(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113360799 A
(43)申请公布日 2021.09.07

(21)申请号 202110629070.3

(22)申请日 2021.06.03

(71)申请人 深圳红途科技有限公司
地址 518000 广东省深圳市南山区粤海街道滨海社区滨海大道3398号赛西科技大厦17层1706房

(72)发明人 刘新凯

(74)专利代理机构 深圳市精英专利事务所 44242
代理人 李燕娥

(51)Int.Cl.
G06F 16/955(2019.01)
G06F 16/957(2019.01)
G06F 16/958(2019.01)

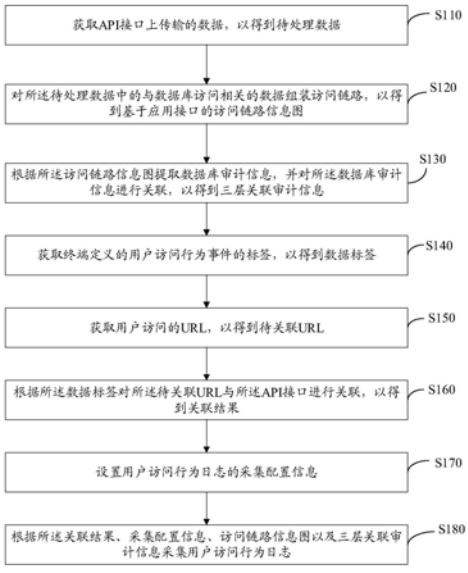
权利要求书2页 说明书20页 附图8页

(54)发明名称

访问行为日志采集方法、装置、计算机设备及存储介质

(57)摘要

本发明实施例公开了访问行为日志采集方法、装置、计算机设备及存储介质。方法包括：获取待处理数据；对待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图；提取数据库审计信息，并对数据库审计信息进行关联，以得到三层关联审计信息；获取终端定义的用户访问行为事件的标签，以得到数据标签；获取用户访问的URL，以得到待关联URL；对待关联URL与API接口进行关联，以得到关联结果；设置采集配置信息；采集用户访问行为日志。通过实施本发明实施例的方法可实现无需对应用进行改造，不需要在代码层面进行二次开发，对原有的业务不存在入侵性，维护成本低，对用户访问数据的行为关联到数据库层面。



CN 113360799 A

1.访问行为日志采集方法,其特征在于,包括:

获取API接口上传输的数据,以得到待处理数据;

对所述待处理数据中的与数据库访问相关的数据组装访问链路,以得到基于应用接口的访问链路信息图;

根据所述访问链路信息图提取数据库审计信息,并对所述数据库审计信息进行关联,以得到三层关联审计信息;

获取终端定义的用户访问行为事件的标签,以得到数据标签;

获取用户访问的URL,以得到待关联URL;

根据所述数据标签对所述待关联URL与所述API接口进行关联,以得到关联结果;

设置用户访问行为日志的采集配置信息;

根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

2.根据权利要求1所述的访问行为日志采集方法,其特征在于,所述获取终端定义的用户访问行为事件的标签,以得到数据标签,包括:

由终端启动浏览器插件,并利用所述浏览器插件拦截用户交互动作的请求信息或响应信息;

由终端根据所拦截的信息定位用户访问指定的页面数据对应的URL及字段;

由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义,以得到用户访问行为事件的标签;

获取用户访问行为事件的标签,以得到数据标签。

3.根据权利要求2所述的访问行为日志采集方法,其特征在于,所述由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义,以得到用户访问行为事件的标签,包括:

由终端对用户访问指定的页面数据对应的URL及字段定义访问事件名称及备注信息,以得到用户访问行为事件的标签。

4.根据权利要求1所述的访问行为日志采集方法,其特征在于,所述根据所述数据标签对所述待关联URL与所述API接口进行关联,以得到关联结果,包括:

对所述待关联URL匹配对应的API接口,以得到目标API接口;

将所述待关联URL对应的字段与所述目标API接口对应的字段进行关联;

将所述数据标签中关于所述待关联URL对应的字段所对应的访问事件名称及备注信息关联至所述目标API接口对应的字段,以得到关联结果。

5.根据权利要求4所述的访问行为日志采集方法,其特征在于,所述对所述待关联URL匹配对应的API接口,以得到目标API接口,包括:

对所述待关联URL进行信息分割,以得到访问URL路径;

对所述访问URL路径进行分割,以得到第一有序数组;

对所述API接口进行信息分割和路径分割,以得到第二有序数组;

遍历所述第一有序数组,并将所述第一有序数组与所述第二有序数组相同位置的值进行对比,以确定所述待关联URL所匹配成功的API接口;

对所述待关联URL所匹配成功的API接口进行打标识,以得到目标API接口。

6.根据权利要求1所述的访问行为日志采集方法，其特征在于，所述采集配置信息包括用户访问行为名称、采集开关、进行用户访问行为日志采集的API接口以及进行用户访问行为日志采集的API接口中的字段。

7.根据权利要求6所述的访问行为日志采集方法，其特征在于，所述根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志，包括：

从访问链路信息图中筛选出与进行用户访问行为日志采集的API接口相关的链路，以得到待判定链路；

判断所述待判定链路中是否存在数据库访问接口节点；

若所述待判定链路中存在数据库访问接口节点，则获取所述数据库访问接口节点相关的三层关联审计信息，并从获取的三层关联审计信息中提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息；

将所述目标信息存储至日志缓存区；

若所述待判定链路中不存在数据库访问接口节点，则通过所述待判定链路提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息，并执行所述将所述目标信息存储至日志缓存区。

8.访问行为日志采集装置，其特征在于，包括：

数据获取单元，用于获取API接口上传输的数据，以得到待处理数据；

链路组装单元，用于对所述待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图；

审计处理单元，用于根据所述访问链路信息图提取数据库审计信息，并对所述数据库审计信息进行关联，以得到三层关联审计信息；

标签获取单元，用于获取终端定义的用户访问行为事件的标签，以得到数据标签；

URL获取单元，用于获取用户访问的URL，以得到待关联URL；

关联单元，用于根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果；

配置设置单元，用于设置用户访问行为日志的采集配置信息；

采集单元，用于根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

9.一种计算机设备，其特征在于，所述计算机设备包括存储器及处理器，所述存储器上存储有计算机程序，所述处理器执行所述计算机程序时实现如权利要求1至7中任一项所述的方法。

10.一种存储介质，其特征在于，所述存储介质存储有计算机程序，所述计算机程序被处理器执行时可实现如权利要求1至7中任一项所述的方法。

访问行为日志采集方法、装置、计算机设备及存储介质

技术领域

[0001] 本发明涉及日志采集方法，更具体地说是指访问行为日志采集方法、装置、计算机设备及存储介质。

背景技术

[0002] 用户访问行为是指用户访问网站或APP等平台的行为，用户访问行为日志是用于记录用户访问行为的日志，用户访问行为日志因为具备审计、行为分析、用户画像、商业推荐等价值被广泛使用，目前针对用户访问行为日志的采集内容主要是用户、访问时间、访问对象、访问数据和访问结果。目前最常用的采集用户访问行为日志的方式包括代码埋点如js/sdk等采集用户访问行为日志，但是这种方式需要对应用进行改造，进行代码层面的二次开发，开发成本高；每次采集日志时，需要涉及二次开发，修改代码并运行，对原有的业务存在入侵性；采集或不采集或新增/减少用户行为日志等，都需要在代码层面进行维护，维护成本高；采集周期长；无法对用户访问数据的行为关联到数据库层面；无法对用户访问数据的应用接口及应用接口路径进行梳理。

[0003] 因此，有必要设计一种新的方法，实现采集用户访问行为日志时，无需对应用进行改造，不需要在代码层面进行二次开发，对原有的业务不存在入侵性，维护成本低，对用户访问数据的行为关联到数据库层面；对用户访问数据的应用接口及应用接口路径进行梳理。

发明内容

[0004] 本发明的目的在于克服现有技术的缺陷，提供访问行为日志采集方法、装置、计算机设备及存储介质。

[0005] 为实现上述目的，本发明采用以下技术方案：访问行为日志采集方法，包括：

[0006] 获取API接口上传输的数据，以得到待处理数据；对所述待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图；根据所述访问链路信息图提取数据库审计信息，并对所述数据库审计信息进行关联，以得到三层关联审计信息；获取终端定义的用户访问行为事件的标签，以得到数据标签；获取用户访问的URL，以得到待关联URL；根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果；设置用户访问行为日志的采集配置信息；根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

[0007] 其进一步技术方案为：所述获取终端定义的用户访问行为事件的标签，以得到数据标签，包括：

[0008] 由终端启动浏览器插件，并利用所述浏览器插件拦截用户交互动作的请求信息或响应信息；由终端根据所拦截的信息定位用户访问指定的页面数据对应的URL及字段；由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义，以得到用户访问行为事件的标签；获取用户访问行为事件的标签，以得到数据标签。

[0009] 其进一步技术方案为:所述由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义,以得到用户访问行为事件的标签,包括:

[0010] 由终端对用户访问指定的页面数据对应的URL及字段定义访问事件名称及备注信息,以得到用户访问行为事件的标签。

[0011] 其进一步技术方案为:所述根据所述数据标签对所述待关联URL与所述API接口进行关联,以得到关联结果,包括:

[0012] 对所述待关联URL匹配对应的API接口,以得到目标API接口;将所述待关联URL对应的字段与所述目标API接口对应的字段进行关联;将所述数据标签中关于所述待关联URL对应的字段所对应的访问事件名称及备注信息关联至所述目标API接口对应的字段,以得到关联结果。

[0013] 其进一步技术方案为:所述对所述待关联URL匹配对应的API接口,以得到目标API接口,包括:

[0014] 对所述待关联URL进行信息分割,以得到访问URL路径;对所述访问URL路径进行分割,以得到第一有序数组;对所述API接口进行信息分割和路径分割,以得到第二有序数组;遍历所述第一有序数组,并将所述第一有序数组与所述第二有序数组相同位置的值进行对比,以确定所述待关联URL所匹配成功的API接口;对所述待关联URL所匹配成功的API接口进行打标识,以得到目标API接口。

[0015] 其进一步技术方案为:所述采集配置信息包括用户访问行为名称、采集开关、进行用户访问行为日志采集的API接口以及进行用户访问行为日志采集的API接口中的字段。

[0016] 其进一步技术方案为:所述根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志,包括:

[0017] 从访问链路信息图中筛选出与进行用户访问行为日志采集的API接口相关的链路,以得到待判定链路;判断所述待判定链路中是否存在数据库访问接口节点;若所述待判定链路中存在数据库访问接口节点,则获取所述数据库访问接口节点相关的三层关联审计信息,并从获取的三层关联审计信息中提取用户访问行为相关的信息,并统计访问的数据条数和访问数据量,以得到目标信息;将所述目标信息存储至日志缓存区;若所述待判定链路中不存在数据库访问接口节点,则通过所述待判定链路提取用户访问行为相关的信息,并统计访问的数据条数和访问数据量,以得到目标信息,并执行所述将所述目标信息存储至日志缓存区。

[0018] 本发明还提供了访问行为日志采集装置,包括:

[0019] 数据获取单元,用于获取API接口上传输的数据,以得到待处理数据;链路组装单元,用于对所述待处理数据中的与数据库访问相关的数据组装访问链路,以得到基于应用接口的访问链路信息图;审计处理单元,用于根据所述访问链路信息图提取数据库审计信息,并对所述数据库审计信息进行关联,以得到三层关联审计信息;标签获取单元,用于获取终端定义的用户访问行为事件的标签,以得到数据标签;URL获取单元,用于获取用户访问的URL,以得到待关联URL;关联单元,用于根据所述数据标签对所述待关联URL与所述API接口进行关联,以得到关联结果;配置设置单元,用于设置用户访问行为日志的采集配置信息;采集单元,用于根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

