基于变化流量互补测试集的入侵检测系统测试

李 培
(西安邮电学院信息中心 西安 710061)

摘 要 面对目前网络安全产品的层出不穷,入侵检测系统无疑是近年来安全厂商大力研发的产品,同时也是各种规模网络管理用户的必选产品。因此系统性能的提高、入侵检测系统的选择,以及相关的衡量标准,都是研发人员和使用者共同关心的问题。提出了一个基于变化背景流量应用互补测试集的入侵检测系统测试方案以及相应的测试指标,采用 Smartbits,blade 等先进软硬件工具,分别对开源系统 Snort 和某款商业软件 6.0 版本进行了测试,并给出了测试结论。

关键词 变化背景流量,互补测试集,入侵检测系统
中图法分类号 TP393.08 **文献标识码** A

Test of Intrusion Detection System with Mutual Complementary Test Collection Based on Diverse Flux

LI Pei
(Information Center, Xi'an Institute of Post and Telecommunications, Xi'an 710061, China)

Abstract With network security products emerging in endlessly now, intrusion detection system is the product researched and developed by security manufacturer with great effort without doubt, and it is required by users of various scales for administration of network. So how to improve system performance and to choose intrusion detection system, having correlative scale standard or not is problem concerned by both researcher s and users. A test project of intrusion detection system with mutual complementary test collection based on diverse background flux and relevant test index were presented, using advanced software and hardware such as Smartbits, blade to test snort and another business software version 6.0, and the test result was given at last.

Keywords Diverse background flux, Mutual complementary test collection, Intrusion detection system

1 引言

计算机用户都有过感染病毒,影响操作,丢失文件,甚至机器瘫痪,导致系统崩溃的经历,可见计算机的安全问题与我们是多么的戚戚相关。因此,为了达到保护计算机的目的,安全产品是必不可少的,而如何选择安全产品,对广大的计算机使用者来说,使用的技术等都并不关键,真正关心的还是它的功能和性能。

入侵检测系统作为防火墙之后的第二道安全闸门,目前已经成为了网络管理机构的必选产品。面对市场上出现的琳琅满目的入侵检测系统,以及网络上开源的各类入侵检测系统,最终判断其功能和性能还是需要专业的测试。但是,目前都没有一个公认的测试方案以及测试标准,各个厂商的产品自测显然有"王婆卖瓜"之嫌,因此大家都很希望能够如同现今大多数成熟家电一般,可以有统一而简单的标准使我们清楚地了解到产品的功能和性能情况,从而方便选择。

为实现这方面的目标,仅靠少数人或是小的团体都是不行的,需要安全厂商的共同努力。本文提出了一个基于变化流量应用互补测试集的入侵检测系统测试方案,希望可以起到抛砖引玉的效果。同时,为了验证该方案的正确性和可行性,采用该方案对两类不同的检测系统进行了测试。

在进行测试之前,首先要确定测试的指标体系,以便定量评价 IDS 的性能。

2 入侵检测系统测试指标概述

2.1 误报率

由于网络攻击类型众多,而且攻击行为非常复杂,因此在编写 IDS 规则时很难全面而准确地描绘出如此多的攻击行为特征,难免存在着不严密甚至错误之处,从而造成 IDS 对一些网络行为很难定性的现象。在真实的网络中,一些正常的访问连接或虚假攻击的数据包可能存在某些字段正好与 IDS 某条规则能够匹配,从而可能造成误报。

IDS 将正常的访问行为或者伪造的攻击行为报告为攻击行为的情况称之为误报。

在本次测试中,将以下 3 种情况统一看作为误报事件,计入误报率中。

(1)正常的访问行为被误判为攻击行为;

(2)有意伪造的攻击行为(如使用 stick 工具,而攻击实际

到稿日期:2008-07-12 本文得到陕西省自然科学基金项目(SJ08-ZT15)和陕西省教育厅科研项目(08JK245),西安邮电学院中青年教师科技项目基金资助。

