DAPRA测试分析和IDS测试方法研究*)

吕志军 金 毅 赖海光 黄 皓 谢 立
(南京大学计算机软件新技术国家重点实验室 南京210093)
(南京大学计算机科学与技术系 南京210093)

摘 要 对入侵检测系统的测试是一个系统过程,需要研究攻击测试方法、评价标准等多方面内容。美国国防部高级计划研究局(DARPA)对IDS的两次测试是最有影响的测试。通过对这两次测试以及其它测试评估标准的分析,指出了测试IDS系统的技术难点和重点,从可靠性、可用性、速度、精确度等方面提出了进一步评估IDS的方法。

关键词 入侵检测系统,测试

Analysis of DARPA Test and Research of IDS Test Method

LU Zhi-Jun JIN Yi LAI Hai-Guang HUANG Hao XIE Li
(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)
(Department of Computer Science and Technology, Nanjing University, Nanjing 210093)

Abstract Test of IDS is an important process in the research on IDS, which includes many research aspects such as test method, analysis standard, etc. The test of DARPA in 1998, 1999 is the most important test. Firstly, we analyzis these two tests and other tests; then, the important and difficult things in IDS test are pointed out. Finally, we propose new methods in reliable, usable, rate and accurate aspects to evaluate IDS.

Keywords Intrusion detection system, Test

1 引言

随着计算机和网络技术在社会各方面应用的深入发展,计算机网络系统安全已成为当前研究的热点。入侵检测系统(Intrusion Detection System,IDS)正成为网络安全解决方案中的一个重要组成部分,发挥着越来越大的作用。IDS采用计算机科学中的各种技术对入侵行为进行检测,是一个复杂的应用系统。但在评价其检测效果和性能时,需要有对其进行测试的数据和标准。另外,目前有很多攻击方法针对IDS检测方法的特点,采取躲避IDS系统的方法。因此,需要分别从攻击技术、检测技术、可用性等方面对IDS进行性能和功能测试。对IDS进行测试和评估,有助于更好地描述IDS的特征,认识理解IDS的处理方法、所需资源及环境。同时利用测试和评估结果,可以预测、推断IDS发展的趋势,发现和改进系统中存在的问题。

目前,有许多组织和机构研究IDS的测试方法。其中最有影响的是DARPA在1998年和1999年组织的测试,这两次测试首先研究了各种攻击方法和检测方法,并针对这些攻击方法,设计了典型的测试环境和测试数据集,提出了分析、评价测试结果的方法。

本文详细分析了DARPA这两次测试的特点,并通过和其它测试评估标准和方法的分析比较,指出了测试IDS系统的技术难点和重点。然后根据IDS当前研究的重点,从可靠性、可用性、速度、精确度等方面提出了评估IDS的方法,并指出IDS测试的进一步研究方向。

2 攻击方法的种类

网络中存在大量的攻击方法,这些攻击方法具有各自的特点,根据不同的标准可以分成不同种类。根据攻击的特点,攻击有缓冲区溢出、木马程序、病毒、syn flood、口令破解等方法。根据攻击方法和攻击结果,DARPA[1]结合用于入侵检测系统测试的入侵实例,将攻击分为拒绝服务攻击(Denial of Service,DoS)、探测(Probe)、u2r(user to root)、r21(remodte to local)和Data攻击5种攻击类型。其中,DoS攻击指耗尽系统资源或使系统无法正常工作;Probe攻击指收集主机信息以便后继攻击使用,表现为同一行为主体(用户或IP地址)在短时间内对该主机系统信息的连续访问;u2r攻击指本地用户通过非法手段获得系统的高级访问权限;r21攻击指不具有系统访问权限的远程用户非法获得系统的访问权限;Data攻击指从受害主机上取走安全策略规定禁止取走的文件,该攻击通常是通过u2r攻击获得对该文件的访问权,并通过正常的应用程序(如mail或ftp)将文件移走。每种攻击类型中包含多种具体的攻击方法,新出现的攻击方法基本属于这5种攻击类型。

