

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 5 月 7 日 (07.05.2020)



(10) 国际公布号
WO 2020/087758 A1

(51) 国际专利分类号:
H04L 29/06 (2006.01)

(21) 国际申请号: PCT/CN2018/125709

(22) 国际申请日: 2018 年 12 月 29 日 (29.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811303640.4 2018 年 11 月 2 日 (02.11.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

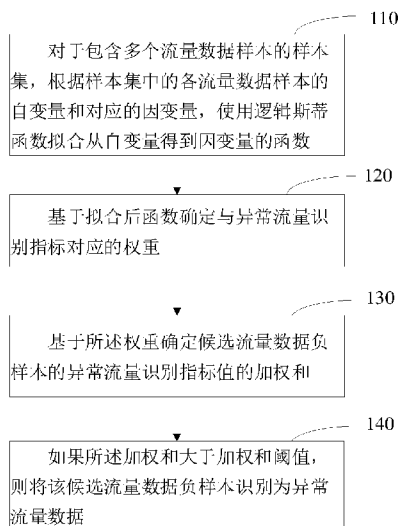
(72) 发明人: 孙家棣(SUN, Jiadi); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
马宁(MA, Ning); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市隆天联鼎知识产权代理有限公司(SHENZHEN LUNGTIN LIANDING INTELLECTUAL PROPERTY AGENT LTD.); 中国广东省深圳市福田区南园路上田大厦 4A 刘抗美, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,

(54) Title: ABNORMAL TRAFFIC DATA IDENTIFICATION METHOD, APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 异常流量数据识别方法、装置、计算机设备和存储介质



110 With regard to a set of samples including a plurality of traffic data samples, according to an independent variable and a corresponding dependent variable of each traffic data sample in the set of samples, use a logistic function to fit a function by which the dependent variable is obtained from the independent variable

120 Determine, on the basis of the fitted function, a weight corresponding to an abnormal traffic identification index

130 Determine, on the basis of the weight, a weighted sum of abnormal traffic identification index values of candidate traffic data negative samples

140 If the weighted sum is greater than a weighted sum threshold, identify the candidate traffic data negative samples as abnormal traffic data

图 1

(57) Abstract: An abnormal traffic data identification method, an apparatus, a computer device, and a storage medium. Said method comprises: with regard to a set of samples including a plurality of traffic data samples, according to an independent variable and a corresponding dependent variable of each traffic data sample in the set of samples, using a logistic function to fit a function by which the dependent variable is obtained from the independent variable (110); determining, on the basis of the fitted function, a weight corresponding to an abnormal traffic identification index (120); determining, on the basis of the weight, a weighted sum of abnormal

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

traffic identification index values of candidate traffic data negative samples (130), the candidate traffic data negative samples being samples which are obtained according to a predetermined rule and from which abnormal traffic data is generated; and if the weighted sum is greater than a weighted sum threshold, identifying the candidate traffic data negative samples as abnormal traffic data (140). This method sufficiently considers different importance of a plurality of abnormal traffic identification indexes during the identification of abnormal traffic data, improving the accuracy of abnormal traffic identification.

(57) 摘要: 一种异常流量数据识别方法、装置、计算机设备和存储介质。该方法包括: 对于包含多个流量数据样本的样本集, 根据样本集中的各流量数据样本的自变量和对应的因变量, 使用逻辑斯蒂函数拟合从自变量得到因变量的函数(110); 基于拟合后函数确定与异常流量识别指标对应的权重(120); 基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和(130), 所述候选流量数据负样本是根据预定规则得到的从中产生异常流量数据的样本; 如果所述加权和大于加权和阈值, 则将该候选流量数据负样本识别为异常流量数据(140)。此方法下, 充分考虑了多个异常流量识别指标在识别异常流量数据过程中重要性的不同, 提高了识别异常流量的准确性。

异常流量数据识别方法、装置、计算机设备和存储介质 技术领域

[0001] 本申请要求2018年11月02日递交、发明名称为“异常流量数据识别方法、装置、介质及电子设备”的中国专利申请CN 201811303640.4的优先权，在此通过引用将其全部内容合并于此。

[0002] 本申请涉及计算机通信技术领域，特别是涉及异常流量数据识别方法、装置、计算机设备和存储介质。

背景技术

[0003] 目前，薅羊毛、黑客攻击等事件的发生都会对互联网服务提供商的设施带来异常的流量。在现有技术中，对异常流量识别是基于用户行为埋点和SDK（software development kit）确定路径重复度、设备前后端埋点占比、IP访问次数等异常流量识别指标与相应的阈值的大小关系来实现的，这些异常流量识别指标的异常流量识别指标值中只要有一个大于与之相应的阈值，就认为该流量是异常的。本申请的发明人意识到，在现有技术，有些流量的某个异常流量识别指标值虽然超过了相应的阈值，但该流量数据是异常流量数据可能性低；有些流量的某个异常流量识别指标值虽然超过了相应的阈值，该流量数据是异常流量数据可能性高。因此每个异常流量识别指标在识别异常流量数据时的重要程度不同，现有技术在实现时无法准确评价每个异常流量识别指标在识别异常流量数据时的重要程度，从而不能基于多个异常流量识别指标综合对一个流量数据进行判断，识别异常流量数据的准确率低。

发明概述

技术问题

问题的解决方案

技术解决方案

[0004] 第一方面，提供了一种异常流量数据识别方法，包括：

[0005] 对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样本的自变

量 and 对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的函数，所述样本集中流量数据样本的多个异常流量识别指标值为自变量，该流量数据样本是否是异常流量数据为因变量；

[0006] 基于拟合后函数确定与异常流量识别指标对应的权重；

[0007] 基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和，所述候选流量数据负样本是根据预定规则得到的从中产生异常流量数据的样本；

[0008] 如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[0009] 第二方面，提供了一种异常流量数据识别装置，包括：

[0010] 拟合模块，被配置为对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样本的自变量和对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的函数；

[0011] 权重获取模块，被配置为基于拟合后函数确定与异常流量识别指标对应的权重；

[0012] 加权模块，被配置为基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和；

[0013] 判断模块，被配置为如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[0014] 第三方面，提供了一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行上述所述异常流量数据识别方法的步骤。

[0015] 第四方面，提供了一种存储有计算机可读指令的计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行上述所述异常流量数据识别方法的步骤。

[0016] 上述异常流量数据识别方法、装置、计算机设备和存储介质，充分并客观地考虑了多个异常流量识别指标在识别异常流量数据时重要性的不同，提高了识别异常流量的准确性。

发明的有益效果

对附图的简要说明

附图说明

- [0017] 图1是根据一示例性实施例示出的一种异常流量数据识别方法的流程图。
- [0018] 图2是根据图1对应实施例示出的一示例性实施例的获取候选流量数据负样本方法的流程图。
- [0019] 图3是根据一示例性实施例示出的获取净化后白名单流量数据的方法的流程图。
- [0020] 图4是根据一示例性实施例示出的所有异常流量识别指标值的上限参照值 x_{max} 获取方法的流程图。
- [0021] 图5是根据一示例性实施例示出的各个类型流量数据的包含关系的示意图。
- [0022] 图6是根据一示例性实施例示出的一种异常流量数据识别装置的框图。
- [0023] 图7是根据图6对应实施例示出的另一种异常流量数据识别装置的框图。
- [0024] 图8是根据图7对应实施例示出的另一种异常流量数据识别装置的框图。
- [0025] 图9是根据一示例性实施例示出的一种用于实现上述异常流量数据识别方法的计算机设备的框图。
- [0026] 图10是根据一示例性实施例示出的一种用于实现上述异常流量数据识别方法的计算机可读存储介质。

