全局的多流量相关异常检测算法

杨 丹，胡光岷，李宗林，姚兴苗

(电子科技大学宽带光纤传输与通信网技术教育部重点实验室 成都 610054)

【摘要】针对现有单链路流量异常检测和全局流量异常检测方法存在的不足，该文提出一种全局的多流量相关异常检测算法。该算法利用同一异常在不同链路或OD流所产生的多个异常流量信号在频率、幅值变化特征等方面具有相似性这一特点，将这种相似性作为检测的依据来检测异常。通过每个OD流或链路的前期流量数据进行下一时刻的流量预测，将实际流量数据减去预测流量值得到异常流量值；通过多个OD流或链路之间的全局相关分析进行流量异常检测。仿真结果表明该文提出的方法能够有效地检测其他单链路和全局异常检测方法无法检测的异常。

关 键 词 ARMA模型； 相关性分析； 全局流量异常检测； 流量预测

中图分类号 TN914.5 **文献标识码** A

The Network-Wide Multi-Traffic Correlative Anomaly Detection

YANG Dan, HU Guang-min, LI Zong-lin, and YAO Xing-miao

(Key Laboratory of Ministry of Education for Broadband Optical Fiber Transmission and Communication Networks, University of Electronics Science and Technology of China Chengdu 610054)

Abstract Aiming at the lack of the single link's anomaly detection and the network-wide traffic's anomaly detection, we propose a network-wide multi-traffic correlative anomaly detection method. This method uses the characteristic that the anomaly signals on different links or origin-destination (OD) flows, produced by one anomaly, are similar in frequency, the transformation characteristic of the amplitude, and so on. And the comparability is used as the evidence of the anomaly detection. In principle, the traffic is forecasted by the previous data on every OD flow or link, the anomaly traffic is obtained by subtracting the real traffic from the forecast data, and lastly, the traffic anomaly is detected by global correlation analysis on all traffics. Simulation result indicates that this kind of method can detect the anomaly.

Key words ARMA model; correlation analysis; network-wide traffic anomaly detect; traffic forecast

网络流量异常指网络的流量行为偏离其正常行为的情形，引起网络流量异常的原因是多种多样的，如网络设备的不良运行、网络操作异常、突发访问(flash crowd)、网络入侵等。异常流量的特点是发作突然，先兆特征未知，可以在短时间内给网络或网上的计算机带来极大的危害(如由特定的攻击程序或蠕虫爆发所引起的突发流量行为)，因此准确、快速地检测网络流量的异常行为，判断引起流量异常的原因，做出合理的响应是保证网络有效运行的前提之一，也成为目前国内外学术界和工业界共同关注的前沿科学问题之一。

现有的网络流量异常检测方法大都以单点或某个局部网络为核心实施对网络流量的监控，通过一维信号的异常变化检测来检测异常[1-3]，但某些网络异常行为(如DDoS)在单条链路上并不一定表现出非常明显的流量异常，检测常常出现误判或漏判。如果将多条链路或多个OD流的流量信号作为一个整体进行研究(进行多时间序列分析)时，异常就有可能显现出来。全局异常检测还能在异常尚未到达被攻击节点时进行检测，为异常情况的应急处理赢得时间。文献[4-5]提出了一种基于PCA(principal component analysis)分解的全局异常检测方法，该方法将网络OD流矩阵或链路流矩阵作为研究对象，通过PCA分解方法，将所有流量分为正常流量部分(OD流或链路间相似性强的部分)和异常部分(相似性弱的部分)，再通过判断异常部分的链路或OD流能量大小来检测异常[5]，可以在一定程度上克服传统局部检测方法存在的缺陷。但该方法有两个方面的不足：一是异常流量(如DDoS)在不同的OD流或链路间具有很强的相似性，采用上述方法可能将其划

收稿日期：2007－04－17；修回日期：2007－09－03
基金项目：国家自然科学基金(60572092)；四川省青年科技基金(04ZQ026-028)
作者简介：杨 丹(1982－)，男，硕士生，主要从事网络流量异常检测与识别方面的研究.

