(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118337466 A
(43)申请公布日 2024.07.12

(21)申请号 202410510287.6

(22)申请日 2024.04.25

(71)申请人 江苏派智信息科技有限公司
地址 212000 江苏省镇江市123

(72)发明人 刘勇 王锦荣

(74)专利代理机构 北京康达联禾知识产权代理事务所(普通合伙) 11461
专利代理师 马娟

(51)Int.Cl.
H04L 9/40(2022.01)

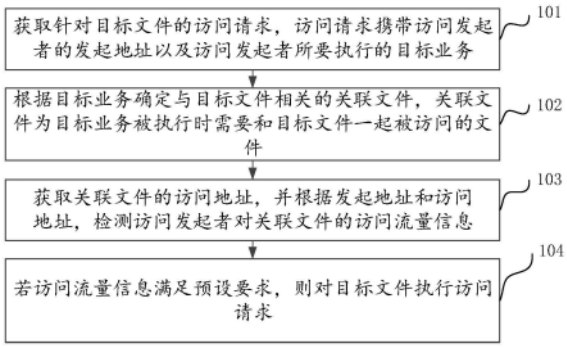
权利要求书3页 说明书10页 附图3页

(54)发明名称

一种基于大数据的信息安全保护方法及系统

(57)摘要

本发明公开了一种基于大数据的信息安全保护方法及系统,该方法包括:获取针对目标文件的访问请求,所述访问请求携带访问发起者的发起地址以及所述访问发起者所要执行的目标业务;根据所述目标业务确定与所述目标文件相关的关联文件,所述关联文件为所述目标业务被执行时需要和所述目标文件一起被访问的文件;获取所述关联文件的访问地址,并根据所述发起地址和访问地址,检测所述访问发起者对所述关联文件的访问流量信息;若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求。本发明能够对大数据信息进行有效保护。



CN 118337466 A

1.一种基于大数据的信息安全保护方法,其特征在于,包括:

获取针对目标文件的访问请求,所述访问请求携带访问发起者的发起地址以及所述访问发起者所要执行的目标业务;

根据所述目标业务确定与所述目标文件相关的关联文件,所述关联文件为所述目标业务被执行时需要和所述目标文件一起被访问的文件;

获取所述关联文件的访问地址,并根据所述发起地址和访问地址,检测所述访问发起者对所述关联文件的访问流量信息;

若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求。

2.根据权利要求1所述的基于大数据的信息安全保护方法,其特征在于,在所述若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求之前,所述方法还包括:

获取所述关联文件对应所述目标业务的参考流量信息;

将所述访问流量信息与所述参考流量信息进行比对;

若所述访问流量信息与所述参考流量信息匹配,则确定所述访问流量信息满足预设要求。

3.根据权利要求2所述的基于大数据的信息安全保护方法,其特征在于,所述参考流量信息包括多个指定时刻的参考流量,所述访问流量信息包括多个时刻的访问流量,在所述若所述访问流量信息与所述参考流量信息匹配,则确定所述访问流量信息满足预设要求之前,所述方法还包括:

若所述多个时刻中包括所述多个指定时刻,则从所述多个时刻的访问流量中筛选出所述多个指定时刻的访问流量;

针对所述多个指定时刻中每一指定时刻,计算所述指定时刻对应的参考流量和访问流量之间的差值,得到多个差值;

将所述多个差值的绝对值进行相加后除以所述多个差值的数量,得到距离评估值;

若所述距离评估值小于或等于预设评估值,则确定所述访问流量信息与所述参考流量信息匹配。

4.根据权利要求2所述的基于大数据的信息安全保护方法,其特征在于,所述方法还包括:

若所述多个时刻中不包括所述多个指定时刻,则根据所述多个时刻的访问流量生成访问流量曲线,以及根据所述多个指定时刻的参考流量生成参考流量曲线;

识别所述访问流量曲线的拐点,并基于所述拐点将所述访问流量曲线划分为多个子流量曲线,其中,所述多个子流量曲线中每一子流量曲线对应一个斜率;

获取每一子流量曲线对应的目标斜率,得到斜率序列;

若所述斜率序列与所述参考流量曲线匹配,则确定所述访问流量信息满足预设要求。

5.根据权利要求4所述的基于大数据的信息安全保护方法,其特征在于,在所述若所述斜率序列与所述参考流量曲线匹配,则确定所述访问流量信息满足预设要求之前,所述方法还包括:

获取所述每一子流量曲线对应的时间段,得到多个目标时间段;

在所述参考流量曲线中获取所述多个目标时间段对应的多个参考斜率;

针对所述每一目标时间段,将所述目标时间段对应的参考斜率和目标斜率进行比对,

得到所述目标时间段对应斜率误差；

若每一所述目标时间段对应斜率误差均不超过误差阈值，则确定所述斜率序列与所述参考流量曲线匹配。

6.根据权利要求4所述的基于大数据的信息安全保护方法，其特征在于，在所述获取每一子流量曲线对应的斜率，得到斜率序列之前，所述方法还包括：

获取所述多个子流量曲线的曲线数量；

若所述多个子流量曲线的曲线数量与所述参考流量曲线匹配，则执行所述获取每一子流量曲线对应的斜率，得到斜率序列的步骤。

7.根据权利要求1所述的基于大数据的信息安全保护方法，其特征在于，在所述若所述访问流量信息满足预设要求，则对所述目标文件执行所述访问请求之后，所述方法还包括：

获取所述目标业务在历史时间段内被执行的过程中，对所述关联文件采集到的第一访问流量信息和对所述目标文件采集到的第二访问流量信息；

基于所述第一访问流量信息和所述第二访问流量信息进行模型训练，得到流量预测模型；

当获取到所述关联文件的当前访问流量信息时，采用所述流量预测模型对所述目标文件的访问流量进行预测，得到预测访问流量信息；

获取目标文件的目标访问地址，并根据所述发起地址和所述目标访问地址检测所述目标文件的当前访问流量信息；

将所述目标文件的当前访问流量信息和所述预测访问流量信息进行比对，得到流量信息误差；

若所述流量信息误差超过信息误差阈值，则关闭所述访问发起者对所述目标文件的访问权限。

8.根据权利要求7所述的基于大数据的信息安全保护方法，其特征在于，所述第一访问流量信息包括多个时刻的第一访问流量，所述第二访问流量信息包括所述多个时刻的第二访问流量，所述基于所述第一访问流量信息和所述第二访问流量信息进行模型训练，得到流量预测模型，包括：

