SDN 架构下基于 ICMP 流量的网络异常检测方法①

史振华[1], 刘外喜[2], 杨家烨[2]

[1](绍兴职业技术学院 信息工程学院, 绍兴 312000)
[2](广州大学 电子信息工程系, 广州 510006)

摘 要: 互联网控制报文协议(Internet Control Message Protocol, ICMP)实时地反映着网络的状态, 当网络故障或受到攻击时, ICMP 报文在整个流量中出现的概率, 以及 ICMP 流量中不同类型的报文比例等特征都会发生变化. 本文利用 ICMP 流量小的特点, 并结合 SDN 架构中控制面可对 ICMP 流量进行集中观察的优势, 采用 SVM 分类的方法, 提出一种轻量级的异常检测机制以改善异常检测的实时性和准确性---AD-ICMP-SDN (Anomaly Detection Method based on ICMP Traffic for SDN). 实验结果证明, AD-ICMP-SDN 在检测准确率和误报率等指标上展现了较好的性能.

关键词: SDN; 异常检测; 支持向量机; ICMP

Anomaly Detection Method Based on ICMP Traffic for SDN

SHI Zhen-Hua[1], LIU Wai-Xi[2], Yang Jia-Ye[2]

[1](Information Engineering Institute, Shaoxing Vocational & Technical College, Shaoxing 312000, China)
[2](Department of Electronic and Information Engineering, Guangzhou University, Guangzhou 510006, China)

Abstract: ICMP (Internet Control Message Protocol) provides a good way to observe the status of network in real time. When the network is in fault or is attacked, the percent of ICMP traffic and the percent of packet type in ICMP characteristics will change. Since the control plane in Software-Defined Networking (SDN) can observe ICMP traffic, and the value of ICMP traffic is also small, this paper proposes a lightweight anomaly detection system based on SVM classification method to improve the accuracy and real-time performance of anomaly detection system. We name it as AD-ICMP-SDN (Anomaly Detection Method based on ICMP Traffic for SDN). The experiment results have shown that AD-ICMP-SDN can effectively improve the accuracy rate and false rate.

Key words: SDN; anomaly detection; support vector machine; ICMP

1 引言

当前, 网络攻击依然经常发生, 如蠕虫、僵尸、分布式拒绝服务攻击(DDOS)等. 这些攻击消耗网络资源, 严重时导致网络瘫痪, 威胁用户安全. 因此, 如何准确地检测网络异常成为一个亟待解决的问题.

过去, 研究者们提出了各种异常检测方法, Chandola V 等人对此作了详细的综述[1]. 当前的检测方法主要分为特征模式(feature-based)和值模式(volume-based)两种. 其中, 值模式是利用整个网络流量的一些参数值的变化来检测异常, 如流量值、流的数量值等. 但单独使用这种方法很难检测到那些对整个网络流量大小几乎没有影响的低速率攻击, 也很难区分由正常应用导致的大流量还是恶意攻击导致的大流量[2]. 因此, 作为该方法的一个补充, 人们又研究基于流量特征模式的方法, 如平均包长度, IP 地址特征分布等. 在这种方法中, 如果要实现高效的检测, 选择合适的特征以及最小的特征集是一个必要前提, 但这很困难[2]. 并且, 上述两种方法的研究对象都是针对

① 基金项目:浙江省教育厅科研项目(Y201534903);广东省自然科学基金(2014A030310349,2014A030313637);2014 年大学生创新训练项目
收稿时间:2015-08-30;收到修改稿时间:2015-10-14

整个网络流量的, 而互联网流量日益变得庞大, 所以, 当前方法中依然存在着检测计算量大、检测实时性差等问题.

在互联网中, 设计 ICMP(Internet Control Message Protocol)的初衷是为了反馈网络使用过程中的故障信息, 例如, 当报文在传递的过程中某子网故障, 或中间路由器无法转发等原因导致目的地不可达, 或目的端口不可达等情形都会在故障点发出 ICMP 差错报文给源端, 而 ICMPv6 更是在此基础上将 IPv4 中的 ARP, DHCP 等功能集成于其中, 使其更加成为 Internet 必不可少的部分. 简而言之, 互联网中的 ICMP 流量充分地、实时地反映着网络的状态, 所以, 我们可以通过分析 ICMP 流量特征的变化来进行异常检测. 更重要的是, ICMP 流量很小, 例如, 在 2009 年 CAIDA 数据集中, 正常情况下的 ICMP 流量仅占整个流量的 0.18%左右. 所以, 分析 ICMP 流量的计算量会比当前基于整个流量的方法小得多, 检测效率和检测实时性都会得到很大的提高.

软件定义网络(Software-Defined Networking, SDN)[3]技术分离了网络的数据平面和控制平面, 为研发网络新应用和未来互联网技术提供了一种新的解决方案. 目前, SDN 已经广泛应用于流量工程(Traffic Engineering), 异常检测(anomaly detection), 流量统计(accounting)等网络流量管理和优化工作. 然而, 目前将 SDN 应用于异常检测方面的研究工作还处于初级阶段.