发明实施例

本发明的实施方式

- [0027] 本申请首先提供了一种异常流量数据识别的方法。本申请的实施终端可以是任何设备，实施终端可以与外部通过网络进行通信，其中可以包括软件、硬件或固件的结合。
- [0028] 图1是根据一示例性实施例示出的一种异常流量数据识别方法的流程图。如图1所示，此方法包括以下步骤：
- [0029] 步骤110，对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样本的自变量和对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的

函数。

[0030] 其中，样本集中流量数据样本的多个异常流量识别指标值为自变量，该流量数据样本是否是异常流量数据为因变量。在一个实施例中，如果流量数据样本为异常流量数据，该因变量值为1，如果流量数据样本不是异常流量数据，该因变量值为0。

[0031] 步骤120，基于拟合后函数确定与异常流量识别指标对应的权重。

[0032] 在一个实施例中，逻辑斯蒂函数为

$$f(x, \theta) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

，其中

$z = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}$, $x = (1, x^{(1)}, \dots, x^{(p)})^T$, $\theta = (\theta_0, \dots, \theta_p)^T$, $\theta_0, \dots, \theta_p$ 是各个异常流量识别指标的权重, $x = (1, x^{(1)}, \dots, x^{(p)})^T$ 是各个异常流量识别指标值。

在一个实施例中，设 $P_i = P(1|X_i; \theta) = \frac{f(X_i; \theta)}{1} = f(X_i; \theta)$, $P(1|X_i; \theta)$ 为条件概率，是指在第 i 个样本 x_i 的各特征值 $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ 和各特征的权重 $\theta_0, \dots, \theta_p$ 下，该样本是一个负样本的概率。相应地， $P(0|X_i; \theta) = 1 - P_i$ 。 $P(0|X_i; \theta)$ 表示第 i 个样本 x_i 的各特征值 $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ 和各特征的权重 $\theta_0, \dots, \theta_p$ 下，该样本是一个正样本的概率。当 $\prod_{i=1}^n [P_i^{y_i} (1 - P_i)^{1-y_i}]$ 值越大时，识别出样本是正负样本的结果与该样本是正负样本的真实属性越接近，说明异常流量的识别准确率越高。所以确定权重 $\theta_0, \dots, \theta_p$ 等同于确定在 $\theta_0, \dots, \theta_p$ 为何值时， $\prod_{i=1}^n [P_i^{y_i} (1 - P_i)^{1-y_i}]$ 具有最大值。

在一个实施例中， $Q(\hat{\theta}_0, \dots, \hat{\theta}_p)$ 为对 $\theta_0, \dots, \theta_p$ 估计的最佳值。即 $Q(\hat{\theta}_0, \dots, \hat{\theta}_p) = \max_{\theta_0, \dots, \theta_p} \{ \prod_{i=1}^n [P_i^{y_i} (1 - P_i)^{1-y_i}] \}$ 。

在一个实施例中，对 $Q(\hat{\theta}_0, \dots, \hat{\theta}_p) = \max_{\theta_0, \dots, \theta_p} \{ \prod_{i=1}^n [P_i^{y_i} (1 - P_i)^{1-y_i}] \}$ 取对数， $Q(\hat{\theta}_0, \dots, \hat{\theta}_p) = \max_{\theta_0, \dots, \theta_p} \{ \ln \prod_{i=1}^n [P_i^{y_i} (1 - P_i)^{1-y_i}] \}$ ，求取 $\ln \prod_{i=1}^n [P_i^{y_i} (1 - P_i)^{1-y_i}]$ 的最大值，最终使用梯度下降法或牛顿迭代法求得 $Q(\hat{\theta}_0, \dots, \hat{\theta}_p)$ 。

[0033] 步骤130，基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和。

[0034] 在一个实施例中，所述候选流量数据负样本是根据预定规则得到的从中产生异常流量数据的样本。

[0035] 步骤140，如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[0036] 在加权和大于加权和阈值的情况下，说明该候选流量数据负样本识别为异常流量数据，就将该异常流量数据识别出来。

[0037] 图2是根据图1对应实施例示出的一示例性实施例的获取候选流量数据负样本方法的流程图。如图2所示，此方法包括以下步骤：

[0038] 步骤210，获取所有用户的流量数据作为流量数据集合，所述流量数据集合包括净化后白名单流量数据和非净化后白名单流量数据。

[0039] 在一个实施例中，白名单流量数据是事先甄别出的、为异常流量数据的可能性低于第一阈值的流量数据，所述净化后白名单流量数据为事先甄别出的、为异常流量数据的可能性低于第二阈值的流量数据，第二阈值低于第一阈值，非净化后白名单流量数据是数据流量集合中所述净化后白名单流量数据以外的流量数据。在一个实施例中，根据经验将流量数据分为黑名单流量数据、白名单流量数据和不确定流量数据。黑名单流量数据是以前记录过的有异常流量行为的或者黑产从业者的流量数据。白名单流量数据是用户、销售人员、行政人员、基金用户的流量数据，这些流量数据是异常流量数据的可能性低。不确定流量数据是流量数据中除了黑名单流量数据和白名单流量数据之外的流量数据。因为白名单流量数据中还有是异常流量数据的可能，可以进一步甄别得到净化后白名单流量数据。这样做的好处在于，可以使最后获取的候选流量数据负样本更符合要求的。

[0040] 步骤220，从所述流量数据集合任取一异常流量识别指标。

[0041] 步骤230，获取与所述异常流量识别指标对应的异常流量识别指标阈值。

[0042] 异常流量识别指标阈值是预先设定好的，当异常流量识别指标值大于所述异常流量识别指标阈值时，就认为该异常流量识别指标值对应的流量数据样本为候选流量数据负样本。

[0043] 步骤240，计算净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与净化后白名单流量数据中流量数据数目的第一比值。

[0044] 步骤250，计算非净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非净化后白名单流量数据中流量数据数目的第二比值，以获取所述第一比值与第二比值之差的绝对值。

[0045] 步骤260，按照第一预定规则改变所述获取的异常流量识别指标阈值，直至所

述绝对值达到最大，对应的异常流量识别指标阈值为该异常流量识别指标对应的目标异常流量识别指标阈值。

[0046] 因为净化后白名单流量数据中候选异常流量数据占比很小，而非净化后白名单流量数据中候选异常流量数据占比很大，所以两者之差的绝对值越大，说明使用该异常流量识别指标阈值判断准确率越高。在一个示例性实施例中，在所述最大绝对值对应的异常流量识别指标阈值为两个或两个以上不同值的情况下，选择最小的第一比值对应的异常流量识别指标阈值为目标异常流量识别指标阈值。这样做的好处在于，最小的第一比值对应的第二比值也最小，比值之差的绝对值占第二比值的比例最大，更能说明该阈值对净化后白名单流量数据和非净化后白名单流量数据中候选异常流量的区分最好。在一个示例性实施例中，所述第一预定规则是将所述异常流量识别指标阈值增加预定步长，所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。

[0047] 步骤270，基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本。

[0048] 目标异常流量识别指标阈值是基于各异常流量识别指标值判断流量数据是否为异常流量数据的准确率最高的异常流量识别指标阈值，所以判断出的候选流量数据负样本最具有代表性。在一个实施例中，在流量数据的任意一个异常流量识别指标值大于对应的目标异常流量识别指标阈值的情况下，判断该流量数据为候选流量数据负样本。这样做的好处在于，能最大程度识别出候选流量数据负样本。

