软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(8):1897−1910 [doi: 10.3724/SP.J.1001.2011.03960]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于条件随机场的 DDoS 攻击检测方法*

刘 运[1+], 蔡志平[1], 钟 平[2], 殷建平[1], 程杰仁[1]

[1](国防科学技术大学 计算机学院,湖南 长沙 410073)

[2](国防科学技术大学 电子科学与工程学院,湖南 长沙 410073)

Detection Approach of DDoS Attacks Based on Conditional Random Fields

LIU Yun[1+], CAI Zhi-Ping[1], ZHONG Ping[2], YIN Jian-Ping[1], CHENG Jie-Ren[1]

[1](College of Computer, National University of Defense Technology, Changsha 410073, China)

[2](College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: xilan724@yahoo.com.cn

Liu Y, Cai ZP, Zhong P, Yin JP, Cheng JR. Detection approach of DDos attacks based on conditional random fields. *Journal of Software*, 2011,22(8):1897−1910. http://www.jos.org.cn/1000-9825/ 3960.htm

Abstract: In recent years, the detection technology based on machine learning algorithms for distributed denial-of-service (DDoS) attacks has made great progress. However, there are still some deficiencies, which are: (1) being unable to make full use of contextual information in both the label and observed features series; (2) making too strong assumptions on the probability distribution of multiple features. Featured with the strong capability in integrating and exploiting contextual information and multiple features, the conditional random fields (CRF) model can be applied to detect DDoS attacks for effectively overcoming the above mentioned problems. A detection approach based on CRF model is proposed in this paper. First, two group of statistics are defined, which include traffic feature conditional entropy (TFCE) and behavior profile deviate degree (BPDD), to depict the characteristics of three types DDoS attacks: TCP flood, UDP flood and ICMP flood. Then, the CRF is trained to build the classification model for the addressed three types of attacks respectively. Lastly, the trained CRF models are used to identify the attacks with model inference. The experimental results demonstrate that the proposed approach can sufficiently exploit the advantages of CRF. The proposed detection approach not only can distinguish between attack traffic and normal traffic accurately, but is also more robust to resist disturbance of background traffic than the similar approaches.

Key words: distributed denial-of-service; conditional entropy; behavior profile; conditional random fields

摘 要: 近年来,基于机器学习算法的分布式拒绝服务(distributed denial-of-service,简称 DDoS)攻击检测技术已取得了很大的进展,但仍存在一些不足:(1) 不能充分利用蕴涵于标记和特征观测序列中的上下文信息;(2) 对多特征的概率分布存在过强的假设.条件随机场模型具有融合利用上下文信息和多特征的能力,将其应用于 DDoS 检测,能够有效地弥补上述不足.提出了一种基于条件随机场的 DDoS 攻击检测方法:首先,定义流特征条件熵(traffic feature conditional entropy,简称 TFCE)、行为轮廓偏离度(behavior profile deviate degree,简称 BPDD)两组统计量,对 TCP

* 基金项目: 国家自然科学基金(61070198, 60970034, 60903040)
收稿时间: 2010-04-21; 修改时间: 2010-08-13; 定稿时间: 2010-10-26

flood,UDP flood,ICMP flood 这 3 类攻击的特点进行描述;然后以此为基础,使用条件随机场,通过对其有效训练,分别为 3 类攻击建立分类模型;最后,通过对模型的有效训练,应用模型推断来完成对 DDoS 攻击的检测.实验结果表明,该方法能够充分发挥条件随机场模型的优势,准确区分正常流量和攻击流量,与同类方法相比,具有更好的抗背景流量干扰的能力.

关键词: 分布式拒绝服务;条件熵;行为轮廓;条件随机场

中图法分类号: TP309 **文献标识码:** A

分布式拒绝服务(distribution denial-of -service,简称 DDoS)攻击是当前主要的一种网络安全威胁,其通过控制多台主机向受害者发送大量突发报文,造成受害者资源被过度消耗,从而无法正常提供服务[1].较之其他攻击,DDoS 攻击发起简单、破坏性强且难以防御.因此,如何实时准确地检测 DDoS 攻击成为一个亟待研究的问题.

近年来,基于机器学习算法的 DDoS 攻击检测技术成为入侵检测领域一个新兴的研究方向.基于机器学习的思想,可将 DDoS 攻击检测看作是区别网络流状态“正常”还是“攻击”的二分类问题,其基本处理流程为:选择适当的分类特征将网络流抽象为一组特征样本序列,并为每个特征样本赋予一个标记,标记取值集合为{正常,攻击},分别代表网络流的两种状态.选择适当的机器学习算法对特征样本进行学习,构建分类模型.使用构建好的分类模型对未标记的待测样本进行分类,推断其最可能的标记.但是由于网络流量本身的突发性[2],且 DDoS 攻击流量总是混杂于庞大的背景流量当中,使得观测到的特征样本值往往呈现出不确定性和模糊性,这直接影响最终的分类精度.

选取有代表性的分类特征能够降低背景流量带来的不确定性和模糊性.然而,DDoS 攻击手段的多样化使得难以单凭某一种攻击特性概括所有类型的 DDoS 攻击.本文针对威胁最大、应用最广的泛洪攻击,深入分析 TCP flood,UDP flood,ICMP flood 这 3 类攻击的内在特点,设计了流特征条件熵(traffic feature conditional entropy,简称 TFCE)、行为轮廓偏离度(behavior profile deviate degree,简称 BPDD)两组统计量来准确表征攻击特征.

另一方面,特征样本之间的时间相关性也可应用于消除特征样本值的不确定性和模糊性.在使用机器学习算法进行攻击检测时,通常的处理方式是直接根据当前的特征观测值推断样本的标记,没有考虑相邻时刻的标记与标记之间、特征与特征之间存在的相关性.实际上,相邻时刻的样本标记之间并非独立的,而是彼此相关的.例如,相邻时刻的特征样本往往具有相同的标记.同样,相邻时刻的特征观测值之间也是相关的,甚至一个时刻的观测值可由与之相邻时刻的观测值来预测.例如,流量预测问题[3].通常把这种分别存在于标记及特征观测值序列中的相关性称为上下文信息.利用上下文信息有助于消除背景流量造成的不确定性和模糊性,提高检测精度.然而,现有学习算法没有有效地同时利用存在于标记及特征观测值序列中的这两类上下文信息,因而难以取得更好的检测结果.此外,对于分类特征是多维向量的情况,一些学习算法总是假设多特征之间满足条件独立性或是某种特定概率分布.然而当真实情况与假设不符时,这些算法的性能往往不尽如人意.

为了克服以上问题,本文提出了一种基于条件随机场(conditional random fields,简称 CRF)模型[4]的 DDoS 攻击检测方法.利用 CRF 强大的融合利用多特征及上下文信息的能力,分别对 TCP flood,UDP flood,ICMP flood 这 3 类攻击建立分类模型,再应用模型推断对待测的网络流样本进行分类,以识别攻击.实验结果表明,该方法能够准确区分正常流量和攻击流量,其检测准确率高于同类方法,特别是抗背景流量干扰的能力更强.

本文第 1 节介绍相关研究成果.第 2 节是 DDoS 攻击特性分析.第 3 节是分类特征的选取.第 4 节讨论基于 CRF 模型的 DDoS 攻击检测方法.第 5 节是实验过程和结论.第 6 节是全文总结和展望.

1 相关研究

当前,DDoS 攻击检测策略可分为两类:基于攻击流特性的检测和基于正常流特性的检测.

