

SDN 环境下基于机器学习算法的 DDoS 攻击检测模型

马乐乐, 束永安

(安徽大学 计算机科学与技术学院, 安徽 合肥 230601)

摘 要: 软件定义网络(SDN)是一种新兴的网络架构,将控制逻辑和转发逻辑分离。SDN 网络中,控制器拥有对网络的全局控制能力。正是由于控制器的这一特性,使得它成为分布式拒绝服务(DDoS)攻击的主要目标。针对这个问题,提出了一种基于机器学习的方法来检测 DDoS 攻击的模型,首先使用熵值检测流是否异常,发出异常警告后提取网络流特征,依次调用 SVM 与 K-means 两种机器学习算法来检测 DDoS 攻击。实验结果表明,所提出的算法能够降低误报率,并且对 DDoS 攻击的检测率和准确率高于原始的 SVM 和 K-means 算法。此外,还通过实验验证了所提出的模型的 CPU 平均利用率低于无熵值检测的 SVM+K-means 算法。

关键词: 软件定义网络;分布式拒绝服务;熵;支持向量机;K 均值

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2018)05-0015-06

A DDoS Attack Detection Model Based on Machine Learning Algorithm in SDN Environment

MA Le-le, SHU Yon-gan

(School of Computer Science and Technology, Anhui University, Hefei 230601, China)

Abstract: The Software Defined Network (SDN) is an emerging network architecture that separates control logic from forwarding logic. In SDN, the controller has a global control of the network. Because of this feature of the controller, making it becomes the main goal of the distributed denial of service (DDoS) attack. Aiming at this problem, this paper proposes a method based on machine learning to detect the DDoS attack model. Firstly, it uses the entropy to check whether the traffic is abnormal. After extracting the abnormal alarm, the network flow feature is extracted, and SVM and K-means are called to detect the DDoS attacks. The experimental results show that the proposed algorithm can reduce the false alarm rate, and the detection rate and accuracy of DDoS attacks are higher than those of the original SVM and K-means. In addition, the experimental results show that the average CPU utilization rate of the proposed model is lower than that of SVM + K-means without entropy detection

Key words: software defined network (SDN); distributed denial of service (DDoS); entropy; support vector machine(SVM); k-means

1 引言

在软件定义网络(SDN)中,控制功能被集中在控制平面的控制器上,控制器具有全局的网络视图,数据平面的数据转发要在控制器的控制管理下进行^[1]。SDN 包括三个平面和两个接口,即:数据平面,控制平面,应用层平面,南向接口和北向接口。

SDN 控制器与数据平面的交换机之间的通信通过南向接口来实现,其中最典型的南向接口是 OpenFlow 协议,与应用层平面通过北向接口 API 连接。OpenFlow 交换机有两种存在形式,即 OpenFlow 专用交换机和 OpenFlow 兼容交换机,OpenFlow 专用交换机仅支持 OpenFlow 协议,OpenFlow 兼容交换机不仅支持 OpenFlow 协议,也支持传统网

络数据链路层和网络层的各种协议^[2].

传统网络中的 DDoS 攻击流量检测是多年来热门的研究领域. 然而, 在基于 SDN 网络中, 关于这个问题的研究仍处于不成熟的阶段. SDN 网络比传统网络更容易受到网络异常流量的影响, 例如分布式拒绝服务 (DDoS) 攻击.

Zahra 等人^[3]提出了基于多层感知器和重力搜索算法的基于流的异常检测系统. 该系统可以以非常高的准确率区分正常流和攻击流. Seyed 等人^[4]使用基于熵值的检测方法, 提出了使用连续窗口的方法来计算熵值偏差, 使用连续窗口减少误报率. 但是 Seyed 等人的工作只做了流量的早期检测, 并没有对流量进行具体的分类. Trung 等人^[5]将模糊逻辑应用于 SDN 中的泛洪攻击防范机制, 但是这个实验是在某一个特定的数据集下得到的结果, 具有偶然性, 不适用于其他数据集, 可扩展性差.

本文所提出的模型首先对流量进行了早期检测, 当检测到异常时再调用 DDoS 攻击检测算法, 提高了流量检测的效率, 降低了控制器的负载. DDoS 攻击检测算法结合了 SVM 与 K-means 算法, 提高了准确率.

