(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116527366 A
(43)申请公布日 2023.08.01

(21)申请号 202310513163.9

(22)申请日 2023.05.08

(71)申请人 清华大学
地址 100084 北京市海淀区清华园
申请人 中国民用航空局空中交通管理局

(72)发明人 张毅 何泓霖 晏松 杨敬轩
陈宝刚 杨锐 马超

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 李丹 栗若木

(51)Int.Cl.
H04L 9/40(2022.01)
H04L 67/1396(2022.01)

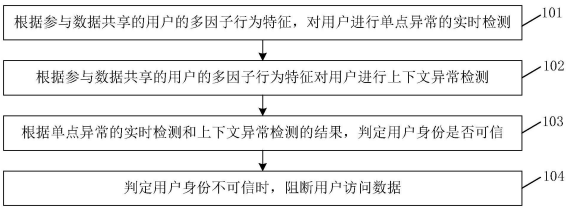
权利要求书2页 说明书8页 附图2页

(54)发明名称

一种用于共享数据访问的用户身份可信辨识的方法

(57)摘要

本文公开一种用于共享数据访问的用户身份可信辨识的方法，包括：根据参与数据共享的用户的多因子行为特征，对用户进行单点异常的实时检测；根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测；根据单点异常的实时检测和上下文异常检测的结果，判定用户身份是否可信；判定用户身份不可信时，阻断用户访问数据；其中，多因子行为特征包括：通过存储数据的系统记录的用户在数据共享过程中的与访问行为相关的特征信息。本发明实施例根据多因子行为特征对用户进行单点异常的实时键合和上下文异常检测，阻断判定为身份不可信的用户的访问操作，提升了数据共享系统的数据安全性。



CN 116527366 A

1.一种用于共享数据访问的用户身份可信辨识的方法,包括:

根据参与数据共享的用户的多因子行为特征,对用户进行单点异常的实时检测;

根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测;

根据单点异常的实时检测和上下文异常检测的结果,判定用户身份是否可信;

判定用户身份不可信时,阻断用户访问数据;

其中,所述多因子行为特征包括:通过存储所述数据的系统记录的用户在数据共享过程中的与访问行为相关的特征信息。

2.根据权利要求1所述的方法,其特征在于,所述多因子行为特征包括以下一项或任意组合:

用户登录时刻、用户授权等级、用户在检测时刻和上一检测时刻的时间段内的上下行流量、用户自登录以来的上下行流量总量、用户的登录互联网地址IP和用户在每一时刻的上下行流量;

其中,所述检测时刻为预先设定的固定检测时长的时刻。

3.根据权利要求1所述的方法,其特征在于,所述对用户进行单点异常的实时检测,包括:

采用预先训练获得的孤立森林模型,对用户进行所述单点异常的实时检测;

其中,所述孤立森林模型的输入包括以下一项或任意组合的所述多因子行为特征:用户登录时刻、用户在当前的检测时刻和上一检测时刻这段时间内的上下行流量、用户自登录以来的上下行流量总量;检测时刻为预先设定的固定检测时长的时刻。

4.根据权利要求3所述的方法,其特征在于,所述判定用户身份是否可信,包括:

所述孤立森林模型输出的异常值得分值大于或等于预先设定的第一异常得分阈值时,判定用户身份异常;

所述孤立森林模型输出的所述异常值得分值小于或等于预先设定的第二异常得分阈值时,判定用户身份正常。

5.根据权利要求3或4所述的方法,其特征在于,所述根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测,包括:

采用长短时记忆神经网络模型,对所述用户进行所述上下文异常检测;

其中,所述长短时记忆神经网络模型的输入包括以下一项或任意组合的所述多因子行为特征:用户授权等级、用户登录时刻、以及用户在当前检测时刻和上一检测时刻这段时间内的上下行流量。

6.根据权利要求5所述的方法,其特征在于,所述判定用户身份是否可信,包括:

所述长短时记忆神经网络模型输出的用户身份异常的概率值大于或等于预先设定的第一异常概率阈值时,判定用户身份异常;

所述长短时记忆神经网络模型输出的用户身份异常的概率值小于或等于预先设定的第二异常概率阈值时,判定用户身份正常。

7.根据权利要求5所述的方法,其特征在于,所述判定用户身份是否可信,包括:

