(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115189963 A
(43)申请公布日 2022.10.14

(21)申请号 202210921799.2

(22)申请日 2022.08.02

(71)申请人 杭州安恒信息技术股份有限公司
地址 310051 浙江省杭州市滨江区西兴街道联慧街188号

(72)发明人 黄章镕 刘博

(74)专利代理机构 杭州华进联浙知识产权代理有限公司 33250
专利代理师 黄文勇

(51)Int.Cl.
H04L 9/40(2022.01)
G06K 9/62(2022.01)

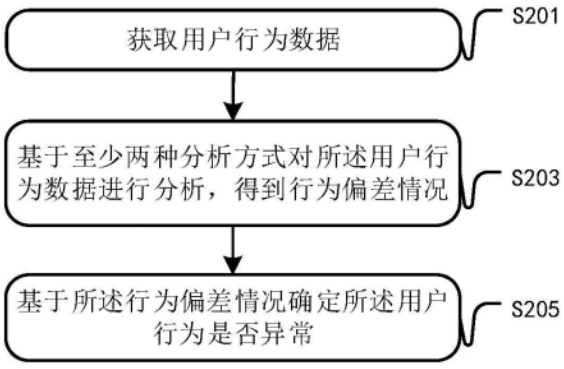
权利要求书2页 说明书12页 附图2页

(54)发明名称

异常行为检测方法、装置、计算机设备及可读存储介质

(57)摘要

本申请涉及一种异常行为检测方法、装置、计算机设备及可读存储介质。所述方法包括：获取用户行为数据；基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况；基于所述行为偏差情况确定所述用户行为是否异常。采用本方法能够从多个维度检测出用户行为存在的偏差，针对用户的疑似异常行为有着更高的反应灵敏度，能够有效应对更加复杂的网络环境和安全威胁场景，提高异常行为检测的准确率并降低误报率。



CN 115189963 A

1.一种异常行为检测方法,其特征在于,所述方法包括:

获取用户行为数据;

基于至少两种分析方式对所述用户行为数据进行分析,得到行为偏差情况;

基于所述行为偏差情况确定所述用户行为是否异常。

2.根据权利要求1所述的方法,其特征在于,所述基于至少两种分析方式对所述用户行为数据进行分析,得到行为偏差情况包括:

基于所述用户行为数据确定用户行为标签;

基于所述用户行为标签与预设条件的比较结果得到所述行为偏差情况。

3.根据权利要求1所述的方法,其特征在于,所述基于至少两种分析方式对所述用户行为数据进行分析,得到行为偏差情况包括:

基于所述用户行为数据获取预设时间内的用户行为特征值,所述用户行为特征值包括用户重复执行某一相同行为的次数;

基于所述用户行为特征值与预设阈值的比较结果得到所述行为偏差情况。

4.根据权利要求1所述的方法,其特征在于,所述基于至少两种分析方式对所述用户行为数据进行分析,得到行为偏差情况包括:

将用户历史行为数据输入至行为基线检测模型中,得到用户行为基线,所述行为基线检测模型通过机器学习训练得到;

基于所述用户行为数据与所述用户行为基线的比较结果得到所述行为偏差情况。

5.根据权利要求1所述的方法,其特征在于,所述基于所述行为偏差情况确定所述用户行为是否异常包括:

基于每种所述行为偏差情况以及对应的权重进行加权计算,得到权重积分;

基于所述权重积分确定所述用户行为是否异常。

6.根据权利要求1所述的方法,其特征在于,所述基于所述行为偏差情况确定所述用户行为是否异常还包括:

基于所述行为偏差情况生成偏差序列数据;

基于所述偏差序列数据生成网络安全矩阵模型标签序列;

将所述标签序列输入至异常检测模型,并输出异常标签序列,所述异常检测模型通过机器学习训练得到;

基于所述异常标签序列确定所述用户行为是否异常。

7.根据权利要求1所述的方法,其特征在于,所述基于所述行为偏差情况确定所述用户行为是否异常还包括:

基于每种所述行为偏差情况以及对应的权重进行加权计算,得到权重积分;

基于所述行为偏差情况生成偏差序列数据;

基于所述偏差序列数据生成网络安全矩阵模型标签序列;

将所述标签序列输入至异常检测模型,并输出异常标签序列,所述异常检测模型通过机器学习训练得到;

基于所述权重积分和所述异常标签序列确定所述用户行为是否异常。

8.一种异常行为检测装置,其特征在于,所述装置包括:

数据获取模块,用于获取用户行为数据;

数据分析模块，用于基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况；

异常确定模块，用于基于所述行为偏差情况确定所述用户行为是否异常。

9.一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述的方法的步骤。

10.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法的步骤。

异常行为检测方法、装置、计算机设备及可读存储介质

技术领域

[0001] 本申请涉及网络安全技术领域，特别是涉及一种异常行为检测方法、装置、计算机设备及可读存储介质。

背景技术

[0002] 随着信息技术的快速发展和集群规模的不断扩大，产生了大量的日志数据。日志数据记录了系统的运行信息，而且，用户的网络行为越来越多样化，当用户在系统上进行操作时，也会产生大量的行为日志。因此对网络安全来说，基于行为日志对用户行为进行识别、判断发现异常行为事件是尤为重要的。

[0003] 现有技术中针对用户行为分析的方法或装置，普遍是对用户行为日志中的数据进行解析来对用户的异常行为进行分析。但是，目前检测异常行为的方法中仅是针对某个单一维度的用户行为进行分析，因此对于较为复杂的安全威胁场景中异常行为的检测效果较弱，检测的准确率也相对较低。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种能够结合多种异常检测方式的异常行为检测方法、装置、计算机设备及可读存储介质。

