关于构造攻击时 IP 地址选取和时戳处理的一种算法

——采用真实流量做背景时

杭强[1,2]，龚俭[1,2]，杨望[1,2]

（1.计算机科学与工程学院，东南大学，南京，211189；2.江苏省计算机网络技术重点实验室，东南大学，南京，211189）

摘 要：入侵检测系统是网络安全的一个主要产品之一,因此研究入侵检测系统的评估有着重要的意义.本文分析了评估测试数据集中背景流量由人工生成时，对攻击流量的地址选取和时戳处理的方法，但是当采用真实流量做背景流量时此方法并不适用，因此提出了一种新的算法，新算法根据真实流量来选取 IP 地址，根据不同的攻击类型分情况处理时戳。

关键字：入侵检测评估；真实流量；地址选取；时戳处理

An algorithm of constructing the attacks at the IP address selecting and timestamp processing

——when the real-world traffic to be background traffic

Hang Qiang[1,2]，Gong Jian[1,2]，Yang Wang[1,2]

(1.School of Computer Science & Engineering, Southeast University, Nanjing, 211189;

2. Jiangsu Provincial Key Laboratory of Computer Network Technology, Southeast University, Nanjing, 211189)

Abstract: Intrusion detection system is a major network security product, so it's an important significance to research the evaluating intrusion detection systems. This paper analyzes the test data how to select the address and process timestamp of the attack traffic, when the background traffic was artificially generated. However, it's not suitable for the situation that the background traffic was the real-world traffic .So a new algorithm was proposed .The new algorithm based on real traffic to select the IP address and processed timestamp under different types of attack.

Keyword: Intrusion Detection;Real-world traffic;IP address selecting;timestamp processing

1 引言

入侵检测系统评估是入侵检测系统领域的一个基本问题。目前已有的 IDS 测试中 MIT 的林肯实验室在 DARPA[1]资助下所开展的 IDS 评估工作是最为领先和完整的。但由于其使用真实的机器代表网络节点，所以只适用于小型的园区网，不能模拟骨干节点的拓扑结构；而驱动应用程序的方法，受制于机器和操作系统性能的限制，无法产生高速的流量。AOLES[2][3][4](Adaptable Offline Evaluation System，自适应离线评估系统)是基于系统能力设计的用于评估基于网络滥用入侵检测系统的测试数据集生成系统,但是其背景数据的语义真实性不足，对于测试有协议识别功能的 IDS 会产生影响。因此本论文的评估系统采用真实流量做背景数据来克服高速流量和语义真实性

作者简介：杭强，（1987-），男，硕士研究生，E-mail: qhang@njnet.edu.cn；龚俭，（1957-），男，教授，博导，E-mail: jgong@njnet.edu.cn；杨望，（1979-）男，讲师，E-mail: wyang@njnet.edu.cn.

问题，

AOLES 对于 IP 地址的选取是根据攻击规则来决定的，若规则中规定了源 IP 或者目的 IP 的范围，则从规定范围中随机的选取一个；如果没规定，则随机的从 0.0.0.0 到 255.255.255.255 的范围内选取一个。其对时戳的处理是根据测试脚本中攻击发生的时间范围随机选取一个时间点作为攻击流中第一个报文的时戳，后面的报文时戳相对于第一个报文时戳一次递增。但是这种方法不适合背景流量使用真实流量的情况。本文讨论的评估系统采用长城宽带在广东省网边界采集下来的真实流量作为背景数据，对原 AOLES 系统的 IP 地址选取和时戳处理算法重新设计。

本文共分六节，第二节主要介绍地址选取算法，第三节介绍时戳处理算法，第四节理论分析新的 IP 地址选取算法，第五节通过实验分析新的时戳处理算法，第六节进行了总结。

2 地址选取算法

AOLES 系统随机生成出的 IP 地址会生成一些根本不会出现在真实网络环境中被路由器采集下来的地址，如私有地址，环回测试地址以及网络地址全 0 或者主机地址全 1 的地址等。因此在构造攻击时的 IP 选取应该依托真实流量来选择。

本系统使用的真实流量来自于边界网的路由器，源宿 IP 地址一个在网内、一个在网外。该真实流量在正交处理（筛选出其中攻击流量）之后，其流量特性可以认为是正常的用户相互交互产生的流量。那么统计出源宿 IP 中属于网外的 IP，对其出现的次数进行排序，出现次数多则是交互频率高的 IP 地址，这样在构造攻击流量时尽量选取这些 IP 地址，所生成出的攻击流量就会混合在这些高交互频率的流中。

