文章编号:1001 - 9081(2006)06 - 1343 - 03

CUSUM 算法在 DDoS 源端检测中的应用

康 健,鞠九滨
(吉林大学 计算机科学与技术学院,吉林 长春 130012)
(kangjian@jlu.edu.cn)

摘 要:在深入分析了 DDoS 源端检测的特点和难点的基础上,引入统计学中非参数改变点检测方法,应用非参数化递归 CUSUM(Cumulative Sum)算法对代表性的源端检测系统 D-WARD 进行了改进。经实验验证,应用 CUSUM 算法的检测系统具有更低的误报率和漏报率,能够适应更复杂的网络检测环境。

关键词:分布式拒绝服务;非参数 CUSUM;D-WARD;源端检测
中图分类号: TP393.08 **文献标识码**:A

Detecting DDoS attacks in source-end network with CUSUM algorithm

KANG Jian, JU Jiu-bin
(*Department of Computer Science & Technology, Jilin University, Changchun Jilin* 130012, *China*)

Abstract: After characteristics and difficult problems in detecting DDoS attacks were analyzed in source-end network, a nonparametric change point detection method in statistics was introduced and D-WARD system, a representative source-end DDoS detection system, was improved with nonparametric recursive CUSUM algorithm. Experiments prove that the improved system is lower in false-positive rate and false-negative rate, which is more accurate and could adapt to more complex network environments.

Key words: DDoS; nonparametric CUSUM; D-WARD; source-end detection

0 引言

随着 Internet 网络服务的广泛应用,网络安全事件频繁发生。其中,分布式拒绝服务(DDoS)攻击已经成为 Internet 网络中最直接有效的攻击方式之一。DDoS 攻击就是利用众多的计算机对一个或多个目标进行有组织的协调的 DOS 攻击,使用客户/服务器技术,攻击者能够调动相当数量的不知情的僵尸主机构成攻击网络,产生远远高于 DOS 攻击成百上千倍的破坏力[1]。DDoS 攻击具有很强的隐蔽性和突发性,使其难于防范。如何有效检测和防御 DDoS 攻击是当前网络安全领域的研究热点之一。

按检测点位置,DDoS 攻击检测可以分为配置在受害端的检测,配置在中间网络(一般是 Internet 核心路由器)的检测和配置在攻击源端的检测三种方式。配置在攻击源端的检测方法具有如下优点:

1) 能够在攻击数据流进入 Internet 网络并在瓶颈处造成拥塞前将其终止。

2) 与受害端的检测相比,更容易追溯到攻击源。

3) 源端网络出口路由器比 Internet 上工作繁忙的核心路由器能够提供更多的用于检测的资源。

源端检测也存在固有的难点。首先,攻击数据流在源端网络中分散并且流量较小,通常的检测方法存在较高的漏报率,因此使用更敏感更精确的检测算法是源端检测系统需要解决的核心问题。其次,源端网络分布广泛,数量众多,只有在广泛配置源端检测系统的情况下,才能对整个 Internet 网络安全产生积极影响。最后,由于攻击流对源端网络通常影响较小,源端网络的管理者往往没有很强的防御 DDoS 攻击的意识。

在 DDoS 攻击的源端检测中,具有代表性的是 D-WARD 系统[2,3]。原始 D-WARD 核心的检测算法具有一定的局限性。对 TCP 和 ICMP 攻击流,它只采用简单的固定阈值方法检测不断变化的网络环境缺乏灵活性和适应性,因而其误报率和漏报率较高。

针对以上缺点和难点,本文引入统计学中非参数改变点检测方法,应用非参数化递归 CUSUM 算法[4,5]改进 D-WARD 系统,并使用自适应的阈值调整算法判定攻击,取得了较好的效果,有效地降低了原系统的误报率和漏报率,改进后的系统能够适应更复杂的网络检测环境。

