一种基于主动探测机制的SYN Flooding攻击检测方法

李海伟[1] 张大方[2] 刘 俊[1] 杨晓波[1]
(湖南大学计算机与通信学院 长沙 410082)[1] (湖南大学软件学院 长沙 410082)[2]

摘 要 SYN Flood给网络正常运行带来极大危害,而已有广泛研究的基于流量自相似性的检测方式对这种小包攻击可能会失效。通过对DAG卡捕获高精度流量样本进行分析,提出一种基于主动探测机制的SYN攻击检测方法。该方法将包对测量背景流量技术应用于异常流量检测中,用夹入背景流长度变化来检测攻击。实验表明,该算法对SYN攻击检测率可达88%。这种基于端到端的检测方法,具有良好的灵活性和可控制性等优点。

关键词 SYN Flooding攻击,自相似性,异常检测,包对
中图法分类号 TP393.08 **文献标识码** A

Active Detecting Method against SYN Flooding Attacks

LI Hai-wei[1] ZHANG Da-fang[2] LIU Jun[1] YANG Xiao-bo[1]
(Department of Computer Science, Hunan University, Changsha 410082, China)[1]
(Department of Computer Software, Hunan University, Changsha 410082, China)[2]

Abstract SYN Flood brings great danger to the normal network operation. Many research studies detect the attack by analyzing the self-similarity of network traffic. However, the method may be ineffective to SYN Flood. By analyzing the high-precision traces which are captured by DAG cards, we proposed a new SYN Flood detection mechanism based on the active detection. It brings the technology of packet-pair to abnormal traffic detection that detects SYN Flood, according to the background flow length change. The method has a 88% SYN attack detection rate from experimental results. This method is based on end-to-end technology which has better flexibility and controllability.

Keywords SYN flooding attack, Self-similarity, Abnormal detection, Packet pair

1 引言

分布式拒绝服务(DDoS)攻击是一种通常难以防范的攻击手段。DDoS通过非法请求占用大量网络资源,给网络的正常运行带来极大危害[1]。SYN Flood是当前最流行的DoS(拒绝服务攻击)与DDoS(分布式拒绝服务攻击)方式之一。该方式利用TCP协议缺陷,发送大量伪造的TCP连接请求,从而使得被攻击方资源耗尽(CPU满负荷或内存不足),暂停服务。目前,DDoS攻击中约有90%是SYN Flooding攻击,所以对SYN Flooding攻击的检测是当前DDoS攻击研究的重点之一。

DDoS检测和防御方法基本可以分为3类:基于协议特征分析的DDoS检测和防御[2]、基于聚积的DDoS检测和防御[3]以及基于网络流量统计模型的DDoS检测和防御[4-8]。大量研究表明,正常的网络流量具有自相似性[9-12]。

近年来,利用网络自相似性检测DDoS攻击的研究得到很大发展[6-8]。自相似性方法首先分别求解正常与发生DDoS攻击时整个时间段的Hurst指数,再计算其偏差值,最后比较设定的门限值与偏差值,从而判定是否发生攻击。例如,文献[6]中用H值改变的方差来确定攻击。文献[7]中提出在发生Flood攻击时,H值会明显减小。文献[8]指出在发生Flood攻击时H值发生显著改变,可用改变量大小来判断攻击,但变化方向不能确定。然而通过对大量DAG卡[13]捕捉高精度流量样本进行分析表明,在发生SYN攻击时,由于攻击包包长较小,流量方差在正常和非正常情况下并没有发生大的变化,且变化方向是不确定的,并不像以前研究所指改变很大或失去自相似性,因此用单纯计算流量H值方法检测SYN攻击,结果是不可靠的,可能造成严重漏警情况。

传统的攻击检测方式是基于被动检测。被动检测具有以下缺点:灵活性小,局限性大,不能按照检测者的意愿进行检测,受到网络机构及检测工具等多方面的限制。因此,本文提出一种新型的基于主动探测机制的检测方式,该方式将包对技术应用于SYN Flood攻击检测中。实验表明,这种新型的端到端的检测方式能较准确地检测出攻击。并且,基于端到端的主动检测方式方便部署,对正常流量干扰不大。

2 检测模型原理

2.1 包对测量背景流量技术简介

包对技术作为一种主动测量技术用于背景流量和可用带宽的测量,已经有了广泛的研究[14]。其基本原理是主动发送

到稿日期:2009-04-24 返修日期:2009-07-09 本文受国家自然科学基金(NSFC)项目(90718008,60673155)资助。
李海伟(1983-),女,硕士生,主要研究方向为网络测试,E-mail:lihaiwei-1028@163.com;张大方(1959-),男,博士,教授,博士生导师,CCF会员,主要研究方向为可信系统与网络、网络与信息安全等;刘 俊(1974-),男,博士生,主要研究方向为网络测量、流量特征分析;杨晓波(1973-),女,博士生,讲师,主要研究方向为可信系统与网络、人工智能、软件理论。