[0020] 本发明还提供了一种计算机设备，所述计算机设备包括存储器及处理器，所述存储器上存储有计算机程序，所述处理器执行所述计算机程序时实现上述的方法。

[0021] 本发明还提供了一种存储介质，所述存储介质存储有计算机程序，所述计算机程序被处理器执行时可实现上述的方法。

[0022] 本发明与现有技术相比的有益效果是：本发明通过获取基于字节码增强技术所采集用户访问行为数据和应用接口传输数据，并基于此数据组装用户访问链路、进行用户与应用及数据库的三层关联，在终端结合浏览器插件对访问页面的需要采集的用户访问行为事件进行定义，并根据定义的标签对三层关联审计信息与API接口行关联，且结合设置的采集配置信息，以进行采集用户访问行为日志，实现采集用户访问行为日志时，无需对应用进行改造，不需要在代码层面进行二次开发，对原有的业务不存在入侵性，维护成本低，对用户访问数据的行为关联到数据库层面；对用户访问数据的应用接口及应用接口路径进行梳理。

[0023] 下面结合附图和具体实施例对本发明作进一步描述。

附图说明

[0024] 为了更清楚地说明本发明实施例技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0025] 图1为本发明实施例提供的访问行为日志采集方法的应用场景示意图；

[0026] 图2为本发明实施例提供的访问行为日志采集方法的流程示意图；

[0027] 图3为本发明实施例提供的访问行为日志采集方法的子流程示意图；

[0028] 图4为本发明实施例提供的访问行为日志采集方法的子流程示意图；

[0029] 图5为本发明实施例提供的访问行为日志采集方法的子流程示意图；

[0030] 图6为本发明实施例提供的访问行为日志采集方法的子流程示意图；

[0031] 图7为本发明实施例提供的访问行为日志采集装置的示意性框图；

[0032] 图8为本发明实施例提供的访问行为日志采集装置的关联单元的示意性框图；

[0033] 图9为本发明实施例提供的访问行为日志采集装置的匹配子单元的示意性框图；

[0034] 图10为本发明实施例提供的访问行为日志采集装置的采集单元的示意性框图；

[0035] 图11为本发明实施例提供的计算机设备的示意性框图。

具体实施方式

[0036] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0037] 应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0038] 还应当理解，在此本发明说明书中所使用的术语仅仅是出于描述特定实施例的目

的而并不意在限制本发明。如在本发明说明书和所附权利要求书中所使用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0039] 还应当进一步理解，在本发明说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

[0040] 请参阅图1和图2，图1为本发明实施例提供的访问行为日志采集方法的应用场景示意图。图2为本发明实施例提供的访问行为日志采集方法的示意性流程图。该访问行为日志采集方法应用于管理服务器中。该管理服务器与终端以及应用服务器进行数据交互，其中，应用服务器采集API接口上传输的数据，并传输至管理服务器，管理服务器对这些数据进行链路组装和三层审计信息的处理，终端则用于用户访问行为事件的定义，管理服务器结合API接口上传输的数据以及终端所定义的用户访问行为事件进行关联，并通过管理服务器设定的采集配置信息，进行用户访问行为日志的采集，另外，在应用服务器上安装有agent，即应用客户端，应用客户端与管理端进行通讯，管理端一般为单独的服务器，通过管理端内设置的采集策略，实时推给应用客户端，结合应用客户端内置的采集开关进行实际采集策略的设定，当有用户通过用户终端发起的数据经过应用客户端的指定接口时，应用客户端采用字节增强技术对数据进行拦截，并采用实际采集策略对拦截的数据进行采集，并对采集到的数据进行缓存。

[0041] 图2是本发明实施例提供的访问行为日志采集方法的流程示意图。如图2所示，该方法包括以下步骤S110至S180。

[0042] S110、获取API接口上传输的数据，以得到待处理数据。

[0043] 在本实施例中，待处理数据是指在API接口上传输的数据，如用户标识、协议、请求和响应等数据，且在该数据上写入了链路ID及访问上下文信息后所形成的数据。

[0044] 具体地，应用服务器上的应用客户端获取来自管理端的采集策略；根据所述采集策略采用开关模式设定实际采集策略；采用字节增强技术对访问数据进行拦截；根据所述实际采集策略结合联动机制对拦截的数据采集用户行为数据、应用传输数据，如用户标识、协议、请求和响应等数据，以得到待处理数据。采集策略包括基于用户标识配置进行数据采集、基于应用服务配置进行数据采集、基于应用接口进行数据采集、根据应用接口的字段进行数据采集、基于所配置的频率进行数据采集以及基于配置的单次数据采集量进行数据采集中至少一种数据采集方式。

[0045] 相对于网络流量采集和应用网关部署采集受限于特定的网络区域和部署位置，采集范围受限，本实施例在部署时以客户端即应用客户端的方式在应用服务主机上进行部署，部署简单，采集范围不受限制，有统一管理的管理端对应用客户端进行统一的配置管理；在管理端可以基于应用服务、用户标识、应用接口、应用接口字段、采集频率、数据采集量配置单个或组合数据采集策略进行用户行为数据、应用传输数据和应用资产数据的采集，实现数据采集的可管控性，通过在管理端更改采集策略，便可更改应用客户端对数据采集的策略，以达到数据采集的可控性。可以快速的配置数据采集策略，无需进行定点数据采集开发，节约了人力，大大缩短开发周期，提升了工作和项目效率；无需埋点和二开的用户行为数据、应用传输数据和应用资产数据的采集，减少了应用的运行维护和对应用系统的影响，从而可以更多的保障应用的正常运行。在部署时以客户端的方式在应用服务主机上进行部署，有统一管理的管理端对应用客户端进行统一的配置管理；对应用系统架构无影

响，对现有的网络架构无任何影响。

[0046] 具体地，上述的基于用户标识配置进行数据采集的数据采集方式是指基于用户标识配置是否采集应用传输数据，指定单个或多个或全部用户可以或不可以采集用户行为数据应用传输数据和应用资产数据，其中用户标识包括访问的用户账号、访问IP、访问mac、访问浏览器等。基于应用服务配置进行数据采集的数据采集方式是指基于应用服务配置是否采集应用传输数据和应用资产数据，指定单个或多个或全部应用服务可以或不可以采集应用传输数据。基于应用接口进行数据采集的数据采集方式是指基于应用接口是否采集应用传输数据和应用资产数据，指定单个或多个或全部应用接口可以或不可以采集应用传输数据。根据应用接口的字段进行数据采集的数据采集方式是指应用接口的字段配置是否采集应用传输数据，主要通过接口名称上的接口字段进行控制，指定单个或多个或全部的应用接口字段可以或不可以采集应用传输数据。基于所配置的频率进行数据采集的数据采集方式是通过采集数据的时间周期进行控制。基于配置的单次数据采集量进行数据采集的数据采集方式主要通过需要采集的数据条数进行控制。以上的数据采集方式任意组合配置是否采集应用传输数据，比如允许采集某个用户访问某个接口的行为数据或不允许采集指定或全部用户访问指定的某个应用服务的行为数据；由此形成采集策略。

[0047] 在本实施例中，实际采集策略是指实际用于采集数据的策略，结合了客户端所传输的采集策略以及设定的开关模式，由此组成的多种数据采集策略，且这些数据采集策略按照设定的优先级依序执行。

[0048] 在本实施例中，所述开关模式包括基于应用服务设置数据采集的开关、基于用户标识设置数据采集的开关、基于应用接口设置数据采集的开关、基于应用接口的字段设置数据采集的开关中至少一种模式；所述开关模式还包括数据采集的周期阀值以及基于单次数据的采集量的过滤规则。

[0049] 具体地，基于应用服务设置数据采集的开关的模式是指当状态为启用时，应用客户端将打开应用服务的数据采集功能，这是采集功能可以正常工作的前置条件；当状态为关闭时，应用客户端将关闭应用服务的采集功能，此应用服务下所有采集功能将全部被关闭。

[0050] 基于用户标识设置数据采集的开关的模式是指当状态为启用时，应用客户端打开基于用户标识的数据采集功能；当状态为关闭时，应用客户端关闭基于用户标识的数据采集功能，但其它采集开关和数据采集不受影响，只是不再基于用户标识采集数据。基于应用接口设置数据采集的开关的模式是指当状态为启用时，默认对所有应用接口执行数据采集，也可以对指定接口的数据定义采集或不采集，其中指定接口可以是单个或多个或全部接口名称结合采集或不采集的标示；当状态为关闭时，默认将关闭基于此应用服务下的应用接口的数据采集功能，但可对指定接口如单个或多个或全部接口名称结合采集的标示的数据进行采集，此时其它采集开关和数据采集不受影响。基于应用接口的字段设置数据采集的开关的模式指当状态为启用时，默认对所有接口和所有字段执行数据采集，对指定接口和指定字段如单个或多个或全部接口名称加上对应接口的字段名称以及不采集的标示的数据不进行采集；当状态为关闭时，将关闭基于应用接口字段的数据采集功能，但其它采集开关和数据采集不受影响，只是不再基于应用接口字段采集应用传输数据，但对指定接口和指定字段如单个或多个或全部接口名称加上对应接口的字段名称以及采集的标示的

数据进行采集，此时其它采集开关和数据采集不受影响。数据采集的周期阀值是通过采集数据的时间周期计数对需要采集的数据进行控制，针对采集的数据主体为单个的应用服务、服务中的应用接口及接口字段；基于单次数据的采集量的过滤规则是针对单个的应用服务、服务中的应用接口及接口字段的数据条数抽取进行过滤，过滤时将采取前置数据抽取、后置数据抽取、随机抽取、从某个特定数据点进行数据抽取等方式进行过滤。

[0051] 管理端与应用客户端建立双向的通讯通道，可以随时发送指令给应用客户端，应用客户端亦可根据指令执行并返回执行结果，应用客户端也可通过这个双向的通讯通道主动发送应用客户端的状态等信息给管理端。初始安装时，应用客户端与管理端建立通讯后，管理端可按照应用客户端群组或按照单个的应用客户端进行采集策略的推送，当数据采集策略有进行更新时，亦会实时进行策略推送，应用客户端针对接收到的数据策略，根据采集策略中的不同采集指令执行不同的数据采集逻辑，从而达到可管可控数据采集的功能。

[0052] 在本实施例中，访问数据是由用户终端发起的数据；拦截的数据是指经过应用客户端的指定接口的数据。

[0053] 具体地，在Java字节码生成之后，根据Java字节码中定义的规则，对已经生成的Java字节码在JVM加载时进行动态修改，增加增强功能的内容，以根据所述增强功能的内容对访问数据进行拦截，以得到拦截的数据。使用字节码增强技术，在Java字节码生成之后，根据Java字节码中定义的规则，对已经生成的Java字节码在JVM(Java虚拟机，Java Virtual Machine)加载的时候，进行动态修改，增加需要的字段或者是方法函数、或者继承实现新的类和接口等。这些动态增加的字段、或者方法函数、或者继承实现新的类和接口，都是增强的功能，增强的功能主要是指对方法函数的请求参数和返回结果进行自动提取的功能。

[0054] 在提取相关数据之前，首先确定所关心的数据在用户一次访问得过程中必定会经过类的方法函数。确定这些方法函数之后，通过字节码增强技术在这些方法函数的调用前和调用后采集数据的逻辑，此时可以根据实际的需求采集到所需要采集到的应用系统，上述的方法函数包括Object first＝methond(String paramer)和/或Object second＝first.methond(String paramer)；首先会将methond(String paramer)方法的主体内容抽取为一个方法名随机字符串的名称的函数中，比如sdfsdfsd(paramer)；接着将methond(String paramer)方法修改为：

```
methond(String paramer){
    befor(){
    }
    sdfsdfsd(paramer);
    after(){
    }
}。
```