[0049] 图3是根据一示例性实施例示出的获取净化后白名单流量数据的方法的流程图。如图3所示，此方法包括以下步骤：

[0050] 步骤310，获取白名单流量数据集合。

[0051] 在一个实施例中，所述白名单流量数据集合包括纯白流量数据和非纯白流量数据，所述纯白流量数据是从所述白名单流量数据集合中事先甄别出的、为异常流量数据的可能性低于第三阈值的流量数据，所述第三阈值低于第二阈值，非纯白流量数据白名单流量数据集合中纯白流量数据之外的流量数据。

[0052] 步骤320，从所述白名单流量数据集合任取一异常流量识别指标。

[0053] 步骤330, 获取与所述任取的异常流量识别指标对应的异常流量识别指标阈值。

[0054] 步骤340, 计算纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与纯白流量数据中流量数据数目的第三比值。

[0055] 步骤350, 计算非纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非纯白流量数据中流量数据数目的第四比值, 以获取所述第三比值与第四比值之差的绝对值。

[0056] 步骤360, 按照第二预定规则改变所述异常流量识别指标阈值, 直至所述绝对值达到最大, 以得到目标异常流量识别指标阈值。

[0057] 在一个实施例中, 在所述最大绝对值对应的异常流量识别指标阈值为两个或两个以上不同值的情况下, 选择最小的第三比值对应的异常流量识别指标阈值为目标异常流量识别指标阈值。这样做的好处在于, 第三比值最小时对应的异常流量识别指标阈值对纯白流量数据和非纯白流量数据中异常流量数据的区分最好。在一个实施例中, 所述第二预定规则是将所述异常流量识别指标阈值增加预定步长, 所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。

[0058] 步骤370, 基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值, 从所述白名单流量数据集合中获取净化后白名单流量数据。

[0059] 在一个实施例中, 将所述白名单流量数据集合中只要有一个异常流量识别指标值大于对应的目标异常流量识别指标阈值的非纯白流量数据, 就从所述白名单流量数据集合中过滤掉, 得到净化后白名单流量数据。在一个实施例中, 将多个异常流量识别指标值按照如下公式归一化到[0,1]区间内:

$$\bar{x} = \begin{cases} \frac{x-x_{\min}}{x_{\max}-x_{\min}}, & x < x_{\max} \\ 1, & x \geq x_{\max} \end{cases}$$

[0060] 其中, x 是与任一待归一化异常流量识别指标的异常流量识别指标值,

$x_{\min}$

是样本集该待归一化异常流量识别指标的所有异常流量识别指标值的最小值,

x_{max}

是该待归一化异常流量识别指标的所有异常流量识别指标值的上限参照值，

$\bar{x}$

是与所述待归一化异常流量识别指标对应的归一化后的特征值。这样做的好处在于，避免了由于不同异常流量识别指标值量度和单位区别太大造成的可比性较低的问题。

[0061] 图4是根据一示例性实施例示出的所有异常流量识别指标值的上限参照值

x_{max}

获取方法的流程图。如图4所示，此方法包括以下步骤：

[0062] 步骤410，确定待归一化的异常流量识别指标。

[0063] 步骤420，获取所述样本集中的所有样本的与所述待归一化的异常流量识别指标对应的异常流量识别指标值。

[0064] 步骤430，对异常流量识别指标值按照递增的方式排序，以建成与所述待归一化的异常流量识别指标对应的异常流量识别指标值排序表。

[0065] 步骤440，计算99%与所述异常流量识别指标值排序表中异常流量识别指标值数目的乘积的数值。

[0066] 步骤450，在99%与所述异常流量识别指标值排序表中异常流量识别指标值数目的乘积的数值为整数的情况下，从异常流量识别指标值排序表中确定以该整数作为次序的异常流量识别指标值为所有异常流量识别指标值的上限参照值

x_{max}

。

[0067] 如果乘积为整数，在异常流量识别指标值排序表中就有与该整数对应的异常流量识别指标值。

[0068] 步骤460，在99%与所述异常流量识别指标值排序表中异常流量识别指标值数目的乘积的数值不为整数的情况下，从异常流量识别指标值排序表中确定以小于99%与所述异常流量识别指标值排序表中异常流量识别指标值数目的乘积的数值的最大整数作为次序的异常流量识别指标值为所有异常流量识别指标值的上

限参照值

x_{max}

。

[0069] 在99%与所述异常流量识别指标值排序表中异常流量识别指标值数目的乘积的数值不为整数的情况下，异常流量识别指标值排序表中就没有与该数值对应的异常流量识别指标值，因此可以选择小于99%与所述异常流量识别指标值排序表中异常流量识别指标值数目的乘积的数值的最大整数作为次序的异常流量识别指标值为所有异常流量识别指标值的上限参照值

x_{max}

。这样做的好处在于，由于选用极端的异常流量识别指标值作为所有异常流量识别指标值的上限参照值的会影响最后归一化的异常流量识别指标值，最终会影响是否是异常流量数据的判断结果，用这种方式选用所有异常流量识别指标值的上限参照值

x_{max}

，消除了极端的异常流量识别指标值对最终异常流量数据识别结果的干扰，提高了异常流量数据识别的准确率。

[0070] 图5是根据一示例性实施例示出的各个类型流量数据的包含关系的示意图。如图5所示，该示意图包括：

[0071] 501流量数据集合、502白名单流量数据、503 净化后白名单流量数据以及504 纯白流量数据。在图5中，为了便于说明本申请的各实施例，给出了各组元相对大小的示例。

[0072] 在一个实施例中，如前所述的异常流量识别指标包括以下14个。这14个异常流量识别指标包括：路径重复度排名、用户风控参数异常率、后端埋点比重、风控IP发散率、风控IP访问账号数、风控IP访问次数、风控IP_WIFI名个数、风控IP累计风险得分、风控IP周期内用户数均值、风控IP周期内用户方差、风控IP周期内访问次数均值、风控IP周期内访问次数方差、周期内手机号段用户登录均值和周期内手机号段用户登录方差。

[0073] 在一个实施例中，本申请还提供了一种异常流量数据识别装置，以下是本申请的装置实施例。

[0074] 图6是根据一示例性实施例示出的一种异常流量数据识别装置的框图。如图6所示，该异常流量数据识别装置包括：

[0075] 拟合模块610，被配置为对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样本的自变量和对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的函数；

[0076] 权重获取模块620，被配置为基于拟合后函数确定与异常流量识别指标对应的权重；

[0077] 加权模块630，被配置为基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和；

[0078] 判断模块640，被配置为如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[0079] 可选的，图7是根据图6对应实施例示出的另一种异常流量数据识别装置的框图。如图7所示，该异常流量数据识别装置还包括：

[0080] 第一流量数据获取模块710，被配置为获取所有用户的流量数据作为流量数据集合，所述流量数据集合包括净化后白名单流量数据和非净化后白名单流量数据，白名单流量数据是为事先甄别出的、为异常流量数据的可能性低于第一阈值的流量数据，所述净化后白名单流量数据为事先甄别出的、为异常流量数据的可能性低于第二阈值的流量数据，第二阈值低于第一阈值，非净化后白名单流量数据是数据流量集合中所述净化后白名单流量数据以外的流量数据；