分到正常空间，不能有效检测；二是网络流量是随时间变化的，且变化量较大，因而异常空间中可能存在许多能量较大的值(并非真正的异常)，而某些异常(如DDoS)的分布式特征，决定了它们在单条链路或OD流的异常能量较小，单纯依靠链路或OD流的异常能量大小判断检测可能会出现漏检或误检。

若同一异常引起多条OD流或链路的流量异常变化，且各OD流或链路的异常流量较小，传统的单链路或单节点流量异常检测和文献[5]提出的全局检测方法无能为力。为此本文提出一种全局的多流量相关异常检测算法，首先通过每个OD流或链路的前期流量数据进行下一时刻的流量预测，将实际流量数据减去预测流量值得到异常流量值，最后通过多个OD流或链路之间的全局相关分析进行流量异常检测。仿真结果表明本文提出的方法能够有效地检测上述异常，且性能优于文献[5]提出的方法。

2 检测方法

网络上的同一异常可能会引起多条OD流或链路的流量异常变化，同一异常产生的异常流量信号在频率、幅值变化特征等方面应具有相似性，这种相似性可以作为检测的依据。但由于网络背景流量相对于异常流量通常比较大(且不同的流之间可能具有较强的相似性)，如果直接计算多条OD流或链路网络流量(包括背景流量和异常流量)的相似特征，那么所获得的往往是背景流量的相似特征，异常流量的相似特征会淹没在巨大的背景流量中。为此必须将网络流量中的背景流量和异常流量区别开来，然后进行检测。

文献[4-5]采用PCA方法将网络流量划分为正常流量和异常流量，但该方法认为多条OD流或链路之间相关性强的是正常流量，相关性差的是异常流量。这一观点对由同一异常引起的多条OD流或链路的流量异常变化是不完全适应的，因为同一异常在不同的链路或OD流上产生的异常流量信号应该具有比正常流量更强的相似性。为此本文采用单个OD流或链路的前期流量数据进行下一时刻的流量预测，将预测得到的数据看作正常流量，将实际流量数据减去预测流量值得到的差看作异常流量值，最后通过多个OD流或链路之间的全局相关分析进行流量异常检测，检测步骤如图1所示。

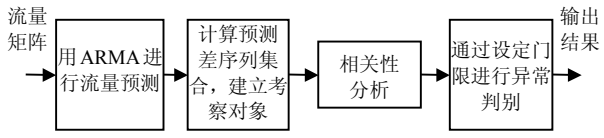


图1 检测步骤

3 流量预测

全局的多流量相关异常检测算法的基础是流量预测，因此，选取一个好的流量预测算法十分重要。本文选择了一种比较成熟的预测算法——AR、MA、ARMA模型预测算法[1-2]。文献[8]首次用到了该预测算法并取得了成功，此后，很多学者都通过该算法对网络流量进行了预测[9-11]，均取得了非常好的效果。这种预测算法的总体思想是：对历史流量序列进行建模，得到该条流量的模型，然后通过这个模型来预测未来的流量。文献[12]给出了该预测算法的具体内容。

4 相关性分析

得到每一条流量的预测值序列之后，将原始流量序列与预测值序列相减，得到每一条流量的预测差序列。这些预测差序列可能由两部分组成：一是预测误差；二是异常。检测方法中已经解释过，异常通常是有一定强度的相关性的(如DDoS攻击)，因此，可以通过相关性分析将异常检测出来。

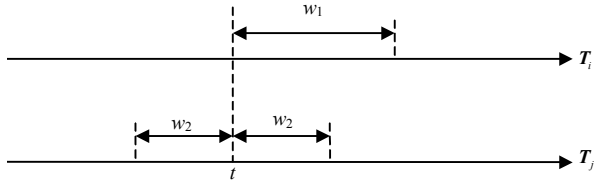


图2 计算相关系数的时窗与滑动时窗

设流量i和流量j的预测差序列分别用向量$\boldsymbol{T}_i$和$\boldsymbol{T}_j$表示；时窗W表示对流量的总采样点数，即向量$\boldsymbol{T}_i$的总长度，w_1表示计算相关系数的向量的大小，时窗w_2表示w_1允许滑动的范围(如图2所示)；$\boldsymbol{T}(t)$表示对向量$\boldsymbol{T}$从时间点t开始截取一段长度为w_1的数据构成的一个新向量。于是，$\boldsymbol{T}_i$和$\boldsymbol{T}_j$在时间点t的相关系数coff(i, j, t)定义为：

$$\mathrm{coff}(i,j,t)=\max_{t_j}\{\mathrm{corrcoef}[T_i(t),T_j(t_j)]\} \qquad (1)$$

