(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117609992 A
(43)申请公布日 2024.02.27

(21)申请号 202311600724.5

(22)申请日 2023.11.27

(71)申请人 南方电网数字电网集团信息通信科技有限公司
地址 510000 广东省广州市黄埔区光谱中路11号2栋3单元12层全层

(72)发明人 张佳发 莫嘉永 邹洪 曾子峰 许伟杰 金浩 江家伟 陈锋

(74)专利代理机构 北京品源专利代理有限公司 11332
专利代理师 严慧

(51)Int.Cl.
G06F 21/55(2013.01)
G06F 21/57(2013.01)
H04L 9/40(2022.01)

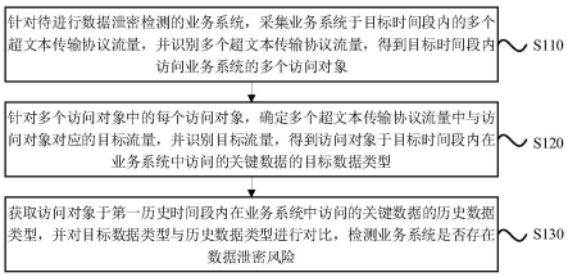
权利要求书2页 说明书12页 附图5页

(54)发明名称

一种数据泄密检测方法、装置及存储介质

(57)摘要

本发明实施例公开了一种数据泄密检测方法、装置及存储介质。该方法可包括:针对待进行数据泄密检测的业务系统,采集并识别业务系统于目标时间段内的多个超文本传输协议流量,得到目标时间段内访问业务系统的多个访问对象;针对多个访问对象中的每个访问对象,确定多个超文本传输协议流量中与访问对象对应的目标流量,并识别目标流量,得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型;获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型,并对目标数据类型与历史数据类型进行对比,检测业务系统是否存在数据泄密风险。本发明的技术方案,可以准确检测业务系统是否存在数据泄密风险。



CN 117609992 A

1.一种数据泄密检测方法,其特征在于,包括:

针对待进行数据泄密检测的业务系统,采集所述业务系统于目标时间段内的多个超文本传输协议流量,并识别所述多个超文本传输协议流量,得到所述目标时间段内访问所述业务系统的多个访问对象;

针对所述多个访问对象中的每个访问对象,确定所述多个超文本传输协议流量中与所述访问对象对应的目标流量,并识别所述目标流量,得到所述访问对象于所述目标时间段内在所述业务系统中访问的关键数据的目标数据类型;

获取所述访问对象于第一历史时间段内在所述业务系统中访问的关键数据的历史数据类型,并对所述目标数据类型与所述历史数据类型进行对比,检测所述业务系统是否存在数据泄密风险。

2.根据权利要求1所述的方法,其特征在于,所述识别所述多个超文本传输协议流量,得到所述目标时间段内访问所述业务系统的多个访问对象,包括:

针对所述多个超文本传输协议流量中的每个超文本传输协议流量,对所述超文本传输协议流量进行识别,得到所述超文本传输协议流量对应的访问对象登录至所述业务系统的登录接口;

基于所述登录接口,得到所述访问对象访问所述业务系统时的会话凭证,并基于所述会话凭证识别出所述访问对象;

根据所述多个超文本传输协议流量分别对应的访问对象,得到于所述目标时间段内访问所述业务系统的多个访问对象。

3.根据权利要求2所述的方法,其特征在于,所述对所述超文本传输协议流量进行识别,得到所述超文本传输协议流量对应的访问对象登录至所述业务系统的登录接口,包括:

对所述超文本传输协议流量进行识别,得到所述超文本传输协议流量对应的访问对象针对所述业务系统的目标登录特征;

获取预先构建出的登录特征库,并从所述登录特征库内存储的各候选登录特征中确定与所述目标登录特征匹配的匹配登录特征,将所述匹配登录特征所对应的登录接口,作为所述访问对象登录至所述业务系统的登录接口。

4.根据权利要求2所述的方法,其特征在于,所述确定所述多个超文本传输协议流量中与所述访问对象对应的目标流量,包括:

获取所述访问对象所对应的会话凭证,并从所述多个超文本传输协议流量中确定与所述会话凭证对应的凭证流量;

将所述凭证流量,作为所述访问对象对应的目标流量。

5.根据权利要求1所述的方法,其特征在于,还包括:

获取所述业务系统在第二历史时间段内含有的关键数据的关键数据类型;

所述对所述目标数据类型与所述历史数据类型进行对比,检测所述业务系统是否存在数据泄密风险,包括:

分别对所述目标数据类型与所述历史数据类型,以及所述目标数据类型与所述关键数据类型进行对比,检测所述业务系统是否存在数据泄密风险。

6.根据权利要求5所述的方法,其特征在于,还包括:

获取所述访问对象于所述目标时间段内在所述业务系统中访问的关键数据的目标数

据流向,以及获取所述业务系统在所述第二历史时间段内含有的关键数据的关键数据流向;

所述分别对所述目标数据类型与所述历史数据类型,以及所述目标数据类型与所述关键数据类型进行对比,检测所述业务系统是否存在数据泄密风险,包括:

分别对所述目标数据类型与所述历史数据类型,所述目标数据类型与所述关键数据类型,以及所述目标数据流向与所述关键数据流向进行对比,并根据得到的对比结果,检测所述业务系统是否存在数据泄密风险。

7.根据权利要求1所述的方法,其特征在于,还包括:

获取预先构建出的关键数据识别库,其中,所述关键数据识别库中存储有预设的各级别和/或各类型下的关键数据;

所述识别所述目标流量,得到所述访问对象于所述目标时间段内在所述业务系统中访问的关键数据的目标数据类型,包括:

对所述目标流量进行识别,得到所述访问对象于所述目标时间段内在所述业务系统中访问的全部数据;

基于所述关键数据识别库,从所述全部数据中识别出关键数据。

8.根据权利要求1所述的方法,其特征在于,在所述检测所述业务系统是否存在数据泄密风险之后,还包括:

在根据得到的检测结果,确定所述业务系统存在所述数据泄密风险的情况下,进行报警提示,和/或,对导致所述数据泄密风险的关键数据进行溯源路径跟踪。

9.一种数据泄密检测装置,其特征在于,包括:

访问对象得到模块,用于针对待进行数据泄密检测的业务系统,采集所述业务系统于目标时间段内的多个超文本传输协议流量,并识别所述多个超文本传输协议流量,得到所述目标时间段内访问所述业务系统的多个访问对象;

目标数据类型得到模块,用于针对所述多个访问对象中的每个访问对象,确定所述多个超文本传输协议流量中与所述访问对象对应的目标流量,并识别所述目标流量,得到所述访问对象于所述目标时间段内在所述业务系统中访问的关键数据的目标数据类型;

