小波及网络异常行为分析*

刘 兰[1,2]，李之棠[1]，李家春[3]，谭晓玲[4]

(1. 华中科技大学 计算机学院，湖北 武汉 430074；2. 广东技术师范学院 电子系，广东 广州 510655；3. 华南理工大学 计算机学院，广东 广州 510641；4. 重庆三峡学院 电子系，重庆 404000)

摘 要：小波分析能将时间域上重叠但频率域上不重叠的信号进行分段，不同的频段(也就是不同的小波系数层次)代表了信号中处于该频段的信号分量，网络流量具有时域中频域重叠的特征。基于小波分解和重构思想，提出采用小波方法对于网络行为中的异常行为进行分析、判别的方法和模型，模型在模拟分析中取得较好的效果。

关键词：异常；小波；神经网络

中图分类号：TP393.06 **文献标志码**：A **文章编号**：1001-3695(2007)04-0318-03

Wavelet and Analysis of Network Traffic Anomalies

LIU Lan[1,2], LI Zhi-tang[1], LI Jia-chun[3], TAN Xiao-ling[4]

(1. *School of Computer Architecture, Huazhong University of Science & Technology, Wuhan Hubei* 430074, *China*; 2. *Dept. of Electron, Guangdong Polytechnic Normal University, Guangzhou Guangdong* 510655, *China*; 3. *School of Computer Science, South China University of Technology, Guangzhou Guangdong* 510641, *China*; 4. *Dept. of Electron, Chongqing Three Gorges University, Chongqing* 404000, *China*)

Abstract: Because of wavelet can process data at different scales which is similar with the real network flow presents fractal / self-similar nature in its scaling behavior. A wavelet based distributed ID model approach to analyses network traffic was developed in detail. The model can post attack characteristic more clearly and, by way of improve the veracity, corresponding network node signals of wavelet decomposition was compared. Attacks and the model can detect attack more precisely than the previous model does.

Key words: anomaly; wavelet; nerve network

0 引言

尽管 Internet 的设计一直在不断地完善，但人们对网络行为许多方面的理解却较少。Internet 技术和管理的多样性、网络规模持续增长性，及其应用和使用方式的变化特性，均对网络行为研究提出了挑战，从而使 Internet 行为学研究从网络管理中分离出来，成为一门独立的网络研究科学。

网络的流量特性是网络行为学分析中所必须考虑的一个重要因素，其在网络协议设计、性能优化和网络设备研究等方面起了至关重要的作用。而当今的计算机网络中，存在着诸多异常因素和计算机黑客变化多端的攻击手段，使得网络的流量特性变得越来越复杂。如何通过对大量数据的统计分析，发现网络系统中的异常流量是一个值得研究的问题。

近年来许多关于网络流量特性的研究结果表明，在真实环境中的网络流量呈现出相当明显的尺度特性[1,2]。网络流量在小时间尺度上呈现出复杂奇异性特征，在大时间尺度上具有长程依赖性(Long Range Dependence, LRD)特征。而小波分析能将时间域上重叠但频率域上不重叠的流量信号进行分段，不同的频段(也就是不同的小波系数层次)代表了信号中处于该尺度的信号分量，网络流量具有时域中频域重叠的特征。

1 小波分析原理

小波分析是信号处理和流量分析的强大工具。小波分析分解过程把信号分解为不同的层，而每一层均以时间作为独立的变量，这样较之傅里叶变换就有了时域定位的能力。

小波函数 $\psi(t)$ 指具有震荡特性，能迅速衰减到 0 的一类函数($\int_{-\infty}^{+\infty}\psi(t)\,dt=0$)，它是一个带通滤波器。通过 $\psi(t)$ 的伸缩和平移后派生出一族函数 $\psi_{a,b}(t)$：

$$\psi_{a,b}(t)=|a|^{-1/2}\psi(t-b/a)\qquad b\in R,\ a\in R,\ a\neq 0 \tag{1}$$

式中：$\psi_{a,b}(t)$ 称连续小波，a 为尺度因子或频率因子，b 为时间因子。

若 $\psi_{a,b}(t)$ 满足式(1)，对于能量有限信号或时间序列 $f(t)\in L_2(R)$，其连续小波变换定义为

$$W_\psi f(a,b)=\langle f,\psi_{a,b}\rangle=|a|^{-1/2}\int f(t)\,\bar{\psi}(t-b/a)\,dt \tag{2}$$

