



基于机器学习的 SDN 网络流量预测与部署策略

刘佳美, 徐巧枝

(内蒙古师范大学 计算机科学技术学院, 呼和浩特 010022)

摘 要: 针对由于网络流量的复杂多变而导致的软件定义网络(SDN)架构控制平面的负载不均问题, 提出一种基于隐马尔科夫优化的最大熵网络流量预测和控制器预部署 PPME 模型。根据协议种类对 SDN 流量进行分类, 利用已捕获的历史数据流, 采用最大熵算法预测未来数据流的分布, 生成控制平面中各类控制器的预部署方案, 并加入隐马尔科夫链对预测方案的时效性进行优化。实验结果表明, 相比于 SVR 模型与 GBRT 模型, 该模型具有更高的预测精度, 且生成的预部署方案能够适应复杂 SDN 环境中的动态变化, 减少了由于突发事件而导致的负载不均和控制器迁移, 缩短了由控制器迁移而产生的网络延迟与响应时间。

关键词: 软件定义网络; 机器学习; 最大熵; 隐马尔科夫; 流量预测; 预部署

开放科学(资源服务)标志码(OSID):



中文引用格式: 刘佳美, 徐巧枝. 基于机器学习的 SDN 网络流量预测与部署策略[J]. 计算机工程, 2020, 46(10): 223-230.

英文引用格式: LIU Jiamei, XU Qiaozhi. Network traffic prediction and deployment strategy based on machine learning for SDN[J]. Computer Engineering, 2020, 46(10): 223-230.

Network Traffic Prediction and Deployment Strategy Based on Machine Learning for SDN

LIU Jiamei, XU Qiaozhi

(College of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010022, China)

[Abstract] To address uneven loads in the control plane of Software Defined Network(SDN) architecture caused by the complexity and variability of network traffic, this paper proposes a network traffic prediction and controller pre-deployment PPME model based on hidden Markov optimization with maximum entropy. The model classifies SDN traffic according to the protocol types, and uses the maximum entropy algorithm to predict the distribution of the future data stream based on the captured historical data stream, so as to generate the pre-deployment scheme of various controllers in the control plane. The timeliness of the prediction scheme is optimized by the introduced hidden Markov chain. Experimental results show that compared with SVR model and GBRT model, the proposed model has higher prediction accuracy, and its generated pre-deployment scheme can adapt to the dynamic changes in complex SDN environment. It reduces the load imbalance and controller migration caused by emergencies, and thus reduces the network delay and response time caused by controller migration.

[Key words] Software Defined Network(SDN); machine learning; maximum entropy; hidden Markov; traffic prediction; pre-deployment

DOI:10.19678/j.issn.1000-3428.0056436

0 概述

软件定义网络(Software Defined Network, SDN)是一种将网络控制和转发功能相分离的新型网络架构^[1], 其可简化网络管理, 提高网络的可编程性和灵活性。近年来, 随着机器学习技术的快速发展, 研究

人员开始尝试在 SDN 中引入机器学习方法, 以提高网络管理和控制效率, 解决传统方法中依赖配件、配置复杂的问题。

分布式 SDN 的架构包括扁平式和层次式^[2]2 种方式, 其中, 扁平式架构将整个网络划分为若干个 SDN 子域, 每个子域内分配一个控制器负责集中管

基金项目: 内蒙古自治区自然科学基金(2012MS0930); 内蒙古自治区高等学校科学研究项目(NJZY18023, NJZY12032)。

作者简介: 刘佳美(1994—), 女, 硕士研究生, 主研方向为机器学习、软件定义网络; 徐巧枝, 副教授。

收稿日期: 2019-10-29 修回日期: 2019-12-02 E-mail: 1124928309@qq.com

理域内交换机,且多个控制器被分别部署在不同的子域内,从而构建一个逻辑上集中、而物理位置上分布的 SDN 架构,如 Onix^[3]、Hyperflow^[4]等。在扁平式架构下,主要采用交换机迁移和重定向流的策略^[5]来平衡多个控制器之间的负载,以减轻控制器的负载,但是频繁迁移交换机将会增加网络延迟和开销,降低网络性能。层次式架构将控制平面划分为垂直的层次,常见的方式为将控制器分为普通控制器和全局控制器两层^[6-7],或者在控制器和交换机之间增加一个中间层^[8],但是这些方案均未对全局控制器或者中间层控制器的过载问题进行深入研究和解决。

文献[9]提出了 SHLB 模型,在控制平面与数据平面之间插入一个中间层,在中间层采用分流策略将不同类型的数据流传输给不同类型的控制器,可有效缓解控制平面负载不均的问题,并采用文献[10]中的实时调度策略解决中间层过载问题。但是全局控制器在发现某个控制器资源负载过轻或过重时,需要合并或者开启新的控制器,在此过程中,将会涉及大量流控制策略的迁移,造成网络延迟的增加,从而影响网络性能。