[0005] 第一方面，本申请提供了一种异常行为检测方法，所述方法包括：

[0006] 获取用户行为数据；

[0007] 基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况；

[0008] 基于所述行为偏差情况确定所述用户行为是否异常。

[0009] 在其中一个实施例中，所述基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况包括：

[0010] 基于所述用户行为数据确定用户行为标签；

[0011] 基于所述用户行为标签与预设条件的比较结果得到所述行为偏差情况。

[0012] 在其中一个实施例中，所述基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况包括：

[0013] 基于所述用户行为数据获取预设时间内的用户行为特征值，所述用户行为特征值包括用户重复执行某一相同行为的次数；

[0014] 基于所述用户行为特征值与预设阈值的比较结果得到所述行为偏差情况。

[0015] 在其中一个实施例中，所述基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况包括：

[0016] 将用户历史行为数据输入至行为基线检测模型中，得到用户行为基线，所述行为基线检测模型通过机器学习训练得到；

[0017] 基于所述用户行为数据与所述用户行为基线的比较结果得到所述行为偏差情况。

[0018] 在其中一个实施例中，所述基于所述行为偏差情况确定所述用户行为是否异常包

括：

[0019] 基于每种所述行为偏差情况以及对应的权重进行加权计算，得到权重积分；

[0020] 基于所述权重积分确定所述用户行为是否异常。

[0021] 在其中一个实施例中，所述基于所述行为偏差情况确定所述用户行为是否异常还包括：

[0022] 基于所述行为偏差情况生成偏差序列数据；

[0023] 基于所述偏差序列数据生成网络安全矩阵模型标签序列；

[0024] 将所述标签序列输入至异常检测模型，并输出异常标签序列，所述异常检测模型通过机器学习训练得到；

[0025] 基于所述异常标签序列确定所述用户行为是否异常。

[0026] 在其中一个实施例中，所述基于所述行为偏差情况确定所述用户行为是否异常还包括：

[0027] 基于每种所述行为偏差情况以及对应的权重进行加权计算，得到权重积分；

[0028] 基于所述行为偏差情况生成偏差序列数据；

[0029] 基于所述偏差序列数据生成网络安全矩阵模型标签序列；

[0030] 将所述标签序列输入至异常检测模型，并输出异常标签序列，所述异常检测模型通过机器学习训练得到；

[0031] 基于所述权重积分和所述异常标签序列确定所述用户行为是否异常。

[0032] 第二方面，本申请提供了一种异常行为检测装置，所述装置包括：

[0033] 数据获取模块，用于获取用户行为数据；

[0034] 数据分析模块，用于基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况；

[0035] 异常确定模块，用于基于所述行为偏差情况确定所述用户行为是否异常。

[0036] 第三方面，本申请提供了一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现上述第一方面中任一项方法的步骤。

[0037] 第四方面，本申请提供了一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述第一方面中任一项方法的步骤。

[0038] 上述异常行为检测方法、装置、计算机设备及可读存储介质，通过获取用户行为数据；基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况；基于所述行为偏差情况确定所述用户行为是否异常。使用至少两种分析方式对用户行为进行分析，可以从多个维度检测出用户行为存在的偏差，针对用户的疑似异常行为有着更高的反应灵敏度，能够有效应对更加复杂的网络环境和安全威胁场景，提高异常行为检测的准确率并降低误报率。

[0039] 本申请的一个或多个实施例的细节在以下附图和描述中提出，以使本申请的其他特征、目的和优点更加简明易懂。

附图说明

[0040] 此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申

请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

[0041] 图1为一个实施例中异常行为检测方法的应用环境图；

[0042] 图2为一个实施例中异常行为检测方法的流程示意图；

[0043] 图3为一个实施例中异常行为检测装置的结构框图。

具体实施方式

[0044] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0045] 除另作定义外，本申请所涉及的技术术语或者科学术语应具有本申请所属技术领域具备一般技能的人所理解的一般含义。在本申请中的“一”、“一个”、“一种”、“该”、“这些”等类似的词并不表示数量上的限制，它们可以是单数或者复数。在本申请中所涉及的术语“包括”、“包含”、“具有”及其任何变体，其目的是涵盖不排他的包含；例如，包含一系列步骤或模块(单元)的过程、方法和系统、产品或设备并未限定于列出的步骤或模块(单元)，而可包括未列出的步骤或模块(单元)，或者可包括这些过程、方法、产品或设备固有的其他步骤或模块(单元)。在本申请中所涉及的“连接”、“相连”、“耦接”等类似的词语并不限定于物理的或机械连接，而可以包括电气连接，无论是直接连接还是间接连接。在本申请中所涉及的“多个”是指两个或两个以上。“和/或”描述关联对象的关联关系，表示可以存在三种关系，例如，“A和/或B”可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。通常情况下，字符“/”表示前后关联的对象是一种“或”的关系。在本申请中所涉及的术语“第一”、“第二”、“第三”等，只是对相似对象进行区分，并不代表针对对象的特定排序。

[0046] 以下所使用的术语“模块”、“单元”等为可以实现预定功能的软件和/或硬件的组合。尽管在以下实施例中所描述的装置较佳地以硬件来实现，但是软件，或者软件和硬件的组合的实现也是可能并被构想的。