基于多个时刻的第一访问流量生成第一流量曲线，并基于所述多个时刻的第二访问流量生成第二流量曲线；

从所述第一流量曲线中提取出第一曲线特征，从所述第二流量曲线中提取出第二曲线特征；

基于所述第一曲线特征和所述第二曲线特征进行模型训练，得到所述流量预测模型。

9.根据权利要求1至8中任一项所述的基于大数据的信息安全保护方法，其特征在于，所述关联文件的数量为多个，多个所述关联文件分别存储在预设分布式系统的不同的节点中。

10.一种基于大数据的信息安全保护系统，其特征在于，包括：

请求获取模块，用于获取针对目标文件的访问请求，所述访问请求携带访问发起者的发起地址以及所述访问发起者所要执行的目标业务；

确定模块，用于根据所述目标业务确定与所述目标文件相关的关联文件，所述关联文件为所述目标业务被执行时需要和所述目标文件一起被访问的文件；

流量检测模块,用于获取所述关联文件的访问地址,并根据所述发起地址和访问地址,检测所述访问发起者对所述关联文件的访问流量信息;

执行模块,用于若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求。

一种基于大数据的信息安全保护方法及系统

技术领域

[0001] 本发明涉及大数据技术领域,具体为一种基于大数据的信息安全保护方法及系统。

背景技术

[0002] 大数据的信息安全保护是确保大数据系统中存储和处理的信息得到保护和安全处理的过程。由于大数据系统涉及的数据量庞大、来源复杂,以及数据处理过程中可能涉及敏感信息,因此信息保护变得尤为重要。

[0003] 目前对大数据的信息安全保护通常是对用户设置访问权限,从而避免一些敏感信息被非法访问,然而,这种方式需要对大量的用户信息进行标注,需要花费大量的人力成本,保护效率较低,并且在用户信息泄露的情况下将使被访问的信息无法得到较好的保护。

发明内容

[0004] 针对现有技术中对大数据的信息保护不到位、保护效率较低的技术问题,本发明提供了一种基于大数据的信息安全保护方法及系统。

[0005] 为实现以上目的,本发明通过以下技术方案予以实现:

[0006] 本发明实施例第一方面,提供一种基于大数据的信息安全保护方法,该方法包括:

[0007] 获取针对目标文件的访问请求,所述访问请求携带访问发起者的发起地址以及所述访问发起者所要执行的目标业务;

[0008] 根据所述目标业务确定与所述目标文件相关的关联文件,所述关联文件为所述目标业务被执行时需要和所述目标文件一起被访问的文件;

[0009] 获取所述关联文件的访问地址,并根据所述发起地址和访问地址,检测所述访问发起者对所述关联文件的访问流量信息;

[0010] 若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求。

[0011] 可选地,在所述若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求之前,所述方法还包括:

[0012] 获取所述关联文件对应所述目标业务的参考流量信息;

[0013] 将所述访问流量信息与所述参考流量信息进行比对;

[0014] 若所述访问流量信息与所述参考流量信息匹配,则确定所述访问流量信息满足预设要求。

[0015] 可选地,所述参考流量信息包括多个指定时刻的参考流量,所述访问流量信息包括多个时刻的访问流量,在所述若所述访问流量信息与所述参考流量信息匹配,则确定所述访问流量信息满足预设要求之前,所述方法还包括:

[0016] 若所述多个时刻中包括所述多个指定时刻,则从所述多个时刻的访问流量中筛选出所述多个指定时刻的访问流量;

[0017] 针对所述多个指定时刻中每一指定时刻,计算所述指定时刻对应的参考流量和访

问流量之间的差值,得到多个差值;

[0018] 将所述多个差值的绝对值进行相加后除以所述多个差值的数量,得到距离评估值;

[0019] 若所述距离评估值小于或等于预设评估值,则确定所述访问流量信息与所述参考流量信息匹配。

[0020] 可选地,所述方法还包括:

[0021] 若所述多个时刻中不包括所述多个指定时刻,则根据所述多个时刻的访问流量生成访问流量曲线,以及根据所述多个指定时刻的参考流量生成参考流量曲线;

[0022] 识别所述访问流量曲线的拐点,并基于所述拐点将所述访问流量曲线划分为多个子流量曲线,其中,所述多个子流量曲线中每一子流量曲线对应一个斜率;

[0023] 获取每一子流量曲线对应的目标斜率,得到斜率序列;

[0024] 若所述斜率序列与所述参考流量曲线匹配,则确定所述访问流量信息满足预设要求。

[0025] 可选地,在所述若所述斜率序列与所述参考流量曲线匹配,则确定所述访问流量信息满足预设要求之前,所述方法还包括:

[0026] 获取所述每一子流量曲线对应的时间段,得到多个目标时间段;

[0027] 在所述参考流量曲线中获取所述多个目标时间段对应的多个参考斜率;

[0028] 针对所述每一目标时间段,将所述目标时间段对应的参考斜率和目标斜率进行比对,得到所述目标时间段对应斜率误差;

[0029] 若每一所述目标时间段对应斜率误差均不超过误差阈值,则确定所述斜率序列与所述参考流量曲线匹配。

[0030] 可选地,在所述获取每一子流量曲线对应的斜率,得到斜率序列之前,所述方法还包括:

[0031] 获取所述多个子流量曲线的曲线数量;

[0032] 若所述多个子流量曲线的曲线数量与所述参考流量曲线匹配,则执行所述获取每一子流量曲线对应的斜率,得到斜率序列的步骤。

[0033] 可选地,在所述若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求之后,所述方法还包括:

