

一种文件泄露检测方法及装置

相关申请的交叉引用

本申请要求于 2021 年 11 月 29 日提交的名称为“一种文件泄露检测方法及装置”的中国专利申请第 202111434476.2 号的优先权，该申请的公开通过引用被全部结合于此。

5 技术领域

本公开涉及数据安全技术领域，尤其涉及一种文件泄露检测方法及装置。

背景技术

企业文件包括项目源代码、配置文件、文档注释以及其他相关文件等，是企业的核心资产之一。然而企业文件面临外部攻击者盗取、内部员工因疏忽上传代码到外部平台、离职员工拷贝等多种泄漏风险，因此及时检测和阻止企业文件泄漏对保护企业核心业务和数据安全非常重要。

发明内容

本公开实施例提供技术方案如下：

第一方面，本公开的实施例提供了一种文件泄露检测方法，包括：

15 获取终端设备上的文件操作事件，所述文件操作事件为对目标文件执行指定操作的事件；

从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径；

根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内容进行映射处理得到所述目标文件的文件指纹，所述文件指纹唯一标识所述目标文件的文件内容；

20 根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库，所述指定文件库用于动态维护需要保护的库文件；

若所述目标文件属于所述指定文件库，则确定所述指定文件库中的文件正在发生泄漏。

作为本公开实施例一种可选的实施方式，所述根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库，包括：

25 确定索引信息库中是否包含所述文件路径；所述索引信息库存储有所述指定文件库中的至少一个文件的文件路径和文件指纹；

若所述索引信息库中包含所述文件路径，则确定所述目标文件属于所述指定文件库；

若所述索引信息库中不包含所述文件路径，则确定所述索引信息库中是否包含所述文件指纹；

若所述索引信息库中包含所述文件指纹，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，在所述索引信息库中不包含所述文件指纹的

5 情况下，所述方法还包括：

根据所述文件路径中是否包含预设字段确定所述目标文件是否来自于预设文件托管平台；

若所述目标文件来自于所述预设文件托管平台，则确定第一文件库，所述第一文件库为所述目标文件所属项目的文件仓库；

10 确定所述第一文件库是否属于所述指定文件库；

若第一文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，在确定所述第一文件库不属于所述指定文件库的情况下，所述方法还包括：

15 根据同级路径中是否包含所述预设字段确定同级文件是否来自于所述预设文件托管平台，所述同级路径为父文件与所述目标文件相同的同级文件的路径；

若所述同级文件来自于所述预设文件托管平台，则确定第二文件库，所述第二文件库为所述同级文件所属项目的文件仓库；

确定所述第二文件库是否属于所述指定文件库；

若第二文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

20 作为本公开实施例一种可选的实施方式，在确定所述第二文件库不属于所述指定文件库的情况下，所述方法还包括：

根据父路径中是否包含所述预设字段同确定至少一级父文件是否来自于所述预设文件托管平台，所述父路径为所述目标文件的父文件的路径；

25 若任一级父文件来自于所述预设文件托管平台，则确定第三文件库，所述第三文件库为来自于所述预设文件托管平台的父文件所属项目的文件仓库；

确定所述第三文件库是否属于所述指定文件库；

若第三文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，所述方法还包括：

在确定所述目标文件属于所述指定文件库的情况下，将所述文件路径和所述文件指纹

添加到所述索引信息库中。

作为本公开实施例一种可选的实施方式，在根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库之前，所述方法还包括：

接收服务器发送的至少一个文件库地址；

- 5 根据所述至少一个文件库地址确定所述指定文件库。

作为本公开实施例一种可选的实施方式，所述方法还包括：

在确定所述指定文件库中的文件正在发生泄漏的情况下，向服务器发送报警信息，所述报警信息用于指示所述指定文件库中的文件正在发生泄漏。

作为本公开实施例一种可选的实施方式，所述方法还包括：

- 10 在确定所述指定文件库中的文件正在发生泄漏的情况下，向服务器发送所述文件操作事件的属性信息；

其中，所述文件操作事件的属性信息包括：所述终端设备的标识信息、所述目标文件的标识信息、登录所述终端设备的账户信息中的至少一个。

作为本公开实施例一种可选的实施方式，所述方法还包括：

- 15 在确定所述指定文件库中的文件正在发生泄漏的情况下，阻止所述文件操作事件的执行。

第二方面，本公开的实施例提供了一种文件泄露检测装置，包括：

检测单元，用于获取终端设备上的文件操作事件，所述文件操作事件为对目标文件执行指定操作的事件；

- 20 提取单元，用于从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径；

获取单元，用于根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内容进行映射处理得到所述目标文件的文件指纹，所述文件指纹唯一标识所述目标文件的文件内容；

- 25 确定单元，用于根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库，所述指定文件库用于动态维护需要保护的業務文件；

处理单元，用于在所述目标文件属于所述指定文件库的情况下，确定所述指定文件库中的文件正在发生泄漏。

作为本公开实施例一种可选的实施方式，所述确定单元，具体用于确定索引信息库中

是否包含所述文件路径；若所述索引信息库中包含所述文件路径，则确定所述目标文件属于所述指定文件库；若所述索引信息库中不包含所述文件路径，则确定所述索引信息库中是否包含所述文件指纹；若所述索引信息库中包含所述文件指纹，则确定所述目标文件属于所述指定文件库