1 D-WARD 检测算法的局限性

以 TCP 类攻击流的检测为例,说明 D-WARD 系统的检测算法。

TCP 类攻击流检测:在正常的 TCP 会话过程中,源主机到目的主机的数据流(TCP_sent_to)是受到反方向上的确认流(TCP_received_from)控制的。目前,主流的 DDoS 攻击工具都可以发动 TCP 类洪泛攻击。攻击发生时,TCP_sent_to > > TCP_received_from。D-WARD 中定义了 TCP_sent_to/TCP_received_from 在正常情况下的最大允许值 max_tcprto,若在实时监测过程中观测到该比值超过 max_tcprto 则判定为攻击。

从分析中可以看出,D-WARD 系统只采用简单的阈值检测方法,没有考虑 DDoS 攻击流的持续变化和攻击流在时间上的累积效应。同时,使用固定的阈值方法检测不断变化的

收稿日期:2005 - 12 - 05;**修订日期**:2006 - 02 - 20 **基金项目**:国家自然科学基金资助项目(90204014)

作者简介:康健(1975 -),男,内蒙古通辽人,讲师,博士研究生,主要研究方向:分布系统和网络安全; 鞠九滨(1935 -),男,黑龙江哈尔滨人,教授,博士导师,主要研究方向:分布系统和网络安全.

网络环境缺乏灵活性和适应性,因而其误报率和漏报率较高。

2 应用 CUSUM 算法提高检测精度

2.1 CUSUM 算法

Internet 网络传输可以看作一个复杂的随机模型,任何传输的异常(比如 DDoS 攻击)都将导致模型的急剧变化。本文的目的就是要检测 TCP_sent_to 与 TCP_received_from 比值的异常改变。目前,有两种方法对这种改变进行检测:1)等长批量检测法,它通过监测在单位时间内被观测量变化的平均值实现检测;2)连续改变点检测方法,它监视的是变量的连续变化情况。其目的是确定观测的时间序列是否符合统计分布,如果不是,找到发生改变的时间点。这种方法[4,5]具有在线实时检测的能力,符合 DDoS 检测的要求。

使用该方法的一个难点是如何确定观测序列的整体分布。到目前为止,关于 Internet 网络整体流量分布规律的研究已经非常广泛。文献[6]中详细讨论了过去几年,Internet 网络流量统计分布的变化过程。事实上,通过一个简单的变量模型无法模拟整个网络流量的统计分布。因此,只能采用一种非确定模型的检测方法。非参数化方法非常适合解决这类问题。文献[4][5]证明对于广域改变点检测问题,CUSUM(Cumulative Sum)算法是接近最优的非参数化方法。文献[9]证明了 CUSUM 算法具有很多其他序列和非参数检测算法中的优点,适合应用在网络异常检测方面。同时,文献[9]也指出该算法计算量很小,能够完全满足实时检测的要求。本文将在此基础上,采用 CUSUM 算法改进 D-WARD 检测系统。

2.2 实现

下面以检测 TCP 类攻击流为例,说明 CUSUM 算法的具体实现。

设 TCP_sent_to 是源端网络中主机到目的主机的数据流,TCP_received_from 是 TCP_sent_to 反方向上的确认流。序列 $\{delta[n], n = 0,1,2,\cdots\}$ 是单位取样时间内 TCP_sent_to 与 TCP_received_from 的差值(式 1)。

$$delta[n] = TCP\_sent\_to[n] - TCP\_received\_from[n], \quad (n = 0,1,\cdots) \tag{1}$$