[0034] 获取所述目标业务在历史时间段内被执行的过程中,对所述关联文件采集到的第一访问流量信息和对所述目标文件采集到的第二访问流量信息;

[0035] 基于所述第一访问流量信息和所述第二访问流量信息进行模型训练,得到流量预测模型;

[0036] 当获取到所述关联文件的当前访问流量信息时,采用所述流量预测模型对所述目标文件的访问流量进行预测,得到预测访问流量信息;

[0037] 获取目标文件的目标访问地址,并根据所述发起地址和所述目标访问地址检测所述目标文件的当前访问流量信息;

[0038] 将所述目标文件的当前访问流量信息和所述预测访问流量信息进行比对,得到流量信息误差;

[0039] 若所述流量信息误差超过信息误差阈值,则关闭所述访问发起者对所述目标文件

的访问权限。

[0040] 可选地,所述第一访问流量信息包括多个时刻的第一访问流量,所述第二访问流量信息包括所述多个时刻的第二访问流量;所述基于所述第一访问流量信息和所述第二访问流量信息进行模型训练,得到流量预测模型,包括:

[0041] 基于多个时刻的第一访问流量生成第一流量曲线,并基于所述多个时刻的第二访问流量生成第二流量曲线;

[0042] 从所述第一流量曲线中提取出第一曲线特征,从所述第二流量曲线中提取出第二曲线特征;

[0043] 基于所述第一曲线特征和所述第二曲线特征进行模型训练,得到所述流量预测模型。

[0044] 可选地,所述关联文件的数量为多个,多个所述关联文件分别存储在预设分布式系统的不同的节点中。

[0045] 本发明实施例第二方面,提供一种基于大数据的信息保护系统,该系统包括:

[0046] 请求获取模块,用于获取针对目标文件的访问请求,所述访问请求携带访问发起者的发起地址以及所述访问发起者所要执行的目标业务;

[0047] 确定模块,用于根据所述目标业务确定与所述目标文件相关的关联文件,所述关联文件为所述目标业务被执行时需要和所述目标文件一起被访问的文件;

[0048] 流量检测模块,用于获取所述关联文件的访问地址,并根据所述发起地址和访问地址,检测所述访问发起者对所述关联文件的访问流量信息;

[0049] 执行模块,用于若所述访问流量信息满足预设要求,则对所述目标文件执行所述访问请求。

[0050] 本发明提供了一种基于大数据的信息安全保护方法及系统。与现有技术相比具备以下有益效果:

[0051] 本实施例提供的方案通过获取针对目标文件的访问请求,访问请求携带访问发起者的发起地址以及访问发起者所要执行的目标业务;再根据目标业务确定与目标文件相关的关联文件,其中,关联文件为目标业务被执行时需要和目标文件一起被访问的文件;然后,获取关联文件的访问地址,并根据发起地址和访问地址,检测访问发起者对关联文件的访问流量信息;若访问流量信息满足预设要求,则对目标文件执行访问请求。由于用户在执行一个业务时通常会访问不同的文件,从而使得同一个业务下需要的文件会存在一定关联性,所以本实施例通过检测与目标文件相关联的关联文件的访问流量信息来确定对目标文件的访问是否异常,从而决定目标文件是否能被访问,对目标文件起到了较好的保护作用,其中,由于访问流量信息不容易被篡改,所以进一步提高了保护力度,另外,避免了对用户信息进行权限设置的操作,也大大提升了保护效率。

附图说明

[0052] 图1是根据一示例性实施例示出的一种基于大数据的信息安全保护方法的应用场景示意图;

[0053] 图2是根据一示例性实施例示出的一种基于大数据的信息安全保护方法的流程图;

[0054] 图3是根据一示例性实施例示出的参考流量曲线示意图;

[0055] 图4是根据一示例性实施例示出的访问流量曲线示意图;

[0056] 图5是根据一示例性实施例示出的一种基于大数据的信息保护系统的原理框图。

具体实施方式

[0057] 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。

[0058] 可以理解的是,在本申请的具体实施方式中,涉及到访问请求、地址等相关的数据,当本申请以上实施例运用到具体产品或技术中时,需要获得用户许可或者同意,且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

[0059] 图1是根据一示例性实施例示出的一种基于大数据的信息安全保护方法的应用环境示意图,如图1所示,该应用环境可以包括服务器100和移动终端200,该服务器100与移动终端200通信连接,该服务器100可以是存储有多个文件以供用户在执行业务时进行访问,其中,对文件的访问可以包括对文件进行内容修改、内容删除、内容上传、内容下载等处理。可选地,该服务器100可以是单独的服务器也可以是服务器集群,在此不做限定。其中,该移动终端200可以用户用来向服务器100发送访问请求的设备。可选地,该移动终端200可以包括但不限于:平板电脑、智能手机、笔记本电脑、影像采集装置等等。可选地,该服务器100和移动终端200的数量可以为一个或多个,在此不做限定。

[0060] 图2是根据一示例性实施例示出的一种基于大数据的信息安全保护方法的流程图,如图2所示,上述方法包括以下步骤:

[0061] 101、获取针对目标文件的访问请求,访问请求携带访问发起者的发起地址以及访问发起者所要执行的目标业务。

[0062] 示例性地,该基于大数据的信息安全保护方法可以应用于上述服务器。

[0063] 其中,访问请求可以为访问发起者向服务器发送的用于申请访问目标文件的请求。其中,该访问请求除了上述携带的信息以外,还可以携带目标文件的文件信息,如目标文件的名称、标识、类型、访问地址等。

[0064] 其中,发起地址可以是访问发起者发送访问请求的网络地址。

[0065] 其中,该目标文件的访问地址可以是网络地址,即可以通过网络协议(如HTTP、FTP、SFTP等)对目标文件进行访问的地址。

[0066] 其中,目标业务为访问发起者当前需要执行的业务。该业务可以是交易业务、项目进度更新业务等等。

[0067] 可以理解的是,目标文件是在目标业务被执行时必定会访问的文件。

[0068] 102、根据目标业务确定与目标文件相关的关联文件,关联文件为目标业务被执行时需要和目标文件一起被访问的文件。

[0069] 示例性地,例如在目标业务执行时需要访问文件1、文件2、文件3以及文件4。其中,文件4为目标文件,那么文件1、文件2、文件3则为关联文件。可选地,可以预先将多个业务中每个业务在执行时需要访问的文件与该业务进行绑定,并将绑定关系存储至服务器本地,