一系列固定间隔的包对，在经过瓶颈链路时，由于受排队机制的影响，包对之间间隔会因为夹入或未夹入背景流量而发生变化，从而可由间隔变化情况来计算当前链路的背景流量。图1是其基本原理图。

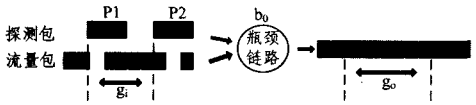

图1 包对测量背景流量模型

如图1所示，发送一对探测包对P1和P2，设初始探测包对的发送时间间隔为 g_i，输出间隔为 g_o，瓶颈链路带宽为 b_0。假如现在时刻为 t，经过时间 ε 后平均背景流量为 b_c，则 b_c 为

$$b_c = \frac{b_0}{\varepsilon}\int_t^{t+\varepsilon} b_c(t)\mathrm{d}t \tag{1}$$

时间 ε 内总的背景流量为

$$V(\varepsilon) = b_0\int_t^{t+\varepsilon} b_c(t)\mathrm{d}t = b_c\varepsilon \tag{2}$$

本文采用背靠背包对（两个连续发送的包）进行测量。设探测包长为 l_c（即P1或P2的长度），则背靠背包对的输入间隔为

$$g_i = l_c/b_0 \tag{3}$$

由于夹入流量包，导致包对间隔增加，设增加的输出间隔为 G，则

$$G = g_o - g_i \tag{4}$$

同时可由业务量变化来求增加的输出间隔 G

$$G = (V(t+g_i)-V(t))/b_0 = (C\cdot(t+g_i)-C\cdot t)/b_0 = Cg_i/b_0 \tag{5}$$

由式(3)和式(5)可得通过输入输出间隔变化来求背景流量 b_c 的公式为

$$b_c = Gb_0/g_i = Gb_0^2/l_c \tag{6}$$

上述计算模型成立的条件是探测包P1和P2在路由器排队时在同一队列中，即处于JQR（joint queuing region）区域，背靠背探测包对保证了这一条件的成立。另外，同其他包对模型一致，本文基于以下假设：分析模型采用单跳模型，排队机制为FCFS。

2.2 基于包对的检测模型介绍

将包对测量背景流量技术应用于攻击检测中。发生SYN Flood攻击时，链路小包数量及到达时间分布异于正常情况，因此影响包对间夹入背景流量的分布状况。基于这种思想，提出基于端到端的攻击检测办法：发送背靠背包对，对每对夹入的背景流量按频段聚合。由于网络流量包长呈双峰分布，因此在SYN攻击（夹入大量小包）和无攻击情况下，流量频度统计会出现相异的特征。检测模型如图2所示。

图2 实时检测模型

下面总结具体实现步骤。

Step1 发送固定包长的背靠背包对，每个包对之间的时间间隔为0.3ms（此时发送的探测包速率与背景流量速率之和与瓶颈链路带宽相当）。

Setp2 捕获包对，依式(6)计算背景流量 b_c。

Step3 对每对包对之间夹入的流量进行分段频度统计，统计方式如表1所列。

表1 夹入背景流量分段方式

组号	1	2	3	4	5
流量分段（字节）	0～250	250～500	500～750	750～1000	1000～1250
组号	6	7	8	9	10
流量分段（字节）	1250～1500	1500～1750	1750～2000	2000～2250	2250～2500

注：由于是背靠背发包，每对包之间最多夹入的流量为2500字节。

Step4 异常比较，对分段频度统计结果求比值，即当前时刻各段频度/与前一时刻各段频度。

Setp5 判定攻击，正常情况下包长属于250～500段流量稀少，异常情况下大量小包聚集，导致该段流量增加。设定经验比较门限值（当前时刻与前一时刻频度统计比）为4，确定是否有攻击。

由图2可以看出，这种检测方式在上一组进行频度统计的同时，即可进行下一组数据检测，具有良好的实时性。将每一次统计结果与前一时刻相除，若改变倍数超过设定的门限值，认为发生攻击；若在有攻击时，下一组比值降低到某一正常水平，认为攻击结束。该模型具有良好的灵活性，可旁路布置在任何需被检测的服务器附近。

