基于 ACK 序号步长的 LDoS 攻击检测方法

吴志军，潘卿波，岳猛
（中国民航大学电子信息与自动化学院，天津 300300）

摘 要：低速率拒绝服务（LDoS, low-rate denial of service）攻击具有极强的隐蔽性，对大数据中心和云计算平台构成潜在的安全威胁。在研究 LDoS 攻击期间网络流量变化的基础上，对数据接收端回传给发送端的 ACK 数据分组进行统计分析，揭示了其序号步长在 LDoS 攻击期间具有的波动特征。采用排列熵的方法提取该特征，提出了一种基于 ACK 序号步长排列熵的 LDoS 攻击检测方法。该方法通过采集发送端收到的 ACK 数据分组，对其序号进行采样并计算步长；再利用对时间敏感性较强的排列熵算法检测出步长突变时刻，达到检测 LDoS 攻击的目的。在实际网络环境中设计和搭建了测试平台并对所提方法进行了验证，实验结果表明，所提方法具有较好的检测性能，取得了较好的检测效果。

关键词：低速率拒绝服务；ACK 序号步长；排列熵算法；检测
中图分类号：TP302
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2018126

Detection method of LDoS attack based on ACK serial number step-length

WU Zhijun, PAN Qingbo, YUE Meng
School of Electronic Information & Automation, Civil Aviation University of China, Tianjin 300300, China

Abstract: Low-rate denial of service (LDoS) attack is a potential security threat to big data centers and cloud computing platforms because of its strong concealment. Based on the analysis of network traffic during the LDoS attack, statistical analysis was given of ACK packets returned by the data receiver to the sender, and result reveals the sequence number step had the characteristics of volatility during the LDoS attack. The permutation entropy method was adopted to extract the characteristics of volatility. Hence, an LDoS attack detection method based on ACK serial number step permutation entropy was proposed. The serial number was sampled and the step length was calculated through collecting the ACK packets that received at the end of sender. Then, the permutation entropy algorithm with strong time-sensitive was used to detect the mutation step time, and achieve the goal of detecting LDoS attack. A test-bed was designed and built in the actual network environment for the purpose of verifying the proposed approach performance. Experimental results show that the proposed approach has better detection performance and has achieved better detection effect.

Key words: low-rate denial of service, ACK serial number step-length, permutation entropy, detection

1 引言

低速率拒绝服务（LDoS, low-rate denial of service）攻击是一种面向 TCP 的 DoS 攻击方式[1-2]。区别于泛洪式攻击，LDoS 攻击主要利用端系统或网络中常用的自适应机制所存在的安全漏洞，通过发送周期性的短脉冲数据流，使网络一直处于不稳定的状态，从而导致链路传输质量差、传输效率低，大大降低了被攻击目标的服务质量。

LDoS 攻击是传统 DoS 攻击的变形。传统 DoS

收稿日期：2017−12−26；修回日期：2018−05−08
基金项目：国家自然基金委员会与中国民航局联合基金资助项目（No.U1533107）；天津市自然基金重点资助项目（No.17JCZDJC30900）
Foundation Items: The Joint Foundation of National Natural Science Foundation and Civil Aviation Administration of China (No.U1533107), The Major Program of Natural Science Foundation of Tianjin (No.17JCZDJC30900)

攻击通过不断向被攻击目标发送高速数据分组，流量突变明显，单从流量分析可以较容易地检测出来。LDoS 攻击则是以一定周期间歇性地发送攻击数据分组，大大降低被发现的概率。一个完整的 LDoS 攻击脉冲可由三元组（T,L,R）组成[3-4]。其中，T 为 LDoS 攻击周期，理想的攻击周期应与 TCP 的超时重传（RTO, retransmission time out）相同，这样可以引起链路的重度拥塞，使 TCP 发送端一直处于超时重传状态，拥塞窗口（CWND, congestion window）始终处于最小值。虽然这种攻击效果最佳，但需要精确估计 RTO 大小，且实际网络复杂多变，所以一般会调整为一个定周期，使链路频繁处于快速重传状态，也能达到良好的攻击效果，本文采用的就是不变周期的攻击方式；L 为攻击脉冲宽度，是一次攻击数据分组持续的时间，一般取 2~3 个 RTT（往返时延）；R 为攻击脉冲的最大速率，一般接近链路瓶颈速率。

LDoS 攻击数据流平均速率低，且完全隐藏在 TCP 流中，不易被察觉，故传统的 DoS 检测方法很难对 LDoS 攻击进行检测和防御。TCP 主要依靠确认（ACK）机制才能使数据顺利传输，所以可以利用 ACK 流量来反映 TCP 的流量变化。因此，本文通过观测 ACK 数据分组，提取其序号步长变化为主要特征，提出一种基于 ACK 序号步长异常波动的 LDoS 检测方法，在实际网络环境真实平台下的实验验证，该方法具有良好的检测效果。

2 相关工作

LDoS 攻击一经发现便引起了相关学者的关注和研究，尤其对 LDoS 攻击的检测手段更是重中之重。目前，对于 LDoS 攻击的检测方法大多是基于网络流量特征的，包括时域特征、频域特征、波形特征等。