基于攻击流特性的检测策略通过分析攻击流特征,识别攻击流来检测攻击.文献[5]针对 DDoS 攻击时产生高流量的特性,通过计算当前流量与正常流量的偏移来识别攻击.这种方法无法区分正常的大流量和 DDoS 攻

击流量,因而误报率较高.文献[6−9]针对 DDoS 攻击时多对一的攻击模式,分别使用新源 IP 地址数目、新源目的 IP 地址对数目、流连接密度(flow connection density,简称 FCD)、流特征分布熵(traffic feature distribution entropy,简称 TFDE)这 4 种统计量来描述攻击源的分散性和攻击目标的集中性.这类方法能够有效地区分大多数攻击流与正常流,但是由于只使用了 IP 地址与端口信息,不能确定具体的 DDoS 攻击类型,并且对于仅通过少量源 IP 地址发起的攻击,这类方法的检测率不高.

基于正常流特性的检测策略通过建立正常流轮廓(profile)来检测偏离正常流轮廓的异常状况.文献[10]假设在正常情况下,网络流入数据包速率与流出数据包速率成一定比例,当发现两个方向的流速出现显著的比例失衡时,则认为有攻击发生.以上假设在某些场景中是不准确的.例如,在实时的音频/视频文件下载中,来自服务器的流速要远大于来自客户端的流速,因此该方法存在误报率高的问题.文献[11−13]基于报文头中 TCP 标志、协议等属性的统计特征为正常流量建立一个正常轮廓,当监测到当前流量的统计特征与正常轮廓发生显著偏离时,则认为有攻击发生.该方法通用性较强,能够识别基于多种攻击模式的 DDoS 攻击.然而,要建立一个包含所有正常流量特征的检测轮廓是十分困难的,这导致一些合法用户流量被误判为攻击流量.

由此可见,仅通过攻击流特性或正常流特性难以准确识别 DDoS 攻击.本文结合两种攻击检测策略:一方面,提出采用条件熵计算源 IP、目的 IP、目的端口 3 个报文头属性的相对分布,以描述 DDoS 攻击多对一的特点;另一方面,使用马尔可夫链为报文的协议和标志属性建立针对正常流量的静态轮廓以及针对实时流量的动态轮廓,再以报文序列相对于两种轮廓的对数似然概率之差来反映报文序列的异常程度.在此基础上,综合两组统计量形成分类特征向量,用于机器学习算法的学习和训练.

当前,用于 DDoS 攻击检测的机器学习算法主要有朴素贝叶斯算法(naïve Bayes,简称 NB)、支持向量机(support vector machine,简称 SVM)及隐马尔可夫模型(hidden Markov model,简称 HMM)等.文献[11]采用异常检测的方式,基于报文头属性的统计特征对正常流量建模,再采用 NB 算法给每个到达的报文打分,评估报文的合法性.文献[8]将一维的 FCD 时间序列转化为多维的 AAR 模型的参数向量序列,再使用 SVM 进行学习和分类.文献[12]基于异常检测机制,使用 HMM 模型描述报文属性的随机变化过程.这些方法在一定程度上提高了检测结果的准确性,然而它们都没有充分利用两类上下文信息.

NB 算法与 SVM 往往只能用于独立分布的数据,本身不能对表征数据相关性的上下文建模;HMM 模型可以利用标记序列的上下文,然而由于其对特征观测序列过强的条件独立性假设[14],导致不能利用观测序列中的上下文信息.此外,对于分类特征是多维向量的情况,NB 算法假设多特征之间是条件独立的;而 HMM 模型通常假设其服从多元高斯、混合多元高斯等一些常用概率分布.但真实网络流量与假设的概率分布往往存在偏差.

本文采用的 CRF 模型可以很好地解决上述学习算法存在的问题.CRF 模型是一种判别的概率无向图模型,具有较强的融合利用上下文信息和多特征的能力[4].近年来,CRF 模型已被成功地应用于图像目标识别[15]、自然语言处理[16]等问题,表现了独特的优势.在网络安全方面,文献[17]等研究开始探讨将 CRF 模型应用于网络入侵检测,并从检测精度和系统效率两个方面验证了 CRF 模型的有效性和实时性.本文针对 DDoS 攻击的特点,将 CRF 模型引入到 DDoS 攻击检测的问题中.本文采用了与文献[17]不同的建模方式,后者将每个数据单元(连接记录)抽象为一个特征序列,再通过 CRF 模型对特征之间的逻辑相关性建模.这种建模方式仅利用了数据内部的相关性,没有考虑数据之间的时间相关性.本文以特征向量的形式表达特征之间的逻辑相关性,再利用 CRF 模型对数据单元(特征样本)之间的时间相关性建模,从而更充分地利用上下文信息,提高检测质量.

2 DDoS 攻击特点分析

Moore 等人通过对受害者响应报文进行分析发现,DDoS 泛洪攻击使用 TCP 报文的比例在 94%以上,而 UDP 报文和 ICMP 报文均在 2%左右[1].因此,我们对泛洪攻击中 TCP flood,UDP flood,ICMP flood 这 3 类主要的攻击方式进行研究.由于当前的许多 DDOS 攻击并不是由具有专业技能的黑客发起的,而是由普通网民运用黑客编写的网络攻击程序发起[18],这使得攻击流量通常具有一定的规律.

本文通过对当前常见的攻击范例以及 MIT 实验室发布的正常数据集和攻击数据集[19]进行分析,得到以下

一些攻击发生时流量的特点,并以此为基础设计检测特征:

(1) 多对一映射.这主要表现在以下 3 个方面:

(a) 攻击者从多个攻击源发起对攻击目标的攻击,因此对于目标网络而言,攻击流量中的源地址相对于目的地址为多对一的映射,而正常流量则具有多对一、一对一、一对多这 3 种映射形式.在很多攻击中,攻击者会随机伪造源地址信息以隐藏身份;常会控制一个大型的僵尸网络向受害者发起攻击.此时,攻击流量较之正常流量具有更为明显的多对一特点.

(b) 合法用户在一定时间段内请求的服务比较单一,而攻击者为了快速地消耗目标资源,通常会向目标主机请求尽可能多的服务,因此常采用随机生成目的端口的发包方式,这也是很多攻击程序的默认设置.此时,目的端口与目的地址之间存在多对一的映射关系.

(c) 在针对某一特定应用服务的攻击中,攻击者会对目标主机某个固定端口发送大量服务请求,例如在对 DNS 服务的攻击中,会出现大量目的端口号为 53(DNS 的默认端口)的报文.此时,源地址与目的端口之间存在多对一的映射关系.

(2) 当发生 TCP flood 攻击时,攻击报文通常携带相同的 TCP 标志.这使得流量中携带该标志的报文比例上升,且该标志连续出现的频率增大.这些标志常见的有 SYN,RST,ACK,SYN/FIN 等.

(3) 根据 MAWI 工作小组[12]的统计,正常情况下,网络流量中 TCP 报文占绝大部分,通常在 80%以上;UDP 报文约占 10%,而 ICMP 报文则更少.当发生 UDP 或 ICMP flood 攻击时,这种关系必定会被破坏.在短时间内,某一种协议的报文比例会有明显增长,且该协议的报文连续出现的频率也会增加.

3 分类特征选取

针对 3 类攻击的特点,本文设计了流特征条件熵、行为轮廓偏离度两组统计量对 TCP flood,UDP flood,ICMP flood 这 3 类攻击分别进行描述.

3.1 流特征条件熵

信息熵是信息论中对随机变量不确定性的一种度量,而条件熵表示在已知第 2 个随机变量 X 的条件下,随机变量 Y 的不确定程度.本文利用条件熵来刻画 DDoS 的多对一特点.

定义 1. 根据信息论的定义,变量 Y 的信息熵为

$$H(Y)=\sum_i p(y_i)\log_2(p(y_i)) \tag{1}$$

其中,$p(y_i)$为变量 Y 各分量的先验概率.

定义 2. 变量 Y 关于变量 X 的条件熵定义为

$$H(Y\mid X)=\sum_j p(x_j)\sum_i p(y_i\mid x_j)\log_2(p(y_i\mid x_j)) \tag{2}$$

其中,$p(y_i|x_j)$为 y_i 关于 x_j 的后验概率.