李 培(1980—),女,硕士,助教,主要研究方向为计算机网络信息安全,E-mail:lipei@xiyou.edu.cn。

并未发生)被误判为攻击行为;

(3)某些指定类型的攻击行为被错误地报为其他类型攻击行为。

计算公式为:

$$\text{误报率}=\frac{\text{正常行为误报事件数}+\text{有意伪造攻击事件数}+\text{错误报送事件数}+\text{待定误报事件数}\times\text{权重}}{\text{攻击测试事件总数}}\times 100\%$$

"待定误报事件"是指以下3种情况:

1)该类行为是否应该算为攻击;

2)攻击名称上存在差异,无法确定是否针对同一漏洞;

3)其他可能出现的争议行为。

权重根据行为的危害程度、出现的频数、对系统的影响程度等因素从0.0~1.0进行变化调整。

2.2 基本漏报率

基本漏报率是在无背景流和不采用IDS逃避技术情况下,使用规则测试集发起确定的攻击,但IDS系统未能报告此攻击行为。产生绝对漏报的原因包括:

1)IDS没有检测此攻击的规则。

2)IDS有关此攻击的规则制作错误,无法正确判断攻击,放过了攻击代码。

基本漏报率如下定义:

$$\text{基本漏报率}=\frac{\text{漏报攻击事件条数}}{\text{攻击测试事件总数}}\times 100\%$$

2.3 特别漏报率

特别漏报率指能够正常准确报告攻击的发生,但是在某些特定情况下无法正常报警,出现漏报攻击行为的情况。

本次测试中根据漏报的原因,分别定义了标准特别漏报率和异常特别漏报率。

(1)标准特别漏报率

标准特别漏报率主要是指不采用任何IDS逃避技术,只在不同字节和不同强度的正常背景流下,使用测试"基本漏报率"的攻击工具测试集进行攻击测试,主要考察IDS自身的性能。

在实际网络应用中,IDS往往是在一定背景流量下工作。在无背景流量下,测试出IDS的漏报率是没有特别大的实际意义。因此,我们提出一种新的测试思想,就是在无背景流量下,用工具集进行攻击测试,记录下IDS报告的所有攻击行为,作为基本数据。然后在不同背景流量下,用同样的工具集进行攻击测试,记录下报告的所有攻击行为,与基本数据进行对比,从而得到在不同背景流量下IDS的标准特别漏报率。

所以,在本次测试中,分别采用不同字长(128字节、512字节、1518字节)不同强度(10M、50M、80M)的流量下,测试IDS的漏报率。

$$\text{某流量压力下标准相对漏报率}=\frac{\text{该流量压力下漏报攻击事件条数}}{\text{攻击测试事件总数}}\times 100\%$$

(2) 异常特别漏报率

异常特别漏报率是采用专门的IDS逃避技术和工具在不同背景流量下进行测试,测出IDS的异常特别漏报率。此指标主要用来反应IDS产品对采用了逃避技术的攻击的检测能力,反映了检测引擎对IP分片包、TCP流重组的能力。采用的逃避技术包括IP包分片和TCP流分段技术、URL混乱技术、混合逃避技术。URL混乱技术是使用Whisker产生的。混合逃避技术包括了ADMmutate变型编码技术。

主要采用以下技术和工具:

1)利用著名的fragroute,使用测试"基本漏报率"的部分攻击工具测试集进行攻击测试,检查IDS检测引擎是否可以进行完整的TCP重组等等,发送攻击。

2)使用URL编码工具Whisker对CGI的攻击代码进行等效变形,检查IDS检测引擎是否会产生漏报。

$$\text{基本漏报率}=\frac{\text{漏报攻击事件条数}}{\text{攻击测试事件总数}}\times 100\%$$

在本次测试中,我们将根据上面的定义,将异常特别漏报率分别定义如下:

$$\text{fragroute技术异常特别漏报率}=\frac{\text{fragroute技术下报告攻击事件条数}}{\text{攻击测试事件总数}}\times 100\%$$

$$\text{whisker技术异常特别漏报率}=\frac{\text{whisker技术下报告攻击事件条数}}{\text{攻击测试事件总数}}\times 100\%$$