以便服务器在确定目标文件和目标业务后,可以根据目标业务和该绑定关系快速差查找到关联业务。

[0070] 其中,关联文件的数量为多个,多个关联文件分别存储在预设分布式系统的不同的节点中。

[0071] 示例性地,例如预设分布式系统可以包括相互通信的多个电子设备,每个电子设备作为一个节点,可以将多个关联文本分别存储在多个电子设备中,从而可以增加非法用户对关联文本的信息的提取难度,大大增加了对信息的保护力度。可选地,每个电子设备可以存储一个或多个关联文本。

[0072] 103、获取关联文件的访问地址,并根据发起地址和访问地址,检测访问发起者对关联文件的访问流量信息。

[0073] 示例性地,例如发起地址为IP地址1,访问地址为IP地址2,则服务器可以实时检测IP地址1与IP地址2进行数据传输时的流量信息,并将该流量信息作为访问流量信息。

[0074] 其中,IP地址是计算机在网络上的唯一标识,用于定位和寻找计算机设备。

[0075] 其中,流量信息可以包括数据传输过程中每个时刻传输的数据包数量或数据传输量。

[0076] 104、若访问流量信息满足预设要求,则对目标文件执行访问请求。

[0077] 在一些实施方式中,该方法还可以包括:

[0078] 若访问流量信息满足预设要求,则不对该访问请求做出任何处理,并输出表征目标文件被非法访问的提示信息,该提示信息可以是音频信息、文本信息、图像信息等等。

[0079] 其中,在步骤104之前,该方法还可以包括:

[0080] S1、获取关联文件对应目标业务的参考流量信息。

[0081] 其中,关联文件对应目标业务的参考流量信息可以是过去正常执行目标业务时在访问该关联文件的过程中采集到的历史流量信息。其中,关联文件对应目标业务的参考流量信息可以预先存储在服务器本地,在需要使用时可以根据关联文件的标识和目标业务的标识从服务器本地中提取出相应的参考流量信息。

[0082] S2、将访问流量信息与参考流量信息进行比对。

[0083] S3、若访问流量信息与参考流量信息匹配,则确定访问流量信息满足预设要求。

[0084] 在一些实施方式中,参考流量信息包括多个指定时刻的参考流量,访问流量信息包括多个时刻的访问流量,在步骤S3之前,该方法还可以包括:

[0085] 若多个时刻中包括多个指定时刻,则从多个时刻的访问流量中筛选出多个指定时刻的访问流量。

[0086] 示例性地,例如多个时刻为t1、t2、t3、t4、t5、t6。t1对应的参考流量为x1、t2对应的参考流量为x2、t3对应的参考流量为x3、t4对应的参考流量为x4、t5对应的参考流量为x5、t6对应的参考流量为x6。多个指定时刻为多个指定时刻为t1、t3、t5。t1对应的访问流量为y1,t3对应的访问流量为y3,t5对应的访问流量为y5。

[0087] 针对多个指定时刻中每一指定时刻,计算指定时刻对应的参考流量和访问流量之间的差值,得到多个差值。

[0088] 沿用上述示例,可以得到t1对应的差值(x1-y1)、t3对应的差值(x3-y3)、t5对应的差值(x5-y5)。

[0089] 将多个差值的绝对值进行相加后除以多个差值的数量,得到距离评估值。

[0090] 沿用上述示例,距离评估值d0可以表示为d0＝(|x1-y1|+|x3-y3|+|x5-y5|)/3。

[0091] 若距离评估值小于或等于预设评估值,则确定访问流量信息与参考流量信息匹配。

[0092] 沿用上述示例,例如预设评估值为d,若距离评估值d0小于或等于d,则可以确定访问流量信息与参考流量信息匹配。

[0093] 可以理解的是,当距离评估值越小时,表明访问流量信息与参考流量信息越接近,当距离评估值小到一定程度(如小于或等于预设评估值),则表明访问流量信息与参考流量信息之间的差距可以忽略不计,此时可以表明访问流量信息与参考流量信息匹配。

[0094] 可见,在本实施方式中,通过针对多个指定时刻中每一指定时刻,计算指定时刻对应的参考流量和访问流量之间的差值,得到多个差值,将多个差值的绝对值进行相加后除以多个差值的数量,得到距离评估值,若距离评估值小于或等于预设评估值,则确定访问流量信息与参考流量信息匹配。从而可以快速精准地确定访问流量信息与参考流量信息是否匹配。

[0095] 在另一些实施方式中,该方法还可以包括:

[0096] 若多个时刻中不包括多个指定时刻,则根据多个时刻的访问流量生成访问流量曲线,以及根据多个指定时刻的参考流量生成参考流量曲线。

[0097] 示例性地,如图3所示,在得到多个指定时刻的参考流量可以将多个指定时刻作为横坐标、参考流量作为纵坐标在一个平面坐标系中得到多个参考流量对应的坐标,如坐标(t1,x1)、坐标(t2,x2)、坐标(t3,x3)坐标(t4,x4)坐标(t5,x5),然后,将上述坐标进行连线处理,可以得到如图3所示的参考流量曲线。

[0098] 如图4所示,在得到多个时刻的访问流量可以将多个时刻作为横坐标、访问流量作为纵坐标在一个平面坐标系中得到多个访问流量对应的坐标,如坐标(t1,y1)、坐标(t1.3,y1.3)、坐标(t2,x2)、坐标(t2.6,y2.6)坐标(t3.8,y3.8),然后,将上述坐标进行连线处理,可以得到如图4所示的访问流量曲线。

