



(51) International Patent Classification:

G06F 12/14 (2006.01) G06F 21/57 (2013.01)
G06F 21/53 (2013.01) H04L 12/24 (2006.01)
G06F 21/56 (2013.01) H04L 12/26 (2006.01)

(21) International Application Number:

PCT/US2018/067983

(22) International Filing Date:

28 December 2018 (28.12.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

16/106,470 21 August 2018 (21.08.2018) US

(71) Applicant: **DIDI RESEARCH AMERICA, LLC**
[US/US]; 450 National Ave., Mountain View, California
94043 (US).

(72) Inventors: **QIN, Xuwen**; 450 National Ave., Mountain
View, California 94043 (US). **REN, Liwei**; 450 National
Ave., Mountain View, California 94043 (US).

(74) Agent: **CHEN, Weiguo**; SHEPPARD MULLIN
RICHTER & HAMPTON LLP, 379 Lytton Avenue, Palo
Alto, California 94301 (US).

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

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: DATA SECURITY RISKS EVALUATION FOR THREAT DETECTION

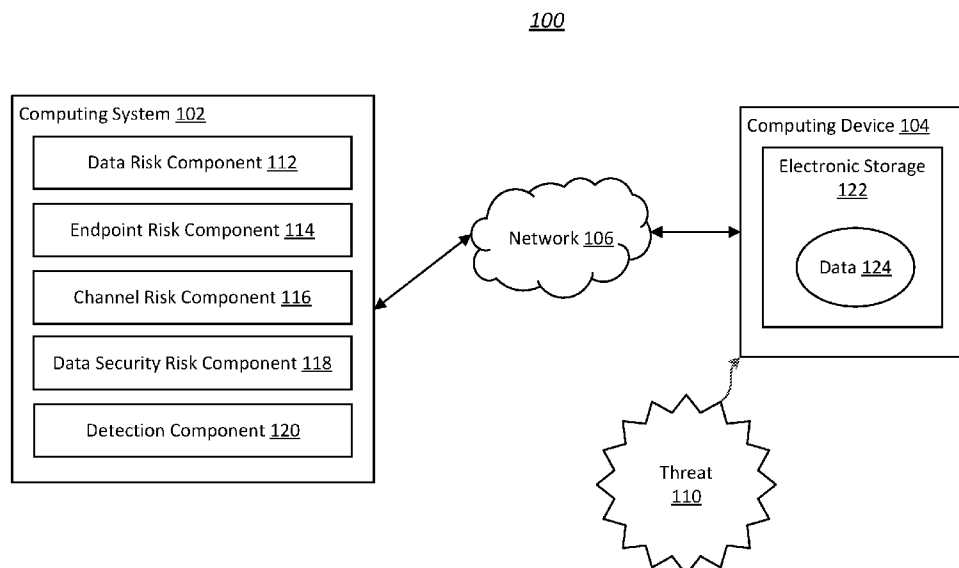


FIG. 1

(57) Abstract: A data risk value for data of an endpoint may be determined. An endpoint risk value for the endpoint may be determined. A channel risk value for a set of channels through which the data is conveyable by the endpoint may be determined. A data security risk value may be determined based on the data risk value, the endpoint risk value, and the channel risk value.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

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

DATA SECURITY RISKS EVALUATION FOR THREAT DETECTION

RELATED APPLICATION

[0001] The application claims the benefit of priority to U.S. Non-Provisional Application No. 16/106,470, filed on August 21, 2018, and entitled “Data Security Risks Evaluation for Threat Detection”, the content of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The disclosure relates generally to evaluating data security risks.

BACKGROUND

[0003] Computing systems may be subject to various security threats, such as data leakage, data corruption, unauthorized access, and/or unauthorized control. Detection of threats based on detection of particular events at a computing system may require individual events and different combinations of events to be coded. Such detection of threats may not provide flexible threat detection. Such detection of threat may not take into account different aspects of a computing system, such as data resting at the computing system, vulnerabilities of the computing system, behavior of users of the computing system, or channels through which the computing system may convey information.

SUMMARY

[0004] One aspect of the present disclosure is directed to a system for evaluating data security risks. The system may comprise one or more processors and a memory storing instructions. The instructions, when executed by the one or more processors, may cause the system to perform: determining a data risk value for data of an endpoint based on a number of classified files within the data and a type of classified files within the data; determining an endpoint risk value for the endpoint based on a user risk value and a cyber security risk value; determining a channel risk value for a set of channels through which the data is conveyable by the endpoint based on a number of channels within the set of channels and a type of channels within the set of channels; and determining a data security risk value based on the data risk value, the endpoint risk value, and the channel risk value.

[0005] Another aspect of the present disclosure is directed to a method for evaluating data security risk. The method may comprise: determining a data risk value for data of an endpoint; determining an endpoint risk value for the endpoint; determining a channel risk value for a set of channels through which the data is conveyable by the endpoint; and determining a data security risk value based on the data risk value, the endpoint risk value, and the channel risk value.

[0006] Yet another aspect of the present disclosure is directed to a system for detecting threats. The system may comprise one or more processors and a memory storing instructions. The instructions, when executed by the one or more processors, may cause the system to perform: determining a data risk value for the data of an endpoint; determining an endpoint risk value for the endpoint; determining a channel risk value for a set of channels through which the data is conveyable by the endpoint; and determining a data security risk value based on the data risk value, the endpoint risk value, and the channel risk value.

[0007] In some embodiments, the data risk value may be determined based on a number of classified files within the data. The data risk value may be determined further based on a type of classified files within the data.

[0008] In some embodiments, the endpoint risk value may be determined based on a user risk value and a cyber security risk value. The user risk value may be determined based on a user behavior associated with the data or the endpoint. The cyber security risk value may be determined based on a number of vulnerabilities of the endpoint.

[0009] In some embodiments, the channel risk value may be determined based on a number of channels within the set of channels. The channel risk value may be determined further based on a type of channels within the set of channels.

[0010] In some embodiments, the data security risk value may be a product of the data risk value, the endpoint risk value, and the channel risk value.

[0011] These and other features of the systems, methods, and non-transitory computer readable media disclosed herein, as well as the methods of operation and functions of the related elements of structure and the combination of parts and economies of manufacture, will become more apparent upon consideration of the following description and the appended claims with reference to the accompanying

drawings, all of which form a part of this specification, wherein like reference numerals designate corresponding parts in the various figures. It is to be expressly understood, however, that the drawings are for purposes of illustration and description only and are not intended as a definition of the limits of the invention. It is to be understood that the foregoing general description and the following detailed description are exemplary and explanatory only, and are not restrictive of the invention, as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Preferred and non-limiting embodiments of the invention may be more readily understood by referring to the accompanying drawings in which:

[0013] FIG. 1 illustrates an example environment for evaluating data security risks, in accordance with various embodiments of the disclosure.

[0014] FIG. 2 illustrates an example triplet model for evaluating data security risks, in accordance with various embodiments of the disclosure.

[0015] FIG. 3 illustrates an example flow of risk value calculations, in accordance with various embodiments of the disclosure.

[0016] FIG. 4 illustrates a flow chart of an example method, in accordance with various embodiments of the disclosure.