与以往的单边研究思路不同,本文在提出网络流量预测算法后,通过收集不同的网络参数,在 SHLB 模型的基础上对其进行改进,并提出基于最大熵和隐马尔科夫算法的流量预测与多控制器预部署 PPME 模型。该模型基于捕获的历史数据流,利用最大熵算法预测未来一段时间内数据流的分布,从而产生控制平面内各类控制器的预部署方案,加入隐马尔科夫链对预测方案进行优化,并采用增强学习算法对实际数据流与预测数据流进行比较。

1 相关工作

1.1 SHLB 模型

SHLB 模型是一个可扩展的、分层的 SDN 控制平面负载均衡模型,该模型在控制平面与数据平面之间增加一个中间层,将 SDN 网络分为控制层、中间层和数据层三层。控制层包含若干个控制器,其接收中间层转发的数据流,按照设定规则对数据流进行处理,并将处理结果反馈给中间层。中间层接收到数据层交换机的消息流后,按类别或者其他规则将其转发给某个控制器,再接收该控制器的处理结果并通知其相应的交换机。数据层由大量的 SDN 交换机组成,其将无法处理的数据流转发给中间层,接收中间层的处理结果并更新流表规则。SHLB 模型架构如图 1 所示。

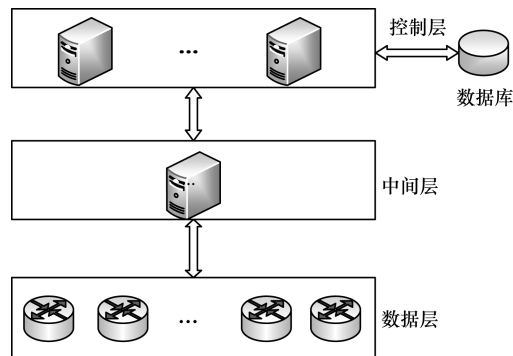


图 1 SHLB 模型架构

Fig. 1 SHLB model architecture

SHLB 模型控制层中的控制器按角色可分为全局控制器和局部控制器,其中,全局控制器负责保障控制层的负载均衡,并对局部控制器进行动态调整。局部控制器一方面接收、处理与响应中间层的数据流和请求,另一方面需周期性地向全局控制器报告实时的资源利用率并接收其命令。

从细粒度的时间角度来看,网络流是一个不断变化的动态变量,但是从一个时间范围来看,其又是相对稳定的。如果负载均衡模型只基于当前时间点考虑这种变化,而不考虑未来数据流的变化趋势,可能会造成控制器的低性价比迁移,即这种迁移并不能保证一定时间段内网络性能的整体提高,而只对当前突发的网络数据流提供帮助,可能造成整体网络延迟的增加。为了解决该问题,本文提出了 PPME 模型,采用增强学习算法对实际数据流与预测数据流进行比较,若偏差大于设定阈值,则重新生成部署方案;否则,仍采用旧方案。在方案运行期间,控制器不再进行迁移,以降低迁移对网络性能的影响。

1.2 最大熵原理

最大熵模型^[11]是一种对数线性模型,其是在给定事件集上挖掘潜在约束条件,并在样本数据有限的情况下,选取一种能够满足已知约束条件,且尽可能使未知事件均匀分布的恰当模型。最大熵模型是采用概率估计的方法,基于信息熵理论^[12]而建立起来的概率统计模型。信息熵作为一个随机事件不确定性的度量,随机事件越不确定,则其熵越大;若随机事件退化为确定事件,则其熵为 0。随机变量 X 的信息熵可以被定义为:

定义 1 假设 $X = \{x_1, x_2, \dots, x_n\}$ 是一个离散随机变量,则其熵 $H(X)$ 表示为:

$$H(X) = - \sum_{i=1}^n p_i \lg p_i \quad (1)$$

其中, n 为离散随机变量可取值的总数量, $p_i = P\{X = x_i\}$ 表示 X 取值为 x_i 的概率,若某一项 $p_i = 0$,即 X 取值为 x_i 的概率为 0。

定义2 随机变量 X 与 Y 的联合分布形成的联合熵用 $H(X, Y)$ 表示, 则其条件熵为:

$$H(X|Y) = H(X, Y) - H(Y) = - \sum_{x,y} p(x,y) \lg p(x|y) \quad (2)$$

定义3 假设 $p(x)$ 、 $q(x)$ 是 X 中取值的 2 个概率分布, 则 p 对 q 的相对熵为:

$$D(p \| q) = \sum_x p(x) \lg \frac{p(x)}{q(x)} = E_{p(x)} \lg \frac{p(x)}{q(x)} \quad (3)$$

随机变量 X 与 Y 的互信息定义为 X 与 Y 的联合分布和独立分布乘积的相对熵, 分别如下所示:

$$I(X, Y) = D(P(X, Y) \| P(X)P(Y)) \quad (4)$$

$$I(X, Y) = \sum_{x,y} p(x,y) \lg \frac{p(x,y)}{p(x)p(y)} \quad (5)$$

1.2.1 最大熵迭代算法

最大熵迭代算法具体步骤为:

输入 特征函数 f_1, f_2, \dots, f_n

输出 最优参数值 w^* , 最优模型 P_w

步骤1 在经验分布 $\tilde{P}(X, Y)$ 、模型 $P_w(y|x)$ 中, 对所有的 $i \in \{1, 2, \dots, n\}$, 取初值 $w_i = 0$, 对每个 $i \in \{1, 2, \dots, n\}$:

1) 令 ∂_i 是方程 $\sum_{x,y} \tilde{P}(x) P(y|x) f_i(x, y) \exp(\partial f^*(x, y)) = E_{\tilde{P}}(f_i)$ 的解, 且 $f^*(x, y) = \sum_{i=1}^n f_i(x, y)$ 。

2) 更新 w_i 值, $w_i \leftarrow w_i + \partial_i$ 。

步骤2 如果不是所有的 w_i 都收敛, 则重复步骤1。

近年来, 最大熵算法不仅被广泛应用于机器翻译、词性标注、语法分析与文本分类等领域^[13-15], 还被应用于构建预测模型。文献[16]运用最大熵算法预测在 RCP2.6 和 RCP8.5 2 个未来气候变暖的情景下, 水杨柳在云南省的潜在分布情况。文献[17]采用最大熵原理构建流量预测算法, 根据历史预测误差对各种单一预测算法进行取舍, 最终得到预测结果的分布情况。通过对实际网络流量进行预测与测量, 结果发现, 与其他预测算法相比, 最大熵算法具有较高的预测精度, 且预测结果更合理。

网络流量会随着时间的空间不断发生变化。在时间维度上, 网络流量会随着一天中的不同时段发生变化, 比如白天的流量高于夜晚的流量, 此外, 网络流量还可能在极短的时间内发生剧烈变化^[18]。本文提出的 PPME 模型在 SDN 控制层中利用最大熵算法, 通过历史网络流量预测未来网络流量, 从而实现对控制器的预部署, 以降低控制器迁移的频率, 提高网络性能。

1.2.2 隐马尔科夫算法

隐马尔科夫算法^[19-20]广泛应用于语言识别、强化学习和生物信息学等领域。隐马尔科夫算法是一个有限离散时间的马尔科夫模型, 该模型假设系统是一个具有隐状态的马尔科夫过程。它可以被定义为一组 $\lambda = (\Pi, A, B)$, Π 是初始概率分布, A 是转移概率矩阵, B 是一个观测概率序列(发射概率)。具体来说, 一个隐马尔科夫算法是由以下要素^[21]定义的:

1) $Q = \{q_1, q_2, \dots, q_N\}$ 是一组状态, 其中, N 为隐藏状态的数量, q_t 为 t 时刻的隐藏状态。

2) $O = \{o_1, o_2, \dots, o_T\}$ 是一组观测值, 其中, T 是观测值的数量, o_T 是 T 时刻的可观测状态。每个观测状态都来源于一个词汇表 $V = \{v_1, v_2, \dots, v_M\}$, 其中, M 为观测值的数量。

3) $A = \{a_{11}, a_{12}, \dots, a_{N1}, \dots, a_{NN}\}$ 是一个 $N \times N$ 的转移概率矩阵。 a_{ij} ($1 \leq i, j \leq N$) 表示状态 i 变为状态 j 的概率, 具体表示方法为:

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i) \quad (6)$$

$$a_{ij} \geq 0, \forall i, j \quad (7)$$

$$\sum_{j=1}^N a_{ij} = 1, \forall i \quad (8)$$

隐马尔科夫算法的状态图如图2所示, 其中, H_i ($i=1, 2, \dots, T$) 为隐藏状态集合, O_i ($i=1, 2, \dots, T$) 为观测值集合。马尔科夫过程位于可观测状态之上, 由当前状态决定, A 只能观测 O_i , 且 O_i 由马尔科夫过程和 B 的隐状态决定。

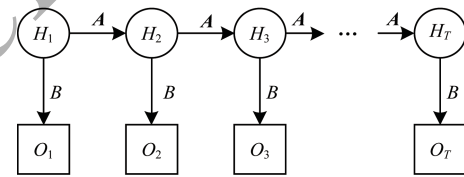


图2 隐马尔科夫算法状态图

Fig. 2 State diagram of hidden Markov algorithm

2 本文模型设计

针对传统方法不能对网络流量的变化特征进行准确描述, 且预测精度较低的问题, 本文提出 PPME 模型, 该模型在控制层引入最大熵和隐马尔科夫算法, 通过学习历史数据流来预测未来一段时间的数据流, 从而提前部署控制器, 以减少控制器的迁移次数, 缩短网络延迟。PPME 模型的设计总共分为 3 个部分, 首先, 收集历史网络数据流并对其进行分类。其次, 利用最大熵算法对未来网络流进行预测, 并产生预测结果, 最后, 利用隐马尔科夫算法对预测结果进行评估, 优化控制器部署。

