

一种 DDoS 攻击复合式检测方法的研究

杨海亮,李震,马天丁,胡毅

(南京水利科学研究院,江苏南京 210029)

摘要:分布式拒绝服务攻击(DDoS)严重影响着网络安全,给网络的应用和发展带来了极大危害。目前,网络流量的自相似性、时间序列分析等已经成为 DDoS 攻击检测中重要的策略和技术。但是,当这些策略和技术单独使用时,DDoS 攻击检测效果并不十分理想。文中提出一种利用网络单边连接密度(OWCD)、网络重尾特性、累积欧几里得距离等方法的复合式检测方法。运用该复合式 DDoS 攻击检测方法进行 DDoS 攻击检测时,能有效地区分出正常流量、DDoS 攻击流量与突发业务流量,从而提高了 DDoS 攻击的检测效率。

关键词:分布式拒绝服务攻击;自相似性;重尾特性;突发业务流

中图分类号:TP393.08

文献标识码:A

文章编号:1673-629X(2015)07-0111-05

doi:10.3969/j.issn.1673-629X.2015.07.024

Research on a Hybrid DDoS Intrusion Detection Method

YANG Hai-liang, LI Zhen, MA Tian-ding, HU Yi

(Nanjing Hydraulic Research Institute, Nanjing 210029, China)

Abstract: Distributed Denials of Service (DDoS) attacks have done great harm to the application and the development of Internet. Currently, the self-similarity of network traffic and time series analysis have been the important strategies and technologies of DDoS attacks detection. But when these strategies and technologies are used individually, the results of DDoS detection are not ideal. A hybrid DDoS intrusion detection method by the OWCD, heavy-tail property and accumulated Euclidean distance is proposed. The result shows that applying the method to detect DDoS attacks, it could distinguish DDoS attacks traffic from normal traffic and burst traffic, to improve the detection rate.

Key words: DDoS; self-similarity; heavy-tail property; burst traffic

1 概述

分布式拒绝服务(Distributed Denial of Service, DDoS)攻击指借助于远程控制技术,发动多台服务器或计算机终端对一个或多个目标机器同时开展拒绝服务攻击,从而成倍地提高拒绝服务攻击的并发请求数量和带宽,以达到快速有效地使目标机器无法提供服务的不法目的。一般来说,攻击者通过漏洞、挂马、社会工程等手段将 DDoS 代理程序安装在大量计算机(俗称“肉鸡”)上,在一个设定的时间由部署在控制服务器上的主控程序向大量“肉鸡”的代理程序发送攻击指令,收到指令后“肉鸡”即开始对目标机器发动拒绝服务攻击。DDoS 攻击通常是不法分子为了攫取政治、经济利益而发动的,加上 DDoS 攻击技术要求不高,因此大量 DDoS 工具应运而生。2013 年 3 月起,国家互联网应急中心针对 rejoice2013、darkness、dark-

shell、imddos、风云、猎鹰、脑残片制造机、雪花、幽幽、残壳、毁灭等典型 DDoS 工具开展抽样监测。监测数据显示,在利用上述典型 DDoS 工具进行的分布式拒绝服务攻击事件中,控制服务器 IP 地址数量 11 650 个,被用于发起 DDoS 的“肉鸡”IP 地址数量超过 14 万个,遭受 DDoS 攻击的目标主机 IP 地址数量 1 619 个,总攻击次数达 6 823 次。

DDoS 攻击是我国互联网基础设施和信息系统正常运转的主要威胁,且呈日益严重的趋势。根据国家互联网应急中心抽样监测数据显示,我国平均每天发生攻击流量超过 1 Gbit/s 的攻击事件 1 802 起。DDoS 同时呈现出一些新的特点,主要表现在以下两个方面。

(1)攻击地址由原来的以虚假源地址为主逐渐转变为以真实地址为主。通过运营商和设备厂商的配合努力,虚假源地址流量攻击的效果越来越弱,攻击者构

收稿日期:2014-09-01

修回日期:2014-12-03

网络出版时间:2015-06-23

基金项目:中央级公益性科研院所基本科研业务费青年基金项目(Y910001)