令 *sip*,*dip*,*dport* 分别表示源地址、目的地址和目的端口这 3 个报文头属性,使用 $H(sip|dip)$,$H(sip|dport)$,$H(dport|dip)$这 3 个条件熵来分别表征 DDoS 的 3 种多对一映射的特点.这 3 个条件熵所组成的向量称为流特征条件熵(TFCE).TFCE 直观地度量了 *sip* 相对于 *dip*,*sip* 相对于 *dport*、*dport* 相对于 *dip* 的发散性,即不确定程度.

以 $H(sip|dip)$为例,TFCE 的计算方法如下:对网络流进行采样,设一个采样周期内到达的报文总数为 S,报文中不同的源地址集合为$\{sip_i|i=1,2,\ldots,N\}$,不同的目的地址集合为$\{dip_i|i=1,2,\ldots,M\}$.定义 M 维矩阵 $A[M]$:$A[i]$表示目的 IP 地址为 dip_i 的报文个数;定义 $N\times M$ 维的矩阵 $B[N][M]$:$B[i][j]$表示源地址为 sip_i、目的地址为 dip_j 的报文个数,则由公式(2)可以得到:

$$H(sip\mid dip)=-\sum_j p(dip_j)\sum_i p(sip_i\mid dip_j)\log_2(p(sip_i\mid dip_j))=-\sum_{j=1}^{M}\frac{A[j]}{S}\sum_{i=1}^{N}\frac{B[i][j]}{A[j]}\log_2\left(\frac{B[i][j]}{A[j]}\right) \tag{3}$$

其中,$p(dip_j)$表示目标地址为 dip_j 的报文集合在报文总数中所占的比率,$p(sip_i|dip_j)$表示在目的地址为 dip_j 的报文

集合中,源地址为 sip_i 的报文所占比率.同理可以计算得到 $H(sip|dport)$和 $H(dport|dip)$.

TFCE 不但能够描述 DDoS 攻击多对一映射的特点,而且能够有效地反映 DDoS 攻击流量增长.

以 $H(sip|dip)$为例,由公式(3)可以得到:

$$H(sip\,|\,dip)=\sum_{j=1}^{M}p(dip_j)H(sip\,|\,dip_j)=\sum_{j=1}^{M}w_jH(sip\,|\,dip_j) \tag{4}$$

其中,$w_j=p(dip_j)=A[j]/S$.因此,$H(sip|dip)$可以表示为 $H(sip|dip_j)$的加权和.

设被攻击主机的 IP 地址为 dip_k.一方面,攻击者采用的源地址越多,在目的地址为 dip_k 的报文集合中,源地址的分布就越分散,因此,公式(4)中的项 $H(sip|dip_k)$越大,$H(sip|dip)$就越大.因此,条件熵能够描述 DDoS 攻击多对一映射的特点.另一方面,攻击流量越大,$A[k]/S$ 就越大,即项 $H(sip|dip_k)$的权值 w_k 越大.由于大多数攻击的源地址比正常流量更趋向分散,有 $H(sip|dip_k)>H(sip|dip_t),t\neq k$,从而可推算出 $H(sip|dip)$越大.因此,条件熵能够反映 DDoS 攻击流量的增长.

3.2 行为轮廓偏离度

正常情况下,网络流量的统计特征在不同时间段内应具有相对的稳定性.当受到 DDoS 攻击时,流量的统计特征会发生变化,表现为某类报文的比例上升、各类型报文的转换频率改变等.为了描述这些变化,本文针对协议、TCP 标志这两个报文属性,采用马尔可夫链分别对正常流量及监控的实时流量建立行为轮廓.正常流量的行为轮廓是对一个较长时间段的数据经过离线统计后生成的结果,称为静态轮廓,记为Ω_1;实时流量的行为轮廓随着新的流量到来而不断更新,称为动态轮廓,记为Ω_2.Ω_2 与Ω_1 的相似性越高,Ω_2 的正常度就越高;反之,则异常度越高.

对一个行为轮廓Ω,可由一个三元组(S,P,Π)来表示.其中,$S=\{s_1,s_2,\ldots,s_n\}$是系统状态集合;$P=(p_{ij})_{n\times n}$ 为状态转移概率矩阵,p_{ij} 表示系统 t 时刻状态为 s_i、t+1 时刻状态为 s_j 的概率;$\Pi=[\pi_1,\pi_2,\ldots,\pi_n]$为初始状态概率向量,$\pi_i$ 表示系统初始时刻状态为 s_i 的概率.

为协议、TCP 标志属性分别建立静态轮廓和动态轮廓,以反映不同类型 DDoS 攻击的特点.对于协议属性,$S=\{TCP,UDP,ICMP\}$;对于 TCP 标志属性,$S=\{0,1,\ldots,63\}$.P 和Π的取值可以从流量数据中学习得到:$p_{ij}=N_{ij}/N_i$,$\pi_i=N_i/N$.N_{ij} 表示流量中第 t 个报文的属性值为 s_i、第 t+1 个报文的属性值为 s_j 的报文对数目;N_i 表示流量中属性值为 s_i 的报文数目;N 为总报文数.假设对网络流量以Δt 为周期进行实时监控,一个周期内到达的报文按时间顺序排列成一个序列 $O=O_1,O_2,\ldots,O_N$,则以 O 为训练数据,依照上述方法可以生成动态轮廓Ω_2.静态轮廓Ω_1 可由网络的正常历史流量训练得到.

在获得网络流量的静态轮廓$\Omega_1=(S,P,\Pi)$和动态轮廓$\Omega_2=(S,P',\Pi')$之后,可分别计算报文序列 O 相对于两个轮廓的出现概率.设报文 O_i 的属性值为 m_i,则由马尔可夫链的时齐性及马尔可夫性[20],序列 O 相对于Ω_1 的出现概率为 $p(O,\Omega_1)=\pi_{m_1}\prod_{i=2}^{N}p_{m_{i-1}m_i}$,相对于$\Omega_2$ 的出现概率为 $p(O,\Omega_2)=\pi'_{m_1}\prod_{i=2}^{N}p'_{m_{i-1}m_i}$.

进一步地,定义Ω_1,Ω_2 关于报文序列 O 的行为轮廓偏离度(BPDD)为

$$BPDD=|\log_2 p(O,\Omega_2)-\log_2 p(O,\Omega_1)|=\left|\log_2\pi'_{m_1}-\log_2\pi_{m_1}+\sum_{i=2}^{M}(\log_2 p'_{m_{i-1}m_i}-\log_2 p_{m_{i-1}m_i})\right| \tag{5}$$

对于协议和 TCP 标志两个报文属性,分别计算出它们的 BPDD,记作 $BPDD[p]$和 $BPDD[f]$.

从公式(5)可知,BPDD 的大小取决于Ω_2 与Ω_1 的相似性及报文序列 O 的长度.对于攻击报文序列而言,序列越长,则Ω_2 与Ω_1 的差异越大,BPDD 的值也就越大;对于正常报文序列而言,序列越长,则Ω_2 与Ω_1 越相似.因此,在长度相当的条件下,正常报文序列的 BPDD 值要小于攻击序列的值.

BPDD 综合考虑了轮廓相似性和序列长度因素,有助于克服因监控周期过短导致的动态轮廓训练不充分的问题.相对于静态轮廓,动态轮廓的统计生成时间要短得多,当某些监控周期内统计到的报文数目很少时,正常流量的Ω_2 与Ω_1 也可能存在较大差异.对于本文研究的泛洪攻击而言,攻击报文序列的长度一般是比较大的.BPDD 融合了序列长度因素,可以避免将这些统计特征显著偏离了静态轮廓的正常短序列误判为攻击.

3.3 特征选取

TFCE 和 BPDD 分别从不同角度对 DDoS 攻击进行了描述,它们的有效组合构成了一个特征二元组.其中,有些特征仅对某类攻击有效,而对其他攻击不敏感.过多的不相关特征会给机器学习算法带来过拟合问题和计算效率降低问题,一种最直接的解决这些问题的办法是,结合不同攻击的特点,从中选取有针对性的分类特征.