初始化所述孤立森林模型输出的异常值得分值为0,按照预设周期获取所述孤立森林模型输出的异常值得分值,获取的异常得分值大于或等于预先设定的第三异常得分阈值时,对大于或等于预先设定的第三异常得分阈值进行累加,将累加结果作为第一异常值得

分值；

初始化所述长短时记忆神经网络模型输出的异常值得分值为0，按照预设周期获取所述长短时记忆神经网络模型输出的异常值得分值，获取的异常得分值大于或等于预先设定的第四异常得分阈值时，对大于或等于预先设定的第四异常得分阈值进行累加，将累加结果作为第二异常值得分值；

所述第一异常值得分值大于预设的一类异常阈值、或所述第二异常值得分值大于预设的二类异常阈值时，判定用户身份异常。

8.根据权利要求5所述的方法，其特征在于，所述对用户进行单点异常的实时检测之前，所述方法还包括：

利用预先设定的第一样本数据集，训练获得所述孤立森林模型；

通过预先设定的第二样本数据集，训练获得所述长短时记忆神经网络模型；

其中，所述第一样本数据集包括一项以上所述多因子行为特征；所述第二样本数据集包括一项以上所述多因子行为特征。

9.根据权利要求8所述的方法，其特征在于，所述多因子行为特征还包括：

所述用户的流量访问类型；

其中，所述流量访问类型包括以下一项或任意组合：网络时间协议NTP、因特网控制报文协议ICMP、对等网络P2P行为、WebSocket、远程连接工具SSH、位图显示的视窗系统X11、关系型数据库管理系统MySql、甲骨文Oracle、安全套接字协议SSL、应用层协议请求HTTP_POST及文件传输协议FTP下载。

10.一种计算机存储介质，所述计算机存储介质中存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求1-9中任一项所述的用于共享数据访问的用户身份可信辨识的方法。

11.一种终端，包括：存储器和处理器，所述存储器中保存有计算机程序；其中，

处理器被配置为执行存储器中的计算机程序；

所述计算机程序被所述处理器执行时实现如权利要求1-9中任一项所述的用于共享数据访问的用户身份可信辨识的方法。

一种用于共享数据访问的用户身份可信辨识的方法

技术领域

[0001] 本文涉及但不限于信息安全技术，尤指一种用于共享数据访问的用户身份可信辨识的方法。

背景技术

[0002] 在当前信息化的时代背景下，数据信息成为了重要的社会资源。同时，为了信息系统的可持续发展，促进企业间合作共赢，企业间数据共享就逐渐成为了一项不可或缺的重要事务。我国相关法律指出，国家保护个人、组织与数据有关的权益，鼓励数据依法合理有效利用，保障数据依法有序自由流动，促进以数据为关键要素的数字经济发展；开展数据处理活动应当加强风险监测，发现数据安全缺陷、漏洞等风险时，应当立即采取补救措施；发生数据安全事件时，应当立即采取处置措施，按照规定及时告知用户并向有关主管部门报告。随着信息化建设的全面推进，社会各界对数据共享需求的增加，对数据安全的重视程度不断增加，已经产生较为成熟的数据安全保护机制。

[0003] 在数据共享过程中，现有数据安全保护机制主要聚焦于对于数据本身的保护机制和对于数据传输链路的保护机制，包括：数据传输前的加密、数据链路监控等。然而对于数据共享过程中的重要组成部分，即进行数据共享的用户双方所采取的安全措施较少，通常只采取用户登录阶段身份认证，传输过程中流量限制等基本措施进行防护，如何进一步提高数据共享过程的安全性成为一个有待解决的问题。

发明内容

[0004] 以下是对本文详细描述的主题的概述。本概述并非是为了限制权利要求的保护范围。

[0005] 本发明实施例提供一种用于共享数据访问的用户身份可信辨识的方法，能够提高数据共享过程的安全性。

[0006] 本发明实施例提供了一种用于共享数据访问的用户身份可信辨识的方法，包括：

[0007] 根据参与数据共享的用户的多因子行为特征，对用户进行单点异常的实时检测；

[0008] 根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测；

[0009] 根据单点异常的实时检测和上下文异常检测的结果，判定用户身份是否可信；

[0010] 判定用户身份不可信时，阻断用户访问数据；

[0011] 其中，所述多因子行为特征包括：通过存储所述数据的系统记录的用户在数据共享过程中的与访问行为相关的特征信息。

[0012] 另一方面，本发明实施例还提供一种计算机存储介质，所述计算机存储介质中存储有计算机程序，所述计算机程序被处理器执行时实现上述用于共享数据访问的用户身份可信辨识的方法。