[0055] 最后在befor()和after()里面采集到的数据会存储在first对象中的某个变量中，这个变量是增强进去的。这样这个变量中的值就可以传递到first.methond(String paramer)中的befor和after方法中，这样就能将用户的数据在一次访问的整个生命周期中串联起来。

[0056] 在提取得数据中，有不同类型的数据、用户标识信息比如用户账号、IP、mac、浏览器等，用户访问时间、访问结果、访问接口、访问数据以及访问次数等，不过这些数据都是分散在用户访问过程中的不同地方，通过上述的方法函数，可以将这些数据汇聚在一起并汇聚到任何的数据接入方，也就是采用字节码增强技术将所有的数据进行拦截汇集。

[0057] 当用户访问和任务调度的访问数据流和接口调用经过应用客户端的Agent时，应用客户端将使用字节码增强技术对流经的数据进行拦截，基于字节码增强技术中获取的也是应用层协议解析后的传输数据，因此无需对协议进行解析和进行内容还原，所以在采集数据时不受加密协议和私有协议的影响。无需埋点和二开的用户行为数据、应用传输数据和应用资产数据采集，减少了应用的运行维护和对应用系统的影响，从而可以更多的保障应用的正常运行；对应用系统架构无影响，对现有的网络架构无任何影响。因在应用服务器上部署了应用客户端，可以直接在应用系统上采集到应用组件、应用主机IP、应用实例，再结合上述两种数据采集，可以采集应用资产数据为应用名称、应用服务名称、应用接口、应用组件、应用主机IP、应用实例、数据库名称、数据库IP、数据库端口、数据库类别等。

[0058] 具体地，根据实际采集策略对拦截的数据流进行数据解析，采集所需的用户行为数据、应用传输数据和应用资产数据。

[0059] 在本实施例中，根据所述实际采集策略内不同的数据采集逻辑对拦截的数据采集用户行为数据、应用传输数据和应用资产数据；其中，不同的数据采集逻辑按照设定的优先级执行数据采集。

[0060] 另外，上述的所述用户行为数据包括用户标识、采集时间、用户的访问行为、访问接口、访问数据条数以及访问数据量，所述用户标识包括用户账号、访问IP、访问mac以及访问浏览器；所述应用传输数据包括用户标识、数据传输的时间、协议中所有字段、协议中所有字段所对应的内容、传输的所有请求数据以及传输的所有响应数据；上述的应用传输数据包括用户标识、数据传输的时间、协议字段如http和RPC(远程过程调用，Remote Procedure Call)等协议所有字段、协议内容如http和RPC(远程过程调用，Remote Procedure Call)等协议字段所有内容、传输的所有请求数据以及传输的所有响应数据。上述的应用资产数据包括用户标识、应用名称、应用服务名称、应用接口、应用组件、应用主机IP、应用实例、数据传输的时间、数据库名称、数据库IP、数据库端口、数据库类别等。

[0061] 在本实施例中，上述的优先级为：

[0062] 当多条策略执行时按照优先级由高到低进行处理，从高到低按照基于应用服务配置进行数据采集的数据采集方式、基于所配置的频率进行数据采集的数据采集方式、基于配置的单次数据采集量进行数据采集的数据采集方式、基于应用接口进行数据采集的数据采集方式、根据应用接口的字段进行数据采集的数据采集方式、基于用户标识配置进行数据采集的数据采集方式进行处理。当多条策略发生冲突时，按照高优先级处理，从高到低按照基于应用服务配置进行数据采集的数据采集方式、基于所配置的频率进行数据采集的数据采集方式、基于配置的单次数据采集量进行数据采集的数据采集方式、基于应用接口进行数据采集的数据采集方式、根据应用接口的字段进行数据采集的数据采集方式、基于用户标识配置进行数据采集的数据采集方式进行处理。当执行的数据采集策略逻辑出现问题时，应用客户端将不执行数据采集指令，返回失败结果至管理端。

[0063] 相对预置埋点和定点二开，本实施例可采集的数据更全面，采集时只需简单配置，

因而使得用户行为数据、应用传输数据和应用资产数据的采集变得更简单和更灵活。

[0064] 在进行数据采集时，会对采集到的数据进行链路ID和上下文关系信息的写入，以形成待处理数据，以便于将用户的访问链路进行展示，以便于对单次用户访问及接口调用之间的单次数据在应用接口、数据库、应用服务、应用之间的传输路径进行梳理。使用节点来标示访问链路信息，其中，节点包括链路ID、父节点编号以及当前节点编号；用户每次访问用户终端时都会生成一个全局唯一的标识，这个标识即为链路ID，用这个链路ID来标识用户的某一次访问，用户的一次访问可能会涉及到多个服务即多个进程的情况中的多个方法接口函数，多个服务之间传递的数据一般分为头部和身体部分，就像http有头部和身体部分，RocketMQ也有MessageHeader，Message Body，身体部分一般放着业务数据，RocketMQ是一个队列模型的消息中间件，在用户访问下一个服务的时候，会将这个全局唯一的链路ID，还有当前节点的上下文关系信息放在通讯数据的头部，传递给下一个服务，每处理完一个服务就会将这个服务下的所有接口函数信息上传到数据接收处汇总数据。其中上下文关系信息使用父节点编号和当前节点编号进行标识，即上下文关系信息包括父节点和当前节点的编号信息；父节点编号和当前节点编号存在严格的顺序关系，使用自然数进行标识，父节点编号一般来自上一个节点的当前节点编号，第一个节点的父节点编号设置为0，当前节点为1，当第二个节点接收到第一个节点的数据时，会从上下文关系中获取第一个节点的当前节点编号，并作为第二个节点的父节点编号，然后基于父节点编号+1号作为当前节点编号，以此类推，直到节点结束，以下为链路访问信息写入的示意图：节点(链路ID(2fa91f5cf3941171)，父节点编号(0)，当前节点编号(1))、节点(链路ID(2fa91f5cf3941171)，父节点编号(1)，当前节点编号(2))、节点(链路ID(2fa91f5cf3941171)，父节点编号(2)，当前节点编号(3))、节点(链路ID(2fa91f5cf3941171)，父节点编号(3)，当前节点编号(4))……

[0065] 上述的链路标识方式较为便捷和清楚。

[0066] S120、对所述待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图。

[0067] 在本实施例中，访问链路信息图是指用户访问应用系统时，请求和响应的应用数据在传输过程中所经过的路径图，也包括单次用户访问及接口调用之间的单次数据在应用接口、数据库、应用服务、应用之间的传输路径。

[0068] 在本实施例中，对于提取与数据库相关的数据过程中，可采用在设定采集策略时通过设定与数据库访问的应用接口进行数据采集，从而确保采集到的待处理数据均为与数据库访问相关的数据，也可不设定采集策略，对采集到的待处理数据再进行过滤。

[0069] 在一实施例中，上述的步骤S120可包括步骤S121～S122。

[0070] S121、对所述待处理数据过滤出与数据库访问相关的数据。

[0071] 在本实施例中，与数据库访问相关的数据是指基于数据库访问的应用接口所采集到的数据。

[0072] 在一实施例中，上述的步骤S121可包括步骤S1211～S1213。

[0073] S1211、提取所述待处理数据中所有数据库访问的应用接口的信息。

[0074] 在本实施例中，数据库访问的应用接口的信息是指基于数据库访问的应用接口所采集到的用户行为数据以及应用传输数据。

[0075] S1212、基于数据库访问应用接口查询关联的数据访问链路信息；

[0076] S1213、提取所有数据库关联的数据访问链路信息的用户行为数据和应用传输数据，以得到与数据库访问相关的数据。

[0077] 采集到的待处理数据汇聚后，将对原始的采集数据进行过滤并提取所需采集数据。

[0078] S122、根据与数据库访问相关的数据组装访问链路，以得到访问链路信息图。

[0079] 在一实施例中，上述的步骤S122可包括步骤S1221～S1224。

[0080] S1221、从与数据库访问相关的数据内提取出相同链路ID的节点，以得到目标节点。

[0081] 在本实施例中，目标节点是指链路ID相同的节点。

[0082] 对待处理数据进行链路ID相同的节点归类，以便于后续应用运维管理的故障分析、数据流分析。

[0083] S1222、提取所述目标节点提取上下文关系信息。

[0084] 在本实施例中，每个目标节点内都带有链路ID、父节点编号以及当前节点编号，因此，当确定目标节点后，根据上下文关系信息是为了确定访问的路径流向。

[0085] S1223、根据所述上下文关系信息进行节点访问顺序的排序，以得到排序结果。

[0086] 在本实施例中，排序结果是指节点访问的顺序，从时间的先后顺序进行排序。

[0087] S1224、根据所述排序结果提取每个节点的请求和响应的应用接口信息，以得到基于应用接口的访问链路信息图。

[0088] 在本实施例中，基于应用接口的访问链路信息图是指每个节点的请求和响应的应用接口的访问路径构成的信息图。

[0089] 数据访问链路可以梳理出用户访问的数据路、接口调用；便于应用运维管理的故障分析、数据流分析等，数据访问链路将梳理出基于用户或接口调用的访问顺序路径及访问的数据，便于安全事件中的安全分析和溯源分析。

[0090] S130、根据所述访问链路信息图提取数据库审计信息，并对所述数据库审计信息进行关联，以得到三层关联审计信息。

[0091] 在本实施例中，三层关联审计信息包括用户访问信息、应用访问对象的相关信息字段、数据库对象、数据库IP、数据库账号的字段信息、数据库访问接口字段信息、操作结果、操作数据行数、操作数据集的字段信息、使用应用接口的访问链路信息、原始的SQL语句信息以及数据库操作行为字段信息。具体地，提取访问用户、用户IP、访问时间、访问结果、数据库访问接口、数据库名称、数据库IP、数据库连接账号、操作对象、操作行为等信息，得到数据库三层关联审计信息。

[0092] 在一实施例中上述的步骤S130可包括步骤S131～S139。

[0093] S131、基于所述访问链路信息图中的单条数据访问链路，从与数据库访问相关的数据中的用户行为数据中获取用户访问数据库的行为数据。

[0094] 在本实施例中，用户访问数据库的行为数据是指用户终端发起访问web应用且web应用，web应用再对数据库访问时对应的行为数据。

[0095] S132、从获取的行为数据中提取用户访问信息以及应用访问对象的相关信息字段。

[0096] 在本实施例中，用户访问信息包括应用层用户、用户IP、访问时间、访问结果字段信息，应用访问对象的相关信息字段包括应用访问对象的应用端IP字段信息。基于单条数据访问链路从用户行为数据中获取到用户访问数据库系统行为数据，并从中提取到应用层用户、用户IP、访问时间、访问结果字段信息。基于单条数据访问链路从用户行为数据中获取到用户访问数据库系统行为数据，并从用户访问数据库系统行为数据中提取应用访问对象的应用端IP字段信息。

[0097] S133、基于所述访问链路信息图中的单条数据访问链路，从与数据库访问相关的数据中的应用传输数据中提取数据库对象、数据库IP以及数据库账号的字段信息；

[0098] S134、基于所述访问链路信息图中的单条数据访问链路，提取数据库访问接口字段信息；

[0099] S135、基于所述访问链路信息图中的单条数据访问链路，从与数据库访问相关的数据中的应用传输数据中提取操作结果、操作数据行数以及操作数据集的字段信息；