[0017] FIG. 5 illustrates a block diagram of an example computer system in which any of the embodiments described herein may be implemented.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0018] Specific, non-limiting embodiments of the present invention will now be described with reference to the drawings. It should be understood that particular features and aspects of any embodiment disclosed herein may be used and/or combined with particular features and aspects of any other embodiment disclosed herein. It should also be understood that such embodiments are by way of example and are merely illustrative of a small number of embodiments within the scope of the present invention. Various changes and modifications obvious to one skilled in the art to which the present invention pertains are deemed to be within the spirit, scope and contemplation of the present invention as further defined in the appended claims.

[0019] The approaches disclosed herein improve technologies for evaluating risks and detecting threats to computing systems. By using a triplet model for evaluating data security risks, flexible threat detection that takes into account different aspects of a computing system may be provided. The triplet model for evaluating security risks may provide for evaluation and detection of threat using risk values for (1) an endpoint, (2) data at the endpoint, and (3) channels through which data is conveyable by the endpoint. By separately determining risks associated with the three different aspects of the computing system, granular measurements of risk may be calculated based on user behavior and endpoint vulnerabilities, and the granular measurements may be weighed or adjusted based on the risks posed by the data and the channels. Separating the risk determination into three elements of the triplet model may facilitate independent changes, updates, or optimization of risk calculations for the separate elements.

[0020] FIG. 1 illustrates an example environment 100 for evaluating data security risks, in accordance with various embodiments. The example environment 100 may include a computing system 102 (e.g., a server) and a computing device 104 (e.g., a client device, desktop, laptop, smartphone, tablet, mobile device). The computing system 102 and the computing device 104 may include one or more processors and memory (e.g., permanent memory, temporary memory). The processor(s) may be configured to perform various operations by interpreting machine-readable instructions stored in the memory. One or both of the computing system 102 and the computing device 104 may include other computing resources or have access (e.g., via one or more connections/networks) to other computing resources.

[0021] The computing system 102 may include a data risk component 112, an endpoint risk component 114, a channel risk component 116, a data security risk component 118, and a detection component 120. The computing system 102 may include other components. The computing system 102 and the computing device 104 may be connected through one or more networks (e.g., a network 106). The computing system 102 and the computing device 104 may exchange information using the network 106. The computing system 102 and the computing device 104 may communicate over the network 106 using one or more communication protocols. The

computing system 102 may be a server of the network 106 and the computing device 104 may be a node of the network 106.

[0022] While the computing system 102 and the computing device 104 are shown in FIG. 1 as single entities, this is merely for ease of reference and is not meant to be limiting. One or more components or functionalities of the computing system 102 or the computing device 104 described herein may be implemented in a single computing device or multiple computing devices. For example, one or more components/functionalities of the computing system 102 may be implemented in the computing device 104 or distributed across multiple computing devices. For instance, the computing device 104 may represent a computing platform, such as an email system and/or a file server, and the components/functionalities of the computing system 102 may be implemented within the computing platform or in one or more other computing devices.

[0023] The computing device 104 may include an electronic storage 122. The electronic storage 122 may refer to a device for storing information, such as information defining computer files. The electronic storage 122 may include one or more storage media in which information may be stored. For example, the electronic storage 122 may include optically readable storage media (e.g., optical disks, etc.), magnetically readable storage media (e.g., magnetic tape, magnetic hard drive, floppy drive, etc.), electrical charge-based storage media (e.g., EPROM, EEPROM, RAM, etc.), solid-state storage media (e.g., flash drive, etc.), or other electronically readable storage media. The electronic storage may be part of the computing device 104 (e.g., integrated into the computing device 104) or removably coupled to the computing device 104.

[0024] The electronic storage 122 may store data 124 and other information. The data 124 may refer to information that is formatted for storage or used by one or more computing devices. For example, the data 124 may include one or more electronic files, executable programs, configuration files, program settings, registry information, or other information stored or used by computing devices. For instance, the data 124 may include one or more classified files. A file may refer to a collection of data or information that has a name (filename). The data 124 may include one or more files of the same type. The data 124 may include files of different types. For example, the

data 124 may include one or more of the following file types: data files, text files, program files, directory files, system files. Other types of files are contemplated. Files within the data 124 may be stored within a single storage media or across multiple storage media. Files within the data 124 may be stored within a single file directory or across multiple file directories. Other types of information within the data 124 are contemplated.

[0025] A classified file may refer to a file associated with one or more classification categories. Classification categories may refer to classes, groupings, or divisions to which files may belong based on contents of the files. Classification categories may identify one or more types of contents of the files. For example, a classification category with which a file is associated may indicate a level or an amount of sensitive information contained within the file. As another example, a classification category with which a file is associated may indicate a level or an amount of classified information contained within the file. Other types of classification categories are contemplated.

[0026] In some embodiments, a file may be associated with multiple classification categories. For example, a file may be associated with multiple types of classification categories. As another example, different portions of a file may be associated with different classification categories. For instance, one part of the file may be associated with a low classification category while another part of the file may be associated with a high classification category.

[0027] The classification categories may determine which users or which groups of users are authorized to access the files. Authorized access of a file may be divided into different types of access. For example, a user's full access to a file may include the user being authorized to open the file, rename the file, add a property to the file, remove a property of the file, change a property of the file, copy the file, delete the file, change the location of the file, share the file, view information in the file, add information to the file, remove information from the file, change information in the file, and otherwise access the file. A user's limited access to a file may include the user being authorized to perform only a subset of activities authorized under full access.

[0028] The computing device 104 may be subject to a threat 110. The threat 110 may refer to potential unauthorized action, occurrence, or event relating to the computing device 104. For example, the threat 110 may include the possibility of the data 124 (or a portion of the data 124) being subject to unauthorized access or modification, such as by an unauthorized user or an unauthorized program that exploits vulnerabilities of the computing device 104, another computing device connected to the computing device 104, or the network 106. For instance, the threat 110 may include an unauthorized user attempting to access the data 124, or a malicious program running on the computing device 104 attempting to destroy or steal the data 124. As another example, the threat 110 may include an unauthorized user or an unauthorized program attempting to install or run unauthorized programs on the computing device 104 or attempting to access an internal network of the computing device 104. As yet another example, the threat 110 may include the user of the computing device 104 improperly using the computing device 104 and/or the data 124. Other types of threats are contemplated.

[0029] The data risk component 112 may be configured to determining a data risk value for data of an endpoint. An endpoint may refer to a device or a node that is connected to a network. An endpoint may communicate across the network with other devices, such as other endpoints, services, or servers. For example, endpoints of a network may include individual computing devices connected to the network, such as desktops, laptops, smartphones, tablets, mobile devices, or other computing devices. For instance, the computing device 104 may be an endpoint of the network 106, and the data risk component 112 may determine a data risk value for the data 124 of the computing device 104.

[0030] A data risk value may refer to a measurement of risk posed by unauthorized action, occurrence, or event relating to data. Risk of data may refer to exposure to danger, harm, loss, or other negative consequence of unauthorized action, occurrence, or event relating to the data. For example, a data risk value for the data 124 including classified files may refer to a measurement of risk posed by unauthorized action, occurrence, or event relating to the classified files. For instance, the data risk value for the data 124 may refer to a quantification of exposure to

danger, harm, loss, or other negative consequence, unauthorized action, occurrence, or event relating to leakage or destruction of the classified files.

