets by estimating the risk to the asset which is a function of the asset's value and the likelihood of the threat to be realized. This process is time consuming and therefore ignores the dynamic nature of the IT infrastructure. This means that the value of assets may change over time and will not be reflected by the risk analysis process results.

Harel et al. [Harel, A., Shabtai, A., Rokach, L., and Elovici, Y., 2012. M-score: A misuseability weight measure. IEEE Trans. on Dependable and Secure Computing, 9(3), 2012, 414-428] initially addressed this challenge and presented a new concept, Misuseability Weight, which assigns a sensitivity score to data, thereby estimating the level of harm that might be inflicted upon the organization when the data is

leaked. Assigning a misuseability weight to a given dataset is strongly related to the way the data is presented (e.g., tabular data, structured or free text) and is domain-specific. Harel et al. [Harel, A., Shabtai, A., Rokach, L., and Elovici, Y., 2012. M-score: A misuseability weight measure. IEEE Trans. on Dependable and Secure Computing, 9(3), 2012, 414-428] focus on mitigating leakage or misuse incidents of data stored in databases (i.e., tabular data) and presented the M-Score, a misuseability weight measure for tabular data.

Vartanian and Shabtai [Shabtai, A., Vartanian, A., 2014. TM-Score: A Misuseability Weight Measure for Textual Content", submitted to IEEE Trans. on Information Forensics and Security] proposes an extension to the misuseability weight concept and specifically focused on textual content. The main goal in Vartanian and Shabtai is to define a misuseability measure, termed TM-Score, for textual content. Using this measure it is possible to estimate the extent of damage that can be caused by an insider that is continuously and gradually exposed to documents. The extent of damage is determined by the amount, type and quality of information to which the insider is exposed. However, there are other IT elements that may be vulnerable and sensitive, except from insiders, such as servers and routers which are affected from different parameters (not necessarily the information to which an element is exposed to) which the prior art does not deal with.

Moreover, the necessity for a full and comprehensive framework that is able to derive a misuseability score for each IT element is rising in order to cope with the challenges of data security in the world of cyber-attacks.

One of the solutions to data security is given by data protection companies, which meets the employees of the organization, learn the roles of each employee and the work that is done by each employee, and then analyze the misuseability of each IT element. However, this solution depends on humans, it takes time and if there is a change in the organization such as a new IT element or a new job, the analysis has to be redone by the analyzer of the data protection company.

It is therefore an object of the present invention to provide a method and a framework that automatically and dynamically derives a misuseability score for every IT component, for supporting the risk analysis process.

Further objects and advantages of this invention will appear as the description proceeds.

Summary of the Invention

The present invention relates to a method for automatically and dynamically deriving a misuseability score for one or more IT element(s), comprising the steps of: collecting data about said one or more IT element(s); analyzing said collected data; deriving from said analyzed data a misuseability score for said one or more IT element; analyzing the connections between said one or more IT elements; and grouping together IT elements with high connection and deriving a misuseability score to said group.

In an embodiment of the invention, an IT element is: a PC, laptop, server, router, smartphone, or user.

In an embodiment of the invention, an IT element is an element which is connected to an organization network and/or to the internet.

In an embodiment of the invention, the misuseability score is derived according to:

$$\mathrm{MScore}(e,c) = \sum_{i=1}^{n} \alpha_i \cdot p_i$$

Wherein , p_i are parameters that are dynamically computed from the data that is collected from the IT element

α_i are constants which define the importance of each parameter;

Mscore(e,c) is the misuseability score;

c is the specific context for which the misuseability score is computed; and

e is the IT element for which the misuseability score is calculated for.

In an embodiment of the invention, the misuseability score for a router is derived from the parameters of: configuration, purpose and activity.

In an embodiment of the invention, the misuseability score for a server is derived from the parameters of: configuration, server purpose and activity.

In an embodiment of the invention, the misuseability score for an end user device is derived from the parameters of: configuration, purpose and activity.

