基于分布式统计时间序列的网络流量分析

孟凡雪[1,2] 刘衍珩[1,2] 吴 静[1,2] 杨书奇[1]
(吉林大学计算机科学与技术学院 长春 130012)[1]
(吉林大学符号计算与知识工程教育部重点实验室 长春 130012)[2]

摘 要 研究网络数据在分布式存储下的相关性,有利于入侵检测整体的学习和指导优化数据的存储。重点研究了网络传输过程中各种类型数据的流量的这种相关性,提出了一种基于分布式统计(DS)的时间序列分析方法:根据网络协议间的关系将数据包分组,分析数量关系并给出报警阈值。仿真实验结果表明,该方法能较好地发现各种网络攻击。

关键词 入侵检测,时间序列,分布式统计,网络流量
中图法分类号 TP393.08 **文献标识码** A

Analysis of Network Traffic Based on Distributed Statistical Time Series

MENG Fan-xue[1,2] LIU Yan-heng[1,2] WU Jing[1,2] YANG Shu-qi[1]
(College of Computer Science and Technology, Jilin University, Changchun 130012, China)[1]
(Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China)[2]

Abstract Studying relationship of distributed storage data would be conductive to the overall intrusion detection learning, thus this relationship could not only be used for intrusion detection learning algorithm, but also supervise to optimize data's storage. According to the analysis focused on relationship of network traffic in the distributed storage, a method based on distributed statistical time series was proposed. According to the relationship of network protocol, this method could group data packet, thus analyzed quantity relationship and gave alarm threshold. The experiment results show that the method can be used to detect network attacks.

Keywords Intrusion detection, Time series, Distributed statistical, Network traffic

1 引言

随着网络规模的扩大和主干网络带宽的增长,网络安全设备的处理速度已经成为影响网络性能的一大瓶颈。如果检测速度跟不上网络数据的传输速度,那么检测系统就会漏掉部分数据包,导致漏报,进而影响检测系统的准确性和有效性[1]。所以在高速网络流量下,数据逐渐呈现分布式存储的趋势。在这种存储体系结构下,研究网络安全变得日益重要。文献[2]通过使用移动代理技术来实现各节点之间既相互协作又相互独立,从而解决了集中式入侵检测系统的单点失败问题。文献[3]提出了一种基于中间件的分布式异常检测系统来对数据进行分流处理。文献[4]提出了基于神经网络的智能网格入侵检测系统,该系统将不同子网内采集的数据根据一定的资源分配算法确定为其提供服务的分析引擎,然后传输给对应的分析引擎进行处理。本文重点分析在分布式存储体系结构下网络数据间的相关性,并根据数据相关特性建立正常网络活动模型。

网络中传输的数据包90%以上是TCP包,所以目前大部分的研究也是针对TCP流量的。主要方法是采用时间序列分析的方法。文献[5]通过对TCP流的宏观测量,根据TCP流的到达模型和流长模型建立了测量误差模型,提出了时间粒度对模型误差的影响。文献[6]通过叶节点路由器来测量网络中SYN包与SYN+ACK包的数量,以此检测SYN Flooding攻击。文献[7]提出在变化点使用累计和检测均值变化来提高检测性能。

本文在分布式存储的网络环境下,针对网络流量特性提出一种基于DS(Distributed Statistical)的时间序列方法:将待检测的网络数据分布存储,在各节点对数据包进行分类统计,研究节点内部及相邻节点间数据的相关性,并根据经验值判断网络流量异常情况。

2 网络流量及TCP协议特性

2.1 流量特性

在对网络流量建模的过程中,一些传统的模型认为:若时间间隔 s 足够大,则当前时间 t 与过去时间 $t-s$ 时的流量是不相关的,即仅考虑 s 较小时分组到达的数据之间的相关性,

到稿日期:2009-08-19 返修日期:2009-10-26 本文受教育部高校博士点基金(20060183043)资助。
孟凡雪(1985-),女,硕士生,主要研究方向为计算机网络安全、机器学习,E-mail:mengfanxue2007@yahoo.cn;刘衍珩(1958-),男,博士,教授,博士生导师,主要研究方向为计算机通信与网络、移动IP、网络QoS;吴 静(1973-),女,博士生,讲师,主要研究方向为计算机通信与网络、网络安全、神经网络。