[0031] In some embodiments, a data risk value for data may be determined based on user input. For example, a user may manually indicate the data risk value for one or more classified files within the data 124 and the data risk component 112 may retrieve the data risk value indicated by the user. In some embodiments, the data risk value may be determined based on a number of classified files within data. For example, the data risk component 112 may traverse the folder(s) containing the classified files within the data 124 and determine the data risk value for the data 124 based on how many classified files are found within the data 124. Larger data risk value may correspond to greater number of classified files.

[0032] In some embodiments, the data risk value may be determined based on type(s) of classified files within the data. For example, the data risk component 112 may identify the type(s) of the classified files within the data 124 and determine the data risk value for the data 124 based on different type(s) of classification categories with which the classified files are associated. Larger data risk value may correspond to higher classification categories (e.g., reflecting a higher sensitivity or the amount of classified information within the files).

[0033] In some embodiments, the data risk value may range between values of zero and one. A “zero” data risk value may indicate that there is no risk posed by unauthorized action, occurrence, or event relating to data. For example, data of publicly accessible information may have a data risk value of zero. A “one” data risk value may indicate the highest risk posed by unauthorized action, occurrence, or event relating to data. A data risk value of one may be determined based on sensitivity or classified nature of information within the data or the amount of sensitive information or classified information within the data. For example, small amount of highly sensitive/classified information may have a data risk value of one. As another example, large amount of low or moderately sensitive/classified information may have a data risk value of one. Other ranges of data risk value are contemplated.

[0034] For example, the determination of the data risk value may include the following calculation: data risk score (RSD) = $1 - e^{-\alpha K}$, where K is the number of

classified files within the data and α is a positive number. The value of α may be configurable, and may be adjusted based on user input or content of the classified files. For instance, value of α may change based on the sensitivity or type of the classified information within the data, or based on the size of the sensitive/classified file(s) within the data. Other values and calculations of data risk values are contemplated.

[0035] The endpoint risk component 114 may be configured to determine an endpoint risk value for an endpoint. For instance, the computing device 104 may be an endpoint of the network 106, and the endpoint risk component 114 may determine an endpoint risk value for the computing device 104.

[0036] An endpoint risk value may refer to a measurement of risk that unauthorized action, occurrence, or event relating to data will occur via an endpoint. For example, an endpoint risk value for the computing device may refer to a measurement of risk that unauthorized action, occurrence, or event relating to the data 124 will occur via the computing device 104. For instance, the endpoint risk value for the computing device 104 may refer to a quantification of possibility or probability that unauthorized action, occurrence, or event relating to the data 124 will occur at or through the computing device 104.

[0037] In some embodiments, the endpoint risk value may be determined based on a user risk value and a cyber security risk value. For example, the endpoint risk value may be determined based on a combination of the user risk value and the cyber security risk value. For instance, the endpoint risk value may be determined as a sum or other combination of the user risk value and the cyber security risk value.

[0038] A user risk value may refer to a measurement of risk that unauthorized action, occurrence, or event relating to data will occur due to a user action or a user inaction. For example, a user risk value may refer to a quantification of possibility or probability that unauthorized action, occurrence, or event relating to the data 124 will occur because of one or more users of the computing device 104. A user may intentionally or unintentionally pose risk to the unauthorized use of data. For example, a user may intentionally or unintentionally leak classified files to unauthorized persons.

[0039] In some embodiments, a user risk value for an endpoint may be determined based on user behavior associated with data or an endpoint. A user behavior may refer to a way in which a user acts with respect to data stored at a particular endpoint, other endpoints, data stored at the particular endpoint, or other data. The endpoint risk component 114 may analyze user behavior with respect to a particular data, other data, a particular endpoint or other endpoint to determine the user risk value for the particular endpoint. The endpoint risk component 114 may analyze a variety of user behaviors to determine the user risk value. For example, the endpoint risk component 114 may take into consideration the number of times the user has visited a malicious website, whether the user has visited a certain number of malicious websites within a given time period, whether the user has previously allowed malware to be installed on an endpoint, whether the user has shared an infected file with other users, where the user is located, the locations to which the user has traveled, the persons or devices with which the user has interacted, status of the user in an organization (e.g., importance/role within the organization, newly hired, recently resigned), or other behaviors of the user.

[0040] Larger user risk value may correspond to higher risk that the user will intentionally or unintentionally allow unauthorized action, occurrence, or event relating to data to occur. Different behaviors of user may be weighed the same or differently in determining the user risk value. For example, a user having recently visited a malicious website may be weighed the same or differently from the user having been recently hired in the determination of the user risk value.

[0041] In some embodiments, the user risk value may range between values of zero and one-hundred. A “zero” user risk value may indicate that there is no risk that unauthorized action, occurrence, or event relating to data will occur due to a user action or a user inaction. A “one-hundred” user risk value may indicate the highest risk that unauthorized action, occurrence, or event relating to data will occur due to a user action or a user inaction. Other ranges of user risk value are contemplated.

[0042] For example, the determination of the user risk value may include the following calculation: user risk score (RSU) = $100 * (1 - e^{-V})$, where $V = \gamma_1 * N + \gamma_2 * f + \gamma_3 * s + \gamma_4 * M + \dots$, each $\gamma_i > 0$ is a configurable parameter providing weights to occurrence of different user behaviors, N is the number of times the user has allowed

unauthorized action, occurrence, or event relating to data to occur (within a given time period), M is the number of times that the user's peers (e.g., co-workers with whom the user interacts or shares data) has allowed unauthorized action, occurrence, or event relating to data to occur (within a given time period), and f and s are flags (having value of zero or one) that indicate whether the user or the user's peers have exhibited the corresponding user behavior (within a given time period). Other user behaviors or factors may be taken into consideration by including other gamma and corresponding flag or number of occurrences into the user risk value calculation. Other values and calculations of user risk values are contemplated.

[0043] A cyber security risk value may refer to a measurement of risk that unauthorized action, occurrence, or event relating to data will occur due to one or more vulnerabilities at an endpoint. For example, a cyber security risk value may refer to a quantification of possibility or probability that unauthorized action, occurrence, or event relating to the data 124 will occur because of one or more vulnerabilities of the computing device 104. A vulnerability of an endpoint may refer to a flaw (in code or design) of an endpoint that creates a potential point of security comprise at the endpoint. A vulnerability of an endpoint may exist due to one or more malicious programs (e.g., malware installed at an endpoint). A vulnerability of an endpoint may exist due to a flaw in software/firmware of the endpoint (e.g., security flaw that has yet to be addressed by a patch or an update).

[0044] In some embodiments, a cyber security risk value for an endpoint may be determined based on a number of vulnerabilities of the endpoint. The endpoint risk component 114 may scan the endpoint to determine the number of vulnerabilities existing at the endpoint, such as the number of malware running on the endpoint or the number of security updates to be applied to the endpoint.

[0045] Larger cyber security risk value may correspond to higher risk that unauthorized action, occurrence, or event relating to data will occur due to the endpoint. Different vulnerabilities of an endpoint may be weighed the same or differently in determining the cyber security risk value. For example, the presence of a malware on an endpoint may be weighed the same or differently from the endpoint having a security update that has not yet been applied. As another example, different malware on the endpoint may be weighed the same or differently, and the weights of

the unapplied security update may be changed based on the type of fix applied by the security update or the duration of time that the security update has been available.