[0013] 再一方面，本发明实施例还提供一种终端，包括：存储器和处理器，所述存储器中保存有计算机程序；其中，

[0014] 处理器被配置为执行存储器中的计算机程序；

[0015] 所述计算机程序被所述处理器执行时实现如上述用于共享数据访问的用户身份可信辨识的方法。

[0016] 本申请技术方案包括：根据参与数据共享的用户的多因子行为特征，对用户进行单点异常的实时检测；根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测；根据单点异常的实时检测和上下文异常检测的结果，判定用户身份是否可信；判定用户身份不可信时，阻断用户访问数据；其中，所述多因子行为特征包括：通过存储所述数据的系统记录的用户在数据共享过程中的与访问行为相关的特征信息。本发明实施例根据多因子行为特征对用户进行单点异常的实时键合和上下文异常检测，阻断判定为身份不可信的用户的访问操作，提升了数据共享系统的数据安全性。

[0017] 本发明的其它特征和优点将在随后的说明书中阐述，并且，部分地从说明书中变得显而易见，或者通过实施本发明而了解。本发明的目的和其他优点可通过在说明书、权利要求书以及附图中所特别指出的结构来实现和获得。

附图说明

[0018] 附图用来提供对本发明技术方案的进一步理解，并且构成说明书的一部分，与本申请的实施例一起用于解释本发明的技术方案，并不构成对本发明技术方案的限制。

[0019] 图1为本发明实施例用于共享数据访问的用户身份可信辨识的方法的流程图；

[0020] 图2为本发明应用示例用于共享数据访问的用户身份可信辨识的方法流程图；

[0021] 图3为本发明应用示例上下文异常检测的示意图。

具体实施方式

[0022] 为使本发明的目的、技术方案和优点更加清楚明白，下文中将结合附图对本发明的实施例进行详细说明。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互任意组合。

[0023] 在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行。并且，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。

[0024] 图1为本发明实施例用于共享数据访问的用户身份可信辨识的方法的流程图，如图1所示，包括：

[0025] 步骤101、根据参与数据共享的用户的多因子行为特征，对用户进行单点异常的实时检测；

[0026] 步骤102、根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测；

[0027] 步骤103、根据单点异常的实时检测和上下文异常检测的结果，判定用户身份是否可信；

[0028] 步骤104、判定用户身份不可信时，阻断用户访问数据；

[0029] 其中，多因子行为特征包括：通过存储数据的系统记录的用户在数据共享过程中的与访问行为相关的特征信息。

[0030] 本发明实施例根据多因子行为特征对用户进行单点异常的实时键合和上下文异常检测,阻断判定为身份不可信的用户的访问操作,提升了数据共享系统的数据安全性。

[0031] 在一种示例性实例中,步骤101和步骤102在实施上没有先后顺序。

[0032] 在一种示例性实例中,本发明实施例中的多因子行为特征包括以下一项或任意组合:

[0033] 用户登录时刻、用户授权等级、用户在检测时刻和上一检测时刻的时间段内的上下行流量、用户自登录以来的上下行流量总量、用户的登录IP、用户在每一时刻的上下行流量;检测时刻为预先设定的固定检测时长的时刻。

[0034] 在一种示例性实例中,本发明实施例对用户进行单点异常的实时检测,包括:

[0035] 采用预先训练获得的孤立森林模型,对用户进行单点异常的实时检测;

[0036] 其中,孤立森林模型的输入包括以下一项或任意组合的多因子行为特征:用户登录时刻、用户在当前的检测时刻和上一检测时刻这段时间内的上下行流量、用户自登录以来的上下行流量总量;检测时刻为预先设定的固定检测时长的时刻。

[0037] 在一种示例性实例中,采用预先训练获得的孤立森林模型,对用户进行单点异常的实时检测时,孤立森林模型输出为每一个检测时刻下每一个用户的异常值得分;本发明实施例判定用户身份是否可信,包括:

[0038] 孤立森林模型输出的异常值得分值大于或等于预先设定的第一异常得分阈值时,判定用户身份异常;

[0039] 孤立森林模型输出的异常值得分值小于或等于预先设定的第二异常得分阈值时,判定用户身份正常;

[0040] 在一种示例性实例中,本发明实施例根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测,包括:

[0041] 采用长短时记忆神经网络模型,对用户进行上下文异常检测;