为短期相关模型。但实际的网络流量具有统计意义上的自相似性[8]。以太网中流量的突发性[9]并不因数据流的累加或测量间隔的加长而迅速弱化，极端情况(如大量的突发数据、长时间没有流量)发生的概率较高[10]，并且网络通信量的文件长度、通信量大小、FTP 与 TCP 的传输时间、数据包到达时间间隔、通信量突发时间长短等呈现重尾(Heavy-Tailness)分布，即非高斯(Non-Gaussian)分布[11]，大量的概率质量集中在分布带的尾部。最简单的重尾分布是 pareto 分布[12,13]。网络通信量的自相似性和重尾特性共同作用，使网络流量体现出长相关性。

针对流量的长相关性，一种改进的方法是对实际的网络流量平稳化[14]，对处理后的流量应用自回归滑动平均模型(ARMA)[15,16]进行研究。本文的 DS 时间序列是将数据包按到达个数分组统计，化解了流量突增和长时间无流量的情况，使得序列值始终在一定的区间内变化。由于数据是分布存储的，并且根据网络活动的不同，网络流量将表现出不同的性质，因此本文使用分布式自回归滑动平均(DARMA)模型来进行分析。

2.2 TCP 协议特性

本文重点是对数据包类型的统计，所以对网络中主流量的 TCP 协议特征进行分析。

TCP 连接建立要经过 3 次握手。当合适的时间粒度内完整 TCP 流的数量比不完整的 TCP 连接数量大时，可以认为该时间粒度内 TCP 的完整性能得到保证[17]。完整的 TCP 流在一定的时间段内各类 TCP 包在数量上存在以下约束[5]：

$$\begin{cases} N_{SYN}=N_{SYN+ACK} \\ 2N_{SYN}=N_{FIN+ACK} \\ N_{RST}\to 0 \end{cases} \tag{1}$$

式中，N_{xxx}表示含带 xxx 标记的 TCP 包的数量。该式表明，在完整的 TCP 连接建立过程中，第一、二次握手的报文数是平衡的。连接正常拆除时，由于是双方关闭，与建立连接的报文是二倍关系。但是在对数据包进行分组时，难免会截断 TCP 流，所以式(1)是近似的。假设主机受到 DoS 攻击，那么它会在很短的时间内接收到大量的 SYN 包。而对于无保护的服务器，若每秒钟发送 500 个 SYN 包[18]；对于安装了专门防火墙的服务器，若每秒钟发送 14000 个 SYN 包便可消耗主机资源[6]，所以当主机接收的 SYN 包的比例增多时视为异常。

RST 包意为重建 TCP 连接。根据 RFC793 可知，在非拥塞、正常交互的情况下，RST 包出现的概率应该很小。TCP 包中的有些标志段是不可同时置位的，如 SYN 和 FIN，不带 ACK 标志的 FIN 包在正常的网络活动中也很少出现，所以当网络中出现此类非正常包时均可视为异常。

3 基于分布式统计时间序列的 DARMA 方法

对时间序列的分析通常是建立模型来描述其行为特征及变化规律。本文采用按数据包个数分组统计得到的 DS 时间序列来建立模型。该序列的样本在单一节点上不存在递增的趋势。为了保证不出现周期成分，根据样本的均值变化分段使用 ARMA 建立模型，因此可将 ARMA 模型扩展为 DARMA 模型，该模型能较好地体现网络活动的分段特性，每段代表一个连续的网络活动，并可根据模型的参数分析相邻节点数据的相关性。