基于时域特征的方法主要通过分析网络流量在时域上的变化，提取正常流量和异常流量的时域特征进行比较，从而判断是否存在 LDoS 攻击。例如，Kwok 等[5]提出了 HAWK 检测方法，通过观测采样时间内高速脉冲的数量，若发现其数量超过阈值则判断存在攻击；Xiang 等[6]提出了利用广义熵和信息距离作为度量指标来量化不同概率分布的网络流量之间的差异，计算广义熵值和信息距离来区分正常流量和 LDoS 攻击流量，以此进行低速率拒绝服务攻击的判定；Yuhei 等[7]提出了一种利用路由器快速分组匹配的功能来检测 DoS 攻击，并证明了突发攻击流量的持续时间可以作为正常流量和 LDoS 攻击流量的区分特性。基于频域特征的方法结合了信号处理技术，通过对时间序列的频域转化，再利用一些经典滤波和检测算法对得到的频域特征加以处理，从而实现对 LDoS 攻击流量的检测和过滤。例如，Cheng 等[8]提出了在频域利用归一化累积功率谱密度（NCPSD, normalized cumulative power spectrum density）检测 LDoS 攻击的方法，利用正常流量和攻击流量 NCPSD 之间的最大距离位置作为检测依据来判断 LDoS 攻击；何炎祥等[9]提出了一种基于小波特征提取的 LDoS 检测方法，利用离散小波变换将网络流量分为高、中、低 3 个频率分量，以此来寻找 LDoS 攻击流量；Paul 等[10]基于费雪统计测试和西格尔统计测试来分析 LDoS 攻击流量的频谱特性，并表明当存在多个复合周期的 LDoS 攻击脉冲时，西格尔统计测试可以更有效地识别低速率拒绝服务攻击。基于相关检测的方法主要根据在发生 LDoS 攻击时，网络系统会产生一系列的一致性特征，此时通过一些相关算法就可从对 LDoS 攻击流进行检测。例如，Wei 等[11]通过计算不同流量之间的皮尔逊相关系数，设定判定阈值来检测 DDoS 攻击。Bhuyan 等[12]利用一种局部秩相关检测（PRCD, partial rank correlation detection）的方法来检测低速率和高速率 DDoS 攻击，一旦从 PRCD 估计中根据预设阈值发现恶意流量，就请求边缘路由器停止将该流量转发到下游路由器，实验表明该方法具有较高的检测准确度。

目前的检测方法基本都是从复杂的背景流量中检测出隐藏的 LDoS 攻击流量，但由于 LDoS 攻击流具有低速率和高隐蔽的特点，现有的检测方法在准确性、实时性、简易性方面都存在一些不足。而 LDoS 在影响正常 TCP 流量的同时也会对 ACK 流量产生一定影响，许多学者基于 ACK 流量的变化也提出了一系列检测方法。例如，Chen 等[13]提出了通过分析网络存在 LDoS 攻击时 3 种流量的异常变化，即 ACK 流量异常分布、ACK 流量异常波动、数据量大小异常变化，制定相应的判决准则对 LDoS 攻击进行检测。本文在对 ACK 流量分析的基础上发现，等时间间隔内，接收端每次确认的数据量（即相邻 2 个时刻 ACK 序号的差值）在正常情况和受到 LDoS 攻击时存在较大差异，可以将这种差异作为检测 LDoS 攻击的特征。因此，提取网络存在 LDoS 攻击时 ACK

序号步长的波动特征，结合排列熵算法判定步长突变时刻并区分正常波动和异常波动，进而判断出 LDoS 攻击是否存在及其攻击时刻。

3　基于 ACK 序号步长分析的攻击检测方法

通过对正常网络和存在 LDoS 攻击时 ACK 数据分组的采样分析发现，ACK 序号步长的波动程度存在明显差异，且在遭受 LDoS 攻击的瞬间，步长存在突变行为。针对这种波动差异，可以用熵值的方法对其进行度量并加以区分，故本文运用对时间敏感性较强的排列熵算法，设定相应阈值来检测步长突变点，以确定 LDoS 攻击的时刻。

3.1　ACK 序号步长特征分析

在数据传输过程中，ACK 序号为接收端期望收到发送端下一个分组段的第一个数据字节的序号，TCP 中接收端一般采用累积确认的方式。即接收端不必对收到的分组逐个发送 ACK 确认，而是在收到几个分组后，对按序到达的最后一个分组发送确认[14]。这样一个字节号为N的ACK意味着直到N的字节（但不包含 N）已经被成功接收。从发送端收到的 ACK 来看，当前 ACK 序号与上次 ACK 序号之差可以代表每次被确认的数据量，正常传输中，每次被确认的数据量基本是恒定的，而受到 LDoS 攻击后，由于要频繁进入快速重传和快速恢复阶段，接收端会回传重复的 ACK 让发送端重传数据，此时 ACK 序号差值将会发生很大变化。具体交互过程如图 1 所示。

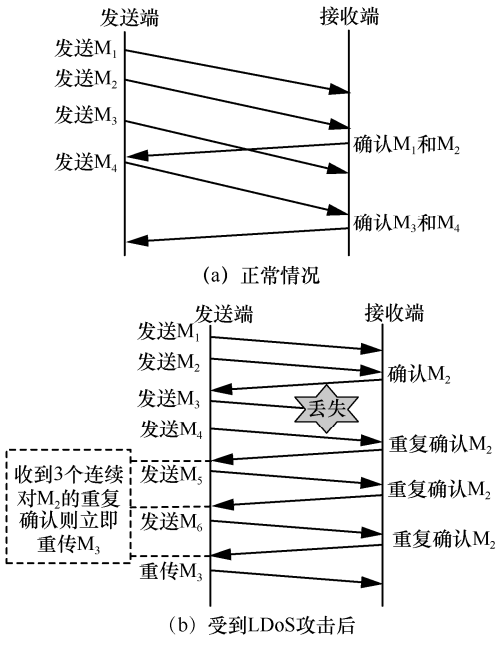

图 1　正常情况和受到 LDoS 攻击后快速重传交互示意

为便于接下来的分析与研究，针对 6 种场景展开研究。

E_1：仅存在单条 TCP 流且没受到任何攻击。

E_2：仅存在单条 TCP 流且受到 LDoS 攻击。

E_3：存在多条相同 RTT 的 TCP 流相互竞争，并添加适当非 LDoS 攻击突变，但不受到 LDoS 攻击。

E_4：存在多条相同 RTT 的 TCP 流相互竞争，且受到 LDoS 攻击。

E_5：存在多条不同 RTT 的 TCP 流相互竞争，并添加适当非 LDoS 攻击突变，但不受到 LDoS 攻击。

E_6：存在多条不同 RTT 的 TCP 流相互竞争且受到 LDoS 攻击。

对各个场景发送端收到的ACK序号进行统计，得到 ACK 序号的时间序列 $S=\{s_1,s_2,\cdots,s_n\}$。为了更好地分析相邻 2 个时刻之间接收端确认数据量的变化情况，对 ACK 序号的序列进行等间隔采样，假设 Δt 为采样间隔，$\hat{s}_i$ 为采样后第 i 个 ACK 的序号值（$1\leqslant i\leqslant n$），则有

$$x(j)=\hat{s}_{i+1}-\hat{s}_i,1\leqslant i\leqslant n,1\leqslant j\leqslant i \tag{1}$$