因为IDS是网络中检测攻击的重要设备,攻击者也将其列为攻击目标,不但采用上述攻击方法直接攻击IDS,还针对IDS检测方法的特点,采用"过载"、"失效"、"欺骗"[2]等方法来躲避IDS的检测。

*)本文研究得到国家863计划2001AA142010(智能入侵检测与预警系统)、2002AA141090(安全服务器)资助。吕志军 博士生,研究方向:信息安全;赖海光 博士生,研究方向:网络与信息安全;金 毅 硕士生,研究方向:网络安全;黄 皓 教授,博导,研究方向:网络安全、分布式计算;谢 立 教授,博导,研究方向:操作系统,信息安全。

3 入侵检测方法

NIST 将入侵定义为"企图对计算机完整性、保密性、可用性的破坏以及绕过计算机或网络安全机制的行为"[3],将入侵检测定义为通过监控发生在计算机系统或网络中的事件,并对这些事件进行分析以识别出入侵的过程。通常 IDS 试图通过观察行为、安全日志、审计数据来检测针对计算机或网络的入侵,这种检测通过手工或专家系统软件对日志或其它网络信息进行分析来完成[4]。

根据检测的数据来源,IDS 分为基于主机和基于网络的入侵检测系统。

基于主机的 IDS 以系统日志、应用程序日志等作为数据源,或通过其他手段(如监督系统调用)从所在的主机收集信息进行分析。主机型入侵检测系统保护的一般是所在的系统。

基于网络的 IDS 分析在网络和操作系统协议栈中传输的数据包,担负着保护整个网段的任务。

根据检测判断的原则,入侵检测可分为异常检测(Anomaly Detection)和误用检测(Misuse Detection)两大类。前者也称为基于行为的入侵检测,以系统、网络、用户或进程的正常行为建立轮廓模型(Profile),即正常行为模式,将与之偏离较大的行为解释成入侵。异常检测方法具有检测系统中新的未知的能力,而网络中不断出现新的攻击方法;因此,异常检测技术一直受到比较多的重视,出现了大量的异常检测技术。

误用检测也称为基于知识或基于签名的入侵检测,根据已知攻击的知识建立攻击特征库,通过用户或系统行为与攻击特征库中各种攻击模式的比较,确定是否发生入侵[5,6]。它的优点是:对入侵行为检测的准确性高,其缺点是:只能发现已知的入侵行为。

IDS 目前研究的重点是解决检测的误报率和漏报率,对未知攻击的检测,高速网络下和分布式攻击检测等问题。

4 DAPRA 测试

虽然 IDS 及其相关技术已获得了很大的进展,但关于 IDS 的性能检测及其相关评测工具、标准以及测试环境等方面的研究工作还很缺乏。

受 DARPA 的委托,林肯实验室分别在1998年、1999年对 DARPA 支持的 IDS 项目进行两次 IDS 离线评估,是迄今为止最有影响的 IDS 评估。

在精心设计的测试网络中,他们对正常网络流量进行了仿真,并开发模拟了大量的攻击[7],将记录下的网络数据、系统日志和主机上文件系统映像等数据,交由参加评估的 IDS 进行离线分析。最后根据各 IDS 提交的检测结果做出评估报告。

4.1 1998年评估[8]

本次评估的目的是测试各 IDS 的关键检测技术和对已知攻击方法的检测能力,并提出公正的评价。通过测试出 IDS 系统的优点和不足,帮助参加测试的系统克服缺点,提高检测的性能。

(1)*测试数据仿真* 1998年测试模拟一个利用路由器和外部网络相连的小型空军基地网。在网络内部安装了三台被攻击主机,分别安装最常受到攻击的操作系统 Linux2.0.27, SUN OS 4.1.4,Sun Solaris 2.5.1。利用软件模拟出一个有1000台主机和100多个用户正常进行的环境,这些模拟主机提供网络中各种常见的服务,各模拟用户自动进行各种网络访问操作。攻击者从外部网络向内部网络发动攻击。

