



(12) 发明专利

(10) 授权公告号 CN 111600750 B

(45) 授权公告日 2022. 10. 21

(21) 申请号 202010391469.8

G06N 3/04 (2006.01)

(22) 申请日 2020.05.11

G06N 3/08 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 111600750 A

(56) 对比文件

CN 101146022 A, 2008.03.19

US 2020076842 A1, 2020.03.05

(43) 申请公布日 2020.08.28

CN 110287439 A, 2019.09.27

(73) 专利权人 北京庭宇科技有限公司

CN 109714322 A, 2019.05.03

地址 100020 北京市朝阳区酒仙桥路甲16号3层3081号

CN 101146021 A, 2008.03.19

审查员 阎洁

(72) 发明人 陶清乾 孙昊 石进 苏文飞

(74) 专利代理机构 南京华恒专利代理事务所

(普通合伙) 32335

专利代理师 宋方园

(51) Int. Cl.

H04L 41/0816 (2022.01)

H04L 43/0876 (2022.01)

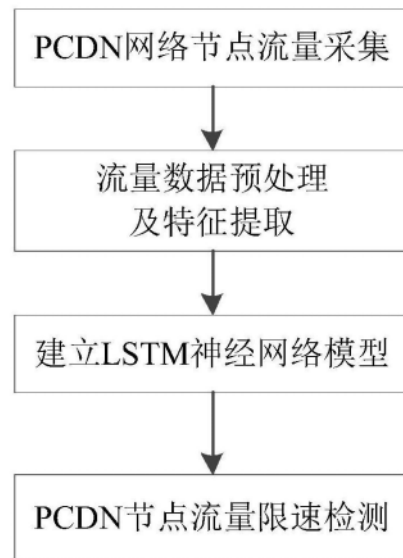
权利要求书2页 说明书4页 附图3页

(54) 发明名称

一种PCDN网络节点流量的限速检测方法及系统

(57) 摘要

本发明公开一种PCDN网络节点流量的限速检测方法及系统,采集正常服务的PCDN网络节点在时间窗口的数据流上行速率,预处理生成数据流速率特征序列;重复执行t次得到t个数据流速率特征序列s,以此特征序列s作为输入数据对LSTM神经网络模型进行训练,得到正常服务的网络通信LSTM神经网络流量模型;采集待测网络的数据流上行速率,预处理生成数据流速率特征序列s';将特征序列s'输入LSTM神经网络模型中,判定其是否存在流量限速的情况。本发明无需事先给出流量识别规则且处理速度快,不影响PCDN节点的正常网络服务;通过输入不同类型的PCDN服务的流量,可获得多种类型的LSTM训练模型,适用于多种PCDN服务场景。



1. 一种PCDN网络节点流量的限速检测方法,其特征在于,包括以下步骤:

(1) 数据采集及处理

采集正常服务的PCDN网络节点在时间窗口 $[t_1, t_n]$ 的数据流上行速率,预处理生成数据流速率特征序列;重复执行 t 次,得到 t 个数据流速率特征序列 s ;

所述步骤(1)包括:

(1.1) 采集正常服务的PCDN网络节点在时间窗口 $[t_1, t_n]$ 的数据流上行速率,得到 n 个上行数据流速率观测值序列 $k = \{k_1, k_2 \dots k_n\}$,其中 k_i 是 t_i 时刻的数据流上行速率, $1 \leq i \leq n$;

(1.2) 预处理数据流速率观测值序列,即:

计算流速率观测值序列 k 的一阶差分 $y = \{y_1, y_2 \dots y_{n-1}\}$,其中 $y_i = k_{i+1} - k_i$, $1 \leq i \leq n - 1$;

计算流速率观测值序列 k 的二阶差分 $z = \{z_1, z_2 \dots z_{n-2}\}$,其中 $z_j = y_{j+1} - y_j$, $1 \leq j \leq n - 2$;

组合差分得到二维流速率特征序列 $s = \begin{pmatrix} y \\ z \end{pmatrix}$;