其中，$x(j)$ 为采样后的 ACK 序号步长。将每 2 个时刻计算的步长值赋予其中较小的时刻，这样预处理后可以得到最终的 ACK 序号步长特征序列。各网络场景 ACK 序号步长波动情况如图 2 所示。

从图 2 可以得到以下结论。

1) 在仅有一条 TCP 流的情况下，正常传输数据时 ACK 序号步长波动幅度很小，即等时间间隔内接收端接收的数据量比较稳定，而受到 LDoS 攻击后，ACK 序号步长波动的较为剧烈且呈现一定的周期性。

2) 在多条 TCP 流的情况下，分为以下 2 种情况介绍。①当存在多条相同 RTT 的 TCP 流同时传输数据时，由于相同的 RTT 使各路径相互竞争较为激烈，ACK 序号步长波动幅度较大，波动剧烈；②当存在多条不同 RTT 的 TCP 流同时传输数据时，由于各路径的时延不同，各路径之间的竞争不如相同 RTT 时激烈，ACK 序号步长波动幅度较小，但波动剧烈，呈现不规则状态。链路受到 LDoS 攻击后，无论场景 E_4 和 E_6，此时发送端反复进入快速重传阶段，接收端每次接收的数据量始终比较少，ACK 序号步长的波动幅度整体降低，但相比场景 E_3 和 E_5（不存在 LDoS 攻击时）排列得更加规则，具有明显的差异，故将其定义为 ACK 序号步长的异常波动。

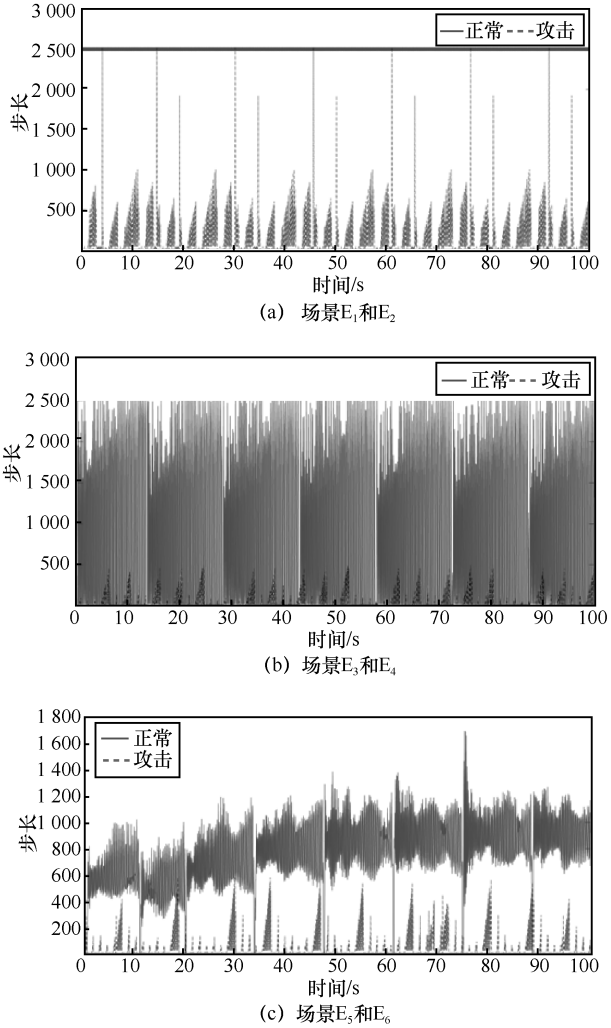

图 2　6 种网络场景的 ACK 序号步长波动对比

根据上述分析结果，可以利用 ACK 序号步长波动的差异性进行 LDoS 攻击的检测。由于实际网络环境中都是多条 TCP 流共存，仅有一条 TCP 流的情况属于特例。因此，在后续研究中，本文主要针对多条 TCP 流的场景（E_3~E_6）的 ACK 序号步长异常波动进行检测分析。

3.2　LDoS 攻击检测

由上述分析可知，受到 LDoS 攻击时 ACK 序号步长的波动范围和规则程度均与正常网络环境下相比有较为明显的差异，呈现出一种混乱无序和规则有序相互突变的现象，采用排列熵值可以清晰地描述这种异常突变。事件分布越无序，熵值越大；事件分布越有序，则熵值越小。

本文目标主要是在 LDoS 攻击的早期时刻检测出 ACK 序号步长的异常波动，从而发现 LDoS 攻击。文献[13]采用“移动极差”（MR, moving range）来分析一组序列的波动程度，移动极差需要计算一个观测窗口内最大值与最小值的差，涉及具体数值的计算且易受到极端值的影响。故这里采用一种基于熵值的突变检测算法——排列熵（PE, permutation entropy）算法。排列熵算法是一种衡量一维时间序列复杂度的平均熵参数[15]，它不涉及具体数值的计算，只关心相空间重构后相邻 2 个数值的大小关系，避免了极端值的影响，对突变具有较好的识别性。