作者简介:杨海亮(1981-),男,硕士研究生,工程师,研究方向为计算机工程、网络安全。

网络出版地址:<http://www.cnki.net/kcms/detail/61.1450.TP.20150623.1051.046.html>

造的虚假源地址流量无法达到攻击目标,因此不得不采用真实地址发起攻击。此处的真实地址有以下几种可能。一种是攻击者通过技术手段控制的“肉鸡”,即感染木马病毒的普通网民的终端机器。普通网民接入带宽逐渐增加,一旦被攻击者控制用作攻击“肉鸡”,将导致更大的危害。第二种是攻击者租用服务器。这些服务器一般性能较高,且位于具有高带宽的 IDC 机房内,所以单台服务器就能够发起高达 Gbit/s 级的攻击流量,危害极为严重。但是如果只使用 IDC 服务器发起真实地址攻击容易被攻击目标防护系统阻断或被相关部门发现,因此“肉夹馍”的攻击方式,即 IDC 真实地址服务器作为主要攻击流量来源,混杂大量“肉鸡”作为掩护成为攻击者发起攻击的首选方式。

2013 年国家互联网应急中心分析处置的典型攻击事件中,真实地址攻击占多数,且多数攻击流量都在数十 Gbit/s,经分析追溯,基本都属于“肉夹馍”的攻击方式^[1]。

(2)反射放大攻击的种类越来越多,攻击强度越来越大。反射攻击者不是直接发起对攻击目标的攻击,而是利用互联网的某些网络服务以及对应开放的服务器,间接对攻击目标发起攻击。通常这些网络服务使用的都是无连接的 UDP 协议,常见的反射放大攻击包括 DNS(Domain Name Service,域名系统)反射放大攻击以及 NTP(Network Time Protocol,网络时间协议)反射放大攻击。DNS 反射放大攻击是指攻击者伪造攻击目标地址向互联网上开放的大量递归服务器发起域名请求,这些递归服务器收到请求后,将会把应答包返回给伪造的攻击目标地址,而且应答包往往比请求包大数十甚至数百倍,从而利用这些递归服务器对攻击目标地址发起反射放大攻击。

2013 年 3 月,国际反垃圾邮件组织 Spamhaus 遭受攻击,攻击者借助现网庞大开放的 DNS 群,利用 DNS 反射技术,将攻击流量轻松放大约 100 倍,峰值达 300 Gbit/s。

NTP 反射放大攻击是指攻击者伪造攻击目标地址向互联网上开放的大量 NTP 服务器发起 monlist 请求,这些 NTP 服务器收到请求后,将会把应答包返回给伪造的攻击目标地址,应答包中包含与 NTP 服务器进行过时间同步的最后 600 个客户端的 IP 地址,比请求包大出数百倍。利用这些 NTP 服务器可对攻击目标地址发起反射放大攻击。美国 CDN 服务商 CloudFlare 声称遭受过高达 400 Gbit/s 流量的 NTP 反射放大攻击。由于反射放大攻击的成本较低,且放大效果显著、追溯困难,且利用的相关服务本身具有开放式的特点,因此难以有效地彻底杜绝,这将对 DDoS 攻击的防护带来更大的挑战^[2]。

(3)域名系统成为 DDoS 攻击的重点目标。域名系统是网民访问网站的入口,由于域名解析系统自身存在的弱点,包括 UDP 协议、固定端口等,针对域名系统的攻击已经成为黑客对网站发起拒绝服务攻击的主要方式,且这类攻击事件一旦得逞往往会导致大量无辜网站受牵连,从而造成严重后果。2013 年 8 月 25 日,黑客对我国的 .CN 国家顶级域名服务器发起大规模的拒绝服务攻击,导致我国大量 .CN 域名解析受到影响,我国大量政府网站、新浪微博等知名网站无法访问或访问缓慢。这也是我国近几年来首次出现的针对国家顶级域名服务器发起的攻击事件,对我国互联网基础设施的安全提出了更高要求。此外,2013 年 8 月 31 日和 9 月 4 日,爱视网(22.cn)域名服务器连续遭受数十 Gbit/s 的拒绝服务攻击,导致其负责解析的数万个域名受到影响。