[0081] 第一识别指标获取模块720，被配置为从所述流量数据集合任取一异常流量识别指标；

[0082] 第一阈值获取模块730，被配置为获取与所述异常流量识别指标对应的异常流量识别指标阈值；

[0083] 第一比值获取模块740，被配置为计算净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与净化后白名单流量数据中流量数据数目的第一比值；

- [0084] 第一绝对值获取模块750，被配置为计算非净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非净化后白名单流量数据中流量数据数目的第二比值，以获取所述第一比值与第二比值之差的绝对值；
- [0085] 第二阈值获取模块760，被配置为按照第一预定规则改变所述获取的异常流量识别指标阈值，直至所述绝对值达到最大，对应的异常流量识别指标阈值为该异常流量识别指标对应的目标异常流量识别指标阈值；
- [0086] 负样本判断模块770，被配置为基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本。
- [0087] 在一个实施例中，所述负样本判断模块770被进一步配置为：在流量数据的任意一个异常流量识别指标值大于对应的目标异常流量识别指标阈值的情况下，判断该流量数据为候选流量数据负样本。
- [0088] 在一个实施例中，所述第二阈值获取模块760被进一步配置为：在所述最大绝对值对应的异常流量识别指标阈值为两个或两个以上不同值的情况下，选择最小的第一比值对应的异常流量识别指标阈值为目标异常流量识别指标阈值。
- [0089] 在一个实施例中，所述第一预定规则包括：将所述异常流量识别指标阈值增加预定步长，所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。
- [0090] 可选的，图8是根据图7对应实施例示出的另一种异常流量数据识别装置的框图。该异常流量数据识别装置还包括：
- [0091] 第二流量数据获取模块810，被配置为获取白名单流量数据集合，所述白名单流量数据集合包括纯白流量数据和非纯白流量数据，所述纯白流量数据是从所述白名单流量数据集合中事先甄别出的、为异常流量数据的可能性低于第三阈值的流量数据，所述第三阈值低于第二阈值，非纯白流量数据是白名单流量数据集合中纯白流量数据之外的流量数据；
- [0092] 第二识别指标获取模块820，被配置为从所述白名单流量数据集合任取一异常流量识别指标；

- [0093] 第三阈值获取模块830，被配置为获取与所述任取的异常流量识别指标对应的异常流量识别指标阈值；
- [0094] 第三比值获取模块840，被配置为计算纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与纯白流量数据中流量数据数目的第三比值；
- [0095] 第二绝对值获取模块850，被配置为计算非纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非纯白流量数据中流量数据数目的第四比值，以获取所述第三比值与第四比值之差的绝对值；
- [0096] 第四阈值获取模块860，被配置为按照第二预定规则改变所述异常流量识别指标阈值，直至所述绝对值达到最大，以得到目标异常流量识别指标阈值；
- [0097] 第三流量数据获取模块870，被配置为基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，从所述白名单流量数据集合中获取净化后白名单流量数据。
- [0098] 在一个实施例中，所述第二预定规则包括：将所述异常流量识别指标阈值增加预定步长，所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。
- [0099] 在一个实施例中，提出了一种计算机设备，执行上述任一所示的异常流量数据识别方法的全部或者部分步骤。该计算机设备包括：至少一个处理器；以及与所述至少一个处理器通信连接的存储器；其中，所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行如上述任一示例性实施例所示出的异常流量数据识别方法。
- [0100] 所属技术领域的技术人员能够理解，本申请的各个方面可以实现为系统、方法或程序产品。因此，本申请的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式（包括固件、微代码等），或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。
- [0101] 下面参照图9来描述根据本申请的这种实施方式的计算机设备900。图9显示的计算机设备900仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任

何限制。

[0102] 如图9所示，计算机设备900以通用计算设备的形式表现。计算机设备900的组件可以包括但不限于：上述至少一个处理单元910、上述至少一个存储单元920、连接不同系统组件（包括存储单元920和处理单元910）的总线930。其中，所述存储单元存储有程序代码，所述程序代码可以被所述处理单元910执行，使得所述处理单元910执行本说明书上述“实施例方法”部分中描述的根据本申请各种示例性实施方式的步骤。存储单元920可以包括易失性存储单元形式的可读介质，例如随机存取存储单元（RAM）921和/或高速缓存存储单元922，还可以进一步包括只读存储单元（ROM）923。存储单元920还可以包括具有一组（至少一个）程序模块925的程序/实用工具924，这样的程序模块925包括但不限于：操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。总线930可以为表示几类总线结构中的一种或多种，包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。计算机设备900也可以与一个或多个外部设备1100（例如键盘、指向设备、蓝牙设备等）通信，还可与一个或者多个使得用户能与该计算机设备900交互的设备通信，和/或与使得该计算机设备900能与一个或多个其它计算设备进行通信的任何设备（例如路由器、调制解调器等等）通信。这种通信可以通过输入/输出（I/O）接口950进行。并且，计算机设备900还可以通过网络适配器960与一个或者多个网络（例如局域网（LAN），广域网（WAN）和/或公共网络，例如因特网）通信。如图所示，网络适配器960通过总线930与计算机设备900的其它模块通信。应当明白，尽管图中未示出，可以结合计算机设备900使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0103] 通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本申请实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是CD-ROM，U盘，移动硬盘等）

中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、终端装置、或者网络设备等）执行根据本申请实施方式的方法。

[0104] 在一个实施例中，提出了一种存储有计算机可读指令的存储介质，该计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行上述异常流量数据识别方法实施例中的步骤。

[0105] 图10是根据一示例性实施例示出的一种用于实现上述异常流量数据识别方法的计算机可读存储介质。如图10所示，包括计算机可读存储介质1000，其上可存储有计算机程序。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，该计算机程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，前述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）等非易失性存储介质，或随机存储记忆体（Random Access Memory, RAM）等。

[0106] 以上所述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。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，本申请专利的保护范围应以所附权利要求为准。

权利要求书

[权利要求 1]

一种异常流量数据识别方法，所述方法包括：

对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样本的自变量和对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的函数，所述样本集中流量数据样本的多个异常流量识别指标值为自变量，该流量数据样本是否是异常流量数据为因变量；

基于拟合后函数确定与异常流量识别指标对应的权重；

基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和，所述候选流量数据负样本是根据预定规则得到的从中产生异常流量数据的样本；

如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[权利要求 2]

根据权利要求1所述的方法，其中，所述候选流量数据负样本是根据如下步骤获得：

获取所有用户的流量数据作为流量数据集合，所述流量数据集合包括净化后白名单流量数据和非净化后白名单流量数据，白名单流量数据是为事先甄别出的、为异常流量数据的可能性低于第一阈值的流量数据，所述净化后白名单流量数据为事先甄别出的、为异常流量数据的可能性低于第二阈值的流量数据，第二阈值低于第一阈值，非净化后白名单流量数据是数据流量集合中所述净化后白名单流量数据以外的流量数据；

从所述流量数据集合任取一异常流量识别指标；

获取与所述异常流量识别指标对应的异常流量识别指标阈值；

计算净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与净化后白名单流量数据中流量数据数目的第一比值；

计算非净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非净化后白名单流量数据中流量数

据数目的第二比值，以获取所述第一比值与第二比值之差的绝对值；
按照第一预定规则改变所述获取的异常流量识别指标阈值，直至所述绝对值达到最大，对应的异常流量识别指标阈值为该异常流量识别指标对应的目标异常流量识别指标阈值；
基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本。

[权利要求 3] 根据权利要求2所述的方法，其中，所述基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本，具体包括：