[0100] S136、基于所述访问链路信息图中的单条数据访问链路，提取使用应用接口的访问链路信息；

[0101] S137、基于所述访问链路信息图中的单条数据访问链路，从与数据库访问相关的数据中的应用传输数据中提取原始的SQL语句信息；

[0102] S138、对原始的SQL语句信息进行词法和语法解析，获取数据库操作行为字段信息；

[0103] S139、存储用户访问信息、应用访问对象的相关信息字段、数据库对象、数据库IP、数据库账号的字段信息、数据库访问接口字段信息、操作结果、操作数据行数、操作数据集的字段信息、使用应用接口的访问链路信息、原始的SQL语句信息以及数据库操作行为字段信息，以得到三层关联审计信息。

[0104] 用户的每一次访问，都将用户标识、数据库接口及操作数据库的行为形成了访问链路，针对数据库的三层关联审计准确度高达100％。

[0105] 本实施例的方法能将应用访问用户对web服务器访问及web服务器对数据库的访问关联起来。

[0106] 通过以下的展示方式和展示字段进行数据库的三层关联审计信息的展示。

[0107] 以列表的方式进行展示，设置筛选条件，筛选字段为应用层用户、用户IP、应用端IP、数据库IP、数据库实例、操作类型，进行三层审计信息的列表展示方式。

[0108] 列表展示字段包括：应用层用户-显示访问应用的用户名称；用户IP-显示访问应用系统的用户端IP地址；访问时间-显示用户访问的时间，显示格式为日期+时分秒；访问结果-显示用户访问的结果，成功或失败；应用端IP-显示调用数据库系统的应用端IP地址；数据库访问接口-显示数据库访问接口名称；数据库连接账号-显示应用端连接数据库的账号名称；数据库IP-显示数据IP地址；数据库实例-显示数据库实例名称或数据库名称；数据库表名-显示数据库表名称；数据库字段名-显示数据库字段名称；操作类型-显示查询或增加或修改或删除等操作类型；操作SQL语句-显示操作数据库的原始SQL语句；操作结果-显示操作结果，成功或是失败；操作数据行数-显示操作数据影响的行数；操作数据集-点击可查看用户访问的样本数据，如查询返回数据、用户更新数据；访问链路-为链接信息，点击可查看用户访问的基于应用接口的数据访问链路信息。

[0109] S140、获取终端定义的用户访问行为事件的标签，以得到数据标签。

[0110] 在本实施例中，数据标签是指对用户访问行为事件进行名称和备注信息的标注所形成的标签，具体包括用户访问URL、针对用户访问URL的访问事件名称、URL字段等备注信息。

[0111] 在一实施例中，请参阅图3，上述的步骤S140可包括步骤S141～S144。

[0112] S141、由终端启动浏览器插件，并利用所述浏览器插件拦截用户交互动作的请求信息或响应信息。

[0113] 在本实施例中，在终端上安装并启动浏览器插件，具体是在用户访问应用的浏览器上安装并启动浏览器插件；当用户访问应用系统并执行交互动作时，可借助浏览器插件拦截用户交互动作的请求信息或响应信息。浏览器插件的核心功能如下：通过管理服务器进行认证管理；通过重写浏览器底层API的请求与接收方法，比如XML Http Request，拦截用户访问应用时的请求和响应数据，并对数据进行定位分析，目前支持ajax、fetch请求及响应类型；在浏览器插件上进行数据标签操作并将标签数据同步到管理平台；支持的浏览器类型：谷歌、火狐、360、QQ、搜狗。在终端的浏览器上通过扩展程序进行插件安装；启动浏览器插件；基于浏览器上弹出交互界面，操作前需要进行认证登录操作：浏览器插件需要进行登录认证，登录认证来自管理服务器的用户管理和角色管理，只有被授权使用浏览器插件的用户才能登录使用。

[0114] 用户在浏览器上登录应用系统，基于访问页面执行请求和响应的交互动作，以下针对请求和响应的使用以下示例进行说明：请求信息是指注册账号输入个人信息，如姓名、手机号码、地址、邮箱，并执行提交动作；响应信息是指查询用户手机号码并获取列表信息，如姓名、手机号码。

[0115] 另外，拦截请求信息时，具体操作如下：以注册账号输入个人信息，如用户账号、姓名、手机号码、年龄，并提交为例说明，此时插件将拦截到用户访问的URL(统一资源定位器，Uniform Resource Locator)，URL格式为‘协议类型://服务器地址[:端口号]/路径/文件名[参数＝值]’，获取请求信息中的数据，如姓名、手机号码、地址、邮箱，以及数据对应的字段。

[0116] 拦截响应信息时，具体操作如下：以用户账号查询并获取用户列表信息，比如用户账号、姓名、手机号码、年龄，为例说明，此时插件将拦截到用户访问的URL，URL格式为‘协议类型://服务器地址[:端口号]/路径/文件名[参数＝值]’，获取响应信息中的数据，如姓名、手机号码，以及数据对应的字段。

[0117] S142、由终端根据所拦截的信息定位用户访问指定的页面数据对应的URL及字段。

[0118] 在本实施例中，由终端从用户访问页面复制指定的数据，并与所拦截的信息通过使用关键字或正则进行双向模糊匹配，当匹配到数据，获取匹配的数据所对应字段和URL，若页面数据上不同字段对应的数据不相同，则进行再次定位；若页面数据上存在不同字段对应的数据有相同值，则从上一次匹配结果中再次执行页面交互和定位动作，以获取用户访问指定的页面数据对应的URL及字段。

[0119] 具体地，对于请求信息的匹配及访问字段和URL定位时，从用户访问页面复制指定的提交数据如文本框/下拉框/单选框/复选框等输入数据，以注册账号时的‘姓名’提交数据作为示例说明，用户输入的‘姓名’信息数据与插件拦截到的请求数据通过使用关键字或

正则进行双向模糊匹配，即姓名会与拦截数据进行模糊匹配，拦截的数据会与姓名进行模糊匹配，如果匹配到数据，则获取匹配的请求数据所对应字段和URL，如果页面上不同字段对应的数据并不相同，一般可以一次定位，大部分时候会定位到一个URL和一个字段，如果页面上存在不同字段对应的数据有相同值，则用户可以在上一次匹配结果中再次执行页面交互和定位动作，需定位字段取数尽量不要与其它字段取数相同，直至完成字段和URL定位。

[0120] 当进行响应数据的匹配及访问字段和URL定位时，从用户访问页面复制响应返回数据如列表字段、概要统计、图形统计等对应数据，以通过员工工号查询并获取员工的个人数据信息作为示例说明，使用查询返回的‘手机号码’数据与插件拦截到的响应数据通过使用关键字或正则进行双向模糊匹配，即手机号码会与拦截数据进行模糊匹配，拦截的数据会与手机进行模糊匹配，如果匹配到数据，则获取匹配的请求数据所对应字段和URL，如果页面上不同字段对应的数据并不相同，一般可以一次定位，大部分时候会定位到一个URL和一个字段，如果页面上存在不同字段对应的数据有相同值，则用户可以在上一次匹配结果中再次执行页面交互和定位动作，需定位字段取数尽量不要与其它字段取数相同，直至完成字段和URL定位。

[0121] S143、由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义，以得到用户访问行为事件的标签。

[0122] 具体地，由终端对用户访问指定的页面数据对应的URL及字段定义访问事件名称及备注信息，以得到用户访问行为事件的标签。

[0123] 当应用系统页面上指定的数据被定位后，则可以对定位到URL字段进行用户访问行为事件的操作，用户访问行为的操作主要是对以下用户访问URL定义一个访问事件及对URL字段进行一个备注说明：

[0124] 用户访问URL事件的备注时，用户可自定义URL接口事件名称，如URL事件名称为‘查询员工信息’；URL字段备注时，可针对手机号码字段写下备注为‘手机号码’。

[0125] S144、获取用户访问行为事件的标签，以得到数据标签。

[0126] 管理服务器从终端获取用户访问URL、针对用户访问URL的访问事件名称、URL字段等备注信息。

[0127] S150、获取用户访问的URL，以得到待关联URL。

[0128] 在本实施例中，待关联URL是指用户访问的URL。

[0129] 管理服务器从浏览器插件获取到用户访问的URL，在根据存储在管理服务器上的API接口数据，根据用户访问URL去匹配对应的API接口。

[0130] S160、根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果。

[0131] 在本实施例中，关联结果是指待关联URL与API接口进行关联所得的结果。

[0132] 在一实施例中，请参阅图4，上述的步骤S160可包括步骤S161～S163。

[0133] S161、对所述待关联URL匹配对应的API接口，以得到目标API接口。

[0134] 在本实施例中，目标API接口是指依据待关联URL生成的数组与依据API接口生成的数组中的字符串数量和值等同；两个数组中的字符串数量为变量，依据API接口生成的数组中对应位置为变量，这符合这两个条件的API接口。

[0135] 在一实施例中，请参阅图5，上述的步骤S161可包括步骤S1611～S1615。

[0136] S1611、对所述待关联URL进行信息分割，以得到访问URL路径。

[0137] 在本实施例中，访问URL路径是指访问URL信息中的指定位置的信息。

[0138] 根据"//"从访问URL中分离出协议信息。以访问URL为'https://主机:端口/app/main/user/query'作为示例，本示例中为https，再根据"/"分割url得出'主机、端口和/app/main/user/query'，其中'/app/main/user/query'为路径。

[0139] S1612、对所述访问URL路径进行分割，以得到第一有序数组。

[0140] 在本实施例中，第一有序数组是指对访问URL路径进行单个字段的划分，以得到的数组。将路径/app/main/user/query再根据"/"逐级分割成/app、/main、/user、/query，得到有序数组A，由此得到第一有序数组。

[0141] S1613、对所述API接口进行信息分割和路径分割，以得到第二有序数组。

[0142] 对于API接口也可按照上述的步骤S1411～S1412进行分割，得到有序数组B，即第二有序数组；此处不再赘述。

[0143] S1614、遍历所述第一有序数组，并将所述第一有序数组与所述第二有序数组相同位置的值进行对比，以确定所述待关联URL所匹配成功的API接口。

[0144] 遍历有序数组A，逐个与有序数组B中相同位置值进行比较，如下两种情况，认为用户访问URL与API接口匹配成功：一种是有序数组A和有序数组B中的字符串数量和值等同；另外一种是有序数组A和有序数组B中的字符串数量为变量，有序数组B中对应位置为变量。

[0145] S1615、对所述待关联URL所匹配成功的API接口进行打标识，以得到目标API接口。

[0146] 对用户访问URL匹配成功的API接口打上对应的标识，标识此用户访问URL与匹配成功的API接口等同。

[0147] S162、将所述待关联URL对应的字段与所述目标API接口对应的字段进行关联。

[0148] 在本实施例中，当用户访问URL与目标API接口关联对应后，此时对比字段名称，将用户访问URL下的字段名称与目标API接口下的字段名称进行，值相同即进行关联，标识用户访问URL下的字段与目标API接口下的字段等同。

[0149] S163、将所述数据标签中关于所述待关联URL对应的字段所对应的访问事件名称及备注信息关联至所述目标API接口对应的字段，以得到关联结果。

[0150] 在本实施例中，根据数据标签中对该待关联URL定义的访问事件名称及URL下字段所定义的备注信息，等值将其关联传递到已关联的目标API接口及相同的字段下。举个例子：用户访问URL为https://10.10.20.33:8443/app/main/user/query，关联到API接口为/app/main/user/query；则自定义的URL事件及字段备注信息如下：URL事件：查询员工信息；URL字段'phone'字段备注：手机号码；关联到的应用接口事件为：查询员工信息；关联到的应用接口下的字段'phone'备注为：手机号码，此处的应用接口是指目标API接口。

[0151] S170、设置用户访问行为日志的采集配置信息。

[0152] 在本实施例中，所述采集配置信息包括用户访问行为名称、采集开关、进行用户访问行为日志采集的API接口以及进行用户访问行为日志采集的API接口中的字段。