正是由于 DDoS 分布式攻击的源头广泛,攻击地址真假结合,攻击组合多样,使 DDoS 这种危害性极大的攻击具有易实现、难检测、难追查和难抵挡的特点。DDoS 攻击的检测与防护已经成为一个世界性的问题。因此,有何有效、快速地检测 DDoS 攻击,并迅速对攻击做出有效防护具有很强的现实意义。文中将 DDoS 攻击分为低速攻击和泛洪攻击。其中,一定时间范围内低平均速率的 DDoS 攻击 RoQ(Reduce of Quality)与正常突发流量(如高考成绩查询时间、世界杯期间流量等)十分相似,难以区分,这成为区分正常流量和 DDoS 攻击的最大难题^[3]。

Yang G 等提出一种利用 TCP 包的大小所占比重和抽样测量的方法来检测低速率的 TCP DDoS 攻击;Shevetkar A 等提出一种在边界路由利用包转发速率的方法来检测 TCP Flow;在国内,任勋益和朱士瑞都提出了一种利用网络流量的自相似性来检测 DDoS 攻击的方法,在时间段内,对网络流量进行小波分析,利用自相似性对其进行分析,利用自相关系数 H 来判断网络是否存在 DDoS 攻击;许晓东等提出了一种利用网络流量的重尾特性的方法来检测 SYN 洪流攻击;徐图等提出了利用网络流量的单边连接密度的参数(方差、自相关系数等)来判断网络是否存在攻击,并对攻击强度进行了研究。国内外的大部分检测方法都是基于通过对网络流量的各种属性进行统计对 DDoS 攻击进行研究,虽然取得了一定的成果,但都存在一定的不足。例如:由于网络流量数据巨大,一些计算方法不能适应大数据环境,所以检测成功率不高,应用范围得不到推广;很难区分正常的突发流量和 RoQ 攻击流量。随着技术的发展,文中提出一种复合式检测方法对正常流量、突发流量、RoQ 攻击流量和泛洪攻击流量进行有效检测。

2 基本知识概念、研究

2.1 网络流 OWCD 时间序列研究

无论是泛洪 DDoS 攻击还是 RoQ 攻击,都是通过攻击手段包耗尽目标主机的网络带宽和系统资源,使其无法与目标机建立完整的双向通信。因此,用于直观反映网络流异常的单边连接密度(One Way Connection Density, OWCD)概念被提出并用于识别 DDoS 攻击。研究表明,正常流量下 OWCD 值较小,在攻击流量下 OWCD 流量值明显增加^[4]。

在时间序列下,对 OWCD 值进行分析。实验结果表明,稳定的正常流的 OWCD 序列近似为白噪声序列,其 OWCD 值在 $[0, 30]$ 的范围内,方差在 $[60, 220]$ 范围内波动,自相关系数很快趋于 0,功率谱密度在方差附近波动,其 OWCD 序列为稳定的序列。而攻击流中,泛洪 DDoS 攻击 OWCD 值在 $(60, 100)$ 范围之内,方差变小很快趋于 0, OWCD 序列的方差会减小,当攻击流淹没整个网络带宽时,方差趋近于 0,自相关系数较正常流有所增加但很快趋于 0,与正常流相似,功率谱密度仍在方差附近上下波动;RoQ 攻击和网络突发流量发生时,OWCD 值在 $[0, 100)$ 范围之内,其均值基本都在 $(30, 100)$,方差会变得很大,通常在 1 000 以上,自相关系数呈规则的周期性变化,而且并不趋近于 0,表现出其序列的非平稳性,功率谱密度极不稳定,出现很高的尖峰,说明功率集中某些点上,而不是像白噪声一样,功率在整个频带上平滑。