排列熵算法需要先对一维时间序列进行相空间重构。由上文假设知，$x(j)$ 为原始 ACK 序号经采样后第 i 个和第 $i+1$ 个 ACK 序号的步长，因此，可以设置一个观测窗口 T_L。对一个观测窗口内的 ACK 序号进行采样、做差，可得到该观测窗口内的 ACK 序号步长序列为

$$\{X(j)=x(1),x(2),\cdots,x(n),j=1,2,3,\cdots,n\}$$

对其进行相空间重构，得到矩阵为

$$\begin{bmatrix} x(1) & x(1+\tau) & \cdots & x[1+(d-1)\tau] \\ x(2) & x(2+\tau) & \cdots & x[2+(d-1)\tau] \\ \vdots & \vdots & \vdots & \vdots \\ x(m) & x(m+\tau) & \cdots & x[m+(d-1)\tau] \\ \vdots & \vdots & \vdots & \vdots \\ x(k) & x(k+\tau) & \cdots & x[k+(d-1)\tau] \end{bmatrix}, m=1,2,3,\cdots,k \tag{2}$$

其中，d 和 $[x(m)\quad x(m+\tau)\quad \cdots\quad x(m+(d-1)\tau)]$ 分别为嵌入维数和时延时间，$k=n-(d-1)\tau$。矩阵中每一行看作一个重构分量，共有 k 个。将 $X(j)$ 的重构矩阵中的第 m 个重构分量 $]x(m)\quad x(m+\tau)\quad \cdots\quad x[m+(d-1)\tau]]$ 按照升序重新进行排列，即

$$x[j+(m_1-1)\tau]\leqslant x[j+(m_2-1)\tau]\leqslant\cdots\leqslant x[j+(m_d-1)\tau] \tag{3}$$

其中，$m_1\quad m_2\quad \cdots\quad m_d$ 表示重构分量中各个元素所在列的索引。若重构分量中存在相等的元素，即

$$x[j-(m_1-1)\tau]=x[j-(m_2-1)\tau] \tag{4}$$

就按照 m_1、m_2 的大小来排序，当 $m_1<m_2$ 时，有

$$x[j-(m_1-1)\tau]\leqslant x[j-(m_2-1)\tau] \tag{5}$$

所以，对于任意一组 ACK 序号步长序列 $X(j)$，相空间重构后所得的矩阵中每一行都可以得到一组符号序列，即

$$F(u)=(m_1,m_2,\cdots,m_d) \tag{6}$$

其中，$u=1,2,3,\cdots,k$（$k\leqslant d!$），m 维相空间映射

不同的符号序列($m_1,m_2,\cdots,m_d$)共有 $d!$种，符号序列 $F(u)$只是其中的一种排列。按照香农熵的形式，计算每一种符号序列出现的概率为 $P_1,P_2,\cdots,P_k$，则 ACK 序号步长序列 $X(j)$ 的 k 种不同符号序列的排列熵可定义为

$$H_P(d)=-\sum_{m=1}^{k}P_m\ln P_m \tag{7}$$

当 $P_m=\frac{1}{d!}$ 时，$H_p(d)$ 取最大值 $\ln(d!)$，为了方便，通常用 $\ln(d!)$ 对 $H_p(d)$ 进行归一化，即

$$0\leqslant H_p=\frac{H_p}{\ln(d!)}\leqslant 1 \tag{8}$$

H_p 值的大小反映了 ACK 序号步长序列 $X(j)$ 的随机程度。H_p 值越小，说明时间序列越规则，波动程度越小；反之，则 ACK 序号步长序列越接近随机值，波动程度越大。H_p 的变化反映并放大了 ACK 序号步长序列的细微变化。

为了更好地分析整个 ACK 序号步长序列的排列熵变化并找到 LDoS 的攻击时刻，本文所采用的处理方法是：将一个观测窗口的 ACK 序号步长组成的时间序列分为若干个长度为 l 的子序列，子序列样本的截止数据点对应的时间为 t。这些子序列相互之间使用最大重叠情形，即将每个子序列向后移动一个数据点得到下一个子序列，计算每个子序列的 PE，再将每个子序列计算得出的排列熵赋值给子序列中间的一个数据点，并把该数据点对应的时间 t_{mid} 作为此子序列排列熵的平均时间点。

通过上述过程描述可知，此算法有 3 个重要参数会对计算结果产生直接影响，分别是嵌入维数 d、时延 τ 和子序列长度 l。若 l 取值太小，那么计算就会失去其统计学意义，但是为了精确检测突变信号，l 取值又不能太大，可以根据实际观测量的大小来确定。d 和 τ 的取值大小直接影响了计算复杂度以及之后的阈值。目前，关于 d 和 τ 的选取主要有 2 种观点[16-17]：① 二者互相独立，d 和 τ 独立确定，通常先运用互信息法确定 τ，当 τ 确定后再利用伪近邻点法选取 d；② 二者相关，典型的方法是关联积分法（C-C 算法），先定义关联积分，再构造统计量，利用构造统计量和时延的关系来确定最优时延和嵌入维数。

本文根据文献[16]对 2 种方法的结果进行比较分析，在异常检测方面观点①（即独立确定参数）的方法更有效。因此，结合本文的攻击检测特点，本文采用观点①的方法来计算嵌入维数 d 和时延 τ 这 2 个参数。

首先采用互信息法计算时延 τ，基本原理是对于 ACK 序号步长组成的离散时间序列 $\{x(t),t=1,2,\cdots,n\}$，选取时延序列 $x(t+\tau)$ 构成新的序列 $y(t)$，通过计算 $x(t)$ 和 $y(t)$ 的相关性来确定时延 τ。二者之间的互信息为

$$I(\tau)=-\sum_i\sum_j P_{xy}(x_i,y_j)\lg[\frac{P_{xy}(x_i,y_j)}{P_x(x_i)P_y(y_j)}] \tag{9}$$