DARMA 模型如下：

$$\begin{cases} x_{t_1}=\phi_0+\phi_1 x_{t_1-1}+\cdots+\phi_{p_1}x_{t_1-p_1}+\varepsilon_{t_1}-\theta_1\varepsilon_{t_1-1}-\cdots- \\ \quad \theta_{q_1}\varepsilon_{t_1-q_1} \\ \quad \vdots \\ x_{t_n}=\phi_0+\phi_1 x_{t_n-1}+\cdots+\phi_{p_n}x_{t_n-p_n}+\varepsilon_{t_n}-\theta_1\varepsilon_{t_n-1}-\cdots- \\ \quad \theta_{q_n}\varepsilon_{t_n-q_n} \end{cases} \tag{2}$$

式中，$1\leqslant t_1\leqslant k_1$，$E(x_{t_1})=e_1$，$k_{n-1}\leqslant t_n\leqslant k_n$，$E(x_{t_n})=e_n$。

式(2)中 x_{t_i} 为 t_i 时刻样本值。模型有 3 个条件：

条件一　$E(\varepsilon_{t_i})=0$，$Var(\varepsilon_{t_i})=\sigma_\varepsilon^2$，$E(\varepsilon_{t_i}\varepsilon_{s_i})=0$，$s_i\neq t_i$，要求各随机干扰序列$\{\varepsilon_{t_i}\}$为零均值白噪声序列；

条件二　$Ex_{s_i}\varepsilon_{t_i}=0$，$\forall s_i<t_i$，这个条件说明当前的随机干扰与过去的序列值无关；

条件三　$|e_i-e_{i-1}|<\alpha$，这个条件说明根据均值的变化对原序列分段，α 是一个给定的常数，可以根据实际经验调整。

DARMA 模型建模步骤如下。

步骤 1　对样本分段。

令 $y=||x_t-\bar{x}|-|x_{t+1}-\bar{x}||$，若 y 值大于 α，继续观察接下来的连续 β 个点，令 $y_i=||x_t-\bar{x}|-|x_{t+i}-\bar{x}||$，$2\leqslant i\leqslant\beta$，若 $y_i>\alpha$，则在 t 点将序列分段，否则从 $t+1$ 点重新计算。该 β 值也由经验给出，$\bar{x}$ 为样本均值。

步骤 2　求出各段观察值样本的自相关系数和偏自相关系数。

$$\hat{\rho}_k=\frac{\sum_{t=1}^{n-k}(x_t-\bar{x})(x_{t+k}-\bar{x})}{\sum_{t=1}^{n}(x_t-\bar{x})^2},\ \forall 0<k<n \tag{3}$$

式中，$\hat{\rho}_k$ 为延迟 k 自相关系数。

$$\hat{\phi}_{kk}=\frac{\hat{D}_k}{\hat{D}},\ \forall 0<k<n \tag{4}$$

式中，$\hat{\phi}_{kk}$ 为延迟 k 偏自相关系数，即在给定中间 $k-1$ 个随机变量 x_{t-1}，x_{t-2}，…，x_{t-k+1} 的条件下，或者说在剔除了中间 $k-1$ 个随机变量 x_{t-1}，x_{t-2}，…，x_{t-k+1} 的干扰之后 x_{t-k} 对 x_t 影响的相关度量。

其中，

$$\hat{D}=\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-2} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & 1 \end{vmatrix}$$

$$\hat{D}_k=\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_2 \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & \hat{\rho}_k \end{vmatrix}$$

式中，$\hat{D}_k$ 是将$\hat{D}$中的第 k 个列向量转换成列向量$(\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_k)'$后的行列式。

步骤 3　根据每段样本的自相关系数和偏自相关系数，选择阶数适当的 DARMA 模型。

如果样本相关系数在最初的 d 阶明显大于 2 倍样本标准差范围，并且几乎 95%的相关系数都落在 2 倍标准差的范围

以内，由非零相关系数衰减为小值波动的过程非常突然，这时通常视为相关系数截尾，截尾阶数为 d；反之，视为相关系数不截尾，即拖尾性。

当自相关系数拖尾、偏自相关系数 p 阶截尾时，该段模型定阶为 ARMA(p,0)，即 AR(p)自回归模型；当自相关系数 q 阶截尾、偏自相关系数拖尾时，该段模型定阶为 ARMA(0,q)，即 MA(q)移动平均模型；当自相关系数和偏自相关系数都拖尾时，应用 SBC 准则，该准则定义为 SBC=$-2\ln$(模型的极大似然函数值)$+\ln(n)$(模型中未知参数个数)，在所有模型中使得 SBC 函数达到最小的模型为相对最优模型，从而确定模型的阶数。