3 实验

3.1 实验环境及样本数据介绍

本研究采用的实验数据来自湖南大学软件学院实时采样流量。实验环境拓扑图如图3所示，其中有超过1000台PC机，在学院出口处配置分路器，用DAG卡捕获整个进出口学院的流量。采用DAG卡捕包的原因如下：1）软件捕包器时间戳误差太大，其捕捉数据不可靠；2）DAG硬件捕捉卡的时间精度[13]可以达到60ns，比精度为毫秒级的Tcpdump等软件捕包器[13]高出5个数量级。

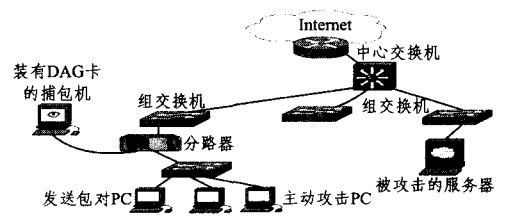

图3 实验环境拓扑图

发送包对。PC连续发送两个1000字节的UDP包作为背靠背探测包，间隔0.3ms发送下一组。攻击机在Windows XP操作系统环境下布署，攻击软件采用几种常见的DoS攻击工具，如HGOD，SYN Flood和多线程SYN攻击工具0.2版等。通过在攻击过程中控制攻击线程数来调节攻击强度，由装有DAG卡的捕包机采样获取4组不同强度的攻击样本。作为对比研究，在每组攻击之后又及时捕获4组无攻击样本。踪迹具体信息如表2所列。

表2 实验采用样本踪迹

		瓶颈链路带宽（Mbps）	包个数	总流量（Mb）	平均流量（Mbps）	平均包长（Bytes）
第一组	5个线程攻击	100	65500	364.03	27.19	630.34
	无攻击	100	65500	403.14	27.71	769.36
第二组	6个线程攻击	100	65500	384.55	49.41	694.54
	无攻击	100	65500	462.37	49.21	882.38

第三组	8个线程攻击	100	65500	339.75	49.50	648.37
	无攻击	100	65500	462.36	49.21	882.36
第四组	10个线程攻击	100	65500	348.87	52.96	665.79
	无攻击	100	65500	459.68	52.21	877.26

3.2 数据分析

3.2.1 确定探测包长

首先确定探测包对的包长。选择探测包长的原则是使有攻击和无攻击情况下夹入背景流量的方差比值最大，即使在发生攻击后背景流量方差波动较大。用不同的探测包长分别对每组踪迹按式(6)计算背景流量及其方差，将每组中的有攻击和无攻击背景流量方差相除，得到图4。

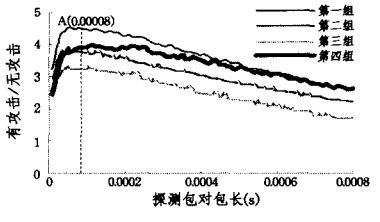

图4 不同探测包长夹入背景流量波动比较图

如图4所示，横轴代表探测包长(每个包长在100Mbps链路上的传输时间)的变化，纵轴代表在每个探测包长有攻击和无攻击时测得的背景流量方差比。可以看出4组曲线均约在 A 点时获得较大比，说明此点有攻击时背景流量方差波动较大，即该点对有攻击状况较敏感。A 点所对应的时间为80μs，链路瓶颈带宽为100M，计算其包长为 0.00008 * 100000000/8=1000 字节。

3.2.2 实验结果

对4组样本踪迹进行实验分析。采用探测包长为1000字节的背靠背包对，此时 g_i 为 0.00008s，瓶颈链路带宽为100M。采用2.1节方法计算，得到图5实验结果。

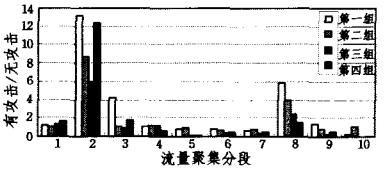

图5 夹入背景流量频度统计比较图

如图5所示，横轴表示按表1所分的10个背景流量段，纵轴表示频度比，每段中4个柱体分别代表第一、二、三、四组数据中有攻击和无攻击频度比。可以看出，第2个分组比最大，均在6倍以上，此时夹入背景流量为250～500字节之间，说明可将夹入背景流量在该段的变化比作为发生SYN攻击的一个特征。经过多组实验，得到从无攻击到有攻击时的经验频度比为4。

为进一步考察该特征是否为从无攻击到有攻击时的特有特征，再对流量在250～500之间纯带攻击和纯无攻击的频度情况进行比较，对表1中的数据重分组，分组方式为循环式，即得到第一组无攻击数据/第二组无攻击数据的第一个比、第二组无攻击数据/第三组无攻击数据的第二个比，依次类推，得到4个比值，对4组异常数据进行同样分组得到4个比值。比较结果如图6所示。