支持向量机(Support Vector Machines, SVM)是在统计学理论基础上发展起来的一种新的机器学习方法, 它通过求取能使两类样本以最大间隔分开的最优分类面来建立分类模型.

综上所述, 本文利用 ICMP 流量小的特点, 并结合 SDN 架构中控制面可对 ICMP 流量进行集中观察的优势, 采用 SVM 分类的方法, 提出一种轻量级的异常检测机制以提高异常检测的实时性和准确性—Anomaly Detection Method based on ICMP Traffic for SDN(AD-ICMP-SDN).

2 相关工作

Tootoonchian.A 等人[4]利用控制器中的路由统计信息, 分析了从不同交换机获取流统计数据的网络负载问题, 从而构建整个网络的流量矩阵. Jose.L 等人[5]在读取交换机的流统计信息后, 采用 Trie 的数据结构设计了一种识别分层高负载流的问题(hierarchical heavy hitters, HHH). Braga.R 等人[6]通过提取流统计信息中与 DDOS 攻击相关的六元组, 采用人工神经网络方法 SOM(Self Organizing Maps)进行降维处理, 从而识别 DDOS 攻击. Mehdi.S 等人[7]重新考虑了几种传统的流量异常检测方法在 SDN 中的应用, 为 SDN 中实现异常检测提供了很好的实际部署经验. 然而这些方法采用的流量特征数据较单一, 仅能针对某种特定的异常.

在 SDN 架构下, K. Giotis 等人[8]融合 OpenFlow 和 sFlow 提出一种可进行异常检测并可减轻异常的机制, 他们在控制平面设计了数据收集模块, 以及独立于数据平面的 sFlow 监测模块. 为了兼顾监测开销和异常检测精度, Ying Zhang 等人[9]提出一种 SDN 架构下对流量进行自适应抽样的 OpenWatch 机制, 其基本思路是: 基于预测的方法, 抽样粒度能够自适应地动态改变时间-空间维度. 刘文懋[10]等人提出了一个分布式的软件定义安全架构(software-defined security architecture, SDSA), 可将安全功能从 SDN 控制器解耦到专有的安全控制器和安全 APP, 提供了全局流和局部数据包层面的检测和防护, 以抵御 SDN 和虚拟化环境中的各类攻击.

Jun 等人[11]提出一种综合利用值模式和报文头部字段信息的 DDOS 检测方法. 他们首先使用流量值做初步检测, 如果流量值超过了阀值, 就针对这些可疑流量进一步分析目的 IP 地址的熵, 源端口的熵以及每秒到达的报文个数. Gao and Wang 等人[12]提出基于 K 平均算法的网络入侵检测机制, 他们用信息熵选择 K 平均算法中初始簇中心以提高检测效率.

而在此前的工作中, 我们已经发现 ICMP 流量也具有自相似特性, 并提出了利用该特性检测网络异常的方法[14].

郑黎明等人[15]针对训练模型难于获取以及部署环境的动态变化性问题, 对 SVM 分类器的选择、使用和训练方法进行了研究. Chandola.V 等人[1]对当前各种利用熵(Entropy)值进行异常检测的方法进行了综述和性能评估. 与上述方法不同的是, 本文以 ICMP 流量中源/目的 IP 地址、ICMP 报文类型等特征的熵作为检测对象, 实现轻量级的检测.

3 基本原理

网络流量异常是指网络流量产生不寻常的变化, 并且可能涵盖多条链路或者路径, 会导致流量激增, 也会导致流量的以下特征发生变化:

1) ICMP 流量中报文的 IP 地址的分布;

2) ICMP 报文类型的分布;

3) 流量中 TCP 、UDP、ICMP、GRE 四种协议的分布;

4) ICMP 流量占整个流量的比例.

考虑到上述这些特点, 本文的 AD-ICMP-SDN 实现异常检测的思路如下: 通过熵(Entropy)来描述前三个流量特征的分散程度; 而通过比例来描述第四个特征. 然后利用SVM分类的方法将正常和异常的特征区分, 达到发现异常的目的.

3.1 信息熵

在信息论中, 熵表示的是不确定性的量度. 设 X 表示一个有着 n 个变量的数据集, 这些变量分别用 x_1, x_2,…… x_n 表示. 它们在数据集中出现的概率分别为 p_1, p_2,…… , p_n, 那么 X 的熵就是如(1)式所示:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad (1)$$

式中, p_i 第 i 个变量的概率. 显然, X 中变量的概率分布越均匀, 熵值会越大, 当某个变量的概率为 1 的话, 熵值会变为 0.