式中 t_j为t在滑动时窗范围内滑动所得到的值。容易看出，由于t_j的滑动，可以得到一系列$\boldsymbol{T}_i(t)$与$\boldsymbol{T}_j(t_j)$的相关系数，而coff(i, j, t)实际上就是这一系列相关系数的最大值。之所以要这样做，是由于同一个异常在不同的流量处，可能因为网络延迟的不同，而导致该异常发生的时间有少许偏差。本文先设置一个滑动时窗，计算完所有的相关系数之后，再取其最大值，就能更准确地判断出异常了。

coff(i, j, t)的值反映了在时间点t流量i和流量j的预测差序列之间的相关性。但是，这里应该考察的是整个网络的情况，因此，要把网络中与被检测节

点相关联的所有流量都考虑在内，将其计算得到的所有coff(i, j, t)叠加起来，然后再归一化，得到一个能反映在时间点t整个网络预测差之间的相关性的相关系数coff(t)：

$$\mathrm{coff}(t)=\frac{1}{\mathrm{num}}\sum_i\sum_j \mathrm{coff}(i,j,t) \tag{2}$$

式中　$0\leqslant t\leqslant W-w_1+1$，$i\neq j$。

通过式(2)得到所有时间点的coff(t)后，再设置一个合适的门限，就能够判断出是否存在网络流量异常，以及出现异常的时间点了。由于网络流量从长期来看存在一定的周期性，因此可以通过对一个历史时间段内的相关系数进行分析得到一个合适的门限。通过长时间对网络流量的研究发现，网络流量的相关系数服从正态分布。设在一个历史时间段内网络流量相关系数的均值为e，标准差为δ，门限系数为α，门限d可写为：