式中：$\bar{\psi}(t)$ 为 $\psi(t)$ 的复共轭函数。式(2)说明小波变换是对 $f(t)$ 按不同尺度进行分解，实质是对 $f(t)$ 用不同滤波器进行滤波，滤波器的脉冲响应为 $|a|^{-1/2}\psi(t-b/a)$。

收稿日期：2006-02-15；修返日期：2006-04-22 基金项目：国家网络与信息安全保障持续发展计划资助项目(2004 1-917-021)；国家“973”计划资助项目(2003CB314805)

作者简介：刘兰(1977-)，女，湖南人，讲师，博士研究生，研究方向为计算机网络安全(liulan@scut.edu.cn)；李之棠(1951-)，男，湖北人，教授，博导，博士，研究方向为网络与信息安全、光互连与光计算；李家春(1968-)，女，湖北人，副教授，博士，研究方向为计算机网络与信息安全；谭晓玲(1969-)，女，重庆人，讲师，学士，研究方向为网络管理和网络行为分析.

小波既具有频率分析的性质, 又能表示发生的时间, 有利于分析确定时间发生的现象, 并具有多分辨率(multi-resolution) , 即多尺度的特点, 可以由粗及精地逐步观察信号, 有利于各分辨度不同特征的提取, 这与网络攻击在不同频率下的不同表现有着相似特性。在实际应用中, 适当地选择基本小波, 使 $W(t)$ 在时域、频域上均比较集中, 便可以使小波技术在时域, 频域上都具有表征信号局部特征的能力, 因此有利于检测信号的瞬态和奇异点。

2　网络异常分析结构模型

2.1　网络异常行为分析及频率特性

网络流量特性是网络性能分析和通信网络规划设计的基础, 准确地描述流量特性对设计高性能网络协议、高效网络拓扑结构、业务量预测与网络规划、高性能价格比的网络设备与服务器、精确的网络性能分析与预测、拥塞管理与流量均衡均有重要意义。因而网络流量特性的研究长期以来受到计算机网络研究人员的高度重视。

而在纷繁复杂的网络行为中, 错误、攻击等流量异常行为在网络中非常流行。在一个短时间内确定、诊断以及应对异常是网络行为学研究的重要内容。准确地确定和诊断异常首先依靠健壮和及时的数据, 其次依靠一种确定的方法将异常信号从数据中分离出来。

为了分析网络异常的流量行为, 需要先分析正常流量的环境特征, 作为发现异常的一个基准。现在有很多研究描述了正常网络环境的一些重要特征, 比如以天和周为单位的网络流量自相似性[3]。笔者在分析实验中发现, 网络流量行为具有较好的时域中频域重叠的特征, 即在某个点测得的流量由不同的网络流量汇集而成, 本文利用小波分析的手段, 将不同频域的流量分解后再作具体的异常分析, 具有较好的效果。

