

FILE LEAKAGE DETECTION METHOD AND DEVICE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the priority to the Chinese patent application No. 202111434476.2 entitled "FILE LEAKAGE DETECTION METHOD AND DEVICE" and filed on November 29, 2021, the disclosure of which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to the technical field of data security, in particular to a file leak detection method and apparatus.

BACKGROUND

[0003] Enterprise files, including project source codes, configuration files, documentation comments, as well as other related files and the like, are one of the core assets of an enterprise. However, enterprise files are confronted with various leak risks, such as stealing by external attackers, code uploading to an external platform by internal staff due to negligence, copying by employees who have left office, and the like, so timely detecting and preventing the leak of enterprise files is of significant importance for protecting core business and data security of enterprises.

SUMMARY

[0004] An embodiment of the present disclosure provides a technical solution as below:

[0005] In a first aspect, an embodiment of the present disclosure provides a file leak detection method, comprising:

[0006] acquiring a file operation event on a terminal device, wherein the file operation event is an event for performing a specified operation on a target file;

[0007] extracting, from the file operation event, a file path of the target file which the file operation event involves;

[0008] searching for file content of the target file in accordance with the file path, and performing a mapping process on the file content of the target file to obtain a file fingerprint of the target file, wherein the file fingerprint uniquely identifies the file content of the target file;

[0009] determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, wherein the specified file libraries are used for dynamically maintaining business files that need to be protected; and

[0010] determining that a file in the specified file libraries is being leaked, if the target file belongs to the specified file libraries.

[0011] As an optional implementation of the embodiment of the present disclosure, the determining whether the target file belongs

to specified file libraries in accordance with the file path and the file fingerprint comprises:

[0012] determining whether the file path is contained in an index information library; wherein the index information library stores a file path and a file fingerprint of at least one file in the specified file libraries;

[0013] determining that the target file belongs to the specified file libraries, if the file path is contained in the index information library;

[0014] determining whether the file fingerprint is contained in the index information library, if the file path is not contained in the index information library; and

[0015] determining that the target file belongs to the specified file libraries, if the file fingerprint is contained in the index information library.

[0016] As an optional implementation of the embodiment of the present disclosure, in a case wherein the file fingerprint is not contained in the index information library, the method further comprises:

[0017] determining whether the target file originates from preset file hosting platforms in accordance with whether a preset field is contained in the file path;

[0018] determining a first file library which is a file warehouse of an item to which the target file belongs, if the target file originates from the preset file hosting platforms;

[0019] determining whether the first file library belongs to the

specified file libraries; and

[0020] determining that the target file belongs to the specified file libraries, if the first file library belongs to the specified file libraries.

[0021] As an optional implementation of the embodiment of the present disclosure, in a case of determining that the first file library does not belong to the specified file libraries, the method further comprises:

[0022] determining whether a same-level file, a parent file of which is the same as that of the target file, originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a same-level path which is a path of the same-level file;

[0023] determining a second file library which is a file warehouse of an item to which the same-level file belongs, if the same-level file originates from the preset file hosting platforms;

[0024] determining whether the second file library belongs to the specified file libraries; and

[0025] determining that the target file belongs to the specified file libraries, if the second file library belongs to the specified file libraries.

[0026] As an optional implementation of the embodiment of the present disclosure, in a case of determining that the second file library does not belong to the specified file libraries, the method further comprises:

[0027] determining whether at least one level of parent file of

the target file originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a parent path which is a path of the parent file;

[0028] determining a third file library which is a file warehouse of an item to which the parent file originating from the preset file hosting platforms belongs, if any level of the parent file originates from the preset file hosting platforms;

[0029] determining whether the third file library belongs to the specified file libraries; and

[0030] determining that the target file belongs to the specified file libraries, if the third file library belongs to the specified file libraries.

[0031] As an optional implementation of the embodiment of the present disclosure, the method further comprises:

[0032] adding the file path and the file fingerprint into the index information library, in a case of determining that the target file belongs to the specified file libraries.

[0033] As an optional implementation of the embodiment of the present disclosure, prior to determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, the method further comprises:

[0034] receiving at least one file library address transmitted by a server;

[0035] determining the specified file libraries in accordance with the at least one file library address.

[0036] As an optional implementation of the embodiment of the

present disclosure, the method further comprises:

[0037] transmitting alarm information to a server, wherein the alarm information is used for indicating that a file in the specified file libraries is being leaked, in a case of determining that a file in the specified file libraries is being leaked.

[0038] As an optional implementation of the embodiment of the present disclosure, the method further comprises:

[0039] transmitting attribute information of the file operation event to a server, in a case of determining that a file in the specified file libraries is being leaked;

[0040] wherein the attribute information of the file operation event comprises at least one of identification information of the terminal device, identification information of the target file, and account information for logging in the terminal device.

[0041] As an optional implementation of the embodiment of the present disclosure, the method further comprises:

[0042] preventing the file operation event from being performed, in a case of determining that a file in the specified file libraries is being leaked.

[0043] In a second aspect, an embodiment of the present disclosure provides a file leak detection apparatus, comprising:

[0044] a detecting unit configured to acquire a file operation event on a terminal device, wherein the file operation event is an event for performing a specified operation on a target file;

[0045] an extracting unit configured to extract, from the file operation event, a file path of the target file which the file

operation event involves;

[0046] an acquiring unit configured to search for file content of the target file in accordance with the file path, and perform a mapping process on the file content of the target file to obtain a file fingerprint of the target file, wherein the file fingerprint uniquely identifies the file content of the target file;

[0047] a determining unit configured to determine whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, wherein the specified file libraries are used for dynamically maintaining business files that need to be protected; and

[0048] a processing unit configured to determine that a file in the specified file libraries is being leaked, if the target file belongs to the specified file libraries.

[0049] As an optional implementation of the embodiment of the present disclosure, the determining unit is specifically configured to determine whether the file path is contained in an index information library; determine that the target file belongs to the specified file libraries, if the file path is contained in the index information library; determine whether the file fingerprint is contained in the index information library, if the file path is not contained in the index information library; and determine that the target file belongs to the specified file libraries, if the file fingerprint is contained in the index information library

[0050] Herein, the index information library stores a file path

and a file fingerprint of at least one file in the specified file libraries.