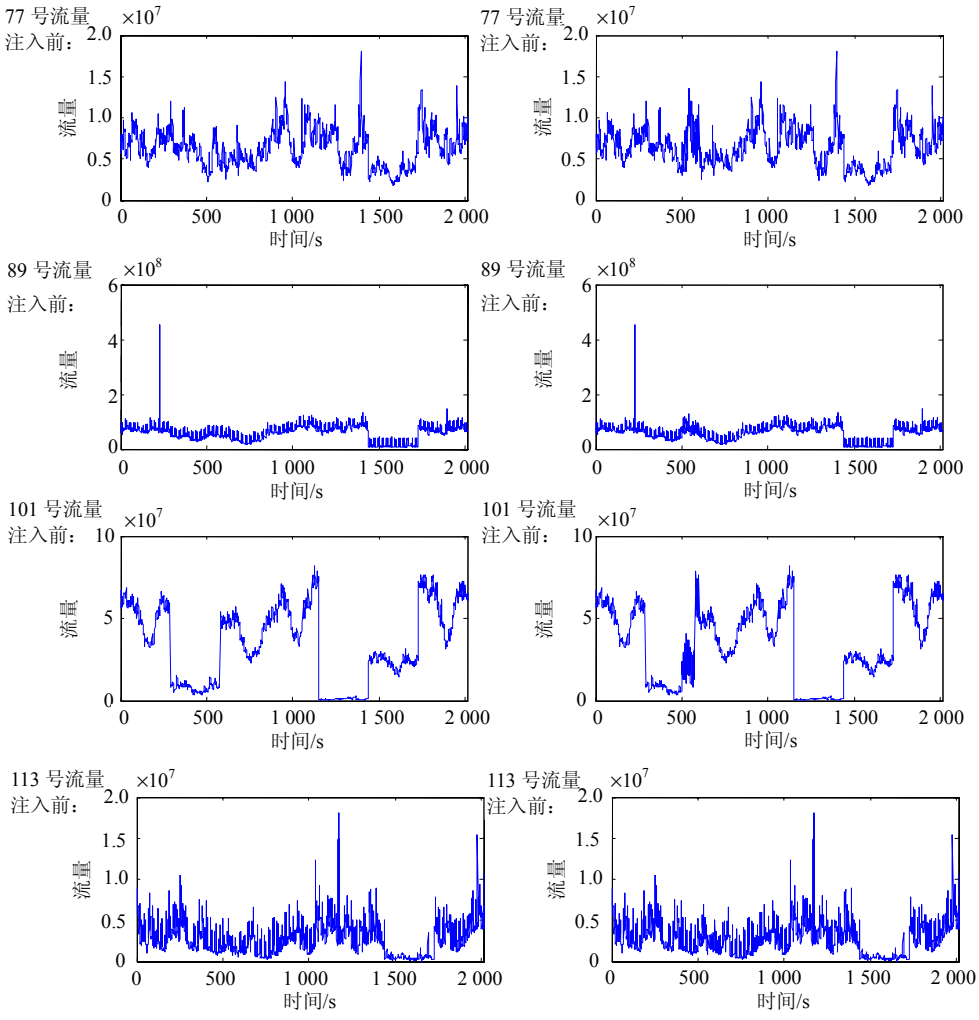
$$d=e+\alpha\delta \tag{3}$$

式中　若取α=2.4，可得到99.6%的检测率。

5　仿　真

仿真中用到的数据采集于美国的Abilene骨干网——美国教育网的骨干网。该网络有12个节点、30条链路。在每个节点上以1%的采样频率采集端到端的数据，并将每5 min采集到的所有数据叠加起来作为一个时间点的数据，这样每周有2 016个时间点。按这种采集方式，网络管理者采集了从2004-03-01～2004-09-10之间总共24周的数据，本文随机地选取第3周的数据作为实验数据，如图3流量注入前的图。

本文首先选择节点5作为受害者网络的路由节点，根据网络拓扑(如图4所示)，在77(从节点7到节点5)、89(从节点8到节点5)、101(从节点9到节点5)、113(从节点10到节点5)、125(从节点11到节点5)和137(从节点12到节点5)号OD流处加入仿真获得的攻击流，且注入攻击的时间为时间点500～600之间。这样，就模拟了一个在时间点500～600之间，由6个节点向一个节点发起的DDoS攻击，图3流量注入后的图是这些OD流加入攻击流量后的流量信号图。