2.2 网络流量重尾分布研究

1997 年,波士顿大学的 Mark E. Crove Ua 和 Azer Bestavros 对网络流量进行了深刻的分析。研究发现在时间范围内,大部分的网络流量都存在自相似性的特点,正常的网络流量基本上都符合重尾分布的特性。而服从重尾分布的随机变量的特点是:大量的小抽样取值和少量的大抽样取值并存,在这些抽样数据集中,虽然大部分的抽样取值是小的,但是对抽样的均值和方差起决定作用的是那些少量的大抽样取值^[5]。

令 X 为一个随机变量,它具有累积分布函数(Cumulative Distribution Function, CDF) $F(x) = P[X \leq x]$ 和余累积分布函数(Complementary CDF, CCDF) $\bar{F}(x) = 1 - F(x) = P[X > x]$ 。定义 $F(x)$ 为重尾分布,如果 $\bar{F}(x) = cx^{-\alpha}$, $0 < \alpha < 2$, 其中 c 为正常数。在对数坐标系中生成该数据集的 CCDF 图,如果在一个大的范围内(几个数量级)是线性递减的,该数据集就符合重尾分布, $\lim_{x \rightarrow \infty} \frac{d \log \bar{F}(x)}{d \log x} = -\alpha$ 。对于大的 x 值,在对数坐标系下重尾分布的 CCDF 是一条斜率为 $-\alpha$ 的斜线, α 为重尾参数^[6-8]。

要把突发流量和攻击流量区分开来,可以用 Matlab 生成它们的 CCDF,然后求重尾参数的值,利用重尾参数的值来判断是否是攻击。当重尾参数 α 的值在 $(0, 2)$ 的时候,网络流量大小符合重尾分布,可以确定网络流为突发业务流;而当 α 的值大于 2 的时候,网络流量的大小不符合重尾分布,网络流量的重尾特性遭到破坏,网络发生攻击。

由于攻击不能模拟网络流量的重尾分布,而由突发业务流引起的突发流量符合重尾分布,所以通过网络流量的重尾特性来区分突发流量和攻击流量。由上述可知,当重尾参数 α 在 0 到 2 之间时,说明网络流量分布很好地符合重尾分布,是正常流量;当 α 大于 2 时,说明网络流量的重尾分布特性遭到破坏,流量为攻击流量。

图 1 和图 2 为两组流量的 CCDF 图,通过计算图 1 得到 $\alpha = 1.66$,由图 2 得到 $\alpha = 2.67$ 。根据突发流量的重尾参数在 0 到 2 之间,而攻击流量的重尾参数大于 2,所以可以很准确地把突发流量和攻击流量区分出来。

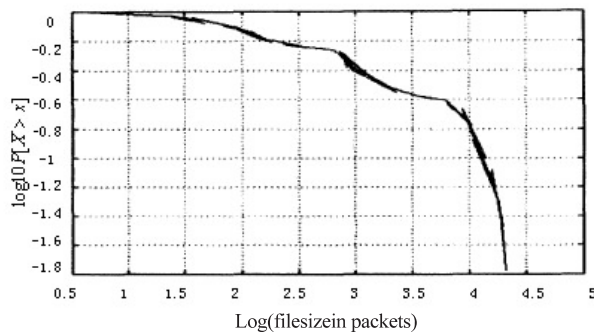


图 1 正常流量 CCDF

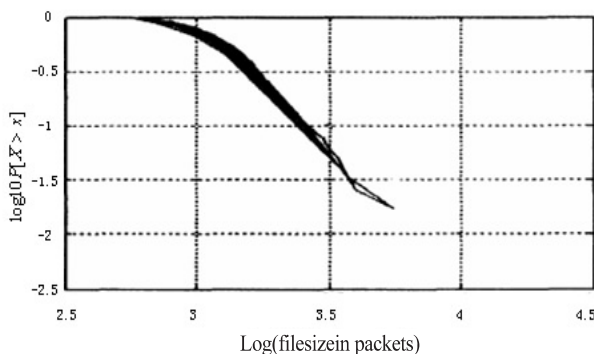


图 2 攻击流量 CCDF

2.3 累计欧几里得距离

Jin Shuyuan 和他的助手提出了运用距离计算来区别 DDoS 流,运用异常流偏离正常流的程度来确定流 DDoS 攻击和使用距离描述偏差程度和攻击强度。在此,把这个参数引用到文中的检测算法,以检测 DDoS 攻击的强度。在网络运行理想情况下,流量包都能正