通常,$delta[n]$ 与网络的规模、主机数量以及取样的时间区间有密切的关系。为使算法更有通用性,减少上述因素的影响,进行下面的变换:

$$smoothed\_fn[0] = 0;$$

$$smoothed\_fn[n] = \alpha * smoothed\_fn[n-1] + (1-\alpha) * TCP\_received\_from[n], (n = 1,2,\cdots)$$

$$X[n] = delta[n]/smoothed\_fn[N] \tag{2}$$

从上面的变换可以看出,$\{delta[n], n = 0,1,2,\cdots\}$ 序列可以用 TCP_received_from 的递归方式来表示,即 $smoothed\_fn[n]$ 变量可以实时的计算和更新。其中 n 指定检测时间序列的序号,α 是自定义常量,$\alpha \in [0\cdots1]$。进行式(2)的变换,$X[n]$ 不再依赖于网络规模和时间区间的变化,而只与当前的数据包传输状态有关。因此可以将 $\{X[n] \mid n = 0,1\cdots\}$ 视为一组稳定的独立的随机过程。

在网络无攻击情况下,计算 $X[n]$ 期望的最大值,记为 max_avg_X,并取:

$$X2[n] = X[n] - \max\_avg\_X \tag{3}$$

即 $\{X2[n] \mid n = 0,1\cdots\}$ 序列在无攻击正常情况下,其值都应为负值。图 1(a) 是 $X2[n]$ 的曲线描述。

为判断是否发生 DDoS 攻击,定义函数:

$$y[n] = (y[n-1] + X2[n])^+$$

$$y[0] = 0$$

其中当 $y > 0$ 时 $y^+ = y$,当 $y \leqslant 0$ 时取 $y^+ = 0$。$y[n]$ 就是 TCP 类洪泛攻击时,算法处理的检测序列。图 1(b) 是 $Y[n]$ 的曲线描述。

从图 1(a) 中看到,在网络无攻击情况下,由观测序列 $X[n]$ 变换得到的序列 $X2[n]$ 其均值小于 0,其对应的图 1(b) 中 $y[n]$ 也趋于 0。当发生攻击时,$X2[n]$ 序列突然增大,均值大于 0,从图 1(b) 中可以观察到 $y[n]$ 的相应变化。

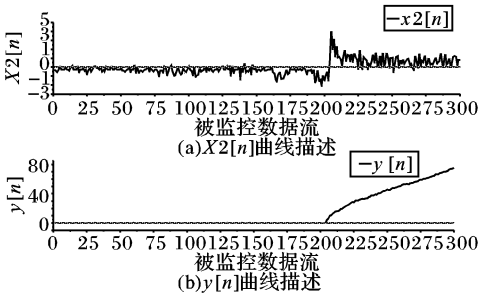

图 1 $X2[n]$ 和 $y[n]$ 曲线描述

2.3 Threshold 的自适应调整

原有的 D-WARD 系统采用简单的固定阈值判断攻击是否发生,无法根据具体情况进行调整,因此误报率和漏报率较高。本文采用文献[9]中提出的阈值确定方法,能够自适应调整 Threshold,提高了检测精度。

设第 n 个检测序列是 $y[n]$,此时对应的判定阈值是 $Threshold[n]$。通过下面的式(5)计算得到 $Threshold[n]$。

$$Threshold[n] = (\alpha + 1)\bar{\mu}_{n-1} \tag{5}$$

其中,$\alpha \in [0,1]$,$\bar{\mu}_{n-1}$ 是前 $n-1$ 个检测序列 $y[i]$($i \in [0,1,\cdots,n-1]$)使用指数加权平均(EWMA)算法得到的均值。

如果 $y[n] > Threshold[n]$,则判定为攻击发生。可见,判定阈值 Threshold 能够根据具体的网络环境自适应调整,改进后的系统比较有效地降低了误报率和漏报率。

3 性能评价

3.1 实验环境配置

实验环境由两个子网组成:10.60.26.* 和 10.60.46.*。取 26 网段作为攻击的源端网络,46 网段为外网,并取 46 段中的某一台主机为被攻击主机。在源端网络中配置了应用 CUSUM 算法的新系统。在源端网络中,选择若干主机配置典型的 DDoS 攻击工具 TFN,进行真实的 DDoS 攻击实验。