[0153] 通过自定义管理，可以针对用户行为日志的采集可以达到即配即采的效果，即配置后即生效并可采集日志的效果，采集周期短，见效快。

[0154] 针对用户访问URL关联到的目标API接口、目标API接口上的访问事件名称及应用

接口字段备注，此时管理服务器还不会针对用户的访问行为日志进行采集，需要在管理服务器上进行采集配置和管理，如下：

[0155] 用户访问行为名称定义：用户可以重新定义一个访问行为名称，也可以使用此前用户在浏览器上定义的访问事件名称，如‘查询员工信息’。

[0156] 采集开关：选择‘开’则代表采集用户访问行为日志，选择‘关’则关闭用户访问行为日志的采集。

[0157] 采集的应用接口选择：因为用户访问行为的操作已经定位到用户访问URL并最终定位到系统应用接口上，因此需要选择用户访问行为对应的应用接口。

[0158] 采集的应用接口字段选择：选择完成目标API接口后，可以针对目标API接口下的字段进行选择，此处选择的字段将在日志中进行记录，并可根据需要记录用户访问的具体数据，也可针对选择的字段记录用户访问的样本数据。

[0159] S180、根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

[0160] 当在管理服务器上完成用户访问行为日志的采集配置后，此时管理服务器将会根据日志采集配置并结合应用客户端上传的原始采集数据、三层关联审计信息、访问链路信息图等采集用户访问行为日志。

[0161] 在一实施例中，请参阅图6，上述的步骤S180可包括步骤S181～S185。

[0162] S181、从访问链路信息图中筛选出与进行用户访问行为日志采集的API接口相关的链路，以得到待判定链路。

[0163] 在本实施例中，待判定链路是指访问链路信息图中与进行用户访问行为日志采集的API接口相关的链路。

[0164] 具体地，获取采集配置信息中的进行用户访问行为日志采集的API接口和选择的字段；根据该API接口查找此API接口对应的每一条用户访问链路。

[0165] S182、判断所述待判定链路中是否存在数据库访问接口节点。

[0166] 具体地，根据用户访问的链路查找待判定链路对应的整个链路中是否存在数据库访问接口节点。

[0167] S183、若所述待判定链路中存在数据库访问接口节点，则获取所述数据库访问接口节点相关的三层关联审计信息，并从获取的三层关联审计信息中提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息。

[0168] 在本实施例中，目标信息是指数据库访问接口节点相关的三层关联审计信息中用户访问行为相关的信息的条数以及访问数据量。

[0169] 具体地，获取所述数据库访问接口节点相关的三层关联审计信息，并从中提取以下信息：应用层用户、用户IP、访问时间、访问的应用、访问行为即采集配置信息中定义的用户访问行为名称、应用接口、访问的字段即采集配置信息中选择的字段、访问结果、应用端IP、数据库IP、数据库账号、SQL语句、操作行为、操作数据、操作结果、用户访问链路、请求或响应的数据等，并在此基础上统计访问的数据条数和访问数据量。

[0170] S184、将所述目标信息存储至日志缓存区。

[0171] 在本实施例中，将提取的目标信息发送到日志缓存区如kafka上；所有需要用户访问行为日志的应用或系统均可经过授权从日志缓存区获取用户访问行为日志。

[0172] S185、若所述待判定链路中不存在数据库访问接口节点，则通过所述待判定链路提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，并执行所述步骤S184。

[0173] 通过待判定链路直接提取以下信息：应用层用户、用户IP、访问时间、访问行为即采集配置信息中定义的用户访问行为名称、访问的字段即采集配置信息中选择的字段、访问结果、应用端IP、用户访问链路、请求或响应的数据等，并在此基础上统计访问的数据条数和访问数据量；以此类推，直到将此应用接口所关联的每次用户访问链路及其三层关联信息进行提取并存储。

[0174] 本实施例无需对应用进行改造，不需要在代码层面进行二次开发，不需要开发人员介入；因用户访问行为日志的采集不需要在代码层面进行二次开发，因此也对原有的业务存不存在入侵性；日志采集的新增、编辑、删除，均无需二次开发或修改代码，可以快速的在管理服务器上进行自定义日志采集配置信息和管理；用户访问行为日志中可以对用户访问数据的行为映射到数据库层面；用户访问行为日志中可以对用户访问数据的应用接口及应用接口路径进行梳理。

[0175] 上述的访问行为日志采集方法，通过获取基于字节码增强技术所采集用户访问行为数据和应用接口传输数据，并基于此数据组装用户访问链路、进行用户与应用及数据库的三层关联，在终端结合浏览器插件对访问页面的需要采集的用户访问行为事件进行定义，并根据定义的标签对三层关联审计信息与API接口行关联，且结合设置的采集配置信息，以进行采集用户访问行为日志，实现采集用户访问行为日志时，无需对应用进行改造，不需要在代码层面进行二次开发，对原有的业务不存在入侵性，维护成本低，对用户访问数据的行为关联到数据库层面；对用户访问数据的应用接口及应用接口路径进行梳理。

[0176] 图7是本发明实施例提供的一种访问行为日志采集装置300的示意性框图。如图7所示，对应于以上访问行为日志采集方法，本发明还提供一种访问行为日志采集装置300。该访问行为日志采集装置300包括用于执行上述访问行为日志采集方法的单元，该装置可以被配置于服务器中。具体地，请参阅图7，该访问行为日志采集装置300包括数据获取单元301、链路组装单元302、审计处理单元303、标签获取单元304、URL获取单元305、关联单元306、配置设置单元307以及采集单元308。

[0177] 数据获取单元301，用于获取API接口上传输的数据，以得到待处理数据；链路组装单元302，用于对所述待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图；审计处理单元303，用于根据所述访问链路信息图提取数据库审计信息，并对所述数据库审计信息进行关联，以得到三层关联审计信息；标签获取单元304，用于获取终端定义的用户访问行为事件的标签，以得到数据标签；URL获取单元305，用于获取用户访问的URL，以得到待关联URL；关联单元306，用于根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果；配置设置单元307，用于设置用户访问行为日志的采集配置信息；采集单元308，用于根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

[0178] 在一实施例中，所述标签获取单元304，用于由终端启动浏览器插件，并利用所述浏览器插件拦截用户交互动作的请求信息或响应信息；由终端根据所拦截的信息定位用户访问指定的页面数据对应的URL及字段；由终端对用户访问指定的页面数据对应的URL及字

段进行用户访问行为事件定义，以得到用户访问行为事件的标签；获取用户访问行为事件的标签，以得到数据标签。

[0179] 具体地，所述标签获取单元304，用于由终端对用户访问指定的页面数据对应的URL及字段定义访问事件名称及备注信息，以得到用户访问行为事件的标签。

[0180] 在一实施例中，如图8所示，所述关联单元306包括匹配子单元3061、字段关联子单元3062以及标签关联子单元3063。

[0181] 匹配子单元3061，用于对所述待关联URL匹配对应的API接口，以得到目标API接口；字段关联子单元3062，用于将所述待关联URL对应的字段与所述目标API接口对应的字段进行关联；标签关联子单元3063，用于将所述数据标签中关于所述待关联URL对应的字段所对应的访问事件名称及备注信息关联至所述目标API接口对应的字段，以得到关联结果。

[0182] 在一实施例中，如图9所示，所述匹配子单元3061包括信息分割模块30611、路径分割模块30612、数组获取模块30613、遍历模块30614以及打标识模块30615。

[0183] 信息分割模块30611，用于对所述待关联URL进行信息分割，以得到访问URL路径；路径分割模块30612，用于对所述访问URL路径进行分割，以得到第一有序数组；数组获取模块30613，用于对所述API接口进行信息分割和路径分割，以得到第二有序数组；遍历模块30614，用于遍历所述第一有序数组，并将所述第一有序数组与所述第二有序数组相同位置的值进行对比，以确定所述待关联URL所匹配成功的API接口；打标识模块30615，用于对所述待关联URL所匹配成功的API接口进行打标识，以得到目标API接口。

[0184] 在一实施例中，如图10所示，所述采集单元308包括链路筛选子单元3081、判断子单元3082、第一信息提取子单元3083、存储子单元3084以及第二信息提取子单元3085。

[0185] 链路筛选子单元3081，用于从访问链路信息图中筛选出与进行用户访问行为日志采集的API接口相关的链路，以得到待判定链路；判断子单元3082，用于判断所述待判定链路中是否存在数据库访问接口节点；第一信息提取子单元3083，用于若所述待判定链路中存在数据库访问接口节点，则获取所述数据库访问接口节点相关的三层关联审计信息，并从获取的三层关联审计信息中提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息；存储子单元3084，用于将所述目标信息存储至日志缓存区；第二信息提取子单元3085，用于若所述待判定链路中不存在数据库访问接口节点，则通过所述待判定链路提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息，并执行所述将所述目标信息存储至日志缓存区。

[0186] 需要说明的是，所属领域的技术人员可以清楚地了解到，上述访问行为日志采集装置300和各单元的具体实现过程，可以参考前述方法实施例中的相应描述，为了描述的方便和简洁，在此不再赘述。

[0187] 上述访问行为日志采集装置300可以实现为一种计算机程序的形式，该计算机程序可以在如图11所示的计算机设备上运行。

[0188] 请参阅图11，图11是本申请实施例提供的一种计算机设备的示意性框图。该计算机设备500可以是服务器，其中，服务器可以是独立的服务器，也可以是多个服务器组成的服务器集群。

[0189] 参阅图11，该计算机设备500包括通过系统总线501连接的处理器502、存储器和网络接口505，其中，存储器可以包括非易失性存储介质503和内存储器504。

[0190] 该非易失性存储介质503可存储操作系统5031和计算机程序5032。该计算机程序5032包括程序指令，该程序指令被执行时，可使得处理器502执行一种访问行为日志采集方法。

[0191] 该处理器502用于提供计算和控制能力，以支撑整个计算机设备500的运行。

[0192] 该内存储器504为非易失性存储介质503中的计算机程序5032的运行提供环境，该计算机程序5032被处理器502执行时，可使得处理器502执行一种访问行为日志采集方法。

[0193] 该网络接口505用于与其它设备进行网络通信。本领域技术人员可以理解，图11中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备500的限定，具体的计算机设备500可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0194] 其中，所述处理器502用于运行存储在存储器中的计算机程序5032，以实现如下步骤：

[0195] 获取API接口上传输的数据，以得到待处理数据；对所述待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图；根据所述访问链路信息图提取数据库审计信息，并对所述数据库审计信息进行关联，以得到三层关联审计信息；获取终端定义的用户访问行为事件的标签，以得到数据标签；获取用户访问的URL，以得到待关联URL；根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果；设置用户访问行为日志的采集配置信息；根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

[0196] 其中，所述采集配置信息包括用户访问行为名称、采集开关、进行用户访问行为日志采集的API接口以及进行用户访问行为日志采集的API接口中的字段。

[0197] 在一实施例中，处理器502在实现所述获取终端定义的用户访问行为事件的标签，以得到数据标签步骤时，具体实现如下步骤：

[0198] 由终端启动浏览器插件，并利用所述浏览器插件拦截用户交互动作的请求信息或响应信息；由终端根据所拦截的信息定位用户访问指定的页面数据对应的URL及字段；由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义，以得到用户访问行为事件的标签；获取用户访问行为事件的标签，以得到数据标签。

[0199] 在一实施例中，处理器502在实现所述由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义，以得到用户访问行为事件的标签步骤时，具体实现如下步骤：

[0200] 由终端对用户访问指定的页面数据对应的URL及字段定义访问事件名称及备注信息，以得到用户访问行为事件的标签。

[0201] 在一实施例中，处理器502在实现所述根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果步骤时，具体实现如下步骤：

[0202] 对所述待关联URL匹配对应的API接口，以得到目标API接口；将所述待关联URL对应的字段与所述目标API接口对应的字段进行关联；将所述数据标签中关于所述待关联URL对应的字段所对应的访问事件名称及备注信息关联至所述目标API接口对应的字段，以得到关联结果。