2.2　网络异常分析结构设计

在系统中, 考虑到小波在将复杂的问题化简方面的优势, 提出网络异常的一种小波分析的方法。该网络异常分析的系统结构如图 1 所示。

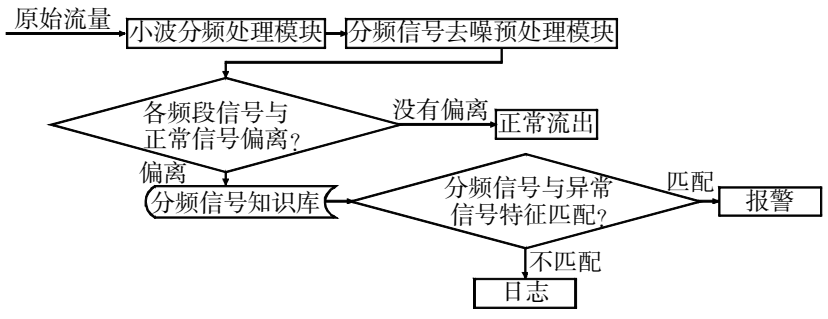

图 1　异常分配系统结构图

本文利用小波分析分解过程将信号分解为不同的层, 而每一层均以时间作为独立的变量, 这样较之傅里叶变换就有了时域定位的能力。较低层的滤波信息较少, 因此可以认为是多种频率的复杂聚合体, 将其用信号的低频表示。而与之相反, 信号的高频部分则显示了信号的纹理细节部分, 体现了信号的细微变化。在处理原始信号时, 频率是通过式(1) 中的 a 尺度因子来调节的。

通过调节 a 尺度因子的范围, 将从原始信号分解之后的导出信号分为以下三部分:

(1) L(低频) 信号

该部分所占有的数据量占原始信号的很少部分, 大概是 0.4%。它可以反映流量以天或星期为单位的较长时间特性。

(2) M(中频) 信号

这部分的信号在 0 周围波动。其数据量大概占原始信号的 3% 左右。

(3) H(高频) 信号

这部分信号需要进行除噪处理, 因为这些变化很快, 很“短”的信号通常是一些对分析网络异常没有帮助的噪声。可以通过将所有超过某个阈值的设置为 0, 从而消除噪声信号。

将原始信号作了初步的分频处理后, 接下来的工作是在分频信号中进行异常流量的分析。对于流量异常诊断重要的一个前期步骤应该是先分析正常流量的特征, 作为发现异常的一个基准, 这需要建立信号知识库。有很多研究已经表明, 在正常情况下, 网络流量具有长程相关和自相似的特性[1,2,5]。该试验研究也表明, 通过小波分频处理后的信号依然很好地保持了这样的特性。

异常分频知识库的建立可以通过小波分解异常信号, 得到高/中/低频频带分析来建立异常信号特征, 将已知的攻击特征进行编码, 存入知识库中。同时也对于相对稳定的正常工作的周围环境信号进行分析, 建立一个对应“正常活动”的系统或用户的正常轮廓, 对高/中/低频频带信号设立阈值。这样在知识库中就记录了分频的异常流量和正常流量特征的知识库。对于系统的即时流量经过前期处理后, 即可对照知识库中的规则进行匹配处理, 如果匹配即启动相应措施, 进行报警。

2.3　网络异常分析系统实现

2.3.1　实验研究

在实验研究中, 笔者应用了不同的小波系统来确定如何最好地揭示已记录的异常。通过限制流量和采集到的 SNMP 数据, 从而做到在较好的时间尺度上的分析。通过实验比较, 选择一个较好的小波系统来有效地揭示正常信号和异常信号的潜在特征。

在实验分析中, 将网络流量数据分为不同的组。第一组包含了一些时间跨度较长的攻击事件。这些数据采集时间间隔较长, 持续的时间也较长。另外的一些实验数据为一些短程的包括网络错误、攻击等异常行为的流量数据。这些短程异常行为在一般的监测方式中较难被发现。这是因为它们与正常的突发网络行为极为相似。通过小波分析的分频方法, 将一些无关的信息过滤后, 中、高频带的信号将能较好地揭示出这些短程异常行为。

在实践过程中, 发现在不同的网络节点上采集的数据对于异常的表征能力是不一样的。当采集点距离异常源较近时, 异常在小波分频后的分析图中表征较为明显; 当距离异常源较远时, 异常将由于一些其他流量的混合干扰而变得模糊, 使得分离异常的任务变得相对困难。结合神经网络的知识, 对这些流量进行一些预处理, 使得系统不管是在离异常点近或者远的测量点均有较好的效果。