共提供三种离线测试数据源:网络数据包、solaris 机器上获得的 SUN 的基本安全模块(BSM)提供的审计数据和3台被攻击机器晚间备份的磁盘数据。

整个测试数据集包括7个星期的训练数据和两个星期的测试数据。训练数据中已标示出攻击类型;测试数据中增加了训练数据中不存在的攻击方法。

攻击方法共有35种,200多次。其中 Prob 攻击5种43次,DoS 攻击有11种17次,r2l 攻击11种38次,u2r 攻击8种22次。

(2)*评价的标准* 采用受试者作业特征曲线 ROC(Receiver operating characteristic)来评价各种 IDS,ROC 曲线表示检测率和误报率的关键,X 轴是每天的误报率,Y 轴是检测率。在分析测试结果时,分别分析各个待测的 IDS 系统针对四种攻击类型的 ROC 曲线。

(3)*测试的结果* 1998年的测试结果表明当时的研究系统能以较低的误报率,很好地检测已知的攻击。但检测未知的攻击方法的能力不足,即当测试数据中的攻击方法和训练数据中的攻击方法有本质不同时,这些攻击经常被漏掉。需要进一步研究以较低的误报率检测新攻击的方法;第二,测试数据不完整,缺少针对 Windows/NT 的攻击检测测试。

4.2 1999年评估[9]

1999年评估主要测试各 IDS 检测事先未知的隐秘攻击的能力,并分析系统漏检新攻击的原因。

(1)*测试数据* 整个测试网络由内网和外网组成。每天大概有400M 字节的数据。

测试数据集包括4种数据源:内外网的网络数据、Solaris 的安全模块(BSM)数据、Solaris 和 Windows nt 主机中收集的审计数据,以及 4台主机夜间的安全有关的备份数据。

共有3个星期的训练数据集和两个星期的测试数据集。其中第1个和第3个星期是没有攻击的训练数据,用来训练异常检测系统;第2个星期的数据是含有43个攻击例子的训练数据,用来训练误用检测系统;第4和第5星期是测试数据。

1999年测试在1998年已有的4种类型的基础上,增加了 Data 攻击类型,总共有64种,212次攻击。其中 Probe 有37次(8种),DoS 有65次(16种),r2l 有56次(16种),u2r 有37次(12种),data 有13次(4种)。

(2)*评价的标准* 1999年采用两种新型分析方法:(1)对每个系统漏掉的和高计分的假警报进行分析,明确为什么系统会漏掉特定的攻击以及为什么会引起假警报。(2)允许参与者随意地提交讨论信息,帮助安全分析员明确攻击和响应的重要特征。

(3)*测试的结果* 此次测试有8个 IDS 系统参加,基于网络的 IDS 对已知的 probe 和 dos 攻击检测率比较好,基于主机的 IDS 对 Solaris u2r 攻击检测率比较高,对事先未知的隐秘攻击和 Windows nt 攻击的检测率较低。

1999年的评估经过对各种系统的检测结果分析,得出以下结论:1)当攻击具有特定特征,或在日志文件中具有和正常数据不同的事件序列时,系统对此类攻击检测率高。2)当攻击利用 IDS 未监控的协议和服务时,系统会漏检这类攻击。

此次评估表明:建立测试过程和环境,很费时间和资源;需要能自动产生数据流和发动攻击,并能集成评分软件的整套测试工具。

4.3 DARPA 对 IDS 测试的扩展[10]

DARPA 在1998、1999测试的基础上,进一步扩展了测试平台:1)建立一个林肯自动实时信息保障测试平台,提供一套供 IDS 开发机构建立开发和测试环境的工具集;2)建立了两上包含多个攻击方法和步骤的攻击过程数据集 LLDOS1.0,LLDOS2.0;其中 LLDOS1.0包含了利用 IPSweep 探测被攻击系统的 sadmind 服务,利用该服务的缓冲区溢出漏洞获得获得三台被攻击系统的 ROOT 权限,然后分别在这三台机器上安装 DDOS 攻击工具的攻击程序,并在其中的1台机器上安装控制程序,然后通过控制程序控制攻击程序发动 DDOS 攻击;3)进一步分析1999测试的数据,提供检测各种攻击的特征和模型。LLDOS2.0采用比 LLDOS1.0更隐秘的方法实现攻击。