步骤 4　估计模型中的未知参数的值。

每段中待估参数共有 $p+q+2$ 个，分别为 $\phi_1,\cdots,\phi_p,\theta_1,\cdots\theta_q,\mu,\sigma_\varepsilon^2$，其中 μ 是序列均值，σ_ε^2 是随机干扰序列方差。实际中经常使用条件最小二乘估计法，使残差平方和达到最小的那组参数值即为最小二乘估计值。并假定当 $t<0$ 时 $x_t=0$，即假定过去未观测到的序列值等于零。

残差平方和方程定义如下：

$$Q(\tilde{\beta})=\sum_{i=1}^{n}\varepsilon_t^2=\sum_{t=1}^{n}(x_t-\phi_1x_{t-1}-\cdots-\phi_px_{t-p}+\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q})^2$$

$$=\sum_{i=1}^{n}[x_t+\sum_{i=1}^{t}\pi_ix_{t-1}]^2 \tag{5}$$

式中，n 表示样本个数，π_k 由式(6)递归定义：

$$\begin{cases}\pi_0=1\\ \pi_k=\sum_{j=1}^{k}\theta_j\pi_{k-j}-\varphi_j\end{cases} \tag{6}$$

在本文中，由于使用了分段建模，因此可以提供过去未观测到的序列值，即当 $t<0$ 时，由本节点的前一段模型或前一节点来提供 x_t 的值，残差平方和方程可以扩展为：

$$Q(\tilde{\beta})=\sum_{i=1}^{n}\varepsilon_t^2=\sum_{i=1}^{n}[x_t+\sum_{i=1}^{\infty}\pi_ix_{t-1}]^2 \tag{7}$$

π_k 由式(8)递归定义：

$$\begin{cases}\pi_0=1\\ \pi_k=\sum_{j=1}^{k}\theta_j'\pi_{k-j}-\varphi_j'\end{cases} \tag{8}$$

式中，$\theta_j'=\begin{cases}\theta_j, & 1\leqslant j\leqslant q\\ 0, & j>q\end{cases}$，$\varphi_j'=\begin{cases}\varphi_j, & 1\leqslant j\leqslant p\\ 0, & j>p\end{cases}$

式(7)使用了更多的样本值来估计参数，可以使模型更精准。

4　实验与结果分析

在一台服务器上随机截获网络层数据包，将得到的数据经过分类统计后作为 DS 时间序列的样本，考虑样本的分布式存储，分析数据之间的相关性。截包持续的时间为 1h，得到数据包 316719 个。随机选取其中 300000 个包进行分析。

将数据包按不同数目分组。由于消除了网络流量的突发性，不同粒度组中，被分组截断的 TCP 流数与 TCP 流的长度和 TCP 流的重尾分布性有关。文献[19]指出当前大多数 TCP 连接都是短暂的(通常传输量在 10kB 以下)，约为 7 个报文段长度。通过观察在不同的统计粒度下网络主流量 ACK 包分类的关系，发现对数据包进行 500 个/组、600 个/组、800 个/组和 1000 个/组的不同分组粒度下得到的序列特性大致一样，所以在以下的实验中采用 500 个包作为一组。

4.1　ACK 包流量分析

由于 TCP 建立连接后，传送的数据包 ACK 位都被置位，因此网络中传输的大部分是 ACK 包，所以对 ACK 序列进行重点分析。图 1 中横坐标代表分组后的第 i 组包，纵坐标代表统计的类型占该分组的百分比。如点(x,y)代表经分组后第 x 组中统计的类型占该分组 500 个包的 y%。