2 系统模型

本文提出的系统模型包括三部分: 异常检测模块、流量特征提取模块和攻击检测模块. 具体框架如图 1 所示.

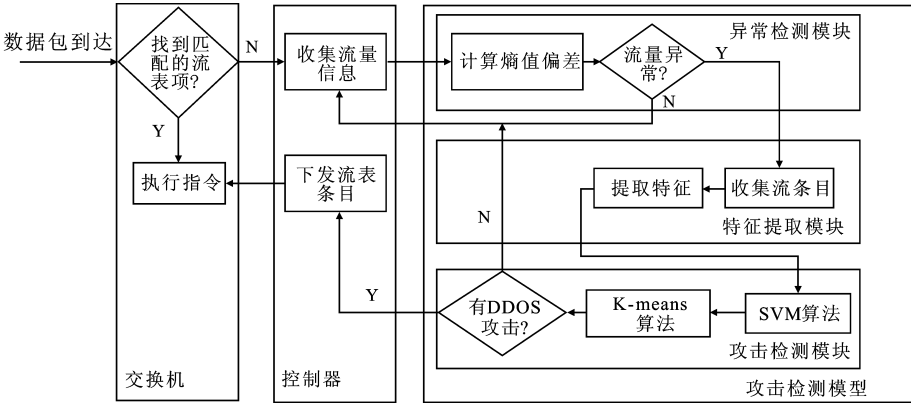


图 1 系统框架

2.1 异常检测模块

我们使用基于熵值检测的方法来检测流, 熵用于测量流量的随机性. 将熵用于 DDoS 检测的主要原因是它能够测量到达网络的数据包的随机性, 随机性越高熵值越高. 在这里, 使用熵检测流量时有两个重要的参数, 即窗口大小和阈值. 窗口大小是基于时间段或数据包数量确定的. 在该窗口内计算流量的熵, 以测量数据包中的随机性. 要判断流量是否异常, 还需要一个阈值. 如果计算得到的熵低于阈值则可以判断流量为异常. 熵的计算公式如下:

$$H = - \sum_{i=1}^n p_i \log p_i \tag{1}$$

n 表示一个窗口大小 T 内数据包的数量, p_i 表示窗口中每个目的 IP 地址的数据包出现的概率. 当每个 IP 地址只出现一次时, 熵值将取得其最大值. 通常 DDoS 攻击的目的主机只有一个, 则大量的数据包将被发送到该目的主机. 因此目的 IP 地址的熵将降到最低. 我们设定了一个实验阈值 C . 如果连续 5 个窗口的熵值下降到该阈值以下, 则认为网络流

量存在异常, 发出异常警告. 在本文中, 我们设定窗口大小为 50 个数据包, 连续测试 5 个窗口, 这样一共检测了 250 个数据包. 选择 5 个连续窗口是为了避免偶然性结果, 不受意外事件的干扰, 降低误报率.

但是在某些情况下, 正常的网络操作也会导致熵值比正常值低, 所以这种方法不能确定网络中是否有攻击流量. 所以当异常检测模块检测到异常时就发出警告, 调用流量特征提取模块提取流量特征, 为进一步的 DDoS 攻击检测做准备.

2.2 流量特征提取模块

流量特征提取模块部署在 SDN 控制器中, 每隔 P 秒向 OpenFlow 交换机发送请求消息查询数据平面中所有流的状态. 参数 P 可由管理员手动设置, 这里设定为 5 s. 当从 SDN 控制器中接收流之后, 流量特征提取模块解析流, 并且提取描述网络流的基本特征. 本文选取以下 4 个特征作为 DDoS 攻击检测算法的输入: (1) 数据包的数量 (2) 数据包的平均大小 (3) 字节数 (4) 数据包到达的时

间间隔,DDoS攻击流通过向网络中发送大量的无用数据包来消耗CPU资源和链路资源,使得网络无法处理正常的流请求。DDoS攻击流通常只含有一个数据包,并且数据包的到达时间间隔很短,并且其字节数明显大于正常流的字节数。因此,选择这四个特征可以很好地描述DDoS攻击流。提取特征后将其输入到攻击检测模块做进一步的DDoS攻击检测。

