一种新型的基于网络流量自相似性的DDoS入侵诊测方法*)

罗光春　林　夏　卢显良　张　骏
(电子科技大学信息中心　成都610054)

A New Method of DDoS Intrude Detection Based on Self-Similarity of Network Traffics

LUO Guang-Chun　LIN Xia　LU Xian-Liang　ZHANG Jun
(Information Centre of UEST of China, Chengdu, China 610054)

Abstract　This paper presents a new method of DDoS Intrude Detection Based on Self-Similarity of Network Traffics based on analysis of parameter of self-similar, which includes Hurst parameter, Holder parameter (Time variable function H(t)), we do research on the affect of H parameter change brought by DDoS attack. And we discover the DDoS attack can be detected in some extent by measure the change of H parameter, as it showed by the research result this network traffic based method can detected DDoS attack and is more reliable on the recognition of all kinds of DDoS attack than any other method based on character recognition.

Keywords　Intrude detection, DDoS, Self-similarity, Multi-fractal

每年全球因计算机网络的安全系统被破坏而造成的经济损失达数百亿美元,近年来出现的分布式拒绝服务攻击(DDoS)更是使网络安全状况令人担忧。DDoS攻击通过操纵傀儡机来实现对目标机的攻击,由于其潜伏期长、隐蔽性高、攻击并发程度高,对网络安全造成极大危害。2000年以来大量国内外许多网站及网络系统在DDoS攻击下瘫痪使人们认识到对DDoS入侵的防范是目前网络安全中很重要的一个组成部分。

近年来也出现了许多针对DDoS攻击的检测和防范措施,但是现有的检测和防范措施大多是基于特征匹配的检测,它对普通的入侵检测比较适用,但是面对攻击原理和方式都迥然不同的DDoS攻击就显得无能为力了。为此,我们在研究了大量网络流量特性和DDoS攻击特性后提出了一种基于网络流量自相似性的DDoS检测方法,利用对真实业务流量和攻击流量进行自相似分析,可以避免对网络数据报内容进行过滤而降低检测效率。

1　DDOS攻击以及现有的防范方法

DDoS攻击如图1所示:黑客通过安装木马等手段来控制攻击控制机进而控制大量的攻击傀儡机,并在傀儡机上安装DDoS攻击程序,攻击程序统一受黑客的控制发动攻击,当大量傀儡机同时发动攻击时,大量的数据包发送到受害计算机,造成其网络拥塞或服务崩溃。

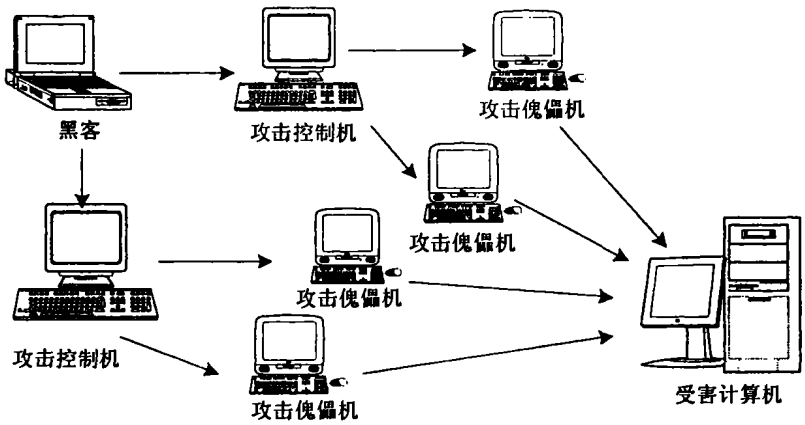

图1

攻击者最常使用的分布式拒绝服务攻击程序有:smurf、trinoo、tfn、tfn2k及stacheldraht。常用的拒绝服务攻击类型是:SYN flood, UDP flood, ICMP flood和TCP flood。目前针对DDoS攻击尚未有非常有效的解决方法,现有的方法只有在攻击到达被攻击对象所在子网后才能进行。因为要分析大量分组内容,在DDoS攻击迅速爆发时具有计算量大、占用资源过多、无法及时准确预测等局限性。