In an embodiment of the invention, the misuseability score for a user is derived from the type of services and machines that said user accesses as well as the type of data and information that the user is exposed to.

In an embodiment of the invention, the misuseability score for a user is derived from general behavioral patterns of the user.

In an embodiment of the invention, the connection between one or more element(s) is derived from the amount of traffic or activity between said one or more element(s) and from the rate of activity and connection between said one or element(s).

Brief description of the drawings

- Fig. 1 schematically shows an example for the concept of the present invention; and

- Fig. 2 schematically summarizes the connections among the misuseability scores of the various IT components and presents the derivation tree of misuseability scores.

Detailed description of the embodiments of the invention

The present invention relates to a method and a framework that automatically and dynamically derives a misuseability score for every IT component (e.g., PC, laptop, server, router, smartphone, and user or any other element that can be connected to organization network or to the internet). The dynamic framework of the present invention supports the risk analysis process. The misuseability score encapsulates the potential damage that can be caused to the organization in case that an asset is compromised and misused, for example, as part of a cyber-attack.

In the present invention the misuseability concept is extended to a full and comprehensive framework that is able to derive a misuseabilty score for each IT element including personal computers, servers, smartphones, databases, routers, switches and users. In addition the method of the present invention also recognize groups of IT element which are highly connected and thus, the method and the framework of the present invention derives a misuseability score for such groups.

In addition the present invention also derives misuseability score for IoT networks (Internet of Things). IoT is in fact the network of all the things and element that are connected to the internet and or to the organization network, for example: smart watch, air conditioning system and cars. The present invention derives a misuseability score to every element that is connected and which may be a destination for attacks.

Fig. 1 schematically shows an example of the concept of the present invention. For each IT element (e.g., PC, laptop, server, router, smartphone, and user), a misuseability score is derived. For example, in group 10, for laptop 11, a

misuseability of 6 is derived, for user 12, a misuseability of 7 is derived and for smartphone 13, a misuseability of 3 is derived. In addition, in group 19 , for server 14 a misuseability of 8 is derived, for server 15, a misuseability of 9 is derived and for server 16, a misuseability of 5 is derived. The routers 17 and 18 connects between the IT element of group 10 and 19. Router 17 has a misuseability score of 7 and router 18 has a misuseability score of 5. The framework of th present invention recognizes the high connection between groups 10 an 19 and therefore marks groups 10 and 19 as a community 20. The framework of the present invention also derives an aggregative misuseability score for the community 20, which is based on the misuseability of each one of the elements in the community. By analyzing the connectivity level among the elements it is possible to identify clusters of elements that are not only highly connected but also have a high misuseability score as a group and therefore should be carefully analyzed and protected.

The misuseability score is a quantitative measure that is calculated for each IT element based on data collected from the element itself (using a dedicated agent) and /or externally - from the network traffic. The collected data may be any data and/or information that affects the vulnerability of the IT element, for example, the location of the IT element and the functionality of the IT element. The type of data that is collected may be in common to all the IT elements such as the information that is exposed to the IT element, or the data may be unique and specific to each IT element, for example: if the IT element is a user - the age of the user is relevant, however if the IT element is a server than the age is not relevant but the volume of data traffic is relevant. The present invention automatically collects the relevant data for each IT element and derives the misuseability score accordingly.

The misuseability score is derived by applying a predefined function on a set of relevant parameters that are calculated from the collected data. Thus, given a set of parameters, p1, p2, ..., pn that are dynamically computed from the data that is collected from the IT element, the misuseability score of the IT element can be computed, for example, according to the following equation:

$$\mathrm{MScore}(e,c) = \sum_{i=1}^{n} \alpha_i \cdot p_i$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are constants which define the importance of each parameter pi. These constants should be provided by the security officer or derived automatically. Mscore(e,c) is the misuseability score, c in the equation refers to the specific context for which the misuseability score is computed, and e in the equation refers to the IT element for which the misuseability score is calculated for. For example, the context can be determined by the day of week and part of day for which the misuseability score is derived. In addition, α_i may be different for different context.