3.2 误报率和漏报率比较实验

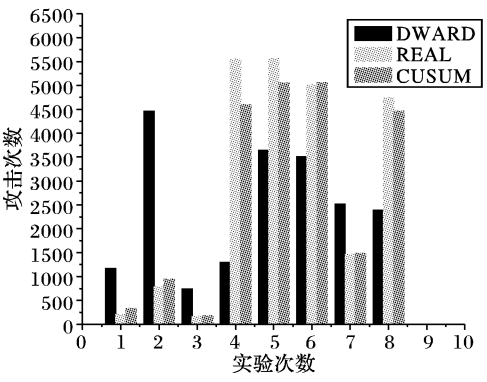

图 2 改进前后系统检测性能比较

在实验 1 中,对原 D-WARD 系统和应用 CUSUM 算法改进的系统在不同的网络环境下,比较误报率和漏报率的情况。

实验过程中,前三组数据是变化协议组合和流量强度,但不发送攻击数据流而得到的,而后五次取样则是在含 DDoS 攻击的网络环境下进行的。实验结果如图 2 所示。

从图 2 可以看出,原 D-WARD 系统在实验样本 1、2、3、7 中误报比较严重。特别是在样本 2 中,原 D-WARD 系统的误报数是真实攻击数的 6.45 倍。而在样本 4 中,原 D-WARD 系统又存在较大的漏报率,其漏报数占总攻击数目的 77.23%,而应用 CUSUM 算法改进的系统,其检测结果和真实的攻击曲线相接近,效果较好。

3.3　定量检测比较实验

实验 2 从定量的角度比较两种检测算法的检测性能。实验 2 中,在 DDoS 攻击的真实网络环境下抽取五次样本。实验过程中,对 DDoS 攻击数据流的强度进行变换,任何两个样本之间攻击强度都存在很大差异。实验结果得出两个系统各自检测到的针对受害主机的攻击数据流占源端网络与该主机之间被监控的总通信量的百分比,并与真实的攻击流量占总通信量的百分比进行比较。三者之间的关系见图 3。

从图 3 中可以看出,原 D-WARD 算法的检测曲线和真实情况相差较大 ,其中在样本 3 中,原 D-WARD 系统误报数是真实攻击数目的 1.87 倍,而在样本 1 中,其漏报数占真实总攻击数目的 49.72%,造成了较高的误报率和漏报率。应用 CUSUM 算法的系统得到的检测曲线和真实情况比较接近,能够较真实的反映出 DDoS 攻击情况。

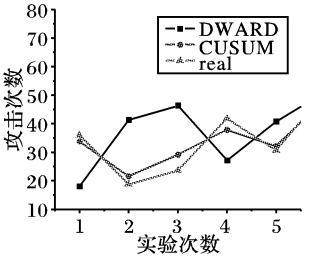

图 3　攻击流量占总流量的百分比

4　同类工作

MULTOPS[7] 系统根据需要可以配置在攻击的源端和受害端。通过监测本网与外网之间的流量特征,MULTOPS 系统判断 DDoS 攻击。它使用树形结构的收缩和扩张来确定攻击的发生和位置。与本文改进的 D-WARD 系统相比,MULTOPS 没有考虑不同协议(TCP,UDP,ICMP 等)的数据流的特征差异,即网络中出现的 TCP 正常数据流流量较大,而 MULTOPS 阈值选取相对较小时则会判断为攻击,发生误报;反之,若阈值设置过大则发生漏报。在本文应用CUSUM算法的系统中考虑到了逆向反馈流的影响,同时把在某时间点以前连续变化的若干监测变量的累积效应也考虑进来,能够有效降低系统的误报率和漏报率。

文献[8]中使用 CUSUM 算法对单位时间内数据流中出现的新的源 IP 地址数目进行监测,把新的 IP 数目的异常增大作为判定发生攻击的依据。但只使用单一的 IP 数目增加作为攻击发生的判定条件,没有将网络的综合变化情况考虑进来,使这种方法存在较大的误报率。

5　结语

本文针对 DDoS 攻击源端检测的特点和难点,引入统计学中非参数改变点检测方法,应用非参数化递归 CUSUM 算法对 D-WARD 系统进行了改进,并在实际网络环境中使用真实 DDoS 攻击工具进行了多次实验。实验结果表明,在源端网络的出口路由器上配置应用 CUSUM 算法的新系统能比较有效地检测和早期遏制 DDoS 攻击,适应更复杂的网络检测环境。

参考文献:

[1] STEIN L, STUART J. The World Wide Web Security FAQ[EB/OL]. http://www.w3.org/security/faq, February 4, 2002.