其中，$P_{xy}(x_i,y_j)$ 是 x_i 和 y_j 的联合分布概率，$I(\tau)$ 是关于时延 τ 的函数。

$I(\tau)$ 的极小值表示 $x(t)$ 和 $y(t)$ 的最大可能不相关，在相空间重构时，采用 $I(\tau)$ 的第一个极小值对应的时间 τ 作为最佳时延时间，这里计算结果为 τ=2；其次，采用伪近邻点法计算嵌入维数 d。其基本思想是：混沌时间序列可看作高维相空间运动轨迹在一维空间的投影，高维中不相邻的 2 个点在一维空间的投影可能成为相邻点，称为伪近邻点，重构相空间的过程即从一维时间序列中恢复高维混沌运动的轨迹，随着嵌入维数的增大，轨迹逐渐打开，伪近邻点也逐渐消失。在某维数 d_0 处，伪近邻点百分比会突降至 0 附近，且趋于稳定，此时的 d_0 即最佳嵌入维数。对于 ACK 序号步长序列，将 d 从最小嵌入维数 2 开始，计算伪近邻点比例，直到伪近邻点比例小于 5%或不随 d 的增大而减小时，可认为混沌吸引子完全打开，此时 d_0=3，作为最佳嵌入维数。

基于上述分析，采用 ACK 序号步长波动特征的 LDoS 攻击检测流程如图 3 所示。

其主要步骤为：①以一定的观测时间窗口采集 ACK 数据分组，并提取序号特征；②以 Δt 为采样间隔，对 ACK 序号进行采样并计算步长；③以长度为 l 的滑动窗口将步长序列截取为若干个子序列，子序列之间采用最大重叠情形；④计算每个子序列的排列熵，并与阈值比较，若小于预设阈值，则报警检测到 LDoS 攻击，否则滑动一个数据点再次运行步骤③，直到所有子序列计算完毕。

4　实验及结果分析

为了验证采用排列熵算法检测 LDoS 攻击的性能，本文在实际网络平台中搭建了测试环境，对所提 LDoS 攻击检测方法进行了测试。

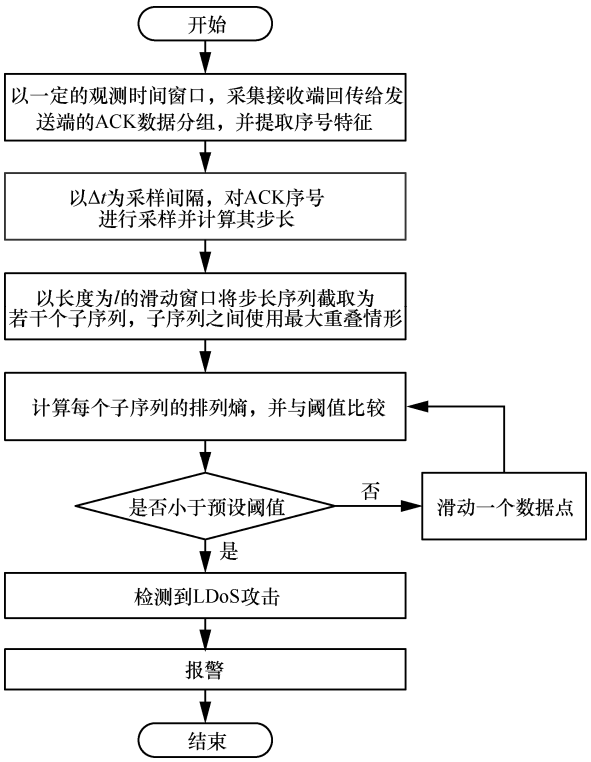

图 3　基于 ACK 序号步长波动特征的 LDoS 检测流程

4.1　实验环境

本文方法的网络测试环境是参考美国莱斯大学的 Knightly 教授指导的研究团队设计的网络环境[1-2]，它是一个哑铃形状的拓扑结构，如图 4 所示。该测试环境由 7 台计算机、一个交换机、一个路由器和一个服务器组成。其中，交换机与路由器之间的瓶颈链路为 10 Mbit/s。主机 1~主机 3 为合法用户，链路带宽均为 100 Mbit/s，单向时延均为 20 ms。傀儡机 4~傀儡机 6 在控制机 7 的控制下发送 LDoS 攻击流量，攻击参数为：攻击速率 R=10 Mbit/s，攻击脉宽 L=200 ms，攻击周期 T=1.2 s。

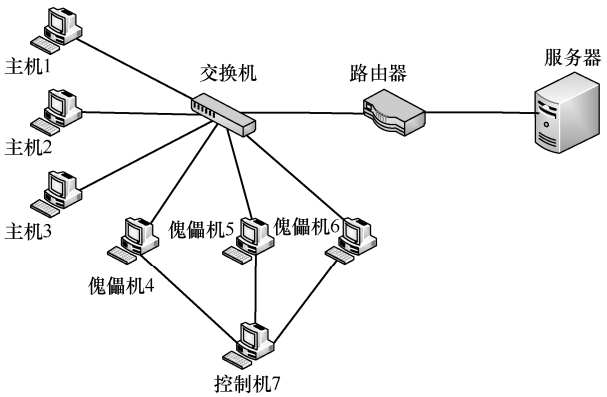

图 4　实验环境

实验环境所用检测算法的相关参数如表 1 所示。

表 1　　测试环境所用检测算法相关参数

参数	取值
嵌入维数 d	3
时延 τ	2
子序列长度 l	100

将检测机制部署在每个合法客户端，按照一定的观测窗口（这里设置观测时间窗口 T_L=100 s），利用 Wireshark 抓取分组工具采集回传的 ACK 数据分组，提取其序号特征，对 ACK 序号以 500 ms 的采样间隔进行采样，利用式(1)计算采样后的 ACK 序号步长，形成 ACK 序号步长序列。