2.1 数据流收集与分类

2.1.1 数据预处理

本文从 MAWI Working Group 获取了 12 月 \times 30 天 \times 24 小时的流量集,该流量集已被广泛应用于流分类,可以帮助研究人员评估其流量检测方法^[22]。数据集中的数据特征包括产生数据包的时间、传输控制层的协议、目标地址、发送地址以及数据包配置信息(数据包大小、端口号与 MAC 地址)等。对于本文来说,影响流量预测的主要因素是数据包的数量、时间和协议类型,因此,PPME 模型以数据包大小、数据包协议名称和时间作为特征值,通过对数据进行整理,生成了网络流数据矩阵,具体如图 3 所示,其中,行表示每秒内各个协议数据包的数量,列表示各种协议类型。在单位时间内,不同协议的数据流数量差别较大,如图 4 所示,其中,HTTP 数据流最多,且 HTTPS、UDP 和 ICMP 数据流相差不大,而 IPv6、UDPv6 及其他数据流较少,利用该流量分布指导控制层中控制器的分类及部署。

	数据流						其他数据流
	HTTP	HTTPS	UDP	ICMP	IPv6	UDPv6	
1 s	5.0	2.0	1.0	3.0	0.0	0.0	0.0
2 s	3.0	7.0	0.0	2.0	0.0	0.0	0.0
3 s	0.0	4.0	0.0	0.0	0.0	0.0	0.0
4 s	0.0	4.0	1.0	0.0	0.0	0.0	0.0
5 s	8.0	0.0	3.0	2.0	0.0	0.0	0.0
6 s	19.0	2.0	1.0	0.0	0.0	0.0	0.0
7 s	1.0	0.0	0.0	0.0	0.0	0.0	0.0
8 s	0.0	0.0	2.0	0.0	0.0	0.0	0.0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n s	0.0	4.0	0.0	0.0	0.0	0.0	0.0

图 3 网络流数据矩阵

Fig. 3 Network flow data matrix

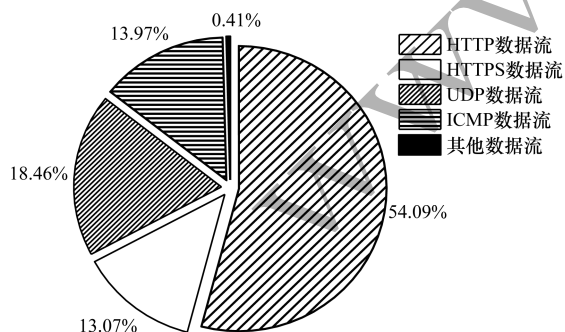


图 4 各种协议数据流的分布

Fig. 4 Distribution of various protocol data flows

2.1.2 数据流分类

关于数据流的分类,部分研究采用机器学习的方法,如文献[23]提出的 TrAdaBoost 模型,其是

AdaBoost 模型的改进,在解决多级网络流量的分类方面具有较好的性能。实验比较了 TrAdaBoost 模型与本文模型对所得数据流的识别效果,如表 1 所示。从表 1 可以看出,在准确度与所需时间方面,本文 PPME 模型更具有优势。

表 1 TrAdaBoost 模型与本文模型的比较

Table 1 Comparison between TrAdaBoost model and the proposed model

模型	准确度/%	时间/s
TrAdaBoost 模型	96.2	45.0
本文 PPME 模型	98.9	32.0

2.2 数据流预测

在了解一个随机事件的概率分布时,预测满足所有的约束且不进行任何条件下的主观假设,概率分布最均匀时预测风险最小,从而得到概率分布的熵最大,即最大熵数学模型。因为网络流复杂多变,且很多人为的主观预测很难得到验证,通过客观的历史数据构造概率矩阵,以求得最大熵情况下的概率矩阵作为当时的预测结果,可以最大程度地提高预测的准确性。

由于数据包的数量能够客观反映数据流量,因此 PPME 模型将数据包的数量和数据包的协议类型作为主要特征参数,并设计了数据流预测算法,数据流预测算法为基于最大熵的网络流预测算法,其中, G 为数据包矩阵, N 为数据包总数, Z 为概率矩阵, E 为最小常数误差。最大熵网络流预测算法的具体步骤为:

步骤 1 将训练矩阵转化为概率矩阵,即 $Z(i,j) = G(i,j)/N$ 。

步骤 2 将训练矩阵乘以其转置矩阵,得到概率方阵,即 $Z_1 = Z_0 \times Z_0^T$ 。

步骤 3 对概率方阵进行迭代幂运算, $Z_2 = Z_1 \times Z_1^T$, $Z_3 = Z_2 \times Z_2^T$, ..., $Z_n = Z_{n-1} \times Z_{n-1}^T$ 。