- 5 其中，所述索引信息库存储有所述指定文件库中的至少一个文件的文件路径和文件指纹。

作为本公开实施例一种可选的实施方式，所述确定单元，还用于在所述索引信息库中不包含所述文件指纹的情况下，根据所述文件路径中是否包含预设字段确定所述目标文件是否来自于预设文件托管平台；若所述目标文件来自于所述预设文件托管平台，则确定第
10 一文件库，所述第一文件库为所述目标文件所属项目的文件仓库；确定所述第一文件库是否属于所述指定文件库；若第一文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，所述确定单元，还用于在确定所述第一文件库不属于所述指定文件库的情况下，根据同级路径中是否包含所述预设字段确定同级文件
15 是否来自于所述预设文件托管平台，所述同级路径为父文件与所述目标文件相同的同级文件的路径；若所述同级文件来自于所述预设文件托管平台，则确定第二文件库，所述第二文件库为所述同级文件所属项目的文件仓库；确定所述第二文件库是否属于所述指定文件库；若第二文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，所述确定单元，还用于在确定所述第二文件
20 库不属于所述指定文件库的情况下，根据父路径中是否包含所述预设字段同确定至少一级父文件是否来自于所述预设文件托管平台，所述父路径为所述目标文件的父文件的路径；若任一级父文件来自于所述预设文件托管平台，则确定第三文件库，所述第三文件库为来自于所述预设文件托管平台的父文件所属项目的文件仓库；确定所述第三文件库是否属于所述指定文件库；若第三文件库属于所述指定文件库，则确定所述目标文件属于所述指定
25 文件库。

作为本公开实施例一种可选的实施方式，所述确定单元，还用于在确定所述目标文件属于所述指定文件库的情况下，将所述文件路径和所述文件指纹添加到所述索引信息库中。

作为本公开实施例一种可选的实施方式，所述确定单元，还用于在根据所述文件路径、所述文件指纹以及所述索引信息库确定所述目标文件是否属于所述指定文件库之前，接收

服务器发送的至少一个文件库地址；根据所述至少一个文件库地址确定所述指定文件库。

作为本公开实施例一种可选的实施方式，所述处理单元，还用于在确定所述指定文件库中的文件正在发生泄漏的情况下，向服务器发送报警信息；所述报警信息用于指示所述指定文件库中的文件正在发生泄漏。

- 5 作为本公开实施例一种可选的实施方式，所述处理单元，还用于在确定所述指定文件库中的文件正在发生泄漏的情况下，向服务器发送所述文件操作事件的属性信息；

其中，所述文件操作事件的属性信息包括：所述终端设备的标识信息、所述目标文件的标识信息、登录所述终端设备的账户信息中的至少一个。

- 10 作为本公开实施例一种可选的实施方式，所述处理单元，还用于在确定所述目标文件属于所述指定文件库的情况下，阻止所述文件操作事件的执行。

第三方面，本公开实施例提供了一种电子设备，包括：存储器和处理器，所述存储器用于存储计算机程序；所述处理器用于在调用计算机程序时，使得所述电子设备实现第一方面或第一方面任一种可选的实施方式所述的文件泄露检测方法。

- 15 第四方面，本公开实施例提供一种非瞬态的计算机可读存储介质，当所述计算机程序被处理器执行时，使得所述处理器实现第一方面或第一方面任一种可选的实施方式所述的文件泄露检测方法。

第五方面，本公开实施例提供一种计算机程序产品，当所述计算机程序产品在计算机上运行时，使得所述计算机实现第一方面或第一方面任一种可选的实施方式所述的文件泄露检测方法。

- 20 第六方面，本公开实施例提供一种计算机程序，当所述计算机程序被处理器执行时，使得所述处理器实现第一方面或第一方面任一种可选的实施方式所述的文件泄露检测方法。

附图说明

- 25 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本公开的实施例，并与说明书一起用于解释本公开的原理。

为了更清楚地说明本公开实施例或相关技术中的技术方案，下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍，对于本领域普通技术人员而言，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

- 图 1 为本公开实施例提供的文件泄露检测方法的场景架构图；
- 图 2 为本公开实施例提供的文件泄露检测方法的步骤流程图之一；
- 图 3 为本公开实施例提供的文件泄露检测方法的步骤流程图之二；
- 图 4 为本公开实施例提供的文件泄露检测方法的步骤流程图之三；
- 5 图 5 为本公开实施例提供的文件泄露检测方法的步骤流程图之四；
- 图 6 为本公开实施例提供的文件泄露检测方法的步骤流程图之五；
- 图 7 为本公开实施例提供的文件泄露检测方法的步骤流程图之六；
- 图 8 为本公开实施例提供的文件泄露检测方法的步骤流程图之七；
- 图 9 为本公开实施例提供的文件泄露检测方法的步骤流程图之八；
- 10 图 10 为本公开实施例提供的文件泄露检测装置的示意图；
- 图 11 为本公开实施例提供的电子设备的硬件结构示意图。

具体实施方式

为了能够更清楚地理解本公开的上述目的、特征和优点，下面将对本公开的方案进行进一步描述。需要说明的是，在不冲突的情况下，本公开的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本公开，但本公开还可以采用其他不同于在此描述的方式来实施；显然，说明书中的实施例只是本公开的一部分实施例，而不是全部的实施例。