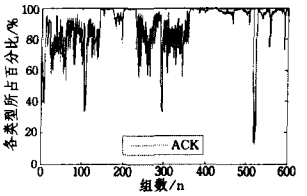

图 1　正常网络活动下的 ACK 包流量

经过计算，每组中 ACK 包个数的均值在 380 至 440 之间，ACK 包占网络流量的 76%～88%。当连续几个点的 ACK 流量小于 60%时，说明网络中正过多传送其他类型数据包，可以进一步观察以决定是否报警。由图 1 中可以观察到在第 100,300 和 520 组数据时有极小点，但流量小于 60%的点都不超过 10 个点，可以视为正常。通过验证实际的网络传输情况，这 3 个极小点是监听 BT 程序的端口大量传输 UDP 包所造成的。

①考虑单一节点，即以上统计数据存储在一个节点上。

根据 ACK 均值的变化对其分段，分段后的 ACK 序列没有明显的周期性，也没有明显的线性关系，并且自相关系数很快收敛到两倍标准差内，因此可以认为是平稳的。使用 DARMA 建模，定义 α 为 30，β 为 10，得到如下模型：

$$\begin{cases}x_{t_1}=383.85896+0.5816x_{t_1-1}+\varepsilon_{t_1}\\ x_{t_2}=438.20792+0.6816x_{t_2-1}+\varepsilon_{t_2}\\ x_{t_3}=394.31038+0.6236x_{t_3-1}+\varepsilon_{t_3}\\ x_{t_4}=482.38606+0.57261x_{t_4-1}+\varepsilon_{t_4}\\ x_{t_5}=306.67054+0.97868x_{t_5-1}+\varepsilon_{t_5}\end{cases} \tag{9}$$

式中，$1\leqslant t_1\leqslant 108$，$E(x_{t_1})=385.1852$；$109\leqslant t_2\leqslant 264$，$E(x_{t_2})=448.6538$；$265\leqslant t_3\leqslant 343$，$E(x_{t_3})=401.1139$；$344\leqslant t_4\leqslant 516$，$E(x_{t_4})=485.9595$；$517\leqslant t_5\leqslant 600$，$E(x_{t_5})=454.6071$。

由式(9)可以看出，该模型各分段的阶数是一致的，均为 AR(1)模型，参数根据均值的变化有所不同。该分段模型建立后，也可以反映当前网络活动情况。当形成正常网络模型后，可将后续接收到的数据包单独建模，得到的参数与已有模型匹配。存在某一段模型的参数在 95%置信区间内与其匹配则无异常。

②考虑相邻节点之间的关系。

每个节点的模型是分段的，将本节点模型的最后一段参数传递给下一节点，与下一节点模型的第一段参数比较。根据 TCP 流的长相关性，这两部分的差异不会很大。

4.2　RST 包流量分析

图 2 显示 RST 包的数量并不趋近于 0。Arlitt 等人[20]以 HTTP 为依托研究 RST 报文产生的机理和行为，发现不同的浏览器和服务器的行为会影响其数量。Windows XP 和 IE6.0 使用RST 终止连接的比例高达 60.4%，因此实际观察到的 RST 包的数量由于浏览器的异常行为会高于期望值，同时会减少 FIN+ACK 包的数量。若其范围不超过总体流量的 5%(2*4%*60.4%)，则视为正常。

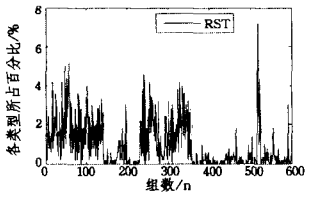

图 2　正常网络活动下的 RST 包流量

4.3　TCP 连接的建立和关闭的流量分析

图 3 显示 TCP 连接建立过程的流量特性。整体上 SYN 包与 SYN+ACK 包的数量基本一致，这符合 TCP 连接建立的唯一性。即使在 SYN 包数量突增的高峰点，也不超过整体网络流量的 4%，证明网络活动的正常性。尽管由于理论及实际上的 RST 包数量异常，SYN+ACK 包与 FIN+ACK 包的数量在图 3 仍基本符合 2 倍关系。