[0047] 本申请实施例提供的异常行为检测方法，可以应用于如图1所示的应用环境中。其中，终端102通过网络与服务器104进行通信。数据存储系统可以存储服务器104需要处理的数据。数据存储系统可以集成在服务器104上，也可以放在云上或其他网络服务器上。本申请实施例中，可以由终端102获取用户行为数据，也可以是由服务器104获取用户行为数据。在其他实施例中，还可以是由终端102或服务器104任意一端获取用户行为数据之后发送至另外一端。基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况。所述两种分析方式可以由终端102或服务器104任意一端执行，也可以由其中一端执行至少一种分析方式，另一端执行其他分析方式，本申请对终端102和服务器104各自执行哪种和多少种分析方式不作限制。得到行为偏差情况后，可以由终端102或服务器104任意一端基于所述行为偏差情况确定所述用户行为是否异常。当确定所述用户行为异常时，可以由终端102或服务器104任意一端生成异常报告信息或对异常行为用户进行封禁、拉黑处理。其中，终端102可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑、物联网设备和便携式可穿戴设备，物联网设备可为智能音箱、智能电视、智能空调、智能车载设备等。便携式可穿戴设备可为智能手表、智能手环、头戴设备等。服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

[0048] 在现有技术中，通常一些组织内部会为内网搭建AD域(AD全称为Active Directory(活动目录)，通过在域内安装AD来实现域内资源的集中、统一管理)，AD域中可以包括域控设备和内网设备，域控设备可以为域控主机或域控服务器，内网设备可以为内网主机或内网服务器。通过域控设备对内网设备进行集中式管理。现有技术中针对AD域的威胁检测采用的方式主要是：基于日志和流量审计并结合预设规则来判断AD域或AD域帐号是否存在威胁，或通过监控蜜罐帐号的活动结合预设规则来判断AD域或AD域帐号是否存在威胁。这样的方法或装置需要针对特定的威胁场景，比如Kerberoasting(一种域口令攻击方法)、DCShadow(针对AD基础架构的一种攻击技术)或者流量加密降级等场景设定专门的规则，需要大量的专家知识，并缺乏灵活性。另外，目前针对AD域攻击的用户异常行为分析仅局限在浅层的异常行为，例如删除动作次数、登录时间、登录地址、访问对象、违规查询敏感信息等，因此，目前针对AD域安全威胁场景的用户行为分析技术中暂时不具备检测较复杂的安全威胁场景的能力。

[0049] 基于此，本申请实施例中，如图2所示，提供了一种异常行为检测方法，本实施例以该方法应用于终端102进行举例说明，可以理解的是，该方法也可以应用于服务器104，还可以应用于包括终端102和服务器104的系统，并通过终端102和服务器104的交互实现。本实施例中，该方法包括以下步骤：

[0050] S201：获取用户行为数据。

[0051] 本申请实施例中，为了检测用户行为是否存在异常，需要通过获取用户行为数据并对所述用户数据进行分析。所述用户行为数据包括目标用户数据源日志、用户实体数据。所述用户行为数据还可以包括用户流量数据。获取用户行为数据包括获取AD域内的用户数据、获取用户行为行为日志数据。获取用户行为数据还可以包括获取内网设备数据和获取组织单位数据中的至少一种。在本申请实施例中，获取的用户行为数据中可能存在数据类型、数据格式并不统一、规范的数据，为提高用户数据分析的效率，获取用户行为数据还可以包括获取标准化用户行为数据。所述获取标准化用户行为数据包括对所述用户行为数据进行数据类型转换、字段名映射转换后获取标准化用户行为数据。举例说明，获取的用户行为数据中可能包括如“2022-05-05 11:11:11”的时间字段，所述时间字段为String类型，需要将数据类型转换为DateTime类型。在一些实施例中，为了提高搜索目标用户数据，在将所述用户行为数据标准化之后，还包括储存所述用户行为数据，以及为所述用户行为数据建立索引。具体而言，可以将标准化之后的用户行为数据储存在ClickHouse(列式存储数据库)中，并为所述用户行为数据中的时间字段(startTime)和帐户字段(srcUserName)建立索引。

[0052] S203：基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况。

[0053] 通常情况下，用户的行为虽然较为复杂多变，但是在用户与数据库、服务器的交互过程中产生的行为数据可以一定程度上反映出用户的行为规律。对于AD域等内网的用户行为而言则具有更加明显和易捕捉的规律。基于此，对用户的行为数据进行分析时可以从多个维度进行综合分析，如果只进行某一个维度的用户行为数据分析则难以全面、准确地获取用户的行为规律，也可能因某一维度的分析方式存在的误差导致用户异常行为检测的准确率下降。本申请实施例中，基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况。所述至少两种分析方式可以包括用户行为基线分析、用户行为画像分析、用户

行为标签分析、用户异常行为权重分析、机器学习检测模型分析中的至少两种分析方式。所述行为偏差情况可以包括用户行为数据与预设值的偏差情况，也可以包括当前用户数据与历史用户数据的偏差情况。当然，所述偏差情况还可以包括基于用户行为数据与预设值和历史用户数据偏差情况确定的最终偏差情况。