[2] MIRKOVIC J. D-WARD: DDoS Network Attack Recognition and Defense[D]. CSD of UCLA, 2002,

[3] MIRKOVIC J. D-WARD: Source-End Defense Against Distributed Denial-of-Service Attacks[M]. CSD of UCLA, 2003. 101 - 125.

[4] BASSEVILLE M, NIKIFOROV IV. Detection of Abrupt Changes: Theory and Application[M]. Prentice Hall, 1993.

[5] BRODSKY BE , DARKHOVSKY BS . Nonparametric Methods in Change point Problems[M]. Kluwer Academic Publishers, 1993.

[6] FELDMANN A. Characteristics of TCP Connection Arrivals[R]. ATT Technical Report, 1998.

[7] GIL TM, POLETTO M. Multops: a data-structure for bandwidth attack detection[A]. Proceedings of the 10th USENIX Security Symposium[C], 2001.

[8] TAKADA HH, HOFMANN U. Application and Analyses of Cumulative Sum to Detect Highly Distributed Denial of Service Attacks using Different Attack Traffic Patterns [EB/OL]. http://www.ist-intermon.org/dissemination/newsletter7.pdf, 2004.

[9] SIRIS VA, PAPAGALOU F. Application of anomaly detection algorithms for detecting SYN flooding attacks[A]. Proc. of the Conf. on Global Telecommunications [C], 2004.

(上接第 1342 页)

虑到人类视觉系统(HVS) 的特性,实现了水印的自适应嵌入。另外,水印的提取与检测只依赖与一组密钥,不需要其他附加条件,是一种水印的盲提取方案。

可以看到,该算法相对于传统算法避免了存储大信息量的嵌入轨迹矩阵,因此今后的工作将是在此基础上深入讨论如何增加信息嵌入量而不降低算法的隐蔽性,从而可以很好的实现大容量的信息隐藏。

参考文献:

[1] TEFAS A, PITAS I. Image Authentication Using Chaotic Mixing Systems [J]. IEEE International Symposium On Circuit Systems, 2000, I(5):216 -219.

[2] SEKERIDOU TS, SOLACHIDIS V, KOLAIDIS NN, *et al.* Statistical Analysis Of A Wate-marking System Based on BernoulliChaotic Sequences[J]. Signal Processing, 2001, 81: 1273 - 1293.

[3] MASUD N, AIHARA K. Cryptosystems with Discretized Chaotic Maps[J]. IEEE TransCircuit and Systems, 2002, 49(1) :28 - 40.

[4] 丁玮, 闫伟齐, 齐东旭. 基于 Arnold 变换的数字图像置乱技术[J]. 计算机辅助设计与图形学学报, 2001, 13(4):338 - 341.

[5] 张志明, 王磊, 郑应平. 一种基于混沌序列的时空域数字水印算法[J] . 计算机应用研究, 2003, (4):52 - 54.

[6] 吴崇明, 王晓丹. 数字水印系统的鲁棒性和常见的攻击[J]. 空军工程大学学报(自然科学版), 2002, 3(1):90 - 93.

[7] 刘彤, 裘正定. 数字水印相关检测的可靠性研究[J]. 电子学报, 2002, 30(5):658 - 688.

[8] 王朔中, 张新鹏, 张开文. 数字密写与密写分析[M]. 北京: 清华大学出版社, 2005.