A set of possible misuseability measures are presented:

1. Router M-Score (RM-Score)

A router misuseability score depends on the amount and quality of information that is handled by the router as well as on the activation of various capabilities of the router. For example, potentially, the more data that is handled by the router the higher the misuseability score. Data that is sent from / to a sensitive machine (e.g., a machine with a high misuseability score) via the router increase the misuseability score of the router. In addition, a specific capability that is activated on the router (e.g., a VPN, VLAN) increases the attractiveness of the router and the potential misuseability. In addition, the type and version of the router may influence the misuseability score because old / outdated operating system or firmware may result in a more vulnerable device.

Therefore the RM-Score is derived from the following parameters:

- Configuration: Vendor, Operating System (OS) type, OS version, Known vulnerabilities, Activated functionalities (e.g., segmentation, tunneling/encryption, quality of service)

- Purpose: Location (gateway or internal), Importance of connecting networks, Sensitivity level of IP addresses (for example, according to the computed misuseability score of the IP address)
- Activity: Amount of data transmitted, Number of distinct IP addresses, Percentage of encrypted traffic

2. Server M-Score (SM-Score)

A server misuseability score depends on the volume of the activity of the server, the number of connected users, the type and importance of the services that are provided by the server and the properties of the server such as the type and version of the operating system and open ports and running services.

Therefore, the SM-Score is derived from the following parameters:

- Configuration: OS type, OS version, Known vulnerabilities, Number and type of open ports, Running services, Number of users including administrators
- Server purpose: Importance of the service provided by the server, Misuseability score of the connected hosts (HM-Score), Location of server (e.g., internal, DMZ)
- Activity: Number of hosts served by the server, Volume of activity, Volume and type of network traffic;

3. Host M-Score (HM-Score)

The Host M-Score is computed for each end-user device which may include personal computers, laptops and smartphones.

Similar to a sever misuseability score the host misuseability score depends on the volume of the activity, the number of users, the type and importance of the services that are used and the properties of the host such as the type of the device, type and

version of the operating system and open ports, running services, and used services such as secured remote connection.

The host misuseability score may be derived from the following parameters:

- Configuration: Host type, OS type, OS version, Known vulnerabilities, Number and type of open ports, Running services, Number of users including administrators;
- Server purpose: Importance of the service provided by the server, Misuseability score of the users (UM-Score), Location of host;
- Activity: Running applications and services, Number of users using the host machine, Volume of activity, Volume and type of network traffic, Connected networks (both wired and wireless);

4. Tabular Data M-Score (DM-Score)

The DM-Score measure [disclosed at Harel, A., Shabtai, A., Rokach, L., and Elovici, Y., 2012. M-score: A misuseability weight measure. IEEE Trans. on Dependable and Secure Computing, 9(3), 2012, 414-428] estimates the extent of damage that can be caused by an insider that is continuously and gradually exposed to tabular data; i.e., datasets (e.g., result sets of relational database queries). The DM-Score is mainly influenced by the number of entities exposed to the insider (i.e., number of records), the number of properties available on each entity (i.e., number of attributes), the value of properties, and the anonymity level which is regarded as the effort that is required in order to fully identify a specific entity in the data.

5. Textual M-Score (TM-Score)

The TM-Score measure [disclosed in Shabtai, A., Vartanian, A., 2014. TM-Score: A Misuseability Weight Measure for Textual Content", submitted to IEEE Trans. on Information Forensics and Security] estimates the extent of damage that can be caused by an insider that is continuously and gradually exposed to documents. The

extent of damage is determined by the amount, type and quality of information to which the insider is exposed. This is done by deriving an accumulated TM-Score each time that the user is exposed to a document (e.g., opening a file, printing a file, copying a file to a storage device). The TM-Score is accumulated in the sense that it considers the documents that the insider was exposed to in the past as well as the recently exposed document. The main challenge in deriving the accumulated TM-Score is the identification of the residual information in the recently exposed document (i.e., identifying exactly what is the true new information in the document with relation to the previously exposed documents) and its contribution to the accumulated TM-Score. It can be said that the TM Score is an extension the DM Score.