(2) 数据训练

以步骤(1)中的数据流速率特征序列 s 作为输入数据,对LSTM神经网络模型进行训练,得到正常服务的网络通信LSTM神经网络流量模型;

(3) 数据检测

采集待测网络的数据流上行速率,预处理生成数据流速率特征序列 s' ;将特征序列 s' 输入步骤(2)中已经训练完成的LSTM神经网络模型中,判定其是否存在流量限速的情况。

2. 根据权利要求1所述的PCDN网络节点流量的限速检测方法,其特征在于:所述步骤(2)具体包括:

(2.1) 以流速率特征序列 s 作为输入训练LSTM神经网络模型;

(2.2) 将流速率特征序列 s 编码成隐变量,采用Attention机制对隐变量分配权重生成包含数据流信息的上下文变量;

(2.3) 采用sigmoid做为激活函数,将上下文变量归一化得到初始网络模型;

(2.4) 使用步骤(1)得到的流速率特征序列反复训练,采用梯度下降反向传播误差,不断调整模型中各个节点和连接的权重,得到最终的LSTM训练模型。

3. 根据权利要求1所述的PCDN网络节点流量的限速检测方法,其特征在于:所述步骤(3)中数据检测的具体过程包括:

(3.1) 采集待检测的PCDN网络节点的数据流上行速率,得到数据流速率观测值序列 k' ,按照步骤(1)中的方法预处理该序列得到二维流速率特征序列 s' ;

(3.2) 将二维流速率特征序列 s' 输入步骤(2)生成的LSTM神经网络模型中,即:将特征序列 s' 编码成隐变量,采用Attention机制对隐变量分配权重生成包含数据流信息的上下文变量,采用sigmoid函数作为激活函数,即可得到检测结果。

4. 一种实现权利要求1至3任意一项所述的PCDN网络节点流量的限速检测方法的系统，其特征在于：包括数据采集模块、数据训练模块和数据检测模块；
所述数据采集模块提取节点流速率，预处理获得流特征序列；
所述数据训练模块对流特征序列进行训练，形成正常服务下的网络通信流量模型；
所述数据检测模块通过所得网络通信流量模型对待测网络流量进行检测，判定其是否存在网络限速情形。

一种PCDN网络节点流量的限速检测方法及系统

技术领域

[0001] 本发明涉及网络安全技术,具体涉及一种PCDN网络节点流量的限速检测方法及系统。

背景技术

[0002] PCDN建立在传统的CDN系统基础上,在骨干网层次保留了原有的CDN系统的架构和功能,在边缘节点引入了P2P技术来进行文件及流媒体的共享,实现了P2P技术与CDN传输的结合。在CDN骨干网仍旧继承了CDN的内容缓存机制、全局负载均衡机制、骨干网内容分发流程、认证计费相关机制等基本技术特征。PCDN在核心节点根据P2P协议对内容(包括文件和流)做切片处理,P2P用户将根据这些规则来完成P2P共享,P2P在边缘层的引入大大降低了边缘服务器的压力,提高了文件传输和流媒体传输的效率。P2P技术充分利用了用户的闲置上行带宽,这样运营商可以通过更少的边缘服务器,提供更多的业务量为更多的用户服务。

[0003] PCDN产品通常采用月95峰值计费方式,即用户节点IP结算按95峰值带宽计费,在一个自然月内,取每5分钟有效带宽值进行降序排列,然后把带宽数值前5%的点去掉,剩下的最高带宽就是95峰值带宽即计费值。在实际应用中,可能会存在多个用户IP共享固定带宽的情况,比如两个IP地址(IP1和IP2)共享1G带宽。在PCDN调度期间满负荷的情况下,用户这两个IP的月95峰值流量约1G。然而,部分用户利用计费方式的特点,在调度期间的前半段时间限制IP2的速率,使得IP1满负荷运行,在调度期间的下半段时间限制IP1的速率,使得IP2满负荷运行。在这种情况下,IP1和IP2的月95峰值分别是1G,即该用户两个IP的计费流量为2G。由于用户私自限制PCDN节点的速率,造成PCDN产品服务的带宽波动,严重影响PCDN业务的稳定性。

