第 32 卷　第 20 期　　计　算　机　工　程　　2006 年 10 月
Vol.32　№ 20　　Computer Engineering　　October 2006
·安全技术·　文章编号：1000—3428(2006)20—0140—03　文献标识码：A　中图分类号：TP306+.2

一种针对网络信息安全系统的测试方法

刘　莹[1,2]，田　野[1,2]
(1. 中国科学院计算技术研究所，北京 100080；2. 中国科学院研究生院，北京 100039)

摘　要：随着网络的高速发展，网络信息安全产品日益成熟，但是针对这类产品的测试方法的研究却仍不完善。文章基于 ServerScope 测试仪的软硬件结构，提出了一种新的测试网络信息安全产品的方法。该方法继承了 ServerScope 软件结构上面向用户的特点，同时利用 ServerScope 硬件刀片服务器的特点，在系统中同时模拟客户端和服务端，易生成网络信息安全系统所特有的交互式负载。该方法弥补了传统回放方法负载单一的不足，并且简化了复杂的测试环境，使得测试更加易行、有效。

关键词：网络信息安全系统；ServerScope；交互式负载；性能评测

Performance Evaluation for Network Information Security System

LIU Ying[1, 2], TIAN Ye[1,2]
(1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080;
2. Graduate School of Chinese Academy of Sciences, Beijing 100039)

【Abstract】With the fast development of network, network information security systems (NISS) make a long progress. But performance evaluation for NISS is found to be challenging. The paper illustrates a new method based on the architecture of ServerScope to test NISS. To generate the interactive network traffic especially needed by NISS, some blades can play the role of the servers and some do the clients. This method makes up the disadvantage of replaying the same traffic and simplifies the testing environment so that it ensures that the implementation of test is easy and efficient.

【Key words】Network information security system; ServerScope; Interactive workload; Performance evaluation

网络信息安全系统主要是指用于保证用户的网络系统和信息安全系统正常运行的各种软硬件产品，包括防火墙、入侵检测系统(IDS)、信息加密系统、安全认证系统、防病毒系统和安全评估系统等。它们可以及时发现网络攻击行为、病毒以及不正当的内容，并有效地帮助系统应对突发事件。

虽然网络信息安全系统得到了长足的发展，但是针对它们的性能评价方法却仍不完善。由于缺少统一的、成熟的测试方法、标准和测试软件，大多数用户并不能切实地对产品进行实际的横向对比。

本文利用ServerScope[1]性能测试仪的可扩展性和面向用户的特点，提出了一种简单、有效的测试方法。

1 相关工作

网络信息安全系统按照在网络中的接入形式，主要分为串接设备和旁路设备，因此测试方法也可分为串接测试和旁路测试。

串接测试，主要用来测试以防火墙为代表的串接设备。这类设备多以静态的方式侦测进出网络内部的通信参数，符合条件的通信予以放行，否则截断。串接测试比较简单，与传统的测试路由器、交换机等网络设备方法基本相同，只是测试负载不仅包括单一的数据包，也包括完整的网络连接。

旁路测试，主要用来测试 IDS、内容安全软件等具有旁路监听性质的设备。这类设备以动态的方式侦测进出网络内部的通信内容，不符合条件的通信内容予以相应处置。由于该类产品的旁路特点，通常采用以下两种测试方法：(1)回放网络流量的方法，即用网络监听工具将网络流量记录到文件中，然后用回放工具将文件的内容朴素地发送到网络中。这种方法简单易行，但是负载的内容缺乏灵活性，且负载特征不具有网络流量的普遍性；(2)如图 1 所示，分别用两台主机产生前景目标流和背景噪声流，两种负载再通过交换设备混合为测试需要的网络流量，发送至测试系统。第 2 种方法虽然在某种程度上弥补了第 1 种方法的不足，但是用于测试的硬件设备较多，测试环境较为复杂，不易操作。