确地进行双边握手,所有采样点的 OWCD 值均为 0,表示 $Y_n = (0, 0, \dots, 0)$, 需要计算 $Da(X_n, Y_n)$ 。其中, X_n 表示 OWCD 时间序列。现实情况是,在正常网络环境中,不是所有的流量包都能建立双边连接,例如,一方网络故障、网络拥塞、通信中断等等,还有可能存在计算 OWCD 出现的误差,使得计算得到的 Da 值不为 0。通过大量实验分析,得到一个确定的值 β , 当 $Da > \beta$ 时,可以确定网络发生攻击,并可根_据 Da 的值的范围确定攻击强度。文中只把通过 Da 的值来决定攻击强度这个模块运用到检测中去。

要计算出 X_n 和 Y_n 的距离有很多方法,其中累积欧几里得距离方法应用最为广泛。

设两个向量 X_n 和 Y_n , 累积欧几里得距离:

$$Da(X, Y) = \sqrt{\sum_{i=1}^n (\sum_{u=1}^i x_u - \sum_{u=1}^i y_u)^2}$$

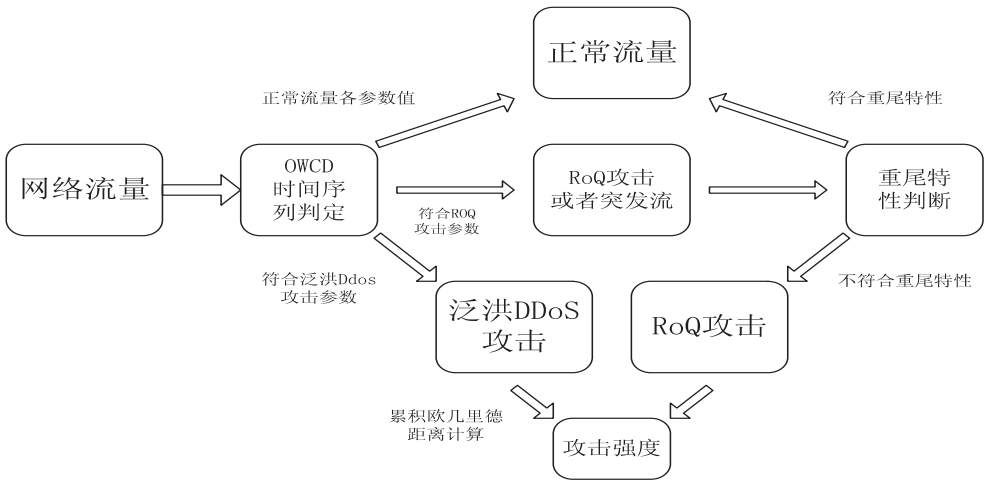


图 3 检测方法

通过对 OWCD 时间序列均值、方差、自相关系数、功率谱密度四个参数的计算,可以把正常流量和攻击流量中比较稳定的正常流量和泛洪 DDoS 攻击流区分出来,但当网络中并存正常突发流量和 RoQ 攻击流量时,由于突发流的 OWCD 序列的属性特征值跟 RoQ 攻击的有很大的相似性,所以,单独使用 OWCD 序列来判断是否发生 DDoS 攻击容易产生较高的误报率。由于正常的网络流量是符合重尾分布的,而 DDoS 攻击不能体现流量重尾特性,当网络发生 DDoS 攻击时,网络流量的重尾参数 α 大于 2。针对这种情况,可以对流量进行进一步的重尾特性测试。当重尾参数 α 小于 2 时,认为网络流量为正常突发流量,当 α 大于 2 时,认为网络发生 RoQ 攻击,还可以继续对网络流量的 OWCD 时间序列进行研究,对其进行累积欧几里得距离计算,得以求得攻击强度^[11-14]。

