基于小波技术的网络异常流量检测与实现*)

肖志新[1] 杨岳湘[2] 杨 霖[1]

(吉首大学网络中心 湖南吉首 416000)[1] (国防科技大学计算机学院 长沙 410073)[2]

摘 要 随着互联网的迅速发展和各种互联网应用的普及,对网络资源的需求和使用也迅速增加,网络中也包含大量的蠕虫、端口扫描、DDoS、网络滥用等异常流量,因此有效、快速检测这些异常流量的手段是必须的。本文通过采集路由端口上的NetFlow流,建立了一个利用小波方法分析网络异常流量的框架,通过此方法可以较为理想地检测出网络中的异常流量,在实际应用中取得了较为理想的效果。最后用小波方差分析了不同的小波函数在检测网络异常流量时的效能。.

关键词 网络异常,NetFlow,小波

Anomaly Detection and Implementation Based on Wavelet

XIAO Zhi-Xin[1] YANG Yue-Xiang[2] YANG Lin[1]

(Center of Network, Jishou University, Jishou 416000)[1] (College of Computer, National University of Defense Technology, Changsha 410073)[2]

Abstract A tremendous demand for an ever increasing reliance on network resources has led to a rapid rise in those resources. The network are faced with more and more threaten from Worm, port scan, DDos and network abuse, so the effective and rapid measure is required. This paper setup a framework to analyze network anomalies based on wavelet by using NetFlow. And we have evaluated various wavelets for their effectiveness at detecting different kinds of anomalies. The system has gained ideal effect.

Keywords Network Anomaly, NetFlow, Wavelet

1 前言

随着互联网的迅速发展和各种互联网应用的普及,网络已成为人们日常工作生活中不可或缺的信息承载工具。然而,伴随着网络的正常使用流量,网络上形形色色的异常流量也随之而来,网络安全已变得越来越重要,网络面临来自蠕虫、端口扫描、DDoS、网络滥用的越来越多来威胁。因此,对于一个大型校园网络的管理者来说,有效、快速而准确地检测出异常流量是非常必要的。目前,有各种不同的异常流量检测方法,本文通过采集某校园网路由器上的NetFlow流,利用小波滤波器将信号分解到不同的频带上,在高频段上反映信号的瞬时变化,而在低频段上反映信号的缓慢变化。在包含异常流量的网络中,异常的起始点往往伴随着信号的突变,而小波变换能方便地检测出信号的突变点,因此可以将小波变换应用于网络异常流量的检测。

2 背景知识

2.1 NetFlow的相关知识

NetFlow是Cisco发展的流量统计协议,其工作原理是:NetFlow利用标准的交换模式处理数据流的第一个IP包数据,生成Netflow缓存,随后同样的数据基于缓存信息在同一个数据流中进行传输,不再匹配相关的访问控制等策略,Netflow缓存同时包含了随后数据流的统计信息。也就是说一个"流"(Flow)是特定来源和目的单向流量信息,即源IP、目的IP、源Port、目的Port四个属性相同的数据包为一个Flow。这些统计被有效地使用在网络管理、计费及网络规划等,目前广受各家厂商支持。

2.2 小波分析方法

小波是近十几年才发展起来并迅速应用到信号处理和图像处理等众多领域的一种数学工具,小波属于时-频分析的一种方法,小波分析比Fourier分析有许多本质性的进步。它具有多分辨率分析的特点,而且在时-频两域都具有表征信号局部特征的能力,是一种窗口大小固定不变但其形状可改变、时间窗和频率窗都可以改变的时频局部化分析方法。即在低频部分具有较高的频率分辨率和较低的时间分辨率,在高频部分具有较高的时间分辨率和较低的频率分辨率,很适合于探测正常信号中夹带的瞬态反常现象并展示其成分,所以被誉为分析信号的显微镜。

小波变换能有效地从信号中提取信息,通过伸缩和平移等运算功能对函数或信号进行多尺度分析(Multi-resolution Analysis),可以解决Fourier变换不能解决的许多困难问题。小波变换可得到低频系数(或叫作近似系数)和高频系数(或叫作细节系数),其中低频系数反映原始信号的轮廓,而高频系数则是反映信号的细节。根据信号理论知道,信号中的奇异性往往是由于频率突变造成,也就是说,这种奇异性一般是通过频率的异常变化而反映出来,那么只要把信号中的频率变化情况提取出来,就可以发现信号中奇异性,并确定奇异性发生的时间。