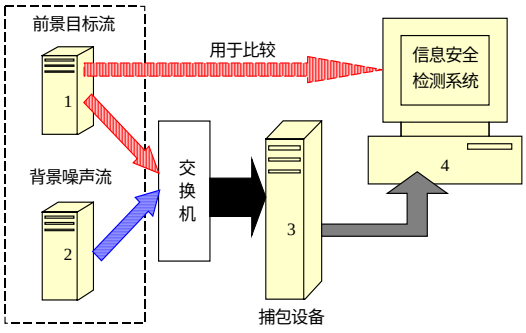


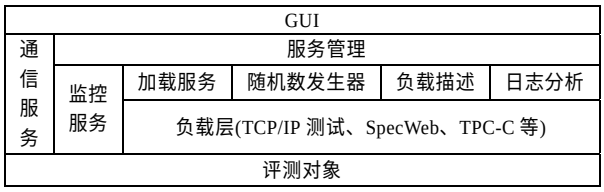
图 1　常用的网络信息安全软件的测试环境

2 我们的工作

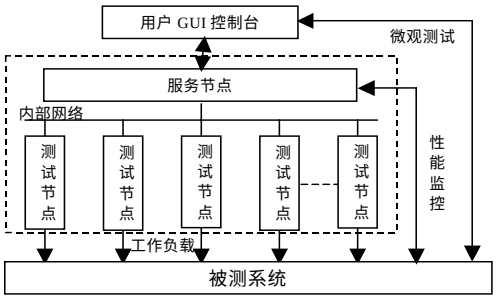
ServerScope 是一种面向用户的，旨在评测用户系统实际性能的评测系统。图 2(a)为 ServerScope 的软件结构，分为控

基金项目：国家“863”计划基金资助项目(2002AA104410)；中科院计算所创新课题资助项目(20026030)
作者简介：刘　莹(1978－)，女，博士生，主研方向：计算机系统的性能评测，网络性能评测，大规模数据密集型应用性能评测；田　野，博士生
收稿日期：2005-12-28　**E-mail：**liuy@ncic.ac.cn

制层、管理层、服务层和负载层，它们分别与图 2(b)硬件结构中的服务节点和测试节点相对应。



(a)软件结构



(b)硬件结构

图 2 ServerScope 系统结构

ServerScope 可以按照用户需求定制测试负载，插件式增减测试负载程序，动态调节负载比例，自定义负载流量模型，具有低成本运作、可扩展性、对象客户化和易用性等特点。硬件结构中的测试节点采用刀片式计算机，具有高密度、易部署、简化连接、易于管理和低成本的特点。

下面将根据 ServerScope 的良好的可扩展的软件结构和硬件特点，详细介绍一种简单、灵活的测试网络信息安全系统的方法。

2.1 测试方法

研究表明，传统的测试交换机、路由器的方法对于测试IDS等旁路设备是不够的，它只能说明系统的基本网络处理能力，并不能全面地测试系统的信息安全监测的性能[2~4]。所以ServerScope已有的TCP/IP负载发生器[5]虽然具有发包长度可变、按长度比例发送、速率可调等特点，却不能满足该类系统的特征测试。为此，添加背景噪声流和前景目标流两个模块，如图 3 所示。

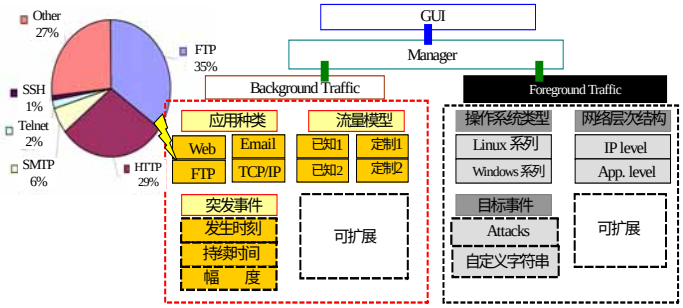


图 3 测试软件的体系结构

背景噪声流模块分为网络应用种类、网络流量模型、网络流量突发事件 3 个部分，用以描述和模拟正常的网络情况。其中，根据实际测量和测试需要，设计了HTTP、FTP、E-mail[6]等网络应用，加上原有的TCP/IP负载发生器，一起作为背景噪声流的网络负载。在网络流量模型部分中提供马尔可夫模型、自相似模型等已知模型，用户也可以自行定义流量模型，设定负载类型间的比例关系。突发事件的提出源于网络流量自身的突发性以及模拟一些特殊环境，用户可以根据突发事件的发生时刻、持续时间以及流量幅度进行描述。

前景目标流模块产生的是一些有针对性的异常行为，包括入侵事件、用户自定义的匹配字符串等。这里，用户可以根据提供的操作系统种类和发生的网络层次进行选择。目标事件的收集和整理是有相当难度的，需要花费大量时间。因此这里可以参考前人的一些做法[7]。

2.2 负载的比较

在通常的网络设备测试中，一般是由测试方直接向被测方发送负载进行测试，而在测试旁路设备时，测试负载不再是单一的测试请求，而是用户与服务器间的交互过程。

图 1 所示的测试环境虽然可以生成交互式的测试负载，但是硬件环境较为复杂，需要多台测试主机，因此测试代价也比较高。ServerScope 的硬件特点刚好解决了测试内容和测试环境的两方面需求。测试时，将一部分刀片作为服务器节点，把另一部分刀片作为客户端节点发送请求。刀片体积小，性价比高，这种方法不仅从软件上实现客户端和服务器的真实交互，弥补了回放方法的不足，更从硬件上节省了测试开销，使得测试更加简单易行。