6. User M-Score (UM-Score)

In general, the user M-Score is derived from the type of services and machines that the user accesses as well the type of data and information that the user is exposed to. It can also be derived from general behavioral patterns of the user (e.g., a salesperson who travels a lot and connects to many WiFi networks). Therefore, the UM-Score is computed from the following parameters:

- Demographic features: Age, Role in the organization, Seniority
- Activity: Number of systems and services that the user accesses and the misuseability score of such systems, Volume of activity (i.e., how often the user access the systems and services), Volume and type of generated network traffic (e.g., encrypted traffic), Connected networks (both wired and wireless)
- Sensitivity of data: User's derived DM-Score, User's derived TM-Score

Fig. 2 summarizes the connections among the misuseability scores of the various IT components and presents the derivation tree of misuseability scores. It can be seen that the Router misuseability score 201 (Mscore) affects the Host Mscore 202, and

the server Mscore 203. The server Mscore 203 affects the host Mscore 202 and the host Mscore 202 affects the User Mscore 204 as well as the data Mscore 206 and the Textual Mscore 205 affrects the user Mscore 204.

As illustrated by Fig. 1, the IT infrastructure of an organization can be represented as a graph. The nodes of the graph indicate an IT element and are assigned with attributes such as type of IT element and its computed misuseability score. The links may indicate the strength of the connection between elements. The strength of the connection between elements can be derived for example, from the amount of traffic or activity between two elements (e.g., a server and a router, or a PC and a server) and from the rate of activity and connection between the two elements.

Given such a graph, various graph clustering and community detection algorithms may be applied in order to identify clusters of IT elements that as a group are highly connected and highly misuseable. Note that each individual element in the identified group not necessarily have the highest misuseability score. The security officer of the organization should focus on those highly misuseable sets and increase their protection, for example by raising the awareness of the users, adding monitoring security measures and lowering the thresholds of existing security measures (although it may come on the account of a higher false positive rate).

In addition, an anomaly detection can be applied on the misuseability score of each IT element. An anomaly detection process can be applied in order to learn, for each context, the acceptable misuseability score of the IT element and identify significant deviations from the misuseability level. For example, when the tunneling functionality of a router is disabled, the misuseability score of the router should be reduced significantly, a fact that should raise an alert to the security officer.

Claims

1. A method for automatically and dynamically deriving a misuseability score for one or more IT element(s), comprising the steps of:
   a. collecting data about said one or more IT element(s);
   b. analyzing said collected data;
   c. deriving from said analyzed data a misuseability score for said one or more IT element;
   d. analyzing the connections between said one or more IT elements; and
   e. grouping together IT elements with high connection and deriving a misuseability score to said group.
2. A method according to claim 1, wherein an IT element is: a PC, laptop, server, router, smartphone, or user.
3. A method according to claim 1, wherein an IT element is an element which is connected to an organization network and/or to the internet.
4. A method according to claim 1, wherein the misuseability score is derived according to:

$$\mathrm{MScore}(e,c) = \sum_{i=1}^{n} \alpha_i \cdot p_i$$

   Wherein , p_i are parameters that are dynamically computed from the data that is collected from the IT element
   α_i are constants which define the importance of each parameter;
   Mscore(e,c) is the misuseability score;
   c is the specific context for which the misuseability score is computed; and
   e is the IT element for which the misuseability score is calculated for.

5. A method according to claim 1, wherein the misuseability score for a router is derived from the parameters of: configuration, purpose and activity.
6. A method according to claim 1, wherein the misuseability score for a server is derived from the parameters of: configuration, server purpose and activity.

7. A method according to claim 1, wherein the misuseability score for an end user device is derived from the parameters of: configuration, purpose and activity.
8. A method according to claim 1, wherein the misuseability score for a user is derived from the type of services and machines that said user accesses as well as the type of data and information that the user is exposed to.
9. A method according to claim 8, wherein the misuseability score for a user is derived from general behavioral patterns of the user.
10. A method according to claim 1, wherein the connection between one or more element(s) is derived from the amount of traffic or activity between said one or more element(s) and from the rate of activity and connection between said one or element(s).