2.3 攻击检测模块

本文采用组合机器学习算法来进行DDoS攻击检测,因为单个机器学习算法在不同的训练样本上容易产生较大的误差,因此,通过使用组合机器学习算法,将机器学习算法的结果组合之后再输出有利于减少总误差^[6]。支持向量机SVM(Support Vector Machine)是一种简单的分类算法,在统计样本量较少的情况下,也能获得很好的统计效果,具有较好的鲁棒性。K-Means算法容易理解和实现,能够在已有聚类样本较少的情况下对样本进行分类,并且该算法具有迭代优化功能,可以克服少量样本的不准确性,时间复杂度低。本文具体采用SVM和K-means算法相结合的方法来检测DDoS攻击。

2.3.1 SVM算法简介

假设给定数据集

$S = \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n)\}$, 其中 $\vec{x}_i \in R^n, y \in \{+1, -1\}$ 。SVM算法试图找到一个能最大功能地分离两个数据组的最优超平面。SVM分类器为:

$$f(\vec{x}) = \text{sign}(\langle \vec{w}, \vec{x}_i \rangle + b) \quad (2)$$

本文只考虑数据集线性可分的情况,线性分类超平面方程为

$$\langle \vec{w}, \vec{x}_i \rangle + b = 0 \quad (3)$$

W 为权重系数, b 为偏差量。当 $y_i = 1, \langle \vec{w}, \vec{x}_i \rangle + b \geq 1$; 当 $y_i = -1, \langle \vec{w}, \vec{x}_i \rangle + b \leq -1$ 。因此最优超平面可表示为

$$y_i (\langle \vec{w}, \vec{x}_i \rangle + b) \geq 1 \quad (4)$$

因此,现在的问题就是找到最小的 $\|\vec{w}\|$, 同时仍然满足(4)的条件。该问题等价于找到最小的 $\frac{1}{2} \|\vec{w}\|^2$ 同时满足式(4)的条件,这是一个二次规划优化问题,可表示为:

$$\begin{cases} \min \varphi(\vec{w}) = \frac{1}{2} \|\vec{w}\|^2 \\ y_i (\langle \vec{w}, \vec{x}_i \rangle + b) \geq 1 \quad i = 1, 2, \dots, n \end{cases} \quad (5)$$

引入拉格朗日乘数, α_i 上述约束问题可以定义为

$$\min_{\vec{w}, b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=0}^n \alpha_i [y_i (\langle \vec{w}, \vec{x}_i \rangle + b) - 1] \right\} \quad (6)$$

为了构建最优超平面,必须找到鞍点。因此,可以使用标准的二次规划方法来解决这个问题。鞍点最终表示为训练向量的线性组合,即

$$\vec{w} = \sum_{i=1}^n \alpha_i y_i \vec{x}_i \quad (7)$$

通过估计所有支持向量的平均值 A_w 来定义偏移量 b 。偏移量的计算公式为

$$b = \frac{1}{A_w} \sum_{i=1}^{A_w} (y_i - \langle \vec{w}, \vec{x}_i \rangle) \quad (8)$$

支持向量机(SVM)是一个快速有效的分类器,但是对于在模糊区域的可疑数据点,SVM分类器无法具体分类。为了解决这个问题,我们使用K-means算法处理这些数据点。

2.3.2 K-Means算法简介

给定一组样本 $(x_1, x_2, \dots, x_n)$, K-means算法将 n 个样本值划分为 k 个集群,这 k 个集群的中心点为 $c = \{c_1, c_2, \dots, c_k\}$ 。对于每一个样本点 x_i , 计算其到所有 K 个集群中心点的距离,其中每个样本点属于具有最近距离平均值的集群,以此作为集群的原型,最小化集群内的平方和,即集群中每个样本点与 K 个集群的中心点的距离之和^[7]。集群内平方和的计算公式如下:

$$\sum_{i=1}^k \sum_{x \in c_i} \|x - c_i\|^2 \quad (9)$$

每个数据点属于一个集群,当所有的数据点都被分配给对应的集群时重新计算集群的平均值以确定最优中心点,并且在此步骤中,中心点更新可能导致数据点被分配给另一个集群。重复该过程,直到集群中心点不变,表明达到收敛。