当出现 TCP flood 攻击时,流量中 TCP 标志的统计特征会发生改变,相应地,*BPDD*[*f*]会有明显增长.但 *BPDD*[*p*]不会有明显变化,因为正常流量中绝大多数都是 TCP 报文.因此,*BPDD*[*f*]比 *BPDD*[*p*]更能反映其特点.对于 UDP 或 ICMP flood,当攻击发生时,*BPDD*[*p*]会显著增大.此外,3 类攻击都具有多对一映射的特点,这可以通过 TFCE 来描述.特别地,由于 ICMP 报文通常不包含端口信息,因此 ICMP flood 攻击只有源地址到目的地址一种多对一映射关系,这样,只需计算 $H(sip|dip)$即可.针对 3 类攻击的特征选取结果见表 1.

Table 1 Selection of features for three types of attacks

表 1 3 类攻击的特征选取

Types of attacks	TFCE	*BPDD*[·]
TCP flood	ALL	*BPDD*[*f*]
UDP flood	ALL	*BPDD*[*p*]
ICMP flood	$H(sip\|dip)$	*BPDD*[*p*]

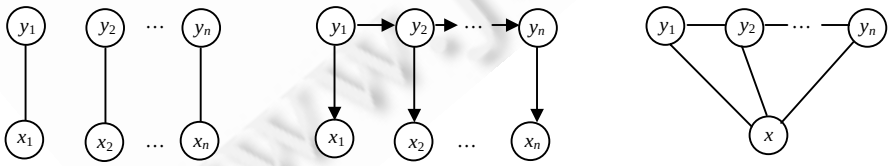
4 基于 CRF 的 DDoS 攻击检测方法

4.1 攻击检测问题描述

对网络流进行时间间隔为Δt的采样,计算每次采样的分类特征值,获得一个特征样本序列 $x=\{x_i,i=0,1,\ldots,n\}$.DDoS 攻击的检测问题可归结为给定一个特征样本序列 x,估计其对应的状态标记序列 $y=\{y_i,i=0,1,\ldots,n\}$,其中,y_i为 x_i 对应的标记,其取值集合为{+1,–1}:“+1”代表“正常”状态,“–1”代表“攻击”状态.估计 x 对应的最优标记 $\hat{y}$ 的过程通常归结为寻找极大后验(maximum a posteriori,简称 MAP)估计的过程,即 $\hat{y}=\arg\max_y p(y|x)$.由贝叶斯定律,后验概率可以等价地写成

$$p(y|x)\propto p(y,x)=p(y)p(x|y) \tag{6}$$

在现有的用于 DDoS 攻击检测的机器学习算法中,NB 算法与 SVM 的标记过程通常是在单个观测样本上进行的,即对 $p(y_i|x_i)$建模,不考虑与 x_i 相邻的其他样本,因此难以利用标记序列和观测序列中的上下文信息,如图 1(a)所示.而 HMM 模型作为一种生成式(generative)模型,分别对公式(6)中的状态先验分布 $p(y)$和似然分布 $p(x|y)$建模.由马尔可夫性质,有 $p(y_i|y_j,j\neq i)=p(y_i|y_{i-1})$,可知样本 x_i 的标记依赖于相邻样本 x_{i-1} 的标记.因此,HMM 可以利用标记序列的上下文信息.然而,HMM 模型在对似然分布 $p(x|y)$建模时,需要枚举所有可能的观测序列,这是十分困难的.为了保证计算的可行性,HMM 模型通常需要假定观测过程条件独立,即 $p(x|y)=\prod_i p(x_i|y_i)$.这使得 HMM 不能利用观测序列的上下文信息,如图 1(b)所示.此外,对于分类特征为多维向量的情况,通常假设 $p(x_i|y_i)$服从多元高斯、混合多元高斯等一些常用概率分布,而网络流量动态多变的特点使得分类特征很难严格服从假定分布.



(a) SVM (b) HMM (c) CRF

Fig.1 Graphic models of SVM, HMM and CRF

图 1 SVM,HMM,CRF 的图形模型

为了更好地利用上下文信息和多特征、提高检测的准确性,本文将 CRF 模型引入 DDoS 攻击的检测问题中.CRF 模型是一种判别式模型,它直接对公式(6)中的后验概率 $p(y|x)$建模,可以克服以上学习算法存在的不足.

4.2 CRF模型

CRF 是一种判别式概率无向图模型,它具有表达数据长距离依赖性和交叠性特征的能力,并且由于使用全局优化技术,克服了最大熵马尔可夫模型中的标记偏置问题[4],是一种很好的处理序列数据分割与标记问题的统计机器学习模型.Lafferty[4]给出的 CRF 的正式定义如下:

定义 3. 设 $G=\langle V,E\rangle$是一个无向图,$Y=\{y_v|v\in V\}$是以 G 中结点 v 为索引的随机变量 y_v 构成的集合.如果给定 X,每个随机变量 y_v 的条件概率在图 G 中满足马尔可夫性质: $p\{y_v|\{y_w\}_{w\neq v},X\}=p\{y_v|\{y_w\}_{w\sim v},X\}$,则称$(X,Y)$是一个条件随机场(CRF),其中,$w\sim v$ 表示 w 与 v 在图 G 中为相邻结点.

由 Hammersley-Clifford 定理[21]可知,在给定观测序列 x 的条件下,标记序列 y 的后验概率分布 $p(y|x)$具有 Gibbs 分布形式:

$$p_\theta(y\mid x)=\frac{1}{Z(x)}\exp\left(\sum_{c\in C}\varphi_c(y_c,x,\theta)\right) \tag{7}$$

其中, $Z(x)=\sum_y\exp\left(\sum_{c\in C}\varphi_c(y_c,x,\theta)\right)$为归一化因子;当子集 $c\subseteq G$ 中的每对结点总相邻时,称 c 是一个基团(clique),C 为图 G 所有基团的集合;$\varphi_c(\cdot)$为定义在基团上的带参数θ的势函数.通过势函数实现对图 G 中局部上下文信息建模,然后将局部上下文信息通过图结构传播,间接得到大范围的上下文信息.

当用该模型来建模序列数据时,最常用的假设为图 G 中标记变量 y 的形状为一阶链形式,如图 1(c),则基团集合 C 即为 G 中的结点和连接两个相邻结点的边.相应地,模型中的势函数$\varphi_c(\cdot)$包括定义在结点上的单位置势函数与定义在边上的双位置势函数两种形式.此时,公式(7)可以具体化为

$$p_\theta(y\mid x)=\frac{1}{Z(x)}\exp\left(\sum_{i\in V}(\varphi_i(y_i,x,\lambda)+\varphi_{i,i-1}(y_i,y_{i-1},x,\mu))\right) \tag{8}$$

其中,$\varphi_i(\cdot)$和$\varphi_{i,i-1}(\cdot)$分别为单位置势函数和双位置势函数,$\theta=\{\lambda,\mu\}$为模型参数.

4.3 基于CRF模型的DDoS攻击检测

公式(7)给出了 CRF 模型的一般数学描述,从中可以看出,CRF 的具体实现需要结合特定的应用背景.本文提出的基于 CRF 的 DDoS 攻击检测模型的实现流程包含以下 4 个重要步骤:(1) 图结构选择:确定图 G 的结构;(2) 势函数定义:结合 DDoS 攻击的特点定义势函数$\varphi_c(\cdot)$的具体形式;(3) 模型训练:利用训练数据估计参数θ的取值;(4) 模型推断:应用训练好的模型推断待测样本的标记.其中,前 3 个步骤是离线处理过程,第 4 步是在线处理过程.