需要说明的是，为了便于清楚描述本公开实施例的技术方案，在本公开的实施例中，采用了“第一”、“第二”等字样对功能和作用基本相同的相同项或相似项进行区分，本领域技术人员可以理解“第一”、“第二”等字样并不是在对数量和执行次序进行限定。例如：第一特征图像集合和第二特征图像集合仅仅是为了区分不同的特征图像集合，而不是在对特征图像集合的顺序等进行限定。

在本公开实施例中，“示例性的”或者“例如”等词用于表示作例子、例证或说明。本公开实施例中描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言，使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。此外，在本公开实施例的描述中，除非另有说明，“多个”的含义是指两个或两个以上。

相关技术中普遍使用的文件泄露检测方式为：通过检测文件中是否包含指定关键字识别文件是否为隐私文件，进而确定是否正在发生文件泄露。然而，通过是否包含指定关键字识别文件是否为隐私文件的过程中，若指定关键字范围太小，则会产生漏报，进而导致无法检测到隐私文件泄露，而若指定关键字范围太大，则会产生误报，影响正常的文件操作事件的执行。

本公开实施例提供的文件泄露检测方法包括：首先获取终端设备上对目标文件执行指定操作的事件的文件操作事件，然后从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径，再根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内容进行映射处理得到所述目标文件的文件指纹，最后根据所述文件路径和所述文件指纹确定所述目标文件是否属于用于动态维护需要保护的业务的指定文件库，以及在所述目标文件属于所述指定文件库的情况下，确定所述指定文件库中的文件正在发生泄漏。相比于通过指定关键词进行文件泄露的检测，本公开实施例提供的文件泄露检测方法可以通过文件的文件路径和文件指纹两个维度确认文件是否属于指定文件库，因此本公开实施例可以在检测文件是否属于指定文件库的过程中减少漏检测和误检测，进而提升文件泄露检测的准确率。

以下首先对本公开实施例提供的文件泄露检测方法的场景架构进行说明。

参照图 1 所示，本公开实施例提供的文件泄露检测方法的场景架构中包括服务器 11 和终端设备 12。

其中，服务器 11 包括：隐私文件管理模块 111 和通信模块 112。隐私文件管理模块 111 对外暴露有应用程序接口（Application Programming Interface, API）。管理人员可以通过隐私文件管理模块 111 对外暴露的 API 配置 Git 项目地址、SVN 项目地址等不想被外传的文件所在的文件库地址，以及对已配置的文件库地址进行增加、删除、修改等操作。通信模块 112 用于向终端设备 11 发送管理人员配置的文件库地址以及接收终端设备 11 上报的指定文件库中的文件泄漏的报警信息、泄露指定文件库中的文件的操作事件的属性信息等信息。

终端设备 12 包括：事件监控模块 121、文件检索模块 122、索引信息库 123、文件库识别模块 124 以及通信模块 125。事件监控模块 121 用于检测对终端设备上的所有进程事件进行检测，并确定各个进程事件是否为文件外发、文件拷贝等有可能会泄露指定文件库中的文件的操作事件，以及在确定某一事件为可能会泄露指定文件库中的文件的事件时，

将该操作事件的相关信息发送至文件检索模块 122，以进一步确定是否正在泄露指定文件库中的文件。文件检索模块 122 获取可能会泄露指定文件库中的文件的操作事件所操作的目标文件的文件路径和文件指纹，并基于索引信息库 123 中存储的路径和哈希值确定目标文件是否为指定文件库中的文件，以及在确定目标文件为指定文件库中的文件时，确定指定文件库中的文件正在发生泄漏。文件库识别模块 124 用于在文件检索模块 122 确定索引信息库 123 中不包括目标文件的文件路径和文件指纹时，进一步确定目标文件的同级文件、父文件所属项目的文件仓库是否属于指定文件库，以及在目标文件的同级文件或父文件所属项目的文件仓库属于指定文件库时，确定指定文件库中的文件正在发生泄漏，并将目标文件的文件路径和文件指纹添加到索引信息库 123 中。通信模块 125 用于接收服务器 11 发送的文件库地址以及向服务器 11 发送报警信息、文件操作事件的属性信息等信息。

基于上述内容，本公开实施例提供了一种应用于终端设备的文件泄露检测方法，参照图 2 所示，本公开实施例提供的文件泄露检测方法包括如下步骤：

S21、获取终端设备上的文件操作事件。

其中，所述文件操作事件为对目标文件执行指定操作的事件。

本公开实施例中的指定操作可以包括各个可能会造成指定文件库中的文件泄露的文件操作。示例性的，指定操作可以包括：拷贝文件的操作，将文件上传至网络平台的事件、将文件发送至外网的操作等。

本公开实施例中的文件可以为任意类型的文件。例如：文件可以包括源代码、文档、配置文件、图像、视频文件、音频文件、编译产物等。

可选的，终端设备可以实时检测终端设备上的各个操作事件是否为涉及文件的操作事件，当确定某一操作事件为涉及文件的操作事件时，进一步确定该操作事件所涉及的操作是否为指定操作，当确定该操作事件所涉及的操作为指定操作，则确定该操作事件为本公开实施例中的文件操作事件。

S22、从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径。

具体的，目标文件的文件路径是指读取目标文件时所历经的存储线路。例如：读取某一文件名称为“初稿”的文件时，依次需要经历的存储线路包括：系统 C 盘、Users 文件夹、zhangsan 文件夹、Desktop 文件夹、KP2116216.4 文件夹，则该文件的文件路径为：C:\Users\zhangsan\Desktop\KP2116216.4\初稿。