[0051] As an optional implementation of the embodiment of the present disclosure, the determining unit is further configured to, in a case wherein the file fingerprint is not contained in the index information library, determine whether the target file originates from preset file hosting platforms in accordance with whether a preset field is contained in the file path; determine a first file library which is a file warehouse of an item to which the target file belongs, if the target file originates from the preset file hosting platforms; determine whether the first file library belongs to the specified file libraries; and that the target file belongs to the specified file libraries, if the first file library belongs to the specified file libraries.

[0052] As an optional implementation of the embodiment of the present disclosure, the determining unit is further configured to, in a case of determining that the first file library does not belong to the specified file libraries, determine whether a same-level file, a parent file of which is the same as that of the target file, originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a same-level path which is a path of the same-level file; determine a second file library which is a file warehouse of an item to which the same-level file belongs, if the same-level file originates from the preset file hosting platforms; determine whether the second file library belongs to the specified file libraries; and determine that the

target file belongs to the specified file libraries, if the second file library belongs to the specified file libraries.

[0053] As an optional implementation of the embodiment of the present disclosure, the determining unit is further configured to, in a case of determining that the second file library does not belong to the specified file libraries, determine whether at least one level of parent file of the target file originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a parent path which is a path of the parent file; determine a third file library which is a file warehouse of an item to which the parent file originating from the preset file hosting platforms belongs, if any level of the parent file originates from the preset file hosting platforms; determine whether the third file library belongs to the specified file libraries; and determine that the target file belongs to the specified file libraries, if the third file library belongs to the specified file libraries.

[0054] As an optional implementation of the embodiment of the present disclosure, the determining unit is further configured to the file path and the file fingerprint into the index information library, in a case of determining that the target file belongs to the specified file libraries.

[0055] As an optional implementation of the embodiment of the present disclosure, the determining unit is further configured to, prior to determining whether the target file belongs to specified file libraries in accordance with the file path, the file fingerprint, and the index information library, receive at least

one file library address transmitted by a server; and determine the specified file libraries in accordance with the at least one file library address.

[0056] As an optional implementation of the embodiment of the present disclosure, the processing unit is further configured to transmit alarm information to a server, wherein the alarm information is used for indicating that a file in the specified file libraries is being leaked, in a case of determining that a file in the specified file libraries is being leaked.

[0057] As an optional implementation of the embodiment of the present disclosure, the processing unit is further configured to transmit attribute information of the file operation event to a server, in a case of determining that a file in the specified file libraries is being leaked;

Wherein the attribute information of the file operation event comprises at least one of identification information of the terminal device, identification information of the target file, and account information for logging in the terminal device.

[0058] As an optional implementation of the embodiment of the present disclosure, the processing unit is further configured to prevent the file operation event from being performed, in a case of determining that a file in the specified file libraries is being leaked.

[0059] In a third aspect, an embodiment of the present disclosure provides an electronic device, comprising: a memory configured to store a computer program and a processor configured to, when calls

the computer program, cause the electronic device to implement the file leak detection method according to the first aspect or any one of the optional implementations of the first aspect.

[0060] In a fourth aspect, an embodiment of the present disclosure provides a non-transitory computer-readable storage medium, a computer program which, when executed by a processor, causes the processor to implement the file leak detection method according to the first aspect or any one of the optional implementations of the first aspect.

[0061] In a fifth aspect, an embodiment of the present disclosure provides a computer program product which, when executed on a computer, causes the computer to implement the file leak detection method according to the first aspect or any one of the optional implementations of the first aspect.

[0062] In a sixth aspect, an embodiment of the present disclosure provides a computer program which, when executed by a processor, causes the processor to implement the file leak detection method according to the first aspect or any one of the optional implementations of the first aspect.

BRIEF DESCRIPTION OF THE DRAWINGS

[0063] The accompanying drawings herein are incorporated in and form a part of the description; they illustrate embodiments consistent with the present disclosure and, together with the description, serve to explain the principles of the present disclosure.

[0064] In order to more clearly illustrate the technical solutions in the embodiments of the present disclosure or in the related art, the accompanying drawings that need to be used in the description of the embodiments or the related art will be briefly described below; for one of ordinary skill in the art, other related drawings can also be obtained in accordance with these drawings without paying creative efforts.

[0065] Fig. 1 shows a scene architecture diagram of a file leak detection method provided by an embodiment of the present disclosure;

[0066] Fig. 2 shows a first flowchart of steps of a file leak detection method provided by an embodiment of the present disclosure;

[0067] Fig. 3 shows a second flowchart of steps of a file leak detection method provided by an embodiment of the present disclosure;

[0068] Fig. 4 shows a third flowchart of steps of a file leak detection method provided by an embodiment of the present disclosure;

[0069] Fig. 5 shows a fourth flowchart of steps of a file leak detection method provided by an embodiment of the present disclosure;

[0070] Fig. 6 shows a fifth flowchart of steps of a file leak detection method provided by an embodiment of the present disclosure;

[0071] Fig. 7 shows a sixth flowchart of steps of a file leak

detection method provided by an embodiment of the present disclosure;

[0072] Fig. 8 shows a seventh flowchart of steps of a file leak detection method provided by an embodiment of the present disclosure;

[0073] Fig. 9 shows a eighth flowchart of steps of a file leak detection method provided by an embodiment of the present disclosure;

[0074] Fig. 10 shows a schematic diagram of a file leak detection apparatus provided by an embodiment of the present disclosure;

[0075] Fig. 11 shows a schematic hardware structure diagram of an electronic device provided by an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0076] In order that the above-mentioned objectives, features and advantages of the present disclosure may be understood in a more clearly manner, the solutions of the present disclosure will be further described as below. It should be noted that, in a case of no conflict, the embodiments and features in the embodiments of the present disclosure may be combined with each other.

[0077] In the following description, many specific details are set forth in order to facilitate a thorough understanding of the present disclosure, yet the present disclosure may be implemented in other ways different from those described herein; it is apparent that the embodiments in the description are merely part of the

embodiments of the present disclosure, rather than all of the embodiments.