数据泄密风险检测模块,用于获取所述访问对象于第一历史时间段内在所述业务系统中访问的关键数据的历史数据类型,并对所述目标数据类型与所述历史数据类型进行对比,检测所述业务系统是否存在数据泄密风险。

10.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机指令,所述计算机指令用于使处理器执行时实现如权利要求1-8中任一所述的数据泄密检测方法。

一种数据泄密检测方法、装置及存储介质

技术领域

[0001] 本发明实施例涉及数据处理技术领域，尤其涉及一种数据泄密检测方法、装置及存储介质。

背景技术

[0002] 随着数据资产日益丰富，对于企业而言，准确检测内部使用的业务系统是否存在数据泄密风险，以提高企业网络环境的整体安全性，至关重要。

[0003] 目前，主要是基于内容识别进行数据泄密检测。但是，这一实现方案无法有效区分正常的数据访问和异常的数据泄密，这会导致大量误报，亟待改进。

发明内容

[0004] 本发明实施例提供了一种数据泄密检测方法、装置及存储介质，可以准确检测业务系统是否存在数据泄密风险。

[0005] 根据本发明的一方面，提供了一种数据泄密检测方法，包括：

[0006] 针对待进行数据泄密检测的业务系统，采集业务系统于目标时间段内的多个超文本传输协议流量，并识别多个超文本传输协议流量，得到目标时间段内访问业务系统的多个访问对象；

[0007] 针对多个访问对象中的每个访问对象，确定多个超文本传输协议流量中与访问对象对应的目标流量，并识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型；

[0008] 获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，并对目标数据类型与历史数据类型进行对比，检测业务系统是否存在数据泄密风险。

[0009] 根据本发明的另一方面，提供了一种数据泄密检测装置，包括：

[0010] 访问对象得到模块，用于针对待进行数据泄密检测的业务系统，采集业务系统于目标时间段内的多个超文本传输协议流量，并识别多个超文本传输协议流量，得到目标时间段内访问业务系统的多个访问对象；

[0011] 目标数据类型得到模块，用于针对多个访问对象中的每个访问对象，确定多个超文本传输协议流量中与访问对象对应的目标流量，并识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型；

[0012] 数据泄密风险检测模块，用于获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，并对目标数据类型与历史数据类型进行对比，检测业务系统是否存在数据泄密风险。

[0013] 根据本发明的另一方面，提供了一种计算机可读存储介质，其上存储有计算机指令，该计算机指令用于使处理器执行时实现本发明任意实施例所提供的任一的数据泄密检测方法。

[0014] 本发明实施例的技术方案，针对待进行数据泄密检测的业务系统，采集业务系统

于目标时间段内的多个超文本传输协议流量,并识别多个超文本传输协议流量,得到目标时间段内访问业务系统的多个访问对象,全面识别数据的访问对象,便于精确追踪泄密对象;针对多个访问对象中的每个访问对象,确定多个超文本传输协议流量中与访问对象对应的目标流量,并识别目标流量,得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型,精准识别每个访问对象的访问内容;获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型,并对目标数据类型与历史数据类型进行对比,检测业务系统是否存在数据泄密风险,与历史数据类型进行对比,快速识别访问数据类型异常的访问对象,快速检测发生数据泄密访问对象。上述技术方案,通过超文本传输协议流量准确识别出访问对象,进而通过对比访问对象当前的目标数据类型以及日常的历史数据类型,检测访问对象是否对业务系统进行了数据泄密,相较于单纯的内容识别,这有助于有效区分正常的数据访问和异常的数据泄密,进而可准确检测业务系统是否存在数据泄密风险,避免误报。

[0015] 应当理解,本部分所描述的内容并非旨在标识本发明的实施例的关键或是重要特征,也不用于限制本发明的范围。本发明的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0016] 为了更清楚地说明本发明实施例中的技术方案,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,下面描述中的附图仅仅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。

[0017] 图1是根据本发明实施例提供的一种数据泄密检测方法的流程图;

[0018] 图2是根据本发明实施例提供的另一种数据泄密检测方法的流程图;

[0019] 图3是根据本发明实施例提供的又一种数据泄密检测方法的流程图;

[0020] 图4是根据本发明实施例提供的又一种数据泄密检测方法中的可选示例的流程图;

[0021] 图5是根据本发明实施例提供的一种数据泄密检测装置的结构框图;

[0022] 图6是实现本发明实施例的数据泄密检测方法的电子设备的结构示意图。

具体实施方式

[0023] 为了使本技术领域的人员更好地理解本发明方案,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,所描述的实施例仅仅是本发明一部分的实施例,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,都应当属于本发明保护的范围。

[0024] 需要说明的是,本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。“目标”、“原始”等的情况类似,在此不再赘述。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤

或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0025] 需要说明的是,本发明的技术方案中,所涉及到的用户个人信息的采集、收集、更新、分析、处理、使用、传输、存储等方面,均符合相关法律法规的规定,被用于合法的用途,且不违背公序良俗。对用户个人信息采取必要措施,从而防止对用户个人信息数据的非法访问,维护用户个人信息安全、网络安全和国家安全。

[0026] 图1是本发明实施例所提供的一种数据泄密检测方法的流程图。本实施例可适用于数据泄密检测的情况,尤其可以适用于访问对象维度的数据泄密检测的情况。该方法可以由本发明实施例提供的数据泄密检测装置来执行,该装置可以由软件和/或硬件的方式实现,该装置可以集成在电子设备上,该电子设备可以是各种用户终端或服务器,该电子设备可以是运行下述的业务系统的设备。

[0027] 参见图1,本发明实施例的方法具体包括如下步骤:

[0028] S110、针对待进行数据泄密检测的业务系统,采集业务系统于目标时间段内的多个超文本传输协议流量,并识别多个超文本传输协议流量,得到目标时间段内访问业务系统的多个访问对象。

[0029] 其中,业务系统可理解成为了办理相关业务所构建的系统。

[0030] 超文本传输协议(HyperText Transfer Protocol,HTTP)流量可理解为在网络上进行信息交换的流量,例如对信息进行上传、浏览和下载等所产生的流量。

[0031] 目标时间段可理解为要进行数据泄密检测的全部或部分时间段,可根据实际情况设置,在此不做具体限制。

[0032] 访问对象可理解为在目标时间段内对业务系统进行数据访问的对象。

[0033] 采集业务系统在目标时间段内的多个超文本传输协议流量,并通过对多个超文本传输协议流量进行识别,得到在目标时间段内针对业务系统进行访问的所有访问对象,防止遗漏。需要强调的是,这些访问对象已授权执行本方法的电子设备采集自身在业务系统的超文本传输协议流量。

[0034] S120、针对多个访问对象中的每个访问对象,确定多个超文本传输协议流量中与访问对象对应的目标流量,并识别目标流量,得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型。