4.4 DARPA 测试的特点

DARPA 测试主要根据各种攻击方法的特点,将其分类,并通过模拟这些攻击方法的数据,来测试 IDS 检测能力的完整性;其中攻击种类丰富,能有效测试 IDS 的检测能力。并且在后续研究中,进一步提供检测 IDS 系统综合检测攻击事件序列的能力。

但 DARPA 测试还存在以下不足:

1)测试数据的生成不完善。首先,测试数据是先分别产生背景数据和攻击数据,然后将两者合起来。这种方法使攻击数据没有真实地分布在背景数据中,攻击数据和背景数据差别明显,减少了背景数据对检测结果的影响。

2)未进行性能测试。在测试时,仅测试了系统在正常情况下的检测能力,而未检测系统在高负载,多事务量,不同攻击密度下的检测效果。没有考虑黑客使用欺骗 IDS 的攻击方法,如数据包分片、数据采用变化的速率等。

3)非实时测试。它只是离线测试,只能检测 IDS 的检测的准确性,而不是实时在线测试,无法检测 IDS 的资源消耗性,实时性,协作性等问题。另外,DARPA 测试没有测试 IDS 系统融合多数据源,检测分布式攻击的能力。

4)缺少对报警信息的关联分析测试。现在很多 IDS 支持综合分析各低层 IDS 的输出信息,分析这些报警信息之间的关系,找出攻击过程,降低误报,形成更高层的报警信息。

5)测试角度有待讨论。这次测试,DARPA 专门研究模拟了很多攻击方法,是从攻击方法的角度去检测 IDS,而不是从 IDS 检测方法的角度去检测。

6)分析评价方法不完整。在分析测试结果时,主要注意了检测率和误报率,而没有考虑检测结果的可信率,即当发生报警时,该报警是真正攻击的比率。当可信率较低时,容易让系统管理员因误报太多,而漏掉对真正攻击的检查。

5 相关的测试方法和测试标准评价

Puketza 等人[11]在1994年开始研究评估 IDS 系统的方法,研究了可以实现自动攻击仿真的软件平台。IBM 的 Debar 等人[12]在1998年建立一个实时测试平台,环境里包括几个客户机和服务器来测试 IDS,指出仿真正常网络流量和攻击是评估工作中的难点。IBM 的 Zurich 研究实验室[13]在2000年开发了一套 IDS 测评工具,提出要从 IDS 设计的角度去检测 IDS 区分入侵和正常行为的能力。

目前还有其它许多机构开展对 IDS 的测试,Neohapsis[14]公司提出开放式安全评估标准(Open Security Evaluation Criteria,OSEC),它首先为网络安全产品提出一个基本的核心测试标准集;并在此基础上,增加了性能和安全性方面的测试标准。OSEC 对网络入侵检测产品的评估重点在于入侵检测产品的能力和本身安全性上,评估主要集中在设备完整性(入侵检测产品本身抗攻击的能力)、检测能力(在有、无干扰数据流情况下的检测能力)、抗躲避能力等方面。这是一种商业性测试,测试主要集中在产品的实用性上。

NSS Group[15]是欧洲一家独立的网络安全测试机构,对入侵检测系统的评估主要从产品的结构、系统的安装、安全策略的更新和分发、检测能力、结果分析四个方面进行。在检测能力方面,主要测试各 IDS 产品在高负载、躲避技术下的检测能力。现已对入侵检测系统的测试评估已经进行了四次,对数十种 IDS 产品进行了测试比较,并生成了相应的测试报告。

在已有的测试 IDS 的各种方法和机构中,有的偏重于测试 IDS 对各种攻击 测能力(DARPA 测试),有的偏重于 IDS 的有效性和易使用性。但都缺乏对各检测方法之间的比较,没有有效的测试结果分析。

6 新的测试方法

目前 IDS 的研究获得了很大的进展,在研究检测传统攻击方法的基础上,根据网络环境(高速网络、大数据量等特点)、攻击方法发展的特点,研究重点已逐渐扩展到采用智能的方法检测新攻击,多数据源融合技术检测分布式攻击,采用并行的方法处理高速数据网络,检测躲避检测的隐秘攻击方法等方面。