3 具体方法

3.1 用小波分解检验流量异常

在本文的系统中,用开源软件flow-tools采集校园网三

*)基金项目:国防预研基金项目(514040601);湖南省教育厅科学研究项目(03C344)。肖志新 副教授,主要研究方向:计算机网络安全。杨岳湘 教授,主要研究方向:网络管理与安全。

层交换机某接口的流量，并计算出采样间隔相应的流数、包数和字节数，然后对采集到的信号（原始流量）进行小波分解。采样的间隔与我们分析的异常流量相关，即间隔长度应依赖与我们所期望研究的异常流量的持续时间。如果定义异常流量的持续时间为 t_0，采样间隔的长度为 t_1，在理想情况下期望 $q=t_0/t_1\approx1$，如果 q 值太小，异常流量可能会被淹没而丢失；如果值 q 太大，将检测到太多的我们并不关心的“异常”。

本文中关注持续时间为数秒到数分钟的异常流量，因此将采样确定为 1 分钟，并且这个长度便于我们进行分析和部署。随后建立一个横轴为时间，竖轴为流（包、字节）数的原始时间序列信号，并将它交由 matlab 的 wavelet 工具，采用 Db4 小波对流量进行分析，根据分析流量的长度，对流量进行 7 层小波分解。将多尺度分析得到的小波分解系数分为三个部分：低频、中频和高频部分。低频部分即尺度分量，它反映了流量每天周期性变化的规律；中频部分包含了 5、6、7 层的细节分量，有着大量对流量异常检测的“有用”信息；高频部分包含了前 4 层的细节分量，这部分主要是噪音，对异常流量的检测没有太多的作用。

实验结果如图 1 所示。图中，展示了三天的流量采集，以 Db4 小波进行分解，并以 byte 为参数对异常流量的检验结果。可以看出，在用红色矩形框标注（即我们实验注入异常流量）的时间点，非常明显检测到了流量异常。另外，由于学生宿舍每晚 10:30－11:00 左右停电，造成流量较为迅速的下降，在图中也可以较为明显的发现。可以看出小波分析对异常流量的检测有着较为理想的效果。

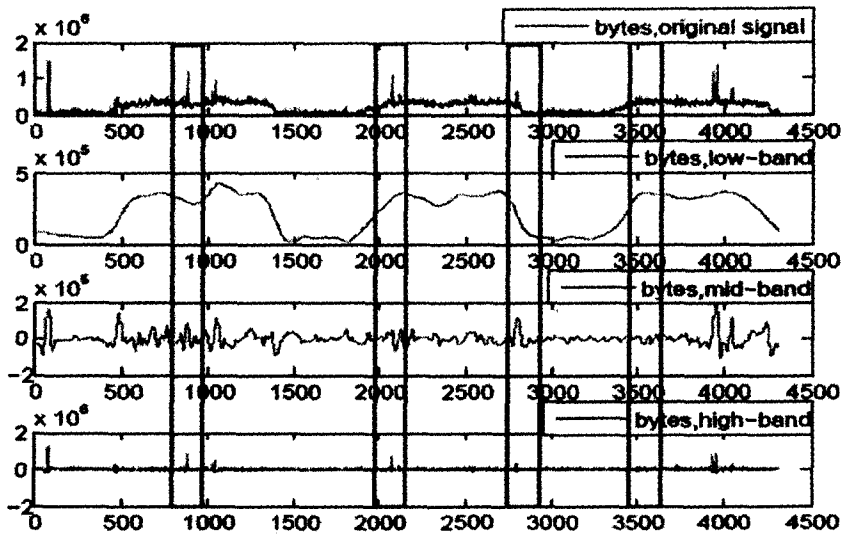

图 1　以 byte（字节）数为参数的小波分解

3.2　不同小波函数的效能

选取不同的小波函数会对分析的数据结果有较大的影响，会导致不同的分析结果。在小波函数中，通常由以下几个特性来决定不同的小波函数类型：(1)小波函数，尺度函数 ϕ 的支撑长度，表示当时间或频率趋向于无穷大时，支撑长度收敛的速度；(2)对称性；(3)小波函数和尺度函数 ϕ 消失阶矩数；(4)正则性。