4.3.1 图结构的选择

在 DDoS 攻击检测中,某时刻样本的标记可能与之前多个时刻的样本标记相关,因此,理论上可以定义高阶的 CRF 模型,实现对大范围的上下文建模,提高算法的检测性能.但是,高阶 CRF 模型通常需要更加复杂的计算.后续的实验表明,采用一阶的 CRF 模型足以取得满意的结果.因此,兼顾算法的检测性能以及检测效率,本文假设某时刻标记只与前一时刻的标记相关,即满足一阶马尔可夫性质.相应的 CRF 图结构 G 为一阶链形式,如图 1(c)所示,模型的数学描述如公式(8).观测序列 x 为特征样本的时间序列.

4.3.2 势函数的定义

势函数用以对图 G 中的局部上下文建模,定义合适的势函数是 CRF 模型中的一个关键问题.在本文定义的 CRF 模型中,主要涉及单位置和双位置势函数的定义.

单位置势函数通常定义为一些判别的分类器.为了能够无缝地(seamlessly)融入到 CRF 模型中,这些判别的分类器应具有闭合的数学表达形式.本文采用对数回归(logistic regression,简称 LR)分类器定义单位置势函数[22].对于一个二分类问题,LR 写成

$$\varphi_i(y_i,x,\lambda)=\log_2\left(\frac{1}{1+\exp(-y_i\lambda^T x_i)}\right) \tag{9}$$

其中,x_i 为位置 i 处的特征观测值,$\lambda=[\lambda_1,\lambda_2,\ldots,\lambda_d]^T$ 为模型参数向量,d 为向量 x_i 的维数.

除了具有闭合的数学表达式以外,采用 LR 分类器定义单位置势函数的另一个重要原因是,它对不同类型观测数据的适应性:与传统的线性判别分析(linear discriminant analysis,简称 LDA)相比,LR 不需要观测数据满足高斯分布等限制条件,因此可以适应更多类型的观测数据.

双位置势函数可用来实现对给定观察数据、相邻节点的标记间的上下文信息建模.一般来说,为了不断地消耗受害者的资源,DDoS 攻击流量在时间上具有连续性,因此,相邻时刻的样本往往具有相同的标记.为了对这种重要的上下文信息建模,本文采用一种推广的 Ising/Potts 模型[23]来定义双位置势函数:

$$\varphi_{i,i-1}(y_i,y_{i-1},x,\mu)=y_i,y_{i-1}\mu^T g_{i,i-1}(x) \tag{10}$$

其中,$g_{i,i-1}(x)$表示从观测序列 x 提取的位置对$(i,i-1)$且维数为 q 的特征向量,它通过将特征向量 x_i 和 x_{i-1} 首尾相连而得到;$\mu=[\mu_1,\mu_2,\ldots,\mu_q]^T$ 为模型参数向量.Ising/Potts 模型会对两个相邻位置标记不相同的情况产生一个惩罚代价$\mu^T g_{i,i-1}(x)$,这样可使得最终的检测结果趋于平滑.

4.3.3 模型训练和推断

确定完模型的基本形式后,要使用 CRF 模型进行攻击检测,通常还要进行如下两步:第 1 步是模型训练,通过对训练样本集的学习,确定模型中的参数,形成 CRF 分类模型;第 2 步是模型推断,利用生成的 CRF 模型对标记未知的样本进行分类,达到攻击检测的目的.

不同于其他学习算法,CRF 的训练和推断是以一个样本序列为单位的.CRF 模型的训练可描述为:给定训练特征样本集 $D=(x^m,y^m)_{m=1}^M$ 及预定义的势函数,采用某种优化算法调整模型的参数θ,使其最终满足事先定义的某种训练标准.其中,x^m 为第 m 个训练样本序列,y^m 为 x^m 对应标记序列.本文采用最常用的极大似然估计作为参数估计标准.极大似然估计选择使对数-似然函数 $L(\theta)=\log_2\left(\prod_{m=1}^M p(y^m\mid x^m,\theta)\right)$ 极大化的参数取值 $\hat{\theta}$ 作为θ的最优估计.

模型推断任务可描述为:给定一个待标记的特征样本序列 x,使用训练得到的 CRF 模型来推断它最大可能的状态标记序列 $\hat{y}$,即

$$\hat{y}=\arg\max_y p_\theta(y\mid x)=\arg\max_y\sum_{i\in V}(\varphi_i(y_i,x,\lambda)+\varphi_{i,i-1}(y_i,y_{i-1},x,\mu)) \tag{11}$$

当前,常用的模型训练算法有 IIS[4],GTB[24],Quasi-Newton[25]等;模型推断算法有 Viterbi[4],BP[26]等.本文采用的训练算法为 Quasi-Newton,推断算法为 Viterbi.对于图结构为一阶链的 CRF 模型,Viterbi 算法不仅可以获得 MAP 估计的全局最优解,而且其时间复杂度比同类算法更低,这对攻击检测的实时性需求是十分有益的.

对于 TCP flood,UDP flood,ICMP flood 这 3 类攻击,由于选取的分类特征不同,因而训练样本集也不同.分别从 3 个样本集中学习得到一个 CRF 分类模型,然后将它们组合起来,以应对不同的攻击行为.

5 实 验

5.1 实验数据集及评价标准

本文实验采用了 4 个攻击数据集,其中两个取自 MIT 林肯实验室 2000 年的 DDoS 数据集 LLDoS1.0 和 LLDoS2.0.2[19].它们是一段 TCP flood 攻击流量,攻击报文的源地址和目的端口均为随机生成,携带的标志为 ACK.另外两个攻击数据集来源于本文设计的攻击实验.实验运行在一个 100Mbps 的共享局域网环境下,由 3 台安装了 TFN2K 攻击软件的主机同时向 1 台受害主机发起攻击,攻击一共发起两次,一次为 UDP flood,一次为 ICMP flood,持续时间各约 2 分钟,报文流速分别约为每秒 1.2×10^4 个和每秒 7.5×10^3 个.为了检验本文方法对由小范围地址发起的攻击的识别能力,两次攻击均采取真实源地址和固定目的端口的发包方式.使用 pcap 在受害主机方捕获所有报文,得到两个攻击数据集.由于以上攻击数据集仅包含攻击流量,为了实验的一般性,在攻击

数据集的基础上添加背景流量,背景流量分别取自林肯实验室的正常流量数据集[19]及某校园骨干网流量.同时,使用背景流量分别为协议和标志属性建立静态轮廓.

本文使用检测率(detection ratio,简称 DR)、误报率(false alarm ratio,简称 FR)和总错误率(error ratio,简称 ER)等 3 个评价指标对实验结果进行评估.

设 *TP* 表示被正确标记的正常测试样本数,*FP* 表示被错误标记的正常测试样本数,*TN* 表示被正确标记的攻击测试样本数,*FN* 表示被错误标记的攻击测试样本数,则

$$DR=\frac{TN}{TN+FN},\ FR=\frac{FP}{TP+FP},\ ER=\frac{FN+FP}{TP+FP+TN+FN} \tag{12}$$