[0099] 识别访问流量曲线的拐点,并基于拐点将访问流量曲线划分为多个子流量曲线,其中,多个子流量曲线中每一子流量曲线对应一个斜率。

[0100] 其中,拐点可以是指访问流量曲线中斜率发生变化的点。

[0101] 示例性的,以图4为例,拐点可以包括坐标(t1.3,y1.3)和坐标(t2.6,y2.6),因此,可以将坐标(0,0)到坐标(t1.3,y1.3)之间的曲线作为一个子流量曲线,将坐标(t1.3,y1.3)到坐标(t2.6,y2.6)之间的曲线作为一个子流量曲线,将坐标(t2.6,y2.6)到坐标(t3.8,y3.8)之间的曲线作为一个子流量曲线。

[0102] 获取每一子流量曲线对应的目标斜率,得到斜率序列。

[0103] 沿用上述示例,例如坐标(t1.3,y1.3)到坐标(t2.6,y2.6)之间的子流量曲线,可以计算该子流量曲线对应的目标斜率为(y2.6-y1.3)/(t2.6-t1.3)。同理,可以通过上述方式计算到每一子流量曲线对应的目标斜率,并将多个目标斜率根据时间从前到后的顺序进行排序,即可得到斜率序列。

[0104] 若斜率序列与参考流量曲线匹配,则确定访问流量信息满足预设要求。

[0105] 可选地,在若斜率序列与参考流量曲线匹配,则确定访问流量信息满足预设要求

之前,方法还包括:

[0106] 获取每一子流量曲线对应的时间段,得到多个目标时间段。

[0107] 沿用上述示例,例如多个目标时间段包括第一时间段(0-t1.3),第二时间段(t1.3-t2.6)、第三时间段(t2.6-t3.8)。

[0108] 在参考流量曲线中获取多个目标时间段对应的多个参考斜率。

[0109] 针对每一目标时间段,将目标时间段对应的参考斜率和目标斜率进行比对,得到目标时间段对应斜率误差。

[0110] 沿用上述示例,例如可以从访问流量曲线中根据第一时间段得到第一斜率k1、根据第二时间段得到第二斜率k2、根据第三时间段得到第三斜率k3,然后,可以从参考流量曲线中根据第一时间段得到第四斜率k4、根据第五时间段得到第五斜率k5、根据第六时间段得到第六斜率k6。

[0111] 然后,针对第一时间段,可以得到斜率误差为|k1-k4|,针对第二时间段,可以得到斜率误差为|k2-k5|,针对第三时间段,可以得到斜率误差为|k3-k6|。

[0112] 若每一目标时间段对应斜率误差均不超过误差阈值,则确定斜率序列与参考流量曲线匹配。

[0113] 沿用上述示例,例如|k1-k4|、|k2-k5|、|k3-k6|均小于误差阈值k,则可以确定斜率序列与参考流量曲线匹配。

[0114] 可选地,在获取每一子流量曲线对应的斜率,得到斜率序列之前,方法还包括:

[0115] 获取多个子流量曲线的曲线数量。

[0116] 若多个子流量曲线的曲线数量与参考流量曲线匹配,则执行获取每一子流量曲线对应的斜率,得到斜率序列的步骤。

[0117] 示例性地,可以根据参考流量曲线中的拐点对参考流量曲线进行参考子流量曲线的划分,以使每个参考子流量曲线对应一个斜率。然后获取参考子流量曲线的数量,将参考子流量曲线的数量与多个子流量曲线的曲线数量进行比对,若二者的数量一致,则执行获取每一子流量曲线对应的斜率,得到斜率序列的步骤。若二者的数量不一致,则不执行该访问请求。从而提升了对访问请求的处理效率。

[0118] 在本实施方式中,通过在多个时刻中不包括多个指定时刻时,根据多个时刻的访问流量生成访问流量曲线,以及根据多个指定时刻的参考流量生成参考流量曲线,并根据访问流量曲线和参考流量曲线的匹配情况确定对目标文件的访问是否异常,从而可以更加灵活对访问请求进行处理,提高对目标文件的保护效率。

[0119] 在一些实施方式中,在步骤104之后,还可以包括:

[0120] 获取目标业务在历史时间段内被执行的过程中,对关联文件采集到的第一访问流量信息和对目标文件采集到的第二访问流量信息。

[0121] 基于第一访问流量信息和第二访问流量信息进行模型训练,得到流量预测模型。

[0122] 其中,可以将第二访问流量信息作为模型的预测目标进行模型训练,即训练好的流量预测模型可以根据输入的第一访问流量信息输出相应的第二访问流量信息。其中,第一访问流量信息可以作为训练集,第二访问流量信息可以作为验证集。

[0123] 可以理解的是,本实施方式中的模型训练方法可以采用常用的预测模型训练方法,具体可以采用线性回归、决策树、神经网络等方法,故不在此赘述。

[0124] 当获取到关联文件的当前访问流量信息时,采用流量预测模型对目标文件的访问流量进行预测,得到预测访问流量信息。

[0125] 获取目标文件的目标访问地址,并根据发起地址和目标访问地址检测目标文件的当前访问流量信息。

[0126] 将目标文件的当前访问流量信息和预测访问流量信息进行比对,得到流量信息误差。

[0127] 示例性地,可以将当前访问流量信息与预测访问流量信息之间的差值,确定为流量信息误差。

[0128] 若流量信息误差超过信息误差阈值,则关闭访问发起者对目标文件的访问权限。

[0129] 在本实施方式中,通过将目标文件的当前访问流量信息和预测访问流量信息进行比对,得到流量信息误差,若流量信息误差超过信息误差阈值,则关闭访问发起者对目标文件的访问权限,从而可以实时对目标文件进行精准地信息保护,防止目标文件被非法访问。

[0130] 其中,第一访问流量信息包括多个时刻的第一访问流量,第二访问流量信息包括多个时刻的第二访问流量。步骤“基于第一访问流量信息和第二访问流量信息进行模型训练,得到流量预测模型”的具体实施方式可以包括:

[0131] 基于多个时刻的第一访问流量生成第一流量曲线,并基于多个时刻的第二访问流量生成第二流量曲线。

[0132] 从第一流量曲线中提取出第一曲线特征,从第二流量曲线中提取出第二曲线特征。

[0133] 可选地,第一曲线特征和第二曲线特征可以包括但不限于曲线轮廓图像、曲线中的各个时间段的斜率、曲线中的极值等等。

[0134] 基于第一曲线特征和第二曲线特征进行模型训练,得到流量预测模型。

[0135] 在本实施方式中,通过从第一流量曲线中提取出第一曲线特征,从第二流量曲线中提取出第二曲线特征,并基于第一曲线特征和第二曲线特征进行模型训练,得到流量预测模型,从而可以利用参考流量信息和访问流量信息中更多的特征来进行模型训练,进而提升流量预测模型的准确性。

[0136] 可见,在本实施例中,通过获取针对目标文件的访问请求,访问请求携带访问发起者的发起地址以及访问发起者所要执行的目标业务;再根据目标业务确定与目标文件相关的关联文件,其中,关联文件为目标业务被执行时需要和目标文件一起被访问的文件;然后,获取关联文件的访问地址,并根据发起地址和访问地址,检测访问发起者对关联文件的访问流量信息;若访问流量信息满足预设要求,则对目标文件执行访问请求。由于用户在执行一个业务时通常会访问不同的文件,从而使得同一个业务下需要的文件会存在一定关联性,所以本实施例通过检测与目标文件相关联的关联文件的访问流量信息来确定对目标文件的访问是否异常,从而决定目标文件是否能被访问,对目标文件起到了较好的保护作用,其中,由于访问流量信息不容易被篡改,所以进一步提高了保护力度,另外,避免了对用户信息进行权限设置的操作,也大大提升了保护效率。

[0137] 在本实施例中,如图5所示,还提供一种基于大数据的信息保护系统,该基于大数据的信息保护系统300包括:

[0138] 请求获取模块310,用于获取针对目标文件的访问请求,上述访问请求携带访问发

起者的发起地址以及上述访问发起者所要执行的目标业务；

[0139] 确定模块320，用于根据上述目标业务确定与上述目标文件相关的关联文件，上述关联文件为上述目标业务被执行时需要和上述目标文件一起被访问的文件；

[0140] 流量检测模块330，用于获取上述关联文件的访问地址，并根据上述发起地址和访问地址，检测上述访问发起者对上述关联文件的访问流量信息；

[0141] 执行模块340，用于若上述访问流量信息满足预设要求，则对上述目标文件执行上述访问请求。

[0142] 在一些实施方式中，执行模块340，还用于：

[0143] 获取上述关联文件对应上述目标业务的参考流量信息；

[0144] 将上述访问流量信息与上述参考流量信息进行比对；

[0145] 若上述访问流量信息与上述参考流量信息匹配，则确定上述访问流量信息满足预设要求。

[0146] 在一些实施方式中，上述参考流量信息包括多个指定时刻的参考流量，上述访问流量信息包括多个时刻的访问流量，执行模块340，具体还用于：

[0147] 若上述多个时刻中包括上述多个指定时刻，则从上述多个时刻的访问流量中筛选出上述多个指定时刻的访问流量；

[0148] 针对上述多个指定时刻中每一指定时刻，计算上述指定时刻对应的参考流量和访问流量之间的差值，得到多个差值；

[0149] 将上述多个差值的绝对值进行相加后除以上述多个差值的数量，得到距离评估值；

[0150] 若上述距离评估值小于或等于预设评估值，则确定上述访问流量信息与上述参考流量信息匹配。

[0151] 在一些实施方式中，执行模块340还用于：

[0152] 若上述多个时刻中不包括上述多个指定时刻，则根据上述多个时刻的访问流量生成访问流量曲线，以及根据上述多个指定时刻的参考流量生成参考流量曲线；

[0153] 识别上述访问流量曲线的拐点，并基于上述拐点将上述访问流量曲线划分为多个子流量曲线，其中，上述多个子流量曲线中每一子流量曲线对应一个斜率；

[0154] 获取每一子流量曲线对应的目标斜率，得到斜率序列；

[0155] 若上述斜率序列与上述参考流量曲线匹配，则确定上述访问流量信息满足预设要求。

[0156] 在一些实施方式中，执行模块340，具体还用于：

[0157] 获取上述每一子流量曲线对应的时间段，得到多个目标时间段；

[0158] 在上述参考流量曲线中获取上述多个目标时间段对应的多个参考斜率；

[0159] 针对上述每一目标时间段，将上述目标时间段对应的参考斜率和目标斜率进行比对，得到上述目标时间段对应斜率误差；

[0160] 若每一上述目标时间段对应斜率误差均不超过误差阈值，则确定上述斜率序列与上述参考流量曲线匹配。

[0161] 在一些实施方式中，执行模块340，具体还用于：

[0162] 获取上述多个子流量曲线的曲线数量；

[0163] 若上述多个子流量曲线的曲线数量与上述参考流量曲线匹配,则执行上述获取每一子流量曲线对应的斜率,得到斜率序列的步骤。

[0164] 在一些实施方式中,该系统300还包括:预测模块,该预测模块用于:

[0165] 获取上述目标业务在历史时间段内被执行的过程中,对上述关联文件采集到的第一访问流量信息和对上述目标文件采集到的第二访问流量信息;

[0166] 基于上述第一访问流量信息和上述第二访问流量信息进行模型训练,得到流量预测模型;

[0167] 当获取到上述关联文件的当前访问流量信息时,采用上述流量预测模型对上述目标文件的访问流量进行预测,得到预测访问流量信息;

[0168] 获取目标文件的目标访问地址,并根据上述发起地址和上述目标访问地址检测上述目标文件的当前访问流量信息;

[0169] 将上述目标文件的当前访问流量信息和上述预测访问流量信息进行比对,得到流量信息误差;

[0170] 若上述流量信息误差超过信息误差阈值,则关闭上述访问发起者对上述目标文件的访问权限。

[0171] 在一些实施方式中,上述第一访问流量信息包括多个时刻的第一访问流量,上述第二访问流量信息包括上述多个时刻的第二访问流量;该预测模块具体用于:

[0172] 基于多个时刻的第一访问流量生成第一流量曲线,并基于上述多个时刻的第二访问流量生成第二流量曲线;

[0173] 从上述第一流量曲线中提取出第一曲线特征,从上述第二流量曲线中提取出第二曲线特征;

[0174] 基于上述第一曲线特征和上述第二曲线特征进行模型训练,得到上述流量预测模型。

[0175] 在一些实施方式中,上述关联文件的数量为多个,多个上述关联文件分别存储在预设分布式系统的不同的节点中。

[0176] 基于同一方面构思本公开实施例还提供一种电子设备,包括:

[0177] 存储器,其上存储有计算机程序;

[0178] 处理器,用于执行上述存储器中的上述计算机程序,以实现上述实施例中任意一项上述的一种基于大数据的信息安全保护方法的步骤。

[0179] 以上述依据本申请的理想实施例为启示,通过上述的说明内容,相关工作人员完全可以在不偏离本项申请技术思想的范围内,进行多样的变更以及修改。本项申请的技术性范围并不局限于说明书上的内容,必须要根据权利要求范围来确定其技术性范围。

[0180] 尽管已经示出和描述了本发明的实施例,对于本领域的普通技术人员而言,可以理解在不脱离本发明的原理和精神的情况下可以对这些实施例进行多种变化、修改、替换和变型,本发明的范围由所附权利要求及其等同物限定。

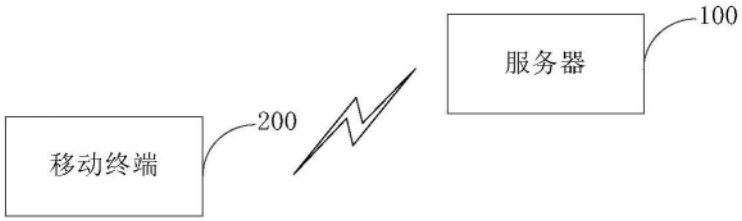

服务器
100
移动终端
200

图1

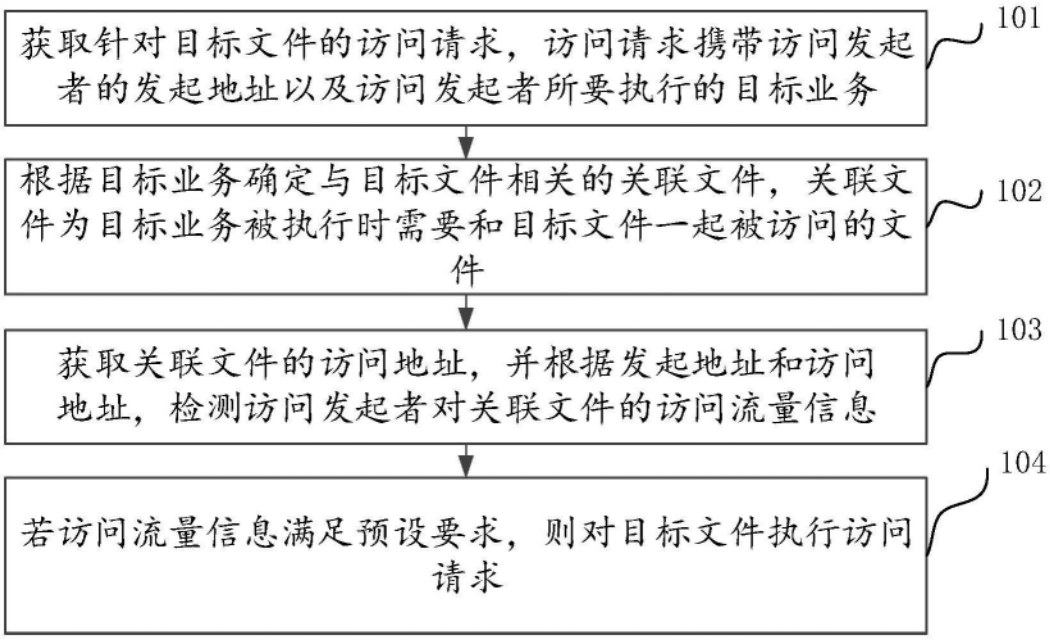

获取针对目标文件的访问请求，访问请求携带访问发起者的发起地址以及访问发起者所要执行的目标业务
101
根据目标业务确定与目标文件相关的关联文件，关联文件为目标业务被执行时需要和目标文件一起被访问的文件
102
获取关联文件的访问地址，并根据发起地址和访问地址，检测访问发起者对关联文件的访问流量信息
103
若访问流量信息满足预设要求，则对目标文件执行访问请求
104