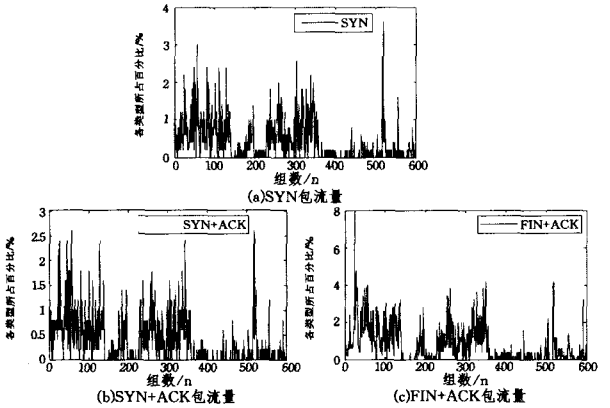

图 3　正常网络活动下 SYN 包、SYN+ACK 包与 FIN+ACK 包的关系

4.4　UDP 包和 ICMP 包的关系分析

图 4 显示 ICMP 包与 UDP 包在同一段有峰值。通过对比可知，ICMP 包负责 UDP 包的差错传输控制。由于该 ICMP 包所占比例一直很小，说明网络活动是正常的。同时说明，除了峰值，UDP 包约占网络流量的 10%，ICMP 包是 UDP 包的 10%左右。

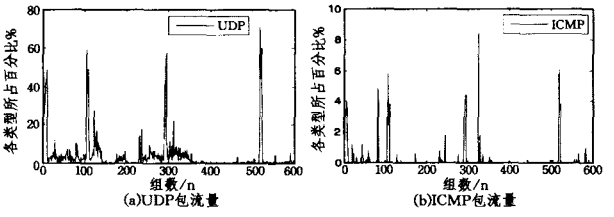

图 4　正常网络活动下 UDP 包与 ICMP 包的关系

4.5　实验结果分析

由上述实验结果可以看出，与传统的时间序列分析方法相比，本文提出的 DS 时间序列方法能够体现网络流量中协议的特性。

1)TCP 流是网络的主流量，且 ACK 包在检测数据中占主导地位。但是考虑到其他协议(如 UDP)可能突发传送数据，所以当超过一定阈值的连续的 ACK 流量低于正常值时才应考虑报警。该阈值应根据实际经验和网络环境进行调整，以控制漏报率和误警率。若 ACK 的均值在正常范围内，为了分析网络活动情况，可以使用 DARMA 方法建立模型，通过与已有模型匹配来判断当前网络行为是否正常。

2)重建连接的 RST 包的数量在正常的网络流量下应该趋于 0。但由于服务器和浏览器的影响，使得网络中的一部分 TCP 连接通过 RST 包来终止。本文通过实验数据给出了 RST 包数量的上限。

3)实验的分组粒度能够保证 TCP 流的完整性。TCP 连接过程中的 3 次握手和终止时双方关闭连接的特性在 DS 时间序列中体现为：

$$\begin{cases} N_{SYN} \geqslant N_{SYN+ACK} \\ 2N_{SYN+ACK} \geqslant N_{FIN+ACK} \end{cases} \tag{10}$$

式中，SYN+ACK 包的数量小于 SYN 包的数量，这是 SYN 包的丢失和重传造成的；SYN+ACK 包的数量 2 倍于 FIN+ACK 包的数量，这是 RST 终止连接造成的。

4)ICMP 包传送差错报文以及其他控制信息[21]。若其数量超过一定比例，也可视为网络异常。UDP 是面向无连接的协议，通过目的地址和端口通知对方主机。UDP 和 ICMP 的关系是：若目的主机没有程序在特定的端口上侦听，UDP 通知 ICMP 模块，“ICMP 目的地址不能到达；目的端口不能到达”消息被返回给发送方，并且将该 UDP 包丢弃。通过上述实验结果也可知 UDP 包的数量与 ICMP 包的数量有一定的相关性。

4.6　模型检测

在同等的实验环境和正常的网络流量下，使用攻击工具 FakePing 和 ddnsf 来观察流量统计情况。随机选取 14min 截获的 204004 个数据包进行分析。