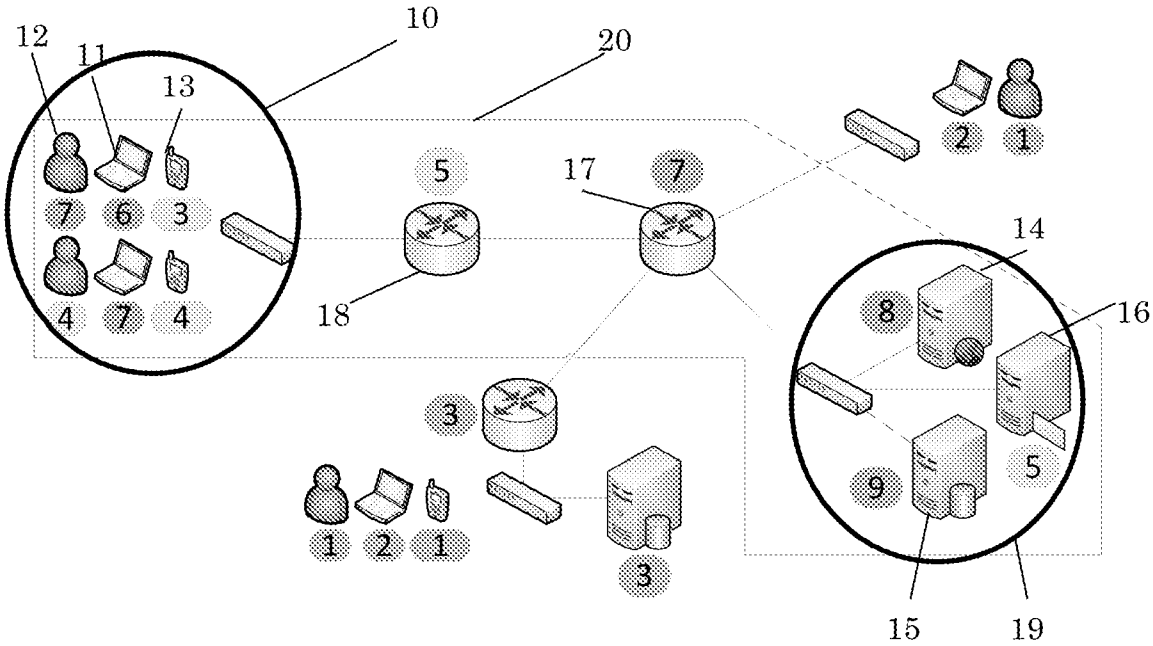

Fig. 1

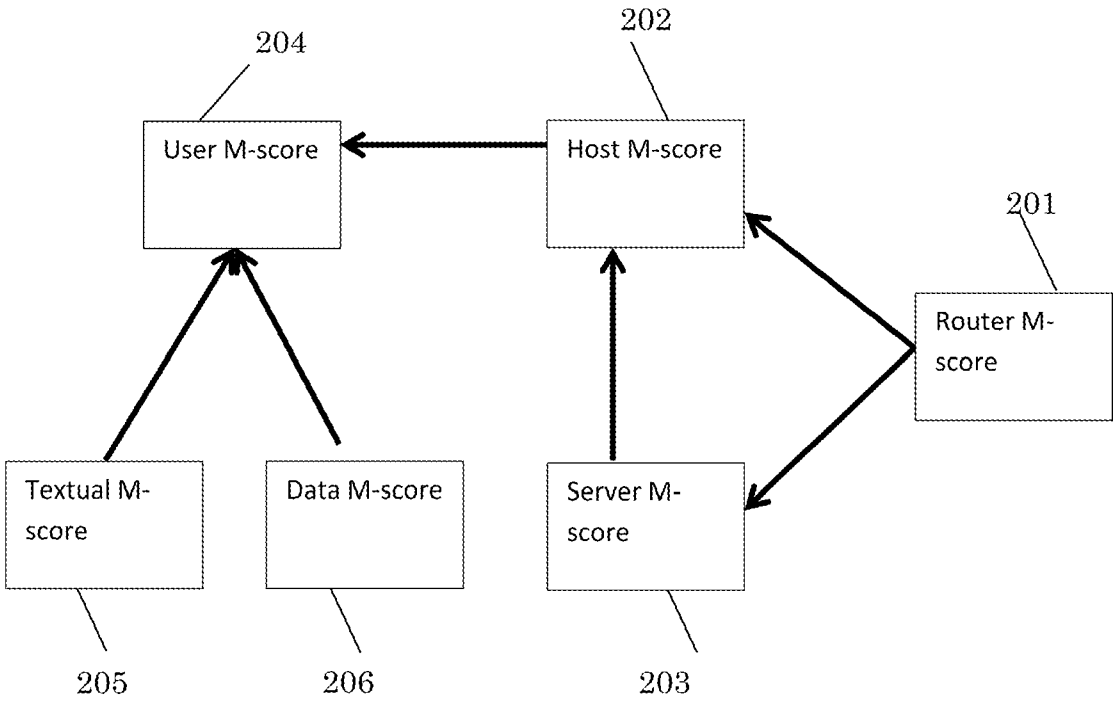
204
202
User M-score
Host M-score
201
Router M-score
Textual M-score
Data M-score
Server M-score
205
206
203

Fig. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL2015/050769

A. CLASSIFICATION OF SUBJECT MATTER
IPC (2015.01) G06F 21/00

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC (2015.01) G06F 21/00

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Databases consulted: PATENTSCOPE
Search terms used: audit score vulnerability element functionality

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2002147803 A1 Dodd et al. 10 Oct 2002 (2002/10/10) ¶¶7,8,17,25,30-32	1-10
A	US 2004083389 A1 Yoshida 29 Apr 2004 (2004/04/29)	8,9
A	US 2010281543 A1 Golomb et al. 04 Nov 2010 (2010/11/04) The whole document	1-10
A	US 2013031634 A1 McClure et al. 31 Jan 2013 (2013/01/31) The entire document	1-10

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

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 Nov 2015	29 Nov 2015