[0035] 其中,针对多个访问对象中的每个访问对象,该访问对象对应的目标流量可理解为该访问对象在业务系统中访问所产生的超文本传输协议流量。

[0036] 关键数据可理解为业务系统中不可泄密的数据,例如敏感数据、机密数据或是核心数据等。需要强调的是,这些访问对象已授权执行本方法的电子设备识别自身在业务系统中访问的关键数据。

[0037] 目标数据类型可理解为关键数据的类型,例如敏感数据、机密数据或是核心数据等,这里以机密数据为例,进一步例如可以是技术机密数据、战略机密数据或是架构机密数据等。

[0038] 针对多个访问对象中的每个访问对象,识别访问对象在多个超文本传输协议流量中对应的目标流量,得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型,将每个访问对象与其访问的关键数据的目标数据类型相对应,方便找到信息泄露

的访问对象。

[0039] S130、获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，并对目标数据类型与历史数据类型进行对比，检测业务系统是否存在数据泄密风险。

[0040] 其中，第一历史时间段可理解为访问对象在过去历史时间段内在业务系统上访问关键数据的时间段，例如可以是过去1个月、过去2个月或者过去半年等，可根据实际情况进行设置，在此不做具体限制。

[0041] 历史数据类型可理解为访问对象在第一历史时间段内访问的关键数据的全部数据类型。

[0042] 每个访问对象在业务系统中获取关键数据的数据类型是有限的，针对每个访问对象，通过历史数据类型可以知道访问对象经常访问或者可以访问的关键数据的数据类型，通过对访问对象的目标数据类型与历史数据类型进行对比，就可以知道访问对象是否访问其他关键数据的数据类型，进一步可以检测业务系统是否存在数据泄密风险。

[0043] 本发明实施例的技术方案，针对待进行数据泄密检测的业务系统，采集业务系统于目标时间段内的多个超文本传输协议流量，并识别多个超文本传输协议流量，得到目标时间段内访问业务系统的多个访问对象，基于超文本传输协议流量，实现了访问对象的精准识别，这与后续步骤结合，便于确定后续的关键数据的访问是正常的数据访问或是异常的数据泄密；针对多个访问对象中的每个访问对象，确定多个超文本传输协议流量中与访问对象对应的目标流量，并识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型，精准识别每个访问对象的访问内容；获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，并对目标数据类型与历史数据类型进行对比，检测业务系统是否存在数据泄密风险，目标数据类型与历史数据类型的对比，准确识别关键数据访问异常的访问对象。上述技术方案，通过超文本传输协议流量准确识别出访问对象，进而通过对比访问对象当前的目标数据类型以及日常的历史数据类型，检测访问对象是否对业务系统进行了数据泄密，相较于单纯的内容识别，这有助于有效区分正常的数据访问和异常的数据泄密，进而可准确检测业务系统是否存在数据泄密风险，避免误报。

[0044] 一种可选的技术方案，在针对待进行数据泄密检测的业务系统，采集业务系统于目标时间段内的多个超文本传输协议流量，并识别多个超文本传输协议流量，得到目标时间段内访问业务系统的多个访问对象之前，上述数据泄密检测方法，还包括：获取预先构建出的关键数据识别库，其中，关键数据识别库中存储有预设的各级别和/或各类型下的关键数据；识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型，包括：对目标流量进行识别，得到访问对象于目标时间段内在业务系统中访问的全部数据；基于关键数据识别库，从全部数据中识别出关键数据。

[0045] 其中，关键数据识别库可理解为集合了业务系统中关键数据信息的数据库，尤其可理解为存储有预设的各级别和/或各类型下的关键数据的数据库，可选的，其中的各级别例如可包括极关键级、关键级、较关键级和低关键级等，各类型例如可包括敏感数据、核心数据或是机密数据等。

[0046] 针对每个访问对象中的任一访问对象，对其对应的目标流量进行识别，得出该访

问对象于目标时间段内在业务系统中访问的全部数据,将该全部数据与关键数据识别库中的数据进行对比,即可识别出该访问对象访问的全部数据中的关键数据。

[0047] 通过以上步骤,可以准确得到每个访问对象对应访问的关键数据,使数据泄密检测可以识别的具体泄密对象。

[0048] 另一种可选的技术方案,在检测业务系统是否存在数据泄密风险之后,还包括:在根据得到的检测结果,确定业务系统存在数据泄密风险的情况下,进行报警提示,和/或,对导致数据泄密风险的关键数据进行溯源路径跟踪。

[0049] 其中,溯源路径跟踪可以理解为通过数据泄密检测,追溯和追踪数据泄密的路径和来源。

[0050] 在检测到存在数据泄密风险之后,进行报警提示和/或对导致数据泄密风险的关键数据进行溯源路径跟踪,可以有效防止泄密范围进一步扩大,减少损失。

[0051] 图2是本发明实施例中提供的另一种数据泄密检测方法的流程图。本实施例以上述各技术方案为基础进行优化。在本实施例中,可选的,识别多个超文本传输协议流量,得到目标时间段内访问业务系统的多个访问对象,具体可包括:针对多个超文本传输协议流量中的每个超文本传输协议流量,对超文本传输协议流量进行识别,得到超文本传输协议流量对应的访问对象登录至业务系统的登录接口;基于登录接口,得到访问对象访问业务系统时的会话凭证,并基于会话凭证识别出访问对象;根据多个超文本传输协议流量分别对应的访问对象,得到于目标时间段内访问业务系统的多个访问对象。其中,与上述各实施例相同或相应的术语的解释在此不再赘述。

[0052] 参见图2,本实施例的方法具体可以包括如下步骤:

[0053] S210、针对待进行数据泄密检测的业务系统,采集业务系统于目标时间段内的多个超文本传输协议流量。

[0054] S220、针对多个超文本传输协议流量中的每个超文本传输协议流量,对超文本传输协议流量进行识别,得到超文本传输协议流量对应的访问对象登录至业务系统的登录接口。

[0055] 其中,登录接口可以理解为对业务系统进行登录访问的入口,例如网页、应用程序(Application,APP)或小程序等。

[0056] S230、基于登录接口,得到访问对象访问业务系统时的会话凭证,并基于会话凭证识别出访问对象。

[0057] 其中,会话凭证可理解为用于临时授权访问对象进行数据访问的凭证。

[0058] 访问对象通过登录接口登录业务系统后,业务系统可将会话凭证返回至登录接口,访问对象基于会话凭证进行后续的一系列操作,同时也可基于会话凭证识别出访问对象。

[0059] S240、根据多个超文本传输协议流量分别对应的访问对象,得到于目标时间段内访问业务系统的多个访问对象。

[0060] S250、针对多个访问对象中的每个访问对象,确定多个超文本传输协议流量中与访问对象对应的目标流量,并识别目标流量,得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型。