S23、根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内

容进行映射处理得到所述目标文件的文件指纹。

其中，所述文件指纹唯一标识所述目标文件的文件内容。

可选的，所述目标文件的文件指纹可以为目标文件的哈希（Hash）值。

即，先根据目标文件的文件路径读取目标文件，然后对所述目标文件的文件内容进行

5 映射处理得到所述目标文件的文件指纹。

S24、根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库。

其中，所述指定文件库用于动态维护需要保护的**业务文件**。

本公开实施例中的需要保护的**业务文件**是指不想被泄露的文件。需要保护的**业务文件**可以包括：企业源代码、文档、配置文件、图像、视频文件、音频文件、编译产物等文件。

10 所述指定文件库用于动态维护需要保护的**业务文件**是指：可以动态的对指定文件库中的文件进行添加、修改、删除，从而动态的对需要保护的**业务文件**进行管理。

作为本公开实施例一种可选的实施方式，在上述步骤 S24（根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库）包括：

确定索引信息库中是否包含所述文件路径和所述文件指纹；其中，所述索引信息库中

15 存储有所述指定文件库中的至少一个文件的文件路径和文件指纹；

若索引信息库中包含所述文件路径和/或所述文件指纹，则确定所述目标文件属于指定文件库。

可选的，索引信息库可以通过下表 1 所示方式存储指定文件库中的文件的路径和指纹。

表 1

文件名称	文件路径	文件哈希值
文件A	文件路径A	文件指纹A
文件B	文件路径B	文件指纹B
文件C	文件路径C	文件指纹C
⋮	⋮	⋮

20

如上表 1 所示，索引信息库可以为 Key-Value 数据库，将文件的文件名称作为 key，将文件的路径和指纹作为 value 进行存储。

可选的，根据所述文件路径、所述文件指纹以及所述索引信息库确定所述目标文件是否属于指定文件库的实现方式可以包括：

遍历所述索引数据库中的文件路径和文件指纹，若所述索引数据库中包括所述文件路径和/或所述文件指纹，则确定目标文件属于指定文件库，而若所述索引数据库中不包括所述文件路径或所述文件指纹，则确定目标文件不属于指定文件库。

需要说明的是，本公开实施例中文件的路径是指文件的全路径。即，文件路径中包含了文件的名称，因此确定根据所述文件路径、所述文件指纹以及索引信息库确定所述目标文件是否属于指定文件库可以理解为根据目标文件的名称、所述文件路径、所述文件指纹以及索引信息库确定所述目标文件是否属于指定文件库。

作为本公开实施例一种可选的实施方式，在上述步骤 S24（根据所述文件路径和所述文件指纹确定所述目标文件是否属于所述指定文件库）之前，本公开实施例提供的文件泄露检测方法还包括：

接收服务器发送的至少一个文件库地址；

根据所述至少一个文件库地址确定所述指定文件库。

即，终端设备需要先从服务器获取配置了哪些文件库中的文件是需要避免泄露的，具体的获取方式为：接收服务器发送的文件库地址，将各个文件库地址标识的文件确定为指定文件库。

在上述步骤 S24 中，若确定所述目标文件属于所述指定文件库，则执行如下步骤 S25。

S25、确定所述指定文件库中的文件正在发生泄漏。

本公开实施例提供的文件泄露检测方法包括：首先获取终端设备上对目标文件执行指定操作的事件的文件操作事件，然后从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径，再根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内容进行映射处理得到所述目标文件的文件指纹，最后根据所述文件路径和所述文件指纹确定所述目标文件是否属于用于动态维护需要保护的业务的指定文件库，以及在所述目标文件属于所述指定文件库的情况下，确定所述指定文件库中的文件正在发生泄漏。相比于通过指定关键词进行文件泄露的检测，本公开实施例提供的文件泄露检测方法可以通过文件的文件路径和文件指纹两个维度确认文件是否属于指定文件库，因此本公开实施例可以在检测文件是否属于指定文件库的过程中减少漏检测和误检测，进而提升文件泄露检测的准确率。

作为本公开实施例一种可选的实施方式，在图 2 所示实施例的基础上，参照图 3 所示，上述步骤 S24（根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库）

的一种实现方式包括如下步骤 S31 至 S34:

S31、确定索引信息库中是否包含所述文件路径。

其中,所述索引信息库存储有所述指定文件库中的至少一个文件的文件路径和文件指纹。

- 5 具体的,可以遍历索引信息库中存储的各个文件路径,判断所述索引信息库中是否具有与所述文件路径完全相同的文件路径,以确定所述索引信息库中是否包含所述文件路径。

在上步骤 S31 中,若确定所述索引信息库中包含所述文件路径,则执行如下步骤 S32。

S32、确定目标文件属于指定文件库。

在上步骤 S31 中,若确定所述索引信息库中不包含所述文件路径,则执行如下步骤 S33。

- 10 S33、确定索引信息库中是否包含所述文件指纹。

同样,可以遍历索引信息库中存储的各个文件指纹,判断所述索引信息库中是否具有与所述文件指纹完全相同的文件指纹,以确定所述索引信息库中是否包含所述文件指纹。

在上步骤 S33 中,若确定所述索引信息库中包含所述文件指纹,则执行如下步骤 S34。

S34、确定所述目标文件属于所述指定文件库。

- 15 一方面,上述实施例在所述索引信息库中不包含所述文件路径时,还会进一步确定所述索引信息库中是否包含所述文件指纹,并在所述索引信息库中包含所述文件指纹的情况下确定所述目标文件属于所述指定文件库,因此上述实施例可以避免文件先将所述指定文件库中的文件拷贝到其它存储位置(文件路径发生变化),然后再通过拷贝、外发等方式泄露该文件,进而提升指定文件库中的文件的安全性。