[0054] S205：基于所述行为偏差情况确定所述用户行为是否异常。

[0055] 本申请实施例中，当步骤S203中基于至少两种分析方式对用户行为数据进行分析后，基于每一种分析方式都可以确定出相应的用户行为偏差情况。所述偏差情况可以包括用户行为是否存在偏差、用户行为偏差程度。所述用户行为偏差程度可以包括用户数据与预设数据或用户历史数据的偏差值。本申请实施例中，所述基于所述行为偏差情况确定所述用户行为是否异常可以包括，对所有分析方式确定的用户行为偏差情况赋予权重并进行加权计算得到权重积分，如果所述权重积分达到预设阈值则确定所述用户行为异常。在其他实施例中，也可以将行为偏差情况输入至训练好的检测模型中，输出用户行为异常检测结果。其他实施例中，还可以使用机器学习方式获得用户行为异常预测模型，将所述用户行为偏差情况输入所述用户行为异常预测模型，输出用户行为异常预测结果，并基于所述行为异常预测结果确定所述用户行为是否异常。可以理解的，在其他一些实施例中，还可以包括基于所述权重积分和异常预测结果进行综合分析来确定用户行为是否异常。

[0056] 本申请实施例中，所述检测模型可以包括用于大数据处理的计算引擎，也可以是利用机器学习方式训练得到的模型组件，还可以是多种模型或计算引擎的结合，其中，所述机器学习方式可以包括异常检测算法模型。具体而言，所述异常检测算法模型可以包括三西格玛算法模型、孤立森林算法模型、自编码器算法模型等等，本申请在此不做限制。

[0057] 本申请实施例中提供的异常行为检测方法，通过获取用户行为数据；基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况；基于所述行为偏差情况确定所述用户行为是否异常。使用至少两种分析方式对用户行为进行分析，可以从多个维度检测出用户行为存在的偏差，针对用户的疑似异常行为有着更高的反应灵敏度，能够有效应对更加复杂的网络环境和安全威胁场景。另一方面，以至少两种分析方式对用户行为数据进行分析，也减小了单一分析方式可能存在的误差，提高了异常行为检测的准确率并降低误报率。

[0058] 另外值得一提的是，本申请提供的异常行为检测方法，对内网行为监测具有天然的优势，因为内网用户行为具有一定的规律性，遵守组织内部行为守则，当用户行为出现明显变化时，即可通过本申请实施例提供的异常行为检测方法快速、准确地检测出来并判断是否是异常行为，所以本申请提供的异常行为检测方法针对内网用户行为，尤其是AD域内的用户行为分析技术具有较好的灵活性和准确性。

[0059] 在异常的用户行为中，存在某些标签化的用户行为，对于这些标签化行为设置预设条件，并以所述预设条件为标准对用户行为数据进行分析，可以有效检测出用户的行为偏差情况。本申请实施例中，在步骤S203中，所述基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况包括：

[0060] S301：基于所述用户行为数据确定用户行为标签。

[0061] S303：基于所述用户行为标签与预设条件的比较结果得到所述行为偏差情况。

[0062] 本申请实施例中，所述基于所述用户行为数据确定用户行为标签包括根据预设检

测项目对用户行为数据进行检测，得到的检测结果即为用户行为数据标签。所述预设条件包括对应于所述预设检测项目设置的用户正常行为数据。具体的，所述预设检测项目可以包括用户登录时来源IP与上一次登录时的来源IP是否相同，对应于所述预设检测项目的预设条件为来源IP相同，则分析用户的行为数据确定当前登录来源IP与上一次的来源IP是否相同作为用户行为标签。在其他实施例中，所述预设检测项目也可以包括用户是否修改访问控制策略，对应的预设条件为未修改访问控制策略，基于分析用户的行为数据确定是否修改访问控制策略作为用户行为标签。所述预设检测项目还可以包括用户行为中是否存在RC4加密降级且用户实体操作系统是否为Windows7及以上，对应的预设条件为用户行为不存在RC4加密降级且用户实体操作系统不为Windows7及以上，基于分析用户的行为数据确定用户行为和操作系统是否满足所述预设检测项目作为用户行为标签。将所述用户行为标签与所述预设条件进行比较，将相同或不同的比较结果作为用户行为偏差情况。可以理解的，上述预设检测项目仅作为举例说明，并不限定本申请实施例中只包括上述预设检测项目，其他用于检测用户行为偏差的项目也可以作为预设检测项目用于确定用户行为标签以及预设条件。另一方面，在其他实施例中，基于所述用户行为标签与预设条件的比较结果得到所述行为偏差情况可以包括，基于所有比较结果得到所述行为偏差情况，也可以基于所有预设检测项目中相应的任意数量的比较结果作为所述行为偏差情况，本申请对此不作限制。

[0063] 本申请实施例中，通过确定用户行为标签，并基于所述用户行为标签与预设条件的比较结果得到行为偏差情况，可以快速、有效地得到用户的行为偏差情况，且设置预设条件可以有针对性地对某些偏差情况进行专项检测，使得到的用户偏差情况可以根据实际需要灵活调整，提高了异常行为检测的效率。

[0064] 在某些异常用户行为中，会出现用户短期内重复执行某一行为的情况，例如用户短期内频繁尝试登录等异常行为。对于此类型的用户行为，可以通过分析用户行为数据进行相应的检测。本申请实施例中，在步骤S203中，所述基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况包括：