[0061] S260、获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数

据类型,并对目标数据类型与历史数据类型进行对比,检测业务系统是否存在数据泄密风险。

[0062] 本发明实施例的技术方案,通过每个访问对象对应的超文本传输协议流量,得到每个访问对象对应的登录接口,基于每个访问对象对应的登录接口,确定每个访问对象对应的访问凭证,通过访问凭证即可识别出对应的访问对象。实现了对每个超文本传输协议流量对应的访问对象的精准识别。

[0063] 一种可选的技术方案,对超文本传输协议流量进行识别,得到超文本传输协议流量对应的访问对象登录至业务系统的登录接口,包括:对超文本传输协议流量进行识别,得到超文本传输协议流量对应的访问对象针对业务系统的目标登录特征;获取预先构建出的登录特征库,并从登录特征库内存储的各候选登录特征中确定与目标登录特征匹配的匹配登录特征,将匹配登录特征所对应的登录接口,作为访问对象登录至业务系统的登录接口。

[0064] 其中,目标登录特征可理解为访问对象登录业务系统时所涉及到的特征,例如可以是统一资源定位符(Uniform Resource Locator,URL)特征和对象名特征等。

[0065] 登录特征库可理解为存储有各候选登录特征的数据库,该各候选登录特征分别对应有各自的登录接口。

[0066] 对超文本传输协议流量进行识别,得到超文本传输协议流量对应的访问对象针对业务系统的目标登录特征,获取预先构建出的登录特征库,并从登录特征库内存储的各候选登录特征中确定与目标登录特征匹配的匹配登录特征,将匹配登录特征所对应的登录接口,作为访问对象登录至业务系统的登录接口,例如可以是对超文本传输协议流量进行识别解析,提取出超文本传输协议流量中的URL特征和对象名特征,并将这两个特征传入到预先构建的登录特征库中进行匹配,从而得到该超文本传输协议流量对应的登录接口。

[0067] 通过特征匹配可以准确得到每个访问对象对应的登录接口。

[0068] 另一种可选的技术方案,确定多个超文本传输协议流量中与访问对象对应的目标流量,包括:获取访问对象所对应的会话凭证,并从多个超文本传输协议流量中确定与会话凭证对应的凭证流量;将凭证流量,作为访问对象对应的目标流量。

[0069] 其中,凭证流量可以理解为访问对象通过会话凭证进行数据访问所产生的流量。

[0070] 从多个超文本传输协议流量中确定出与会话凭证对应的凭证流量,该凭证流量即为该会话凭证所表征的访问对象对应的目标流量,由此保证了目标流量确定的准确率和效率。

[0071] 图3是本发明实施例中提供的再一种数据泄密检测方法的流程图。本实施例以上述各技术方案为基础进行优化。在本实施例中,可选的,数据泄密检测方法还包括获取业务系统在第二历史时间段内含有的关键数据的关键数据类型,对目标数据类型与历史数据类型进行对比,检测业务系统是否存在数据泄密风险具体可包括:分别对目标数据类型与历史数据类型,以及目标数据类型与关键数据类型进行对比,检测业务系统是否存在数据泄密风险。其中,与上述各实施例相同或相应的术语的解释在此不再赘述。

[0072] 参见图3,本实施例的方法具体可以包括如下步骤:

[0073] S310、针对待进行数据泄密检测的业务系统,采集业务系统于目标时间段内的多个超文本传输协议流量,并识别多个超文本传输协议流量,得到目标时间段内访问业务系统的多个访问对象。

[0074] S320、针对多个访问对象中的每个访问对象，确定多个超文本传输协议流量中与访问对象对应的目标流量，并识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型。

[0075] S330、获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，获取业务系统在第二历史时间段内含有的关键数据的关键数据类型。

[0076] 其中，第二历史时间段可理解为业务系统在过去历史时间段内在业务系统上访问关键数据的时间段，例如可以是过去1个月、过去2个月或者过去半年等。第一历史时间段与第二历史时间段可以是相同或是不同的时间段，可根据实际情况进行设置，在此不做具体限制。

[0077] 关键数据类型可理解为关键数据的数据类型，例如可以是敏感数据、机密数据或是核心数据等，这里以机密数据为例，进一步例如可以是技术机密数据、战略机密数据或是架构机密数据。

[0078] S340、分别对目标数据类型与历史数据类型，以及目标数据类型与关键数据类型进行对比，检测业务系统是否存在数据泄密风险。

[0079] 其中，若访问对象访问的目标数据类型与该访问对象在第一历史时间段中所访问的历史数据类型不同，或是超出业务系统在第二历史时间段中关键数据类型的范围，则表明存在数据泄密风险。

[0080] 本发明实施例的技术方案，通过将访问对象访问的目标数据类型与该访问对象在第一历史时间段中所访问的历史数据类型进行对比判断该访问对象是否有泄密行为，通过判断访问对象访问的目标数据类型是否在业务系统第二历史时间段中关键数据类型的范围，判断该访问对象是否访问超限。进一步提高了数据泄密检测的准确性。

[0081] 在此基础上，一种可选的技术方案，上述数据泄密检测方法，还包括：

[0082] 获取访问对象于目标时间段内在业务系统中访问的关键数据的目标数据流向，以及获取业务系统在第二历史时间段内含有的关键数据的关键数据流向；

[0083] 分别对目标数据类型与历史数据类型，以及目标数据类型与关键数据类型进行对比，检测业务系统是否存在数据泄密风险，包括：分别对目标数据类型与历史数据类型，目标数据类型与关键数据类型，以及目标数据流向与关键数据流向进行对比，并根据得到的对比结果，检测业务系统是否存在数据泄密风险。

[0084] 其中，目标数据流向可理解为目标数据从一个源头传送或复制到一个目标地点的过程。

[0085] 通过对访问对象的目标数据流向与业务系统在第二历史时间段内的关键数据流向进行对比，即可判断出目标对象的目标数据流向是否异常，进一步提高数据泄密检测的准确性。

[0086] 为了更好地理解上述的各个技术方案，下面结合具体示例进行示例性说明。如图4所示，示例性的，具体步骤如下：

[0087] 步骤1：收集业务系统中的超文本传输协议流量。

[0088] 步骤2：对超文本传输协议流量进行解析，提取出URL特征和对象名特征。然后，通过预先构建的特征库与提取出URL特征和对象名特征进行匹配，识别超文本传输协议流量对应的登录接口，并基于登录接口获取会话凭证，再基于会话凭证从收集来的全部超文本