[0078] It should be noted that, for the convenience of clearly describing the technical solutions of the embodiments of the present disclosure, in the embodiments of the present disclosure, wordings such as "first", "second", etc. are employed to distinguish the same items or similar items with basically the same functions and effects, and those skilled in the art can understand that wordings such as "first", "second", etc. do not make definitions on the quantity and execution order. For example, the first feature image set and the second feature image set are only used for distinguishing different feature image sets, rather than making definitions on the order of the feature image sets.

[0079] In the embodiments of the present disclosure, expressions such as "as an example" or "for example" are used to mean an example, an instance, or an illustration. Any embodiment or design described as "as an example" or "for example" in the embodiments of the present disclosure is not necessarily to be construed as more preferred or more advantageous than other embodiments or designs. Rather, use of the expressions such as "as an example" or "for example" is intended to present relevant concepts in a concrete fashion. Further, in the description of the embodiments of the present disclosure, the meaning of "a plurality" means two or more, unless otherwise specified.

[0080] The file leak detection method commonly used in the related art is as follows: identifying whether a file is a privacy file

by detecting whether the file contains specified keywords, and then determining whether file leak is occurring. However, in the process of identifying whether the file is the privacy file by whether specified keywords are contained, if the range of specified keywords is too narrow, failure to report will occur, so that the leak of the privacy file cannot be detected, and if the range of specified keywords is too wide, misreport will occur, which will affect the execution of the normal file operation event.

[0081] A file leak detection method provided by an embodiment of the present disclosure comprises: first, acquiring a file operation event on a terminal device, wherein the file operation event is an event for performing a specified operation on a target file; then, extracting, from the file operation event, a file path of the target file which the file operation event involves; subsequently, searching for file content of the target file in accordance with the file path, and performing a mapping process on the file content of the target file to obtain a file fingerprint of the target file; finally, determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, wherein the specified file libraries are used for dynamically maintaining business files that need to be protected; and, determining that a file in the specified file libraries is being leaked, if the target file belongs to the specified file libraries. Compared with the detection of file leak through specified keywords, the file leak detection method provided by the embodiment of the present disclosure can determine whether

a file belongs to specified file libraries in two dimensions, namely through a file path and a file fingerprint of the file, and therefore the embodiment of the present disclosure can reduce miss detection and false detection in the process of detecting whether a file belongs to specified file libraries or not, thereby improving the accuracy of file leak detection.

[0082] First, a scene architecture of a file leak detection method provided by an embodiment of the present disclosure is described as follows.

[0083] Referring to Fig. 1, a scene architecture of a file leak detection method provided by an embodiment of the present disclosure includes a server 11 and a terminal device 12.

[0084] Herein the server 11 includes a privacy file management module 111 and a communication module 112. The privacy file management module 111 has an Application Programming Interface (API) exposed to the outside. Through the API exposed to the outside of the privacy file management module 111, an administrator can configure the file library addresses where the documents which are not intended to be transmitted outside are located, such as a Git item address, a SVN item address, and the like, and perform operations such as addition, deletion or modification on the already configured file library addresses. The communication module 112 is configured to transmit the file library addresses configured by the administrator to the terminal device 11, and to receive information reported by the terminal device 11, such as alarm information that a file in specified file libraries is leaked,

attribute information of an operation event which leaks a file in specified file libraries, and the like.

[0085] The terminal device 12 includes an event monitoring module 121, a file retrieving module 122, an index information library 123, a file library identifying module 124, and a communication module 125. The event monitoring module 121 is configured to detect all process events on the terminal device, and determine whether each process event is an operation event that may leak a file in specified file libraries, such as outbound transmitting of a file, copying of a file, and the like, and to, when it is determined that a certain event is an event that may leak a file in specified file libraries, transmit relevant information of the operation event to the file retrieving module 122, so as to further determine whether a file in specified file libraries is being leaked. The file retrieving module 122 acquires a file path and a file fingerprint of a target file operated by the operation event that may leak a file in specified file libraries, determines whether the target file is a file in the specified file libraries based on the paths and the Hash values stored in the index information library 123, and determines that a file in the specified file libraries is being leaked, when the target file is determined to be a file in the specified file libraries. The file library identifying module 124 is configured to further determine whether a file warehouse of an item, to which a same-level file or a parent file of the target file belongs, belongs to the specified file libraries, when the file retrieving module 122 determines that the file path and the

file fingerprint of the target file are not contained in the index information library 123, and to determine that a file in the specified file libraries is being leaked when the file warehouse of the item, to which the same-level file or the parent file of the target file belongs, belongs to the specified file libraries, and to add the file path and the file fingerprint of the target file into the index information library 123. The communication module 125 is configured to receive the file library addresses transmitted by the server 11 and to transmit information, such as the alarm information, the attribute information of the file operation event, and the like to the server 11.

[0086] Based on the above contents, an embodiment of the present disclosure provides a file leak detection method applied to a terminal device, and referring to Fig. 2, the file leak detection method provided by the embodiment of the present disclosure comprises the following steps:

[0087] S21, acquiring a file operation event on a terminal device.

[0088] Herein, the file operation event is an event for performing a specified operation on a target file.

[0089] The specified operation in the embodiment of the present disclosure may include various file operations that may cause a file in specified file libraries to be leaked. Exemplarily, the specified operation may include: an operation of copying a file, an event of uploading a file to a network platform, an operation of transmitting a file to an external network, and the like.

[0090] The file in the embodiments of the present disclosure may

be any type of files. For example, the file may include a source code, a document, a configuration file, an image, a video file, an audio file, a compilation product, and the like.

[0091] Optionally, the terminal device may detect in real time whether each operation event on the terminal device is an operation event related to a file, further determine whether an operation which the operation event involves is a specified operation, when it is determined that a certain operation event is an operation event related to a file, and determine that the operation event is the file operation event in the embodiment of the present disclosure, when it is determined that the operation which the operation event involves is a specified operation.