异常特别漏报率=fragroute技术异常特别漏报率×权重×50%+whisker技术异常特别漏报率×权重×50%

注:"权重"是指根据用户需求和实际用途,对fragroute和Whisker攻击行为的重视程度,取值范围为0.1~1。

3 测试流量仿真

由于IDS系统自身性能的原因或者IDS所在操作系统性能原因,在一定的流量压力下,操作系统无法继续抓取数据包,IDS检测引擎也开始丢包,无法重组会话,甚至遗漏关键特征数据包而漏报此攻击行为。

因此,网络背景流量对测试结果影响很大。网络背景流量中的各个数据包大小、内容、协议类型和流量大小等决定了IDS的处理方式,在很大程度上决定了性能指标的有效性和真实性。

为了更加真实地反映IDS在实际应用中所表现出来的真实性能,测试中采用了比照真实网络流量的模拟流量。真实的网络流量包含了多种协议流,每种协议流都占有一定的比例。所以,在模拟背景流量时,不仅需要考虑模拟的网络带宽大小,还需要考虑不同大小、不同协议类型的数据包在数据流中所占的比例。在测试前,我们抓获了校园网一段时期内的实际网络流量,通过分析,计算出了各类TCP协议所占的比例,然后确定需要产生哪些协议流以及各类协议流所占的比例,如表1所列。

表1 流量分析表

	流量(K)	比例
HTTP	19722000000	88.576%
POP	44122000	0.198%
SMTP	522369000	2.346%
Finger	362940	0.002%
Telnet	79684000	0.358%
FTP	1897000000	8.520%
	22265537940	100%

按照表1所列比例,采用Smartbits产生相应比例的背景流量作为测试流量。

和误报率指标不同，本次测试中漏报率的测试采用两个指标：基本漏报率和特别漏报率。

4 测试过程

4.1 测试目的

入侵检测产品有百兆、千兆之分，本次测试的是百兆IDS；入侵检测产品一般分为基于主机和基于网络两类，本次测试对象为基于网络的IDS。入侵检测产品采用的方法有特征检测（误用检侧）或异常检测两种，本次测试对象为采用特征检测的IDS。

本次测试对象是某著名商业公司的IDS 6.0和snort 2.3.0，将针对该测试对象进行功能和性能测试。功能测试主要是检查和对比一下两个IDS对各种网络攻击的检测能力；性能测试主要是检查在不同字长、不同强度背景流量下，两个IDS对网络入侵攻击事件的检测能力。