4 实验分析

为了验证检测方法的有效性,利用 MIT/LL 林肯

经过大量分析研究,当 Da 值在 $(600, 900)$ 时,为轻度攻击; $[900, 1\ 200]$ 时为中度攻击; $(1\ 200, 1\ 800)$ 为强度攻击;当 $Da \geq 1\ 800$ 时,网络资源耗尽。

3 复合式检测方法研究

根据研究发现,时间序列分析技术是检测泛洪 DDoS 攻击的较好策略,但是在区分突发流量和 RoQ 攻击流量上存在明显的缺陷^[9-10],而网络流量的重尾特性在区分突发流量和攻击流量上有明显优势。文中结合时间序列分析和网络流量重尾特性技术的优点,提出一种 DDoS 攻击复合式检测方法。为了更好地防范和解决 DDoS 攻击,通过计算 OWCD 时间序列的累积欧几里得距离来得到攻击强度。具体检测方法如图 3 所示。

实验室 DAPRA 工作组提供的两组 DDoS 攻击数据集作为基础研究数据,该数据集包括泛洪攻击、RoQ 攻击、正常流量和突发流量,适合文中检测方法的测试。表 1 为林肯实验室 DDoS 数据集。

表 1 实验数据集

Dataset	Start time	End time	File Sizes/kB
LLS-DDoS-1.0	2000-3-7 22:51	00:35	119 221
LLS-DDoS-2.0	2000-4-17 14:45	16:28	64 561

对 MIT 林肯实验室提供的正常流量数据进行每 5 min 的流量聚合,选取 50 段正常流量作为背景流量:第一组实验对这 50 段流量加入 DDoS 攻击(10 组泛洪攻击,10 组 RoQ 攻击);第二组实验对这 50 段流量加入 DDoS 攻击(10 组泛洪攻击,10 组 RoQ 攻击)和突发流量(10 组突发流量),并且用传统匹配方法(自相似性)和文中提出的复合式检测方法分别进行检测。

由表 2 可以看出,当网络流量中只存在 DDoS 攻击时,传统的匹配方法和文中提出的方法都能很好地

检测出 DDoS 攻击。再对流量中存在 DDoS 攻击流量和突发流量的 50 段流量进行检测,检测情况如表 3 所示。

表 2 加入 DDoS 流量攻击检测性能表

检测方法	检测异常	实际攻击	正确检测	错误检测	检测率/%	误报率/%
传统匹配方法	21	20	17	4	85	19
文中方法	20	20	18	2	90	10

表 3 加入突发流量和 DDoS 攻击性能对比表

检测方法	检测异常	实际攻击	正确检测	错误检测	检测率/%	误报率/%
传统匹配方法	28	20	17	11	85	39.3
文中方法	21	20	18	3	88	11.3

由表 3 可以看出,当网络中存在突发流量和 DDoS 攻击时,相对于传统的匹配方法,文中提出的复合式 DDoS 攻击检测方法能更好、更有效地检测出 DDoS 攻击,误报率更低。

5 结束语

文中提出的复合式 DDoS 攻击检测方法是对网络流量数据进行离线操作分析,如果把该方法应用到实际的网络环境中,在检测速率、检测效率方面有待进一步实验测试和提高;另外,还可把基于重尾特性的检测方法应用到网络流量的其他属性,例如包的持续时间、流速、包的大小等,对 DDoS 攻击进行更加有效的检测。等技术成熟还可应用到对其他网络攻击的检测中去,这将有更大的实际意义。

参考文献:

[1] 徐图,何大可. 网络流单边连接密度的时间序列分析

(上接第 110 页)

analysis[J]. Proceedings of the IEEE,1992,80(7):1079-1092.

[7] Gasser A, Hazem R. An automatic text reader using neural networks[C]//Proceedings of the Canadian conference on electrical and computer engineering. [s. l.]:[s. n.],1993:92-95.