这种方法生成出的流量和 AOLES 生成的流量在测试 IDS 后期的关联性分析产生影响。算法流程见图 1 和图 2：

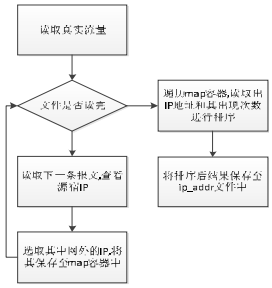

图 1 IP 统计算法流程图

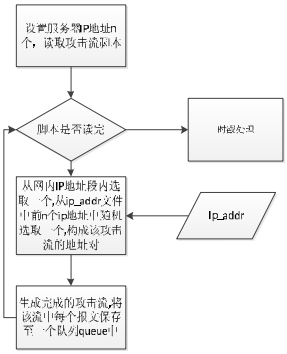

图 2 IP 选取算法流程图

3 时戳处理算法

测试数据集的时戳主要有两点作用：一是用于描述攻击的时间特性，不同类型攻击的时间特性是不同的，如 DOS 攻击一般集中在一个很小的时间段内产生大量报文，其时间间隔很短；SCAN 攻击一般根据其扫描策略的不同,有可能时间间隔是相同的。二是用于播放控制，以时戳为参考来调整各报文之间的时间间隔来满足不同测试场景的压力测试需求。

因此 AOLES 的完全随机生成算法就不合理了，其无法满足攻击的时间特性。新算法则根据攻击类型的不同对其时间分别处理：DOS 攻击根据其攻击强度算出其流内报文时间间隔 Δ；SCAN 攻击根据其策略算出流内报文时间间隔 Δ——快速扫描或者慢速扫描使用平均时间间隔，随机扫描使用随机函数依次生成下个报文的时间间隔；交互式攻击的流内报文时间间隔通过符合泊松过程的函数发生器产生，文献[5][6]都提出了流到达速率可以采用平稳的泊松过程进行模拟。

首先需要计算泊松过程的函数发生器的参数 λ，λ 表示在单位时间内随机时间的发生率，在此则表示单位时间内攻击报文的个数。由于 DOS 攻击和 SCAN 攻击其流内报文间隔的处理有特殊性，所有在计算 λ 时应将 DOS 攻击和 SCAN 攻击看成单报文。得到公式：

N（总报文数）=N（DOS 攻击流个数）+N（SCAN 攻击流个数）+N（交互式攻击流中报文总数）

λ=N（总报文数）/T（攻击持续时间）

当攻击流的 IP 地址填完之后会构成一个攻击队列，队首元素表示该攻击流中的第一个攻击报文，

队尾元素表示该攻击流中的最后一个攻击报文。构造一个最小堆，堆内元素为一个结构体，存储的是DOS攻击和SCAN攻击的队列序号以及该队列的队首元素表示的攻击报文的时戳。

算法流程图见图3：

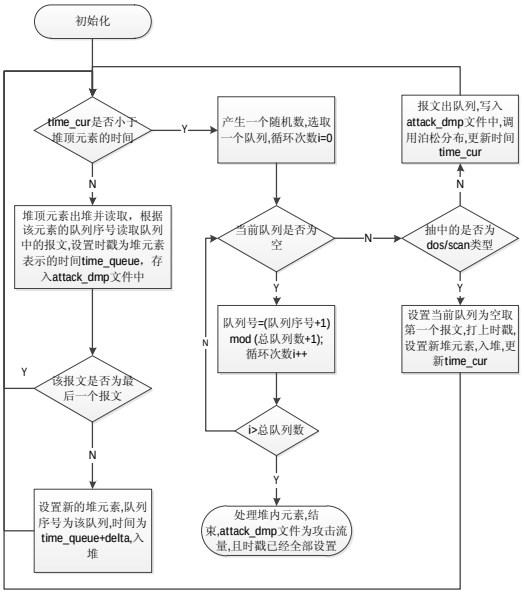

图3 时戳处理算法流程图

4 IP地址的理论分析

正常的互联网上如果出现以下 IP 地址属于非正常现象：

表1 非法IP地址统计

类型	总数
回环地址	2^{24}
D类地址	$16*2^{24}$
E类地址	$8*2^{24}$
E类之后的地址	$8*2^{24}$
私有地址	$2^{24}+16*2^{16}+2^{16}$
广播地址	$(2^{8}-2)+2^{15}+2^{22}$
主机地址	$2^{24}+2^{16}+2^{8}$