传输协议流量中找到与该会话凭证所表征的访问对象对应的目标流量,以从该目标流量中解析出该访问对象访问的关键数据,并确定关键数据的目标数据类型。

[0089] 步骤3:基于关键数据识别库,对超文本传输协议流量中的明文、编码后的数据、压缩文件、文档以及图片等各类可能存在的关键数据进行识别。具体的,超文本传输协议流量中的压缩文件、编码后的数据、文档以及图片等数据,可基于业务系统内置的文件解析模块和光学字符识别(Optical Character Recognition,OCR)模块进行解析得到。解析后的数据结合关键数据识别库,即可识别出超文本传输协议流量中包含的关键数据。

[0090] 步骤4:结合关键数据识别,建立业务系统中每个访问对象与关键数据的关联,以及业务系统与关键数据的关联。

[0091] 业务系统中访问对象与关键数据的关联:指的是每个访问对象在业务系统中获取的关键数据的关键数据类型是有限的,通过一段时间的统计后,可形成对应的关联表。

[0092] 业务系统与关键数据的关联:指的是业务系统中所包含的关键数据的关键数据类型和关键数据流向,在没有业务变更的情况是保持稳定的,因此基于一段时间的统计后,可形成业务系统涉及的关键数据类型以及关键数据流向(如跨境或是跨省等)。

[0093] 步骤5:根据策略,判断关键数据是否泄密。

[0094] 策略:策略指的是当访问对象的数据传输行为违背了访问对象历史的访问行为或业务系统所涉及的关键数据类型和关键数据流向,则产生告警。

[0095] 步骤6:对导致数据泄密风险的关键数据进行溯源路径跟踪;以及,优化各类策略和规则库,降低误报率。

[0096] 以上具体示例通过访问对象维度的精准识别与安全策略判定,极大地提升了业务系统中关键数据泄密风险的检测能力,使数据资产安全性达到一个全新的高度。

[0097] 图5为本发明实施例中提供的数据泄密检测装置的结构框图,该装置用于执行上述任意实施例所提供的数据泄密检测方法。该装置与上述各实施例的数据泄密检测方法属于同一个发明构思,在数据泄密检测装置的实施例中未详尽描述的细节内容,可参考上述数据泄密检测方法的实施例。参见图5,该装置具体可以包括:访问对象得到模块410、目标数据类型得到模块420以及数据泄密风险检测模块430。

[0098] 其中,访问对象得到模块410,用于针对待进行数据泄密检测的业务系统,采集业务系统于目标时间段内的多个超文本传输协议流量,并识别多个超文本传输协议流量,得到目标时间段内访问业务系统的多个访问对象;

[0099] 目标数据类型得到模块420,用于针对多个访问对象中的每个访问对象,确定多个超文本传输协议流量中与访问对象对应的目标流量,并识别目标流量,得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型;

[0100] 数据泄密风险检测模块430,用于获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型,并对目标数据类型与历史数据类型进行对比,检测业务系统是否存在数据泄密风险。

[0101] 可选的,访问对象得到模块还包括:

[0102] 登录接口得到子模块,用于针对多个超文本传输协议流量中的每个超文本传输协议流量,对超文本传输协议流量进行识别,得到超文本传输协议流量对应的访问对象登录至业务系统的登录接口;

[0103] 访问对象识别子模块,用于基于登录接口,得到访问对象访问业务系统时的会话凭证,并基于会话凭证识别出访问对象;

[0104] 访问对象得到子模块,用于根据多个超文本传输协议流量分别对应的访问对象,得到于目标时间段内访问业务系统的多个访问对象。

[0105] 在此基础上,一种可选的,登录接口得到子模块包括:

[0106] 目标登录特征得到单元,用于对超文本传输协议流量进行识别,得到超文本传输协议流量对应的访问对象针对业务系统的目标登录特征;

[0107] 登录接口确定单元,用于获取预先构建出的登录特征库,并从登录特征库内存储的各候选登录特征中确定与目标登录特征匹配的匹配登录特征,将匹配登录特征所对应的登录接口,作为访问对象登录至业务系统的登录接口。

[0108] 另一种可选的,目标数据类型得到模块还包括:

[0109] 凭证流量确定子模块,用于获取访问对象所对应的会话凭证,并从多个超文本传输协议流量中确定与会话凭证对应的凭证流量;

[0110] 目标流量确定子模块,用于将凭证流量,作为访问对象对应的目标流量。

[0111] 可选的,数据泄密检测装置还包括:

[0112] 关键数据类型获取模块,用于获取业务系统在第二历史时间段内含有的关键数据的关键数据类型;

[0113] 数据泄密风险检测模块还包括:

[0114] 数据泄密风险检测子模块,用于分别对目标数据类型与历史数据类型,以及目标数据类型与关键数据类型进行对比,检测业务系统是否存在数据泄密风险。

[0115] 在此基础上,可选的:

[0116] 数据流向获取模块,用于获取访问对象于目标时间段内在业务系统中访问的关键数据的目标数据流向,以及获取业务系统在第二历史时间段内含有的关键数据的关键数据流向;

[0117] 数据泄密风险检测子模块还包括:

[0118] 数据泄密风险检测单元,用于分别对目标数据类型与历史数据类型,目标数据类型与关键数据类型,以及目标数据流向与关键数据流向进行对比,并根据得到的对比结果,检测业务系统是否存在数据泄密风险。

[0119] 可选的,数据泄密检测装置还包括:

[0120] 关键数据识别库构建模块,用于获取预先构建出的关键数据识别库,其中,关键数据识别库中存储有预设的各级别和/或各类型下的关键数据;

[0121] 目标数据类型得到模块还包括:

[0122] 数据得到子模块,用于对目标流量进行识别,得到访问对象于目标时间段内在业务系统中访问的全部数据;

[0123] 关键数据识别子模块,用于基于关键数据识别库,从全部数据中识别出关键数据。

[0124] 可选的,数据泄密检测装置还包括:

[0125] 溯源路径跟踪模块,用于在根据得到的检测结果,确定业务系统存在数据泄密风险的情况下,进行报警提示,和/或,对导致数据泄密风险的关键数据进行溯源路径跟踪。

[0126] 本发明实施例所提供的数据泄密检测装置,通过访问对象得到模块,针对待进行

数据泄密检测的业务系统,采集业务系统于目标时间段内的多个超文本传输协议流量,并识别多个超文本传输协议流量,得到目标时间段内访问业务系统的多个访问对象,全面识别数据的访问对象,便于精确追踪泄密对象;通过目标数据类型得到模块,针对多个访问对象中的每个访问对象,确定多个超文本传输协议流量中与访问对象对应的目标流量,并识别目标流量,得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型,精准识别每个访问对象的访问内容;通过数据泄密风险检测模块,获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型,并对目标数据类型与历史数据类型进行对比,检测业务系统是否存在数据泄密风险,与历史数据类型进行对比,快速识别访问数据类型异常的访问对象,快速检测发生数据泄密访问对象。本发明实施例所提供的数据泄密检测装置可以有效区分正常的数据访问和异常的数据泄密,进而可准确检测业务系统是否存在数据泄密风险,避免误报。