3.2 支持向量机 SVM

对于给定的入侵检测审计数据 (x_1, y_1), (x_2, y_2),…,(x_n, y_n),,x 为特征空间的输入数据, $x \in R^d$, y 为类标志, $y \in \{0, 1\}$, n 为样本数, d 为输入维数. 如果 y 的值为 0, 表示对应的样本为正常; 如果 y 的值为 1, 表示对应的样本为异常, 即有入侵发生.

根据 SVM 理论, 对于待分类样本存在一个超平面, 使得两类样本完全分开, 如图 2 所示. 图中“1”和“0”分别代表两类样本, 实线为分类超平面(hyper plane). SVM 旨在寻找最优超平面(optimal hyper plane)以最大间隔分隔两类样本. 其中, 带圈的“1”和“0”样本决定类分隔面上, 称为支持向量.

求解最优超平面可看成解二次规划问题:

$$Min \frac{\|w\|^2}{2} + C(\sum_{i=1}^{N} \vartheta_i) \qquad (2)$$

$$s.t.\ y_i(x \cdot w + b) \geq 1 - \vartheta_i,\ i = 1, 2, 3 \ldots\ldots N$$

式中, ϑ 是松弛变量, C 为惩罚因子, 是人工指定的常数, 起到控制对错分样本进行惩罚程度的作用. 求解式(2)可转化为求解其对偶问题:

$$MaxW(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j k(x_i, x_j)\ \ s.t.$$

$$\sum_{i=1}^{N} y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \ldots\ldots N$$

上式中, α_i 为拉格朗日乘子; 对应于 $\alpha_i > 0$ 的向量称为支持向量; $k(x_i, x_j) = \Phi(x_i)^T, \Phi(x_i)$ 为核函数. 选择不同核函数, 可以生成不同 SVM, 常用核函数有以下 4 种:

(1)线性核函数 Linear: K(x,y)=$x \cdot y$;

(2)多项式核函数 polynomial: K(x,y)=[($x \cdot y$)+1]d;

(3) 高斯径向基核函数 RBF: K(x,y)=exp(-|x-y|^2/d^2)

(4) 二次核函数 quadratic: K(||x-xc||)=exp{-||x-xc||^2/(2*σ)^2) }

简单地说, SVM 是升维和线性化的过程. 一般情况下这会增加计算的复杂性. 但是核函数的特性, 可以隐式地将非线性的训练数据映射到高维空间中, 从而减少了计算的难度.

所以, 利用SVM分类方法进行异常检测的基本思路是: 通过训练数据, SVM 获得一个优化参数的分类器, 然后对另一组待检测数据进行分类. 若线性可分, 则直接分析处理. 若线性不可分, 则使用非线性映射算法, 即核函数. 其将低维空间中线性不可分的数据映射到高维空间中. 在原低维空间中线性不可分的样本, 在高维空间中有了线性可分的可能.

3.3 SDN 架构下的异常检测

如图 1 所示, AD-ICMP-SDN 主要包括两个组件: 流表管理和异常检测. NOX[12]是最早实现控制器功能的网络操作系统, AD-ICMP-SDN 的流表管理和异常检测也是在NOX上开发的应用. 下面分别介绍各个模块的具体功能.

流表管理组件: OpenFlow 交换机在 NOX 上注册成功之后, NOX 维护网络资源的状态: 链路性能状态、节点级的拓扑、包的到达过程、用户需求趋势等. 根据到达的数据包向 OpenFlow 交换机安装流表项.

异常检测组件: 主要包含以下 3 个模块:

1) ICMP 流量采集模块: 采集的数据包括: ICMP

报文中源/目的 IP 地址、ICMP 报文类型、传输层四种协议 TCP/UDP/ICMP/GRE、ICMP 报文数量. 流量采集的时间间隔使用固定时长. 周期太长, 则检测到异常流量并进行处理的时延也更长; 周期太短, 将会增加 NOX 和 OpenFlow 交换机的处理开销. 本文中, 该时间间隔设为 5 秒.

2) 数据训练: 按照如上所述的 SVM 原理, 需要通过对 ICMP 已知流量数据训练, 获得一个优化参数的分类器, 参数主要包括核函数和超平面计算方法.

3) SVM 分类(检测): 详见 4.2 节的分析.

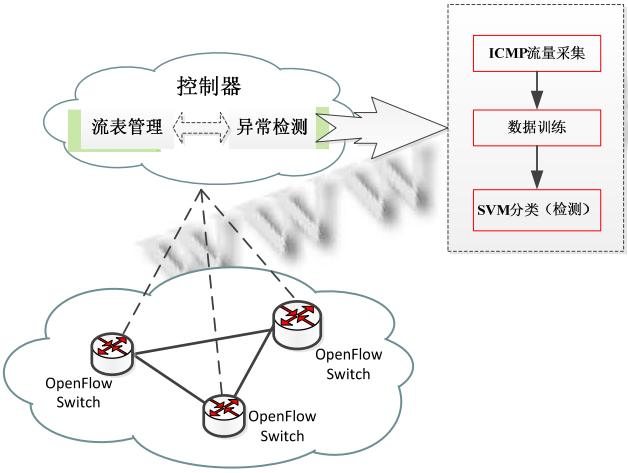


图 1　AD-ICMP-SDN 的系统流程图

4　实验与结果分析

4.1 实验方案

4.1.1 实验环境设置

我们利用 NetFPGA 开发板搭建实际的 SDN 网络, 实验网络的拓扑由 105 个节点和 386 条链路组成. 其中包括 5 台 OpenFlow 交换机和 100 台主机, 100 台主机分布于 5 个子网, 分别与五台 OpenFlow 交换机进行连接.