图2

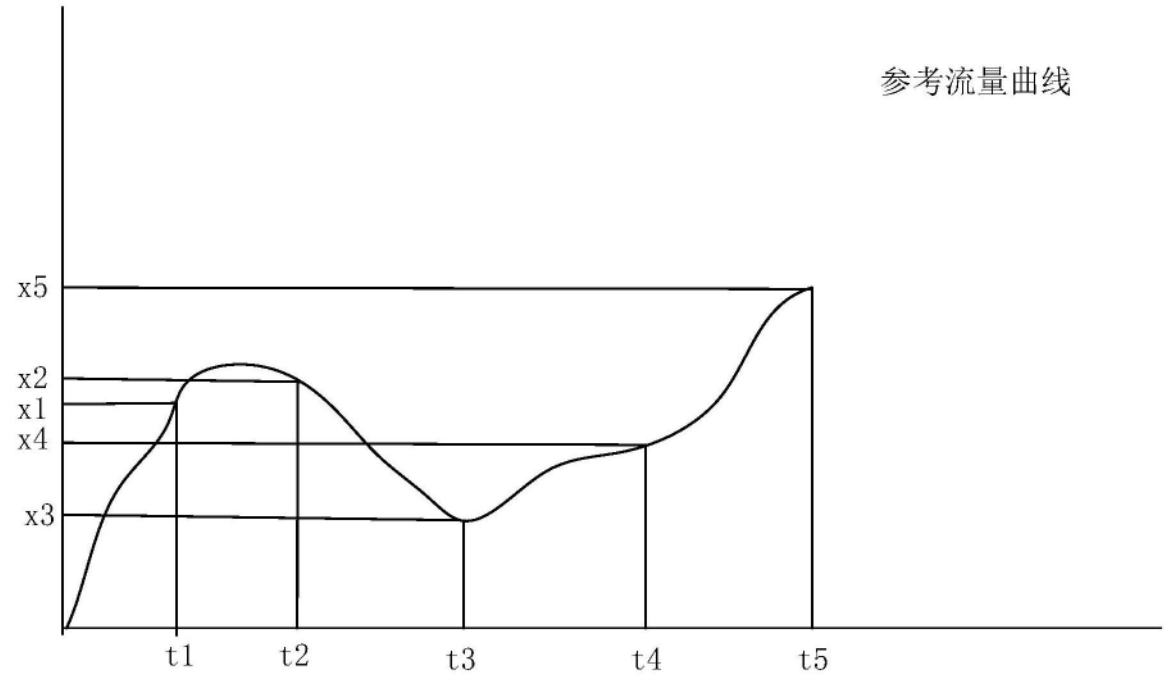

参考流量曲线
x5
x2
x1
x4
x3
t1
t2
t3
t4
t5

图3

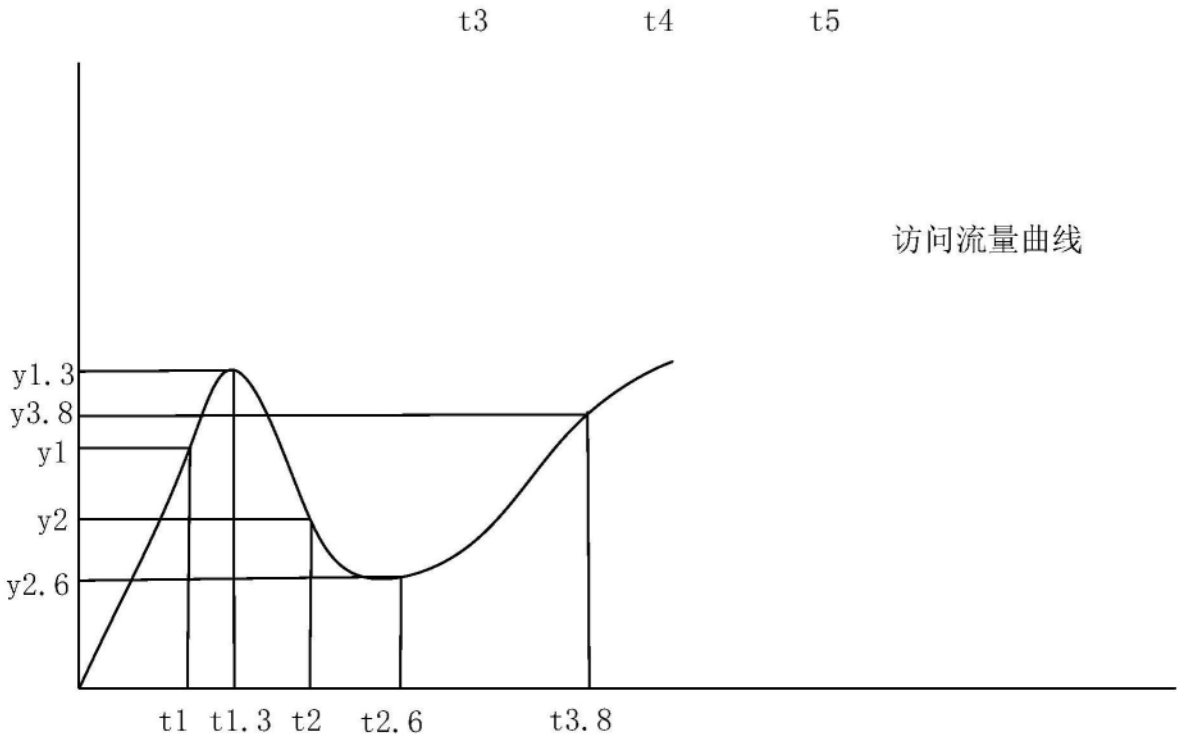

t3
t4
t5
访问流量曲线
y1.3
y3.8
y1
y2
y2.6
t1
t1.3
t2
t2.6
t3.8

图4

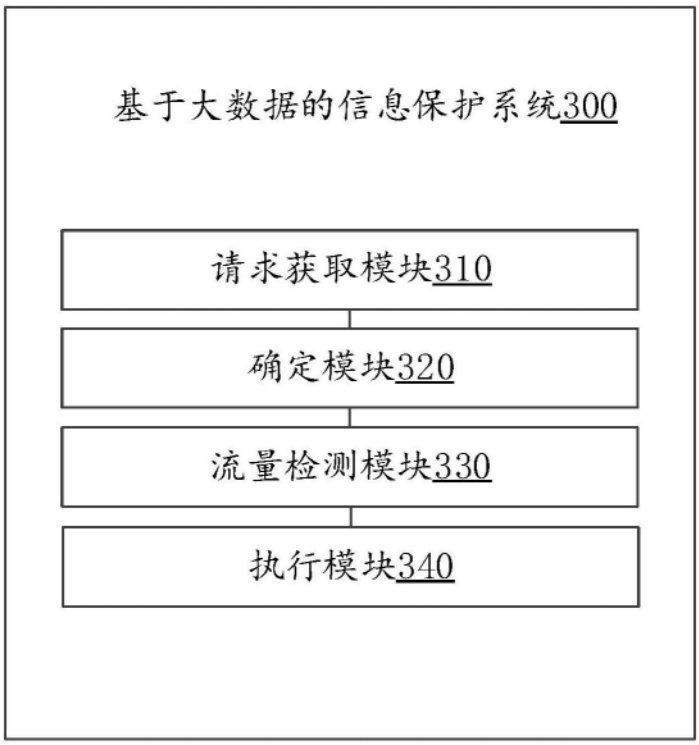


图5