针对DDoS入侵特点,现有的解决办法有以下几种:

· 在网络上建立一个过滤器(filter)或侦测器(sniffer)诊测特征字串,过滤嫌疑的数据。

*)本文由国家九七三(项目号973-1-4-2)和电子科技大学青年基金支持。罗光春　博士研究生;林　夏　硕士研究生;卢显良　教授,博士生导师;张　骏　讲师。

· 使用 Fnd_DDoS 等软件检测特定端口的使用情况，拒绝对特定端口的访问。

· 在 Router 上对异常源 IP 地址设限，拒绝对嫌疑的客户服务。

· 监控异常半联接，异常的数据包数量、尺寸。

· 限制网络流量。

但是这几种 DDoS 入侵诊测与防范措施只是简单检测网络连接、数据源、流量、数据内容、数据包等网络元素的异常情况。由于检测的单一性，只能对特定类型的几种类型的 DDoS 入侵进行有限的防御，但是在网络攻击手段发展变化后就无能为力了。如监控特征字符串和 UDP 端口的办法，当 DDoS 入侵改变了标识或选择另外的 UDP 端口进行通讯，系统就无法诊测；又如限制流量的办法，对流量设置一个阈值，对突然增大的流量就判断为攻击行为，这种检测手段对 DDoS 入侵的判断力极低，非常容易造成误判。判断它对正常情况下的网络流可能出现的暂时流量高峰也将视作攻击。再如使用 Find_DDoS 软件可以有效地防止本机不会成为已知 DDoS 攻击的帮凶，但是无法防止自己被攻击。

为解决此类检测方法的弊端，从根本上解决问题，就必须正确判断网络流量性质、探询入侵的本质特点。为此本文提出了建立网络流量的自相似模型，通过分析网络流量的自相似及多重分形特性准确判断是否出现 DDoS 入侵。

2 网络流量的自相似性

2.1 对自相性的研究

网络自相似性揭示了网络流量在多时间尺度范围具有相对恒定的相关性的特点。自1993年以来，大量的证据显示实际网络中的数据传输是长范围相关的(long-range dependent，LRD)[1,10]、自相似的(self-similar)[2,4]即网络流量在本质上是自相似的。这是20世纪90年代高速网络领域的一个重大发现，它摒弃了对瞬时网络流量大小的简单分析，取而代之的是对一个时间段的网络流量内在的相关性分析，出现了描述自相似程度的指标：H 参数和相应的分析模型与工具。

随后的研究又发现，真实的突发业务不是严格自相似的，而是具有多重分形(Multi-fractal)的性质[3~6]，自相似业务建模只能检验其渐近自相似特性的有效程度，为便于描述突发业务多重分形性质引入了表示突发程度的 Holder 指数，及表示多重分形程度的 Legendre 多重分形谱函数 $f_L(\alpha)$[9]，Holder 指数是 $f_L(\alpha)$波形曲线取得最大值时的 α 值，有关 Holder 指数的信息都包含在了多分形谱函数 f_G[7,8]中，f_G 描述了 Holder 指数与其期望值的方差。对突发业务的多重分形分析有助于理解相同 Hurst 参数的业务源在统计复用时可能出现的不同结果，并对业务的突发度给出更为精细的分析。

f_L 与 f_G 具有同样的物理意义，但 f_G 式中有二个参变量，估计方法比较复杂，f_L 相比之下要容易计算一些。但是两种方法都涉及到计算量大的困难，难以在实际的网络环境中实时的检测 DDoS 攻击的发生与否。因此，本文采用粗粒化的多分形分析方法来对网络流量数据快速序列分析，快速计算其 Hurst 参数和 Holder 指数及时变函数 H(t)。

2.2 自相似和多重分形过程的定义及其性质

目前存在许多对自相似过程的定义，且它们并非完全等价，这里采用文[11]中的定义方法。若一连续时间过程 $Y=\{Y(t),t\in T\}$ 满足

$$Y(t)\overset{d}{=}a^{-H}Y(at),t\in T,a>0,0\leqslant H<1 \tag{1}$$

则称 Y 为自相似过程，等式的含义是指其有限维分布相同，H 称为 Hurst 系数。Y(t)经典的实例即分数布朗运动(Fractional Brownian Motion：FBM)。