4.1.2 实验数据集

本文利用人工拟合流量进行测试, 流量由 3 种成分构成: 正常流量、高斯噪音流量和异常流量. 整个模拟实验运行 190 个周期, 每个周期流量注入的时间为 20 秒.

正常流量的产生: 内容的请求过程服从泊松过程, 亦即, 请求的间隔时间是指数分布; 用户的访问模式遵循 Zipf 分布. 就是说, 如果用 Pr{Ck}表示第 k 级受欢迎程度的内容被请求到的概率, 那么它遵循以下规律: $Pr\{Ck\} \propto k^{(-\alpha)}$, α 就称为是 Zipf 参数(Zipf parameter (α)), 本文中 α=0.7.

本文的异常检测目标有两种: 端口扫描和 SYN Flood, 其注入正常流量的过程为: 从第 140 到 160 个周期, 每隔 2 个周期注入一次端口扫描攻击, 持续时间 5 秒; 第 170 到 190 周期, 每隔 3 个周期注入 SYN flood 攻击, 持续时间 8 秒.

端口扫描是一种重要的攻击, 通常具有如下特征:

1) 流的数量剧增, 其中大部分有相同的源地址.

2) 大量失败响应, 多数扫描请求会失败, 导致大量的如 TCP RST、ICMP 目的地不可达等响应.

上述特征分别在 4.2.1、4.2.2 和 4.2.4 节的实验结果中得到验证.

SYN Flood 导致目的路由器为那些伪造源主机建立了大量的连接队列, 并且由于没有收到 ACK 需要一直维护着它们, 造成了资源的大量消耗而不能向正常请求提供服务, 最后甚至导致路由器崩溃. 服务器要等待超时(Time Out)才能回收已分配的资源.

上述特征分别在 4.2.2、4.2.3 节的实验结果中得到验证.

4.1.3 实验设计

本文分别利用 ICMP 流量中的源/目的 IP 地址的熵; ICMP 报文类型的熵; 传输层四种协议 TCP/UDP/ICMP/GRE 的熵; ICMP 流量的比例等来进行异常检测.

同时, 在 SVM 参数设置方面, 一次实验中只改变核函数和超平面计算方法两个参数中一个, 另一个保持默认设置. 其中核函数包括: 线性核函数 Linear, 二次核函数 quadratic, 多项式核函数 polynomial, 高斯径向基核函数 RBF. 超平面计算方法包括: 二次规划方法 QP, 序列最小优化算法 SMO, 最小二乘法 LS.

本文采用如下性能指标评估实验结果:

1) 准确率: 异常被检测为异常的概率;

2) 误报率: 正常被检测为异常的概率.

4.2 实验结果

4.2.1 ICMP 流量中源/目的 IP 地址的熵

在流量中, 基于 IP 地址可以将主机之间的连接关系分为以下四种:

CC: Concentrated origin and concentrated destination (1 对 1);

CD: Concentrated origin and dispersed destination

(1 对多);

DC: Dispersed origin and concentrated destination (多对 1);

DD: Dispersed origin and dispersed destination (多对多).

在正常流量下, 四种情况的比例会保持相对稳定: 1)如果是 CD 突然增加, 表示服务器已经被攻陷, 其需要向各个请求服务的主机发布无法服务的 ICMP 报文, 如, 端口不可达、目的地址不可达等. 2)如果是 DC 突然增加, 表示服务器正在被攻击. 如在 DDoS 攻击模式下, 攻击者雇佣很多主机假装向服务器发出了服务请求, 服务器向这些主机发出了响应. 但这些主机可能会向服务器报告 ICMP 差错报文, 如, 端口不可达、目的地址不可达等.