4.2 测试环境框图

测试环境由3个部分组成，如图1所示。

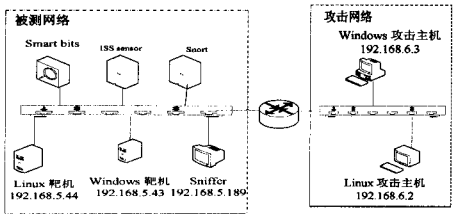

图1 测试环境框图

1）被测网络：由以下几部分组成

Windows靶机：系统中保留常见的漏洞、后门，IP地址为：192.168.5.43；

Linux靶机：对外开放WWW，FTP，Telnet，Finger等服务；

Smartbits：按照实际网络中各网络协议数据流量的比例，产生不同字节数和不同流量大小，模拟正常背景流量；

Sniffer：用于抓取被测网络中的流量，进行攻击特征数据包分析；

10M/100M自适应HUB：保证了整个被测网络可在0M～100M流量环境中，测试Beta1版的各项能力和指标。

2）攻击网络由以下几部分组成

Windows攻击机：预装多种攻击手段向Windows靶机和Linux靶机发起攻击；

Linux攻击机：安装各种Linux攻击工具和脚本向Windows靶机和Linux靶机发起攻击；

Windows靶机：安装blade软件，实现145种攻击手段和数据的回放；

D-Link 100M交换机：连接两台攻击机，保证攻击顺利进行；

Cisco 2600路由器：将两个网段链接，保证测试顺利进行。

4.3 测试工具

本次测试采用的攻击工具为Blade，内置有646种攻击脚本和特征，其详细攻击分类如表2所列。

Fragroute进行IDS逃避技术检测，即首先运行Linux下fragroute，建立攻击回路。然后，运行攻击测试脚本。

在Linux下采用Whisker可以进行CGI的攻击代码变形攻击。

表2 Blade攻击脚本列表

攻击类型	数量（单位个）
backdoor	280
Dns	11
Dos	81
FTP	22
Finger	1
HTTP	190
IGMP	4
IMAP	6
LPRng	4
NetBIOS	14
Nmap	3
RPC	24
Rlogin	1
SMTP	2
SNMP	1
Trace	1
RAS	1

4.4 测试过程

（1）正常攻击检测功能测试

在没有任何背景流量下测试IDS产品对一般攻击类型的检测能力。它可以反映出IDS检测引擎处理数据包的能力。在此项测试中，我们采用Windows靶机上运行的blade软件释放145种攻击方式和数据，得到IDS的误报率和基本漏报率。

（2）强度性能测试

该项测试可以反映出在大负荷背景流量下，IDS对各种攻击行为的检测性能，以及该IDS的负荷能力。本次测试使用Smartbits产生相应的64，128，256，512，1024，1518字节10M，50M，80M背景流量，每组背景流量按照表1产生相应比例的协议流，计算标准特别漏报率。

（3）逃避技术检测功能测试

该项测试反映IDS产品对采用了逃避技术的攻击行为的检测能力，反映了检测引擎IP分片包、TCP流重组能力。采用的逃避技术包括IP包分片和TCP流分段技术、URL混乱技术、混合逃避技术。URL混乱技术是使用Whisker产生的。混合逃避技术包括了ADMmutate变型编码技术。在此项测试中，将得到IDS的异常特别漏报率。

5 测试结果与分析

经过近一个月IDS测试，我们得到如下测试结果。

Realsecure和Snort的误报率如图2所示。

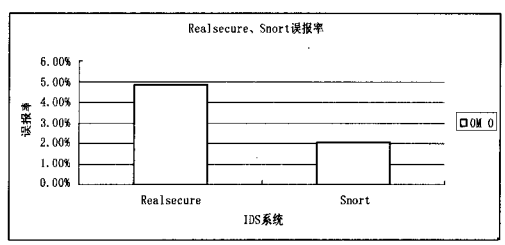

图2 Realsecure，Snort误报率

由此图可以看出，Realsecure的误报率高于Snort。

（下转第122页）

结束语 多模式分类作为传统算法的外延正在引起越来越多的研究和应用人员的重视。现实世界中混合数据的广泛存在也对数据分析的普适性提出更高的要求。本文将离散化处理和图形模式表达有机地结合在一起，构造局部性能最优的多模式贝叶斯分类模型，并给出了结构学习和参数学习的基本思路。在UCI机器学习数据集上的实验结果证明了本算法的合理性和有效性。进一步研究的内容包括先验信息的传播方式、多模式理论在回归分析、聚类和神经网络等方面的推广应用。

参 考 文 献

[1] Savakis K. Bayesian network structure learning and inference in indoor vs. utdoor image classification // Proceedings of International Conference on Pattern Recognition. 2004:479-482

[2] Peter J F. Bayesian network modelling through qualitative patterns. Artificial Intelligence, 2005, 163:233-263

[3] Zhiqiang Y, Rebecca N W. Privacy-Preserving Computation of Bayesian Networks on Vertically Partitioned Data. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(9): 1253-1264

[4] Luis M C. A Scoring Function for Learning Bayesian Networks Based on Mutual Information and Conditional Independence Tests. The Journal of Machine Learning Research, 2006, 7: 2149-2187

[5] Jennifer N, David J. Relational Dependency Networks. The Journal of Machine Learning Research, 2007, 8:653-692

