

基于自相似特性的局域网流量异常评估方法

李国栋¹, 秦 涛¹, 褚 懿²

(1 西安交通大学, 陕西 西安 710049; 2 中国武装警察部队北京总队司令部 通信网络管理中心, 北京 100039)

摘 要: 针对现有的局域网评估模型不能提供局域网系统中数据传输安全的有效信息的缺陷, 文中提出了一种通过分析局域网网关流量的自相似性变化, 来对局域网系统的数据传输安全进行合理评估的方法. 该方法通过选择合适的时间尺度, 利用 whittle 估计器法计算多种测度的自相似系数, 判断测度值是否偏离正常值范围来综合分析局域网流量的异常状况. 在此基础上, 提出了改进的局域网系统安全评估模型. 实验证明, 此方法能够反映网络攻击等多种原因导致的流量异常变化, 提供给管理员直观的数据传输安全态势图.

关键词: 网络安全评估; 网络流量; 自相似; whittle 估计器

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2011)09-0220-05

Study on Security Evaluation for Data Transmission in LAN System

LI Guo-dong¹, QIN Tao¹, CHU Yi²

(1 Xi'an Jiaotong University, Xi'an 710049, China;

2 Chinese People's Armed Police Force of Beijing, Beijing 100039, China)

Abstract: Aiming at the deficiency that unable to provide effective security information about network data transmission information in current LAN security evaluation system, a security evaluation method, which analyzes the self-similarity of traffic on router, is proposed. After choose the suitable time-scale, whittle estimator is applied for calculate the self-similar parameters of four evaluate metrics. We study the abnormal situation of traffic on LAN by calculate the distances between the self-similar parameters and normal values. A reformative model for evaluates the LAN's security is given on the base of before-mentioned method. Experiment results show that our method is able to evaluate the abnormal change on traffic caused by several reasons. It provide system administrator with intuitive security curve that expresses the risk of network system data transmission.

Key words: network security assessment; traffic pattern; self-similar; whittle estimator

1 引言

近年来,伴随着互联网规模的飞速发展,网络安全事件的数目呈急剧上升的势头. 人们对网络安全技术的关注与日俱增. 网络安全与风险评估已经成为防火墙、入侵检测技术、安全审计之外网络安全领域的一个新的研究热点.

目前对入侵检测的研究多注重于检测数据源特征和用户行为异常. 虽然基于网络的入侵检测系统

能够同时保护多台网络主机,例如 ISS RealSecure Engine,但是要对每个数据包进行解析,在规模较大的局域网中的执行效率和在交换式局域网中的配置都是面临的难点. 现有的网络入侵检测系统很少把局域网的流量特性作为检测数据源. 局域网中的流量状况很大程度反映了局域网中主机的网络服务可用性,许多因素都将导致主机网络访问性能的降低,例如许多网络攻击、硬件设备故障等.

现在,网络基本性能和服务质量参数的评估、基

收稿日期: 2011-02-01; 修回日期: 2011-03-05

基金项目: 国家自然科学基金项目(60574087); 国家“八六三”计划项目(2008AA01Z415); 国家“十一五”科技支撑计划重点课题(2006BAK11B02)

于系统安全漏洞的威胁评估等领域的研究已经取得了一些成果^[1,2,3]. 文献[4]给出了一种通过计算流量样本的自相似系数 H , 结合网络攻击的威胁权重来对网络中数据传输的安全性进行估算的方法. 然而适用的前提是所有的底层路由器均采用统一的权重标准^[4], 而权重标准的确定仅仅由经验得出, 而且估算数据传输路径的安全时, 对报文需经过的所有路由器均要进行评估, 在实用中是相对困难的. 文中认为在一个规模相对较小的局域网或者自治系统中, 进行较为全面的安全评估是更为可行和合理的.

文中通过研究局域网内多种流量测度的自相似性变化规律, 以及分析多测度自相似系数偏离正常值范围的情况, 给出了一种评估局域网的数据传输安全的方法.

2 局域网流量的自相似性分析

2.1 相似性定义及自相似系数的计算方法

定义: 一个随机过程 $x(t)$, $-\infty < t < \infty$, 如果这个随机过程在时间上压缩或者扩展时, 统计特性不变, 称为是统计相似的. 对于一个随机过程, 如果满足 $X(t) = a^{-H}x(at)$, 就可以称为是自相似的, 参数 H 称为自相似系数或 Hurst 系数. 根据上面的定义, 自相似的网络流量是指无论以哪个时间尺度作为基本测量单位去观察, 网络流量都具有相似的特征.