[0065] S401：基于所述用户行为数据获取预设时间内的用户行为特征值，所述用户行为特征值包括用户重复执行某一相同行为的次数。

[0066] S403：基于所述用户行为特征值与预设阈值的比较结果得到所述行为偏差情况。

[0067] 本申请实施例中，对用户行为数据进行分析，统计预设时间段内用户行为数据中重复某一行为的次数作为用户行为特征值，将所述用户行为特征值与预设阈值进行比较得到比较结果作为行为偏差情况。具体的，所述用户行为特征值可以包括预设时间内用户账户尝试登录次数、预设时间内用户账户登录失败次数、预设时间内用户登录来源IP的个数。对于上述用户行为分别设置预设阈值，将所述用户行为特征值与所述预设阈值进行比较并将小于、等于或高于预设阈值作为用户行为偏差情况。在其他实施例中，可以理解的，也可以将等于和/或高于预设阈值的比较结果作为行为偏差情况。上述用户行为特征值仅作为举例说明，并不限定本申请实施例中只包括上述行为特征值，其他用于检测用户行为偏差的特征值也可以作为得到行为偏差情况的依据，本申请对此不作限制。

[0068] 本申请实施例中，通过对用户在预设时间内重复某一行为的次数作为行为特征值，并将所述特征值与预设阈值的比较结果作为行为偏差情况。对于短时间内高频次的用

户异常行为有着较好的检测效果，可以基于与预设阈值的比较结果快速、有效、直观地得到用户行为偏差情况。

[0069] 在用户异常行为分析中，还可以根据用户当前行为与历史行为的偏差情况分析用户是否存在异常行为。本申请实施例中，可以通过机器学习的方式检测用户历史行为并得出用户行为基线作为检测用户当前行为的依据，在步骤S203中，所述基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况包括：

[0070] S501：将用户历史行为数据输入至行为基线检测模型中，得到用户行为基线，所述行为基线检测模型通过机器学习训练得到。

[0071] S503：基于所述用户行为数据与所述用户行为基线的比较结果得到所述行为偏差情况。

[0072] 本申请实施例中，所述用户历史行为数据包括目标用户历史数据源日志、用户历史实体数据。在其中一些实施例中，所述用户历史行为数据还可以包括用户历史流量数据。使用行为基线检测模型可以对用户历史行为数据进行检测，并输出用户行为基线。所述用户行为基线用于表征用户在历史行为中的行为规律。所述行为基线检测模型可以使用现有技术中的非监督学习算法训练得到，可以包括关联算法、聚类算法、降维算法等。可以理解的，所述行为基线检测模型也可以是多种算法组合训练得到的模型，本申请对此不作限制。下面通过聚类算法和关联算法两个实施例进行举例说明。

[0073] 实施例一

[0074] 本申请实施例中，通过某用户日常工作产生的历史行为数据确定用户的访问时间基线。首先获取用户过去30天的登录时间和注销时间数据作为用户历史行为数据。将DateTime类型的数据进行转换，如登录时间数据为2022-01-0108:31:11，则按式(1)进行转换：

[0075] $8\times60\times60+31\times60+11=30671$ (1)

[0076] 将所有用户登录、注销的时间按式(1)的方式进行转换后可以得到以秒为单位的时间戳数据。利用K-means算法(聚类算法)将所述时间戳数据按照上班时间、下班时间分为2类，或按上午上班、午休下班、下午上班、下午下班分为4类。其中两个样本数据的距离可以选用绝对值距离计算方法得到，输出聚类质心作为该用户的访问时间基线。基于用户在某类时间的登录、注销时间数据与访问时间基线的差值作为用户行为偏差情况。

[0077] 实施例二

[0078] 本申请实施例中，通过某用户访问资源列表的历史行为数据确定用户的访问资源基线。受限获取用户过去30天内所有访问的目的IP列表数据作为用户历史行为数据。以天数为单位，将所述用户历史行为数据进行划分。采用Aprior算法(关联算法)挖掘该帐户访问资源的频繁项集。为了缩短计算时间，将频繁10项集作为输出结果，而不完全计算完所有频繁项集。最大频繁项集长度小于10的，以最大频繁项集作为输出结果。所述输出结果作为用户访问资源基线。基于用户当前访问的资源IP数据与访问资源基线的交集数量作为用户行为偏差情况。

[0079] 本申请实施例中，通过行为基线检测模型对用户历史行为数据进行分析检测，并得到用户行为基线，基于用户行为数据与所述用户行为基线的比较结果得到所述行为偏差情况，可以有效检测出用户当前行为与历史行为的偏差情况，对于用户偏离行为基线较大

的异常行为可以通过行为偏差情况准确地展示出来。

[0080] 本申请实施例中，在得到用户行为偏差情况之后，可以根据所述偏差情况确定用户行为是否存在异常。在步骤S205中，基于所述行为偏差情况确定所述用户行为是否异常包括：

[0081] S601：基于每种所述行为偏差情况以及对应的权重进行加权计算，得到权重积分。

[0082] S603：基于所述权重积分确定所述用户行为是否异常。

[0083] 本申请实施例中，每种用户行为偏差情况对应的权重用于表示该用户偏差情况对于用户行为异常判定的重要程度。对所述行为偏差情况赋予权重的方法可以包括层次分析法、模糊法、模糊层次分析法或专家评价法等，本申请对此不做限制。在某些实施例中，在所述基于每种所述行为偏差情况以及对应的权重进行加权计算，得到权重积分之前，还包括对每种所述行为偏差情况进行评分，所述评分用于衡量该用户行为数据与预设数据或历史数据的偏差程度。所述进行评分可以是专家打分，也可以是根据不同的用户行为偏差情况预设不同的评分标准自动评分。在其他实施例中，也可以不对所述偏差情况进行评分，此时则可以认为所有用户行为的偏差情况均为评分相同的非零数值。对所有用户行为偏差情况赋予权重之后，基于所有偏差情况的评分分数和权重进行加权计算并得到权重积分。所述基于所述权重积分确定所述用户行为是否异常，可以是当所述权重积分高于预设权重阈值时判定所述用户行为异常。在其他的实施例中，也可以是基于权重积分得到所述用户行为偏差情况的置信度，当所述置信度达到预设置信度阈值时判定所述用户行为异常。