所以, 利用 ICMP 流量中源/目的 IP 地址的熵检测异常的思路如下:

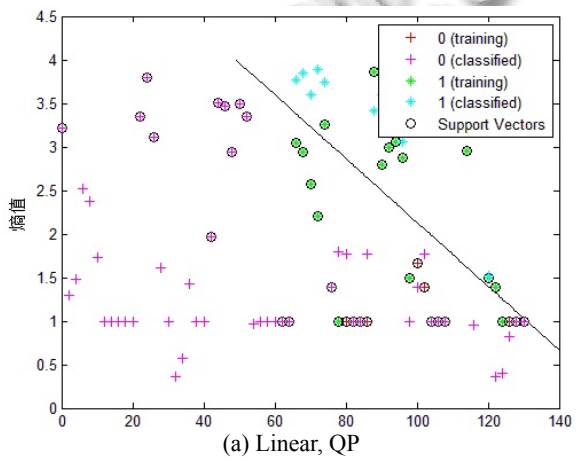
1) 由于攻击的存在, IP 地址池的规模会变大, 也更加分散一些, 导致熵的变化.

2) IP 地址空间中源地址和目的地址的映射关系会发生变化.

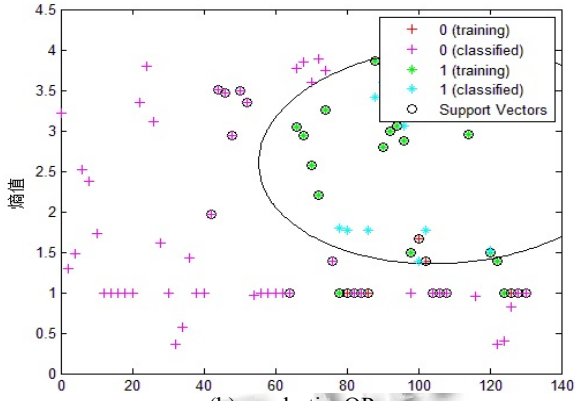
下面, 本节分别从源 IP 地址、目的 IP 地址两个方面进行分析.

①ICMP 流量报文源 IP 地址的熵

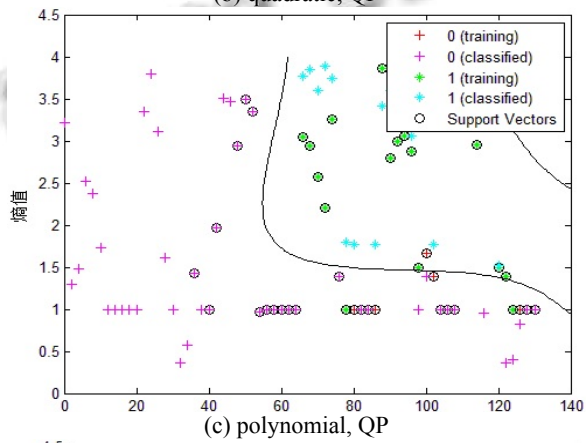
对于不同 SVM 核函数和超平面计算方法, 基于 ICMP 流量报文源 IP 地址熵的 SVM 分类结果如图 2 所示. 我们看到, 利用核函数的映射作用, 对异常和正常样本实现了非线性可分.