[0203] 在一实施例中，处理器502在实现所述对所述待关联URL匹配对应的API接口，以得

到目标API接口步骤时，具体实现如下步骤：

[0204] 对所述待关联URL进行信息分割，以得到访问URL路径；对所述访问URL路径进行分割，以得到第一有序数组；对所述API接口进行信息分割和路径分割，以得到第二有序数组；遍历所述第一有序数组，并将所述第一有序数组与所述第二有序数组相同位置的值进行对比，以确定所述待关联URL所匹配成功的API接口；对所述待关联URL所匹配成功的API接口进行打标识，以得到目标API接口。

[0205] 在一实施例中，处理器502在实现所述根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志步骤时，具体实现如下步骤：

[0206] 从访问链路信息图中筛选出与进行用户访问行为日志采集的API接口相关的链路，以得到待判定链路；判断所述待判定链路中是否存在数据库访问接口节点；若所述待判定链路中存在数据库访问接口节点，则获取所述数据库访问接口节点相关的三层关联审计信息，并从获取的三层关联审计信息中提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息；将所述目标信息存储至日志缓存区；若所述待判定链路中不存在数据库访问接口节点，则通过所述待判定链路提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息，并执行所述将所述目标信息存储至日志缓存区。

[0207] 应当理解，在本申请实施例中，处理器502可以是中央处理单元(Central Processing Unit,CPU)，该处理器502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。其中，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。

[0208] 本领域普通技术人员可以理解的是实现上述实施例的方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成。该计算机程序包括程序指令，计算机程序可存储于一存储介质中，该存储介质为计算机可读存储介质。该程序指令被该计算机系统中的至少一个处理器执行，以实现上述方法的实施例的流程步骤。

[0209] 因此，本发明还提供一种存储介质。该存储介质可以为计算机可读存储介质。该存储介质存储有计算机程序，其中该计算机程序被处理器执行时使处理器执行如下步骤：

[0210] 获取API接口上传输的数据，以得到待处理数据；对所述待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图；根据所述访问链路信息图提取数据库审计信息，并对所述数据库审计信息进行关联，以得到三层关联审计信息；获取终端定义的用户访问行为事件的标签，以得到数据标签；获取用户访问的URL，以得到待关联URL；根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果；设置用户访问行为日志的采集配置信息；根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志。

[0211] 其中，所述采集配置信息包括用户访问行为名称、采集开关、进行用户访问行为日志采集的API接口以及进行用户访问行为日志采集的API接口中的字段。

[0212] 在一实施例中，所述处理器在执行所述计算机程序而实现所述获取终端定义的用户访问行为事件的标签，以得到数据标签步骤时，具体实现如下步骤：

[0213] 由终端启动浏览器插件，并利用所述浏览器插件拦截用户交互动作的请求信息或响应信息；由终端根据所拦截的信息定位用户访问指定的页面数据对应的URL及字段；由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义，以得到用户访问行为事件的标签；获取用户访问行为事件的标签，以得到数据标签。

[0214] 在一实施例中，所述处理器在执行所述计算机程序而实现所述由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义，以得到用户访问行为事件的标签步骤时，具体实现如下步骤：

[0215] 由终端对用户访问指定的页面数据对应的URL及字段定义访问事件名称及备注信息，以得到用户访问行为事件的标签。

[0216] 在一实施例中，所述处理器在执行所述计算机程序而实现所述根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果步骤时，具体实现如下步骤：

[0217] 对所述待关联URL匹配对应的API接口，以得到目标API接口；将所述待关联URL对应的字段与所述目标API接口对应的字段进行关联；将所述数据标签中关于所述待关联URL对应的字段所对应的访问事件名称及备注信息关联至所述目标API接口对应的字段，以得到关联结果。

[0218] 在一实施例中，所述处理器在执行所述计算机程序而实现所述对所述待关联URL匹配对应的API接口，以得到目标API接口步骤时，具体实现如下步骤：

[0219] 对所述待关联URL进行信息分割，以得到访问URL路径；对所述访问URL路径进行分割，以得到第一有序数组；对所述API接口进行信息分割和路径分割，以得到第二有序数组；遍历所述第一有序数组，并将所述第一有序数组与所述第二有序数组相同位置的值进行对比，以确定所述待关联URL所匹配成功的API接口；对所述待关联URL所匹配成功的API接口进行打标识，以得到目标API接口。

[0220] 在一实施例中，所述处理器在执行所述计算机程序而实现所述根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志步骤时，具体实现如下步骤：

[0221] 从访问链路信息图中筛选出与进行用户访问行为日志采集的API接口相关的链路，以得到待判定链路；判断所述待判定链路中是否存在数据库访问接口节点；若所述待判定链路中存在数据库访问接口节点，则获取所述数据库访问接口节点相关的三层关联审计信息，并从获取的三层关联审计信息中提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息；将所述目标信息存储至日志缓存区；若所述待判定链路中不存在数据库访问接口节点，则通过所述待判定链路提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息，并执行所述将所述目标信息存储至日志缓存区。

[0222] 所述存储介质可以是U盘、移动硬盘、只读存储器(Read-Only Memory，ROM)、磁碟或者光盘等各种可以存储程序代码的计算机可读存储介质。

[0223] 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专

业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

[0224] 在本发明所提供的几个实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的。例如，各个单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。

[0225] 本发明实施例方法中的步骤可以根据实际需要进行顺序调整、合并和删减。本发明实施例装置中的单元可以根据实际需要进行合并、划分和删减。另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以是两个或两个以上单元集成在一个单元中。

[0226] 该集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分，或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，终端，或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。

[0227] 以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以权利要求的保护范围为准。

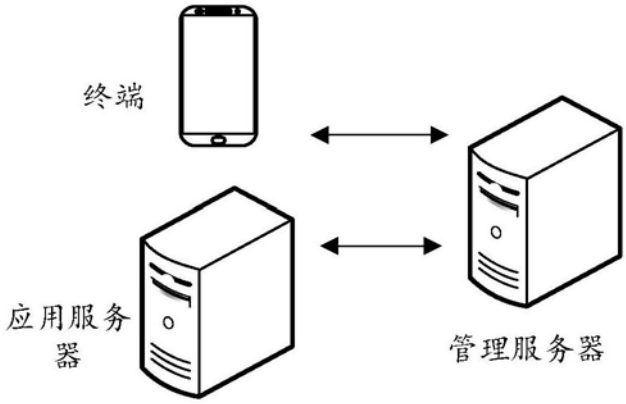

终端
应用服务器
管理服务器

图1

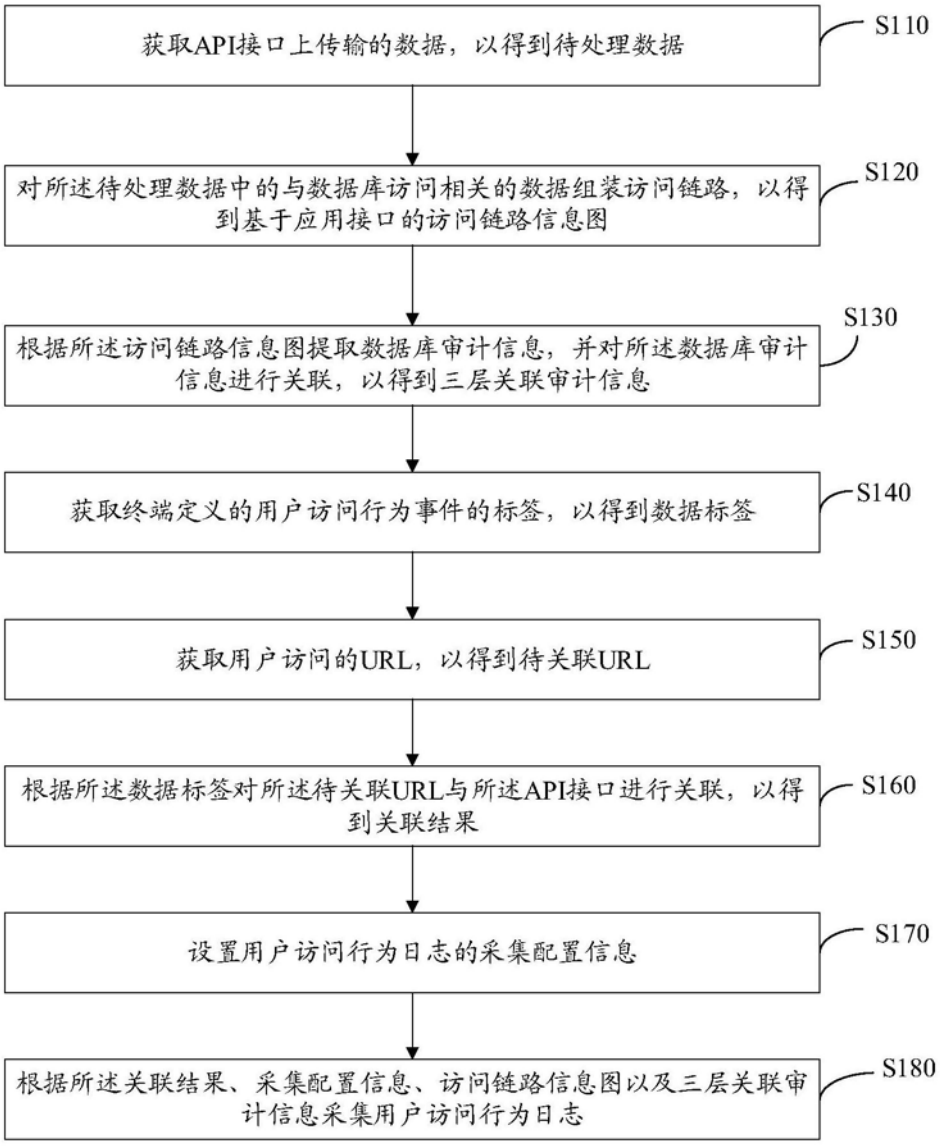

获取API接口上传输的数据，以得到待处理数据
S110
对所述待处理数据中的与数据库访问相关的数据组装访问链路，以得到基于应用接口的访问链路信息图
S120
根据所述访问链路信息图提取数据库审计信息，并对所述数据库审计信息进行关联，以得到三层关联审计信息
S130
获取终端定义的用户访问行为事件的标签，以得到数据标签
S140
获取用户访问的URL，以得到待关联URL
S150
根据所述数据标签对所述待关联URL与所述API接口进行关联，以得到关联结果
S160
设置用户访问行为日志的采集配置信息
S170
根据所述关联结果、采集配置信息、访问链路信息图以及三层关联审计信息采集用户访问行为日志
S180