[0042] 其中,长短时记忆神经网络模型的输入包括以下一项或任意组合的多因子行为特征:用户授权等级、用户登录时刻、以及用户在当前检测时刻和上一检测时刻这段时间内的上下行流量。

[0043] 在一种示例性实例中,采用长短时记忆神经网络模型,对用户进行上下文异常检测时,长短时记忆神经网络模型的输出为每一个检测时刻下对应的用户身份异常的概率值;本发明实施例判定用户身份是否可信,包括:

[0044] 长短时记忆神经网络模型输出的用户身份异常的概率值大于或等于预先设定的第一异常概率阈值时,判定用户身份异常;

[0045] 长短时记忆神经网络模型输出的用户身份异常的概率值小于或等于预先设定的第二异常概率阈值时,判定用户身份正常。

[0046] 孤立森林模型输出的异常值得分值小于或等于预先设定的第二异常得分阈值时,判定用户身份正常。

[0047] 在一种示例性实例中,本发明实施例判定用户身份是否可信,包括:

[0048] 初始化孤立森林模型输出的异常值得分值为0,按照预设周期获取孤立森林模型输出的异常值得分值,获取的异常得分值大于或等于预先设定的第三异常得分阈值时,对大于或等于预先设定的第三异常得分阈值进行累加,将累加结果作为第一异常值得分值;

[0049] 初始化长短时记忆神经网络模型输出的异常值得分值为0,按照预设周期获取长短时记忆神经网络模型输出的异常值得分值,获取的异常得分值大于或等于预先设定的第四异常得分阈值时,对大于或等于预先设定的第四异常得分阈值进行累加,将累加结果作为第二异常值得分值;

[0050] 第一异常值得分值大于预设的一类异常阈值、或第二异常值得分值大于预设的二类异常阈值时,判定用户身份异常。在一种示例性实例中,本发明实施例之前还包括:

[0051] 利用预先设定的第一样本数据集,训练获得孤立森林模型;

[0052] 通过预先设定的第二样本数据集,训练获得长短时记忆神经网络模型;

[0053] 其中,第一样本数据集包括一项以上多因子行为特征;第二样本数据集包括一项以上多因子行为特征。

[0054] 在一种示例性实例中,本发明实施例多因子行为特征还包括:

[0055] 用户的流量访问类型;

[0056] 其中,流量访问类型包括以下一项或任意组合:网络时间协议(NTP)、因特网控制报文协议(ICMP)、对等网络(P2P)行为、WebSocket(是HTML5开始提供的一种浏览器与服务器间进行全双工通讯的网络技术)、远程连接工具(SSH)、位图显示的视窗系统(X11)、关系型数据库管理系统(MySql)、甲骨文(Oracle)、安全套接字协议(SSL)、应用层协议请求(HTTP_POST)及文件传输协议(FTP)下载。

[0057] 本发明实施例还提供一种计算机存储介质,计算机存储介质中存储有计算机程序,计算机程序被处理器执行时实现上述用于共享数据访问的用户身份可信辨识的方法。

[0058] 本发明实施例还提供一种终端,包括:存储器和处理器,存储器中保存有计算机程序;其中,

[0059] 处理器被配置为执行存储器中的计算机程序;

[0060] 计算机程序被处理器执行时实现如上述用于共享数据访问的用户身份可信辨识的方法。

[0061] 以下通过应用示例对本发明实施例进行简要说明,应用示例仅用于陈述本发明实施例,并不用于限定本发明的保护范围。

[0062] 应用示例

[0063] 与相关技术采用静态用户身份识别及管理机制不同,本应用示例基于用户长时间进行数据共享的多因子行为特征,构造数据共享过程中的多维度用户身份动态辨识算法,提高了数据共享过程的安全性。

[0064] 本发明应用示例方法包括:对参与数据共享的用户进行单点异常的实时检测;即实时的检测所有参与数据共享的用户的行为是否异常,在一种示例性实例中,本发明应用示例可以采用孤立森林模型进行单点异常的实时检测;在一种示例性实例中,本发明应用示例可以采用相关技术中的其他种类的单点异常的检测算法实现单点异常的实时检测;单点异常(Global Outliers),也可以称为全局异常,即某个点与全局大多数点都不一样,那么这个点构成了单点异常;

[0065] 本发明应用示例方法包括:对参与数据共享的用户,进行自登录以来的行为进行上下文异常检测;即判断用户自登录以来的行为是否具有时间维度上的异常性;在一种示例性实例中,本发明应用示例可以采用长短时记忆神经网络模型进行上下文异常检测;在