自相似过程的另一种定义方法更适用于时间序列分析。设 $X=\{X(i),i\geqslant 1\}$为平稳序列，其 m 阶平滑过程定义为

$$X^{(m)}(k)=\frac{1}{m}\sum_{i=(k-1)m+1}^{km}X(i),k=1,2,\cdots \tag{2}$$

如果序列 X 是某一满足式(1)过程 Y 的差分过程，即 X(i)=Y(i+1)−Y(i)，则对所有的 m，有

$$X\overset{d}{=}m^{1-H}X^{(m)} \tag{3}$$

称满足式(3)的平稳序列为严格自相似的。如果平稳序列 X 满足当 $m\to\infty$时式(3)成立，则称 X 是渐近自相似的。

对一个随机序列自相似特性的检验通常不是考查其有限维的分布，而是考查其绝对值的矩，称之为粗粒化(Coarse Grain)的方法。令

$$\mu^{(m)}(q)=E|X^{(m)}|^q=E|\frac{1}{m}\sum_{i=1}^{m}X(i)|^q,m=1,2,\cdots \tag{4}$$

如果 X 是自相似的，则有

$$\log\mu(m)(q)=\beta(q)\log m+C(q) \tag{5}$$

其中 $\beta(q)$是 q 的线性函数。

特别地，对于满足 $X=dm^{1-H}X(m)$的严格自相似序列 X，则有

$$\beta(q)=q(H-1) \tag{6}$$

显然，利用(5)、(6)两式还可给出自相似过程的第三种定义，而且由式(6)还可以推广到 $\beta(q)$不是 q 的线性函数的情形，即多重分形过程。此时 $H(q)=1+\beta(q)/q$ 。在 H(q)图像中可以看出曲线是单调下降的，在较高的阶数下计算出来的 H 参数较小。

下面定义多重分形。称一连续时间过程 $Y=\{Y(t),t\in T\}$是多重分形的，即

$$Y(t)\overset{d}{=}a^{-H(t)}Y(at),t\in T,\forall\ a>0 \tag{7}$$

其中 H(t)称为时变的尺度系数或 Holder 指数，表征该过程的局部奇异性。与自相似模型不同，在多重分形中尺度函数 $\beta(q)$与 q 未必是线性关系，因此它实际上是自相似模型的推广，其 Holder 指数及多重分形谱描述了比自相似过程的 Hurst 系数更为丰富的信息。

要判断一个离散随机过程是自相似还是多重分形的，仅仅检验序列的二阶统计特性是不够的，必须分析业务的高阶统计特性。传统的 R/S 分析、方差-时间分析和 IDC 等方法可用来检验序列的长时相关特性，但是它们都是基于二阶统计量的分析方法，因此不足以判定序列是否具有多重分形性质。真实的突发业务不可能是严格自相似的，业务建模只能检验其渐近自相似特性的有效程度。因此我们要对公式(4)(5)(6)中的 q 取不同值以保证检验其高阶特性。

3 DDoS 入侵检测试验

DDoS 入侵是一种人为的大规模数据流，它破坏了网络流量的自相似性和多重分形性，表现为自相似的 H 参数和 H(t)函数的异常变化。可以利用这种变化来准确识别是否出现 DDoS 入侵，以及 DDoS 入侵的规模，DDoS 的类型等信息。下面我们通过试验来说明 DDoS 攻击对网络流量自相似性的影响。网络环境：100兆以太网，100兆路由器，100兆交换机。4台安装 Win2K 的攻击计算机。1台安装 RedHat Linux 的检测计算机。