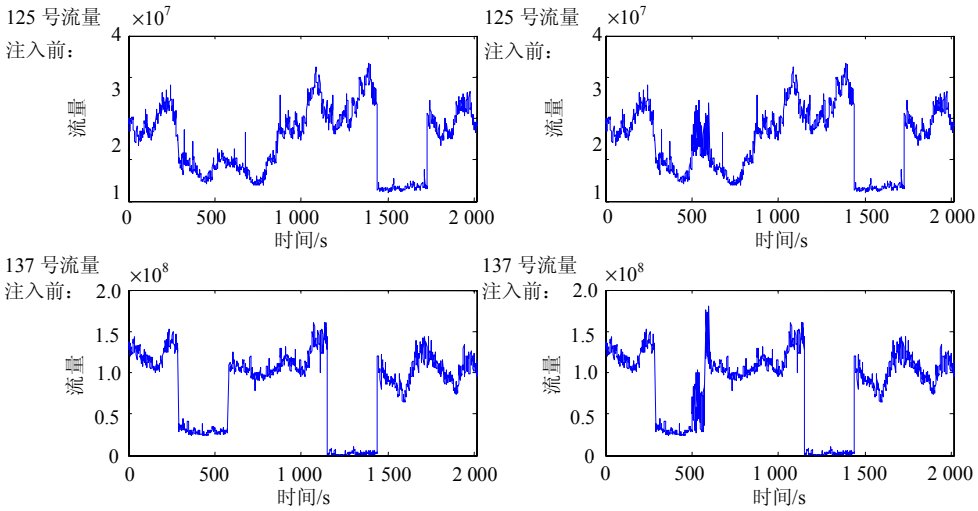


图3 注入攻击的6条流量在注入前后的比较

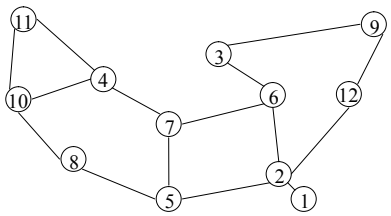
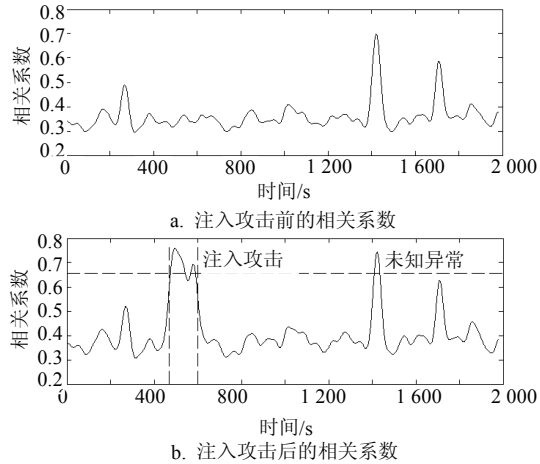


图4 Abilene骨干网网络拓扑结构图

直接从加入攻击后的流量信号来看，虽然攻击数据的加入引起了流量在时间点500～600之间有一定程度的加强，但是与其他时刻的流量比较起来，其强度并不明显。因此可以说，类似这种DDoS攻击，直接从流量情况来看，难以得出异常结论。但是通过本文方法分析后，根据最终得到如图5b所示的相关系数时序图，可以很容易的判断出异常并确定其位置。从图5b发现时间点500～600之间有一个异常，这就是人为注入的异常。此外，在时间点1 400附近也存在一个异常，这是一个未知的异常。从图3中原始流量的情况看，每条流量在1 400附近流量强度都有较大变化，可能原始流量在此处本身就包含了异常。作为比较，给出通过本文方法分析原始流量数据得到的相关系数时序图(如图5a)。

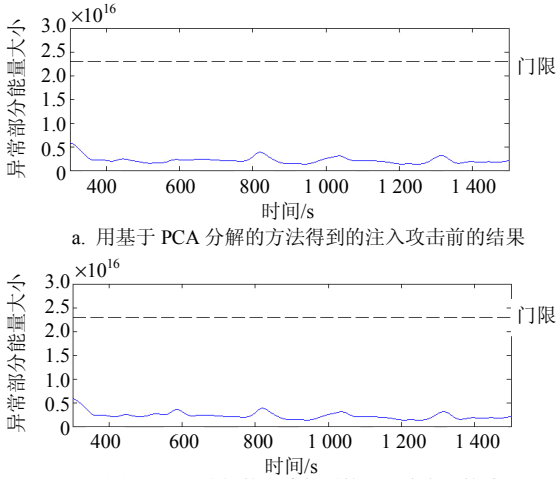
为了将本文方法与文献[4-5]的基于PCA分解的方法作一个比较，此处用基于PCA分解的方法对同样的数据(原始流量和注入攻击均相同)进行了实验仿真，得到的结果如图6所示。该仿真的实验程序以及参数设置、门限设定等完全参照文献[5]中介绍的基于PCA分解的全局流量异常检测法。从分析仿真结果图可知，人为注入的异常(500～600之间)虽然引起了流量一定程度的变化，但是变化并不显著；而且，从整个时间范围来看，根本无法判断该处存在异常。



a. 注入攻击前的相关系数

b. 注入攻击后的相关系数

图5 注入攻击前后相关系数的比较



a. 用基于 PCA 分解的方法得到的注入攻击前的结果

b. 用基于 PCA 分解的方法得到的注入攻击后的结果

图6 用基于PCA分解的方法得到的结果

6 总 结

本文提出了一种基于流量预测的全局多流量相关异常检测算法。该方法首先通过有效的流量预测算法，得到较为逼近于正常流量的流量预测值。然后，将这个预测值与流量真实值相减，得到一个包

含异常部分的预测差。最后，通过相关性分析，可将隐藏在预测差中的具有强相关性的异常部分显露出来，从而达到异常检测的目的。实验仿真证明该方法能克服现有检测方法的缺陷，对于类似DDoS攻击的强相关、低流量异常具有非常好的检测效果。

参 考 文 献

[1] CHENG C M, KUNG H T, TAN K S. Use of Spectral Analysis in Defense Against DoS Attacks[C]//Global Telecommunications Conference. [S.l.]: IEEE Press, 2002: 2143-2148.