图2

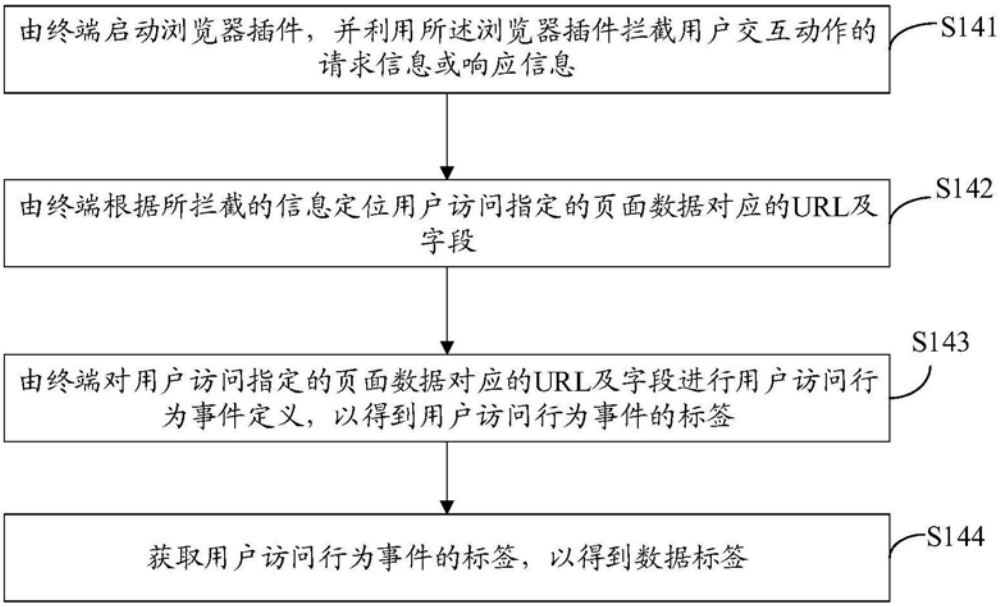

由终端启动浏览器插件，并利用所述浏览器插件拦截用户交互动作的请求信息或响应信息
S141
由终端根据所拦截的信息定位用户访问指定的页面数据对应的URL及字段
S142
由终端对用户访问指定的页面数据对应的URL及字段进行用户访问行为事件定义，以得到用户访问行为事件的标签
S143
获取用户访问行为事件的标签，以得到数据标签
S144

图3

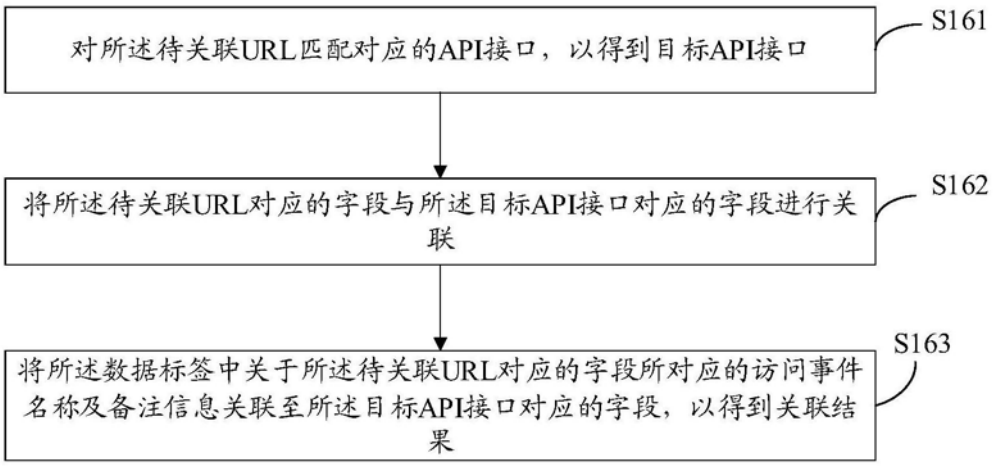

对所述待关联URL匹配对应的API接口，以得到目标API接口
S161
将所述待关联URL对应的字段与所述目标API接口对应的字段进行关联
S162
将所述数据标签中关于所述待关联URL对应的字段所对应的访问事件名称及备注信息关联至所述目标API接口对应的字段，以得到关联结果
S163

图4

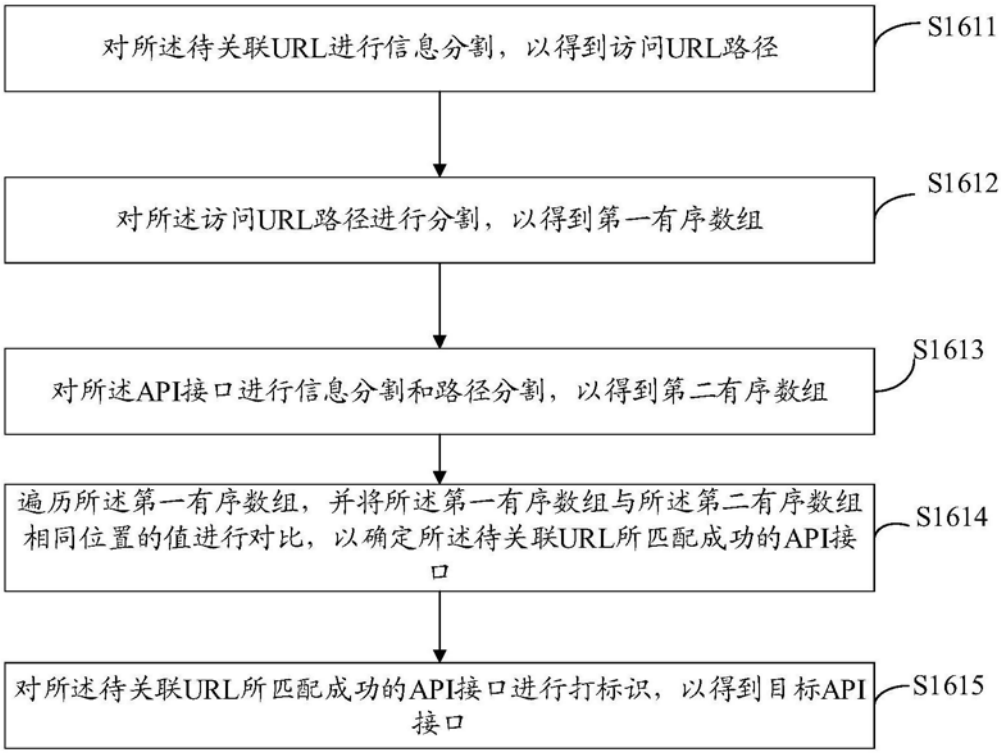

对所述待关联URL进行信息分割，以得到访问URL路径
S1611
对所述访问URL路径进行分割，以得到第一有序数组
S1612
对所述API接口进行信息分割和路径分割，以得到第二有序数组
S1613
遍历所述第一有序数组，并将所述第一有序数组与所述第二有序数组相同位置的值进行对比，以确定所述待关联URL所匹配成功的API接口
S1614
对所述待关联URL所匹配成功的API接口进行打标识，以得到目标API接口
S1615

图5

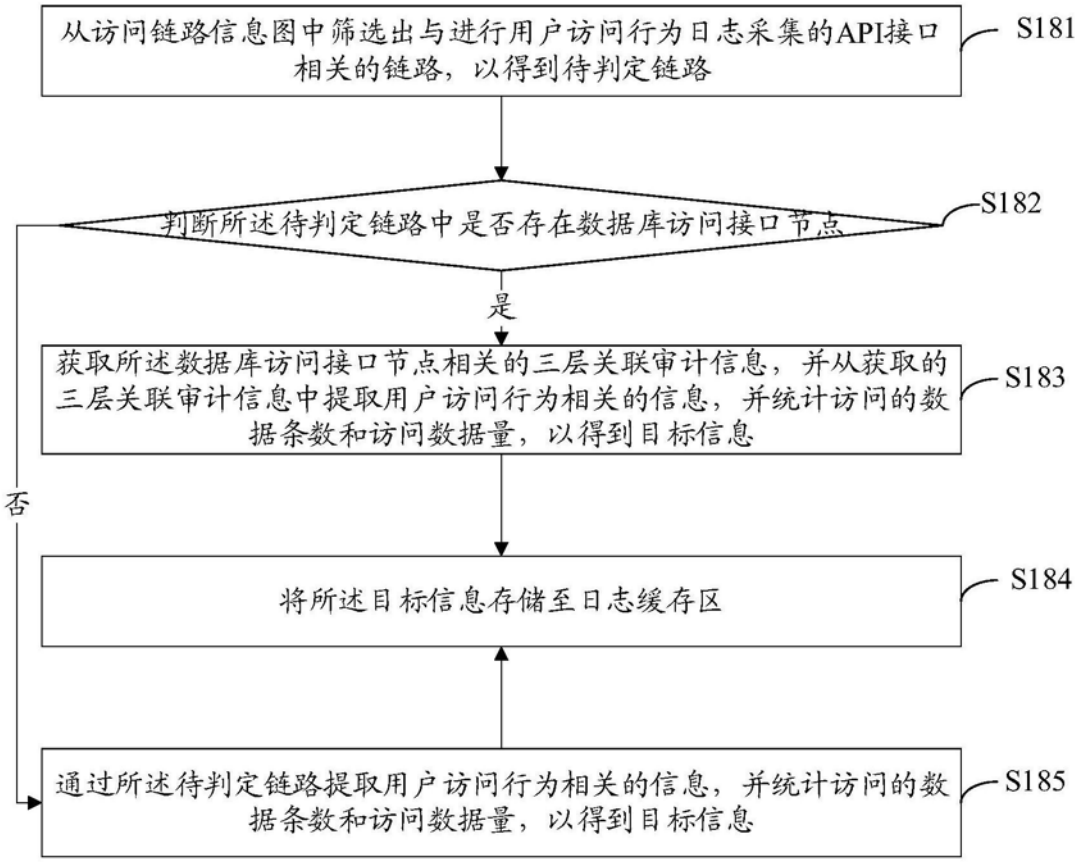

从访问链路信息图中筛选出与进行用户访问行为日志采集的API接口相关的链路，以得到待判定链路
S181
判断所述待判定链路中是否存在数据库访问接口节点
S182
是
否
获取所述数据库访问接口节点相关的三层关联审计信息，并从获取的三层关联审计信息中提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息
S183
将所述目标信息存储至日志缓存区
S184
通过所述待判定链路提取用户访问行为相关的信息，并统计访问的数据条数和访问数据量，以得到目标信息
S185

图6

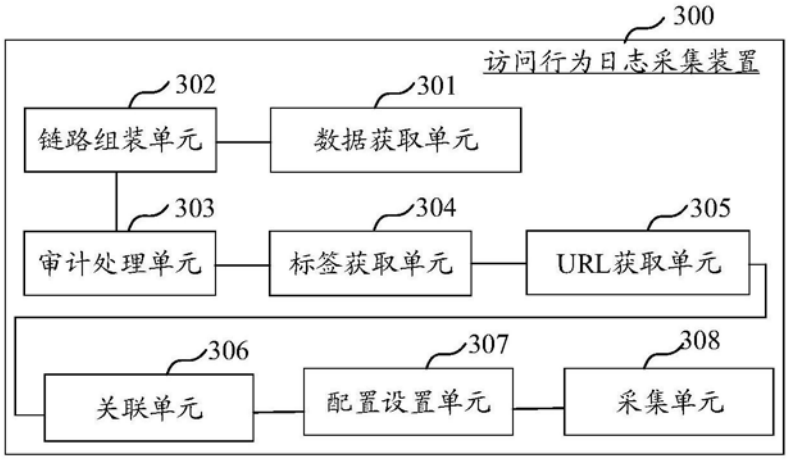

300
访问行为日志采集装置
302
链路组装单元
301
数据获取单元
303
审计处理单元
304
标签获取单元
305
URL获取单元
306
关联单元
307
配置设置单元
308
采集单元