4.2　实验内容

基于 ACK 序号步长波动特征的 LDoS 攻击检测实验主要包括 2 个内容。

1) 对单条 TCP 流是否受到 LDoS 攻击进行检测。

2) 对多条 TCP 流是否受到 LDoS 攻击进行检测。

为了比较其他攻击对 ACK 序号步长的影响，在实验 1)和实验 2)中还加入了传统泛洪拒绝服务（FDoS, flood DoS）攻击进行对比实验。

针对不同网络情况进行大量实验，定义检测的性能指标包括检测率、漏报率、误报率。其中，检测率 $=\dfrac{\text{检测到LDoS攻击的实验次数}}{\text{实验总次数}}$，漏报率 $=\dfrac{\text{存在LDoS攻击但未检测出来的实验次数}}{\text{实验总次数}}$，误报率 $=\dfrac{\text{不存在LDoS攻击但检测出含有攻击的实验次数}}{\text{实验总次数}}$。

1) 单条 TCP 流

实验首先验证本文检测方法在只有一条 TCP 流存在下受到 LDoS 攻击时的检测效果，即仅让主机 1 与服务器建立 TCP 连接，控制机控制傀儡机按照预设参数在 50 s 时发起 LDoS 攻击和 FDoS 攻击。利用表 1 中参数的排列熵算法对得到的 ACK 序号步长序列计算其排列熵，计算结果如图 5 所示。

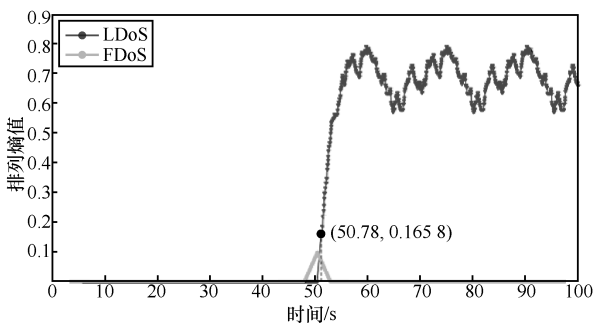

图 5　单条 TCP 流正常状况下和受到 LDoS 攻击时 ACK 序号步长排列熵

在正常情况下，由于接收端接收窗口的限定，接收端每次接收的数据量基本恒定，回传给发送端的 ACK 序号步长波动范围规则有序，故其排列熵值一直处于最低的 0 状态，直到 50 s 时遭受 DoS 攻击。当受到 FDoS 攻击时，由于 FDoS 攻击使网络流量瞬间激增，链路发生严重拥塞，致使网络进入瘫痪状态。因此 ACK 流量在 50 s 处发生突变，ACK 序号的步长开始混乱，其排列熵值在 50 s 处突变增大，之后由于链路瘫痪，ACK 数据分组不再回传，ACK 序号步长一直为 0 不再波动，所以排列熵值降到 0，代表链路再次进入平稳状态（其实此时是空白链路，没有数据交互）。当受到 LDoS 攻击时，LDoS 攻击也会导致 ACK 流量突变，排列熵值在（50.78, 0.165 8）时产生熵值跳变点，之后由于不断的快速重传，ACK 序号步长波动剧烈，其排列熵值一直处于较高状态，与 FDoS 攻击相比具有明显差别，基于此可以区分 2 种攻击类型。而且预测 LDoS 攻击时刻与实际遭受攻击时刻相差 0.78 s，在允许误差范围之内。该方法可在早期判断出攻击时刻，具有一定时效性。

2) 多条 TCP 流

针对多条 TCP 流共存情况下受到 LDoS 攻击时验证本文方法的有效性，分为 2 种情况：① 多条路径具有相同 RTT；② 多条路径具有不同 RTT，即让主机 1~主机 3 同时与服务器建立 TCP 连接，共用瓶颈链路传输数据，攻击流量同样在 50 s 时发起。由于 DoS 攻击会降低一个路由域内所有节点的服务质量，各链路均表现出一样的异常特征，因此选取其中一条链路，得到一个观测窗口内的 ACK 序号步长序列。

然后对其进行排列熵算法处理。由之前的分析可知，判定是否发生 LDoS 攻击需要设定一个阈值。根据切比雪夫不等式原理[18]：设 X 为随机变量，其数学期望为 μ，标准差为 σ，则对于任意 z>0，有

$$P\{|X-\mu|\geqslant z\sigma\}\leqslant\frac{1}{z^2} \tag{10}$$

由切比雪夫不等式可知，对任意分布的数据样本，落在样本数据 z 个标准差之内的数据的概率至少是 $1-\frac{1}{z^2}$。即对于服从任何分布的样本数据，“几乎所有”的值都会接近平均。

依据切比雪夫不等式来设置 ACK 序号步长排列熵的阈值，对于 ACK 序号步长的排列熵样本，无论其服从何种分布，都满足切比雪夫不等式，即

$$P\{|H_p-\mu_p|\geqslant z\sigma_p\}\leqslant\frac{1}{z^2} \tag{11}$$

其中，μ_p 为排列熵的均值，σ_p 为排列熵的标准差，z 为任意正数。本文取 z=5，即排列熵偏离均值 5 倍的标准差的样本的概率小于 $\frac{1}{25}$（即 4%），也就是有大约 96%的数据落在 $\mu_p\pm5\sigma_p$ 置信区间内。因此，可以通过计算正常状态下 ACK 序号步长排列熵的均值 μ_p 和标准差 σ_p 来设定检测门限，将正常状态下 ACK 序号步长排列熵的 $\mu_p\pm5\sigma_p$ 设置为置信区间来区别正常状态和受到攻击状态。

本实验中，通过计算 2 种情况（相同 RTT 和不同 RTT）正常状态下（0～50 s）ACK 序号步长对应的排列熵，可分别得其均值为 μ_{p_1}=0.926 2、μ_{p_2}=0.937 7，标准差 σ_{p_1}=0.031 7、σ_{p_2}=0.032 3。根据切比雪夫不等式，分别采用置信区间下限 $\mu_{p_1}-5\sigma_{p_1}$=0.767 7 和 $\mu_{p_2}-5\sigma_{p_2}$=0.775 9 作为检测阈值，如图 6 和图 7 所示。