一种示例性实例中，本发明应用示例可以采用相关技术中的其他种类的上下文异常检测的检测算法实现上下文异常的检测；上下文异常，又被称之为情景异常，指的是一个对象在某一个情景下和大部分对象的行为差异比较大，这个数据对象就是在这个上下文或者这个情景下的异常。

[0066] 单点异常的实时检测和上下文异常检测可以获得相应的一个异常值得分，异常值得分达到预先设定的各检测相应的异常得分阈值时，即认为用户身份不可信；对用户访问数据的行为，执行预先设定的阻断措施，本发明应用示例工作流程如图2所示。

[0067] 本发明应用示例用于单点异常的实时检测和上下文异常检测的模型，其建立都需要首先收集所有授权用户在一定时间内的多因子行为特征，用以学习，从而获得相关参数；多因子行为特征为用户在数据共享过程中，通过计算机系统能够获得的用户访问存储数据的系统时的多类数据，主要包括：用户登录时刻、用户授权等级、用户在检测时刻和上一检测时刻的时间段内的上下行流量、用户自登录以来的上下行流量总量、用户的登录IP、用户在每一时刻的上下行流量等。这些多因子行为特征也是后续实时检测过程中所需要检测的相关数据，输入到模型中，模型依据上述过程输出相应异常得分值，用于对用户身份进行可信辨识。

[0068] 在一种示例性实例中，本应用示例采用孤立森林模型进行单点异常的实时检测时，孤立森林模型的输入为固定时刻(固定的进行用户行为是否异常检测的周期时长)下所有用户的多因子行为特征，根据数据共享过程的基本特点，兼顾辨识效率、准确性等因素，选择如：用户登录时刻、用户在当前检测时刻和上一检测时刻这段时间内的上下行流量、用户自登录以来的上下行流量总量等特征项作为模型的输入。孤立森林模型的输出为固定时刻下每个用户的异常值得分，得分值在0-1之间，若异常值得分大于预设的第一异常得分阈值，例如趋近于1，则认为对应用户在当期时刻行为存在异常(用户身份不可信)；若异常值得分小于或等于第二异常得分阈值，则认为用户不存在异常(用户身份可信)。

[0069] 在一种示例性实例中，本发明应用示例可以采用长短时记忆神经网络模型进行上下文异常检测时，不同用户的上下文异常检测的用户是互不相同的，即针对每一用户都包含相应的检测模型参数，其输入为当前检测时刻用户的多因子行为特征，即包含用户自登录以来行为特征信息的隐变量，兼顾辨识效率、准确性等因素，选择如：用户授权等级、用户登录时刻以及用户在当前检测时刻和上一检测时刻这段时间内的上下行流量等多因子行为特征项作为模型的输入。长短时记忆神经网络模型的输出为固定时刻下对应用户身份异常的概率值，其在0-1之间，若概率值大于预设的异常概率阈值，例如趋近于1，则认为对应用户在当前时刻行为存在异常；若概率值小于或等于异常概率阈值，例如趋近于0，则认为用户不存在异常。

[0070] 每一用户登录时，初始化其两类异常值得分S_1，S_2均为0，在每一时刻，通过上述两类模型输出的异常值结果分别为s_1，s_2，当s_1超过设定阈值d_1时，$S_1 \leftarrow S_1+s_1$，当s_2超过设定阈值d_2时，$S_2 \leftarrow S_2+s_2$；若在更新后，S_1超过一类异常阈值K_1或S_2超过二类异常阈值K_2，判定用户身份异常，即认为当前用户身份不可信；在一种示例性实例中，判定用户身份异常，采取相应的阻断措施。本发明应用示例从数据共享过程中数据使用双方的角度出发，对数据共享过程的安全进行考虑，以数据使用者的多因子行为特征作为依据，动态辨识用户身份的可信性，加强了数据共享过程的安全性。在一种示例性实例中，本应用示例基于孤立森林和长

短时记忆网络分别对用户群体、用户单体的时序特征进行异常分析，相比于仅基于流量阈值检测的方法，能够检测出更多具有潜在风险的行为。

[0071] 在一种示例性实例中，本应用示例单点异常的实时检测的孤立森林模型的建立方法如下：

[0072] 输入第一样本数据集(正常的数据集)$X=\{x_1,x_2,\cdots,x_i,\cdots,x_n\}$，$\forall x_i \in X, x_i=\{x_{i1},\cdots,x_{id}\}\in R^d$；其中，X为所有用户在一段时间内的访问行为数据，n为数据集样本数，d为行为多因子行为特征的维数；