- 20 另一方面,本公开是实施例先检测文件路径是否属于检索数据库,且在文件路径是否属于检索数据库时直接确定文件路径是否属于检索数据库,不再进一步确定文件指纹是否属于检索数据库,因此上述实施例还可以减少文件检索过程中的数据处理量,进而提升数据文件泄露检测的效率。

- 25 作为本公开实施例一种可选的实施方式,在图 3 所示实施例的基础上,参照图 4 所示,在所述索引信息库中不包含所述文件指纹的情况下,本公开实施例提供的文件泄露检测方法还包括如下步骤 S41 至 S44:

S41、根据所述文件路径中是否包含预设字段确定所述目标文件是否来自于预设文件托管平台。

示例性的,预设字段可以包括: .git 和 .svn。

具体的，基于研发、运维等需求，企业常常将文件上传到的内部文件托管平台，在需要使用这些文件时可以直接从内部文件托管平台拉取这些文件，而从文件托管平台拉取的文件的路径均会被写入相应的字段。例如：将代码托管到 git 平台或 svn 平台，则从 git 平台或 svn 平台拉取的代码的路径中会字段“.git”或“.svn”。

5 具体的，根据所述文件路径中是否包含预设字段确定所述目标文件是否来自于预设文件托管平台，包括：

 当所述文件路径中包含所述预设字段，则确定所述目标文件属于预设文件托管平台；

 当所述文件路径中不包含所述预设字段，则确定所述目标文件不属于预设文件托管平台。

10 在上述步骤 S41 中，若确定所述目标文件来自于所述预设文件托管平台，则执行如下步骤 S42。

 S42、确定第一文件库。

 其中，所述第一文件库为所述目标文件所属项目的文件仓库。

 具体的，当预设字段为.git 时，可以通过解析.git 目录获取目标文件所属项目的文件仓库地址，从而确定第一仓库，当预设字段为.svn 时，可以通过解析.svn 目录获取目标文件所属项目的文件仓库地址，从而确定第一仓库。

15 S43、确定第一文件库是否属于指定文件库。

 具体的，可以通过比对第一文件库的地址是否属于指定文件库的地址，从而确定第一文件库是否属于所述指定文件库。

20 在上述步骤 S43 中，若确定所述第一文件库属于所述指定文件库，则执行如下步骤 S44。

 S44、确定所述目标文件属于所述指定文件库。

 上述实施例在所述索引信息库中不包含所述文件指纹的情况下，进一步根据文件路径中是否包含预设字段确定所述目标文件是否来自于预设文件托管平台，在目标文件来自于所述预设文件托管平台的情况下，获取目标文件所属项目的文件仓库，判断目标文件所属项目的文件仓库是否属于指定文件库，以及在目标文件所属项目的文件仓库属于指定文件库的情况下，确定所述目标文件属于所述指定文件库，因此上述实施例可以防止先将指定文件库中的敏感文件上传文件托管平台，然后在从文件托管平台下载并泄露敏感文件，进而更加全面的检测出对指定文件库中的文件的泄露行为。

 作为本公开实施例一种可选的实施方式，本公开实施例提供的文件泄露检测方法还包

括：

在确定所述第一文件库属于所述指定文件库的情况下，将所述文件路径和所述文件指纹添加到所述索引信息库中。

在确定所述目标文件属于所述指定文件库时，将所述文件路径和所述文件指纹添加到
5 所述索引信息库中，可以方便后直接根据索引信息库检测出指定文件库中的文件的泄露行为，进而提升文件泄露检测的效率。

作为本公开实施例一种可选的实施方式，在图 4 所示实施例的基础上，参照图 5 所示，在确定所述第一文件库不属于所述指定文件库的情况下，本公开实施例提供的文件泄露检测方法还包括如下步骤 S51 至 S54：

10 S51、根据同级路径中是否包含预设字段确定同级文件是否来自于预设文件托管平台。
其中，所述同级路径为父文件与所述目标文件相同的同级文件的路径。

例如：文件“初稿”为本公开实施例中的目标文件，文件“初稿”的文件路径为：
C:\Users\zhangsan\Desktop\ KP2116216.4\初稿，则文件“初稿”的父文件为：KP2116216.4，
若文件“初稿”的父文件 KP2116216.4 下还存在文件名为“2 稿”的文件，则文件“2 稿”为文
15 件“初稿”的同级文件，文件“2 稿”的路径 C:\Users\zhangsan\Desktop\ KP2116216.4\2 稿为本公开实施例中的同级路径。