[0004] PCDN技术目前还在快速发展期,PCDN技术主要应用于互联网主要厂商的业务中,然而PCDN市场资源良莠不齐,缺乏对PCDN网络节点的有效监控。

[0005] 现有的其他领域中,基于规则的检测方法由于规则的有限性并不能适应PCDN业务的需求,目前专门针对PCDN网络节点流量限速检测尚无有效的技术手段。基于PCDN网络节点资源的市场现状和PCDN服务对节点资源稳定性的要求,设计一种PCDN网络节点流量的限速检测方法及系统,引导PCDN网络节点市场的健康发展。

发明内容

[0006] 发明目的:本发明的目的在于解决PCDN网络节点管控上的空白,提供一种PCDN网络节点流量的限速检测方法及系统,本发明能在不影响PCDN业务的条件下快速检测PCDN网络节点限速情况,同时通过对不同类型的业务(如视频类业务、图像类业务、存储类业务、计算类业务等)流量训练,该方法能够支持多种形式的调度业务限速检测。

[0007] 技术方案:本发明的一种PCDN网络节点流量的限速检测方法,包括以下步骤。

[0008] (1) 数据采集模块:采集正常服务的PCDN网络节点在时间窗口 $[t_1, t_n]$ 的数据流上行速率,预处理生成数据流速率特征序列;重复执行 t 次,得到 t 个数据流速率特征序列;

[0009] (2) 数据训练模块:以步骤(1)中的数据流速率特征序列作为输入数据,对LSTM神经网络模型进行训练,得到正常服务的网络通信LSTM神经网络流量模型;

[0010] (3) 数据检测模块:采集待测网络的数据流上行速率,预处理生成数据流速率特征序列;将特征序列输入步骤(2)中已经训练完成的LSTM神经网络模型中,判定其是否存在流量限速的情况;

[0011] 进一步的,所述步骤(1)具体包括以下步骤:

[0012] (1.1) 采集正常服务的PCDN网络节点在时间窗口 $[t_1, t_n]$ 的数据流上行速率,得到n个上行数据流速率观测值序列 $k = \{k_1, k_2 \dots k_n\}$,其中 k_i 是 t_i 时刻的数据流上行速率, $1 \leq i \leq n$;

[0013] (1.2) 预处理数据流观测值序列,即:

[0014] 计算流速率观测序列k的一阶差分 $y = \{y_1, y_2 \dots y_{n-1}\}$,其中 $y_i = k_{i+1} - k_i$, $1 \leq i \leq n-1$;

[0015] 计算流速率观测序列k的二阶差分 $z = \{z_1, z_2 \dots z_{n-2}\}$,其中 $z_j = k_{j+1} - k_j$, $1 \leq j \leq n-2$;

[0016] 组合差分得到二维流速率特征序列 $s = \begin{pmatrix} y \\ z \end{pmatrix}$ 。

[0017] 进一步的,所述步骤(2)具体包括以下步骤:

[0018] (2.1) 以流速率特征序列s作为输入训练LSTM神经网络模型;

[0019] (2.2) 将流速率特征序列s编码成隐变量,采用Attention机制对隐变量分配权重生成包含数据流信息的上下文变量;

[0020] (2.3) 采用sigmoid做为激活函数,将上下文变量归一化得到初始网络模型;

[0021] (2.4) 使用步骤(1)得到的流速率特征序列反复训练,采用梯度下降反向传播误差,不断调整模型中各个节点和连接的权重,得到最终的LSTM训练模型。

[0022] 进一步的,所述步骤(3)具体包括以下步骤:

[0023] (3.1) 采集待检测的PCDN网络节点的数据流上行速率,得到数据流观测值序列 k' ,按照步骤(1)中的方法预处理该序列得到二维流速率特征序列 s' ;