2.3.3 DDoS攻击检测算法

DDoS攻击检测算法

(1) 训练SVM分类器 $f(\vec{x}) = \text{sign}(\langle \vec{w}, \vec{x}_i \rangle + b)$ 。

(2) 记录流量 x_i 。

(3) 根据最优超平面 $y_i (\langle \vec{w}, \vec{x}_i \rangle + b) \geq 1$ 将流

量分为正常流和异常流。

(4)记录 SVM 分类器无法准确分类的流量信息。

(5)根据其距离最优超平面的位置首先将流量分为正常或异常两类集群。

(6)分别计算两个集群的平均值。

(7)重新分配每个样本点,将其分配给具有最小平均值的集群。

(8)重复步骤(6),直至上一次迭代中的样本分配与当前迭代中的样本分配一致。

(9)基于流量集群将流量分为正常流量或攻击流量。

(10)将获取的值与类标签一起添加在训练数据集中。

3 实验分析

本节首先描述实验环境设置,然后对攻击检测算法的实验结果进行了分析和评估。

3.1 实验设置

本文使用 Mininet 网络模拟器来模拟真实的 SDN 网络环境,实验使用基于 Java 实现的开源控制器 OpenDayLight,操作系统为 Ubuntu14.04,在 i5 CPU 和 4G RAM 计算机上完成。

在本文的实验中,我们使用经典的 DDoS 工具 TFN 来发起 DDoS 攻击,该工具能够使用僵尸网络模拟 DDoS 攻击,分别向目标网络发送 TCP flood 和 ICMP flood.使用如图 2 所示的用 Mininet 创建的拓扑图来进行实验,网络 1 中有 50 台主机,包括正常主机和僵尸主机。

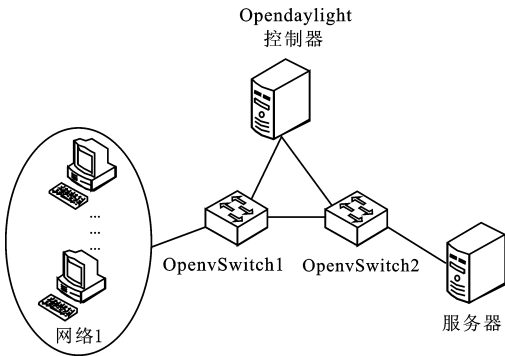


图 2 实验拓扑图

3.2 实验结果

本节对实验结果进行了全面的分析,在进行试验之前,首先分别使用正常数据集和异常的攻击数据集来训练 SVM 和 K-means 分类器.将 SVM+K-

means 算法、SVM 算法与 K-means 算法以同样的方式写入 SDN 控制器中并在不同的训练样本下进行实验。

TP(True Positive)指异常流量被正确识别的概率,TN(True Negative)指正常流量被正确识别的概率,FP(False Positive)指正常流量被错误识别为异常流量的概率,FN(False Negative)指异常流量被错误识别为正常流量的概率^[8].我们使用三个参数来评估算法的性能。

(1)检测率(DR:Detection Rate):检测率是被正确识别的攻击流的数量除以数据集中所有攻击流的数量的比值,它的大小与随机误差相关.检测率越高说明识别 DDoS 攻击的精度越高。

$$DR = \frac{TP}{TP + FN} \tag{10}$$

(2)准确率(ACC:Accuracy):准确率是正常流和攻击流分别被正确识别的数量除以数据集中的总体流的数量的比值,它的精度与系统误差相关.准确率越高,流量分类更精确。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

误报率(FAR:False Alarm Rate):误报率是被识别为攻击流的正常流的数量与数据集中所有正常流的数量的比值.噪声或其他错误配置都会影响误报率的值.误报率越低,分类效果越好。

$$FAR = \frac{FP}{FP + TN} \tag{12}$$

图 3 揭示了三种算法的检测率,SVM+K-means 方法的检测率最高为 95.35%,SVM 和 K-means 方法的检测率分别为 93.11%和 92.35%.三种算法的准确率如图 4 所示,使用 SVM+K-means 方法的准确率最高为 95.60%,而 SVM 和 K-means 方法的准确率分别为 92.83%和 92.99%.图 5 揭示