纵轴仍然为频度比，横轴是重分组后的4个组，每组中的两个柱体分别代表正常流量频度比和异常流量频度比。由图中可以看出，在正常或异常情况下，无论背景流量是否相同，包对夹入背景流量在250～500间的波动都不大，均小于3倍，再一次说明该段流量显著增加是有攻击发生时的特征。

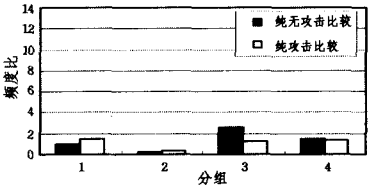

图6 背景流量在250～500间的频度比较图

由上面两个实验分析可得出结论：只有在从无攻击到有攻击变化时，夹入长度为250～500字节的背景流量才显著增加，可用来判断是否发生攻击。增加比超过设定经验门限值4倍时，发生攻击；小于4倍时，则认为没有发生攻击。

3.3 异常变化原因分析

大量研究表明，网络流量包长呈双峰分布，包长小于50字节和大于1400字节的包约占80%。因此，在无异常情况下，夹入长度为250～500字节背景流概率相对较小。在SYN Flood攻击情况下，容易造成小包连续到达，从而使250～500字节的流量增多(即这些流量是多个小包叠加的结果)。为此，对小包的到达时间进行研究。

研究方法：将大包从总的流量中剔掉，只对小包进行时间到达分析，用变异系数CV来表征小包的聚集情况。在有攻击和无攻击情况下均值和方差均会发生一定变化，因此不能单独用来揭示变化程度，可用平均变化来表征聚集程度。设小包到达时间为 t，CV定义如下：

$$CV(t)=\sqrt{D(t)}/E(t) \tag{7}$$

将有攻击和无攻击情况进行比较：

$$CV\text{比}=CV(\text{有攻击})/CV(\text{无攻击}) \tag{8}$$

用上面公式对4组踪迹的小包到达时间CV值分别进行计算，如表3所列。

表3 有攻击和无攻击小包到达时间CV比值

	5个线程	6个线程	8个线程	10个线程
CV比	1.472289	1.577061	1.580559	1.461728

由表3可以看出，比值均超过1，说明在有攻击时，小包的到达时间间隔波动较大，即小包呈现连续聚集到达，不连续包之间间隔增大。此现象也可由图7直观说明。

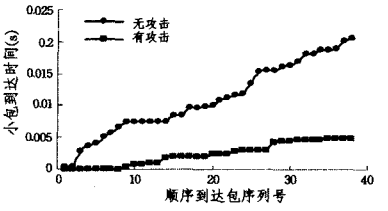

图7 有攻击和无攻击小包到达时间分布比较图

如图7所示，在无攻击时，小包到达分布较均匀，时间上连续到达情况较少。而有攻击时，小包明显呈现分段连续到达。因此，造成夹入长度250～500字节背景流量显著增加的原因是小包连续到达，形成多个包聚集。这种小包聚集现象也可以解释图5中第8个分段比值也明显增大的原因：包长呈双峰分布，1500字节的包占一定数量，又在有攻击情况下，250～500之间流量增多，造成1750～2000之间流量也明显增多。

4 模型性能评估

检测算法有2个性能指标：检测率，即算法检测到的攻击次数与实际发生攻击的次数的比率；误报率，即把正常行为误认为异常攻击的次数与实际发生攻击的次数的比率。为考察本文检测方法的效率，对实时抓取的100组数据进行了比较分析，其中前10组和最后10组无攻击，中间80组有攻击。采用的分析探测包长为1000字节，设定夹入长度250～500字节之间改变的门限值仍然为4，得到的结果如表4所列。可以看出，本模型有较好的检测率，但存在误报率偏高的缺点。

表4 检测算法性能分析

	检测率	误报率
基于包对的检测技术	88%	3%

结束语 本文提出一种新型的端到端的SYN Flood攻击检测方法。发送背靠背固定间隔的包对来观察有无异常时夹入背景流量的变化，通过对校园网实时采样高精度流量样本进行分析，得出如下结论：(1)探测包长为1000字节时异常特征变化较明显；(2)在有异常时，夹入包对间长度为250～500字节间的流量显著增加；(3)出现该异常现象的原因是SYN Flood攻击时大量小包连续到达，小包聚集形成夹入长度250～500字节的流量增加。本方法具有实时性好、算法简单、易于实施及节省资源等优点，且算法新颖，并有望实现端到端的自动攻击检测。