3 测试结果

测试对象为 Snort 的 2.1.2 版本，一种基于规则匹配的网络入侵检测系统。参与测试的节点配置为两颗 Intel(R) XEON(TM) CPU 1.80GHz，2GB内存，运行RedHat7.3(kernel 2.4.18-14smp)，100Mbps网卡(以下测试节点配置均同此)。

3.1 基本网络处理性能测试

实验中使用 ServerScope 的 TCP/IP 发包负载发生器进行测试。首先分别发送不带匹配字符串的 64B 和 1 518B 的 TCP 包，观察 snort 的基本捕包性能，然后发送带有字符串“ 123 ”的同样大小、数量、速率的数据包测试 snort 的规则匹配性能。测试时，测试节点与被测节点直联。

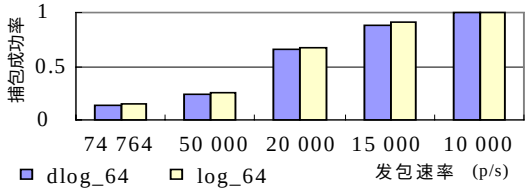
规则文件为 snort.rule，规则内容为“ alert tcp any any -> 10.10.99.98 any (content:"123";msg:"456";) ”。执行以下 4 种命令，如表 1 所示。

表 1 测试命令

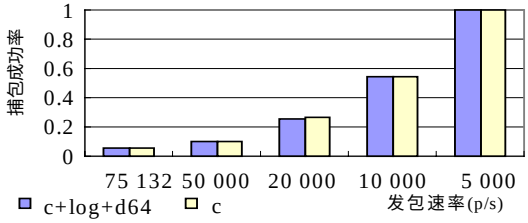
命令名称	内容
log_64/1518	*./snort –l ./log*
dlog_64/1518	*./snort –l ./log –d*
c	*./snort –c snort.rule*
c+log+d64/1518	*./snort –c snort.rule –l /log -d*

(1)64B 小包

每次发送 100 万个包进行测试，测试结果如图 4 所示。



(a)测试 snort 的捕包性能



(b)测试 snort 的规则匹配性能

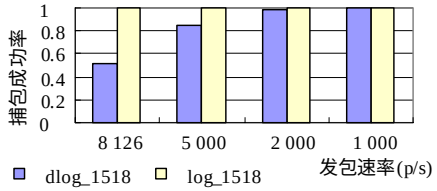
图 4 用 64B TCP 包测试 snort 性能

当发包速率为 75 132pps，网络带宽占用率为 38.47%，此时 snort 的捕包成功率只有 14%，对应的规则匹配成功率

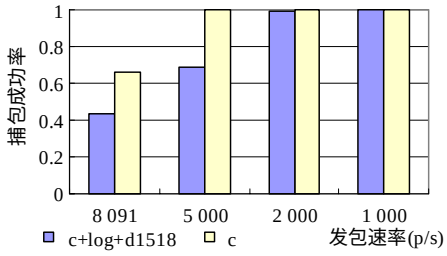
只有 5%；调节发包速率，使发包速度降到 50 000pps 时，捕包成功率升至 24%，而对应的规则匹配成功率仍不到 10%；当发包速率降为 20 000pps 时，性能有明显好转，捕包成功率达到了 66%，规则匹配性能也上升到 26%。当发包速率减为 10 000pps 时，没有丢包现象，而此时规则匹配性能仍不乐观，直到发包速率降为 5 000pps 时，规则匹配成功率为 100%，但此时的网络吞吐率只有 2.56%。

(2)1 518B 大包

每次发送 10 万个包进行测试，测试结果如图 5 所示。由于大包的网络延迟较大，因此性能明显好于小包，并且由于数据包的 payload 所占包长度比例较大，因此有无参数“d”性能差别很大。在无需记录 payload 内容时，发送速率达到线速 8 126pps 的情况下，捕包记录成功可达到 100%，这一结果可以从 64B 小包的实验中得到解释。当需要记录包的数据内容时，log 的成功率只有 50%，此时的规则匹配记录成功率分别为 44%(+d)和 66%(-d)。当发包速度调整至 2 000pps 时，二者的成功率都接近 100%，此时网络带宽的吞吐率约为 24.3%。