在这里，对流量只进行三层分解，用小波分析的第三层细节系数用小波方差分析进行度量。方差与标准差是测度离中趋势的最重要最常用的量。总体方差是一组资料中各数值与其算术平均数离差平方和的平均数。通常用 σ^2 表示。总体标准差则是总体方差的平方根，用 σ 表示。方差在统计推断上具有较佳的统计与数学性质，使得方差成为最重要的离中趋势测度量。样本方差是一组资料中各数值与其算术平均数离差平方和的平均数，通常用 S^2 表示。样本标准差则是样本方差的平方根，用 S 表示。

设被测量统计量为 X，该统计量前 n 个单位时间内平均抽样为 $x_1,x_2,\cdots,x_n$。则该统计量的样本均值为 $\bar{x}=\sum_{i=1}^{n}x_i$。标准方差是测量数据的偏差，如果数据离平均值近，则置信区间较窄，对于 n 个数据值样本标准差对总体标准差的无偏估计定义为：

$$S_n=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}=\sqrt{\frac{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}{n-1}}$$

$$=\sqrt{\frac{\sum_{i=1}^{n}x_i^2-\bar{x}\sum_{i=1}^{n}x_i}{n-1}}$$

对于每个测度，系统只要维护 3 个值：样本均值、样本累加和、样本平方累加和。样本均值和标准差能为流量特性的总体均值构造一个置信区间。样本均值的标准偏差 $S_{\bar{x}n}=\frac{S_n}{\sqrt{n}}$。如果从相同的流量中重复选择样本，并计算每个样本的均值，则这个统计量表明期望的变化量。由中心极限定理：对于大于 n 的样本，其均值服从均值等于流量总体的均值，标准差为 S_{σ_n} 的正态分布。因此，可以构造总体均值的置信区间 μ 为

$$|\mu-\bar{x}|\leqslant z_a\frac{S_n}{\sqrt{n}}$$

z_a 是标准正态分布给定 a 的分位数。如果当前样本测度值满足上式要求，说明当前流量正常，若不满足上式要求，则说明当前流量异常。

a 的取值可以根据实际情况来确定，只需计算 $|x_i-\bar{x}|$ 与标准差的比值，以减少计算量。通过计算偏离均值的程度来判断是否发生了异常，例如，对于正态分布，95.45% 的数据包含在均值附近的 2 倍标准差内，可以通过这种简化的了方法来判断是否发生了异常。

具体方法是：首先计算出一个窗口内所有样本值的标准差 S 和均值 $\bar{x}$，然后计算每一个样本点 x_i 与均值的偏离 $|x_i-\bar{x}|$；最后计算：

$P=|x_i-\bar{x}|/S$

在流量正常时，P 值较小，一般小于1；如果 P 的值超过2（这仅仅为经验值，根据实际情况的不同，P 值在2附近浮动），认为有异常流量发生或结束。

那么一个理想的小波函数在检测流量时，应该是：在异常流量发生或结束时应该有较大的偏离，而在其他情况下，偏离较小。

同时我们以100个数据为一个总体样本，即以100个采样点（100分钟）为一个分析检测窗口，为方便观察，统一将分析检测窗口的第50时间点作为异常流量的发生时间。

由于总体样本较小（100时间点），故只需进行三层分解，其中第1、2层看作高频部分，第3层看作中频部分，是我们最感兴趣的部分，也是包含奇异性信息最丰富的部分。我们以byte为参数，对小波分析的第三层细节系数用小波方差分析进行度量。采用四种小波函数：Db3、Db6、Coif3、Dmey，分别

表1 不同小波函数的效能

	均值 $\bar{x}$	异常流量发生点第三层系数 x	方差 S	$P=\|x_i-\bar{x}/S$
Db3	−366.3	7.11×10^4	3.67×10^4	1.95
Db6	−313.8	1.25×10^5	4.68×10^4	2、68
Coif3	362.0	1.95×10^5	5.15×10^4	3.78
Dmey	340.4	1.76×10^5	5.40×10^4	3.25

用这四种小波函数对SYN flooding异常流量进行检测分析，实验结果如表1，可以发现，Coif3具有最好的检测效果。我们还对其他异常流量进行了分析，发现coif3均具有较好的检测效果，限于篇幅，这里不作详细介绍。