[0092] S22, extracting, from the file operation event, a file path of the target file which the file operation event involves.

[0093] Specifically, the file path of the target file refers to a storage route through which the target file is read. For example, when a certain file entitled 'first draft' is read, the storage route to be went through sequentially includes: system disk C, Users folder, zhangsan folder, Desktop folder, KP2116216.4 folder, then the file path of the file is: C:\Users\zhangsan\Desktop\KP2116216.4\first draft.

[0094] S23, searching for file content of the target file in accordance with the file path, and performing a mapping process on the file content of the target file to obtain a file fingerprint of the target file.

[0095] Herein, the file fingerprint uniquely identifies the file

content of the target file.

[0096] Optionally, the file fingerprint of the target file may be a Hash value of the target file.

[0097] That is, first reading the target file in accordance with the file path of the target file, and then performing the mapping process on the file content of the target file to obtain a file fingerprint of the target file.

[0098] S24, determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint.

[0099] Herein, the specified file libraries are used for dynamically maintaining business files that need to be protected.

[00100] The business files that need to be protected in the embodiment of the present disclosure refer to files that are not intended to be leaked. Business files that need to be protected may include an enterprise source code, a document, a configuration file, an image, a video file, an audio file, a compilation product, and the like. The specified file libraries being used for dynamically maintaining business files that need to be protected means that, addition, modification, and deletion may be dynamically performed on the files in the specified file libraries, so as to dynamically manage the business files that need to be protected.

[00101] As an optional implementation of the embodiment of the present disclosure, the above step S24 (determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint) comprises:

[00102] determining whether the file path and the file fingerprint are contained in an index information library; wherein the index information library stores a file path and a file fingerprint of at least one file in the specified file libraries;

[00103] determining that the target file belongs to the specified file libraries, if the file path and/or the file fingerprint are/is contained in the index information library.

[00104] Optionally, the index information library may store the paths and the fingerprints of the files in the specified file libraries in the manner as shown in Table 1 below.

Table 1

File Name	File Path	File Hash Value
File A	File Path A	File Fingerprint A
File B	File Path B	File Fingerprint B
File C	File Path C	File Fingerprint C
⋮	⋮	⋮

[00105] As shown in Table 1 above, the index information library may be a Key-Value database, wherein the file name of the file is referred as the key and the path and the fingerprint of the file are stored as the value.

[00106] Optionally, an implementation for determining whether the target file belongs to specified file libraries in accordance with the file path, the file fingerprint, and the index information library may comprise:

[00107] traversing the file paths and the file fingerprints in the index database, determining that the target file belongs to the specified file libraries, if the file path and/or the file fingerprint are/is contained in the index database, and determining that the target file does not belong to the specified file libraries, if the file path or the file fingerprint is not contained in the index database.

[00108] It should be noted that, the path of a file in the embodiment of the present disclosure refers to a full path of the file. That is, the file path includes the name of the file, and thus determining whether the target file belongs to the specified file libraries in accordance with the file path, the file fingerprint, and the index information library may be understood as determining whether the target file belongs to the specified file libraries in accordance with the name of the target file, the file path, the file fingerprint, and the index information library.

[00109] As an optional implementation of the embodiment of the present disclosure, prior to the above step S24 (determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint), the file leak detection method provided by an embodiment of the present disclosure further comprises:

[00110] receiving at least one file library address transmitted by a server; and

[00111] determining the specified file libraries in accordance with the at least one file library address.

[00112] That is, the terminal device needs to first acquire, from the server, the configuration as to files in which file libraries need to be prevented from being leaked, and the specific acquisition manner is as follows: receiving the file library addresses transmitted by the server, and determining the files identified by the respective file library addresses as the specified file libraries.

[00113] In the above step S24, if it is determined that the target file belongs to the specified file libraries, the following step S25 is performed.

[00114] S25, determining that a file in the specified file libraries is being leaked.

[00115] A file leak detection method provided by an embodiment of the present disclosure comprises: first, acquiring a file operation event for performing a specified operation on a target file on a terminal device; then, extracting, from the file operation event, a file path of the target file which the file operation event involves; subsequently, searching for file content of the target file in accordance with the file path, and performing a mapping process on the file content of the target file to obtain a file fingerprint of the target file; finally, determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, wherein the specified file libraries are used for dynamically maintaining business files that need to be protected; and determining that a file in the specified file libraries is being leaked, if the target file belongs to the

specified file libraries. Compared with the detection of file leak through specified keywords, the file leak detection method provided by the embodiment of the present disclosure can determine whether a file belongs to specified file libraries in two dimensions, namely through a file path and a file fingerprint of the file, and therefore the embodiment of the present disclosure can reduce miss detection and false detection in the process of detecting whether a file belongs to specified file libraries or not, thereby improving the accuracy of file leak detection.

[00116] As an optional implementation of the embodiment of the present disclosure, on the basis of the embodiment as shown in Fig. 2, referring to Fig. 3, one implementation of the above step S24 (determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint) comprises the following steps S31 to S34:

[00117] S31, determining whether the file path is contained in the index information library.

[00118] Herein, the index information library stores a file path and a file fingerprint of at least one file in the specified file libraries.

[00119] Specifically, the respective file paths stored in the index information library may be traversed, and it is determined whether there is a file path that is completely the same as the file path in the index information library, so as to determine whether the file path is contained in the index information library.

[00120] In the above step S31, if it is determined that the file

path is contained in the index information library, the following step S32 is performed.

[00121] S32, determining that the target file belongs to the specified file libraries.

[00122] In the above step S31, if it is determined that the file path is not contained in the index information library, the following step S33 is performed.

[00123] S33, determining whether the file fingerprint is contained in the index information library.

[00124] Likewise, the respective file fingerprints stored in the index information library may be traversed, and it is determined whether there is a file fingerprint that is completely the same as the file fingerprint in the index information library, so as to determine whether the file fingerprint is contained in the index information library.

[00125] In the above step S33, if it is determined that the file fingerprint is contained in the index information library, the following step S34 is performed.

[00126] S34, determining that the target file belongs to the specified file libraries.