[0024] (3.2) 以二维流速率特征序列 s' 作为输入将步骤(2)生成的LSTM神经网络模型中,即:将特征序列 s' 编码成隐变量,采用Attention机制对隐变量分配权重生成包含数据流信息的上下文变量,采用sigmoid函数作为激活函数,即可得到检测结果。

[0025] 有益效果:与现有技术相比,本发明具有如下显著优点:

[0026] (1) 本发明以正常运行的PCDN网络节点流作为输入,通过LSTM训练得到训练模型,这种基于LSTM的方法与基于规则的检测方法相比,无需事先给出流量识别规则,且处理速度快,不影响PCDN节点的正常网络服务。

[0027] (2) 本发明具有很强的PCDN业务适配性。不同的PCDN服务有不同的流量特征,这些特征根据PCDN调度时间有很大的关系。通过输入不同类型的PCDN服务的流量,可获得多种类型的LSTM训练模型。

附图说明

[0028] 图1为本发明的整体流程示意图;

[0029] 图2为实施例中某视频类业务流量图;

- [0030] 图3为实施例中循环神经网络结构；
 [0031] 图4为实施例中LSTM的循环单元；
 [0032] 图5为实施例中待测网络流量图。

具体实施方式

[0033] 下面对本发明技术方案进行详细说明,但是本发明的保护范围不局限于所述实施例。

[0034] 如图1所示,本实施例的一种PCDN网络节点流量的限速检测方法,其具体实施步骤如下:

[0035] (1) 采集某视频类业务的PCDN网络节点24小时的数据流速率,提取数据流速率特征,形成流特征序列,该视频类业务的网络流量如图2所示;每隔5分钟采集一次正常服务下的PCDN网络节点流速率,采集时间窗口为24小时,得到288个流速率观测值序列 $X = \{X_1, X_2, \dots, X_{288}\}$;计算流速率观测序列的一阶差分 $Y = \{Y_1, Y_2, \dots, Y_{287}\}$,其中 $Y_i = \{X_{i+1} - X_i\}$, $1 \leq i \leq 287$;计算流速率观测序列的二阶差分 $Z = \{Z_1, Z_2, \dots, Z_{286}\}$,其中 $Z_i = \{Y_{i+1} - Y_i\}$, $1 \leq i \leq 286$;组合差分得到二维流速率特征序列 $K = \begin{bmatrix} Y_i \\ Z_i \end{bmatrix} = \begin{bmatrix} X_{i+1} - X_i \\ X_{i+2} - X_i \end{bmatrix}$, $0 \leq i \leq 286$;

[0036] (2) 重复执行步骤(1) n次,得到n个二维数据流速率特征序列 $S = \{K_i\}$, $1 \leq i \leq n$;

[0037] (3) 以步骤(2)中的n个流速率特征序列作为输入,对LSTM神经网络模型进行训练,得到正常服务的网络通信LSTM神经网络流量模型;以流速率特征序列S作为输入训练LSTM神经网络模型;将流速率特征序列S编码成隐变量,采用Attention机制对隐变量分配权重生成包含数据流信息的上下文变量;采用sigmoid做为激活函数,将上下文变量归一化得到初始网络模型;使用步骤(2)得到的n个流速率特征序列反复训练,采用梯度下降反向传播误差,不断调整模型中各个节点和连接的权重,得到最终的LSTM训练模型;图3所示为循环神经网络结构,图4所示为LSTM的循环单元。

[0038] 其中本实施例的LSTM网络结构设计如下:LSTM的循环模块包含四个神经网络层,涉及到三个门激活函数 σ_1 、 σ_2 以及 σ_3 ,两个输出激活函数 ϕ_1 和 ϕ_2 ,符号 π 和 Σ 分别表示向量点乘和加法。