参 考 文 献

[1] Chang R K C. Defending against flooding-based distributed denial-of-service attack: a tutorial[J]. IEEE Comm Magazine, 2002, 40(10): 42-51

[2] Wang H N, Zhang D L, Shin K G. Detecting SYN flooding attacks[C]// Proc. of the 21st Annual Joint Conf. of the IEEE Computer and Communications Societies. 2002, 3: 1530-1539

[3] Jin C, Wang H N, Shin K G. Hop-Count filtering: An effective defense against spoofed DDoS traffic[C]// Proc. of the 10th ACM Conf. on Computer and Communications Security. 2003: 30-41

[4] Kim Y W, Lau W C, Chuah M C, et al. Packetscore: Statistical-based overload control against distributed denial-of-service attacks[C]// Proc. of the 23rd Annual Joint Conf. of the IEEE Computer and Communications Societies. 2004, 4: 2594-2604

[5] 严芬，王佳佳，陈轶群，等. 一种轻量级的SYN Flooding攻击检测方法[J]. 计算机科学，2008，35(9)：72-75

[6] Rohani M F, Maarof M A, et al. An implementation of LoSS detection with second order statistical model[C]// Proceedings of the Postgraduate Annual Research Seminar. FSKSM, UTM, 2007

[7] Li M. Change trend of averaged hurst parameter of traffic under DDOS flood attacks[J]. Computer & Security, 2006, 25(3): 213-220

[8] 任勋益，王汝传，王海艳. 基于自相似检测DDOS攻击的小波分析方法[J]. 通信学报，2006，27(5)：6-11

[9] Leland W, Taqqu M, Willinger W. On the self-similar nature of Ethernet traffic(Extended Version)[J]. IEEE/ACM Trans on Networking, 1994, 2(1): 1-15

[10] Paxson V, Floyd S. Wide area traffic : the failure of poisson modeling[J]. IEEE/ACM Tracns on Networking, 1995, 3(3): 226-244

[11] Dang T D, Molnar S. On the Effects of Non-stationarity in Long Range Dependent Tests[R]. Budapest, Hungary: Budapest Univ Technology and Economics, 1999

[12] Abry P, Veitch D. Wavelet analysis of long range dependent traffic[J]. IEEE Trans on Infor Theory, 1998, 44(1): 2-15

[13] Arlos P, Fiedler M. A method to estimate the timestamp accuracy of measurement hardware and software tools[C]// Passive and Active Measurement Workshop. 2007: 197-206

[14] Hu Ningning, Steenkiste P. Evaluation and characterization of available bandwidth probing techniques[J]. IEEE Journal on Selected Areas in Communications, 2003, 21(6): 879-894

(上接第82页)

[7] Toh C K. Associativity-based routing for ad hoc mobile networks[J]. Wireless Personal Communications, 1997, 4(3): 103-139

[8] Chakeres I D, Perkins C E. Dynamic MANET On-demand (DYMO) Routing [EB/OL]. draft-ietf-manet-dymo-17, Internet Draft, March 2009

[9] The IETF MANET Working Group. Charter Website [EB/OL]. http://www.ietf.org/html.charters/manet-charter.html

[10] Perkins C, Belding-Royer E, Das S. Ad hoc On-demand Distance Vector (AODV) Routing[EB/OL]. http://www.ietf.org/rfc/rfc3561.txt, July 2003

[11] Perkins C, Royer E. Ad hoc On-Demand Distance Vector Routing[C]// 2nd IEEE Workshop on Mobile Computing Systems and Applications. New Orleans, LA, February 1999: 90-100

[12] Sommer C, Dressier F. The DYMO Routing Protocol in VANET Scenarios[C]// Vehicular Technology Conference 2007. 2007: 16-20

[13] Chakeres I D, Belding R E M. PAC: perceptive admission control for mobile wireless networks[C]// Quality of Service in Heterogeneous Wired/Wireless Networks. Piscataway: IEEE Press, 2004: 18-26

[14] Nasipuri A, Castaneda R, Das S R. Performance of multipath routing for on-demand protocols in mobile ad hoc networks[J]. ACM/Kluwer Mobile Networks and Applications (MONET), 2001, 6(4): 339-349

[15] Marina M K, Das S R. On-demand multipath distance vector routing in ad hoc networks[C]// Proc. IEEE International Conference on Network Protocols (ICNP). Mission Inn, Riverside, California, Nov. 2001: 14-23

[16] NS-2[EB/OL]. http://www.isi.edu/nsnam/ns