[0046] In some embodiments, the cyber security risk value may range between values of zero and one-hundred. A “zero” cyber security risk value may indicate that there is no risk that unauthorized action, occurrence, or event relating to data will occur due to an endpoint or a vulnerability at an endpoint. A “one-hundred” cyber security risk value may indicate the highest risk that unauthorized action, occurrence, or event relating to data will occur due to an endpoint or a vulnerability at an endpoint. Other ranges of user risk value are contemplated.

[0047] For example, the determination of the cyber security risk value may include the following calculation: cyber security risk score (RSCS) = $100 * (1 - e^{-V})$, where $V = \gamma_1 * N + \gamma_2 * f_2 + \gamma_3 * f_3 + \gamma_4 * f_4 + \gamma_5 * f_5 + \gamma_6 * K + \gamma_{7,1} * N_1 + \gamma_{7,2} * N_2 + \gamma_{7,3} * N_3 \dots$, each $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_{7,1}, \gamma_{7,2}, \gamma_{7,3} > 0$ are configurable parameters providing weights to different vulnerabilities, N is the number of one or more types of vulnerabilities detected at the endpoint, f_2, f_3, f_4, f_5 are flags (having value of zero or one) that indicate whether certain vulnerabilities are detected at the endpoint, K is the number of one or more types of vulnerabilities detected at the user’s peer endpoints, and N_1, N_2, N_3 are the numbers of particular activities (e.g., visits to safe external websites, visits to risky external website, reception of files from unknown sources) performed at the endpoint. Other vulnerabilities or factors may be taken into consideration by adding other gamma and corresponding flag or number of occurrences into the cyber security risk value calculation. Other values and calculations of cyber security risk values are contemplated.

[0048] The channel risk component 116 may be configured to determine a channel risk value for a set of channels through which data of an endpoint is conveyable by the endpoint. For instance, the computing device 104 may be an endpoint of the network 106, and the channel risk component 116 may determine a channel risk value for a set of channels through which the data 124 of the computing device 104 is conveyable by the computing device 104.

[0049] A set of channels may refer to one or more channels through which an endpoint may convey data. For example, a set of channels for the computing device 104 may include one or more channels through which the computing device 104 may

convey some or all of the data 124. A channel may refer to a path through which information may flow. A channel may refer to the medium through which information may flow or a program that is used to convey information through a medium. For example, a set of channels of the computing device 104 may include wired or wireless connection, peripheral connectors (e.g., USB connector), email program, texting program, virtual chat program, or video conferencing program. Other types of channels are contemplated.

[0050] A channel risk value may refer to a measurement of risk that unauthorized action, occurrence, or event relating to data will occur via a set of channels. For example, a channel risk value for the computing device may refer to a measurement of risk that unauthorized action, occurrence, or event relating to the data 124 will occur via one or more channels of the computing device 104. For instance, the channel risk value for the computing device 104 may refer to a quantification of possibility or probability that unauthorized action, occurrence, or event relating to the data 124 will occur at or through one or more channels of the computing device 104.

[0051] In some embodiments, a channel risk value may be determined based on a number of channels within the set of channels. For example, the channel risk component 116 determine the number of channels through which the computing device 104 may exchange information relating to the data 124 and determine the channel risk value for the computing device 104 based on the number of channels. Larger channel risk value may correspond to greater number of channels.

[0052] In some embodiments, the channel risk value may be determined based on type(s) of channels within the set of channels. A type of channels may refer to a category of channels, such as categories of medium through which information may flow, categories of programs that are used to convey information through a medium, or categories of security associated with different channels (e.g., unsecured channel, lowly secured channel, moderately secured channel, highly secured channel). For example, the channel risk component 116 may identify the type(s) of channels by which the computing device 104 may convey the data 124 or information relating to the data, and determine the channel risk value for the computing device 104 based on different type(s) of channels of the computing device 104.

[0053] In some embodiments, the channel risk value may range between values of zero and one. A “zero” channel risk value may indicate that there is no risk that unauthorized action, occurrence, or event relating to data will occur via the channel(s) of the endpoint. For example, the channel(s) of the endpoint may be protected by security measure(s) to protect leakage of classified files. A “one” channel risk value may indicate the highest risk that unauthorized action, occurrence, or event relating to data will occur via the channel(s) of the endpoint. Other ranges of channel risk value are contemplated.

[0054] For example, the determination of the channel risk value may include the following calculation: channel risk score (RSC) = $1 - e^{-\beta * M}$, where M is the number of channels through which unauthorized action, occurrence, or event relating to data may occur and β is a positive number. The value of β may be configurable, and may be adjusted based on the type of the channels. For instance, value of β may change based on security measures in place to prohibit unauthorized action, occurrence, or event relating to data to occur through a channel. For example, an email program may be secured using scanners to prevent leakage of classified files and have a lower β value than a chat program, which may not be secured or have less extensive security measures than the email program. As another example, the determination of the channel risk value may include the following calculation: $RSC = 1 - e^{-V}$, where $V = \mu_1 + \mu_2 + \dots + \mu_M$, $\mu_j > 0$, $j=1, \dots, M$, M is the number of channels, and $\mu_1 + \mu_2 + \dots + \mu_M$ represent risk values for different channels. Other values and calculations of channel risk values are contemplated.

[0055] The data security risk component 118 may be configured to determine a data security risk value based on the data risk value, the endpoint risk value, and the channel risk value. For example, the data security risk value may be determined based on a combination of the data risk value, the endpoint risk value, and the channel risk value. For instance, the data security risk value may be determined as a product or other combination of the data risk value, the endpoint risk value, and the channel risk value. The data security risk component 118 may take other information or factors into account in determining a data security risk value.

[0056] A data security risk value may refer to a comprehensive measurement of risk posed by unauthorized action, occurrence, or event relating to data at an endpoint.

A data security risk value may take into account: the data risk value (measurement of risk posed by unauthorized action, occurrence, or event relating to data), the endpoint risk value (measurement of risk that unauthorized action, occurrence, or event relating to data will occur via an endpoint), and the channel risk value (measurement of risk that unauthorized action, occurrence, or event relating to data will occur via a set of channels).

[0057] Calculation of the data security risk value (RSDS) as a product of the data risk value, the endpoint risk value, and the channel risk value ($RSDS = RSD * RSE * RSC$) may provide a data security risk value that ranges between values of zero and two-hundred. The endpoint risk value (RSE, combination of user risk value and cyber security risk value) may provide granular risk measurement (values ranging from zero to two-hundred) based on user behavior and endpoint vulnerabilities, while the data risk value (RSD, ranging from zero to one) and the channel risk value (RSC, ranging from zero to one) may determine how much of the endpoint risk value should be taken into account when detecting data leaking threats.

[0058] Thus, the risk values may be determined based on a triplet model including three separate elements: (1) the data at the endpoint, (2) the endpoint, and (3) the channels of the endpoint. Granular measurements of risk may be calculated based on user behavior and endpoint vulnerabilities, and the granular measurements may be weighed or adjusted based on the risks posed by both the data and the channel. The use of the triplet model for evaluating data security risks may provide for more flexible and nuanced threat detection than threat detection based on recognition of problematic events. The use of the triplet model for evaluating data security risks may enable tailoring of threat detection to different security policies with different rules relating to the data risk value, the endpoint risk value, the channel risk value, or the data security risk value.