在流量数据的任意一个异常流量识别指标值大于对应的目标异常流量识别指标阈值的情况下，判断该流量数据为候选流量数据负样本。

[权利要求 4] 根据权利要求2所述的方法，其中，所述改变所述异常流量识别指标阈值直至所述绝对值达到最大，对应的异常流量识别指标阈值为该异常流量识别指标对应的目标异常流量识别指标阈值，包括：

在所述最大绝对值对应的异常流量识别指标阈值为两个或两个以上不同值的情况下，选择最小的第一比值对应的异常流量识别指标阈值为目标异常流量识别指标阈值。

[权利要求 5] 根据权利要求2所述的方法，其中，所述第一预定规则包括：将所述异常流量识别指标阈值增加预定步长，所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。

[权利要求 6] 根据权利要求2所述的方法，其中，所述净化后白名单流量数据是根据如下步骤获得：

获取白名单流量数据集合，所述白名单流量数据集合包括纯白流量数据和非纯白流量数据，所述纯白流量数据是从所述白名单流量数据集合中事先甄别出的、为异常流量数据的可能性低于第三阈值的流量数据，所述第三阈值低于第二阈值，非纯白流量数据是白名单流量数据集合中纯白流量数据之外的流量数据；

从所述白名单流量数据集合任取一异常流量识别指标；

获取与所述任取的异常流量识别指标对应的异常流量识别指标阈值；
计算纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与纯白流量数据中流量数据数目的第三比值；
计算非纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非纯白流量数据中流量数据数目的第四比值，以获取所述第三比值与第四比值之差的绝对值；
按照第二预定规则改变所述异常流量识别指标阈值，直至所述绝对值达到最大，以得到目标异常流量识别指标阈值；
基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，从所述白名单流量数据集合中获取净化后白名单流量数据。

[权利要求 7] 根据权利要求6所述的方法，其中，所述第二预定规则包括：将所述异常流量识别指标阈值增加预定步长，所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。

[权利要求 8] 一种异常流量数据识别装置，所述装置包括：
拟合模块，被配置为对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样本的自变量和对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的函数；
权重获取模块，被配置为基于拟合后函数确定与异常流量识别指标对应的权重；
加权模块，被配置为基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和；
判断模块，被配置为如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[权利要求 9] 根据权利要求8所述的装置，其中，所述装置还包括：
第一流量数据获取模块，被配置为获取所有用户的流量数据作为流量数据集合，所述流量数据集合包括净化后白名单流量数据和非净化后白名单流量数据，白名单流量数据是为事先甄别出的、为异常流量数据的可能性低于第一阈值的流量数据，所述净化后白名单流量数据为

事先甄别出的、为异常流量数据的可能性低于第二阈值的流量数据，第二阈值低于第一阈值，非净化后白名单流量数据是数据流量集合中所述净化后白名单流量数据以外的流量数据；

第一识别指标获取模块，被配置为从所述流量数据集合任取一异常流量识别指标；

第一阈值获取模块，被配置为获取与所述异常流量识别指标对应的异常流量识别指标阈值；

第一比值获取模块，被配置为计算净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与净化后白名单流量数据中流量数据数目的第一比值；

第一绝对值获取模块，被配置为计算非净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非净化后白名单流量数据中流量数据数目的第二比值，以获取所述第一比值与第二比值之差的绝对值；

第二阈值获取模块，被配置为按照第一预定规则改变所述获取的异常流量识别指标阈值，直至所述绝对值达到最大，对应的异常流量识别指标阈值为该异常流量识别指标对应的目标异常流量识别指标阈值；

负样本判断模块，被配置为基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本。

[权利要求 10] 根据权利要求9所述的装置，其中，所述负样本判断模块被进一步配置为：在流量数据的任意一个异常流量识别指标值大于对应的目标异常流量识别指标阈值的情况下，判断该流量数据为候选流量数据负样本。

[权利要求 11] 根据权利要求9所述的装置，其中，所述第二阈值获取模块被进一步配置为：在所述最大绝对值对应的异常流量识别指标阈值为两个或两个以上不同值的情况下，选择最小的第一比值对应的异常流量识别指标阈值为目标异常流量识别指标阈值。

[权利要求 12]

根据权利要求9所述的装置，其中，所述装置还包括：

第二流量数据获取模块，被配置为获取白名单流量数据集合，所述白名单流量数据集合包括纯白流量数据和非纯白流量数据，所述纯白流量数据是从所述白名单流量数据集合中事先甄别出的、为异常流量数据的可能性低于第三阈值的流量数据，所述第三阈值低于第二阈值，非纯白流量数据是白名单流量数据集合中纯白流量数据之外的流量数据；

第二识别指标获取模块，被配置为从所述白名单流量数据集合任取一异常流量识别指标；

第三阈值获取模块，被配置为获取与所述任取的异常流量识别指标对应的异常流量识别指标阈值；

第三比值获取模块，被配置为计算纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与纯白流量数据中流量数据数目的第三比值；

第二绝对值获取模块，被配置为计算非纯白流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非纯白流量数据中流量数据数目的第四比值，以获取所述第三比值与第四比值之差的绝对值；

第四阈值获取模块，被配置为按照第二预定规则改变所述异常流量识别指标阈值，直至所述绝对值达到最大，以得到目标异常流量识别指标阈值；

第三流量数据获取模块，被配置为基于各异常流量识别指标值和各自对应的目标异常流量识别指标阈值，从所述白名单流量数据集合中获取净化后白名单流量数据。

[权利要求 13]

一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行：

对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样

本的自变量和对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的函数，所述样本集中流量数据样本的多个异常流量识别指标值为自变量，该流量数据样本是否是异常流量数据为因变量；

基于拟合后函数确定与异常流量识别指标对应的权重；

基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和，所述候选流量数据负样本是根据预定规则得到的从中产生异常流量数据的样本；

如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[权利要求 14]

根据权利要求13所述的计算机设备，其中，所述候选流量数据负样本是根据如下步骤获得：

获取所有用户的流量数据作为流量数据集合，所述流量数据集合包括净化后白名单流量数据和非净化后白名单流量数据，白名单流量数据是为事先甄别出的、为异常流量数据的可能性低于第一阈值的流量数据，所述净化后白名单流量数据为事先甄别出的、为异常流量数据的可能性低于第二阈值的流量数据，第二阈值低于第一阈值，非净化后白名单流量数据是数据流量集合中所述净化后白名单流量数据以外的流量数据；

从所述流量数据集合任取一异常流量识别指标；

获取与所述异常流量识别指标对应的异常流量识别指标阈值；

计算净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与净化后白名单流量数据中流量数据数目的第一比值；

计算非净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非净化后白名单流量数据中流量数据数目的第二比值，以获取所述第一比值与第二比值之差的绝对值；

按照第一预定规则改变所述获取的异常流量识别指标阈值，直至所述绝对值达到最大，对应的异常流量识别指标阈值为该异常流量识别指

标对应的目标异常流量识别指标阈值；

基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本。

[权利要求 15] 根据权利要求14所述的计算机设备，其中，所述基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本，具体包括：

在流量数据的任意一个异常流量识别指标值大于对应的目标异常流量识别指标阈值的情况下，判断该流量数据为候选流量数据负样本。

[权利要求 16] 根据权利要求14所述的计算机设备，其中，所述第一预定规则包括：将所述异常流量识别指标阈值增加预定步长，所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。