[0127] 值得注意的是,上述数据泄密检测装置的实施例中,所包括的各个单元和模块只是按照功能逻辑进行划分的,但并不局限于上述的划分,只要能够实现相应的功能即可;另外,各功能单元的具体名称也只是为了便于相互区分,并不用于限制本发明的保护范围。

[0128] 图6示出了可以用来实施本发明的实施例的电子设备10的结构示意图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备(如头盔、眼镜、手表等)和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本发明的实现。

[0129] 如图6所示,电子设备10包括至少一个处理器11,以及与至少一个处理器11通信连接的存储器,如只读存储器(ROM)12、随机访问存储器(RAM)13等,其中,存储器存储有可被至少一个处理器执行的计算机程序,处理器11可以根据存储在只读存储器(ROM)12中的计算机程序或从存储单元18加载到随机访问存储器(RAM)13中的计算机程序,来执行各种适当的动作和处理。在RAM13中,还可存储电子设备10操作所需的各种程序和数据。处理器11、ROM12以及RAM13通过总线14彼此相连。输入/输出(I/O)接口15也连接至总线14。

[0130] 电子设备10中的多个部件连接至I/O接口15,包括:输入单元16,例如键盘、鼠标等;输出单元17,例如各种类型的显示器、扬声器等;存储单元18,如磁盘、光盘等;以及通信单元19,例如网卡、调制解调器、无线通信收发机等。通信单元19允许电子设备10通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0131] 处理器11可以是各种具有处理和计算能力的通用和/或专用处理组件。处理器11的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的处理器、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。处理器11执行上文所描述的各个方法和处理,例如数据泄密检测方法。

[0132] 在一些实施例中,数据泄密检测方法可被实现为计算机程序,其被有形地包含于计算机可读存储介质,例如存储单元18。在一些实施例中,计算机程序的部分或者全部可以经由ROM12和/或通信单元19而被载入和/或安装到电子设备10上。当计算机程序加载到RAM13并由处理器11执行时,可以执行上文描述的数据泄密检测方法的一个或多个步骤。备

选地,在其他实施例中,处理器11可通过其他任何适当的方式(例如,借助于固件)而被配置为执行数据泄密检测方法。

[0133] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、以及至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、以及该至少一个输出装置。

[0134] 用于实施本发明的方法的计算机程序可以采用一个或多个编程语言的任何组合来编写。这些计算机程序可以提供给通用计算机、专用计算机或是其他可编程数据处理装置的处理器,使得计算机程序当由处理器执行时使流程图和/或框图中所规定的功能/操作被实施。计算机程序可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行并且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0135] 在本发明的上下文中,计算机可读存储介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的计算机程序。计算机可读存储介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。备选地,计算机可读存储介质可以是机器可读信号介质。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0136] 为了提供与用户的交互,可以在电子设备上实施此处描述的系统和技术,该电子设备具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给电子设备。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0137] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)、区块链网络和互联网。

[0138] 计算系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器,又称为云计算服务器或云

主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务中,存在的管理难度大,业务扩展性弱的缺陷。

[0139] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发明中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本发明的技术方案所期望的结果,本文在此不进行限制。

[0140] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,根据设计要求和其他因素,可以进行各种修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

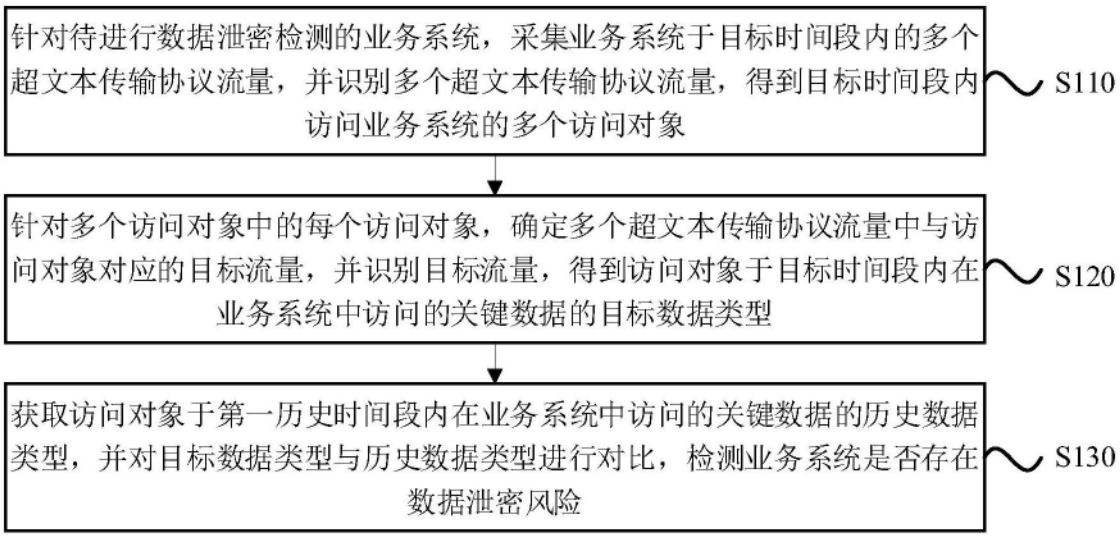

针对待进行数据泄密检测的业务系统，采集业务系统于目标时间段内的多个超文本传输协议流量，并识别多个超文本传输协议流量，得到目标时间段内访问业务系统的多个访问对象
S110
针对多个访问对象中的每个访问对象，确定多个超文本传输协议流量中与访问对象对应的目标流量，并识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型
S120
获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，并对目标数据类型与历史数据类型进行对比，检测业务系统是否存在数据泄密风险
S130