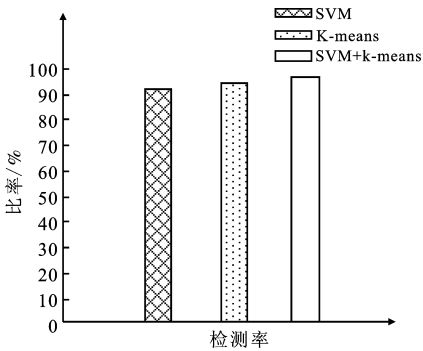


图 3 三种算法下的检测率对比图

了三种算法的误报率,我们可以看出使用本文所提

出的算法得到的误报率最小,仅有 4.16%,而 SVM 和 K-means 的误报率分别为 7.45%和 6.36%. 因此可以得出我们提出的方法的性能优于原始的 SVM 算法和 K-means 算法.

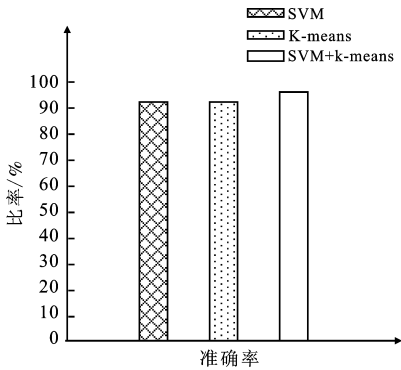


图 4 三种算法下的准确率对比图

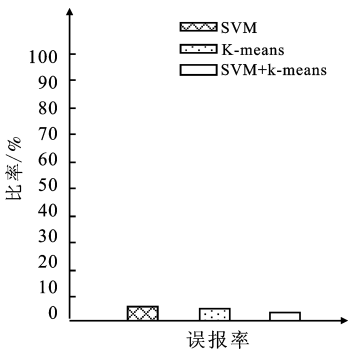


图 5 三种算法下的误报率对比图

此外,我们还对本文提出的模型的 CPU 的平均利用率进行了分析. 将本文提出的模型与没有早期检测的 SVM+K-means 算法的 CPU 的平均利用率进行了比较. 在本文所提出的模型中,首先利用熵对网络中的流进行早期检测,熵值的计算简单,消耗的 CPU 资源较少. 在我们的方法中,当在熵值检测模块中检测到流量异常时才调用机器学习算法来对流量进行进一步的检测. 因此,我们所提出的模型具有更小的 CPU 利用率. 如图 6 所示,我们所提出的方法的 CPU 的平均利用率最低,约为 25%. 而没有早期检测的 SVM+K-means 方法的 CPU 平均利用率分别约为 38%. 可以得出,我们所提出的方法可以有效防止 OpenFlow 交换机和 SDN 控制器的内存和 CPU 在 DDoS 攻击下过载,提高系统的鲁棒性.

4 结束语

本文我们提出了在 SDN 网络中检测 DDoS 攻击的框架模型,通过以上的实验从三个方面即检测率,准确率和误报率对实验结果进行了分析和评估,

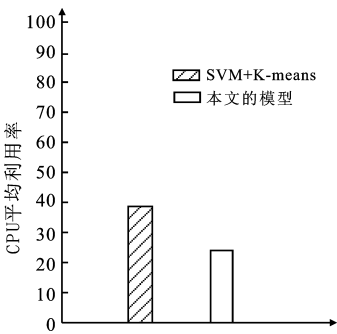


图 6 有无熵值早期检测的 CPU 平均利用率对比图

提高了检测率和准确率,降低了误报率和 CPU 的平均利用率.

参考文献:

[1] Kreutz D, Ramos F M V, Verissimo P E, et al. Software-defined networking: a comprehensive survey[J]. Proceedings of the IEEE, 2015, 103(1):14-76.

[2] Nunes B A A, Mendonca M, Nguyen X N, et al. A survey of software-defined networking: past, present, and future of programmable networks[J]. IEEE Communications Surveys & Tutorials, 2014, 16 (3): 1617-1634.