[权利要求 17] 一种存储有计算机可读指令的计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行：对于包含多个流量数据样本的样本集，根据样本集中的各流量数据样本的自变量和对应的因变量，使用逻辑斯蒂函数拟合从自变量得到因变量的函数，所述样本集中流量数据样本的多个异常流量识别指标值为自变量，该流量数据样本是否是异常流量数据为因变量；

基于拟合后函数确定与异常流量识别指标对应的权重；

基于所述权重确定候选流量数据负样本的异常流量识别指标值的加权和，所述候选流量数据负样本是根据预定规则得到的从中产生异常流量数据的样本；

如果所述加权和大于加权和阈值，则将该候选流量数据负样本识别为异常流量数据。

[权利要求 18] 根据权利要求17所述的计算机可读存储介质，其中，所述候选流量数据负样本是根据如下步骤获得：

获取所有用户的流量数据作为流量数据集合，所述流量数据集合包括净化后白名单流量数据和非净化后白名单流量数据，白名单流量数据是为事先甄别出的、为异常流量数据的可能性低于第一阈值的流量数

据，所述净化后白名单流量数据为事先甄别出的、为异常流量数据的可能性低于第二阈值的流量数据，第二阈值低于第一阈值，非净化后白名单流量数据是数据流量集合中所述净化后白名单流量数据以外的流量数据；

从所述流量数据集合任取一异常流量识别指标；

获取与所述异常流量识别指标对应的异常流量识别指标阈值；

计算净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与净化后白名单流量数据中流量数据数目的第一比值；

计算非净化后白名单流量数据中异常流量识别指标值大于所述异常流量识别指标阈值的流量数据数目与非净化后白名单流量数据中流量数据数目的第二比值，以获取所述第一比值与第二比值之差的绝对值；

按照第一预定规则改变所述获取的异常流量识别指标阈值，直至所述绝对值达到最大，对应的异常流量识别指标阈值为该异常流量识别指标对应的目标异常流量识别指标阈值；

基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本。

[权利要求 19] 根据权利要求18所述的计算机可读存储介质，其中，所述基于各异常流量识别指标值和各对应的目标异常流量识别指标阈值，判断非净化后白名单流量数据是否为候选流量数据负样本，具体包括：

在流量数据的任意一个异常流量识别指标值大于对应的目标异常流量识别指标阈值的情况下，判断该流量数据为候选流量数据负样本。

[权利要求 20] 根据权利要求18所述的计算机可读存储介质，其中，所述第一预定规则包括：将所述异常流量识别指标阈值增加预定步长，所述预定步长根据异常流量识别指标阈值的上限、下限、以及预设的步长数决定。