如果协方差平稳过程 $\{X_i\}$ 的自相关函数 $\gamma(k)$ 不收敛, 即 $\lim_{n \rightarrow \infty} \sum_{k=1}^n \gamma(k) \rightarrow \infty$, 则称平稳过程 $\{X_i\}$ 具有长相关特性(Long Range Dependence, LRD). 在长相关过程中, 距离较远的数据的相关性是不可忽略的, 而传统基于马尔可夫(Markov)过程的网络流量模型认为长距离的相关性为零, 即 $\lim_{n \rightarrow \infty} \sum_{k=1}^n \gamma(k) \rightarrow 0$, 只具有短相关特性.

对于一个有限方差过程, 当 $H = 0.5$ 时, 过程完全不具备长相关特性, 当 $0.5 < H < 1$ 时, 过程出现长相关特性; 当 $0 < H < 0.5$ 时, 过程呈现短相关特性. 可以证明, 渐进二阶自相似过程实际上是长相关过程的一个子集^[5]. 在下文中, 在不致混淆的前提下, 对长相关和自相似的概念不加区分.

计算有限或无限时间序列的自相关指数有很多方法^[6]. 其中聚合序列方差法(aggregated variance)、周期图方法、R/S方法、Whittle估计器(Whittle estimator)都是计算网络流量自相似特性

的常用方法^[7]. 与其他方法相比较, Whittle估计器方法除了能够判断网络流量是否具有自相似特性, 还能够在此基础上估计 Hurst 系数的精确值, 并计算相应的置信区间. 使用此方法计算随机序列的 Hurst 系数时, 首先要计算观测数据的频谱图.

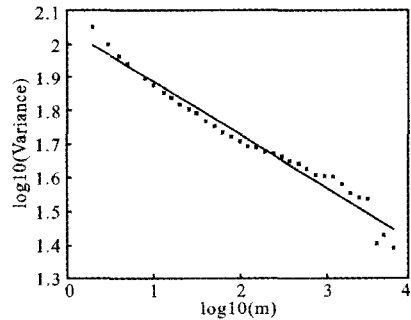
$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{j=1}^N X_j e^{i\lambda j} \right|^2 \quad (1)$$

其中, λ 是频率, N 是序列的长度. $I(\lambda)$ 是对谱密度的估计, 频谱图应大致与 $|\lambda|^{1-2H}$ 成正比, 然后计算使 $Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda; \eta)} d\lambda$ 取最小值的 η , 确定 $f(\lambda, \eta)$ 的形式, 进而得到 H .

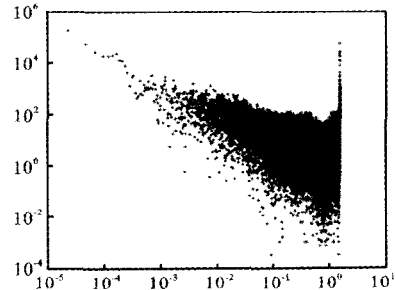
2.2 局域网流量自相似系数的计算

先前的研究表明^[8], 包括局域网、广域网、ISDN、VBR 视频业务、ATM 网络中传输的视频会议业务等相当广范围的网络流量数据存在着自相似现象. 在 CNSIS(Center for Networked Systems and Information Security) 局域网一段正常流量数据, 即以单位时间尺度内数据包数形成的随机序列上, 分别使用聚合序列方差法^[2]和 Whittle 方法.

根据聚合序列方差法作出的对数图求得斜率 $\beta = -0.1571$, $H = 0.9214$, 如图 1(a). Whittle 方法计算出的频谱图如图 1(b) 所示, 求得 $H = 0.8427$. 两种方法得出相同的结论即数据具有较明显的自相似特性.



(a) 聚合序列方差法画出的 $\log \text{Var}X(m) - \log m$



(b) Whittle 估计法计算出的频谱图

图1 两种方法计算自相关系数结果图

将不同协议的数据包流量作为研究对象, 发现

TCP 协议、UDP 协议、ARP 协议的网络流量数据同样具有各自独立的较明显的自相似特性. 因此, 把不同类型协议的数据包作为局域网流量安全评估的测度, 可以分析恶意的攻击引发各测度的自相似系数的变化情况.