在上述步骤 S51 中，若确定所述同级文件来自于所述预设文件托管平台，则执行如下步骤 S52。

S52、确定第二文件库。
20 其中，所述第二文件库为所述同级文件所属项目的文件仓库。

S53、确定所述第二文件库是否属于所述指定文件库。
在上述步骤 S53 中，若确定所述第二文件库属于所述指定文件库，则执行如下步骤 S54。

S54、确定所述目标文件属于所述指定文件库。
上述实施例在确定所述第一文件库不属于所述指定文件库的情况下，进一步根据同级
25 路径中是否包含所述预设字段确定同级文件是否来自于所述预设文件托管平台，以及在同级文件来自于所述预设文件托管平台的情况下，获取同级文件所属项目的文件仓库，并判断同级文件所属项目的文件仓库是否属于指定文件库，以及在同级文件所属项目的文件仓库属于指定文件库的情况下，确定所述目标文件属于所述指定文件库，因此上述实施例可以进而更加全面的检测出对指定文件库中的文件的泄露行为。

作为本公开实施例一种可选的实施方式，本公开实施例提供的文件泄露检测方法还包括：

在确定所述第二文件库属于所述指定文件库的情况下，将所述文件路径和所述文件指纹添加到所述索引信息库中。

- 5 在确定所述第二文件库属于所述指定文件库时，将所述文件路径和所述文件指纹添加到所述索引信息库中，可以方便后直接根据索引信息库检测出指定文件库中的文件的泄露行为，进而提升文件泄露检测的效率。

作为本公开实施例一种可选的实施方式，在图 5 所示实施例的基础上，参照图 6 所示，在确定所述第二文件库不属于所述指定文件库的情况下，本公开实施例提供的文件泄露检测方法还包括如下步骤 S61 至 S64：

S61、根据父路径中是否包含所述预设字段同确定至少一级父文件是否来自于所述预设文件托管平台。

其中，所述父路径为所述目标文件的父文件的路径。

示例性的，至少一级父路径可以为 3 级。即，依次向上 3 级目录查找。

- 15 在上述步骤 S61 中，若任一级父文件来自于所述预设文件托管平台，则执行如下步骤 S62。

S62、确定第三文件库。

其中，所述第三文件库为来自于所述预设文件托管平台的父文件所属项目的文件仓库。

S63、确定所述第三文件库是否属于所述指定文件库。

- 20 在上述步骤 S63 中，若确定所述第三文件库属于所述指定文件库，则执行如下步骤 S64。

S64、确定所述目标文件属于所述指定文件库。

上述实施例在确定所述第二文件库不属于所述指定文件库的情况下，进一步确定至少一级父路径中是否包含预设字段，在任一级父路径包含预设字段的情况下，获取该级父路径对应的父文件所属项目的文件仓库，判断父文件所属项目的文件仓库是否属于指定文件库，以及在父文件所属项目的文件仓库属于指定文件库的情况下，确定所述目标文件属于所述指定文件库，因此上述实施例可以进而更加全面的检测出对指定文件库中的文件的泄露行为。

25 作为本公开实施例一种可选的实施方式，本公开实施例提供的文件泄露检测方法还包括：

在确定所述第三文件库属于所述指定文件库的情况下，将所述文件路径和所述文件指纹添加到所述索引信息库中。

在确定所述第三文件库属于所述指定文件库时，将所述文件路径和所述文件指纹添加到所述索引信息库中，可以方便后直接根据索引信息库检测出指定文件库中的文件的泄露行为，进而提升文件泄露检测的效率。

作为本公开实施例一种可选的实施方式，在图 1 所示实施例的基础上，参照图 7 所示，在确定所述指定文件库中的文件正在发生泄漏的情况下，本公开实施例提供的文件泄露检测方法还包括：

S71、向服务器发送报警信息。

其中，所述报警信息用于指示所述指定文件库中的文件正在发生泄漏。

当发现指定文件库中的文件正在发生泄漏时及时向服务器发送报警信息，可以使信息安全人员及时对泄露行为进行阻止。

作为本公开实施例一种可选的实施方式，在图 1 所示实施例的基础上，参照图 8 所示，在确定所述指定文件库中的文件正在发生泄漏的情况下，本公开实施例提供的文件泄露检测方法还包括：

S81、向服务器发送所述文件操作事件的属性信息。

其中，所述文件操作事件的属性信息包括：所述终端设备的标识信息、所述目标文件的标识信息、登录所述终端设备的账户信息中的至少一个。

上述实施例在发现指定文件库中的文件正在发生泄漏时，将终端设备的标识信息、所述目标文件的标识信息、登录所述终端设备的账户信息发送至服务器，因此上述实施例可以方便对泄露行为进行追责。

作为本公开实施例一种可选的实施方式，在图 1 所示实施例的基础上，参照图 9 所示，在确定所述指定文件库中的文件正在发生泄漏的情况下，本公开实施例提供的文件泄露检测方法还包括：

S91、阻止所述文件操作事件的执行。

上述实施例在发现指定文件库中的文件正在发生泄漏时，阻止所述文件操作事件的执行，因此上述实施例可以在发现指定文件库中的文件正在发生泄漏时，及时阻止指定文件库中的文件发生泄漏。

基于同一发明构思，作为对上述方法的实现，本公开实施例还提供了一种文件泄露检

测装置，该装置实施例与前述方法实施例对应，为便于阅读，本装置实施例不再对前述方法实施例中的细节内容进行逐一赘述，但应当明确，本实施例中的文件泄露检测装置能够对应实现前述方法实施例中的全部内容。

本公开实施例提供了一种文件泄露检测装置，图 10 为该文件泄露检测装置的结构示意图，如图 10 所示，该文件泄露检测装置 100 包括：

检测单元 101，用于获取终端设备上的文件操作事件，所述文件操作事件为对目标文件执行指定操作的事件；

提取单元 102，用于从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径；

10 获取单元 103，用于根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内容进行映射处理得到所述目标文件的文件指纹，所述文件指纹唯一标识所述目标文件的文件内容；