[0059] The separation of the risk determination into three elements of the triplet model may facilitate independent changes, updates, or optimization of risk calculations for the separate elements. For instances, factors taken into account when calculating the data risk value, the endpoint risk value, or the channel risk value may independently be changed. Factors taken into account when calculating the data risk value, the endpoint risk value, or the channel risk value may be changed to reflect the

desired security policies. The triplet model for evaluating data security risks may merge into a single view risk arising from the classification of data, the vulnerabilities of an endpoint, the user behavior, and the channels. The single view may enable computer analysis of different aspects of a computing system for threat detection while providing a comprehensive view of how different aspects of the computing system contribute to the overall risk faced by the computing system.

[0060] The detection component 120 may be configured to detect a threat (e.g., the threat 110) based on the data security risk value. For example, the detection component 120 may detect a threat based on the data security risk value satisfying a threat detection criterion. A threat criterion may refer to one or more rules or standards by which a threat is detected. For instance, a threat may be detected based on the data security risk value being the same as or greater than a threat threshold. The detection component 120 may take other information or factors into account in detecting a threat.

[0061] A threat detected by the detection component 120 may refer to potential unauthorized action, occurrence, or event relating to computing device 104. For example, the threat 110 may refer to potential unauthorized action, occurrence, or event relating to the data 124 of the computing device 104, such as leakage or destruction of the data 124. Detection of other threats are contemplated.

[0062] The threat threshold may be static or dynamic. The threat threshold may be set by a user (e.g., user defining the value of the threat threshold). The threat threshold may be automatically set based on occurrence of one or more events. For example, the threat threshold may be lowered based on the computing device 104 or the network 106 operating in a high-security mode and raised based on the computing device 104 or the network 106 operating in a low-security mode.

[0063] In some embodiments, the detection of a threat, the data security risk value, or values underlying the data security risk value may be presented within a user interface. For example, based on a threat being detected based on the data security risk value satisfying a threat detection criterion, the data security risk value may be presented within a user interface. The user interface may also provide values of the data risk value, the endpoint risk value, and the channel risk value. The endpoint risk value may be broken out into the user risk value and the cyber security risk value.

Different values that make up the data security risk value may be presented differently (e.g., in different fonts, in different colors).

[0064] The presentation of different values that make up the data security risk value may enable analysis of which area(s) of security needs to be improved. For example, a high data security risk value for an endpoint may be the result of a high user risk value. To reduce the risk of threat for the endpoint, the user may be required to attend training on proper computing behavior to reduce the user risk value. The data of the endpoint may be limited to non-classified files or files with low-sensitivity to reduce the data risk value. The channels available at the endpoint may be limited (e.g., reduce the number of channels, increase security measures in place for the channels) to reduce the channel risk value.

[0065] In some embodiments, remedial measures may be suggested or taken based on detection of threats. Based on detection of a threat, information relating to data, data risk value, endpoint, endpoint risk value, channel, channel risk value, or data security risk value may be analyzed to determine what actions may be taken to reduce or remove the threat. For example, one or more aspects of data, endpoint vulnerabilities, user behavior, or channels may be tagged for further view or analysis. One or more changes to data, endpoint, user behavior, or channels may be suggested or automatically taken to reduce the data security risk value.

[0066] FIG. 2 illustrates an example triplet model 200 for evaluating data security risks, in accordance with various embodiments of the disclosure. The triplet model 200 include three elements: a data 202, an endpoint 204, and a channel 206. The data 202 may represent risk due to confidential data at rest in an endpoint. The endpoint 204 may represent risk due to actors at the endpoints, including user(s) at the endpoint and vulnerabilities (e.g., malware, unpatched security flaw) at the endpoint. The channel 206 may represent risk due to channel(s) which may allow unauthorized access of data at the endpoint. Individual elements 202, 204, 206 of the model 200 may contribute risk to potential threat at an endpoint. Risk values associated with individual elements 202, 204, 206 may be separately analyzed and combined together to form a comprehensive model for evaluating data security risks. Risk values associated with individual element 202, 204, 206 may be determined independently of each other. Risk values associated with individual elements 202, 204, 206 may be

calculated using separate sets of algorithm. Individual sets of algorithm may be modified (e.g., changed, updated, improved) independently of each other. For example, factors taken into consideration for determination of risk values associated with the data 202 may be changed to include additional factors without impacting calculation of risk values for the endpoint 204 or the channel 206.

[0067] FIG. 3 illustrates an example flow 300 of risk value calculations, in accordance with various embodiments of the disclosure. The flow 300 may include calculations 302, 304, 306, 308, 310, 312 of different risk values for an endpoint. The calculation 302 may include a calculation of a data risk value (data risk score, RSD). The data risk value may range from zero to one, with zero being the lowest risk value and one being the highest risk value.

[0068] The calculation 304 may include a calculation of a user risk value (user risk score, RSU). The user risk value may range from zero to one hundred, with zero being the lowest risk value and one hundred being the highest risk value.

[0069] The calculation 306 may include a calculation of a cyber security risk value (cyber security risk score, RSCS). The cyber security risk value may range from zero to one hundred, with zero being the lowest risk value and one hundred being the highest risk value.

[0070] The calculation 308 may include a calculation of an endpoint risk value (endpoint risk score, RSE) based on a combination of the user risk value and the cyber security risk value. For example, the endpoint risk value may be the sum of the user risk value and the cyber security risk value. The endpoint risk value may range from zero to two hundred, with zero being the lowest risk value and two hundred being the highest risk value.

[0071] The calculation 310 may include a calculation of the channel risk value (channel risk score, RSC). The channel risk value may range from zero to one, with zero being the lowest risk value and one being the highest risk value.

[0072] The calculation 312 may include a calculation of a data security risk value (data security risk score, RSDS) based on a combination of the data risk value, the endpoint risk value, and the channel risk value. For example, the data security risk value may be the product of the data risk value, the endpoint risk value, and the channel risk value. The data security risk value may range from zero to two hundred,

with zero being the lowest risk value and two hundred being the highest risk value. Other ranges of risk values and other calculations of risk values are contemplated.

[0073] FIG. 4 illustrates a flowchart of an example method 400, according to various embodiments of the present disclosure. The method 400 may be implemented in various environments including, for example, the environment 100 of FIG. 1. The operations of the method 400 presented below are intended to be illustrative. Depending on the implementation, the method 400 may include additional, fewer, or alternative steps performed in various orders or in parallel. The method 400 may be implemented in various computing systems or devices including one or more processors.

[0074] With respect to the method 400, at block 410, a data risk value for data of an endpoint may be determined. At block 420, an endpoint risk value for the endpoint may be determined. At block 430, a channel risk value for a set of channels may be determined. The data may be conveyed by the endpoint through the set of channels. At block 440, a data security risk value may be determined based on the data risk value, the endpoint risk value, and the channel risk value. At block 450, a threat may be detected based on the data security risk value.

[0075] One or more blocks of the method 400 may be performed by one or more computer components that are the same as or similar to the components of the computing system 102 shown in FIG. 1. For example, the block 410 may be performed by a computer component the same as or similar to the data risk component 112. The block 420 may be performed by a computer component the same as or similar to the endpoint risk component 114. The block 430 may be performed by a computer component the same as or similar to the channel risk component 116. The block 440 may be performed by a computer component the same as or similar to the data security risk component 118. The block 450 may be performed by a computer component the same as or similar to the detection component 120.