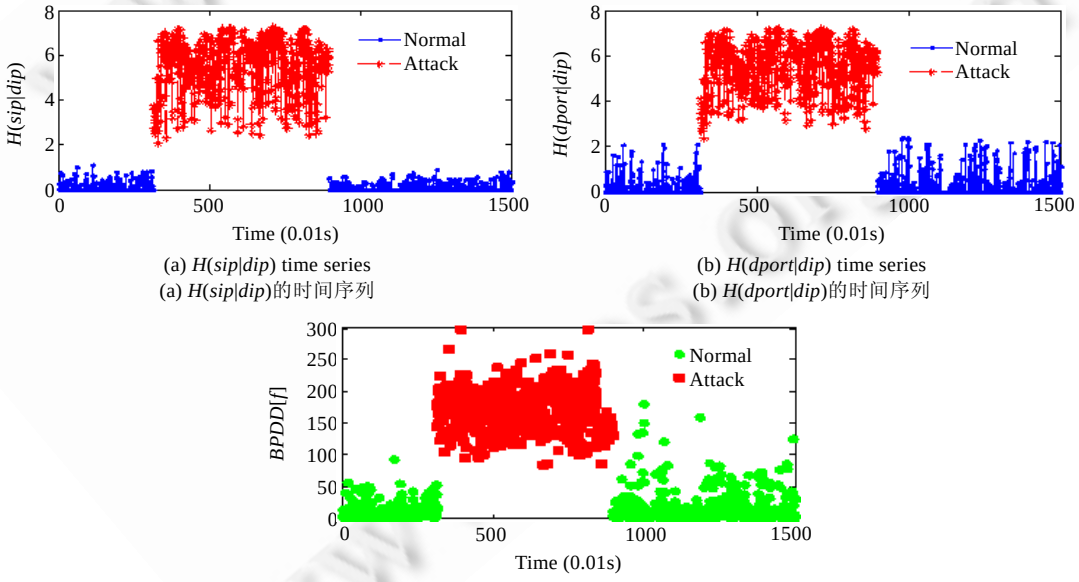
本文的实验分 3 部分,分别对 TCP flood,UDP flood,ICMP flood 这 3 类攻击进行检测,并将检测结果与 SVM,HMM 两种学习算法的检测结果进行比较,以验证本文提出的基于 CRF 的检测方法具有更好的序列建模和识别能力.实验中使用 CRFall[27]工具包来进行 CRF 模型训练和推断.CRFall 也包括对 HMM 模型的实现,多特征之间假设满足混合高斯分布.利用 LIBSVM[28]工具包来完成针对 SVM 的测试.

5.2 对TCP flood攻击的检测

对背景流量和攻击流量进行混合采样,计算每次采样的分类特征值,获得正负两类特征样本集.为了检验本文方法抗背景流量干扰的能力,通过延长采样周期的方式来加大背景流量.背景流量的采样周期(记为 *T*)从 1s 逐步增加至 5s,每次以 1s 的间隔递增,攻击流量的采样周期(记为 *t*)固定为 0.01s.

将 LLDoS1.0 与 LLDoS2.0.2 两个数据集一个用于模型训练,一个用于模型测试.将对前者采样所得的特征样本作为训练数据集的负样本,将对后者采样所得的特征样本作为测试数据集的负样本.

将得到的样本集按时间划分为若干个序列,以样本序列为单位进行 CRF 训练和推断.每个序列的长度应当适中,太长会加大检测的时间,太短则影响上下文信息的利用.本文把训练和测试样本序列的长度设置为 50,则相应的检测时间为 50×0.01=0.5(s),这对持续时间一般为几分钟至几十分钟的 DDoS 攻击是合理的.按照上述不同周期的方式采样,得到 5 组数据,每组数据的训练集和测试集均为连续的 1 500 个样本,即 30 个样本序列.图 2 给出了 *T* 为 1s 时,训练集中 3 个分类特征的时间序列曲线.



(a) $H(sip|dip)$ time series
(a) $H(sip|dip)$的时间序列

(b) $H(dport|dip)$ time series
(b) $H(dport|dip)$的时间序列

(c) $BPDD[f]$ time series
(c) $BPDD[f]$的时间序列

Fig.2 Three feature series in training set when T=1s in the detection of TCP flood attacks

图 2 TCP flood 攻击检测中,当 T=1s 时,训练集的 3 个分类特征时间序列

由于攻击报文的源地址与目的端口均为随机生成,因此两者相对于目的地址具有多对一的特点.从图 2(a)、图 2(b)可以看出,即使背景流量的采样间隔为攻击流量的 100 倍,攻击时段的 $H(sip|dip)$与 $H(dport|dip)$两个条件熵的值与正常时段的取值也有明显差异.此外,从图 2(c)可以看出,由于攻击报文携带的是 ACK 标志,因此流量中 ACK 报文的比例及连续出现的频率都增大.这使得标志属性的动态轮廓偏离了静态轮廓,导致攻击流量的 $BPDD[f]$取值比较高,基本上在 100 以上;而正常流量除了少量样本以外,大部分取值都在 70 以下,这表明 $BPDD[f]$能够较好地区分正常流量和攻击流量.

使用训练集对 CRF 模型进行训练,再用经过训练的模型对测试集进行检测,并将检测结果与 HMM,SVM 两种算法的检测结果进行比较.表 2 给出了 3 种算法分别对 5 组数据的检测结果.

从表 2 可以看出,当 T 较小时,3 种算法的检测效果相差不大.但随着 T 的增大,CRF 的优势逐渐明显,与另两者相比,具有更高的检测率和更低的总错误率;并且检测精度不易受背景流量影响,即使在 T 为 5s 时,仍有 98.5%以上的检测率.这说明 CRF 融合利用多特征和上下文信息的能力确实有助于更好地识别 DDoS 攻击.

Table 2 Detection results of three algorithms on TCP flood attacks (%)

表 2 3 种算法对 TCP flood 攻击的检测结果(%)

		T (×0.01s)				
		100	200	300	400	500
CRF	*DR*	100.0	100.0	99.73	99.48	98.69
	FR	0.00	0.14	0.27	0.14	0.14
	ER	0.00	0.07	0.27	0.33	0.73
HMM	*DR*	100.0	99.21	97.51	97.12	91.75
	FR	0.27	0.54	0.82	1.63	2.04
	ER	0.13	0.67	1.67	2.33	5.20
SVM	*DR*	99.61	98.17	97.38	96.73	93.60
	FR	0.14	0.27	0.54	0.27	1.22
	ER	0.60	1.07	1.60	1.80	3.87

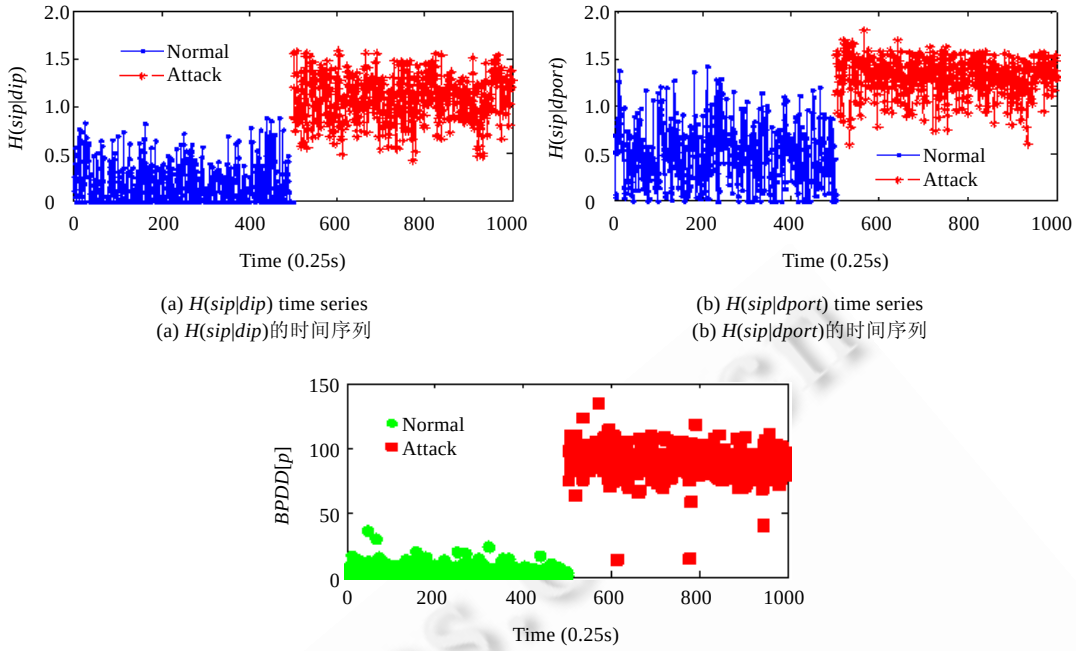
5.3 对UDP flood攻击的检测

将从攻击实验中获得的 UDP flood 攻击数据集与背景流量进行混合采样.将背景流量的采样周期固定为 5s,并逐步加大攻击流量的采样周期,每次以 5ms 的间隔从 5ms 逐步增长至 25ms.采样得到 5 组数据,每组数据的训练集和测试集都取连续的 1 000 个样本,即 20 个样本序列.