[00127] On one hand, when the file path is not contained in the index information library, the above embodiment will further determine whether the file fingerprint is contained in the index information library, and determine that the target file belongs to the specified file libraries, in a case wherein the file fingerprint is contained in the index information library, and therefore the above

embodiment can avoid the file in the specified file libraries from being first copied to another storage position (the file path is changed), and then being leaked by way of copying, outbound transmitting, or the like, thereby further improving security of the files in the specified file libraries.

[00128] On the other hand, it is the case in the embodiment of the present disclosure that whether a file path belongs to the retrieved database is detected first, and whether the file path belongs to the retrieved database is directly determined when detecting whether the file path belongs to the retrieved database, and whether a file fingerprint belongs to the retrieved database is not further determined, so that the above embodiment can further reduce the data processing amount in the file retrieval process, thereby improving the efficiency of data file leak detection.

[00129] As an optional implementation of the embodiment of the present disclosure, on the basis of the embodiment as shown in Fig. 3, referring to Fig. 4, in a case wherein the file fingerprint is not contained in the index information library, the file leak detection method provided by the embodiment of the present disclosure further comprises the following steps S41 to S44:

[00130] S41, determining whether the target file originates from preset file hosting platforms, in accordance with whether a preset field is contained in the file path.

[00131] Exemplarily, the preset field can include: .git and .svn.

[00132] Specifically, based on the demands on research and development and on operation and maintenance, enterprises often

upload files to an internal file hosting platform, which files can be directly pulled from the internal file hosting platform when demands for using these files arise, and the paths of the files pulled from the file hosting platform will all be written into corresponding fields. For example, if a code is hosted to the git platform or the svn platform, the field ". git" or ". svn" will be contained in the path of the code pulled from the git platform or the svn platform.

[00133] Specifically, determining whether the target file originates from preset file hosting platforms, in accordance with whether a preset field is contained in the file path, comprises:

[00134] determining that the target file originates from preset file hosting platforms, when the preset field is contained in the file path;

[00135] determining that the target file does not originate from preset file hosting platforms, when the preset field is not contained in the file path.

[00136] In the above step S41, if it is determined that the target file originates from the preset file hosting platforms, the following step S42 is performed.

[00137] S42, determining a first file library.

[00138] Herein, the first file library is a file warehouse of an item to which the target file belongs.

[00139] Specifically, when the preset field is .git, the file warehouse address of the item to which the target file belongs can be acquired by analyzing the .git directory, so that a first

warehouse is determined, and when the preset field is .svn, the file warehouse address of the item to which the target file belongs can be acquired by analyzing the .svn directory, so that a first warehouse is determined.

[00140] S43, determining whether the first file library belongs to the specified file libraries.

[00141] Specifically, whether the first file library belongs to the specified file libraries or not can be determined by comparing whether the address of the first file library belongs to the addresses of the specified file libraries or not.

[00142] In the above step S43, if it is determined that the first file library belongs to the specified file libraries, the following step S44 is performed.

[00143] S44, determining that the target file belongs to the specified file libraries.

[00144] In a case wherein the file fingerprint is not contained in the index information library, the above embodiment further determines whether the target file originates from preset file hosting platforms, in accordance with whether a preset field is contained in the file path; obtains the file warehouse of the item to which the target file belongs, in a case wherein the target file originates from preset file hosting platforms; determines whether the file warehouse of the item, to which the target file belongs, belongs to the specified file libraries; and determines that the target file belongs to the specified file libraries, in a case wherein the file warehouse of the item, to which the target file

belongs, belongs to the specified file libraries; therefore, the above embodiment can prevent a sensitive file in the specified file libraries from being first uploaded to the file hosting platform and then downloaded from the file hosting platform and leaked, thereby realizing a more comprehensive detection of the leak action of the file in the specified file libraries.

[00145] As an optional implementation of the embodiment of the present disclosure, the file leak detection method provided by the embodiment of the present disclosure further comprises:

[00146] adding the file path and the file fingerprint into the index information library, in a case of determining that the first file library belongs to the specified file libraries.

[00147] Adding the file path and the file fingerprint into the index information library when determining that the target file belongs to the specified file libraries, may facilitate subsequent detection of leak actions of the files in the specified file libraries directly in accordance with the index information library, thereby improving the efficiency in file leak detection.

[00148] As an optional implementation of the embodiment of the present disclosure, on the basis of the embodiment as shown in Fig. 4, referring to Fig. 5, in a case of determining that the first file library does not belong to the specified file libraries, the file leak detection method provided by the embodiment of the present disclosure further comprises the following steps S51 to S54:

[00149] S51, determining whether a same-level file originates from the preset file hosting platforms, in accordance with whether the

preset field is contained in a same-level path.

[00150] Herein, the same-level path is a path of the same-level file, a parent file of which is the same as that of the target file.

[00151] For example, the file "first draft" is the target file in the embodiment of the present disclosure, the file path of the file "first draft" is C:\Users\zhangsan\Desktop\KP2116216.4\first draft, and the parent file of the file "first draft" is KP2116216.4; and if a file with the file name "second draft" is further present under the parent file KP2116216.4 of the file "first draft", the file "second draft" is the same-level file of the file "first draft", and the path C:\Users\zhangsan\Desktop\KP2116216.4\second draft of the file "second draft" is the same-level path in the embodiment of the present disclosure.

[00152] In the above step S51, if it is determined that the same-level file originates from the preset file hosting platforms, the following step S52 is performed.

[00153] S52, determining a second file library.

[00154] Herein, the second file library is a file warehouse of an item to which the same-level file belongs.

[00155] S53, determining whether the second file library belongs to the specified file libraries.

[00156] In the above step S53, if it is determined that the second file library belongs to the specified file libraries, the following step S54 is performed.

[00157] S54, determining that the target file belongs to the specified file libraries.