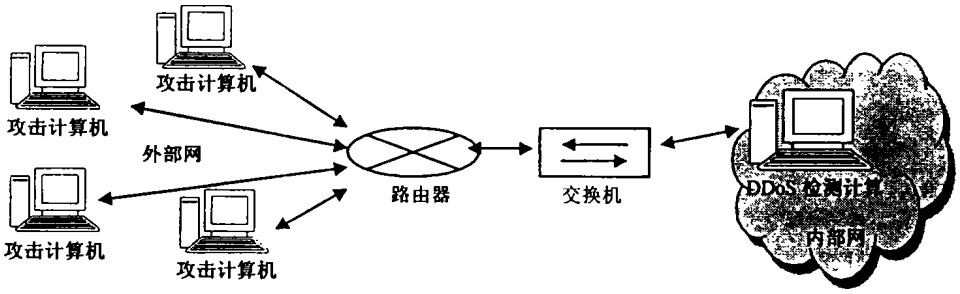

图2 DDoS入侵检测网络结构图

4 对真实业务流量和攻击流量的自相似分析

本节我们对各种真实的突发业务源数据进行统计特征分析，以判定用自相似或多重分形进行DDoS检测的有效性，我们选取电子科技大学网络中心提供的一天的Etherent流量数据加以分析每4分钟的流量累计在一起形成流量数据。限于篇幅，我们以2002年12月20日的一次典型的攻击作为分析的范例。日常的网络流量约为10^5数量级，我们使用DDoS攻击模拟器从3点到6点进行持续时间约为3小时，攻击速率为5kBpS类型为FakeUDP的模拟攻击，利用(5)、(6)两式对全部的业务流量进行自相似和多重分形特征的判定。如果由粗粒化方法得到不同q值在对数坐标下式(5)渐近为直线，则说明多重分形建模是有效的；如果在q的不同取值下由式(6)估计的Hurst系数近似相等，则说明仅用自相似建模即是充分的。图3(a)为正常的流量数据，(b)加上攻击数据的流量数据。

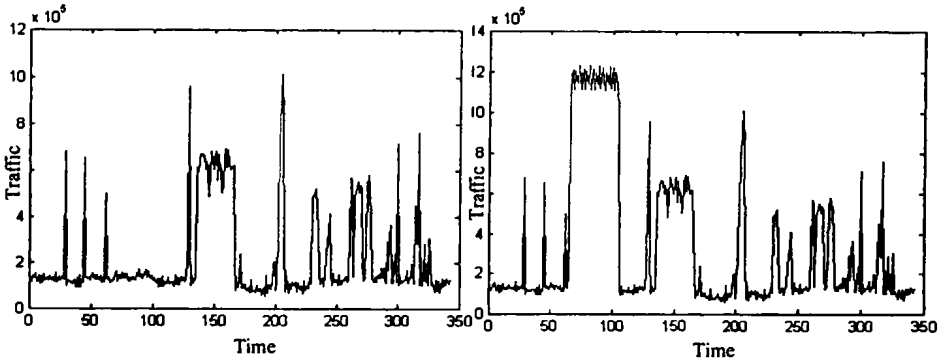

图3 (a)正常的流量数据 (b)加上攻击数据的流量数据

从图3(b)可以看出横坐标从65到108就是时间从3点到6点进行持续时间约为3小时的攻击流量数据。由公式(5)可以得到如图4所示的$\mathrm{Log}\mu^{(m)}(q)$-Log(m)曲线

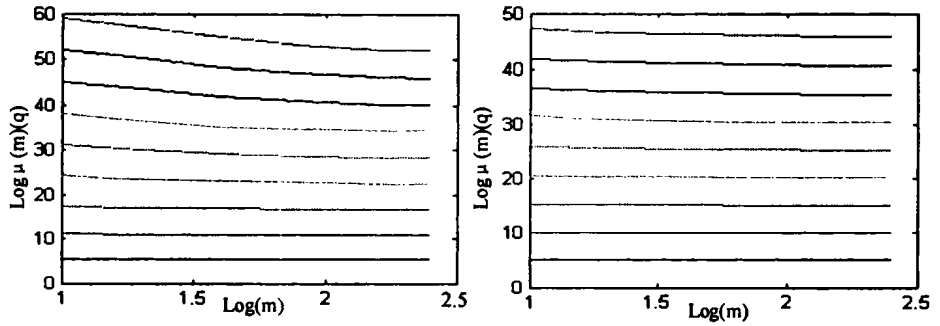

图4 (a)正常流量不同q值$\mathrm{Log}\mu^{(m)}(q)$-Log(m)曲线 (b)攻击流量的不同Q值$\mathrm{Log}\mu^{(m)}(q)$-Log(m)曲线