步骤 4 计算方差 $|Z_n - Z_{n-1}|$,直到方差小于设置的常数 E ,则结果收敛,且在该结果上进行预测时产生的误差最小;否则重复步骤 3。

在最大熵算法中,通过将训练矩阵乘以训练矩阵的转置矩阵可以得到一个概率方阵,通过对该概率方阵进行迭代幂运算,最终可得到一个趋于稳定的各种协议数据流量的比例关系。根据该比例关系对控制器进行预部署,可以提高控制器的利用率,降低控制器迁移过程对网络运行的影响,从而提高网络性能。

2.3 预测结果的优化

由于最大熵算法的计算量非常庞大,目前还没

有有效的方法能彻底解决其计算复杂度问题,导致 PPME 控制器预部署方案的生成需要花费较长时间,对于合理的部署方案,尽可能延长其工作时间是非常有必要的。通过对预测流量与实际流量进行拟合,结果发现,最大熵算法的预测结果在一段时间内具有较为稳定的预测准确度,如图 5 所示。最大熵算法在模拟流量曲线的过程中,流量变化曲线的导数越大,则算法的预测准确度越低,且时效性越差。在理想条件下,导数为 0 且流量曲线不变化时是预测算法的理想状态,但是对于网络流来说,其流量曲线在不断变化,几乎不可能出现导数为 0 的情况,因此单一的最大熵算法无法满足较长时间的预测准确度。为了解决该问题,PPME 模型在控制层引入了隐马尔科夫链,利用隐马尔科夫的五元组(隐含状态 S ,观测状态 O ,初始状态概率矩阵 P 、隐含状态转移概率矩阵 M ,观测状态转移概率矩阵 K),通过对不同阶段的预测结果给予奖励或惩罚,并对历时较长的隐马尔科夫过程给予 γ 平方衰减,以降低其权重。

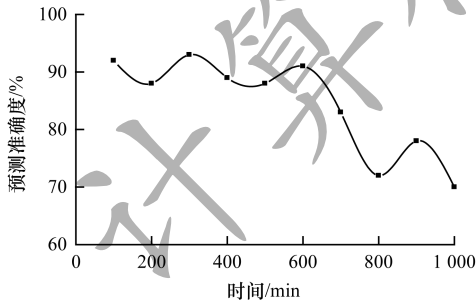


图 5 单一最大熵算法的预测准确度随时间的变化曲线

Fig.5 Change curve of prediction accuracy of the single maximum entropy algorithm over time

不同阶段的训练奖励机制可以在很大程度上提高最大熵算法预测结果的准确度。通过对每个阶段的数据进行最大熵预测,极大地减少了最大熵算法的运算量,提高了运算效率。

在部署周期优化算法中,设置初始指数 W_0 为 1 000,警戒比例 K_0 为 0.618,阻尼指数 K_2 为 1,假设矩阵 Z 有 n 行、 m 列,其具体步骤为:

步骤 1 对矩阵的每行进行遍历,for x in range(n), $x \in [0, n)$ 。

步骤 2 对矩阵的每列进行遍历,for y in range(m), $y \in [0, m)$ 。

步骤 3 初始化:令 $W_x = W_0$ 。

步骤 4 加强学习, $W_x = W_x + K_2 \times Z[x, y]$ 。

步骤 5 如果 $W_x > K_0 \times W_0$,保持使用原有预测

方案,否则重新使用最大熵算法进行网络流量预测,并部署控制器。

实验验证了加入隐马尔科夫算法后的预测拟合效果,如图 6 所示。使用单一的最大熵算法在前 600 min 内,预测准确度较高,但随着时间的延长,预测准确度呈降低趋势;加入隐马尔科夫后该算法仍可以保持较高的预测准确度,这说明了本文算法通过不断调整每个阶段预测结果的准确度,可在多数情况下较准确地预测复杂场景下的网络流量情况,优化最大熵算法的预测结果。

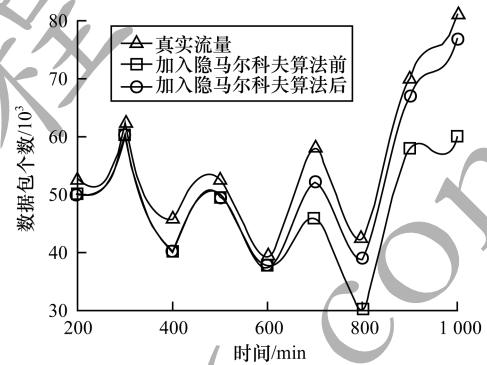


图 6 最大熵算法加入隐马尔科夫算法前后的网络流量对比

Fig.6 Comparison of network traffic before and after adding hidden Markov algorithm to maximum entropy algorithm

3 实验结果评估

3.1 最大熵算法预测结果的准确度

本文对所采集的网络流以 8:2 的比例进行分割,分别作为训练集和测试集。实验将采用最大熵算法对训练集中的数据进行训练,并预测之后时间段的流量。通过比较预测结果与实际流量来评估最大熵算法预测结果的准确度。