[0073] 首先构建一颗孤立树，包括：

[0074] 步骤1、从第一样本数据集X中随机抽选k个样本点构成一个子集X_c，作为根节点；

[0075] 步骤2、从d个维度中随机选择一个维度q，并随机产生一个分割点p满足$\min(x_{iq},x_{iq}\in X_c)<p<\max(x_{iq},x_{iq}\in X_c)$；

[0076] 步骤3、将随机抽选k个样本点中满足$x_{iq}<p, x_{iq}\in X$的样本点放置于左子节点，反之，将随机抽选k个样本点中满足$x_{iq}\geq p, x_{iq}\in X$的样本点放置于右子节点；

[0077] 递归执行上述步骤2和步骤3，直至所有叶子节点都只有一个样本点，或者孤立树达到了指定高度h；

[0078] 循环上述处理，直至生成完整的孤立森林；

[0079] 对于每一个样本点x_i，令其遍历每一颗孤立树，计算其在森林中的平均高度$h(x_i)$，对所有样本点的平均高度做归一化处理，并通过式(1)～(3)计算异常值得分$s(x,n)$：

[0080] $$s(x,\psi)=2^{\frac{E(h(x))}{c(\psi)}} \quad (1)$$

[0081] $$c(\psi)=\begin{cases}2H(\psi-1)-\frac{2(\psi-1)}{\psi}, & \psi>2\\ 1, & \psi=2\\ 0, & \psi<2\end{cases} \quad (2)$$

[0082] $$H(k)=\ln(k)+0.5772156649 \quad (3)$$

[0083] 在利用模型进行身份可信辨识阶段，记录待确认的行为特征数据在每颗树的高度均值，获得每个用户的异常值得分。

[0084] 本应用示例基于长短时记忆网络的上下文异常用户检测模型建立方法如下：

[0085] 输入第二样本数据集(正常数据集)$M=\{m_1,m_2,\cdots,m_i,\cdots,m_n\}$，$\forall x_i \in M, m_i=\{m_{i1},\cdots,m_{id}\}\in R^d$，其中：M为某一用户I在一段时间内的访问行为数据，n为数据集样本数，d为行为特征维数，对其标签y_i记录为0，考虑到异常数据在真实数据共享环境中，特别是敏感性较高的条件下数量极其稀少，根据数据共享过程中异常的主要特点，构造如：周期性的高峰流量、非正常时刻用户登录等异常数据，并对其标签y_i记录为1；

[0086] 构造长短时记忆神经网络模型，网络隐藏层层数为h，第i隐藏层神经元个数为c_i，网络输入维度为d，输出维度为1，激活函数为A(•)，为确保最后一层输出为[0,1]范围内取值，采用Sigmoid函数$S(x)=\frac{1}{1+e^{-x}}$；

[0087] 建立损失函数$L(y,\hat{y})=||y-\hat{y}||_2$；其中，y和$\hat{y}$分别为真实标签和网络输出结

果，据此利用梯度下降方法Adam进行梯度下降，更新神经网络参数，当达到迭代次数上限K时终止，获得模型；

[0088] 在利用模型进行身份可信辨识阶段，用户登录时采用其模型，此后在每个检测时刻输入相应多因子行为特征数据，获得用户的异常得分值。

[0089] 在一种示例性实例中，本发明应用示例对于迭代次数上限，通常根据经验随机设定数值，在训练过程中，根据数据收敛情况，即：相邻两个训练过程的损失函数差异相较于损失函数的比值∈，当∈小于一个极小量(通常采取1^{e-3})时，可以认为模型收敛，终止训练；而当达到K时仍未满足该条件，则进一步加大K值，继续训练，直到收敛。当不可收敛时(可能性极低)，通常需要重新构造网络结构(增加层数等)，或增加神经元个数等，改进模型。

[0090] 如考虑包括流量阈值限制在内的安全检测方法，以基于长短时记忆网络的上下文异常用户检测模型为例，其检测流程图如图3所示。

[0091] 本发明应用示例基于孤立森立和长短时记忆神经网络，设计了数据共享过程中，依据数据共享人员多因子行为特征进行用户身份可信辨识，提升了数据共享过程中的安全性。