从图4可以看出对应不同的q值，粗粒化结果在logm=[1,2.5]内接近于直线，表明多重分形建模是有效的；由于数据长度有限，当logm较大时，选取的q越大，粗粒化估计的方差也越大。对于q取负值的情形，其相应矩的含义不明确，而且统计量的数值稳定性较差，故在此不作分析。用公式(6)可以得到如图5所示的H参数的曲线。

下面，把整个数据区间分成20个子区间来研究各个子区间，横坐标是时间区间号，纵坐标是对应区间的H参数值。各个子区间的H参数时变曲线H(t)，从上到下为q值从3到7对应的5条曲线。

可以看出当有攻击发生时5条曲线比较紧密地凑在一起并且比较靠近高H值，表示其自相似性非常高，H参数几乎不受q值影响而变化表示没有多重分形的特征，上述分析表示在这些时间点遭受到攻击，从而使H参数的特性发生了变化。

正常情况下H(q)曲线比较平缓地下降，如图5(a)表示的整个一天的流量数据的H(q)曲线，在有攻击发生的(a)(b)两图的图像并没有发生多大的变化。图6的对比就比较明显了，

在特定时间段的H(q)曲线表示了那个特定时间段q的取值对H参数的影响。

对前10个区间每个区间绘制H(q)曲线，如图7所示。

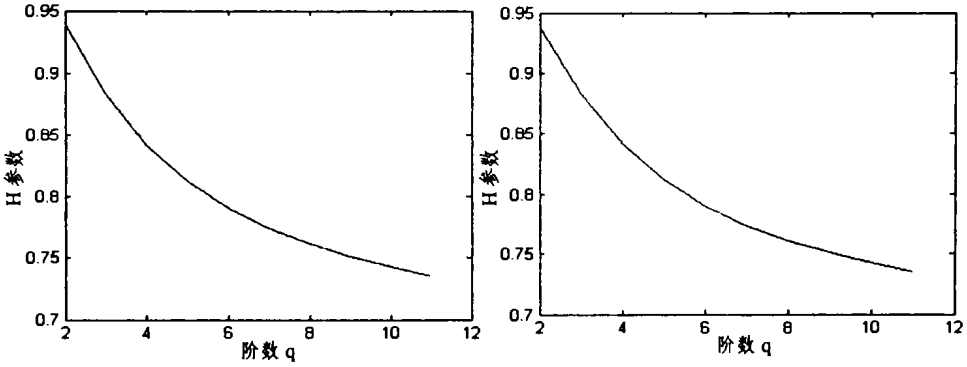

图5 (a)正常流量H(q)曲线 (b)加上攻击数据的H(q)曲线

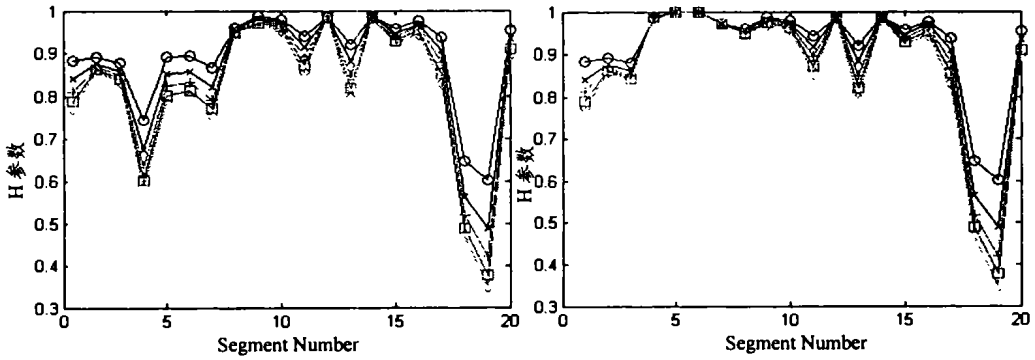

图6 (a)正常流量对不同q值H(t)曲线 (b)加上攻击数据的对不同q值H(t)曲线

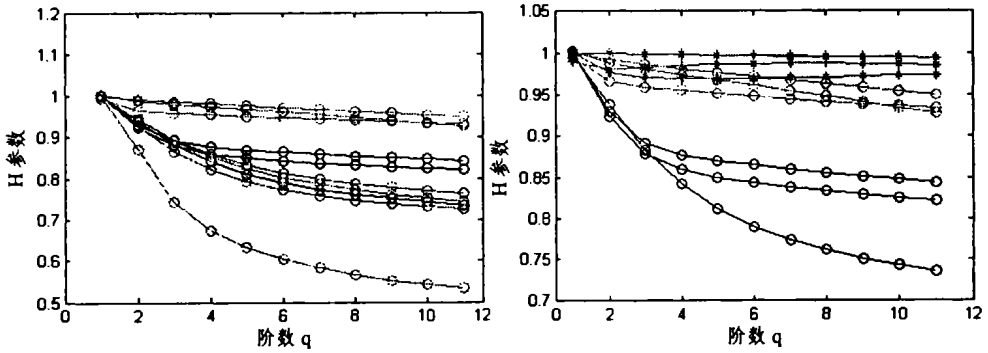

图7 (a)正常流量对不同子区间H(q)曲线 (b)加上攻击数据对不同子区间的H(q)曲线

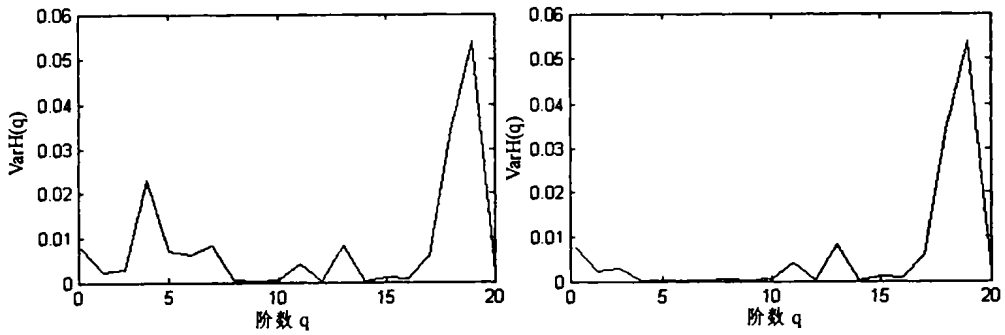