本文以 4 个不同大小的数据集(流量包个数依次为 12 104、63 741、180 040、623 120)为研究对象,实验采用支持向量回归(Support Vector Regression, SVR)模型、梯度提升回归树(Gradient Boosting Regression Tree, GBRT)模型与本文提出的 PPME 模型对其下一时刻流量值的预测情况与实际流量进行对比,如图 7 所示。从图 7 可以看出,对于不同的数据集,预测准确度存在差异,当数据量较小时,3 种算法的预测准确度相差不大,且均能达到 85% 左右,但是随着流量数据的增加,SVR 模型和 GBRT 模型的预测误差逐渐增大,而本文提出的 PPME 模型仍然保持较高的预测准确度,且最大熵的核心是收敛的,因此数据集的大小不会影响最大熵的准确度,在大型复杂的网络场景下,最大熵算法的预测准确度较稳定。

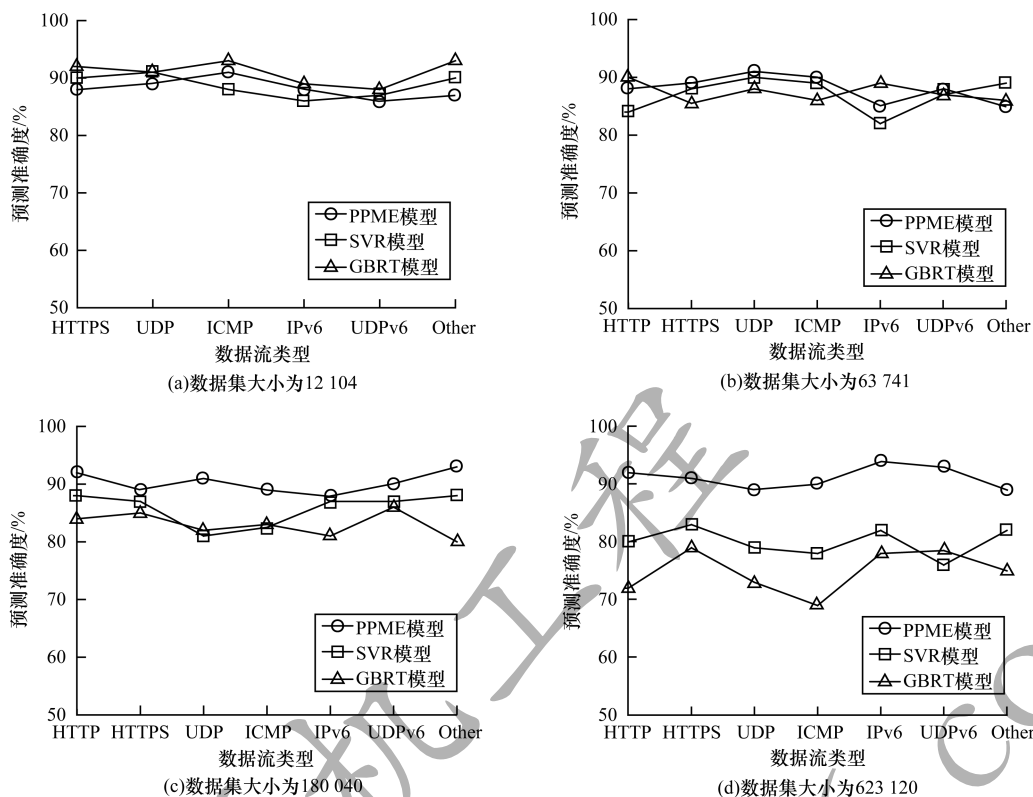


图 7 3 种模型在不同数据集中预测准确度的对比

Fig. 7 Comparison of prediction accuracy of three models in different datasets

3.2 预部署的有效性

本文模型将不同协议的网络流分配到相应类型的控制器中进行处理,实验采用与 SHLB 模型相同的计算机作为控制器,每个控制器所能装载的数据包个数如图 8 所示。从图 8 可以看出,当数据流达到 30 000 个/s 时,控制器负载率达到了 80%,即认为控制器满负载。因此,本文对每个控制器每秒最多分配 30 000 个数据包,然后再进行部署。

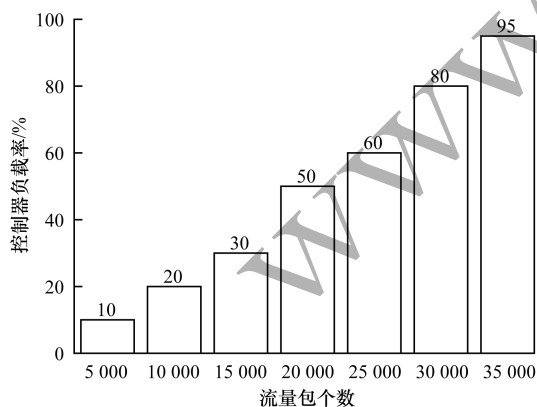


图 8 不同数量数据包的控制器负载率

Fig. 8 Controller load rates for different number of packets

图 9 比较了 PPME 模型与 SHLB 模型部署的控制器个数。考虑到图 4 所示的各协议数据包的分布

情况,相比 SHLB 模型,PPME 模型考虑了一段时间内各类网络流的整体情况,对于数据量巨大的 HTTP 数据流,PPME 模型部署了多余的控制器,避免当 HTTP 数据流增加时控制器发生迁移,而对于数量相对较少的 HTTPS、UDP 及 ICMP 数据流则部署较少的控制器,避免了资源浪费,同时降低了控制器迁移的次数与频率。