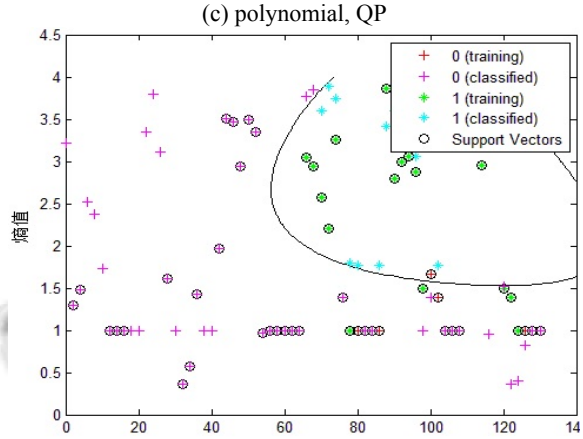
(a) Linear, QP



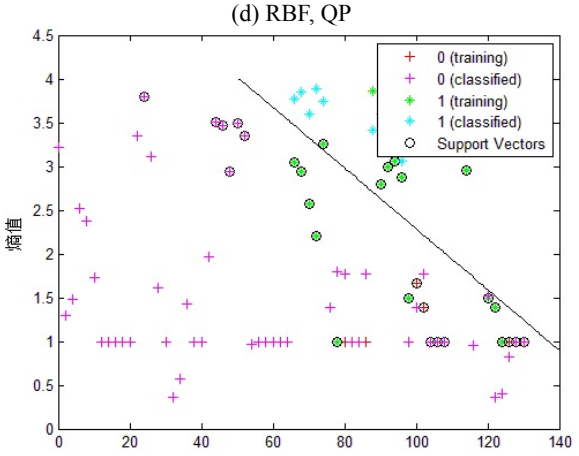
(b) quadratic, QP



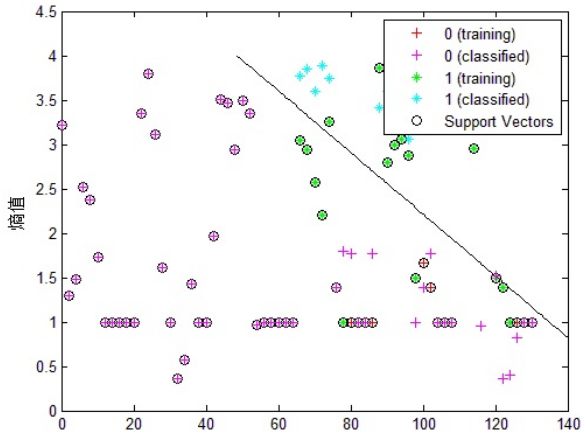
(c) polynomial, QP



(d) RBF, QP



(e) Linear, SMO



(f) Linear, LS

图 2　不同 SVM 核函数和超平面计算方法下的分类结果

对于不同的SVM核函数和超平面计算方法，利用ICMP 流量报文源 IP 地址的熵进行异常检测结果如图 3 所示. 我们可以看到，在超平面计算方法都为 QP 的情况下，多项式核函数 polynomial 展现了更好的性能，检测的准确率达到了 95%，而误报率只有 1.7%. 而对于使用相同核函数 Linear 来说，两种超平面计算方法的准确率几乎一样，而 SMO 的误报率稍微低一些.

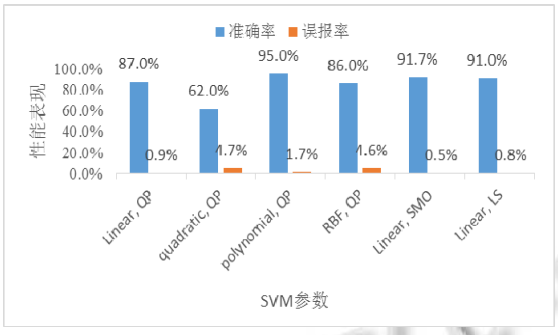


图 3　ICMP 流量中源 IP 地址的熵

②ICMP 流量报文目的 IP 地址的熵

对于不同的SVM核函数和超平面计算方法，利用ICMP 流量报文目的 IP 地址的熵进行异常检测的结果如图 4 所示. 我们可以看到，在超平面计算方法都为 QP 的情况下，多项式核函数 polynomial 展现了更好的性能，检测的准确率达到了 98.5%，而误报率只有 1.9%.

对于使用相同的核函数 Linear 来说，超平面计算方法 LS 的准确率稍微高一点点，达到 78%，而两种超平面计算方法的误报率几乎一样.

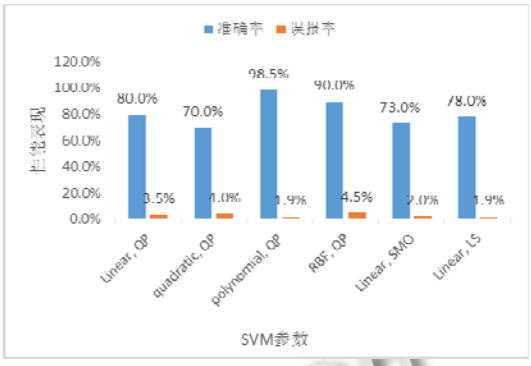


图 4　ICMP 流量中目的 IP 地址的熵

4.2.2 ICMP 报文类型的熵

ICMP 报文的主要类型有两种，即 ICMP 差错报告报文和 ICMP 询问报文.

ICMP 差错报告报主要分为 5 种类型：目的地不可达(Destination unreachable)，源点抑制(Source quench)，更改路由(Redirect)，超过生存时间(TTL exceeded)，参数问题(Parameter problem).

ICMP 询问报文主要分为 4 种类型：回送响应(Echo Reply)，回送请求(Echo Request)，时间戳请求(Timestamp Request)，时间戳应答(Timestamp Reply).

所以，利用 ICMP 报文类型的熵进行异常检测的基本思路如下：正常的情况下，ICMP 流量中各个类型报文的比例相对稳定. 如果网络发生了故障或者受到攻击，某些类型报文的数量将会增加，这样就会引起该比例的变化，其熵值从而也会发生变化. 我们可从 ICMP 各类型报文比例的熵值来检测网络是否出现异常.

上述分析可从如图 5 所示的实验结果得到验证：在 2800 秒之前，ICMP 报文类型的熵值一直维持在 2.25～2.38 之间，而当从 140 周期(2800 秒)开始注入端口扫描攻击开始的，其熵值明显地减小. 其原因正是

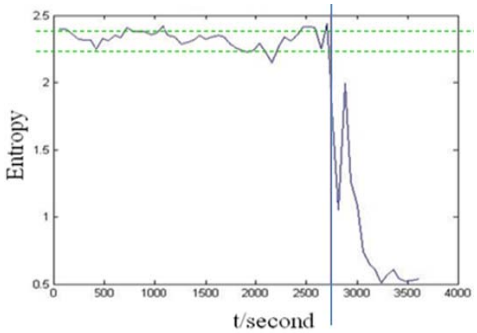


图 5　ICMP 报文类型的熵的异常变化

由端口扫描引起的目的地不可达 ICMP 报文在整体中的比例增加, 使得熵值减小, 从 170 周期开始的 SYN Flood 会导致服务器无法为正常用户提供 TCP 连接服务, 端口不可达 ICMP 报文数量则会激增, 也会使得熵值减小. 因此, 我们以 ICMP 报文类型的熵值是否在 2.25～2.38 区间来判断网络是否发生异常.