图1

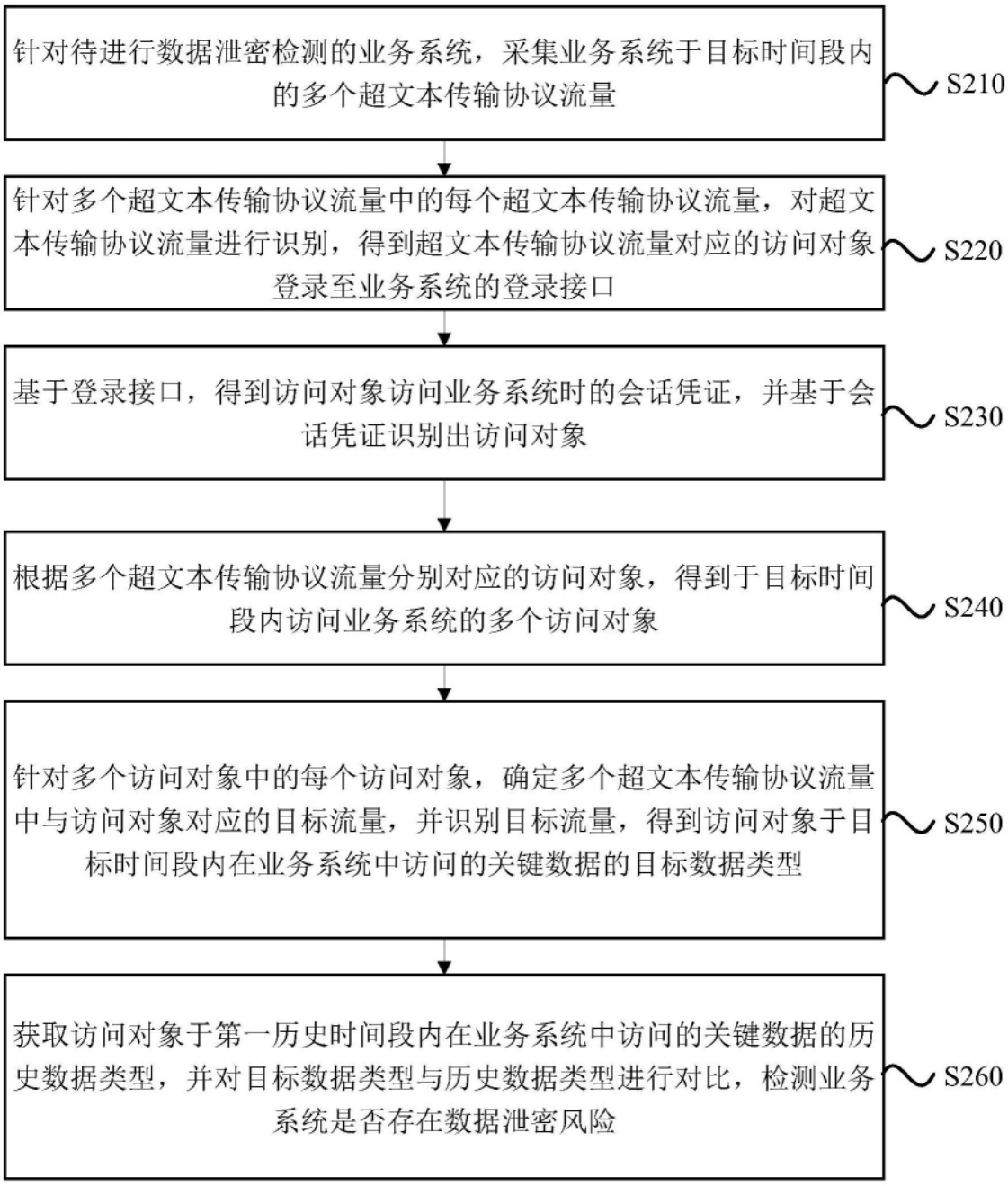

针对待进行数据泄密检测的业务系统，采集业务系统于目标时间段内的多个超文本传输协议流量
S210
针对多个超文本传输协议流量中的每个超文本传输协议流量，对超文本传输协议流量进行识别，得到超文本传输协议流量对应的访问对象登录至业务系统的登录接口
S220
基于登录接口，得到访问对象访问业务系统时的会话凭证，并基于会话凭证识别出访问对象
S230
根据多个超文本传输协议流量分别对应的访问对象，得到于目标时间段内访问业务系统的多个访问对象
S240
针对多个访问对象中的每个访问对象，确定多个超文本传输协议流量中与访问对象对应的目标流量，并识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型
S250
获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，并对目标数据类型与历史数据类型进行对比，检测业务系统是否存在数据泄密风险
S260

图2

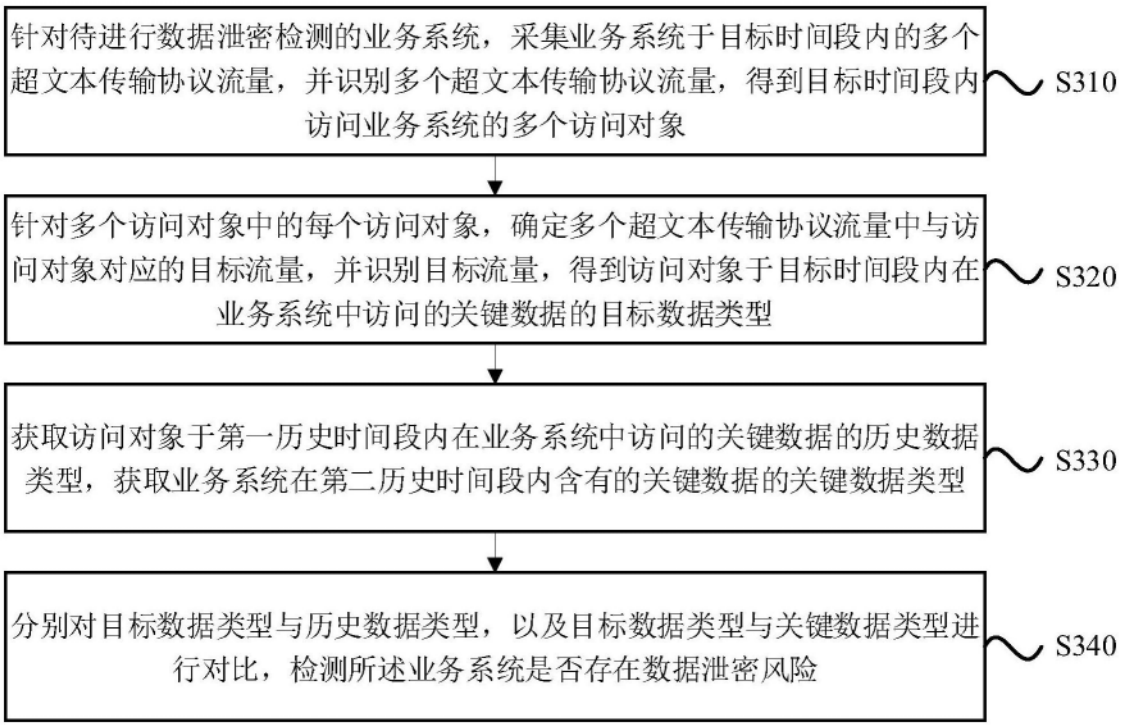

针对待进行数据泄密检测的业务系统，采集业务系统于目标时间段内的多个超文本传输协议流量，并识别多个超文本传输协议流量，得到目标时间段内访问业务系统的多个访问对象
S310
针对多个访问对象中的每个访问对象，确定多个超文本传输协议流量中与访问对象对应的目标流量，并识别目标流量，得到访问对象于目标时间段内在业务系统中访问的关键数据的目标数据类型
S320
获取访问对象于第一历史时间段内在业务系统中访问的关键数据的历史数据类型，获取业务系统在第二历史时间段内含有的关键数据的关键数据类型
S330
分别对目标数据类型与历史数据类型，以及目标数据类型与关键数据类型进行对比，检测所述业务系统是否存在数据泄密风险
S340