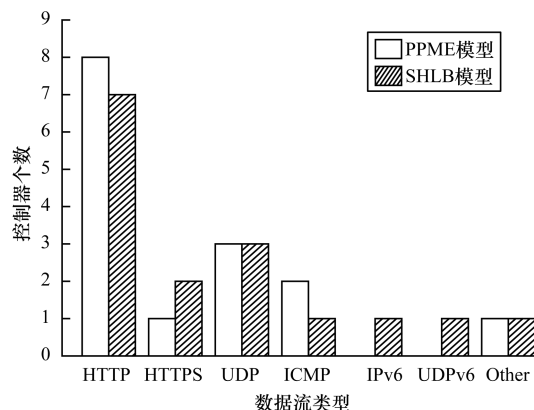


图 9 PPME 模型与 SHLB 模型控制器分配情况对比

Fig. 9 Comparison of controller allocation between PPME model and SHLB model

在进行控制器分配后,对于多组数据,实验通过发送包工具发送不同速率的流量数据包对 PPME

模型和SHLB模型进行测试,中间层实体资源利用情况和网络响应时间如图10和图11所示。实验结果表明,PPME模型在网络发生特殊状况后,系统的处理速度和性能均更好,且承受突发状况的能力更强。

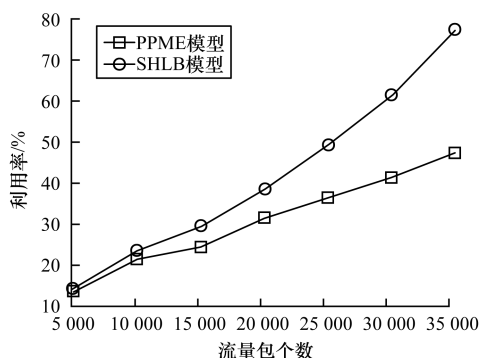


图10 中间层内存利用率

Fig.10 Memory utilization of middle layer

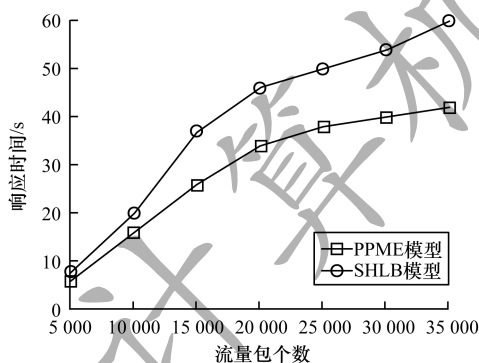


图11 中间层网络响应时间

Fig.11 Response time of middle layer network

4 结束语

本文基于最大熵和隐马尔科夫算法构建了PPME模型,通过对不同类型的网络数据包数量进行预测,为SDN模型中的控制器提供预部署方案的理论依据。PPME模型能够实现在复杂网络情景下对流量的准确预测,同时,将网络流量以协议类型进行分类,能够达到对SDN中多控制器实现预部署的目的。实验结果表明,PPME模型对网络流量的预测准确度达到85%以上,实现了对SDN中控制器的预部署策略,有效缓解网络中的拥堵问题,提高SDN环境中的资源利用率和控制器的工作效率,且保证了SDN控制网络的稳定性。下一步将在云环境下,将各控制器的决策控制情况汇报给云端,并在云端利用EM高斯方差进行聚类分析。