不同 SVM 参数下的实验结果如图 6 所示, 我们可以看到, 在超平面计算方法都为 QP 的情况下, 多项式核函数的检测准确率为 76%, 而线性核函数为 60%. 对于使用相同核函数 Linear 来说, 将超平面算法修改为 SMO 和 LS, 它们的检测准确率分别为 76%和 68%.

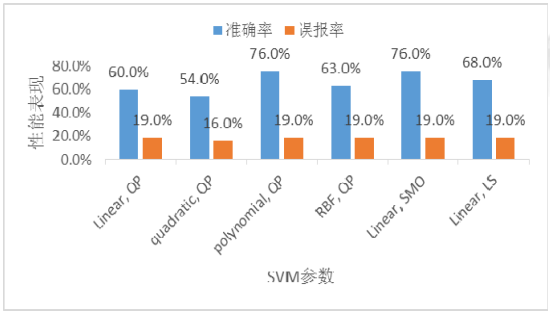


图 6 ICMP 报文类型的熵

4.2.3 传输层协议的熵

传输层主要包括 TCP 、UDP、ICMP、GRE 等四种协议. 在正常的网络中, TCP 协议数量最大, 一般都会占到 80%以上, 而 ICMP 一般维持在 0.18%左右[13]. 而当网络中发生故障时, 或者受到攻击, ICMP 报文的数量将会增加. UDP 协议是一种面向非连接的数据报, 发生 UDP 风暴攻击时, 该协议的出现概率也会提高. 在正常的情况下, 这四种协议的比例都是比较稳定的, 但由于 TCP 占比很大, 所以他们的熵值比较稳定也比较小. 当出现网络异常或攻击时, 四种协议出现的概率会表现出异常, 我们可以通过熵值的变化发现异常.

上述分析可从如图 7 所示实验结果得到验证: 与

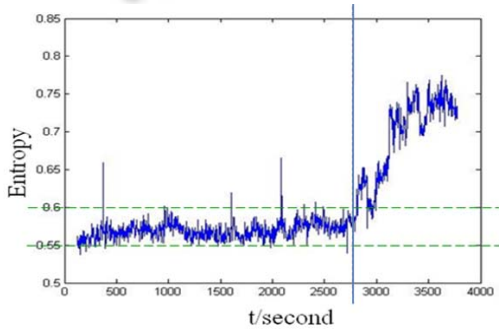


图 7 传输层协议的熵的异常变化

图 5 相对应的是, 在 2800 秒之前, 传输层协议的熵值一直维持在 0.55～0.6 的正常区间, 而从 2800 秒开始注入端口扫描以及 SYN Flood 攻击后, 由于攻击导致的 ICMP 报文数量的快速增加, 其比例从 0.18 上升到 1.5%以上, 那么整体熵值也就快速地增加. 因此, 我们以传输层协议的熵值是否在 0.55～0.6 区间来判断网络是否发生异常.

不同 SVM 参数下的实验结果如图 8 所示, 通过观察, 我们可以得到如下结论:

1) 默认设置下的结果(Linear, QP), 检测准确率是 61%, 误报率是 5.1%.

2) 使用二次核函数(quadratic, QP), 检测准确率大大上升, 达到 96%, 误报率只有 2.7%.

3) 使用多项式核函数(polynomial, QP), 检测准确率是 78%, 误报率是 4.9%.

4) 使用高斯径向基核函数(RBF, QP), 检测准确率是 84%, 误报率 4%.

5) 使用序列最小优化超平面算法(Linear, SMO) , 准确率是 71%, 误报率是 3.9%.

6) 使用最小二乘算法(Linear, LS), 使用 LS 算法后, 准确率是 65%, 误报率是 3.7%.

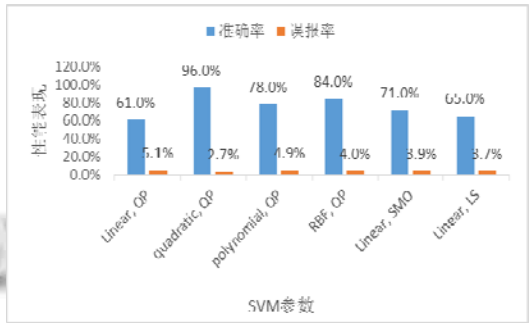


图 8 传输层协议的熵

4.2.4 ICMP 流量的比例

熵值只是描述变量的平均消息量, 是一个衡量各成分相对比例的参数. 如果出现以下情况: ICMP 协议出现的概率增加至接近正常 TCP 出现的概率, 而 TCP 的概率下降到对应正常情况的 ICMP 协议概率, 那么整体熵值将不会发生变化, 仍然是正常的熵值. 但是实际上, 网络已经受到了攻击. 如前所述, ICMP 流量正常情况下的比例一直维持在 0.18%左右, 当异常时, 其比例会激增. 所以, 我们通过分析 ICMP 报文在整个流量中比例变化, 来发现网络异常. 不同 SVM 参数

下的实验结果如图 9 所示.