[0076] One or more blocks of the method 400 may correspond to calculations performed to determine risk values of one or more elements of the triplet model 200 shown in FIG. 2 for evaluating data security risks and to one or more calculations of the flow 300 of risk value calculations shown in FIG. 3. For example, the block 410 may correspond to calculation(s) performed to determine risk value of the data 202

(the calculation 302). The block 420 may correspond to calculation(s) performed to determine risk value of the endpoint 204 (the calculations 304, 306, 308). The block 430 may correspond to calculation(s) performed to determine risk value of the channel 206 (the calculation 310). The block 440 may correspond to calculation(s) performed to determine the overall data security risk value of the triplet model 200 (the calculation 312).

[0077] FIG. 5 is a block diagram that illustrates a computer system 500 upon which any of the embodiments described herein may be implemented. The computer system 500 includes a bus 502 or other communication mechanism for communicating information, one or more hardware processors 504 coupled with bus 502 for processing information. Hardware processor(s) 504 may be, for example, one or more general purpose microprocessors.

[0078] The computer system 500 also includes a main memory 506, such as a random access memory (RAM), cache and/or other dynamic storage devices, coupled to bus 502 for storing information and instructions to be executed by processor(s) 504. Main memory 506 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor(s) 504. Such instructions, when stored in storage media accessible to processor(s) 504, render computer system 500 into a special-purpose machine that is customized to perform the operations specified in the instructions. Main memory 506 may include non-volatile media and/or volatile media. Non-volatile media may include, for example, optical or magnetic disks. Volatile media may include dynamic memory. Common forms of media may include, for example, a floppy disk, a flexible disk, hard disk, solid state drive, magnetic tape, or any other magnetic data storage medium, a CD-ROM, any other optical data storage medium, any physical medium with patterns of holes, a RAM, a DRAM, a PROM, an EPROM, a FLASH-EPROM, NVRAM, any other memory chip or cartridge, and networked versions of the same.

[0079] The computer system 500 may implement the techniques described herein using customized hard-wired logic, one or more ASICs or FPGAs, firmware and/or program logic which in combination with the computer system causes or programs computer system 500 to be a special-purpose machine. According to one embodiment, the techniques herein are performed by computer system 500 in response to

processor(s) 504 executing one or more sequences of one or more instructions contained in main memory 506. Such instructions may be read into main memory 506 from another storage medium, such as storage device 508. Execution of the sequences of instructions contained in main memory 506 causes processor(s) 504 to perform the process steps described herein.

[0080] For example, the computing system 500 may be used to implement the computing system 102 or one or more components of the computing system 102 shown in FIG. 1. As another example, the process/method shown in FIG. 4 and described in connection with this figure may be implemented by computer program instructions stored in main memory 506. When these instructions are executed by processor(s) 504, they may perform the steps as shown in FIG. 4 and described above. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions.

[0081] The computer system 500 also includes a communication interface 510 coupled to bus 502. Communication interface 510 provides a two-way data communication coupling to one or more network links that are connected to one or more networks. As another example, communication interface 510 may be a local area network (LAN) card to provide a data communication connection to a compatible LAN (or WAN component to communicate with a WAN). Wireless links may also be implemented.

[0082] The performance of certain of the operations may be distributed among the processors, not only residing within a single machine, but deployed across a number of machines. In some example embodiments, the processors or processor-implemented engines may be located in a single geographic location (e.g., within a home environment, an office environment, or a server farm). In other example embodiments, the processors or processor-implemented engines may be distributed across a number of geographic locations.

[0083] While examples and features of disclosed principles are described herein, modifications, adaptations, and other implementations are possible without departing from the spirit and scope of the disclosed embodiments. Also, the words “comprising,” “having,” “containing,” and “including,” and other similar forms are

intended to be equivalent in meaning and be open ended in that an item or items following any one of these words is not meant to be an exhaustive listing of such item or items, or meant to be limited to only the listed item or items. It must also be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise.

[0084] The embodiments illustrated herein are described in sufficient detail to enable those skilled in the art to practice the teachings disclosed. Other embodiments may be used and derived therefrom, such that structural and logical substitutions and changes may be made without departing from the scope of this disclosure. The Detailed Description, therefore, is not to be taken in a limiting sense, and the scope of various embodiments is defined only by the appended claims, along with the full range of equivalents to which such claims are entitled.

WHAT IS CLAIMED IS:

1. A system for evaluating data security risks, the system comprising:
one or more processors; and
a memory storing instructions that, when executed by the one or more processors, cause the system to perform:
determining a data risk value for data of an endpoint based on a number of classified files within the data and a type of classified files within the data;
determining an endpoint risk value for the endpoint based on a user risk value and a cyber security risk value;
determining a channel risk value for a set of channels through which the data is conveyable by the endpoint based on a number of channels within the set of channels and a type of channels within the set of channels; and
determining a data security risk value based on the data risk value, the endpoint risk value, and the channel risk value.
2. The system of claim 1, wherein:
the user risk value is determined based on a user behavior associated with the data or the endpoint;
the cyber security risk value is determined based on a number of vulnerabilities of the endpoint; and
the data security risk value is a product of the data risk value, the endpoint risk value, and the channel risk value.
3. A system for evaluating data security risks, the system comprising:
one or more processors; and
a memory storing instructions that, when executed by the one or more processors, cause the system to perform:
determining a data risk value for data of an endpoint;
determining an endpoint risk value for the endpoint;
determining a channel risk value for a set of channels through which the data is conveyable by the endpoint; and

determining a data security risk value based on the data risk value, the endpoint risk value, and the channel risk value.

4. The system of claim 3, wherein the data risk value is determined based on a number of classified files within the data.
5. The system of claim 4, wherein the data risk value is determined further based on a type of classified files within the data.
6. The system of claim 5, wherein the endpoint risk value is determined based on a user risk value and a cyber security risk value.
7. The system of claim 6, wherein the user risk value is determined based on a user behavior associated with the data or the endpoint.
8. The system of claim 7, wherein the cyber security risk value is determined based on a number of vulnerabilities of the endpoint.
9. The system of claim 8, wherein the channel risk value is determined based on a number of channels within the set of channels.
10. The system of claim 9, wherein the channel risk value is determined further based on a type of channels within the set of channels.
11. The system of claim 10, wherein the data security risk value is a product of the data risk value, the endpoint risk value, and the channel risk value.
12. A method for evaluating data security risks, the method comprising:
 - determining a data risk value for data of an endpoint;
 - determining an endpoint risk value for the endpoint;
 - determining a channel risk value for a set of channels through which the data is conveyable by the endpoint; and

determining a data security risk value based on the data risk value, the endpoint risk value, and the channel risk value.

13. The method of claim 12, wherein the data risk value is determined based on a number of classified files within the data.

14. The method of claim 13, wherein the data risk value is determined further based on a type of classified files within the data.

15. The method of claim 14, wherein the endpoint risk value is determined based on a user risk value and a cyber security risk value.

16. The method of claim 15, wherein the user risk value is determined based on a user behavior associated with the data or the endpoint.

17. The method of claim 16, wherein the cyber security risk value is determined based on a number of vulnerabilities of the endpoint.

18. The method of claim 17, wherein the channel risk value is determined based on a number of channels within the set of channels.

19. The method of claim 18, wherein the channel risk value is determined further based on a type of channels within the set of channels.

20. The method of claim 19, wherein the data security risk value is a product of the data risk value, the endpoint risk value, and the channel risk value.