(a)测试 snort 的捕包性能



(b)测试 snort 的规则匹配性能

图 5 用 1518B TCP 包测试 snort 性能

(3)小结

小包和大包的测试表明，snort 处理小包的能力还是十分有限的，攻击者可以利用系统的这一缺陷将攻击行为隐藏在大量的小包中，使其不易发现。

3.2 连接处理性能测试

包测试虽然速度快、数量大，但是真正的攻击往往是隐匿在完整的 TCP 连接中，所以基于连接的性能测试比包测试更加有意义。基于连接的监测也逐渐被 IDS 开发者所重视。

这里选用了 WebStone2.5 作为测试工具。WebStone 作为一种较为成熟的 Web 服务器测试工具，主要用来测试 Web 服务器的单位时间请求数和吞吐率。它自带 PCM 网站的历史网页作为负载，用户也可以根据自己的需求增减负载，并可以按照网页的大小、类型定制负载和动态/静态请求，这里常用的是 GET 请求方式。测试时选用 3 个节点分别安装 Webclient、Webmaster 和 Appache2.0.49，作为客户端 Cnode、控制节点 Mnode 和服务节点 Snode。并将 3 节点与被测的 snort 系统通过 hub 连接，让 snort 监听 Cnode 与 Snode 间的交互流量。

实验中根据 Zipf 分布，定义文件的被访问概率，具体情况如表 2 所示。

表 2 WebStone 负载分配

文件名称	Ratio %	Filesize
/wbtree/223_1.gif	35	223<1k
/wbtree/6040_1.htm	50	6 040<10k
/wbtree/41518_1.htm	14	41 518<100k
/wbtree/11426_1.htm	9	150 260 <1 000k
/wbtree/test.htm	1	123

测试结果如表 3 所示。当激活 log 操作时，需要记录数据包，消耗大量的磁盘 IO 操作，所以此时记录到 session.log 文件中的连接数相对较少，当建立 18 200 次连接时，记录成功率约只有 48.9%。但是随着客户端请求数的增加，在省去 log 操作的情况下，session.log 的记录也出现丢失现象，当建立 24 343 次连接时，丢失率为 8%。

表 3 WebStone 测试 snort 结果

Client num.	总连接数	A+L	A-L	Con./s	服务器端吞吐率 (Mbps)	客户端吞吐率 (Mbps)
1	18 200	8 900	18 200	305.37	22.60	22.73
2	22 337	10 135	22 291	387.43	28.62	28.74
4	24 343	11 238	22 377	405.72	30.52	30.56

4 总结

本文提出了一种新的测试网络信息安全系统的测试方法。该方法利用 ServerScope 的可扩展性，从软件上实现测试负载的多样性，从硬件上简化了测试的复杂性。它弥补了回放方法负载单一的缺点，并根据流量模型和突发事件模拟真实网络环境。它通过插件的方式任意增减负载的种类，从而增强了测试的灵活性，为测试提供了更广阔的测试内容的空间；同时它通过用刀片作为服务节点，既实现了测试流的交互性，又简化了测试环境的部署。目前该测试方法已经成功地应用到 snort 等安全系统的测试中，并取得了较好的效果。

笔者将在下一步工作中有针对性地开发一些负载发生软件，使之更加接近用户的测试需求。此外，负载发生性能也是影响测试结果的一个主要因素，对测试软件体系结构的探讨和性能的深入优化也是下一步工作的重点。

参考文献

1 焦丽梅, 孙凝晖, 褚兴军. ServerScope: 一种新的性能评测系统的提出[J]. 计算机工程与应用, 2003, 39(13): 55-57.

2 Hall M, Wiley K. Capacity Verification for High Speed Network Intrusion Detection Systems[C]. Proceedings of the Recent Advances in Intrusion Detection, 2002.

3 Antonatos S, Anagnostakis K G, Markatos E P. Generating Realistic Workloads for Network Intrusion Detection Systems[C]. Proc. of ACM WOSP' 04, 2004.

4 Antonatos S, Anagnostakis K G, Polychronakis M, et al. Performance Analysis of Content Matching Intrusion Detection Systems[C]. Proceedings of the 4th IEEE/IPSJ Symposium on Applications and the Internet, 2004.

5 刘 莹, 焦丽梅, 孙凝晖. 基于 ServerScope 的 TCP/IP 负载发生器的研究[J]. 小型微型计算机系统, 2005, 26(5): 850-854.

6 Hubona G S, Burton-Jones. Modeling the User Acceptance of E-Mail[C]. IEEE Proceedings of the 36th Hawaii International Conference on System Sciences, 2003.

7 Undercoffer, Pinkston J. Modeling Computer Attacks: A Target-centric Ontology for Intrusion Detection[C]. proc. of RAID'03, 2003.