Name and mailing address of the ISA: Israel Patent Office Technology Park, Bldg.5, Malcha, Jerusalem, 9695101, Israel Facsimile No. 972-2-5651616	Authorized officer MAUDA Nissim Telephone No. 972-2-5651733

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IL2015/050769

Patent document cited search report	Publication date	Patent family member(s)	Publication Date
US 2002147803 A1	10 Oct 2002	US 2002147803 A1	10 Oct 2002
		AU 2002243763 A1	12 Aug 2002
		AU 2002244083 A1	12 Aug 2002
		US 2002104014 A1	01 Aug 2002
		US 7340776 B2	04 Mar 2008
		US 2007250935 A1	25 Oct 2007
		US 7712138 B2	04 May 2010
		WO 02061544 A2	08 Aug 2002
		WO 02061544 A3	10 Oct 2002
		WO 02062049 A2	08 Aug 2002
		WO 02062049 A3	21 Nov 2002
US 2004083389 A1	29 Apr 2004	US 2004083389 A1	29 Apr 2004
		US 7337155 B2	26 Feb 2008
		JP 2004147067 A	20 May 2004
		JP 3948389 B2	25 Jul 2007
US 2010281543 A1	04 Nov 2010	US 2010281543 A1	04 Nov 2010
		US 8549649 B2	01 Oct 2013
		WO 2010126733 A1	04 Nov 2010
US 2013031634 A1	31 Jan 2013	US 2013031634 A1	31 Jan 2013
		US 8997234 B2	31 Mar 2015
		CN 103828298 A	28 May 2014
		EP 2737664 A1	04 Jun 2014
		EP 2737664 A4	22 Jul 2015
		WO 2013016577 A1	31 Jan 2013