[00158] In a case of determining that the first file library does not belong to the specified file libraries, the above embodiment further determines whether the same-level file originates from preset file hosting platforms in accordance with whether the preset field is contained in the same-level path; obtains the file warehouse of the item to which the same-level file belongs, in a case wherein the same-level file originates from the preset file hosting platforms; determines whether the file warehouse of the item, to which the same-level file belongs, belongs to the specified file libraries; and determines that the target file belongs to the specified file libraries, in a case wherein the file warehouse of the item, to which the same-level file belongs, belongs to the specified file libraries, therefore, the above embodiment can realize a more comprehensive detection of the leak actions of the files in the specified file libraries.

[00159] As an optional implementation of the embodiment of the present disclosure, the file leak detection method provided by the embodiment of the present disclosure further comprises:

[00160] adding the file path and the file fingerprint into the index information library, in a case of determining that the second file library belongs to the specified file libraries.

[00161] Adding the file path and the file fingerprint into the index information library when determining that the second file library belongs to the specified file libraries, may facilitate subsequent detection of leak actions of the files in the specified file libraries directly in accordance with the index information

library, thereby improving the efficiency in file leak detection.

[00162] As an optional implementation of the embodiment of the present disclosure, on the basis of the embodiment as shown in Fig. 5, referring to Fig. 6, in a case of determining that the second file library does not belong to the specified file libraries, the file leak detection method provided by the embodiment of the present disclosure further comprises the following steps S61 to S64:

[00163] S61, determining whether at least one level of parent file originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a parent path.

[00164] Herein, the parent path is a path of the parent file of the target file.

[00165] Exemplarily, at least one level of parent file may be three levels. That is, sequentially looking up three levels in directory upward.

[00166] In the above step S61, if it is determined that any level of parent file originates from the preset file hosting platforms, the following step S62 is performed.

[00167] S62, determining a third file library.

[00168] Herein, the third file library is a file warehouse of an item to which the parent file originating from the preset file hosting platforms belongs.

[00169] S63, determining whether the third file library belongs to the specified file libraries.

[00170] In the above step S63, if it is determined that the third file library belongs to the specified file libraries, the following

step S64 is performed.

[00171] S64, determining that the target file belongs to the specified file libraries.

[00172] In a case of determining that the second file library does not belong to the specified file libraries, the above embodiment further determines whether the preset field is contained in at least one level of parent path; obtains the file warehouse of the item to which the parent file corresponding to this level of parent path belongs, in a case wherein the preset field is contained in at least one level of parent path; determines whether the file warehouse of the item, to which the parent file belongs, belongs to the specified file libraries; and determines that the target file belongs to the specified file libraries, in a case wherein the file warehouse of the item, to which the parent file belongs, belongs to the specified file libraries. Therefore, the above embodiment can realize a more comprehensive detection of the leak actions of the files in the specified file libraries.

[00173] As an optional implementation of the embodiment of the present disclosure, the file leak detection method provided by the embodiment of the present disclosure further comprises:

[00174] adding the file path and the file fingerprint into the index information library, in a case of determining that the third file library belongs to the specified file libraries.

[00175] Adding the file path and the file fingerprint into the index information library when determining that the third file library belongs to the specified file libraries, may facilitate subsequent

detection of leak actions of the files in the specified file libraries directly in accordance with the index information library, thereby improving the efficiency in file leak detection.

[00176] As an optional implementation of the embodiment of the present disclosure, on the basis of the embodiment as shown in Fig. 1, referring to Fig. 7, in a case of determining that a file in the specified file libraries is being leaked, the file leak detection method provided by the embodiment of the present disclosure further comprises:

[00177] S71, transmitting alarm information to a server.

[00178] Herein, the alarm information is used for indicating that a file in the specified file libraries is being leaked.

[00179] Timely transmitting alarm information to the server upon realizing that a file in the specified file libraries is being leaked may cause the information security personnel to stop the leak actions in time.

[00180] As an optional implementation of the embodiment of the present disclosure, on the basis of the embodiment as shown in Fig. 1, referring to Fig. 8, in a case of determining that a file in the specified file libraries is being leaked, the file leak detection method provided by the embodiment of the present disclosure further comprises:

[00181] S81, transmitting attribute information of the file operation event to a server.

[00182] Herein, the attribute information of the file operation event comprises at least one of identification information of the

terminal device, identification information of the target file, and account information for logging in the terminal device.

[00183] Upon realizing that a file in the specified file libraries is being leaked, the above embodiment transmits the identification information of the terminal device, the identification information of the target file, and the account information for logging in the terminal device to the server; therefore, the above embodiment can facilitate investigation and affixing of the responsibilities for the leak actions.

[00184] As an optional implementation of the embodiment of the present disclosure, on the basis of the embodiment as shown in Fig. 1, referring to Fig. 9, in a case of determining that a file in the specified file libraries is being leaked, the file leak detection method provided by the embodiment of the present disclosure further comprises:

[00185] S91, preventing the file operation event from being performed.

[00186] Upon realizing that a file in the specified file libraries is being leaked, the above embodiment prevents the file operation event from being performed; therefore, the above embodiment can timely prevent the files in the specified file libraries from being leaked, upon realizing that a file in the specified file libraries is being leaked.

[00187] Based on the same inventive concept, as an implementation of the above method, the embodiment of the present disclosure further provides a file leak detection apparatus which corresponds

to the above-mentioned method embodiment; for the sake of convenience in reading, details in the above-mentioned method embodiment are not described again in the apparatus embodiment, but it should be clear that the file leak detection apparatus in the present embodiment can correspondingly implement all contents in the above-mentioned method embodiment.

[00188] The embodiment of the present disclosure provides a file leak detection apparatus, and Fig. 10 shows a schematic diagram of the file leak detection apparatus; as shown in Fig. 10, the file leak detection apparatus 100 comprises:

[00189] a detecting unit 101 configured to acquire a file operation event on a terminal device, wherein the file operation event is an event for performing a specified operation on a target file;

[00190] an extracting unit 102 configured to extract, from the file operation event, a file path of the target file which the file operation event involves;

[00191] an acquiring unit 103 configured to search for file content of the target file in accordance with the file path, and perform a mapping process on the file content of the target file to obtain a file fingerprint of the target file, wherein the file fingerprint uniquely identifies the file content of the target file;

[00192] a determining unit 104 configured to determine whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, wherein the specified file libraries are used for dynamically maintaining business files that need to be protected; and

[00193] a processing unit 105 configured to determine that a file in the specified file libraries is being leaked, in a case wherein the target file belongs to the specified file libraries.