[0084] 本申请实施例中，基于对不同的用户行为偏差情况赋予权重的方式，通过获取权重积分判断用户行为是否异常。对于不同的行为偏差情况来说，其客观上对判断用户行为是否异常的影响大小并不一定相同。本申请实施例可以通过赋予不同权重的方式，来体现出不同的用户行为偏差情况对判定结果的影响程度，使判定结果更加真实可靠。

[0085] 为了使步骤S203中对用户行为数据进行分析的过程更加直观，本申请实施例还提供一种基于用户数据生成用户画像的方法，在步骤S203中，所述基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况还包括：

[0086] S701：基于至少两种分析方式对应的所述用户行为数据生成用户行为画像；

[0087] S703：基于所述用户行为画像得到行为偏差情况。

[0088] 本申请实施例中，可以对用户行为数据进行统计分析并形成用户行为画像。示例性的，在某些实施例中，如果所述至少两种分析方式中包括上述实施例中基于用户行为标签、基于用户行为特征值、基于用户行为基线的分析方式，则生成的用户画像如表1所示：

[0089]

用户是否修改访问控制策略	Y
流量中是否存在 RC4 加密降级且域内实体操作系统为 Win7 及以上。	Y
用户登录失败次数	30
用户短时间内登录来源 IP 的个数	3
用户访问的时间段基线	8:00～17:00

[0090] 表1

[0091] 通过表1所示的用户行为画像可以直观地展示出用户各种不同的行为数据，再结合预设条件、预设阈值、用户行为基线即可得到行为偏差情况。在一些实施例中，可以对每一个维度的用户行为预设权重，权重的取值可以为0-1。例如，可以将“用户是否修改访问控制策略”该维度的用户行为权重设置为0.7。用户行为与用户画像对应的偏离程度可以被划分成高、中、低三个等级，对应的权重为0.9，0.6，0.3，再对所有维度的特征进行加权平均计算，得到权重积分，基于所述权重积分确定所述用户行为是否异常。

[0092] 本申请实施例中，通过生成用户行为画像，可以直观、有效地展示出用户在不同维度下的行为数据，可以提高对用户行为数据分析的效率，从整体上也提高了用户异常行为的检测效率。

[0093] 本申请实施例中，还提供一种在步骤S205中，基于所述行为偏差情况确定所述用户行为是否异常的方法，包括：

[0094] S801：基于所述行为偏差情况生成偏差序列数据。

[0095] S803：基于所述偏差序列数据生成网络安全矩阵模型标签序列。

[0096] S805：将所述标签序列输入至异常检测模型，并输出异常标签序列，所述异常检测模型通过机器学习训练得到。

[0097] S807：基于所述异常标签序列确定所述用户行为是否异常。

[0098] 本申请实施例中，可以基于行为偏差情况判断所述用户行为是否存在偏差。具体的，当所述用户行为数据满足预设条件、超出预设阈值或偏离用户行为基线，则认为所述用户行为数据对应的行为偏差情况存在偏差。所述基于所述行为偏差情况生成偏差序列数据包括根据时序顺序生成偏差序列数据。将所述偏差序列数据命中网络安全矩阵的模型标签并生成网络安全矩阵模型标签序列。所述网络安全矩阵可以包括ATT&CK模型、零信任模型或以其为基础迭代更新后的网络安全模型中的任意一个，本申请对此不作限制。将所述标签序列输入至异常检测模型，并输出异常标签序列，所述异常检测模型通过机器学习训练得到。在一些实施例中，所述异常检测模型可以包括孤立森林、鲁棒随机切割森林、自编码器或常见的监督学习算法和模型，用于训练所述检测模型的数据集包括用户行为历史数据，所述机器学习模型的训练方法为现有技术，此处不再赘述。基于所述异常检测模型输出的异常标签序列，可以根据是否满足预设异常条件来确定所述用户行为是否异常。例如，输出的异常标签序列为用户在多次登录失败后，在非常见的来源IP登录成功，并访问了不常见的资源后，进行特权登录操作，修改了访问控制策略，上述异常标签序列全部满足预设的异常检测条件时，确定用户行为异常。可以理解的，在一些实施例中，也可以将所述异常检测条件作为异常检测模型的训练条件之一，则异常检测模型可以直接输出所述用户行为是否异常的结果。

[0099] 本申请实施例中，通过使用机器学习算法可以高效、准确地分析用户不同行为的偏差情况，针对在某一个维度的用户异常行为并不明显的情况，而以人工的方式难以进行综合判断时，本申请实施例可以从多个维度利用机器学习的异常检测模型进行有效检测，对于复杂的攻击场景有着较高的检测效率和准确度。

[0100] 为了进一步提高异常行为检测的准确度，本申请实施例还提供一种综合考虑偏差情况确定用户行为是否异常的方法，在步骤S205中，基于所述行为偏差情况确定所述用户行为是否异常的方法，还包括：