[8] Congedo G,Dimauro G,Impedovo S,et al. Segmentation of numeric strings[C]//Proceedings of the third international conference on document analysis and recognition. [s. l.]:[s. n.],1995:1038-1041.

[9] 张闯,蔺志青,肖波,等. 适用于银行票据手写数字串切分的滴水算法[J]. 北京邮电大学学报,2006,29(1):13-

[J]. 四川大学学报:工程科学版,2007,39(3):136-140.

[2] 程光,龚俭,丁伟. 基于抽样测量的高速网络实时异常检测模型[J]. 软件学报,2003,14(3):594-599.

[3] 程光,龚俭,丁伟. 网络测量及行为学研究综述[J]. 计算机工程与应用,2004,40(27):1-8.

[4] 许晓东,杨海亮,朱士瑞. 基于重尾特性的 SYN 洪流检测方法[J]. 计算机工程,2008,34(22):179-181.

[5] 巩永旺. 基于扫描流量统计的本地网蠕虫检测方法[J]. 计算机技术与发展,2011,21(7):145-148.

[6] 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.

[7] 孙知信,李清东. 基于源目的 IP 地址对数据库的防范 DDoS 攻击策略[J]. 软件学报,2007,18(10):2613-2623.

[8] 孙知信,唐益慰,程媛. 基于改进 CUSUM 算法的路由器异常流量检测[J]. 软件学报,2005,16(12):2117-2123.

[9] 母军臣,甘志华,许宏云. 基于动态包过滤的 RoQ 攻击防御策略[J]. 电脑知识与技术,2007,1(6):1532-1533.

[10] 谢逸,余顺新. 新网络环境下应用层 DDoS 攻击的剖析与防御[J]. 电信科学,2007,23(1):89-93.

[11] Yang G, Gerla M, Sanadidi M Y. Defense against low rate TCP attacks: dynamic detection and protection[C]//Proceedings of network04. New York, NY, USA: ACM,2004:189-198.

[12] Shevetkar A, Anantharam K, Ansari N. Low rate TCP denial-of-service attack detection at edge routers[J]. IEEE Communications Letters,2005,9(4):363-365.

[13] Guirguis M, Bestavros A, Matta I. Exploiting the transients of adaptation for RoQ attacks on internet resources[C]//Proc of the 12th IEEE international conference on network protocols. [s. l.]:IEEE,2004:184-195.

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

16.

[10] 李小园,田刚,封超. 印刷公式中粘连字符的切分[J]. 科学技术与工程,2011,11(3):628-632.

[11] 刘赛,王江晴,张振绘. 一种用于脱机手写体女书字符切分的方法[J]. 计算机应用研究,2011,28(3):1188-1190.

[12] 何耘嫻. 印刷体文档图像的中文字符识别[D]. 秦皇岛:燕山大学,2011.

[13] 马瑞. 非限制手写字字符分割中相关技术与算法的研究[D]. 南京:南京理工大学,2007.

[14] 李兴国,高炜. 基于滴水算法的验证码中粘连字符分割方法[J]. 计算机工程与应用,2014,50(1):163-166.

[15] 靳简明,丁晓青,彭良瑞,等. 印刷维吾尔文本本切割[J]. 中文信息学报,2005,18(5):76-83.

一种DDoS攻击复合式检测方法的研究

作者：[杨海亮](#)，[李震](#)，[马天丁](#)，[胡毅](#)，[YANG Hai-liang](#)，[LI Zhen](#)，[MA Tian-ding](#)，[HU Yi](#)
作者单位：[南京水利科学研究院, 江苏 南京, 210029](#)
刊名：[计算机技术与发展](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：2015(7)

引用本文格式：[杨海亮](#). [李震](#). [马天丁](#). [胡毅](#). [YANG Hai-liang](#). [LI Zhen](#). [MA Tian-ding](#). [HU Yi](#) [一种DDoS攻击复合式检测方法的研究](#)[期刊论文]-[计算机技术与发展](#) 2015(7)