[0039] 输入门层(input gate)学习记忆新信息并存储在细胞状态中,对应的数学公式为 $I_t = \sigma_2(w_i[H_{t-1}, x_t, y_t, z_t] + b_i)$,其中, σ_2 是sigmoid函数, w_i 是输入门神经元的权重, H_{t-1} 是上一个循环单元的输出, x_t 是当前时点的上行网络流量, y_t 是当前时点上行网络流量的一阶差分,即 $y_t = x_t - x_{t-1}$, z_t 是当前时点上行网络流量的二阶差分, b_i 是输入门的误差。

[0040] 遗忘门层(forget gate)忘记细胞状态中的部分信息,对应的数学公式为 $F_t = \sigma_1(w_f[H_{t-1}, x_t, y_t, z_t] + b_f)$,其中, σ_1 是sigmoid函数, w_f 是遗忘门层神经元的权重, b_f 是遗忘门的误差。

[0041] 输出门层(output gate)激活产生输出,对应的数学公式为 $O_t = \sigma_3(w_o[H_{t-1}, x_t, y_t, z_t] + b_o)$,其中, σ_3 是sigmoid函数, w_o 是输出门层神经元的权重, b_o 是输出门层的误差。

[0042] (4) 采集待测PCDN网络节点24小时的数据流速率,提取其流速率特征,形成未知网络数据流的流特征序列;待测流量如图5所示。

[0043] (5) 以步骤(4)中的流特征序列作为输入,将其输入步骤(2)中的LSTM神经网络模

型,对待测网络数据流进行检测,判定其是否存在网络节点流量限速的情况。将步骤(4)生成的流速率特征序列编码成隐变量,采用Attention机制对隐变量分配权重生成包含数据流信息的上下文变量;采用sigmoid函数作为激活函数,即可得到检测结果。

[0044] 通过上述实施例可以看出,本发明能够保持PCDN业务稳定性的同时,及时发现用户限制PCDN网络节点速率的情形;并且本发明无需事先给出流量识别规则,且处理速度快,不影响PCDN节点的正常网络服务;通过输入不同类型的PCDN服务的流量,可获得多种类型的LSTM训练模型,适用于多种PCDN服务场景。