100

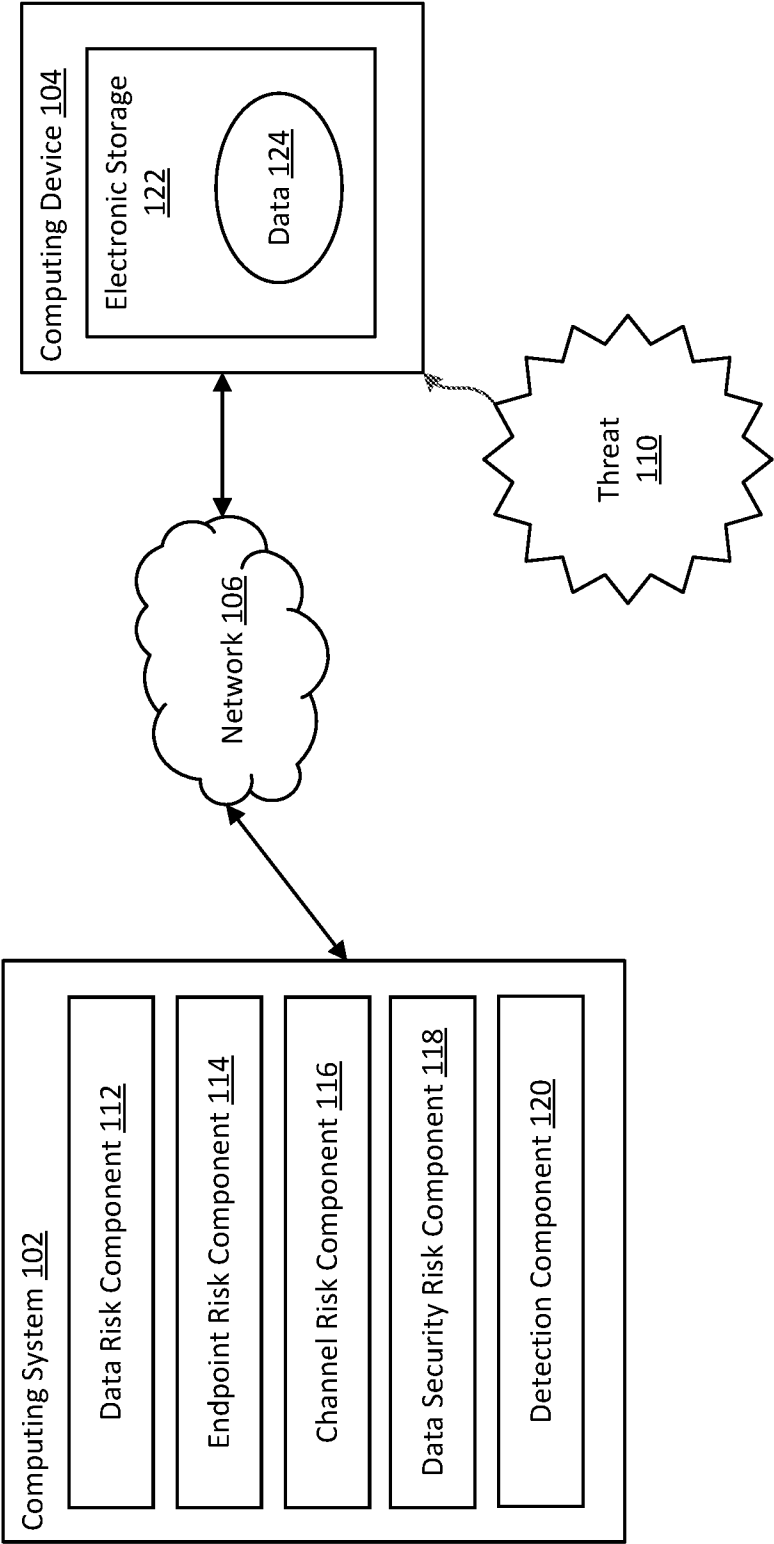


FIG. 1

2/5

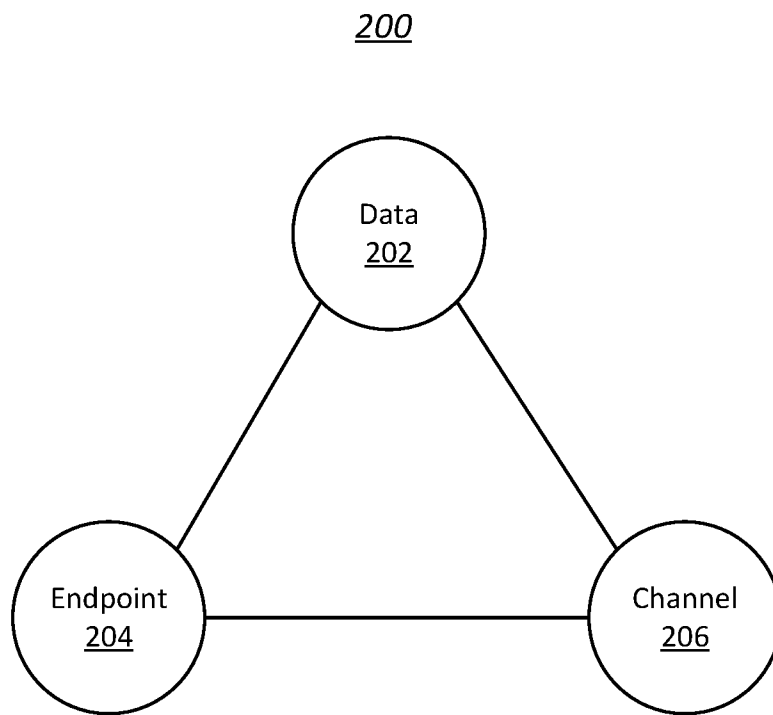


FIG. 2

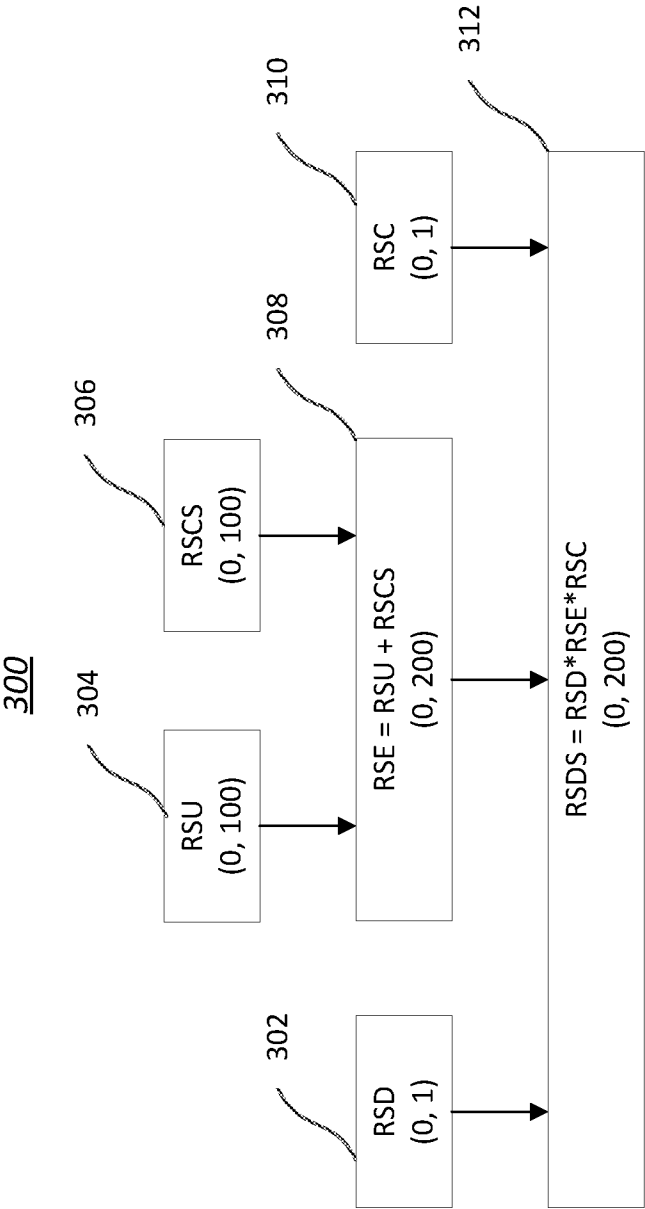
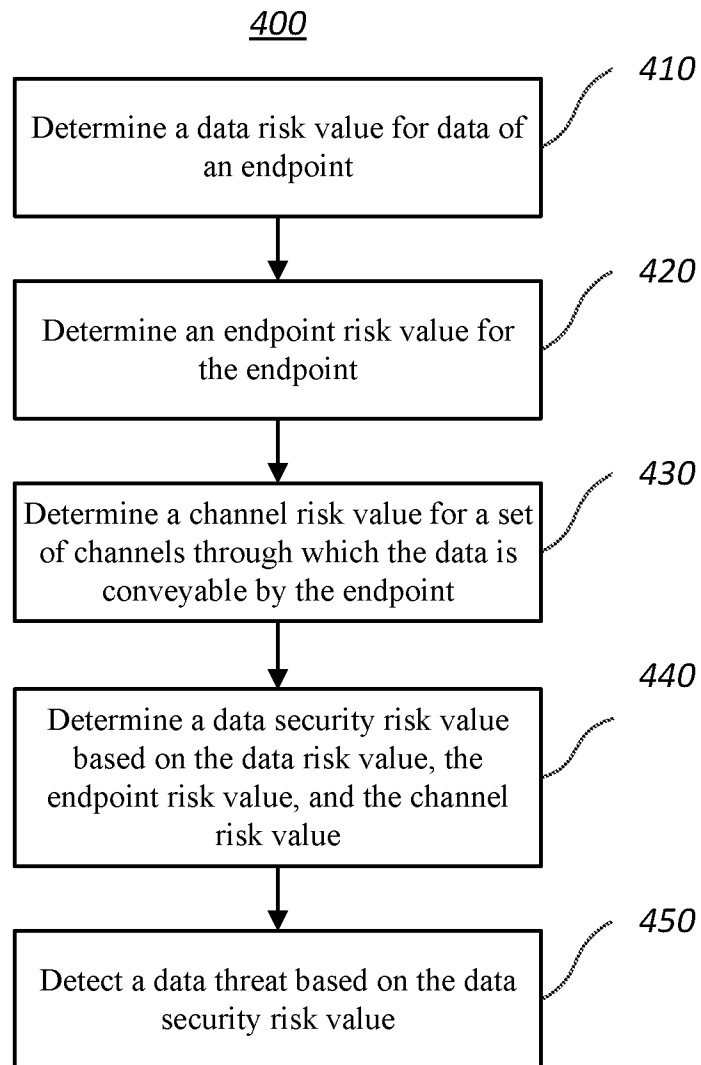


FIG. 3

4/5

**FIG. 4**

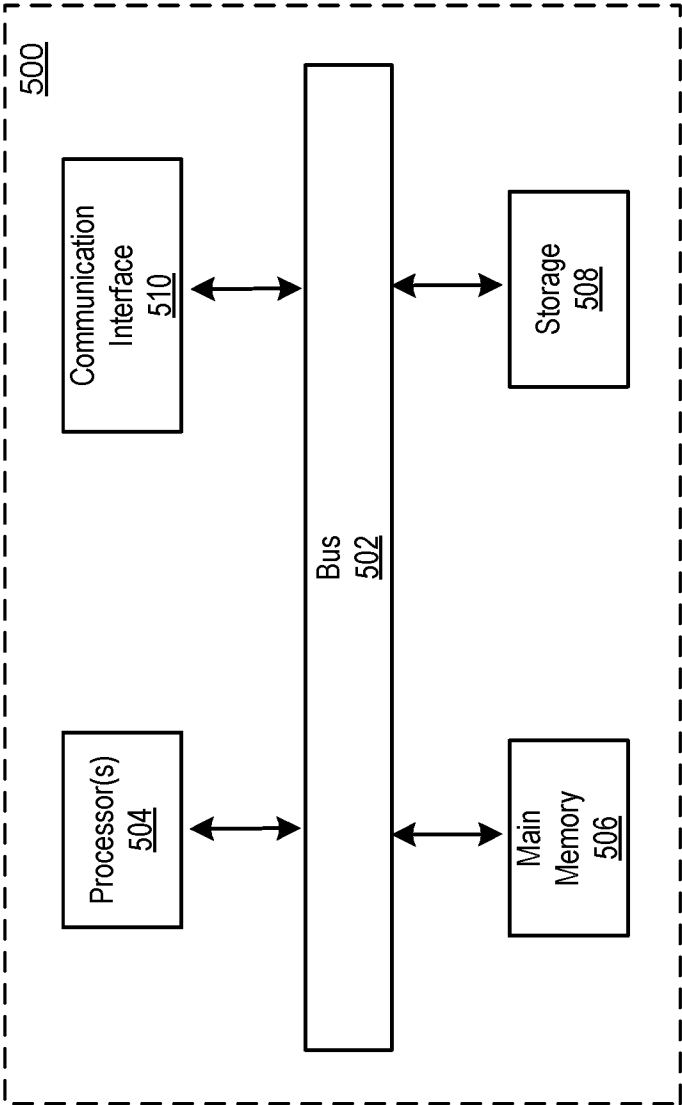


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/067983

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 12/14; G06F 21/53; G06F 21/56; G06F 21/57; H04L 12/24; H04L 12/26 (2019.01)

CPC - G06F 21/577; G06F 21/552; G06F 21/554; G06F 21/561; G06F 21/564; G06F 21/566; H04L 63/1425 (2019.02)

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)

See Search History document

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

USPC - 726/24 (keyword delimited)

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 7,865,958 B2 (LIEBLICH et al) 04 January 2011 (04.01.2011) entire document	1-20
A	US 7,418,733 B2 (CONNARY et al) 26 August 2008 (26.08.2008) entire document	1-20
A	US 2015/0205954 A1 (FILETREK INC.) 23 July 2015 (23.07.2015) entire document	1-20
A	US 2005/0114186 A1 (HEINRICH) 26 May 2005 (26.05.2005) entire document	1-20
A	US 2018/0018602 A1 (MCS2, LLC) 18 January 2018 (18.01.2018) entire document	1-20
A	US 2014/0325670 A1 (RIVENDALE SOFTWARE SOLUTION PRIVATE LIMITED) 30 October 2014 (30.10.2014) entire document	1-20

☐ 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

07 March 2019

Date of mailing of the international search report

25 MAR 2019

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774