如图 5 所示，前半段是 FakePing 攻击，攻击机发送大量的 ICMP 包，同时服务器回应 ICMP 包；后半段是 ddnsf 攻击，攻击机发送大量协议字段为 03 的非正常 IP 包。

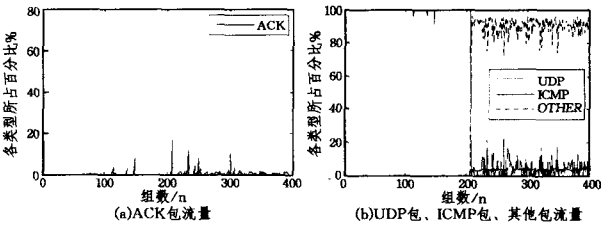

图 5　FakePing 和 ddnsf 攻击下的流量统计情况

图 5 所示的两种攻击的直接结果是 ACK 包的流量严重小于正常范围。当检测到连续 10 个点的 ACK 包流量小于 60%时，即可通知报警。此时通过查询其他类型包的流量，可知不是短期突发的 UDP 包在主要传输，可以判断为是某一类型的 DoS 攻击。

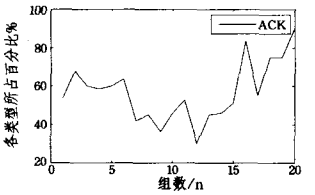

图 6　snot 攻击下的 ACK 包流量

为了测试模型在轻量级攻击下的检测情况。使用攻击工具 snot 观察流量统计情况，随机选取持续攻击 24min 所截获的 12345 个数据包。同样定义 α 为 30，β 为 10，该 ACK 序列不需再分段。由图 6 可知，ACK 包的数量小于正常范围。在计算出 ACK 包的均值是 285.05，只占网络流量的 57%后，可以省去使用 DARMA 建模，从而提高检测的实时性。从图 7

和图 8 可知，ICMP 包数量基本等于 UDP 包数量，并且超出正常范围，SYN 包与 SYN+ACK 包的数量不匹配。这些都可作为异常触发条件，并由此判断攻击类型。

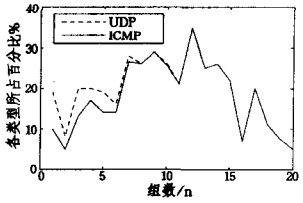

图 7 snot 攻击下 UDP 包与 ICMP 包的关系

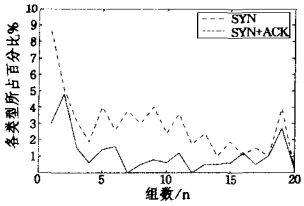

图 8 snot 攻击下 SYN 包与 SYN+ACK 包的关系

结束语 本文针对数据分布式存储的发展趋势，通过对网络数据包的分组统计，提出了一种基于分布式统计(DS)的时间序列方法来分析网络活动；使用平稳时间序列的 DARMA 模型对分类数据中的 ACK 序列建立分段模型，提出了一种分析单一节点和相邻节点之间数据相关性的方法。通过分析各类数据包之间的关系，对其所占比例给出经验阈值并确定置信区间，当统计结果出现异常时选择报警。最后使用 3 种攻击工具对建立的模型进行检测，结果表明此模型能够发现大规模及轻量级的 DoS 攻击，同时能够检测出针对某一特定协议的攻击。未来的工作将针对 ACK 序列在分布节点间的相关性进行研究，并确定前一个节点 DARMA 模型的最后一段与下一节点 DARMA 模型参数匹配的置信区间。

参 考 文 献

[1] 肖政宏，尹浩. 基于网络流量统计分析的入侵检测研究[J]. 微电子学与计算机，2006，23(5)：76-78，82

[2] 张淑英，刘淑芬. 基于移动代理的多层次分布式入侵检测网络预警系统[J]. 计算机应用研究，2006，23(50)：13-15

[3] 陈宁军，倪桂强，潘志松. 基于中间件的分布式网络异常检测系统[J]. 微电子学与计算机，2006，23(z1)：15-17，19

[4] 魏宇欣，武穆清. 智能网格入侵检测系统[J]. 软件学报，2006，17(11)：2384-2394

[5] 龚俭，彭艳兵，杨望. TCP 流的宏观平衡性[J]. 计算机学报，2006，29(6)：1561-1571

[6] Wang H N，Zhang D L，Shin K G. Detecting SYN Flooding Attacks[A]// Proceedings of IEEE Infocom 2002[C]. New York，2002：1530-1539