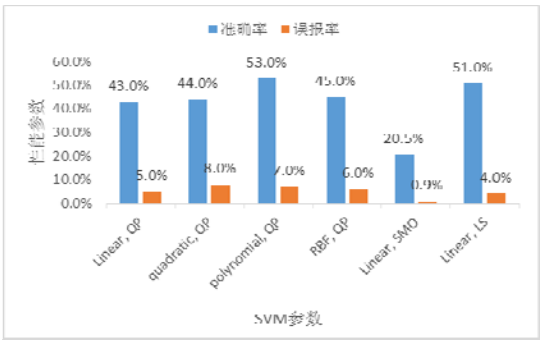


图 9 ICMP 流量的比例

5 结论

本文使用 SVM 的分类方法, 通过分析 ICMP 报文的地址空间的熵特征, ICMP 中各类型报文比例的熵, TCP/ICMP/UDP/GRE 四种协议的出现比例的熵, ICMP 在整个流量中的比例变化等四个指标的变化, 进行网络异常检测. 实验结果表明, 利用上述指标分析来判断网络异常的表现较好.

SDN 是未来互联网架构演进的主要方向, 其控制与数据相分离的思想极大地方便了网络管理. 在此架构下, 还有很多有意义的工作可以做, 如流量控制、流量平衡、防火墙设计等. 同时, 我们未来将对不同的网络异常数据采用不同的核函数和超平面计算方法, 以进一步提高网络异常检测机制的精确性.

参考文献

1 Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Computing Surveys (CSUR), 2009, 41(3): 15.

2 Özçelik İ, Brooks R R. Deceiving entropy based DoS detection. Computers & Security, 2014, 9091(6): 1–7.

3 Open Networking Fundation. SDN. https://www.opennetworking.org [2013-8-3].

4 Tootoonchian A, Ghobadi M, Ganjali Y. OpenTM: Traffic matrix estimator for OpenFlow networks. Passive and Active Measurement. Springer Berlin Heidelberg, 2010: 201–210.

5 Jose L, Yu M, Rexford J. Online measurement of large traffic aggregates on commodity switches. Proc. of the USENIX HotICE workshop. 2011.

6 Braga R, Mota E, Passito A. Lightweight DDoS flooding attack detection using NOX/OpenFlow. 2010 IEEE 35th Conference on Local Computer Networks (LCN). IEEE, 2010. 408–415.

7 Mehdi SA, Khalid J, Khayam SA. Revisiting traffic anomaly detection using software defined networking. Recent Advances in Intrusion Detection. Springer Berlin Heidelberg, 2011: 161–180.

8 Giotis K, Argyropoulos C, Androulidakis G. Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments. Computer Networks, 2014, 62: 122–136.

9 Zhang Y. An adaptive flow counting method for anomaly detection in SDN. Proc. of the Ninth ACM Conference on Emerging Networking Experiments and Technologies. ACM. 2013. 25–30.

10 刘文懋,裘晓峰,陈鹏程,等.面向 SDN 环境的软件定义安全架构.计算机科学与探索,2015,9(1): 63–70.
Liu WM, Qiu XF, Chen PC, et al. SDN oriented software-defined security architecture. Journal of Frontiers of Computer Science & Technology, 2015, 9(1): 63–70.

11 Jun JH, Ahn CW, Kim SH. DDos attack detection by using packet sampling and flow features. Proc. of the 29th Annual ACM Symposium on Applied Computing. New York, USA. 2014.

12 Gao M, Wang N. A network intrusion detection method based on improved k-means algorithm. Advanced Science and Technology Letters, 2014, 53(3): 429–433.

13 Nox Repository Website. http://www.noxrepo.org/.

14 Liu WX, Yan YE. Self-similarity and heavy-tail of ICMP traffic. Journal of Computers, 2012, 7(12): 2948–2954.

15 郑黎明,邹鹏,贾焰,等.网络流量异常检测中分类器的提取与训练方法研究.计算机学报,2012,35(4):719–730.
Zhang LM, Zhou P, Jia Y, et al. How to extract and train the classifier in traffic anomaly detection system. Chinese Journal of Computers, 2012, 35(4): 719–730.

16 Calvert KI, Doar MB, Zegura EW. Modeling internet topology. Communications Magazine, IEEE, 1997, 35(6): 160–163.