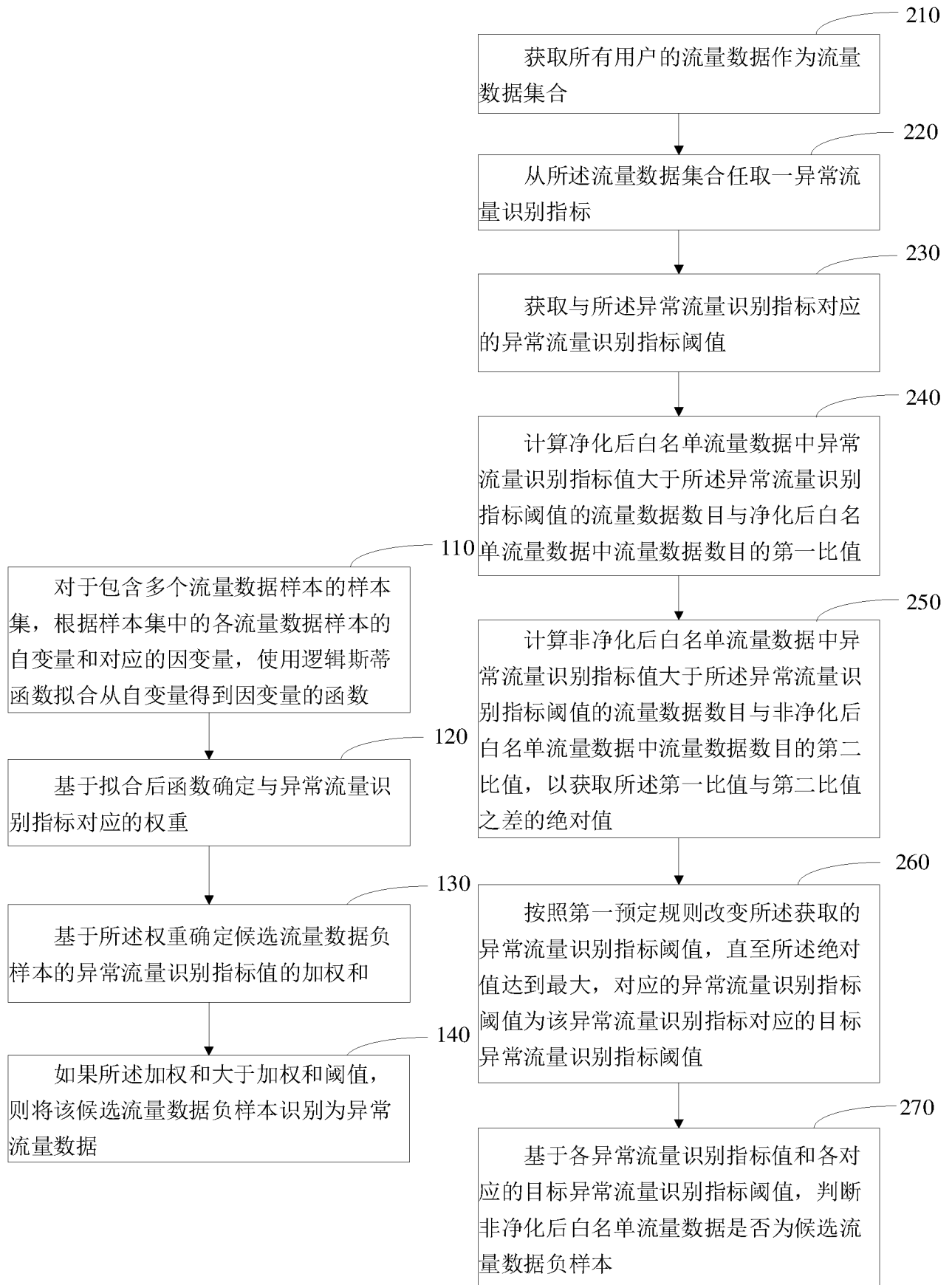


图 1

图 2

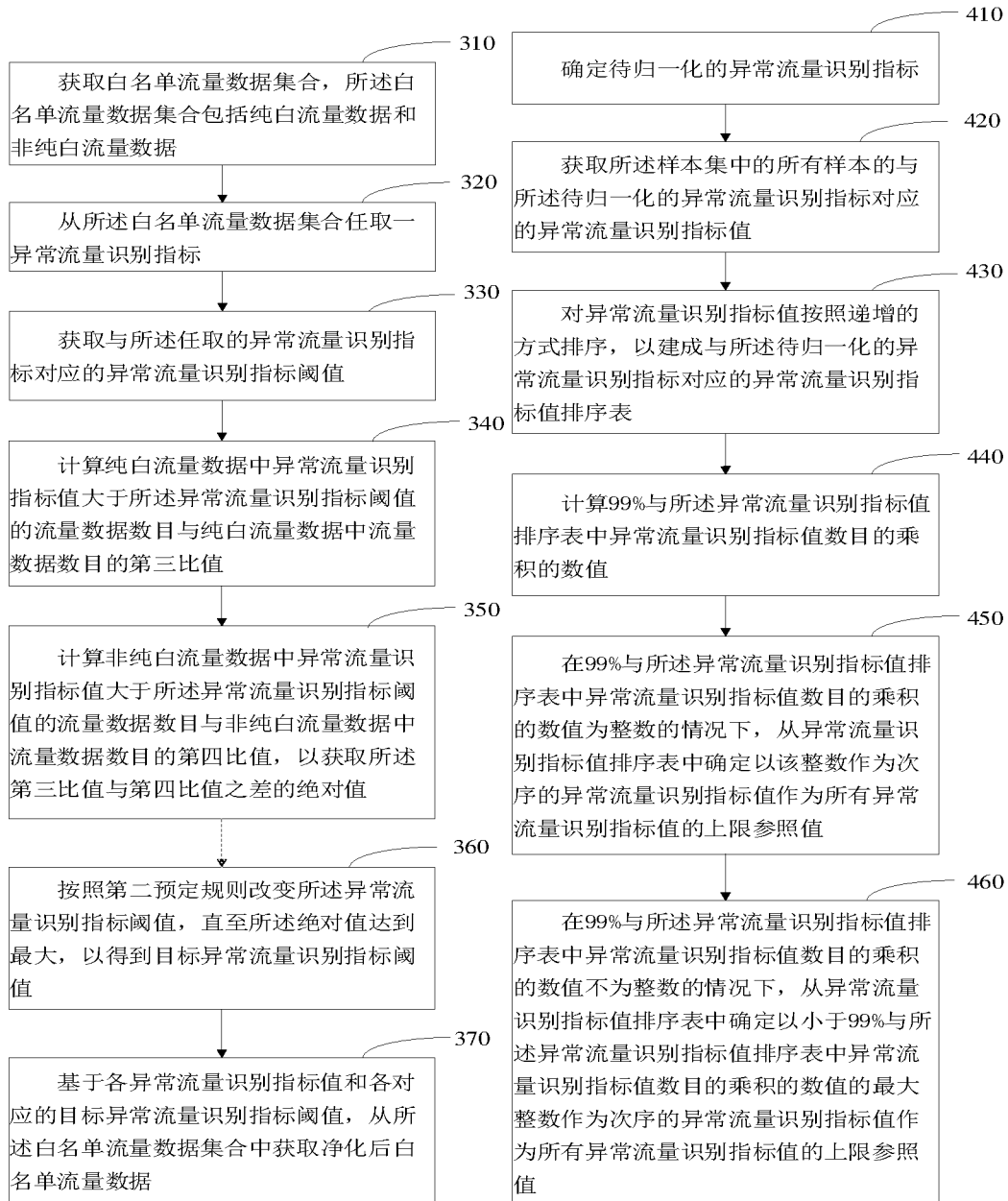


图 3

图 4

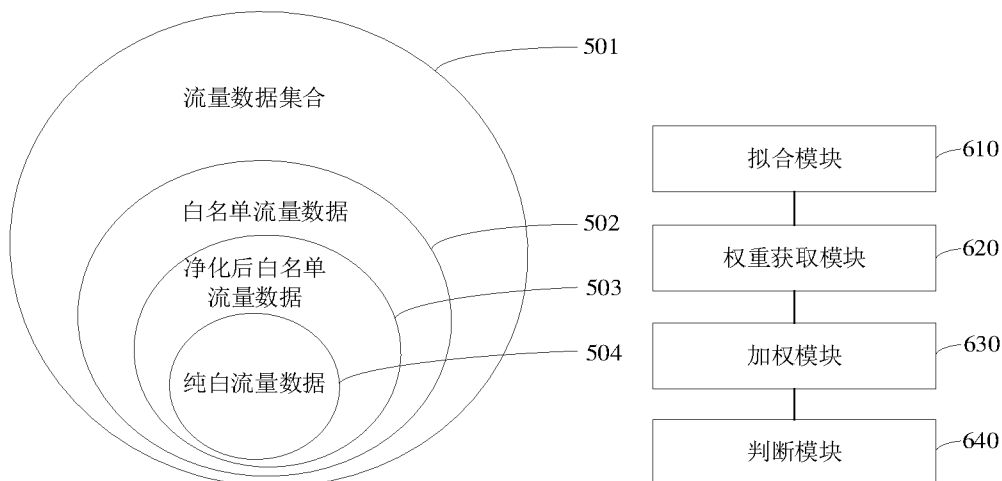


图 5

图 6

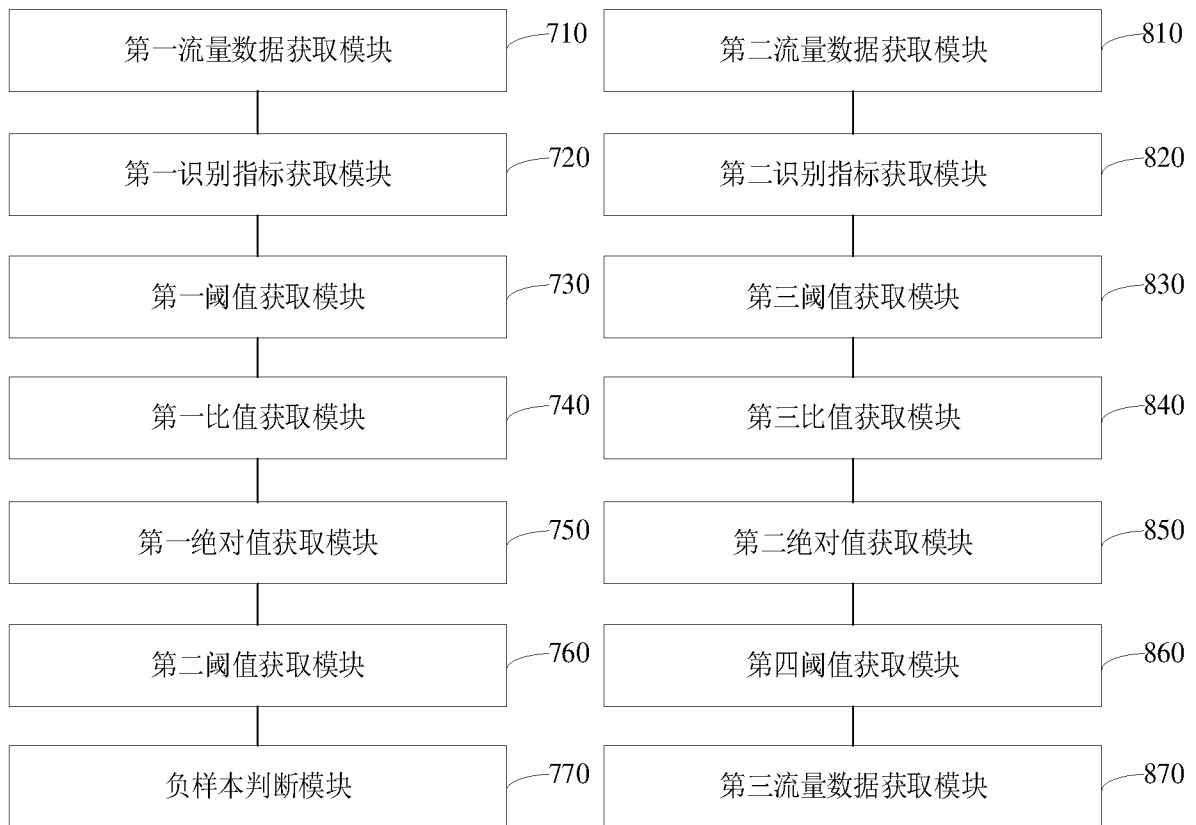


图 7

图 8

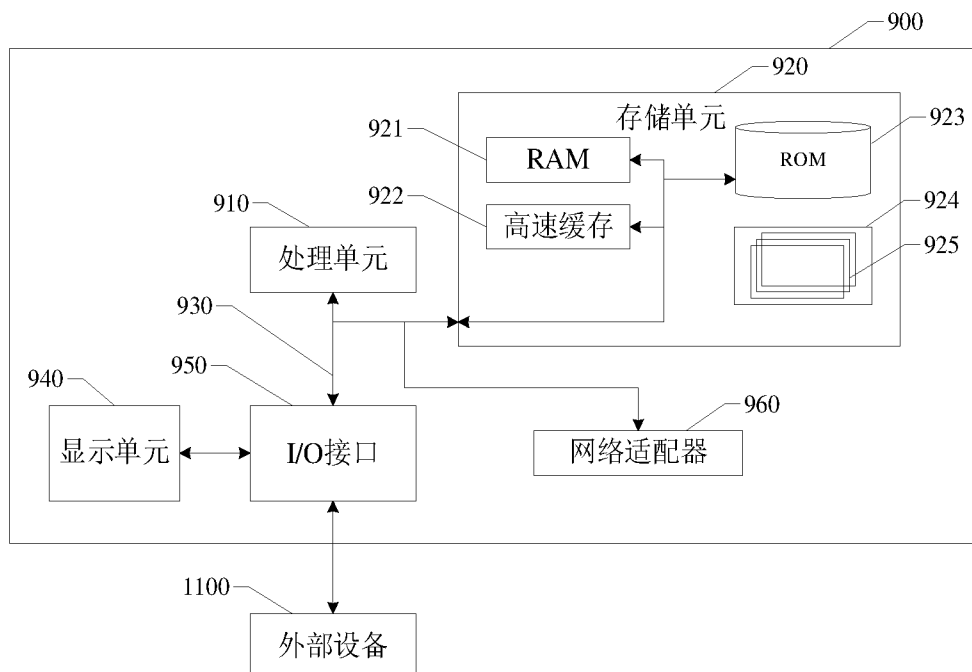


图 9

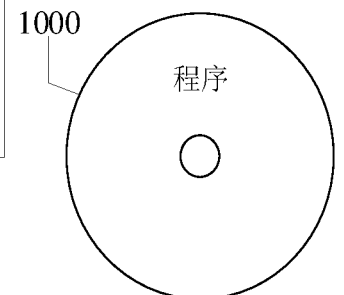


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/125709

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L G06F G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC, IEEE: 异常, 流量, 数据, 识别, 辨识, 判断, 样本, 逻辑斯蒂, 函数, 权重, 阈值, abnormal, traffic, data, identify, sample, logistic, function, weight, threshold

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107067157 A (BEIJING QIYI CENTURY SCIENCE & TECHNOLOGY CO., LTD.) 18 August 2017 (2017-08-18) description, paragraphs [0029]-[0064], and figures 2-4	1-20
X	侯爱华 等 (HOU, Aihua et al.). "基于逻辑回归模型的流量异常检测方法研究 (Research on Traffic Anomaly Detection Method Based on the Logistic Regression Model)" <i>工程数学学报 (Chinese Journal of Engineering Mathematics)</i> , Vol. 34, No. 5, 31 October 2017 (2017-10-31), ISSN: 1005-3085, pp. 479-489	1-20
A	CN 108206046 A (XINHUA SANDA DATA TECHNOLOGY CO., LTD.) 26 June 2018 (2018-06-26) entire document	1-20
A	CN 107169534 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. et al.) 15 September 2017 (2017-09-15) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 July 2019

Date of mailing of the international search report

31 July 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/125709

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2017061322 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 02 March 2017 (2017-03-02) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/125709

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107067157	A	18 August 2017	None			
CN	108206046	A	26 June 2018	None			
CN	107169534	A	15 September 2017	None			
US	2017061322	A1	02 March 2017	US	2019034836	A1	31 January 2019

国际检索报告

国际申请号

PCT/CN2018/125709

A. 主题的分类 H04L 29/06 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																				