图8 (a)正常流量对不同子区间H(q)方差线 (b)加上攻击数据对不同子区间的H(q)方差线

可以看到缺少多重分形特征的攻击数据形成的曲线(用"－＊－"标注的几条曲线)比较平缓而其他的曲线则随q值的增加而逐渐下降。方差分析可以判断出H参数对于q值的稳定程度，并可以作为判断DDoS入侵的依据。下面是方差分析的公式

$$V(q)=\mathrm{Var}H(q)=E([H(q)-EH(q)]^2) \quad (8)$$

对每条H(q)曲线做方差分析可以得到如图8所示的V(q)曲线图像。

观察图8(b)可见DDoS攻击使H(q)曲线方差变得很小。时间轴的第4到7区间正好是模拟DDoS攻击所在的区间，这4点的方差值都在0.001以下，和图8(a)比较起来有较大程度的下降。但是除了DDoS攻击外也有其他原因可能导致方差变小，比如12和14区间的方差也很小，这表明这两点本身的流量就缺乏多分形特点。

5 基于网络流量自相似性的DDoS入侵检测方法

使用第4节的方法，分别使用攻击软件中的ICMP flood，SYN flood，UDP flood，TCP flood四种攻击手段进行各自5000次攻击测试，得出以下判断DDoS入侵的经验条件：

令 $H_d=0.95$，$V_d=2\times10^{-3}$，

$$H_d< H(t)<1 \text{ 且 } 0<V(q)<V_d \quad (9)$$

满足这个公式的时间区间内可以判断有DDoS攻击的存在。根据四种攻击手段进行各自5000次攻击测试的数据统计，此方法的漏判率和误判率如表1所示。

表1

攻击类型	漏判率	误判率
ICMP flood	13.7%	20.2%
SYN flood	14.6%	21.2%
UDP flood	13.1%	21.5%
TCP flood	10.8%	17.4%