参考文献

- [1] TANG T, MHAMDI L, MCLERNON D, et al. Deep learning approach for network intrusion detection in software defined networking [C]//Proceedings of 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM). Washington D. C., USA: IEEE Press, 2016: 258-263.
- [2] SCHMID S, SUOMELA J. Exploiting locality in distributed SDN control [C]//Proceedings of the 2nd ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking. New York, USA: ACM Press, 2013: 121-126.
- [3] KOPONEN T, CASADO M, GUDE N, et al. Onix: a distributed control platform for large-scale production networks [EB/OL]. [2019-09-25]. <https://wenku.baidu.com/view/83a8f8ccdd88d0d233d46a9e.html>.
- [4] TOOTOONCHIAN A, GANJALI Y. HyperFlow: a distributed control plane for OpenFlow [C]//Proceedings of the 2010 Internet Network Management Conference on Research on Enterprise Networking. California, USA: USENIX Association, 2010: 3-6.
- [5] KRISHNAMURTHY A, CHANDRABOSE S P, GEMBER-JACOBSON A. Pratyastha: an efficient elastic distributed SDN control plane [C]//Proceedings of the 3rd Workshop on Hot Topics in Software Defined Networking. New York, USA: ACM Press, 2014: 133-138.
- [6] YAO Haipeng, QIU Chao, ZHAO Chenglin, et al. A multicontroller load balancing approach in software-defined wireless networks [J]. International Journal of Distributed Sensor Networks, 2015(2): 1-8.
- [7] MA Y W, CHEN J L, TSAI Y H, et al. Load-balancing multiple controllers mechanism for software-defined networking [J]. Wireless Personal Communications, 2017, 94(4): 3549-3574.
- [8] BASTA A, BLENK A, BELHAJ HASSINE H, et al. Towards a dynamic SDN virtualization layer: control path migration protocol [C]//Proceedings of the 11th International Conference on Network and Service Management. Washington D. C., USA: IEEE Press, 2015: 354-359.
- [9] LI Lin. Research on load balancing in control plane of SDN [D]. Hohhot: Inner Mongolia Normal University, 2018. (in Chinese)
李琳. SDN网络控制平面负载均衡技术的研究 [D]. 呼和浩特: 内蒙古师范大学, 2018.
- [10] DIXIT A A, HAO F, MUKHERJEE S, et al. ElastiCon: an elastic distributed sdn controller [C]//Proceedings of the 10th ACM/IEEE Symposium on Architectures Networking and Communications Systems. New York, USA: ACM Press, 2014: 17-28.
- [11] WANG Lei. Sentiment analysis of Chinese words based on maximum entropy [J]. Computer Era, 2018(12): 7-11. (in Chinese)

- 王磊. 基于最大熵的中文词语情感分析研究[J]. 计算机时代, 2018(12): 7-11.
- [12] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.
- [13] YU D, HINTON G, MORGAN N, et al. Introduction to the special section on deep learning for speech and language processing[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 4-6.
- [14] HE Ming. A maximum entropy model text classification method based on improved information gain feature selection[J]. Journal of Southwest China Normal University(Natural Science Edition), 2019, 44(3): 113-118. (in Chinese)
- 何明. 一种基于改进信息增益特征选择的最大熵模型文本分类方法[J]. 西南师范大学学报(自然科学版), 2019, 44(3): 113-118.
- [15] SE Chajia, GONGBAO Cairang, CAI Rangjia. A comparative study of Tibetan new words recognition based on maximum entropy and HMM[J]. Journal of Qinghai Normal University(Natural Science Edition), 2018, 34(1): 12-16. (in Chinese)
- 色差甲, 贡保才让, 才让加. 基于最大熵和 HMM 的藏文新词识别对比研究[J]. 青海师范大学学报(自然科学版), 2018, 34(1): 12-16.
- [16] ZHOU Yang, YI Yujun, YANG Yufeng, et al. Predicting geographical distributions of homonofa riparia lour by using maximum entropy[J]. Water Resources and Hydropower Engineering, 2019, 50(5): 73-81. (in Chinese)
- 周扬, 易雨君, 杨雨风, 等. 基于最大熵模型预测水杨柳的潜在生境分布[J]. 水利水电技术, 2019, 50(5): 73-81.
- [17] LIN Tianfeng. A combined model for network traffic forecasting based on maximum entropy principle[J]. Microelectronics & Computer, 2006, 23(8): 147-149. (in Chinese)
- 林天峰. 基于最大熵原理的网络流量预测综合模型[J]. 微电子学与计算机, 2006, 23(8): 147-149.
- [18] HU Yannan. Research on key technologies and related issues of software-defined networks[D]. Beijing: Beijing University of Posts and Telecommunications, 2015. (in Chinese)
- 胡延楠. 软件定义网络关键技术及相关问题的研究[D]. 北京: 北京邮电大学, 2015.
- [19] RABINER L, JUANG B. An introduction to hidden Markov models[J]. IEEE ASSP Magazine, 1986, 3(1): 4-16.
- [20] BEVILACQUA V, DALENO D, CARIELLO L, et al. Pseudo 2D hidden Markov models for face recognition using neural network coefficients[C]//Proceedings of 2007 IEEE Workshop on Automatic Identification Advanced Technologies. Washington D. C., USA: IEEE Press, 2007: 1-9.
- [21] JURAFSKY D, MARTIN J H. Speech and language processing [EB/OL]. [2019-09-25]. <https://max.book118.com/html/2018/0317/157710205.shtm>.
- [22] FONTUGNE R, ABRY P, FUKUDA K, et al. Scaling in Internet traffic: a 14 year and 3 day longitudinal study, with multiscale analyses and random projections[J]. ACM Transactions on Networking, 2017, 25(4): 2152-2165.
- [23] SUN Guanglu, LIANG Lili, CHEN Teng, et al. Network traffic classification based on transfer learning[J]. Computers and Electrical Engineering, 2018, 69: 920-927.

编辑 刘继娟