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) H04L G06F G06Q 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, CNKI, WPI, EP0DOC, IEEE: 异常, 流量, 数据, 识别, 辨识, 判断, 样本, 逻辑斯蒂, 函数, 权重, 阈值, abnormal, traffic, data, identify, sample, logistic, function, weight, threshold																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 107067157 A (北京奇艺世纪科技有限公司) 2017年 8月 18日 (2017 - 08 - 18) 说明书第[0029]-[0064]段、图2-4</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>侯爱华 等, . “基于逻辑回归模型的流量异常检测方法研究,” 《工程数学学报》, 第34卷, 第5期, 2017年 10月 31日 (2017 - 10 - 31), ISSN: 1005-3085, 第479-489页</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108206046 A (新华三大数据技术有限公司) 2018年 6月 26日 (2018 - 06 - 26) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 107169534 A (北京京东尚科信息技术有限公司 等) 2017年 9月 15日 (2017 - 09 - 15) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2017061322 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2017年 3月 2日 (2017 - 03 - 02) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 107067157 A (北京奇艺世纪科技有限公司) 2017年 8月 18日 (2017 - 08 - 18) 说明书第[0029]-[0064]段、图2-4	1-20	X	侯爱华 等, . “基于逻辑回归模型的流量异常检测方法研究,” 《工程数学学报》, 第34卷, 第5期, 2017年 10月 31日 (2017 - 10 - 31), ISSN: 1005-3085, 第479-489页	1-20	A	CN 108206046 A (新华三大数据技术有限公司) 2018年 6月 26日 (2018 - 06 - 26) 全文	1-20	A	CN 107169534 A (北京京东尚科信息技术有限公司 等) 2017年 9月 15日 (2017 - 09 - 15) 全文	1-20	A	US 2017061322 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2017年 3月 2日 (2017 - 03 - 02) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 107067157 A (北京奇艺世纪科技有限公司) 2017年 8月 18日 (2017 - 08 - 18) 说明书第[0029]-[0064]段、图2-4	1-20																		
X	侯爱华 等, . “基于逻辑回归模型的流量异常检测方法研究,” 《工程数学学报》, 第34卷, 第5期, 2017年 10月 31日 (2017 - 10 - 31), ISSN: 1005-3085, 第479-489页	1-20																		
A	CN 108206046 A (新华三大数据技术有限公司) 2018年 6月 26日 (2018 - 06 - 26) 全文	1-20																		
A	CN 107169534 A (北京京东尚科信息技术有限公司 等) 2017年 9月 15日 (2017 - 09 - 15) 全文	1-20																		
A	US 2017061322 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2017年 3月 2日 (2017 - 03 - 02) 全文	1-20																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																				
国际检索实际完成的日期 2019年 7月 22日		国际检索报告邮寄日期 2019年 7月 31日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 田越 电话号码 86-(10)-53961347																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/125709

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	107067157	A	2017年 8月 18日	无	
CN	108206046	A	2018年 6月 26日	无	
CN	107169534	A	2017年 9月 15日	无	
US	2017061322	A1	2017年 3月 2日	US 2019034836 A1	2019年 1月 31日