根据 IDS 技术发展的状况,需要研究评估新的攻击检测方法的标准,重点是评估检测能力以及建立分析检测结果的标准。

针对目前 IDS 评测中存在的问题,同时为了测试我们自己研究的智能 IDS,我们在 DARPA 测试的基础上提出了自己的测试数据、测试方法和评价标准。建立分布式测试数据集和测试环境,测试 IDS 系统在各种情况下的检测能力,并提供对这些检测能力的评价分析标准。这些标准既测试了 IDS 在单一攻击源等正常情况下的检测能力,同时也评估了其在受到攻击、分布式多攻击源等异常情况下的检测、分析和响应能力。

在评测 IDS 时,提出了分别从可信度、处理性能、抗躲避能力、抗攻击能力、响应能力和可用性六个方面进行测试评价的方案。

检测可信度 指 IDS 检测结果的可信程度,这是评估 IDS 的最重要的指标。主要包括误报率、漏报率和报警可信率。如果 IDS 尽量减少误报率,则漏报率就会提高;反之,如果 IDS 尽量减少漏报率,则误报率就会提高。其中报警可信率是指当系统检测到攻击时,该攻击是真攻击的概率,该指标主要影响系统的可用性。当可信率较低时,系统管理员会被大量的误报所困扰。在测试时,不但测试对单个攻击的检测可信度,还测试对攻击序列的检测可信度。同时测试系统对多源信

息的综合检测能力。

处理性能　指IDS处理数据的能力。当IDS的处理性能达不到高速网络的要求时，它就可能因为来不及处理数据，而漏检数据，不能检测到在低速数据流时该系统能检测到的攻击；从而因不能实时检测入侵，而影响整个系统的性能。处理性能测试分强度测试和资源消耗测试两种。强度测试主要评估IDS在强负荷运行状况下检测效果是否受影响，主要检测大负载、高密度数据流量情况下的检测效果。资源消耗测试主要检查IDS占用系统资源的状况，考虑的主要因素是硬盘占用空间、内存消耗、CPU的占用率等，判断IDS系统对慢攻击等持续时间长的攻击的检测能力。

抗躲避能力　由于攻击者为了加大检测的难度甚至绕过IDS的检测，常常会发送一些特别设计的分组，使IDS截获到的信息与接收方正常收到的信息不一致，错误理解它所接收到的数据，不能检测出入侵行为的方法。该方法主要测试IP包碎片重组能力和TCP流重组能力。IP包碎片指通过将一个IP包分片，将攻击的特征分布在几个IP包碎片中。因为分析单个的IP包碎片会导致许多误报和漏报，当NIDS不能正确地重组分片的IP包时，就可能检测不出攻击，所以IP碎片的重组能力影响检测的精确度。TCP流重组是指通过对完整的网络连接进行分析，完成对应用层数据的分析。抗躲避能力直接影响IDS的检测可信度。

抗攻击能力　由于IDS是网络安全防护中的重要手段，所以它也就成为很多入侵者攻击的目标。和其他系统一样，IDS本身也存在安全漏洞，IDS必须能够抵御对它自身的攻击，特别是拒绝服务攻击。由于大多数的IDS是运行在易遭受攻击的操作系统和硬件平台上，这就使得系统的安全性变得特别重要。

响应能力　主要测试IDS系统的实时性、响应方式和身份追踪分析能力。实时性要求IDS必须尽快地分析数据并把分析结果传播出去，以使系统安全管理者能够在入侵攻击尚未造成更大危害以前做出反应，阻止入侵者进一步的破坏活动。实时性不仅要求IDS的处理速度要尽可能地快，而且要求传播、反应检测结果信息的时间尽可能少。在响应方式中，主要测试IDS本身提供的对入侵的阻击方式，以及和其他安全系统的联动能力。攻击者身份追踪分析能力主要测试系统对攻击者信息的日志记录是否完整，以及对这些日志的分析结果的有效性。