2.3.2　数据采集

对于数据流量的采集笔者采用 SNMP 协议采集 MIB 数据[4], SNMP 是由 IETF 提出的, 随着 TCP/IP 成为事实上的协

议标准而广泛被使用。SNMP 主要由三部分组成,即管理者、代理和 MIB。通过 SNMP 协议的标准命令 GetRequest, GetNextRequest 通过代理获得所需的相关 MIB 数据值。基本原理如图 2 所示。

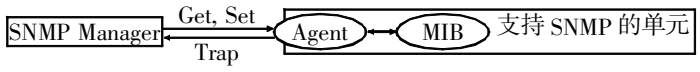

图 2 SNMP 原理图

在该实验中,主要用到网卡的通信流量。其具体数据在 SNMP 中被定义为接口 if 组:

ifInOctets 为接口发送的字节数

ifOutOctets 为接口接收的字节数

ifInUcastPkts 为输入的单播包数

ifOutUcastPkts 为输出的单播包数

ifInNUcastPkts 为输入的非单播包数

ifOutNUcastPkts 为输出的非单播包数

通过 if 组的数据,可以计算出网络接口处发送和接收数据包的传输速率 V:

$$V(\text{发送}) = (\triangle \text{ifOutUcastPkts} + \triangle \text{ifOutNUcastPkts}) / \triangle t \quad (1)$$

$$V(\text{接收}) = (\triangle \text{ifInUcastPkts} + \triangle \text{ifInUcastPkts}) / \triangle t \quad (2)$$

笔者从华南理工大学网络中心的网络监测系统获得 3 ~6 月份期间的几个校内边界路由器的流量数据,数据是对各个网络设备进行 MIB 数据采样获得的。

2.3.3 数据转换

从路由器等网络设备采集到的数据为 RRD(Round Robin Database) 格式。RRD 可以用来储存时间序列的数据,如网络流量,机房温度等。它以一种压缩程度非常高的方式来存储数据。由于试验的需要,对初始数据在系统处理过程中进行适当的转换,使原始数据方便进行算法的处理,得到系统核心处理器所需的数据格式。

2.3.4 实验结果及分析

对于异常有长程异常和短程异常。笔者认为其可以由小波分频后的不同频带进行表征,对于长程的网络异常现象,应该体现在系统处理后流量信号的中低频信号中。

图 3 是 3 ~5 月的校园网络流量及该系统处理后的分解图示。其中这次长程异常在低频部分中能较好地反映出来。从图 3 中可以看到,在 4 月中旬时,低频流量部分的数据与知识库中的长程异常流量特征相匹配,系统及时启动预警机制。

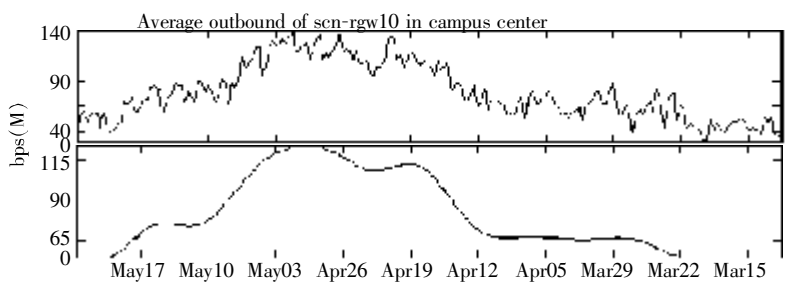

图 3 长程异常分析结果

而现在流行的网络攻击和蠕虫等造成的网络异常一般的特性为短时、突发,需要用细尺度的小波函数来刻画,这时需要分析信号的高频和中频部分。因为高频和中频部分信号区间相对较短,频率的分辨率升高,此时才能有足够的分辨率去发现这些短程攻击。图 4 为 5 月份时,在网络中心对某服务器模拟的一次 DoS 攻击,正如笔者所预计的在有较高频率定位能力的中、高频段系统均可以较好地发现异常网络行为。在原信号中,此次异常行为流量与一些其他的网络行为流量混杂在一起,不容易处理,通过该系统的预处理和分析之后,其表现特征在高频部分得到很好的体现,模型系统及时发现了发生在 5 月 12 日的这次模拟攻击。