[6] Boaz L, Josepha Y, Lev K. On the Classification of a Small Imbalanced Cytogenetic Image Database. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2007, 4(2): 204-215

[7] Gal E, Iftach N, Friedman N. Ideal Parent Structure Learning for Continuous Variable Bayesian Networks. The Journal of Machine Learning Research, 2007, 8:1799-1833

[8] 田凤占，张宏伟，陆玉昌，等. 多模块贝叶斯网络中推理的简化. 计算机研究与发展，2003，40(8)：1230-1237

[9] 胡小建，杨善林，胡笑旋，等. 基于贝叶斯网的决策表系统的优化分解. 计算机研究与发展，2007，44(4)：667-673

[10] Wang L M, Yuan S M. Induction of hybrid decision tree based on post-discretization strategy. Progress in Natural Science, 2004, 14(6): 541-545

[11] Silverman B W. Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability, 1986

[12] Smyth P, Gray A, Fayyad U. Retrofitting decision tree classifiers using kernel density estimation // Proceedings of the 12th International Conference on Machine Learning. Morgan Kaufmann Publishers, 1995:506-514

(上接第99页)

Realsecure，Snort基本漏报率和标准特别漏报率，如图3所示。

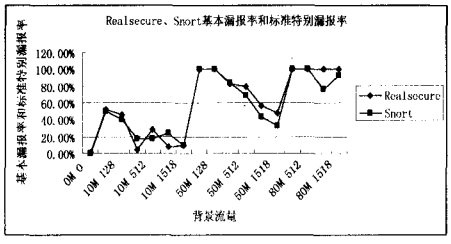

图3 Realsecure，Snort基本漏报率和标准特别漏报率

从Realsecure和Snort的两个漏报率指标可以看出：

(1)在小流量大字节的情况下，Snort的漏报率高于Realsecure。

(2)在大流量大字节的情况下，Realsecure的漏报率高于Snort。

(3)在小字节的情况下，二者漏报率都很高，甚至达到100%。但Snort的表现要比Realsecure好。

Realsecure，Snort异常特别漏报率如图4所示。

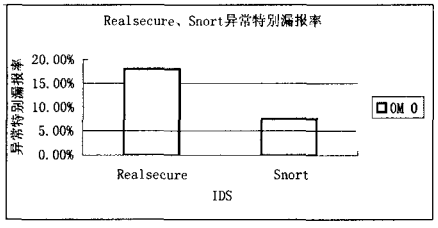

图4 Realscure，Snort异常特别漏报率

由上图可知，Realsecure比Snort的异常特别漏报率要高很多。

以上，根据基于变化流量互补测试集的入侵检测系统测试的思想，应用与之对应的指标体系，通过对Realsecure和Snort两类具有代表性的入侵检测系统的实际测试，得到了关于两者检测入侵行为的测试数据，并进行了相应的对比分析，通过比较，两者在不同的条件下表现出来的性能还是有所区别，但是总体来说各有特长，各有其发挥出色之处。应该说，符合两个系统在实际应用中展现出来的特征，以此验证了该测试方案的正确性和可行性。同时，也通过该测试结果，向使用者表明，使用单一种IDS无法全面解决检测问题，必须相互配合才能真正解决异常检测问题。

参 考 文 献

[1] 郑飞，方敏. 入侵检测技术研究 [J]. 计算机仿真，2004，21(8)：70-73

[2] 张涛，董占球. 网络攻击行为分类技术的研究 [J]. 计算机应用，2004，24 (4)：115-118

[3] 汪洋，龚俭. 入侵检测系统评估方法综述 [J]. 计算机工程与应用，2003，39(32)：171-173

[4] 张雪芹，顾春华，林家骏. 入侵检测技术的挑战与发展 [J]. 计算机工程与设计，2004，25(7)：1096-1099

[5] 蔡忠闽，孙国基，卫军胡，等. 入侵检测系统评估环境的设计与实现 [J]. 系统仿真学报，2002，14(3)：377-380