可见有的网络业务流量本身就缺乏多分形特征，所以导致此方法的误判率较高都在17%以上，由于比起其他几类数据报来TCP数据报的多分形特征较强，所以对TCP flood的误判率也较低。增大H_d、减少V_d都可以使误判率降低但是也会增加漏判率，上述H_d和V_d值是在电子科技大学网络中心的局域网内12月份的最优化取值，对某个特定的网络来讲应该根据情况选取适合自身的H_d和V_d值，并且应该定时调整。

表2

攻击检测率 / 方法 / 攻击方式	NetST 2103	FW3010 PF	黑客愁	NetEye3.0	自相似性检测
ICMP Flood	80%	---	77%	80%	86.3%
SYN Flood	70%	80%	77%	---	85.4%
UDP Flood	85%	---	77%	---	86.9%
TCP Flood	75%	85%	77%	80%	89.2%

6 自相似分析方法与传统方法的对比

自相似分析方法是基于网络流量自相似本质的检测方法，它和基于特征匹配的传统入侵检测方法在检测的基本原理上不同，也使用于不同的检测场合。自相似分析方法用于在受害计算机的子网上检测DDoS攻击，而传统方法是用于在黑客使用的计算机和控制计算机以及傀儡攻击计算机之间的通讯数据检测、驻留的控制程序检测和攻击程序检测。对受害计算机子网上的检测效率和过滤成功率都不高。表2显示了几种主流DDoS检测产品和自相似性的检测方法在检测成功率方面的比较。

自相似性方法有较高的攻击检测率(即较小的漏判率)，比起传统方法有明显提高，但误判率指标还有待提高。

结论 本文提出了一种新型的基于网络流量自相似性的DDoS入侵检测方法，该方法应用粗粒化计算方法对网络流量的自相似参数：Hurst参数、Holder指数详细的分析，揭示了DDoS入侵对网络流量自相似参数的影响。并提出判断DDoS入侵的参数标准。试验表明此方法适合作为检测DDoS入侵的依据并且比传统的DDoS入侵检测系统在检测的准确度有较大提高，而且能够适用于各种类型的DDoS入侵和未来的DDoS变种。未来工作是：1)进一步改进检测算法，降低误判率，使检测系统在网络业务流量多分形性较低的情况下也有较好的检测效果；2)利用自学习算法提高H_d和V_d的自调整能力；3)改进检测算法以提高检测速度，适应实时检测的需要。

参考文献

1 Taqqu M S, Teverovsky V. On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. Preprint Boston University, USA, 1996

2 Popescu A. Traffic Self-Similarity. In: Proc. of the IEEE Intl. Conf. on Telecommunications, Jun. 2001

3 Kargl F, Maier J, Weber M. Protecting web servers from distributed denial of service attacks. In: Proc. of 10th Intl. World Wide Web Conference, May 2001

4 蔡弘，陈惠民，李衍达. 一种新型的通信网络突发业务建模方法—自相似业务. 通信学报，1997，18(11)：51～59

5 许都，李乐民. 网络中业务流的自相似性与线性AR1模型. 电子学报，1999，27：(4)

6 张鹏，廖建新，程时端. 自相似业务量的多重分形分析. 电子学报，2000，28(1)：96～98

7 Flandrin P. Wavelet Analysis and Synthesis of Fractional Brownian Motion. IEEE Transactions on Information Theory, March 1992, 38(2)

8 Halsey T, et al. Fractal measures and their singularities: The characterization of strange sets. Phys. Rev. 1986

9 Reiher P, Prier G, Michel S, Li J. Project D-WARD: DDoS Network Attack Recognition and Defense. UCLA, http://lever.cs.ucla.edu/ddos/, Aug. 2001

10 Leland W, Taqqu M, Willinger W, Wilson D. On the Self-Similar Nature of Ethernet Traffic. IEEE/ACM Transactions on Networking, February 1994, 2(1): 1～15

11 Meadows C. A formal framework and evaluation method for network denial of service. In: Proc. of the 12th IEEE Computer Security Foundations Workshop, June 1999