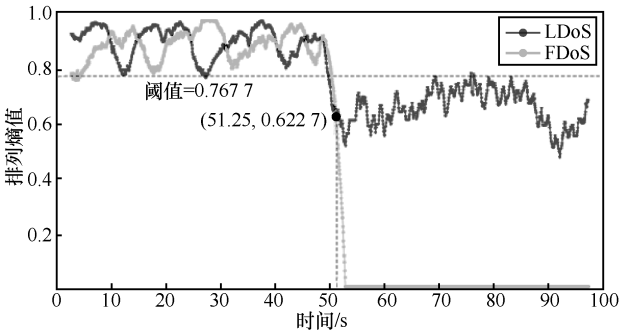

图 6　相同 RTT 的 TCP 流共存时 ACK 序号步长的排列熵

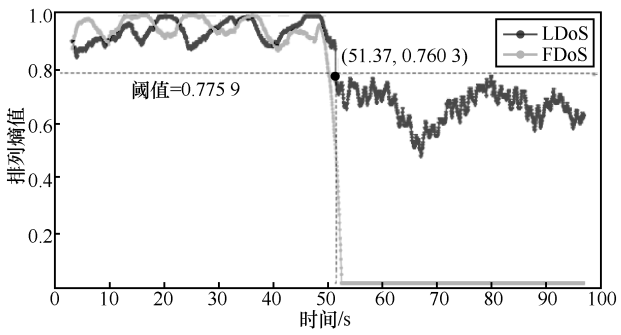

图 7　不同 RTT 的 TCP 流共存时 ACK 序号步长的排列熵

由图 6 和图 7 可知，正常情况下，由于各 TCP 流之间存在相互竞争，导致 ACK 流量分布不均，ACK 序号步长波动混乱，其排列熵数值较高。受到 DoS 攻击后，DoS 攻击会使各链路表现出相同的拥塞状态，ACK 流量反而比正常情况下排列得更有序，所以熵值波动范围比正常情况时要低。当受到 FDoS 攻击时，接收端立刻进行 ACK 重传，ACK 序号步长突降为 0，其排列熵值发生突变，之后由于链路严重阻塞，导致不再有 ACK 数据分组的回传，ACK 序号步长持续为 0，排列有序，排列熵值也一直处于 0 状态；受到 LDoS 攻击时，由于快速重传机制的影响，使 ACK 序号步长不会降至最低，而是在一定范围内波动，其排列熵值相比 FDoS 攻击时要大，但基本不会超过阈值。所设阈值可以有效预测 LDoS 攻击时刻，具有相同 RTT 的 TCP 流共存时预测 LDoS 攻击发生在 51.25 s，与实际受到攻击时刻相差 1.25 s；具有不同 RTT 的 TCP 流共存时预测 LDoS 攻击发生在 51.37 s，与实际受到攻击时刻相差 1.37 s。2 个时间差在允许误差范围之内，可以实现在攻击初期及早发现攻击时刻的目标。而且阈值可以明显区分正常时与受到 LDoS 时 ACK 序号步长排列熵的变化情况。

针对不同网络场景进行 100 次实验，统计各检测性能指标，得到本文方法的检测性能指标和误差结果如表 2 所示。

表 2 不同网络情况下的检测性能指标和误差结果

TCP 流数	阈值	检测率	误报率	漏报率	预测攻击时刻误差/s
单条	0	100.0%	1.0%	0.0%	0.78
多条相同 RTT	0.767 7	96.0%	3.0%	4.0%	1.25
多条不同 RTT	0.775 9	98.0%	7.0%	2.0%	1.37

由表 2 可知，单条 TCP 流时，网络场景最为简单，不存在其他链路的影响，所以检测率最高，可达 100%，误报率、漏报率和预测攻击时刻误差也均比较低；多条 TCP 流时，由于 TCP 流之间存在相互竞争的影响，相同 RTT 的检测率为 96.0%，误报率为 3.0%，漏报率为 4.0%，预测攻击时刻误差为 1.25 s。不同 RTT 的检测率为 98.0%，误报率为 7.0%，漏报率为 2.0%，预测攻击时刻误差为 1.37 s。误差均在可承受范围之内。

4.3 比较分析

根据现实场景，主要利用多条不同 RTT 的 TCP 进行对比实验，对比传统经典检测算法 NCPSD[5] 和基于 ACK 异常流量[8]的检测方法，并对各检测性能指标进行了统计，如表 3 所示。

表 3 各检测算法性能比较

检测方法	报警率	虚警率	漏警率
NCPSD	88.0%	16.0%	12.0%
基于 ACK 异常流量	94.0%	2.0%	6.0%
本文检测方法	98.0%	7.0%	2.0%

由表 3 可知，相比传统基于频域的检测方法，本文检测方法可获得更高的报警率，同时虚警率和漏警率也比较低。基于 ACK 流量异常的检测方法，需要统计较长时间内的 ACK 流量数据，在 LDoS 攻击中期检测效果较为理想，而在 LDoS 攻击初期存在一定的误报率。本文检测方法可在 LDoS 攻击初期实现及时检测，降低了攻击初期的漏报率。

5 结束语

本文分析了网络数据传输过程中 ACK 数据分组序号步长的意义，研究了正常情况和存在 LDoS 攻击时 ACK 序号步长的波动差异，并利用排列熵算法进行特征提取，最后设计了判决准则，实现了对 LDoS 攻击的检测。排列熵算法具有计算简单、抗噪性强、顽健性高的特点，可以有效地发现早期的信号异常突变，具有较好的实时性和稳定性。实验结果表明，基于 ACK 序号步长排列熵的 LDoS 攻击检测方法可以准确判断是否发生攻击，并可以在 LDoS 攻击的初期阶段较准确地找到攻击发起时刻，使相关网络管理人员可以及早地发现攻击，避免造成更大损失，为后续的防御和过滤起到了重要作用。