3 局域网流量的安全评估方法

局域网流量风险指数是选用的网络流量的评估测度函数, 即

$$F_{\text{traffic}} = \sum_{l=1}^4 w_l F_{M_l}, \quad \sum_{l=1}^4 w_l = 1 \quad (2)$$

其中, w_l 为测度 M_l 的相对权重. $M = [\text{总数据包测度}, \text{TCP 协议数据包测度}, \text{ARP 协议数据包测度}, \text{UDP 协议数据包测度}]$. 根据实验网络流量的具体情况, TCP 协议在总流量中占比例较高, 即总流量和 TCP 协议流量的相关性较高, 将权重向量设置为 $w = [0.2, 0.2, 0.3, 0.3]$. F_{M_l} 为根据式 (2) 计算出的网络流量评估测度 l 的风险指数, 如果当前时间段内第 l 个测度的取值在正常值的 95% 置信区间内, 异常度值为 0; 反之为 1.

4 实验及结果分析

在 CNSIS 网络安全实验室中搭建了局域网如图 2 所示.

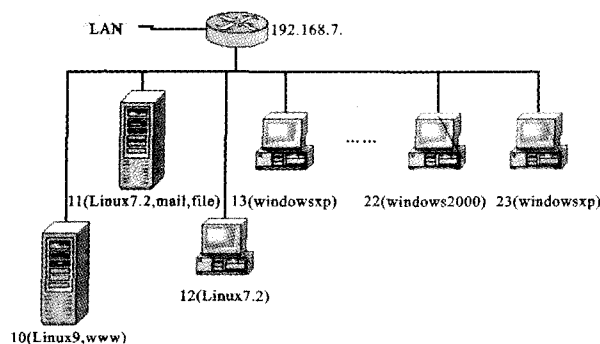


图 2 实验网络拓扑图

其中主机 10 为 Linux9 提供 www 服务器, 11 为 mail 服务器兼做文件服务器, 12 是 Linux 用户, 13 至 23 是 windows2000 或 windowsxp 的普通网络用户, 在网关处取得该局域网的流量数据. 根据文献[2], 分别以 10 秒、1 秒、0.1 秒等几种不同的时间尺度来分析数据, 可以发现: 当时间尺度取 0.1 秒或者更小的时候可以体现出明显的自相似特性, 当时间尺度放大时, 网络流量更接近马尔可夫过程, 当时间尺度放大到 10 分钟以上时, 则体现了一天当中网络访问繁忙和空闲的周期变化. 在实验中, 选用

0.1 秒作为时间尺度, 分析正常网络环境下, 2 小时内网络数据包总数、TCP 协议的数据包数, UDP 协议的数据包数、ARP 协议的数据包数四项指标的自相似特性. 经过 7 天的抓包发现, 四项测度均具有稳定的自相似特性, 自相似系数分别如表 1 所示.

表 1 各测度 Hurst 系数的正常值

协议种类	自相似系数(Hurst)
TCP	0.8442
UDP	0.5948
ARP	0.5206

各测度的正常值虽然相对稳定, 但也不是一成不变的. 当各测度值均在正常范围内时, 新的测度正常值将由原来的正常值和此时的测度值按 0.5 加权得到, 通过这样的方法, 各测度的正常值可以得到更新.

在某天的上午和下午, 分别进行了一次 SYN-flood 拒绝服务攻击和 trino 攻击, Trino 是基于 UDP flood 的攻击软件, 它向被攻击目标主机随机端口发送全零的 4 字节 UDP 包, 被攻击主机的网络性能在处理这些超出其处理能力垃圾数据包的过程中不断下降, 直至不能提供正常服务甚至崩溃.

在实验过程中, 每 2 小时对在该时间段的数据进行一次分析. 在 SYN-flood 拒绝服务攻击和 Trino 攻击过程中四项指标与正常值范围的对应关系如图 3、图 4 所示. 其中虚实两条曲线分别表示自相似系数的置信上限和下限, 它们与竖虚线的交点之间的线段为该度量正常值的 95% 的置信区间^[9].

从图 3、图 4 可以看出, 在进行 SYN-flood 攻击时, 总数据包测度值为 0.8942, TCP 协议数据包测度值为 0.9801, 均明显超过正常值的置信区间. 而在 Trino 攻击过程中, 总数据包指标和 UDP 协议数据包指标超过了正常值的置信区间.

利用第 3 节提到的方法, 就可以得到一天内连续 16 个小时的各指标异常度曲线图和网络数据传输安全性态势图, 如图 5 所示. 例如, 第 2 至第 4 小时的数据传输异常度计算方法如下:

$$F_{\text{traffic}} = \sum_{l=1}^4 w_l F_{M_l} = 0.2 \times 1 + 0.2 \times 1 + 0 \times 0.3 + 0 \times 0.3 = 0.4$$

从图 5 可以看出, 在恶意攻击发生的第 2 至第 4 小时和第 12 至 14 小时, 局域网系统数据传输的安全风险估算值明显偏高. 攻击事件产生了大量规律性的异常数据包, 改变了流量的自相似特性, 评估

模型较好的捕获了这种异常变化,数据传输安全态势图具有直观性,可以准确的反映网络中是否存在

异常,给网络管理员提供参考.