[0101] S901:基于每种所述行为偏差情况以及对应的权重进行加权计算,得到权重积分。

[0102] S903:基于所述行为偏差情况生成偏差序列数据。

[0103] S905:基于所述偏差序列数据生成网络安全矩阵模型标签序列。

[0104] S907:将所述标签序列输入至异常检测模型,并输出异常标签序列,所述异常检测模型通过机器学习训练得到。

[0105] S909:基于所述权重积分和所述异常标签序列确定所述用户行为是否异常。

[0106] 本申请实施例中,步骤S901-步骤S907所述的方法可以参考上述步骤S601、S801-步骤S805的方法,此处不再赘述。在步骤S601得到权重积分以及步骤S907得到异常标签序列后,可以基于所述权重积分和所述异常标签序列确定所述用户行为是否异常。具体的,在一些实施例中,可以将权重积分大于或等于预设偏差阈值的偏差情况,使用步骤S901-步骤S907所述的方法进行检测并输出异常标签序列,将满足预设异常条件的异常标签序列对应的用户行为确定为异常。可以理解的,在另一些实施例中,也可以将满足预设异常条件的异常标签序列对应的用户行为数据,使用步骤S601-步骤S603所述方法进行权重积分计算,再最终确定用户行为是否异常。当然,还可以将同一用户行为对应的用户数据在步骤S601-步骤S603和步骤S901-步骤S907两种判断方式中所得到的结果进行综合分析,最终确定所述用户行为是否异常。所述综合分析可以包括再次使用机器学习算法模型、再次使用权重积分、专家打分或人工判断方法中的至少一种,本申请对此不作限制。

[0107] 本申请实施例中,通过对两种判断方法所得到的结果进行二次判断,使得能够检测的用户异常行为覆盖面更广,能够有效应对更加复杂的网络环境和安全威胁场景,提高异常行为检测的准确率并降低误报率

[0108] 应该理解的是,虽然如上所述的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,如上所述的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0109] 基于同样的发明构思,本申请实施例还提供了一种用于实现上述所涉及的异常行为检测方法的异常行为检测装置1100。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似,故下面所提供的一个或多个异常行为检测装置实施例中的具体限定可以参见上文中对于异常行为检测方法的限定,在此不再赘述。

[0110] 在一个实施例中,如图3所示,提供了一种异常行为检测装置1100,包括:数据获取模块1101、数据分析模块1102和异常确定模块1103,其中:

[0111] 数据获取模块1101,用于获取用户行为数据;

[0112] 数据分析模块1102,用于基于至少两种分析方式对所述用户行为数据进行分析,得到行为偏差情况;

[0113] 异常确定模块1103,用于基于所述行为偏差情况确定所述用户行为是否异常。

[0114] 在一个实施例中,所述数据分析模块1102还用于基于所述用户行为数据确定用户行为标签;基于所述用户行为标签与预设条件的比较结果得到所述行为偏差情况。

[0115] 在一个实施例中，所述数据分析模块1102还用于基于所述用户行为数据获取预设时间内的用户行为特征值，所述用户行为特征值包括用户重复执行某一相同行为的次数；基于所述用户行为特征值与预设阈值的比较结果得到所述行为偏差情况。

[0116] 在一个实施例中，所述数据分析模块1102还用于将用户历史行为数据输入至时序分析算法模型中，得到用户行为基线，所述时序分析算法模型通过机器学习训练得到；基于所述用户行为数据与所述用户行为基线的比较结果得到所述行为偏差情况。

[0117] 在一个实施例中，所述异常确定模块1103还用于基于每种所述行为偏差情况以及对应的权重进行加权计算，得到权重积分；基于所述权重积分确定所述用户行为是否异常。

[0118] 在一个实施例中，所述异常确定模块1103还用于基于所述行为偏差情况判断所述用户行为是否存在偏差；若存在偏差，则基于存在偏差的相应行为偏差情况生成偏差序列数据；对所述偏差序列数据赋予网络安全矩阵模型标签并生成网络安全矩阵模型标签序列；将所述标签序列输入至异常检测模型，并输出异常标签序列，所述异常检测模型通过机器学习训练得到；基于所述异常标签序列确定所述用户行为是否异常。

[0119] 在一个实施例中，所述异常确定模块1103还用于基于所述行为偏差情况判断所述用户行为是否存在偏差；若存在偏差，则基于存在偏差的相应行为偏差情况生成偏差序列数据；对所述偏差序列数据赋予网络安全矩阵模型标签并生成网络安全矩阵模型标签序列；将所述标签序列输入至异常检测模型，并输出异常标签序列，所述异常检测模型通过机器学习训练得到；基于所述权重积分和所述异常标签序列确定所述用户行为是否异常。

[0120] 上述异常行为检测装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

[0121] 在一个实施例中，提供了一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现上述任一项所述的异常行为检测方法的步骤。

[0122] 在一个实施例中，提供了一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现上述任一项所述的异常行为检测方法的步骤。

[0123] 需要说明的是，本申请所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等)，均为经用户授权或者经过各方充分授权的信息和数据。

[0124] 本领域普通技术人员可以理解实现上述实施例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，所述的计算机程序可存储于一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory，ROM)、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器(ReRAM)、磁变存储器(Magnetoresistive Random Access Memory，MRAM)、铁电存储器(Ferroelectric Random Access Memory，FRAM)、相变存储器(Phase Change Memory，PCM)、石墨烯存储器等。易失性存储器可包括随机存取存储器(Random Access Memory，

RAM)或外部高速缓冲存储器等。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等,不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等,不限于此。

[0125] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0126] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对本申请专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请的保护范围应以所附权利要求为准。