结论与进一步的工作 本文提出了一种基于小波技术的网络异常流量检测方法，通过采集校园网流量数据，计算采样点的偏离值，用偏离值的大小来评估是否发生了流量异常，在实际应用中有着较为理想的效果。同时，进一步的研究也是必须的，重点将放在以下几个方面：不同的采样间隔对异常流量检测的影响；小波分解只对信号的低频段进行分解，而对高频段没有进一步的分解，那么使用小波包分解是否有着更为理想的效果；异常流量的分类研究，这将有助于我们对本文方法进行进一步的总结和提升；在开源平台上建立起一个自动检测异常流量的系统等。

参 考 文 献

1 Thottan M, Ji C. Proactive Anomaly Detection Using Distributed Intelligent Agents. IEEE Network, 1998, 12(5): 21～27

2 Alarcon-Aquio V, Barria J A. Anomaly detection in communication networks using wavelets. IEE Proceeding-Communication, 2001, 148(6): 355～362

3 Barford P, Kline J, Plonka D, et al. A signal analysis of network traffic anomalies. In: Proceedings of ACM SIGCOMM Internet Measurement Workshop, 2002

4 Abry P, Veitch D. Wavelet analysis of long range dependent traffic. IEEE Transactions on Information Theory, 1998, 44(1)

5 http://www.cisco.com/en/US/products/ps6601/prod_white_papers_list.html

6 http://www.splintered.net/sw/flow-tools/docs/

（上接第115页）

```
9 compareProcessesGetFromDifferentSource();
10 getHiddenProcessesList();
1  getProcessesFromElement (element e)
2     get the thread t which element e belongs to;
3     get the process p which thread t belongs to;
4     insert the process p in foundProcessList;
```

4 实现与结果

针对本文提出的基于线程调度的进程隐藏检测技术，可设计能检测出目前所有特洛伊木马的隐藏进程的工具软件。

OneTNS是本设计思想在Window平台下的进程检测工具，它针对当前计算机木马程序自身隐藏的特性，通过读取操作系统内核数据结构，来获取可靠的系统状态信息，实现可信任的主机入侵检测与防御。OneTNS目前可以检测已公开流行的具有进程隐藏功能的系统恶意程序包括：

Vanquish，用户态进程隐藏的恶意系统程序。

Hacker Defender，内核态进程隐藏的恶意系统程序。

Fu，基于直接修改内核对象的进程隐藏的恶意系统程序。

Dasher，2005年12月出现的基于多个Windows漏洞的蠕虫，包含了基于内核态系统调用挂钩的进程隐藏功能。

当以上系统恶意程序被安装于Windows操作系统中之后，通过现有的任务管理器等其他的进程检测监视工具不能发现它们的存在，而通过我们的检测软件将会完全探测出它们相关的进程信息，并成功实现对该进程的分析、结束、停止等控制操作。

结束语 基于线程调度的进程隐藏检测技术是针对现有的恶意程序中进程隐藏行之有效的检测方法，但在实现的过程中，同样会面对入侵者的攻击，使得我们通过本方法检测的结果数据被修改。因此，如何提高检测的完整性与可靠性，如何保护安全检测软件自身的行为等，是下一步重点研究的内容。另外，针对特洛伊木马程序的其他隐藏功能，利用操作系统的内核运行机理，也可以提出更进一步的检测方式思想。

参 考 文 献

1 Thimbleby H, Anderson S, Cairns P. A Framework for Modeling Trojans and Computer Virus Infections. The Computer Journal, 1998, 41(7): 444～458

2 Buteler J R II. Detecting Compromises of Core Subsystems and Kernel Functions in Windows NT/2000/XP: M.S. Thesis, University of Maryland, Baltimore County, 2002

3 Butler J, Jeffrey L. Undercoffer and John Pinkston. Hidden Processes: The Implication for Intrusion Detection. In: Proceedings of the 2003 IEEE Workshop on Information Assurance United States Military Academy, West Point, NY, June 2003

4 Levine J G, Grizzard J B, Hutto P W, Owen H L. A Methodology to Characterize Kernel Level Rootkit Exploits that Overwrite the System Call Table. In: Proceedings of IEEE. SoutheastCon, IEEE, 2004. 25～31

5 Levine J, Grizzard J, Owen H. A Methodology to Detect and Characterize Kernel Level Rootkit Exploits Involving Redirection of the System Call Table. In: Second IEEE International Information Assurance Workshop, 2004