由于攻击时使用的是真实源地址和固定目标端口,因此攻击特征为:源地址相对于目的地址及目的端口分别为多对一映射;UDP 报文的比例上升且连续出现的频率增加.以 t 为 25ms 时获取的训练样本集为例,各分类特征的时间序列变化如图 3 所示.

由图 3(a)和图 3(b)可以看出:一方面,条件熵 $H(sip|dip)$与 $H(sip|dport)$在攻击时段的取值总体上比正常时段的取值要大,说明对于即使只使用少量源地址的攻击,TFCE 依然是有效的;另一方面,较大的背景流量使得分类特征值具有一定的模糊性和不确定性,表现为正常时段内某些时刻的特征取值与攻击时段内的特征取值具有相似性.例如图 3(b)中,正常时段中某些时刻的 $H(sip|dport)$取值也比较高.但是正常时段内高特征值连续出现的频率远小于攻击时段的频率,这是一种有用的上下文信息.CRF 模型能够有效利用这种上下文信息,提高检测精度.从图 3(c)可以看出:正常样本的 $BPDD[p]$取值都在 45 以下,分布比较集中;而攻击样本中,除了个别情况以外,取值都在 70 以上.两者差异比较明显,其原因在于,$BPDD[p]$综合考虑了行为轮廓相似性和报文序列长度两方面因素.对于较长的正常报文序列,由于其动态轮廓与静态轮廓较为相似,因此 $BPDD[p]$取值与攻击序列存在差异;而对于较短的正常报文序列,即使其动态轮廓与静态轮廓偏差较大,但因为序列长度的限制,$BPDD[p]$取值也比较小.

分别使用 CRF,HMM,SVM 这 3 种算法对 5 组测试数据进行检测,检测结果见表 3.比较 3 种算法的检测结果可以看出,使用 CRF 利用上下文信息能够在一定程度上降低背景流量造成的模糊性,改善检测质量.特别地,在 t 为 5ms 的测试数据中,与 HMM 和 SVM 相比,CRF 的检测率分别提高了 9.2%和 11.1%.



(a) $H(sip|dip)$ time series
(a) $H(sip|dip)$的时间序列

(b) $H(sip|dport)$ time series
(b) $H(sip|dport)$的时间序列

(c) $BPDD[p]$ time series
(c) $BPDD[p]$的时间序列

Fig.3 Three feature series in training set when t=25ms in the detection of UDP flood attacks

图 3 UPD flood 攻击检测中,当 t=25ms 时,训练集的 3 个分类特征时间序列

Table 3 Detection results of three algorithms on UDP flood attacks (%)

表 3 3 种算法对 UDP flood 攻击的检测结果(%)

		T (ms)				
		5	10	15	20	25
CRF	*DR*	95.78	97.80	98.59	99.40	99.80
	FR	3.13	0.80	0.40	0.40	0.00
	ER	3.70	1.50	0.90	0.50	0.10
HMM	*DR*	86.56	91.42	94.55	95.62	97.80
	FR	6.68	1.20	0.99	0.60	0.40
	ER	10.20	4.90	3.20	2.50	1.30
SVM	*DR*	84.64	94.81	96.36	97.81	98.20
	FR	6.89	3.61	3.17	2.61	1.80
	ER	11.30	4.40	3.40	2.40	1.80

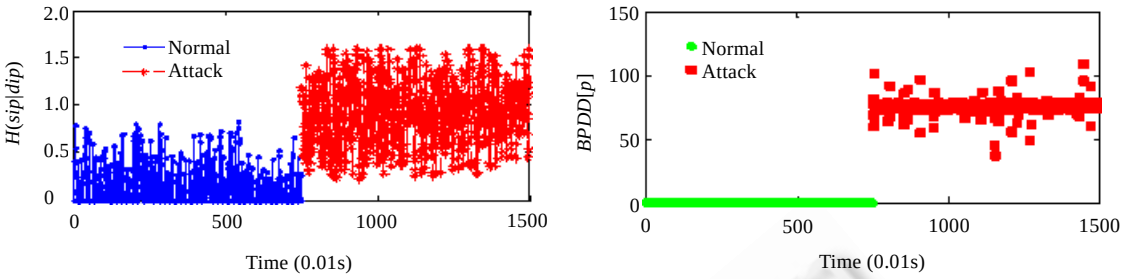
5.4 对ICMP flood攻击的检测

将获取的 ICMP flood 攻击数据集与背景流量混合.攻击流量的采样周期固定为 0.01s,而背景流量的采样周期从 1s 逐步增加至 5s,每次以 1s 的间隔递增.这样得到 5 组数据,每组数据的训练集和测试集均为连续的 1 500 个样本,即 30 个样本序列.图 4 给出当 T 为 1s 时,训练集中各特征取值随时间的变化.

由图 4(a)可以看出,尽管只使用了少量的攻击源地址,但攻击流量中源地址相对于目的地址,其多对一的程度基本上高于正常流量.图 4(b)表明,攻击引起的流量统计特征的变化使得攻击流量的 $BPDD[p]$取值与正常流量取值的差异十分明显.

分别使用 CRF,HMM,SVM 这 3 种学习算法对 5 组测试数据进行检测,检测结果见表 4.比较 3 种算法的检测结果可以看出,随着背景流量的加大,CRF 的优势越大.尽管 3 种算法的检测结果均受到不同程度的影响,但 HMM,SVM 的检测精度明显下降,而 CRF 仍具有 97%以上的检测率.这说明 CRF 能够利用上下文信息降低噪声

的干扰提高检测质量,与 HMM 相比,SVM 具有更好的鲁棒性.



(a) $H(sip|dip)$ time series
(a) $H(sip|dip)$的时间序列

(b) $BPDD[p]$ time series
(b) $BPDD[p]$的时间序列

Fig.4 Feature series in training set when T=1s in the detection of ICMP flood attacks

图 4 ICMP flood 攻击检测中,当 T=1s 时,训练集的各分类特征时间序列

Table 4 Detection results of three algorithms on ICMP flood attacks (%)

表 4 3 种算法对 ICMP flood 攻击的检测结果(%)

		T (×0.01s)				
		100	200	300	400	500
CRF	*DR*	99.73	99.33	98.93	98.27	97.20
	FR	0.00	0.27	0.67	1.07	1.20
	ER	0.13	0.47	0.87	1.40	2.00
HMM	*DR*	97.87	95.60	94.94	93.87	89.35
	FR	1.34	2.14	2.54	3.08	3.47
	ER	1.73	3.27	3.80	4.60	7.07
SVM	*DR*	98.27	97.07	96.54	94.41	91.74
	FR	0.93	1.47	2.80	3.34	4.41
	ER	1.33	2.20	3.13	4.47	6.33

6 结束语

本文提出了一种基于 CRF 模型的 DDoS 攻击检测方法,充分利用 CRF 模型具有的融合利用上下文信息和多特征的优势,有效克服了现有机器学习算法的局限性.设计了 TFCE,BPDD 的统计特征二元组对 TCP flood, UDP flood,ICMP flood 这 3 类攻击的特点进行描述.通过分别为 3 类攻击建立分类模型和有效训练,准确地推断和检测 DDoS 攻击.实验结果表明,该方法能够充分发挥 CRF 模型的优势,准确区分正常流量和攻击流量.与同类方法相比,检测率更高,鲁棒性更好.