确定单元 104，用于根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库，所述指定文件库用于动态维护需要保护的义务文件；

15 处理单元 105，用于在所述目标文件属于所述指定文件库的情况下，确定所述指定文件库中的文件正在发生泄漏。

作为本公开实施例一种可选的实施方式，所述确定单元 104，具体用于确定索引信息库中是否包含所述文件路径；若所述索引信息库中包含所述文件路径，则确定所述目标文件属于所述指定文件库；若所述索引信息库中不包含所述文件路径，则确定所述索引信息
20 库中是否包含所述文件指纹；若所述索引信息库中包含所述文件指纹，则确定所述目标文件属于所述指定文件库

其中，所述索引信息库存储有所述指定文件库中的至少一个文件的文件路径和文件指纹。

作为本公开实施例一种可选的实施方式，所述确定单元 104，还用于在所述索引信息
25 库中不包含所述文件指纹的情况下，根据所述文件路径中是否包含预设字段确定所述目标文件是否来自于预设文件托管平台；若所述目标文件来自于所述预设文件托管平台，则确定第一文件库，所述第一文件库为所述目标文件所属项目的文件仓库；确定所述第一文件库是否属于所述指定文件库；若第一文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，所述确定单元 104，还用于在确定所述第一文件库不属于所述指定文件库的情况下，根据同级路径中是否包含所述预设字段确定同级文件是否来自于所述预设文件托管平台，所述同级路径为父文件与所述目标文件相同的同级文件的路径；若所述同级文件来自于所述预设文件托管平台，则确定第二文件库，所述

5 第二文件库为所述同级文件所属项目的文件仓库；确定所述第二文件库是否属于所述指定文件库；若第二文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，所述确定单元 104，还用于在确定所述第二文件库不属于所述指定文件库的情况下，根据父路径中是否包含所述预设字段同确定至少一级父文件是否来自于所述预设文件托管平台，所述父路径为所述目标文件的父文件的路

10 径；若任一级父文件来自于所述预设文件托管平台，则确定第三文件库，所述第三文件库为来自于所述预设文件托管平台的父文件所属项目的文件仓库；确定所述第三文件库是否属于所述指定文件库；若第三文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

作为本公开实施例一种可选的实施方式，所述确定单元 104，还用于在确定所述目标

15 文件属于所述指定文件库的情况下，将所述文件路径和所述文件指纹添加到所述索引信息库中。

作为本公开实施例一种可选的实施方式，所述确定单元 104，还用于在根据所述文件路径、所述文件指纹以及所述索引信息库确定所述目标文件是否属于所述指定文件库之前，接收服务器发送的至少一个文件库地址；根据所述至少一个文件库地址确定所述指定文件

20 库。

作为本公开实施例一种可选的实施方式，所述处理单元 105，还用于在确定所述指定文件库中的文件正在发生泄漏的情况下，向服务器发送报警信息；所述报警信息用于指示所述指定文件库中的文件正在发生泄漏。

作为本公开实施例一种可选的实施方式，所述处理单元 105，还用于在确定所述指定

25 文件库中的文件正在发生泄漏的情况下，向服务器发送所述文件操作事件的属性信息；

其中，所述文件操作事件的属性信息包括：所述终端设备的标识信息、所述目标文件的标识信息、登录所述终端设备的账户信息中的至少一个。

作为本公开实施例一种可选的实施方式，所述处理单元 105，还用于在确定所述目标文件属于所述指定文件库的情况下，阻止所述文件操作事件的执行。

本实施例提供的文件泄露检测装置可以执行上述方法实施例提供的文件泄露检测方法，其实现原理与技术效果类似，此处不再赘述。

基于同一发明构思，本公开实施例还提供了一种电子设备。图 11 为本公开实施例提供的电子设备的结构示意图，如图 11 所示，本实施例提供的电子设备包括：存储器 111 和处理器 112，所述存储器 111 用于存储计算机程序；所述处理器 112 用于在调用计算机程序时执行上述实施例提供的文件泄露检测方法。本公开实施例中的电子设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体播放器）、车载终端（例如车载导航终端）等等的移动终端以及诸如数字 TV、台式计算机等等的固定终端。

基于同一发明构思，本公开实施例还提供了一种非瞬态的计算机可读存储介质，该计算机可读存储介质上存储有计算机程序，当计算机程序被处理器执行时，使得所述处理器实现上述实施例提供的文件泄露检测方法。

基于同一发明构思，本公开实施例还提供了一种计算机程序产品，当所述计算机程序产品在计算机上运行时，使得所述计算机实现上述实施例提供的文件泄露检测方法。

基于同一发明构思，本公开实施例还提供了一种计算机程序，当计算机程序被处理器执行时，使得所述处理器实现上述实施例提供的文件泄露检测方法。

本领域技术人员应明白，本公开的实施例可提供为方法、系统、或计算机程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本公开可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质上实施的计算机程序产品的形式。

处理器可以是中央处理单元(Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现成可编程门阵列(Field-Programmable Gate Array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

存储器可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。存储器是计算机可读介质的示例。

计算机可读介质包括永久性和非永久性、可移动和非可移动存储介质。存储介质可以

由任何方法或技术来实现信息存储，信息可以是计算机可读指令、数据结构、程序的单元或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。根据本文中的界定，计算机可读介质不包括暂存电脑可读媒体(transitory media)，如调制的数据信号和载波。