[2] ALARCON V, BARRIA J A. Anomaly detection in communication networks using wavelets[J]. IEE-Proceedings-Communications, 2001, 148(6): 355-362.

[3] 孙钦东, 张德运, 郑卫斌, 等. 基于时频分析的分布式拒绝服务攻击的自动检测[J]. 西安交通大学学报, 2004, 38(12): 39-42.
SUN Qin-dong, ZHANG De-yun, ZHENG Wei-bin, et al. Automatic Detection of Distributed Denial of Service Attacks Based on Time-Frequency Analysis[J]. Journal of Xi'an Jiaotong University, 2004, 38(12): 39-42.

[4] LAKHINA A, PAPAGIANNAKI K, CROVELLA M, et al. Structural Analysis of Network Traffic Flows[C]// The joint international conference on Measurement and modeling of computer systems. New York: ACM Press, 2004: 61-72.

[5] LAKHINA A, CROVELLA M, DIOT C. Diagnosing Network-Wide Traffic Anomalies[C]// ACM SIGCOMM Computer Communication Review. New York: ACM Press, 2004: 219-230.

[6] LAKHINA A, CROVELLA M, DIOT C. Characterization of Network-Wide Anomalies in Traffic Flows[C]//The 4th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2004: 201-206.

[7] 罗 华, 胡光岷, 姚兴苗. 基于网络全局流量异常特征的DDoS攻击检测[J]. 计算机应用, 2007, 27(2): 314-317.
LUO Hua, HU Guang-min, YAO Xing-miao. DDoS attack detection based on global network properties of network traffic anomaly[J]. PC Digest Magazine, 2007, 27(2): 314-317.

[8] GROSCHWITZ N, PLOYZOS G. A time series model of long-term NSFNET backbone traffic[C]//ICC'94. [S. l.]: IEEE Press, 1994: 1400-1404.

[9] BASU S, MUKHERJEE A, KLIVANSKY S. Time Series Models for Internet Traffic[C]//INFOCOM'96. San Francisco: [s.n.], 1996: 611-620.

[10] SANG A, LI S Q. A predictability analysis of network traffic[J]. Computer Networks，2002，39(4)：329-345.

[11] PAPAGIANNAKI K, TAFT N, ZHANG Z L, et al. Long-term Forecasting of Internet Backbone Traffic: Observations and Initial Models[C]//INFOCOM 2003. [S.l.]: IEEE Press, 2003: 1178-1188.

[12] 汪荣鑫. 随机过程[M]. 西安: 西安交通大学出版社, 1987.
WANG Rong-xin. Random Process[M]. Xi'an: Xi'an Jiaotong University Press, 1988.

[13] 张 鹏, 胡光岷. 瞬时频率分析的网络流量异常检测[J]. 电子科技大学学报, 2007.
ZHANG Peng, HU Guang-min. Network traffic anomaly detection based on instantaneous frequency analysis[J]. Journal of University of Electronic Science and Technology of China, 2007, 36(5): 1007-1010.

编 辑 张 俊

·科研成果专利介绍·

一种多入多出无线通信基站分集天线装置

多入多出无线通信基站分集天线装置，由天线支架以及沿方位向分布在天线支架支撑杆周围的至少三副有向天线(或有向天线阵列)等构成该装置的基本方案。基于该方案，通过添加天线附近的外加电磁散射体、或架设在天线装置顶端的金属反射面、或围绕天线支架的共用反射器、或用开关天线阵列替换所述的有向天线，就可以构成该装置的各种改进方案。具有空间相关性低、性能优良、结构简单与易于制作等特点，可广泛用于无线通信以获得电波传播空间分集、角度分集或极化分集效益，或用于空时编码发时或接收，或用于多入多出(MI-MO)无线通信系统。