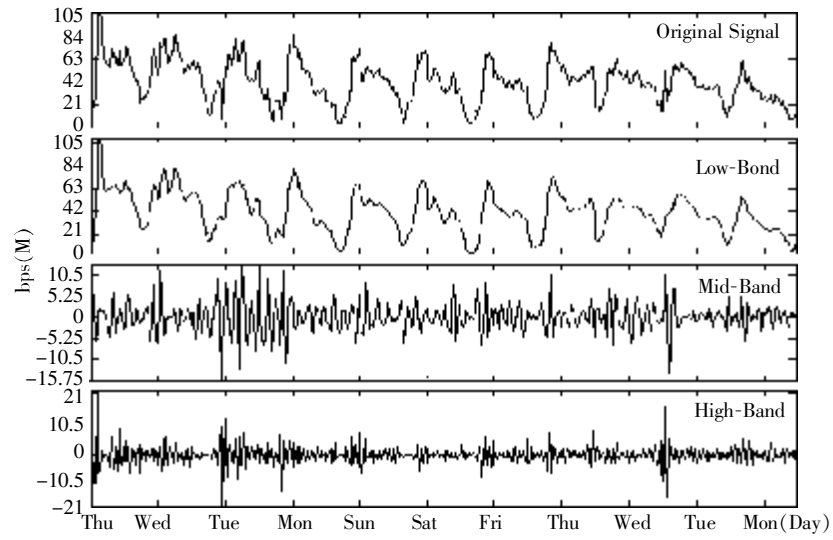

图 4 模拟短程异常分析结果

通过实践证明,该模型系统能初步定位一些异常行为,在丰富了知识库后能较为精准的实时捕获异常并报警。

3 结束语

由于网络流量具有时域中频域重叠的特征,本文利用小波技术在检测网络异常流量方面的优势,设计并初步实现了一种异常行为分析系统。对于网络行为中的攻击行为进行了分析和判别,并在实际的应用中进行了模拟分析,得到较好的实验结果,但系统仍然存在一些需要改进的地方。其中精确描述刻画网络流量特性以及对网络异常知识库的充实是重点需要研究的。

参考文献:

[1] IAN X, WU J, JI C. An unified framework for understanding network traffic using independent wavelet models[C]. [S. l.]: IEEE INFOCOM, 2002: 446-454.

[2] CROVELLA M, KOLACZYK E. Graph wavelets for spatial traffic analysis[J]. **IEEE Computer**, 2003: 1848-1857.

[3] BARFORD P, KLINE J, PLONKA D, *et al*. A signal analysis of network traffic anomalies. ACM SIGCOMM Internet Measurement Workshop[C]. [S. l.]: [s. n.], 2002.

[4] 李之棠,刘兰. 基于 SNMPv3 协议的 VPN 集中管理系统的设计与实现[J]. 小型微型计算机系统, 2001, **22**(9): 249-252.

[5] 陈惠民,蔡弘,李衍达. 突发业务的多重分形建模及其参数估计[J]. 电子学报, 1999, **27**(4): 19-23.

(上接第 317 页)

[9] MAGONI D, PANSIOT J J. Evaluation of Internet topology generators by power-law and distance indicators: proc. of 10th IEEE International Conference on Networks[C]. Singapore: [s. n.], 2002: 401-406.

[10] JARED W, SUGIH J. Inet-3.0: Internet topology generator CSE-TR-456-02, Ann Arbor[R]. Michigan: University of Michigan, 2002.

[11] JIANG Y, FANG B X, HU M Z, *et al*. An example of analyzing the characteristics of a large scale ISP topology measured from multiple vantage points[J]. **Journal of Software**, 2005, **16**(5): 846-856.

[12] 姜誉,方滨兴,胡铭曾,等. 大型 ISP 网络拓扑多点测量及其特征分析实例[J]. 软件学报, 2005, **16**(5): 846-856.