图7

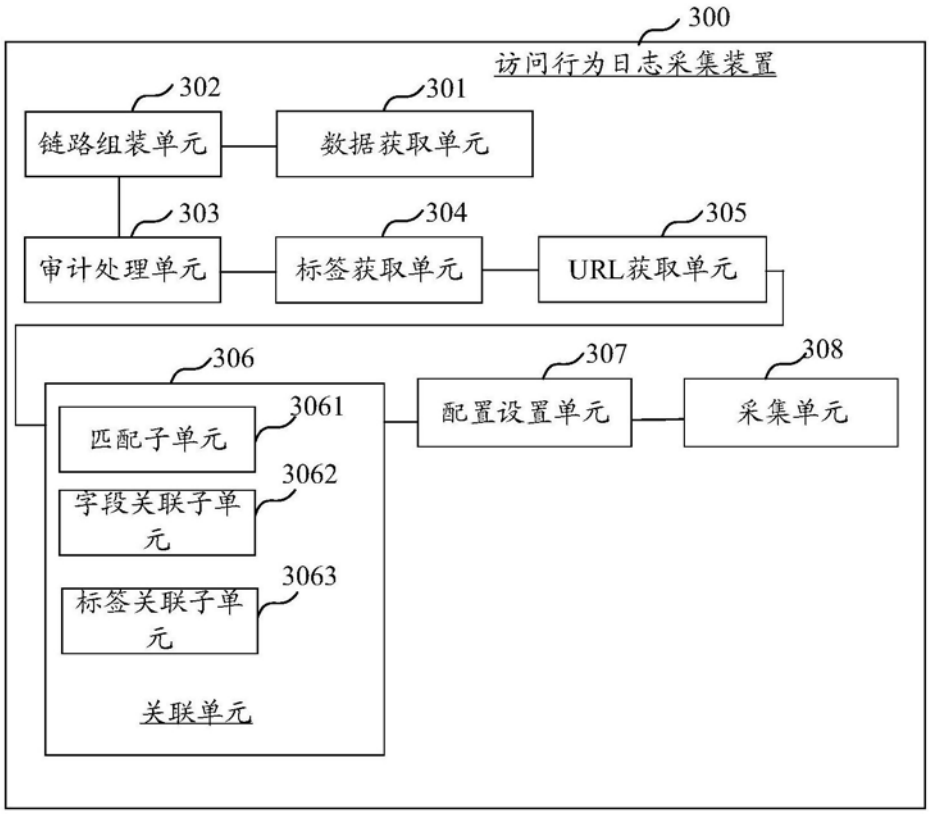

300
访问行为日志采集装置
302
链路组装单元
301
数据获取单元
303
审计处理单元
304
标签获取单元
305
URL获取单元
306
匹配子单元
3061
字段关联子单元
3062
标签关联子单元
3063
关联单元
307
配置设置单元
308
采集单元

图8

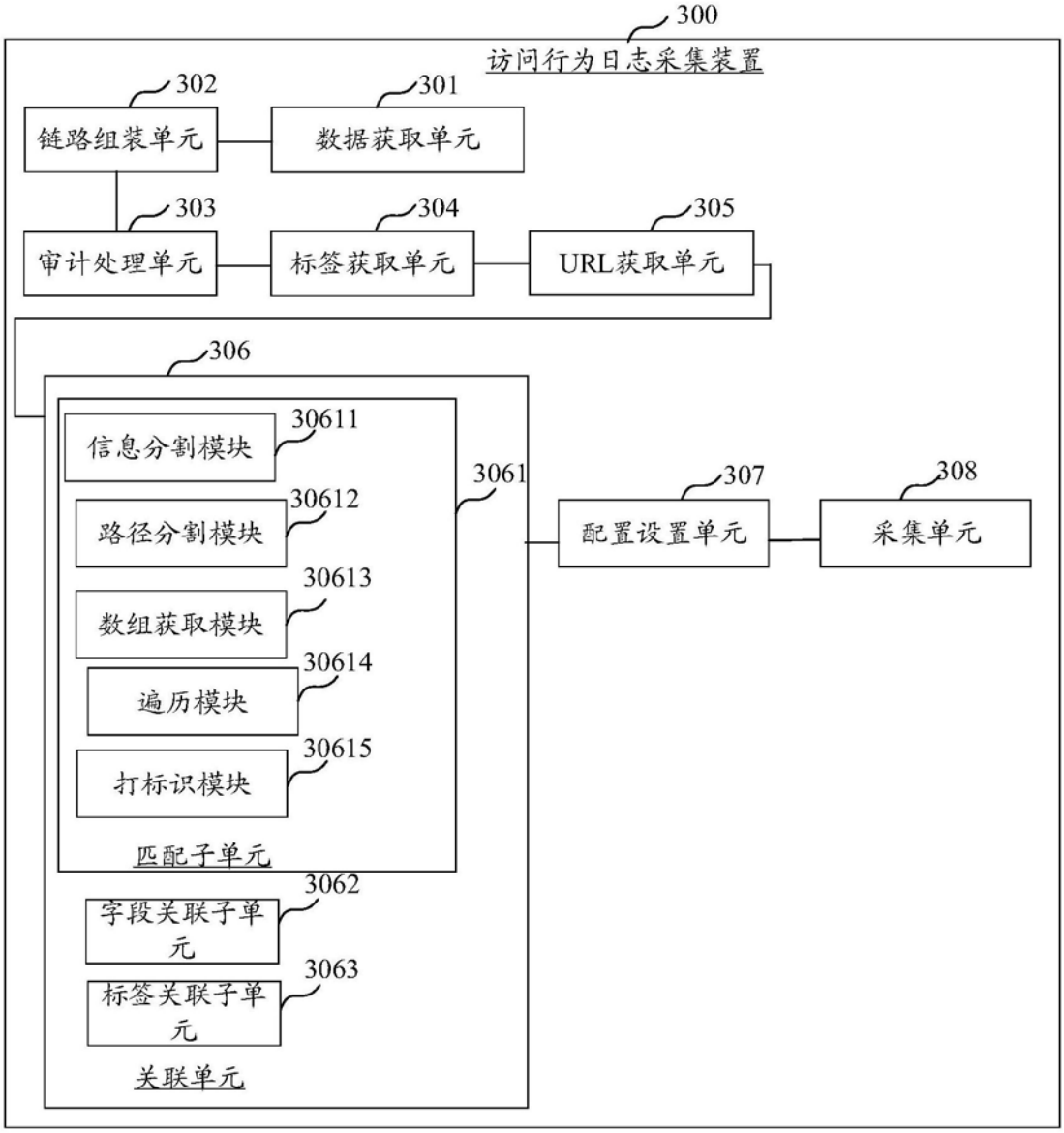

300
访问行为日志采集装置
302
链路组装单元
301
数据获取单元
303
审计处理单元
304
标签获取单元
305
URL获取单元
306
信息分割模块
30611
路径分割模块
30612
数组获取模块
30613
遍历模块
30614
打标识模块
30615
匹配子单元
3061
字段关联子单元
3062
标签关联子单元
3063
关联单元
307
配置设置单元
308
采集单元

图9

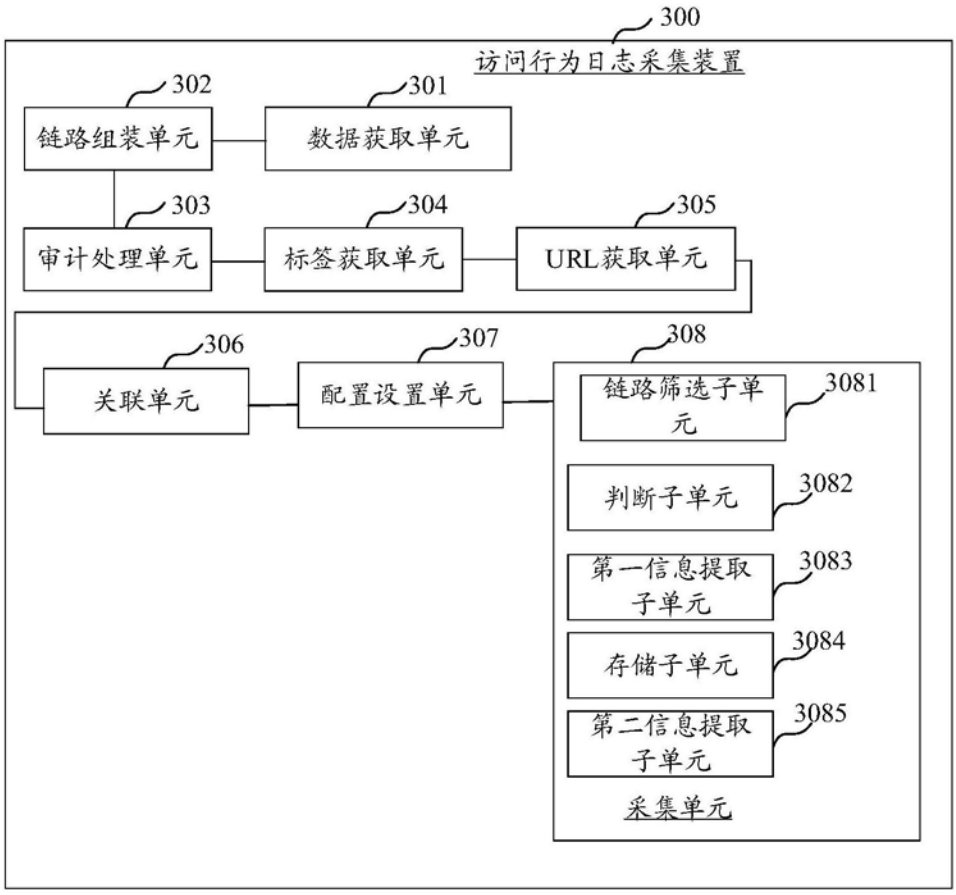

300
访问行为日志采集装置
302
链路组装单元
301
数据获取单元
303
审计处理单元
304
标签获取单元
305
URL获取单元
306
关联单元
307
配置设置单元
308
链路筛选子单元
3081
判断子单元
3082
第一信息提取子单元
3083
存储子单元
3084
第二信息提取子单元
3085
采集单元

图10

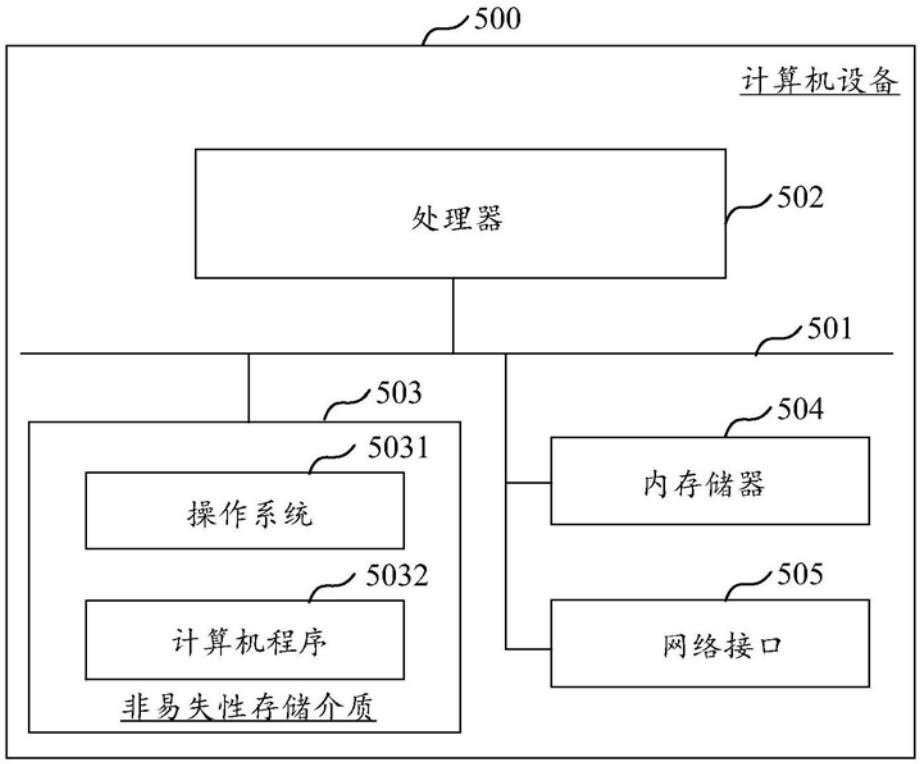

500
计算机设备
处理器
502
501
503
504
5031
操作系统
内存储器
5032
505
计算机程序
网络接口
非易失性存储介质

图11