[3] Jadidi J Z, Muthukkumarasamy V, Sithirasenan E, et al. Flow-based anomaly detection using neural network optimized with gsa algorithm[C]// 2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops. Philadelphia, PA, 2013: 76-81.

[4] Mousavi S M, St-Hilaire M. Early detection of DDoS attacks against SDN controllers[C]// 2015 International Conference on Computing, Networking and Communications (ICNC). Garden Grove, CA, 2015: 77-81.

[5] Van Trung P, Huong T T, Van Tuyen D, et al. A multi-criteria-based DDoS-attack prevention solution using software defined networking[C]// 2015 International Conference on Advanced Technologies for Communications (ATC). Ho Chi Minh City, 2015: 308-313.

[6] Robinson R R R, Thomas C. Ranking of machine learning algorithms based on the performance in classifying DDoS attacks[C]// 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS). Trivandrum, 2015:185-190.

[7] Barki L, Shidling A, Meti N, et al. Detection of distributed denial of service attacks in software defined networks[C]// 2016 International Conference on Advances in Computing, Communications and Informatics

(ICACCD). Jaipur, 2016:2576-2581.

- [8] Tang T A, Mhamdi L, McLernon D, et al. Deep learning approach for network intrusion detection in software defined networking[C]// 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM). Marrakesh, Morocco, 2016: 258-263.

(上接第 14 页)

系统收敛速度的分析,并结合 Matlab 仿真,得出新型协议可以有效提高系统的收敛速度.

参考文献:

- [1] Vicsek T, Czirok T. Novel type of phase transition in a system of self-driven particles[J]. Physical Review Letters, 1995, 75(6): 1226-1229.
- [2] Fax J, Murray R. Information flow and cooperative control of vehicle formations [J]. IEEE Transactions on Automatic Control, 2004, 49(9): 1465-1476.
- [3] Olfati-Saber R, Murray R M. Consensus problems in networks of agents with switching topology and time-delays [J]. IEEE Transactions on Automatic Control, 2004, 49(9): 1520-1533.
- [4] Ren W, Beard R W, Atkins E M. A survey of consensus problems in multi-agent coordination[C]// American Control Conference, 2005. Proceedings of the 2005. USA, IEEE, 2005: 1859-1864.
- [5] Cao Y, Yu W, Ren W, et al. An overview of recent progress in the study of distributed multi-agent coordination [J]. IEEE Transactions on Industrial Informatics, 2013, 9(1): 427-438.
- [6] Bhat S P, Bernstein D S. Finite-time stability of continuous autonomous systems [J]. SIAM Journal on

作者简介:

马乐乐 女, (1994-), 硕士研究生. 研究方向为软件定义网络. E-mail: 1790879656@qq.com.

束永安 男, (1964-), 博士, 教授. 研究方向为无线网络、软件定义网络.

Control and Optimization, 2000, 38(3): 751-766.

- [7] Wang L, Feng X. Finite-time consensus problems for networks of dynamic agents [J]. IEEE Transactions on Automatic Control, 2007, 55(4): 950-955.
- [8] Zuo Z, Lin T. A new class of finite time nonlinear consensus protocols for multiagent systems [J]. International Journal of Control, 2014, 87(2): 363-370.
- [9] Wang L, Chen Z, Liu Z, et al. Finite time agreement protocol design of multiagent systems with communication delays [J]. Asian Journal of Control, 2009, 11(3): 281-286.
- [10] Yilun Shang. Finitetime consensus for multiagent systems with fixed topologies[J]. International Journal of Systems Science, 2009, 43(3): 499-506.
- [11] Godsil C, Royle G F. Algebraic graph theory [M]. New York ; Springer Science & Business Media, 2013.
- [12] Miller R K, Michel A N. Ordinary differential equations [M]. USA, Salt Lake City : Academic Press, 1982.

作者简介:

王晓波 男, (1988-), 博士研究生. 研究方向为多智能体系统一致性研究. E-mail: xbwang1988@163.com.

邢建春 男, (1964-), 博士, 教授. 研究方向为国防工程内部设备及智能化.