[00194] As an optional implementation of the embodiment of the present disclosure, the determining unit 104 is specifically configured to determine whether the file path is contained in an index information library; determine that the target file belongs to the specified file libraries, if the file path is contained in the index information library; determine whether the file fingerprint is contained in the index information library, if the file path is not contained in the index information library; and determine that the target file belongs to the specified file libraries, if the file fingerprint is contained in the index information library

[00195] Herein, the index information library stores a file path and a file fingerprint of at least one file in the specified file libraries.

[00196] As an optional implementation of the embodiment of the present disclosure, the determining unit 104 is further configured to, in a case wherein the file fingerprint is not contained in the index information library, determine whether the target file originates from preset file hosting platforms in accordance with whether a preset field is contained in the file path; determine a first file library which is a file warehouse of an item to which the target file belongs, if the target file originates from the preset file hosting platforms; determine whether the first file

library belongs to the specified file libraries; and to determine that the target file belongs to the specified file libraries, if the first file library belongs to the specified file libraries.

[00197] As an optional implementation of the embodiment of the present disclosure, the determining unit 104 is further configured to, in a case of determining that the first file library does not belong to the specified file libraries, determine whether a same-level file, a parent file of which is the same as that of the target file, originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a same-level path which is a path of the same-level file; determine a second file library which is a file warehouse of an item to which the same-level file belongs, if the same-level file originates from the preset file hosting platforms; determine whether the second file library belongs to the specified file libraries; and determine that the target file belongs to the specified file libraries, if the second file library belongs to the specified file libraries.

[00198] As an optional implementation of the embodiment of the present disclosure, the determining unit 104 is further configured to, in a case of determining that the second file library does not belong to the specified file libraries, determine whether at least one level of parent file of the target file originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a parent path which is a path of the parent file; determine a third file library which is a file warehouse of an item to which the parent file originating from the preset

file hosting platforms belongs, if any level of parent file originates from the preset file hosting platforms; determine whether the third file library belongs to the specified file libraries; and determine that the target file belongs to the specified file libraries, if the third file library belongs to the specified file libraries.

[00199] As an optional implementation of the embodiment of the present disclosure, the determining unit 104 is further configured to add the file path and the file fingerprint into the index information library, in a case of determining that the target file belongs to the specified file libraries.

[00200] As an optional implementation of the embodiment of the present disclosure, the determining unit 104 is further configured to, prior to determining whether the target file belongs to the specified file libraries in accordance with the file path, the file fingerprint, and the index information library, receive at least one file library address transmitted by a server; and determine the specified file libraries in accordance with the at least one file library address.

[00201] As an optional implementation of the embodiment of the present disclosure, the processing unit 105 is further configured to transmit alarm information to a server, in a case of determining that a file in the specified file libraries is being leaked; wherein the alarm information is used for indicating that a file in the specified file libraries is being leaked.

[00202] As an optional implementation of the embodiment of the

present disclosure, the processing unit 105 is further configured to transmit attribute information of the file operation event to a server, in a case of determining that a file in the specified file libraries is being leaked;

[00203] wherein the attribute information of the file operation event comprises at least one of identification information of the terminal device, identification information of the target file, and account information for logging in the terminal device.

[00204] As an optional implementation of the embodiment of the present disclosure, the processing unit 105 is further configured to prevent the file operation event from being performed, in a case of determining that the target file belongs to the specified file libraries.

[00205] The file leak detection apparatus provided by the present embodiment may perform the file leak detection method provided by the above-mentioned method embodiment, and the implementation principles and the technical effects thereof are similar, which are not described herein again.

[00206] Based on the same inventive concept, an embodiment of the present disclosure further provides an electronic device. Fig. 11 is a schematic structural diagram of an electronic device provided by the embodiment of the present disclosure, and as shown in Fig. 11, the electronic device provided in the present embodiment comprises: a memory 111 configured to store a computer program and a processor 112 configured to perform the file leak detection method provided by the above-mentioned embodiment when the computer

program is called. The electronic device in the embodiment of the present disclosure may include, but are not limited to, a mobile terminal, such as a mobile phone, a laptop, a digital broadcast receiver, a PDA (personal digital assistant), a PAD (tablet computers), a PMP (portable multimedia player), an on-board terminal (e.g., an on-board navigation terminal), and the like, as well as a fixed terminal, such as a digital TV, a desktop computer, and the like.

[00207] Based on the same inventive concept, an embodiment of the present disclosure further provides a non-transitory computer-readable storage medium having stored thereon a computer program which, when executed by a processor, causes the processor to implement the file leak detection method provided by the above-mentioned embodiment.

[00208] Based on the same inventive concept, an embodiment of the present disclosure further provides a computer program product which, when executed on a computer, causes the computer to implement the file leak detection method provided by the above-mentioned embodiment.

[00209] Based on the same inventive concept, an embodiment of the present disclosure further provides a computer program which, when executed by a processor, causes the processor to implement the file leak detection method provided by the above-mentioned embodiment.

[00210] As shall be appreciated by those skill in the art, embodiments of the present disclosure may be provided as a method, a system, or a computer program product. Accordingly, the present

disclosure may take the form of an entirely hardware embodiment, an entirely software embodiment, or an embodiment combining software and hardware aspects. Furthermore, the present disclosure may take the form of a computer program product implemented on one or more computer-usable storage medium with computer-usable program codes contained therein.

[00211] The processor may be a Central Processing Unit (CPU), or may be another general processor, a Digital Signal Processor (DSP), an Application Specific Integrated Circuit (ASIC), a Field-Programmable Gate Array (FPGA) or another programmable logic device, a discrete gate or transistor logic device, a discrete hardware component, or the like. A general processor may be a microprocessor or the processor may be any conventional processor or the like.

[00212] The memory may include forms of volatile memory, Random Access Memory (RAM) and/or non-volatile memory, such as Read Only Memory (ROM) or flash memory (flash RAM), in a computer readable medium. The memory is an example of a computer-readable medium.