在后续的工作中,我们将会进一步研究样本序列长度参数对检测效果的影响,给出一种自适应的参数设置方法.

References:

[1] Moore D, Voelker G, Savage S. Inferring Internet denial-of-service activity. In: Proc. of the 10th USENIX Security Symp. Washington, 2001. 9−22. [doi: 10.1145/1132026.1132027]

[2] Jiang H, Dovrolis C. Why is the Internet traffic bursty in short time scales? In: Proc. of the ACM SIGMETRICS 2005. New York, 2005. 241−252. [doi: 10.1145/1071690.1064240]

[3] Sang A, Li SQ. A predictability analysis of network traffic. In: Proc. of the INFOCOM 2000. Tel Aviv, 2000. 342−351. [doi: 10.1109/INFCOM.2000.832204]

[4] Lafferty JD, McCallum A, Pereira FCN. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Brodley C, Danyluk A, eds. Proc. of the 18th Int'l Conf. on Machine Learning (ICML 2001). San Francisco: Morgan Kaufmann Publishers, 2001. 282−289.

[5] Lakhina A, Crovella M, Diot C. Diagnosing network-wide traffic anomalies. In: Proc. of the ACM SIGCOMM 2004. Portland, 2004. 219−230. [doi: 10.1145/1015467.1015492]

[6] Peng T, Leckie C, Ramamohanarao K. Proactively detecting distributed denial of service attacks using source IP address monitoring. In: Proc. of the 3rd Int'l IFIP-TC6 Networking Conf. Athens, 2004. 771−782. [doi: 10.1007/978-3-540-24693-0_63]

[7] Sun ZX, Li QD. Defending DDos attacks based on the source and destination IP address database. Journal of Software, 2007, 18(10):2613−2623 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/18/2613.htm [doi: 10.1360/ jos182613]

[8] Sun QD, Zhang DY, Gao P. Detecting distributed denial of service attacks based on time series analysis. Chinese Journal of Computers, 2005,28(5):767−773 (in Chinese with English abstract).

[9] Lakhina A, Crovella M, Diot C. Mining anomalies using traffic feature distributions. In: Proc. of the ACM SIGCOMM 2005. Philadelphia, 2005. 217−228. [doi: 10.1145/1090191.1080118]

[10] Gil TM, Poletto M. Multops: A data-structure for bandwidth attack detection. In: Proc. of the 10th USENIX Security of Symp. Washington, 2001. 3−14.

[11] Kim Y, Lau WC, Chuah MC, Chao HJ. PacketScore—A statistics-based packet filtering scheme against distributed denial-of-service attacks. IEEE Trans. on Dependable and Secure Computing, 2006,3(2):141−155. [doi: 10.1109/TDSC.2006.25]

[12] Zhou DQ, Zhang HF, Zhang SW, Hu XP. A DDoS attack detection method based on hidden Markov model. Journal of Computer Research and Development, 2005,42(9):1594−1599 (in Chinese with English abstract). [doi: 10.1360/crad20050921]

[13] Noh S, Jung G, Choi K, Lee C. Compiling network traffic into rules using soft computing methods for the detection of flooding attacks. Applied Soft Computing, 2008,8(3):1200−1210. [doi: 10.1016/j.asoc.2007.02.016]

[14] Ghahramani Z. An introduction to hidden Markov models and Bayesian networks. Int'l Journal of Pattern Recognition and Artificial Intelligence, 2001,15(1):9−42. [doi: 10.1142/S0218001401000836]

[15] Zhong P, Wang RS. Using combination of statistical models and multilevel structural information for detecting urban areas from a single gray-level image. IEEE Trans. on Geoscience and Remote Sensing, 2007,45(5):1469−1482. [doi: 10.1109/TGRS.2007. 893739]

[16] McCallum A, Li W. Early results for named entity recognition with conditional random fields, feature induction and Web-enhanced lexicons. In: Proc. of the 7th CoNLL. Edmonton: Morgan Kaufmann Publishers, 2003. 188−191. [doi: 10.3115/ 1119176.1119206]

[17] Gupta KK, Nathn B, Kotagiri R. Layered approach using conditional random fields for intrusion detection. IEEE Trans. on Dependable and Secure Computing, 2008,7(1):35−49. [doi: 10.1109/TDSC.2008.20]

[18] Sun ZX, Tang YW, Zhang W, Gong J, Wang RC. A router anomaly traffic filter algorithm based on character aggregation. Journal of Software, 2006,17(2):295−304 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/17/295.htm [doi: 10.1360/ jos170295]

[19] MIT Lincoln Laboratory. 2000. http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html

[20] Xu M, Chen C, Ying J. A two-layer Markov chain anomaly detection model. Journal of Software, 2005,16(2):276−285 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/16/276.htm

[21] Li SZ. Markov Random Field Modeling in Image Analysis. 3rd ed., London: Springer-Verlag, 2009. 28−30.

[22] Kumar S, Hebert M. Discriminative random fields: A discriminative framework for contextual interaction in cassification. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Nice, 2003. 1150−1157. [doi: 10.1109/ICCV.2003.1238478]

[23] Kumar S, Hebert M. Discriminative fields for modeling spatial dependencies in natural images. In: Proc. of the Conf. on Advances in Neural Information Processing Systems. Vancouver, 2003. 1531−1538.

[24] Dietterich TG, Ashenfelter A, Bulatov Y. Training conditional random fields via gradient tree boosting. In: Proc. of the Int'l Conf. on Machine Learning. Banff, 2004. 217−224. [doi: 10.1145/1015330.1015428]

[25] Sha F, Pereira F. Shallow parsing with conditional random fields. In: Proc. of the Human Language Technology Conf. and North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2003). Edmonton: Morgan Kaufmann Publishers, 2003. 213−220. [doi: 10.3115/1073445.1073473]

[26] Kschischang FR, Frey BJ, Loeliger HA. Factor graphs and the sum-product algorithm. IEEE Trans. on Information Theory, 2001, 47(2):498−519. [doi: 10.1109/18.910572]

[27] Murphy K. CRF toolbox for matlab. 2004. http://www.cs.ubc.ca/~murphyk/Software/CRF/crfGeneralOld.html

[28] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm/

附中文参考文献:

[7] 孙知信,李清东.基于源目的 IP 地址对数据库的防范 DDos 攻击策略.软件学报,2007,18(10):2613−2623. http://www.jos.org.cn/1000-9825/18/2613.htm [doi: 10.1360/jos182613]

[8] 孙庆东,张德运,高鹏.基于时间序列分析的分布式拒绝服务攻击检测.计算机学报,2005,28(5):767−773.

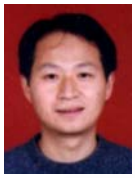
[12] 周东清,张海锋,张绍武,胡祥培.基于 HMM 的分布式拒绝服务攻击检测方法.计算机研究与发展,2005,42(9):1594−1599. [doi: 10.1360/crad20050921]

[18] 孙知信,唐益慰,张伟,宫婧,王汝传.基于特征聚类的路由器异常流量过滤算法.软件学报,2006,17(2):295−304. http://www.jos.org.cn/1000-9825/17/295.htm [doi: 10.1360/jos170295]

[20] 徐明,陈纯,应晶.一个两层马尔可夫链异常入侵检测模型.软件学报,2005,16(2):276−2623. http://www.jos.org.cn/1000-9825/16/276.htm

刘运(1979－),男,湖南怀化人,博士生,主要研究领域为网络安全,人工智能.

殷建平(1963－),男,博士,教授,博士生导师,主要研究领域为算法设计与分析,模式识别,人工智能,网络安全.



蔡志平(1975－),男,博士,副教授,主要研究领域为网络安全,虚拟化.

程杰仁(1974－),男,博士,副教授,主要研究领域为网络安全,人工智能.

钟平(1979－),男,博士,讲师,主要研究领域为图像处理,计算机视觉,模式识别.