可用性　主要包括系统的可护展性、用户界面的可用性、安全检测策略的分配、系统配置的方便性等方面。

我们在DARPA测试的基础上，利用其攻击方法数据，并增加新的攻击方法，在有100个工作节点、千兆网的实际网络环境中，模拟攻击，组成新的测试数据集。并重点用以下几个方法来分别测试IDS在上述各个方面的性能。(1)将测试数据分成新、旧攻击方法，测试IDS系统对已知攻击方法和未知攻击方法的检测率；(2)分别以单网络环境和多网络环境分布式发送测试数据，测试IDS利用数据融合技术检测分布式攻击的能力。(3)根据抗躲避能力的测试要求，建立分片和数据重发工具，增加攻击数据的无规则性。(4)建立新的攻击场景测试数据集合，测试对整个攻击过程的检测成功率，重点测试IDS对一个攻击过程中各个攻击的关联分析程度。

在分析测试结果是，先分析一般速率下的误报率、检测率，并分析报警可信率，以评价IDS系统的基本检测性以和可用性；然后分析在高速网络、分布工环境下，该IDS系统的检测性能。以六个测试标准对IDS分别进行评价，每个测试标准以不同的权值进行计分，最后形成综合评价。在评价时，去除同一个误报的多次重复，减少评估结果的负面影响，提高评测结果的可信度。理想状况是可以自动地对评测结果进行分析，但实际上很难做到这一点。对IDS的实际测试既包含客观的评估，又包含主观的评估。

小结　对IDS进行评测IDS研究中的一个重要领域，它对推动IDS的发展，提高IDS的性能，具有重要的推动意义。本文在分析已有测试的基础上，针对IDS技术发展的特点，从分布式攻击测试、可信度衡量等方面提出了新的测试方法和分析标准。目前，IDS评估还有很多不完善和有待改进的地方，在网络流量仿真、用户行为仿真、攻击特征库的构建、评估环境的实现和评测结果的分析等方面需要进一步的研究。

参考文献

1 DARPA Site http://www.11.mit.edu/IST/ideval/index.html contains information on the 1998 and 1999 evaluations

2 Ptacek T H, Newsham T N. Insertion, evasion, and denial of service: Eluding network intrusion detection. 1998. http://www.secinf.net/info/ids/idspaper/idspaper.html

3 Bace R, Mell P. NIST Special Publication on Intrusion Detection System. March 2001

4 NSA Glossary of Terms Used in Security and Intrusion Detection. SANS Institute, 1999. http:www.sentinelsys.com/glossary.html

5 Lindqvist U, Porras P A. Detecting Computer and Network Misuse through the Production-based Expert System Toolset (P-BEST). In: Proc. of the IEEE Computer Society Symposium on Research in Security and Privacy, 1999. 146～161

6 Paxson V. Bro: A System for Detecting Network Intruders in Real-Time. In: Proc. of the 7th USENIX Security Symposium San Antonio, Texas, Jan. 1998

7 Das K. Attack development for intrusion detection: [Master's Thesis]. Massachusetts Institute of Technology, Cambridee, MA. 2000

8 Lippmann R P, et al. Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation. In: Proc. of the on DARPA Information Survivability Conference and Exposition (DISCEX'00, Hilton Head, Carolina, Jan. 25-27). IEEE Computer Society Press, Los Alamitos, CA, 12-26

9 Lippmann R P, Haines J. Analysis and Results of the 1999 DARPA Off-Line Intrusion Detection Evaluation. Computer Networks, 2000, 34: 579～595

10 Haines J W, Rossey L M, Lippmann R P. Extending the DARPA Off-Line Intrusion Detection Evaluations. DISCEX-Ⅱ, 2000

11 Puketza N J, et al. A methodology for testing intrusion detection systems. IEEE Trans. Softw. Eng. 22(10): 719～729. http://seclab.cs.ucdavis.edu/papers.html

12 Debar H, Dacier M, Wespi A, Lampart S. An experimentation workbench for intrusion detection systems: [Research Reprot RZ 2998 (#93044)]. IBM Research Division, Zurich Research Laboratory, Switzerland, March 1999

13 Alessandri D. Using rule-based activity descriptions to evaluate intrusion-detection systems. In: RAID2000, H. Debar, L. Me, and S. F. Wu, Eds. Springer-Verlag, New York: NY, 2000. 183～196

14 http://osec.neohapsis.com

15 http:www.nss.co.uk