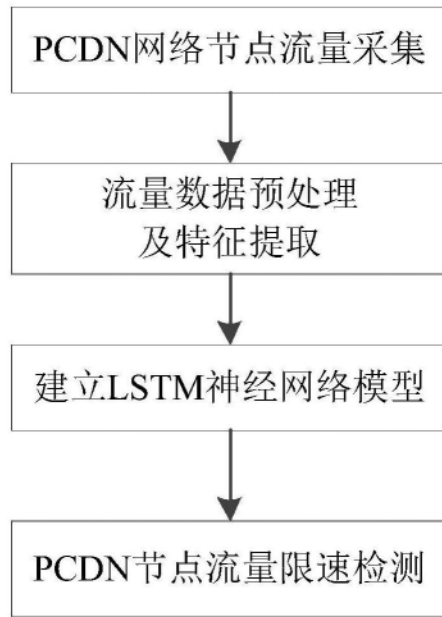


图1

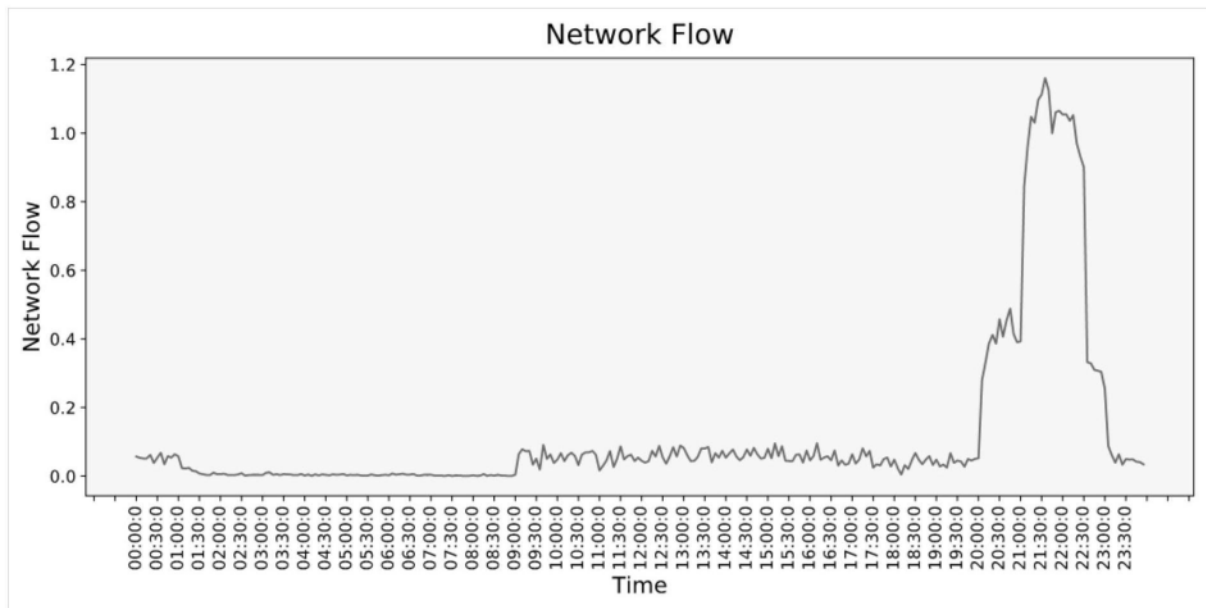


图2

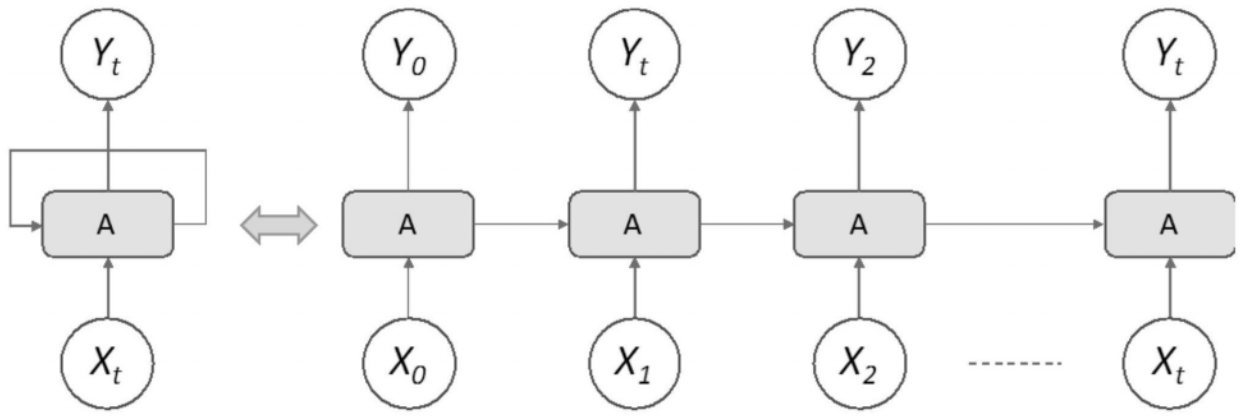