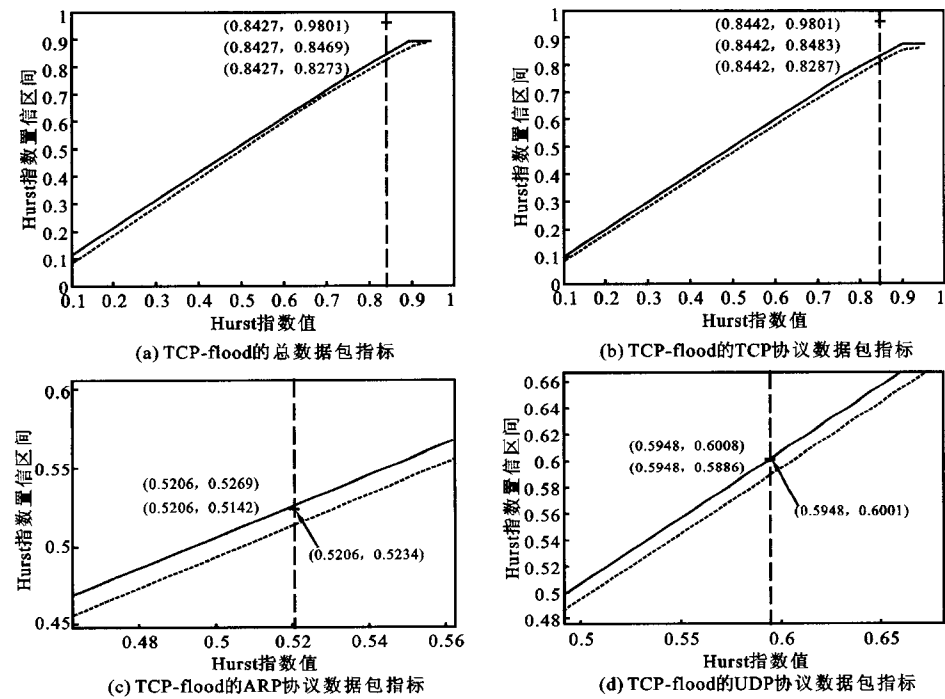


图 3 TCP-flood 攻击时期各测度值偏离正常范围情况

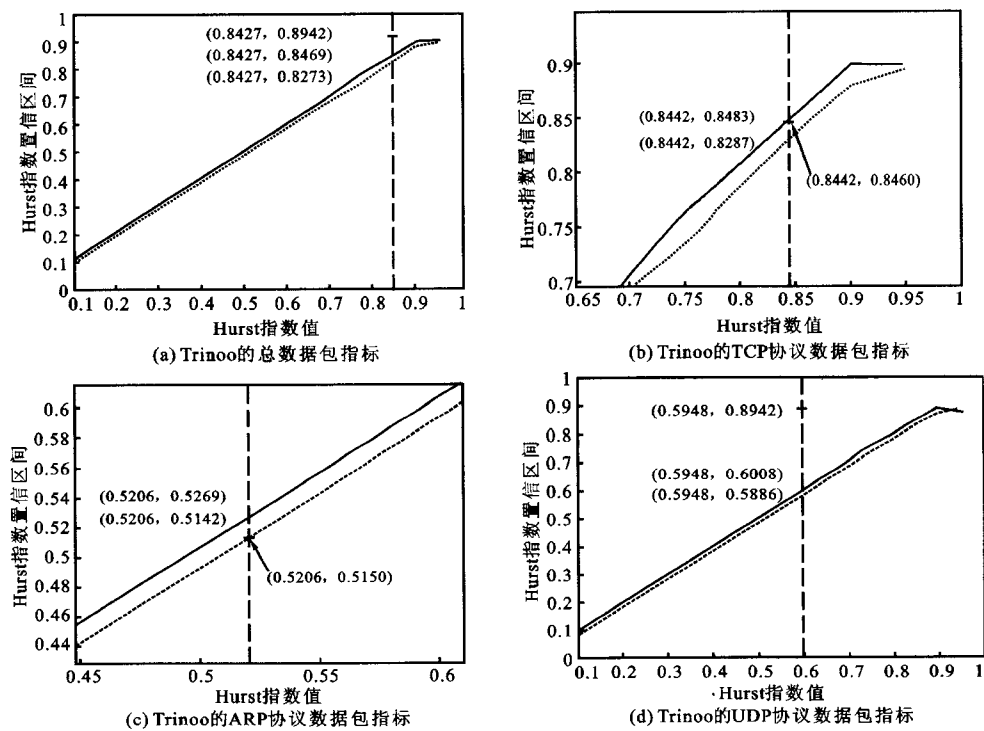


图 4 Trinoo 攻击时期各测度值偏离正常范围情况

5 结束语

文中提出了利用多个网络流量度量的自相似系

数变化来评估流量安全的方法. 实验证明,该方法对于利用不同协议漏洞,引起流量异常变化的拒绝服务攻击有良好的反映效果. 同时,能够直观合理的分

析和评估局域网系统的数据传输安全状况. 下一步的工作是把文中提到的方法与较长期的基于主机安全的局域网评估模型[1]实际结合; 另外实验中的正常值是通过长时间的实验获取的, 如果改进为能够随着网络环境的变化而进行自适应修正, 将会取得更好的效果.

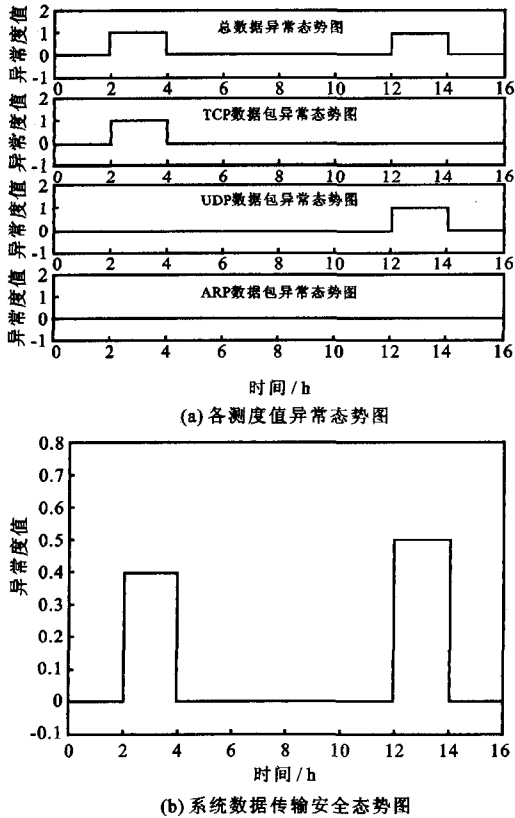


图5 数据传输安全态势图

参考文献:

[1] Paxson V, Almes G, Mathis M, et al. Framework for IP performance metrics. [R]. USA: ACM, RFC2330. 1998.
[2] S Brocklehurst, Littlewood B, Olovsson T, et al. On measurement of operational security. [J]. Aerospace and

Electronic Systems Magazine, 1994, 9 (10): 7-16.

[3] 陈秀真, 郑庆华, 管晓宏, 等. 网络化系统安全态势评估的研究[J]. 西安交通大学学报, 2004, 38(4): 404-408.
[4] Hu Hanping, Guo Wenxuan, Zhang Baoliang, et al. A method of security measurement of the network data transmission[J]. proceeding of the 19th IEEE International Parallel and Distributed Processing Symposium. Washington. DC, USA: IEEE Computer Society, 2005: 1530-2075.
[5] Park k, willinger w. Self-similar network traffic and performance Evaluation Honoken[M]. NY: nj wiley-Interscience, 2000.
[6] Murad S Taquu, Vadim Teverovsky. On estimating the intensity of long-range dependence in finite and infinite variance time series. A Practical Guide to Heavy Tails [Z]. Statistical Techniques and Applications. Birkhauser, Boston: 1998: 177-217.
[7] Will E Leland, Murad S Taquu, Walter Willinger et al. On the self-similar nature of Ethernet traffic (extended version)[J]. IEEE/ACM Transaction on Networking, 1994, 2(1): 1-46.
[8] Jong Suk R Lee, Hae-Duck J Jeong. Practical modeling for generating self-similar VBR video traffic[C]// proceedings of the 5th International Conference Computational Science - ICCS 2005, Atlanta, GA, USA: IEEE, 2005: 22-25.
[9] 汪荣鑫. 数理统计[M]. 西安: 西安交通大学出版社, 1986.

作者简介:

李国栋 男, (1974-), 博士研究生. 研究方向为计算机网络安全.
秦涛 男, (1982-), 博士研究生. 研究方向为计算机网络安全.
褚懿 男, (1972-), 助理工程师. 研究方向为通信及计算机系统.