[7] Siris V A，Papagalou F. Application of Anomaly Detection Algorithms for Detecting SYN flooding attacks[A]// Proceedings of IEEE *Global Telecommunications Conference*[C]. Dallas，2004：2050-2054

[8] 任勋益，王汝传，王海艳. 基于自相似检测 DDoS 攻击的小波分析方法[J]. 通信学报，2006，27(5)：6-11

[9] 石江涛，王永纲，戴雪花. 自相似网络业务流量的研究与实现[J]. 通信学报，2005，26(6)：112-117

[10] 刘武，卢曦，胡保民. 一种以太网流量的仿真方法[J]. 光通信技术，2006，30(5)：7-9

[11] 闻勇，朱光喜. 自相似网络通信量的滑动平均预测[J]. 计算机科学，2006，33(6)：32-34，38

[12] Kherani A A，Kumar A. Long range dependence in network traffic and the closed loop behaviour of buffers under adaptive window control[J]. Performance Evaluation，2005，61(2/3)：95-127

[13] 杨荣根，龚乐君. 网络仿真中的自相似流量发生模型研究[J]. 中国制造业信息化，2006，35(19)：9-11，15

[14] 邹柏贤，姚志强. 一种网络流量平稳化方法[J]. 通信学报，2004，25(8)：14-23

[15] Basu S，Mukherjee A，Klivansky S. Time series models for Internet traffic[A]// Proceedings of INFOCOM[C]. San Francisco，CA，March 1996，2：611-620

[16] Sang A，Li S Q. A predictability analysis of network traffic [A]// Proceedings of IEEE INFOCOM 2000 [C]. Tel-Aviv，Israel，March 2000：342-351

[17] 龚俭，彭艳兵，杨望. 基于 Bloom Filter 的大规模异常 TCP 连接参数再现方法[J]. 软件学报，2006，17(3)：434-444

[18] 林白，李鸥，赵桦. 基于源端网络的 SYN Flooding 攻击双粒度检测[J]. 计算机工程，2005，31(10)：132-134

[19] 赵炯，周其刚，张树京. TCP 启动阶段的建模研究[J]. 计算机工程，2003，29(8)：19-21，31

[20] Arlitt M，Williamsong C. An analysis of TCP reset behavior on the Internet [J]. ACM SIGCOMM Communications Review，2005，35(1)：37-44

[21] Yoo I S. Protocol anomaly detection and verification[A]// 2004 IEEE Workshop on Information Assurance and Security[C]. New York，2004

(上接第 78 页)

[10] Ye D，Parlos A G. Predictive path switching control for improving the quality of service in real-time applications[J]. IEEE Journal Selected Topics in Signal Processing，2007，1(2)：308-318

[11] Javed U，Suchara M，He J，et al. Multipath protocol for delay-sensitive traffic[C]// Proc. International Conference on Communication Systems and Networks. January 2009

[12] Palomar D，Chiang M. A tutorial on decomposition methods for network utility maximization[J]. IEEE J. on Selected Areas in Communications，2006，24(8)：1439-1451

[13] He J，Suchara M，Bresler M，et al. Rethinking Internet traffic management：From multiple decompositions to a practical protocol [C]// Proc. CoNEXT. December 2007

[14] Xu D，Chiang M，Rexford J. Link-state routing with hop-by-hop forwarding can achieve optimal traffic engineering[C]// Proc. IEEE INFOCOM. May 2008

[15] Li R，Ying L，Atilla E. A unified approach to optimizing performance in network serving heterogeneous flows[C]// Proc. IEEE INFOCOM. 2009