图3

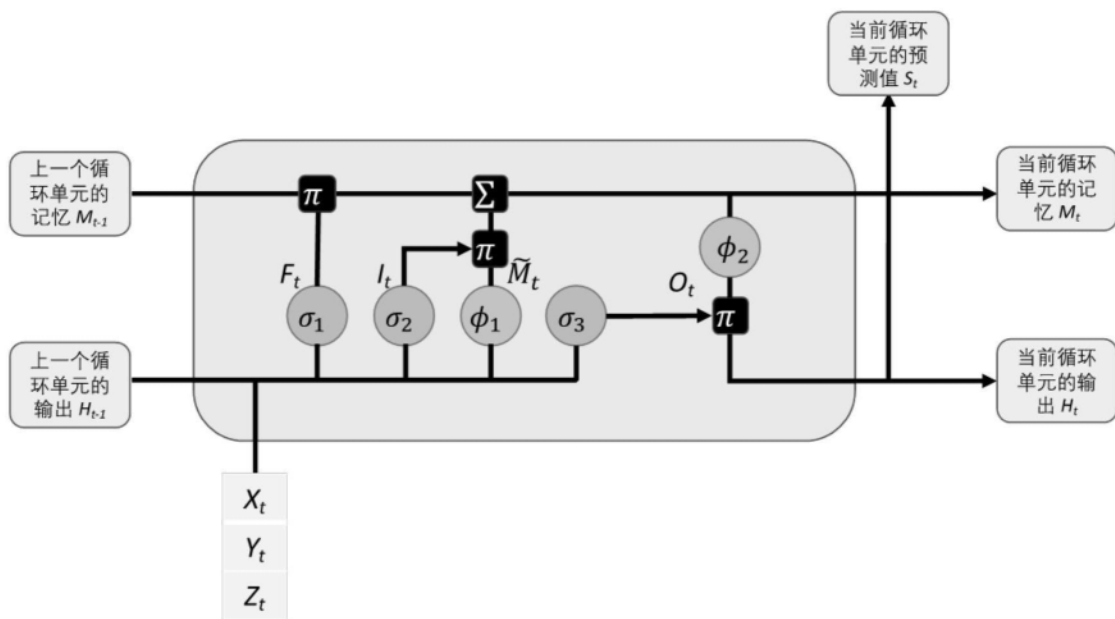


图4

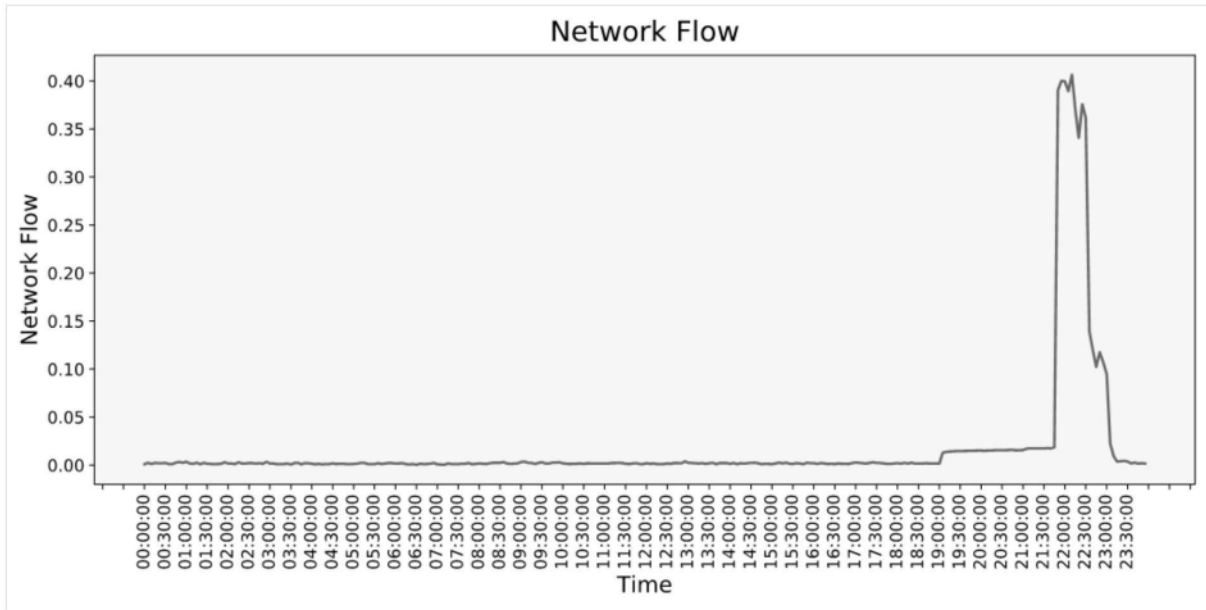


图5