最后应说明的是：以上各实施例仅用以说明本公开的技术方案，而非对其限制；尽管参照前述各实施例对本公开进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本公开各实施例技术方案的范围。

权 利 要 求

1、一种文件泄露检测方法，包括：

获取终端设备上的文件操作事件，所述文件操作事件为对目标文件执行指定操作的事件；

从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径；

根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内容进行映射处理得到所述目标文件的文件指纹，所述文件指纹唯一标识所述目标文件的文件内容；

根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库，所述指定文件库用于动态维护需要保护的業務文件；

若所述目标文件属于所述指定文件库，则确定所述指定文件库中的文件正在发生泄漏。

2、根据权利要求 1 所述的方法，其中，所述根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库，包括：

确定索引信息库中是否包含所述文件路径；所述索引信息库存储有所述指定文件库中的至少一个文件的文件路径和文件指纹；

若所述索引信息库中包含所述文件路径，则确定所述目标文件属于所述指定文件库；

若所述索引信息库中不包含所述文件路径，则确定所述索引信息库中是否包含所述文件指纹；

若所述索引信息库中包含所述文件指纹，则确定所述目标文件属于所述指定文件库。

3、根据权利要求 2 所述的方法，其中，在所述索引信息库中不包含所述文件指纹的情况下，所述方法还包括：

根据所述文件路径中是否包含预设字段确定所述目标文件是否来自于预设文件托管平台；

若所述目标文件来自于所述预设文件托管平台，则确定第一文件库，所述第一文件库为所述目标文件所属项目的文件仓库；

确定所述第一文件库是否属于所述指定文件库；

若第一文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

4、根据权利要求 3 所述的方法，其中，在确定所述第一文件库不属于所述指定文件库的情况下，所述方法还包括：

根据同级路径中是否包含所述预设字段确定同级文件是否来自于所述预设文件托管平台，所述同级路径为父文件与所述目标文件相同的同级文件的路径；

若所述同级文件来自于所述预设文件托管平台，则确定第二文件库，所述第二文件库为所述同级文件所属项目的文件仓库；

确定所述第二文件库是否属于所述指定文件库；

若第二文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

5、根据权利要求 4 所述的方法，其中，在确定所述第二文件库不属于所述指定文件库的情况下，所述方法还包括：

根据父路径中是否包含所述预设字段同确定至少一级父文件是否来自于所述预设文件托管平台，所述父路径为所述目标文件的父文件的路径；

若任一级父文件来自于所述预设文件托管平台，则确定第三文件库，所述第三文件库为来自于所述预设文件托管平台的父文件所属项目的文件仓库；

确定所述第三文件库是否属于所述指定文件库；

若第三文件库属于所述指定文件库，则确定所述目标文件属于所述指定文件库。

6、根据权利要求 3-5 中任一项所述的方法，其中，所述方法还包括：

在确定所述目标文件属于所述指定文件库的情况下，将所述文件路径和所述文件指纹添加到所述索引信息库中。

7、根据权利要求 1-5 中任一项所述的方法，其中，所述方法还包括：

在确定所述指定文件库中的文件正在发生泄漏的情况下，向服务器发送报警信息，所述报警信息用于指示所述指定文件库中的文件正在发生泄漏。

8、根据权利要求 1-5 中任一项所述的方法，其中，所述方法还包括：

在确定所述指定文件库中的文件正在发生泄漏的情况下，向服务器发送所述文件操作事件的属性信息；

其中，所述文件操作事件的属性信息包括：所述终端设备的标识信息、所述目标文件的标识信息、登录所述终端设备的账户信息中的至少一个。

9、根据权利要求 1-5 中任一项所述的方法，其中，所述方法还包括：

在确定所述指定文件库中的文件正在发生泄漏的情况下，阻止所述文件操作事件的执行。

10、一种文件泄露检测装置，包括：

检测单元，用于获取终端设备上的文件操作事件，所述文件操作事件为对目标文件执行指定操作的事件；

提取单元，用于从所述文件操作事件中提取所述文件操作事件所涉及的目标文件的文件路径；

获取单元，用于根据所述文件路径查找所述目标文件的文件内容，并对所述目标文件的文件内容进行映射处理得到所述目标文件的文件指纹，所述文件指纹唯一标识所述目标文件的文件内容；

确定单元，用于根据所述文件路径和所述文件指纹确定所述目标文件是否属于指定文件库，所述指定文件库用于动态维护需要保护的业务文件；

处理单元，用于在所述目标文件属于所述指定文件库的情况下，确定所述指定文件库中的文件正在发生泄漏。

11、一种电子设备，包括：存储器和处理器，所述存储器用于存储计算机程序；所述处理器用于在执行所述计算机程序时，使得所述电子设备实现如权利要求 1-9 中任一项所述的文件泄露检测方法。

12、一种非瞬态的计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，当所述计算机程序被处理器执行时，使得所述处理器实现如权利要求 1-9 中任一项所述的文件泄露检测方法。

13、一种计算机程序产品，当所述计算机程序产品在计算机上运行时，使得所述计算机实现如权利要求 1-9 中任一项所述的文件泄露检测方法。

14、一种计算机程序，当所述计算机程序被处理器执行时，使得所述处理器实现如权利要求 1-9 中任一项所述的文件泄露检测方法。