当发生 LDoS 攻击时，网络中还可能存在其他的异于正常的突变行为，如吞吐量或可用带宽等，未来研究中，可联合多个特征的异常行为进行检测，提高检测准确率。对于检测算法中的 3 个主要参数：子序列长度 l、嵌入维数 d 和时延 τ，目前 l 的选取主要根据实际观测序列长度和人为经验判断，d 和 τ 的选取则采用一些传统的相空间重构参数选取方法。今后还需挖掘各参数之间的潜在相关性，对参数选取方法进行优化，尽量减小误差，提高算法的准确率和运算速度，这也是未来的研究重点。

参考文献：

[1] KUZMANOVIC A, KNIGHTLY E W. Low-rate TCP-targeted denial of service attacks and counter strategies[J]. IEEE/ACM Transactions on Networking, 2006, 14(4): 683-696.

[2] KUZMANOVIC A, KNIGHTLY E W. Low-rate TCP-targeted denial of service attacks: the shrew vs. the mice and elephants[C]//ACM SIGCOMM 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. 2003: 75-86.

[3] 文坤, 杨家海, 张宾. 低速率拒绝服务攻击研究与进展综述[J]. 软件学报, 2014, 25(3): 591-605.
WEN K, YANG J H, ZHANG B. Survey on research and progress of low-rate denial of service attacks[J]. Journal of Software, 2014, 25(3): 591-605.

[4] 何炎祥, 刘陶, 曹强, 等. 低速率拒绝服务攻击研究综述[J]. 计算机科学与探索, 2008, 2(1): 1-19.
HE Y X, LIU T, CAO Q, et al. A survey of low-rate denial-of-service attacks[J]. Journal of Frontiers of Computer Science and Technology, 2008, 2(1): 1-19.

[5] KWOK Y K, TRIPATHI R, CHEN Y, et al. HAWK: halting anomalies with weighted choking to rescue well-behaved TCP sessions from shrew DDoS attacks[C]//International Conference on NETWORKING and Mobile Computing. 2005: 423-432.

[6] XIANG Y, LI K, ZHOU W. Low-rate DDoS attacks detection and trace back by using new information metrics[J]. IEEE Transactions Information Forensics and Security, 2011, 6(2): 426-437.

[7] YUHEI H, JIA Y Z. SATOSHI N. Method for detecting low-rate attacks on basis of burst-state duration using quick packet-matching function[C]//IEEE International Symposium on Local and Metropolitan Area Networks. 2017: 1-2.

[8] CHENG C M, KUNG H, TAN K S. Use of spectral analysis in defense against DoS attacks[C]//IEEE Global Telecommunications. 2002: 2143-2148.

[9] 何炎祥, 曹强, 刘陶, 等. 一种基于小波特征提取的低速率 DoS 检测方法[J]. 软件学报, 2009, 20(4): 930-941.
HE Y X, CAO Q, LIU T, et al. A low-rate Dos detection method based on feature extraction using wavelet transform[J]. Journal of Software, 2009, 20(4): 930-941.

[10] PAUL C, MYONG K, ALEXANDER V. Spectral analysis of low rate of denial of service attacks detection based on fisher and Siegel tests[C]//IEEE International Conference on Communications(ICC). 2016: 1-6.

[11] WEI W, FENG C, XIA Y, et al. A rank correlation based detection against distributed reflection DoS attacks[J]. IEEE Communications Letters, 2013, 17(1): 173-175.

[12] BHUYAN M H, KALWAR A, GOSWAMI A, et al. Low-rate and high-rate distributed DoS attack detection using partial rank correlation[C]//Fifth International Conference on Communication Systems and Network Technologies. 2015: 706-710.

[13] CHEN K, LIU H Y, CHEN X S. Detecting LDoS attacks based on abnormal network traffic[J]. KSII Transactions on Internet and Information Systems, 2012, 6(7): 1831-1853.

[14] FALL K R, RICHARD S W. TCP/IP 详解卷 1: 协议[M]. 北京: 机械工业出版社, 2016.
FALL K R, RICHARD S W. TCP/IP illustrated volume 1: the protocols[M]. Beijing: China Machine Press, 2016.

[15] FENG F Z, RAO G Q, WEI S A. Application and development of permutation entropy algorithm[J]. Journal of Academy of Armored Force Engineering, 2012, 26(2): 34-38.

[16] 饶国强, 冯辅周, 司爱威. 排列熵算法参数的优化确定方法研究[J]. 振动与冲击, 2014, 33(1): 188-193.
RAO G Q, FENG F Z, SI A W. Method for optimal determination of parameters in permutation entropy algorithm[J]. Journal of Vibration and Shock, 2014, 33(1): 188-193.

[17] 王海燕, 盛昭瀚. 混沌时间序列相空间重构参数的选取方法[J]. 东南大学学报(自然科学版), 2000, 30(5): 113-117.
WANA H Y, CHENG Z H, Choice of the parameters for the phase space reconstruction of Chaotic time series[J]. Journal of Southeast University(Natural Science Edition), 2000, 30(5): 113-117.

[18] 刘永斌. 基于非线性信号分析的滚动轴承状态监测诊断研究[D]. 合肥: 中国科学技术大学, 2011.
LIU Y B. Nonlinear signal analysis for rolling bearing condition monitoring and fault diagnosis[D]. Hefei: University of Science and Technology, 2011.

[作者简介]

吴志军（1965−），男，新疆库尔勒人，博士，中国民航大学教授、博士生导师，主要研究方向为网络空间安全。

潘卿波（1992−），男，山西太原人，中国民航大学硕士生，主要研究方向为网络信息安全、低速率拒绝服务攻击的检测。

岳猛（1984−），男，河北沧州人，博士，中国民航大学讲师，主要研究方向为信息安全、云计算、低速率拒绝服务攻击的检测。