根据上表统计,如果采用完全随机法来生成 IP 地址，则出现非法 IP 的概率为 13.80%。这个概率是无法接受的，所以本文所采用的依托真实数据更为合理。

5 时戳处理的实验分析

根据本文所描述的算法设计了一个时戳处理程序，用 AOLES 程序和本程序各生成出一套攻击数据，数据的描述如下：包含 5 个 DOS 攻击，攻击强度为 100 个/秒，每次攻击持续时间为 5 秒；2 个 SCAN 攻击，攻击强度为 100 个/秒，每次攻击持续为 5 秒；包含 2500 个交互式攻击，总持续时间为 1 分钟。

生成的结果以图片显示如下：

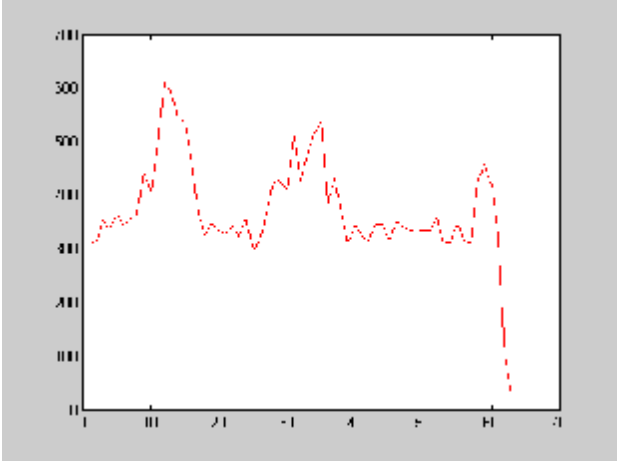

图4 本论文升序生成的数据的时间特性

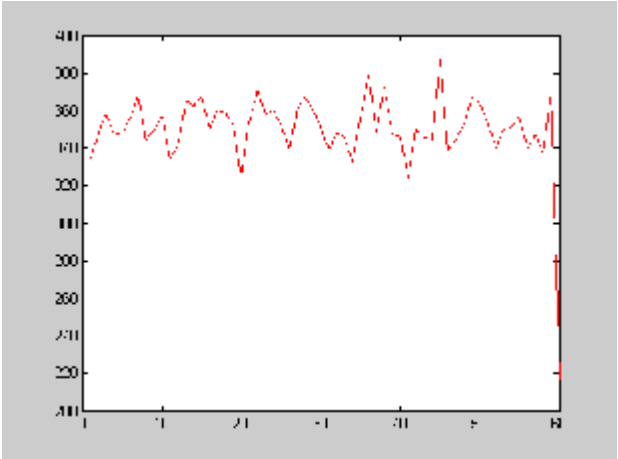

图5 AOLES 程序生成的数据的时间特性

其中横坐标表示时间（以 1s 为单位）,纵坐标表示每 1 秒内出现的报文数.

从图 5 可以看出采用原 AOLES 方法生成出的数据，其每秒报文数(pps)在 320 到 380 之间波动，相对图 4 很平滑。而图 4 正常情况下 pps 在 340 左右,但是图中出现了几个波峰,经过分析报文内容知道这些时刻发生了 DOS 攻击和 SCAN 攻击，导致 pps 陡增。这也正符合 DOS 攻击和 SCAN 攻击的时

间特性。所以本算法相对原 AOLES 系统对于时戳的处理的算法更合理。

6 总结

入侵检测系统在如今的互联网中越来越重要，同时如何去评价一个入侵检测系统的性能也成为了现在科研领域的重要问题。本论文针对测试数据集的背景数据采用真实数据时，对 IP 地址和时戳提出了一种新的算法，依托真实流量中的 IP 地址来选取攻击流量的 IP，有效的避免了攻击数据中出现非法 IP；采用泊松过程函数发生器生成报文时戳,生成出的攻击具有了应用的时间特性。

参考文献

[1]. 1999 DARPA Intrusion Detection Evaluation Data Set [M]. 2009.

[2]. 汪洋. IDS 测试系统 AOLES 的设计与实现[J]. 2004.

[3]. 王晨．IDS 评估系统(AOLES)测试数据集的生成[J]. 2004.

[4]. 杨望．网络入侵检测系统的评估方法研究[J]. 2009.

[5]. Afanasiev, F., Petrov, A., V.Grachev and Sukhov, A. Flow-based analysis of Internet traffic[J]. Russian Edition of Network Computing, 2003. 5 (98). 92-95.

[6]. Barakat, C., Thiran, P., Iannaccone, G., et al. Modeling Internet Backbone Traffic at the Flow Level[J]. IEEE TRANSACTIONS ON SIGNAL PROCESSING. 2003 51(8):2111-2224.