[0092] 本案例使用民航空管局数据库访问日志数据，包含有效数据项5308条；其中，所有用户每次登录到登出过程中产生的上行下行流量数据，数据主要集中在上下行流量均小于1吉比特(GB)的部分，包含数据项4934条，占总数据的93.0％。考虑基于孤立森林模型的异常行为辨识，在这一实例中，正常上下行流量为1GB，流量阈值设置为5GB。

[0093] 根据上下行流量数据进行孤立森林模型建立，模型通过学习可以判断各点的异常性，本发明应用示例，孤立森林能够在访问数据未达到阈值限制时，对存在异常的数据能够进行识别，据此，相应用户的累积异常得分会相应增加，当达到设定阈值时，系统便可对其进行阻断措施。而仅通过阈值限制进行检测的结果，则永远无法检测到行为存在长期异常(如：频繁高流量流通，窃取信息者)但未超出每次访问流量阈值的行为。

[0094] 在一种示例性实例中，本发明应用示例额外考虑用户流量访问类型特征，根据数据集信息，依据出现频率，除未标记访问类型的数据外，数据集中的流量访问类型依次为：NTP、ICMP协议、P2P行为、WebSocket、SSH、X11、MySql、Oracle、SSL、HTTP_POST及ftp下载。通过将这些访问类型取值离散向量化，作为特征进行学习，本发明应用示例，模型所输出的异常额外地考虑了访问类型特征，能够更精确地识别异常行为，作为辅助判断依据。

[0095] 本领域普通技术人员可以理解，上文中所公开方法中的全部或某些步骤、系统、装置中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。在硬件实施方式中，在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分；例如，一个物理组件可以具有多个功能，或者一个功能或步骤可以由若干物理组件合作执行。某些组件或所有组件可以被实施为由处理器，如数字信号处理器或微处理器执行的软件，或者被实施为硬件，或者被实施为集成电路，如专用集成电路。这样的软件可以分布在计算机可读介质上，计算机可读介质可以包括计算机存储介质(或非暂时性介质)和通信介质(或暂时性介质)。如本领域普通技术人员公知的，术语计算机存储介质包括在用于存储信息(诸如计算机可读指令、数据结构、程序模块或其他数据)的任何方法或技术中实施的易失性和非易失性、可移除和不可移除介质。计算机存储介质包括但不限于RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光盘存储、磁盒、磁带、磁盘存储或其他

磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。此外，本领域普通技术人员公知的是，通信介质通常包含计算机可读指令、数据结构、程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据，并且可包括任何信息递送介质。

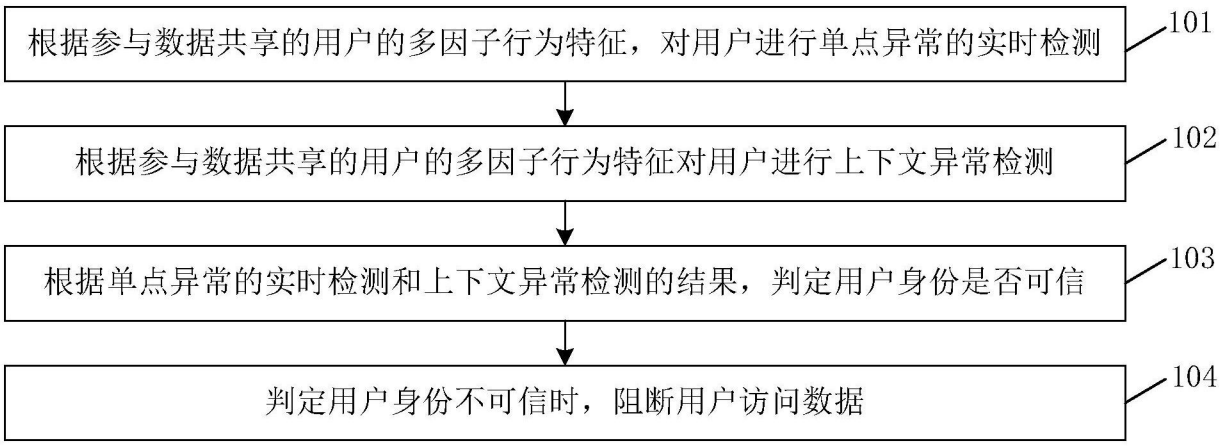

根据参与数据共享的用户的多因子行为特征，对用户进行单点异常的实时检测
101
根据参与数据共享的用户的多因子行为特征对用户进行上下文异常检测
102
根据单点异常的实时检测和上下文异常检测的结果，判定用户身份是否可信
103
判定用户身份不可信时，阻断用户访问数据
104