[00213] Computer readable medium includes permanent and non-permanent, movable and non-movable storage medium. Storage medium may achieve information storage by any method or technology, and information may be computer-readable instructions, a data structure, a unit of a program, or other data. Examples of computer storage medium include, but are not limited to, phase change memory (PRAM), Static Random Access Memory (SRAM), Dynamic Random Access Memory (DRAM), other types of Random Access Memory (RAM), Read Only

Memory (ROM), Electrically Erasable Programmable Read Only Memory (EEPROM), flash memory or other memory technology, Compact Disc Read Only Memory (CD-ROM), Digital Versatile Disks (DVD) or other optical storage, magnetic cassettes, magnetic disk storage or another magnetic storage device, or any other non-transmission medium, which can be used to store information that can be accessed by a computing device. As defined herein, computer-readable medium do not include transitory computer-readable medium, such as a modulated data signal and a carrier wave.

[00214] Finally, it should be noted that, all the above embodiments are only intended to be used for illustrating the technical solutions of the present disclosure, and not for limiting the same; although the present disclosure has been described in detail with reference to the above-mentioned embodiments, those of ordinary skill in the art shall understand that the technical solutions described in the above-mentioned embodiments may still be modified, or some or all of the technical features may be equivalently replaced; however, such modifications or replacements will not cause the essence of the corresponding technical solutions to depart from the scope of the technical solutions of the embodiments of the present disclosure.

CLAIMS:

1. A file leak detection method, characterized by comprising:

acquiring a file operation event on a terminal device, wherein the file operation event is an event for performing a specified operation on a target file;

extracting, from the file operation event, a file path of the target file which the file operation event involves;

searching for file content of the target file in accordance with the file path, and performing a mapping process on the file content of the target file to obtain a file fingerprint of the target file, wherein the file fingerprint uniquely identifies the file content of the target file;

determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, wherein the specified file libraries are used for dynamically maintaining business files that need to be protected; and

determining that a file in the specified file libraries is being leaked, if the target file belongs to the specified file libraries.

2. The method according to claim 1, characterized in that the determining whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint comprises:

determining whether the file path is contained in an index information library; wherein the index information library stores a file path and a file fingerprint of at least one file in the specified file libraries;

determining that the target file belongs to the specified file libraries, if the file path is contained in the index information library;

determining whether the file fingerprint is contained in the index information library, if the file path is not contained in the index information library; and

determining that the target file belongs to the specified file libraries, if the file fingerprint is contained in the index information library.

3. The method according to claim 2, characterized in that in a case wherein the file fingerprint is not contained in the index information library, the method further comprises:

determining whether the target file originates from preset file hosting platforms in accordance with whether a preset field is contained in the file path;

determining a first file library which is a file warehouse of an item to which the target file belongs, if the target file originates from the preset file hosting platforms;

determining whether the first file library belongs to the specified file libraries; and

determining that the target file belongs to the specified file libraries, if the first file library belongs to the specified file libraries.

4. The method according to claim 3, characterized in that in a case of determining that the first file library does not belong to the specified file libraries, the method further comprises:

determining whether a same-level file, a parent file of which is the same as that of the target file, originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a same-level path which is a path of the same-level file;

determining a second file library which is a file warehouse of an item to which the same-level file

belongs, if the same-level file originates from the preset file hosting platforms;

determining whether the second file library belongs to the specified file libraries; and

determining that the target file belongs to the specified file libraries, if the second file library belongs to the specified file libraries.

5. The method according to claim 4, characterized in that in a case of determining that the second file library does not belong to the specified file libraries, the method further comprises:

determining whether at least one level of parent file of the target file originates from the preset file hosting platforms, in accordance with whether the preset field is contained in a parent path which is a path of the parent file;

determining a third file library which is a file warehouse of an item to which the parent file originating from the preset file hosting platforms belongs, if any level of the parent file originates from the preset file hosting platforms;

determining whether the third file library belongs to the specified file libraries; and

determining that the target file belongs to the specified file libraries, if the third file library belongs to the specified file libraries.

6. The method according to any one of claims 3 to 5, characterized in that the method further comprises:

adding the file path and the file fingerprint into the index information library, in a case of determining that the target file belongs to the specified file libraries.

7. The method according to any one of claims 1 to 5, characterized in that the method further comprises:

transmitting alarm information to a server, wherein the alarm information is used for indicating that a file in the specified file libraries is being leaked, in a case of determining that a file in the specified file libraries is being leaked.

8. The method according to any one of claims 1 to 5, characterized in that the method further comprises:

transmitting attribute information of the file operation event to a server, in a case of determining that a file in the specified file libraries is being leaked;

wherein the attribute information of the file operation event comprises at least one of identification information of the terminal device, identification information of the target file, and account information for logging in the terminal device.

9. The method according to any one of claims 1 to 5, characterized in that the method further comprises:

preventing the file operation event from being performed, in a case of determining that a file in the specified file libraries is being leaked.

10. A file leak detection apparatus, characterized by comprising:

a detecting unit configured to acquire a file operation event on a terminal device, wherein the file operation event is an event for performing a specified operation on a target file;

an extracting unit configured to extract, from the file operation event, a file path of the target file which the file operation event involves;

an acquiring unit configured to search for file content of the target file in accordance with the file path, and perform a mapping process on the file content of the target file to obtain a file fingerprint of the target file, wherein the file fingerprint uniquely identifies the file content of the target file;

a determining unit configured to determine whether the target file belongs to specified file libraries in accordance with the file path and the file fingerprint, wherein the specified file libraries are used for dynamically maintaining business files that need to be protected; and

a processing unit configured to determine that a file in the specified file libraries is being leaked, if the target file belongs to the specified file libraries.

11. An electronic device, characterized by comprising: a memory configured to store a computer program and a processor configured to, when executes the computer program, cause the electronic device to implement the file leak detection method according to any one of claims 1 to 9.

12. A computer-readable storage medium characterized by having stored thereon a computer program which, when executed by a computing device, causes the computing device to implement the file leak detection method according to any one of claims 1 to 9.