图3

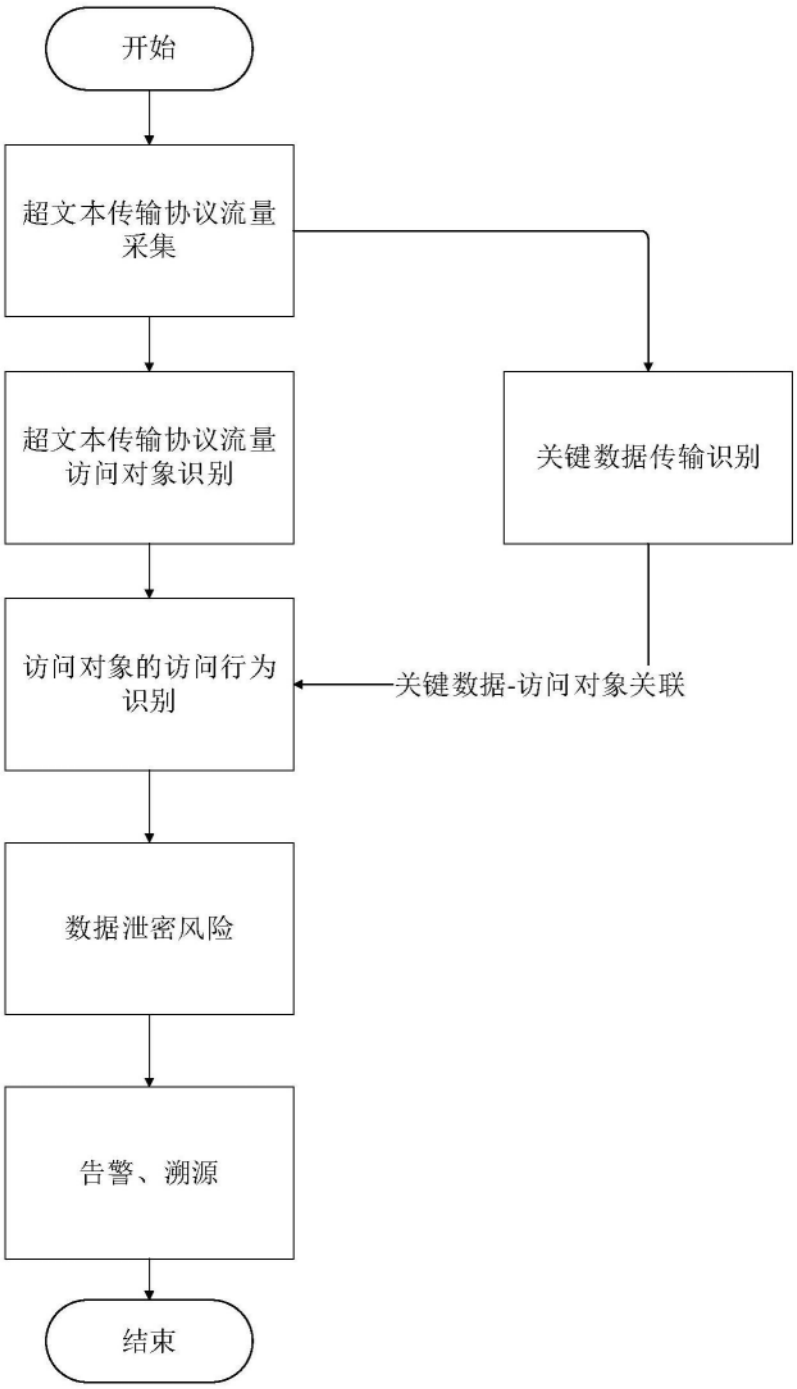

开始
超文本传输协议流量采集
超文本传输协议流量访问对象识别
关键数据传输识别
访问对象的访问行为识别
关键数据-访问对象关联
数据泄密风险
告警、溯源
结束

图4

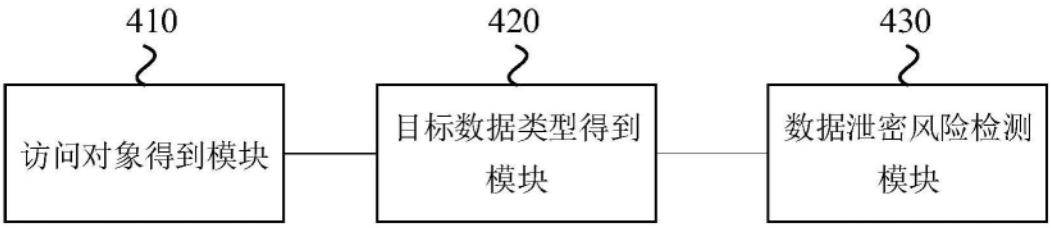

410
420
430
访问对象得到模块
目标数据类型得到模块
数据泄密风险检测模块

图5

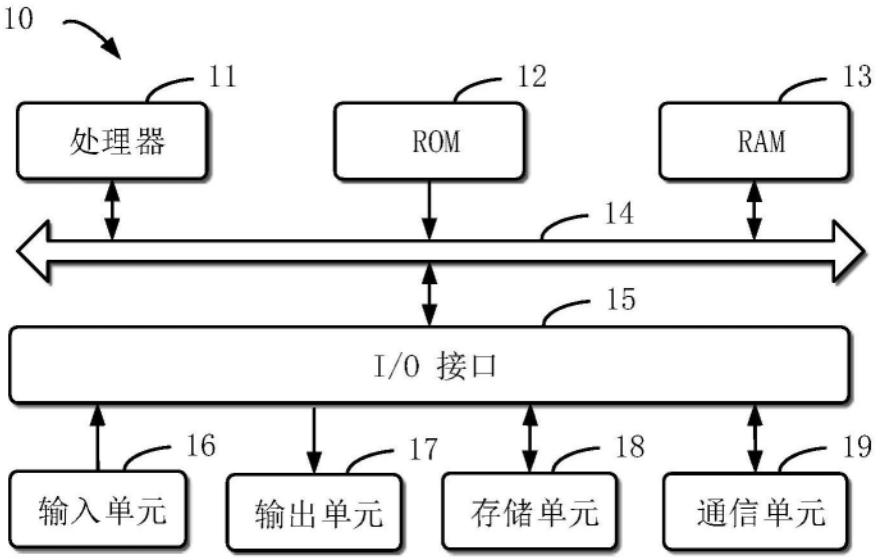

10
11
12
13
处理器
ROM
RAM
14
15
I/O 接口
16
17
18
19
输入单元
输出单元
存储单元
通信单元

图6