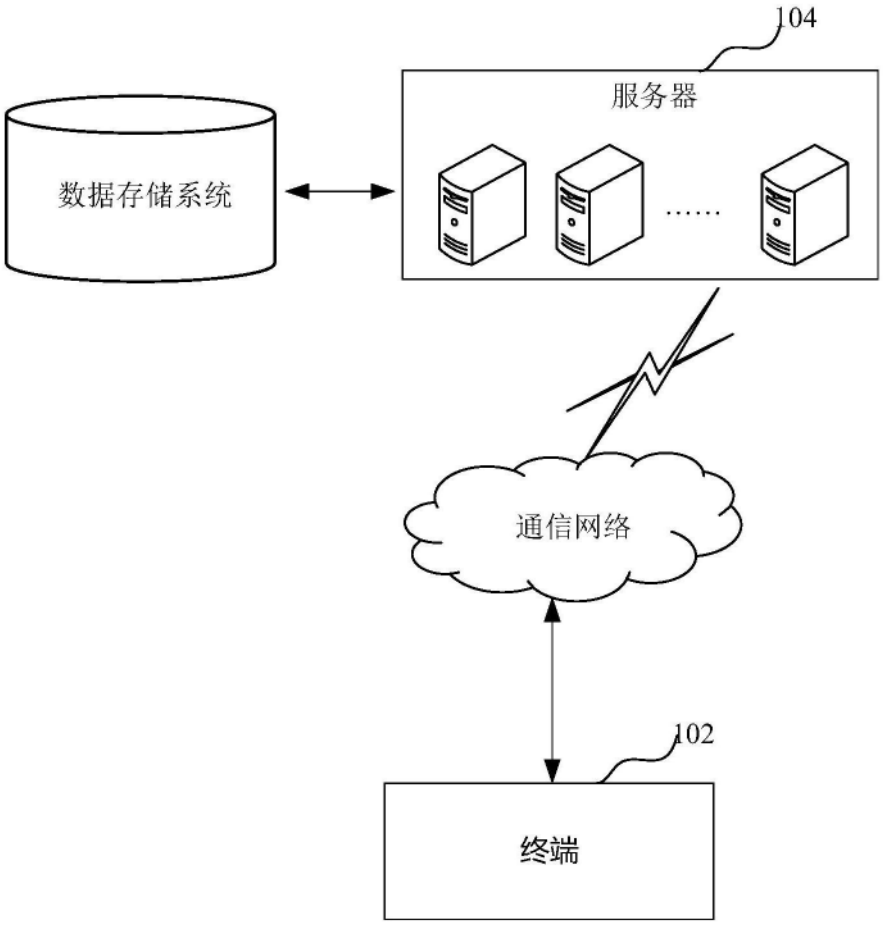

104
服务器
数据存储系统
.....
通信网络
102
终端

图1

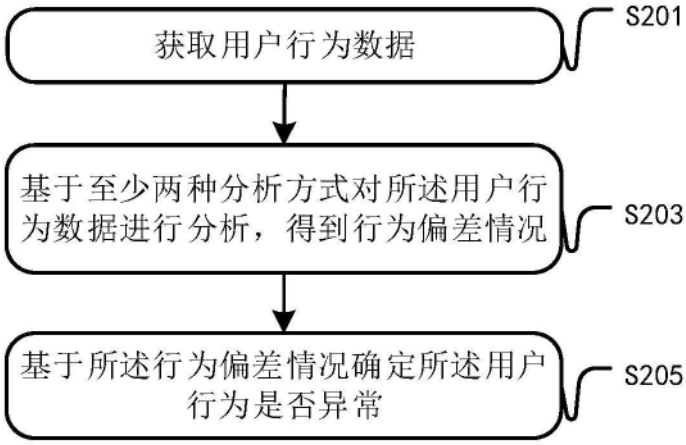

获取用户行为数据
S201
基于至少两种分析方式对所述用户行为数据进行分析，得到行为偏差情况
S203
基于所述行为偏差情况确定所述用户行为是否异常
S205

图2

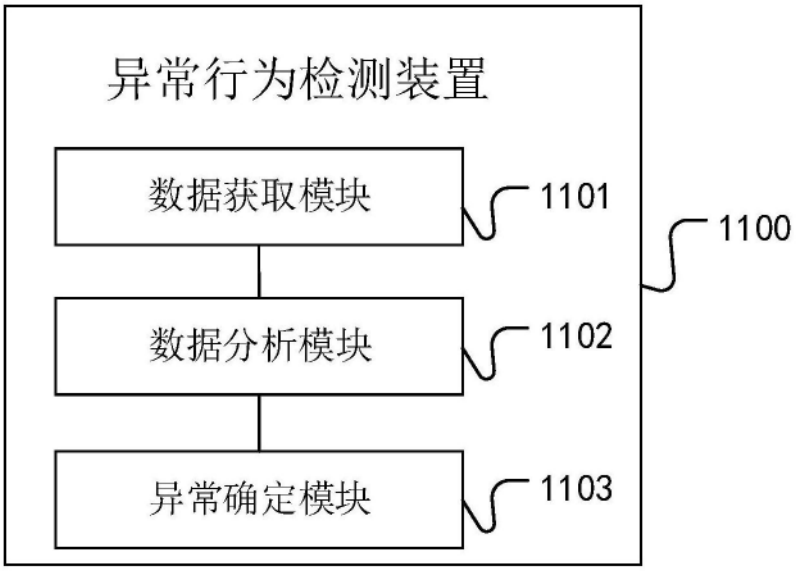


图3