图1

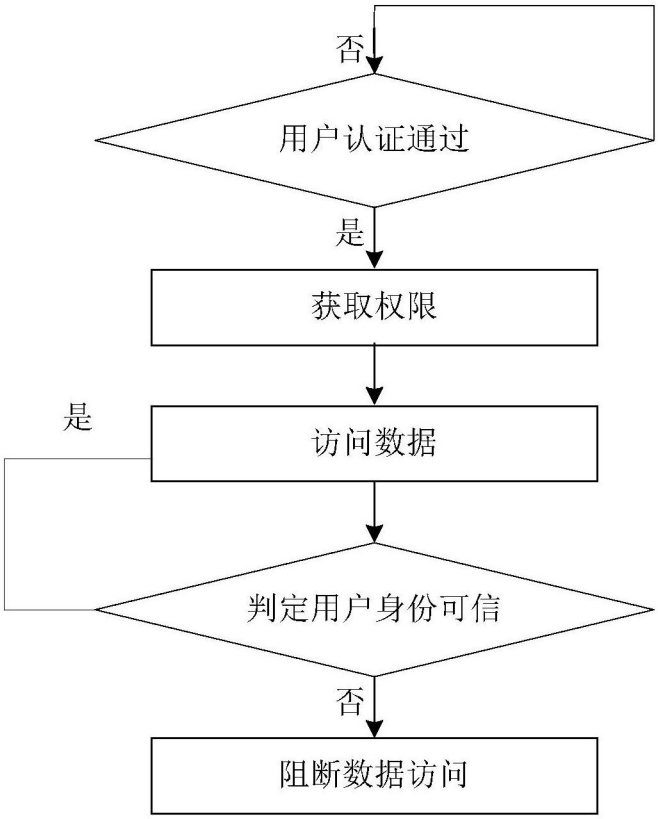

否
用户认证通过
是
获取权限
是
访问数据
判定用户身份可信
否
阻断数据访问

图2

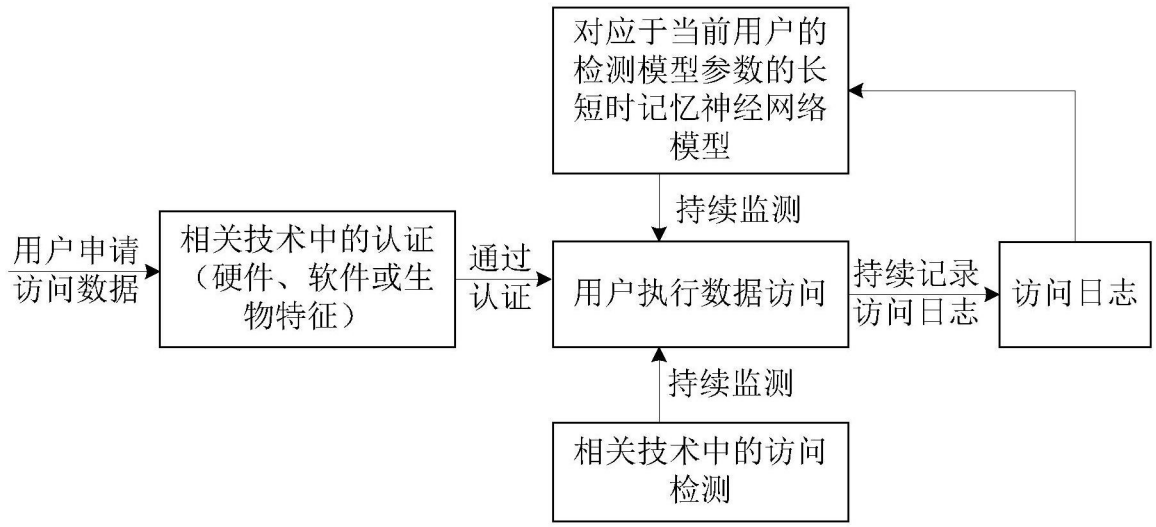

对应于当前用户的检测模型参数的长短时记忆神经网络模型
持续监测
用户申请访问数据
相关技术中的认证（硬件、软件或生物特征）
通过认证
用户执行数据访问
持续记录访问日志
访问日志
持续监测
相关技术中的访问检测

图3