# 入侵检测系统的部署及入侵模式的识别

陈辰 张佐 吴秋峰 (清华大学自动化系 100084)

摘要: 入侵检测是当今网络安全领域的一个热点问题,本文参考一般防火墙的部署配置方法以及入侵检测系统的特点,采用了一种兼顾效率与性能的旁路入侵检测系统部署结构,可适用于安全性要求从低到高的各种应用场合。

关键词: 入侵检测 入侵模式 旁路IDS系统

## 1 引言

### 1.1 概述

入侵检测技术起源于上个世纪80年代初期,是一种试图在攻击发生之前或者起始阶段,准确地预测并阻止攻击发生或继续的半主动和半智能化的技术手段。所谓半主动,是指它可以预测具有一定特征的潜在的安全威胁的发生;而半智能化指的是在一定条件下,系统可以建立起一定的入侵行为规则及推理库,借助于人机交互识别出入侵行为的模式,进而有可能预测某些未知的入侵方法。入侵检测系统既可以部署主机端,以操作系统或者路由、交换节点的日志作为分析数据源,构成基于主机(Host-based)的入侵检测系统;也可以部署在网络节点上,以网络上所有流经检测点的原始数据包作为分析数据源构成基于网络(Network-based)的入侵检测系统。

入侵检测的实施一般分为几个步骤:首先是采集数据,其次建立入侵检测算法模型,最后是入侵检测系统的实施与部署。根据入侵检测系统的种类不同,采集的数据也有所不同。基于主机系统主要是收集网络系统中被保护主机系统日志和审计记录以及路由器日志记录。基于网络系统则通过分布在网络中的智能代理收集相应位置的数据包。根据系统分析方法的不同也可分为滥用(Misuse)检测分析和异常(Anomaly)检测分析,在滥用检测分析中,主要根据事件日志记录的结果以及已知的、确定的攻击特征作为数据源进行判断分析,而异常检测分析则要收集系统状态:如CPU占用率、内存使用情况、进程空闲时间、网络连接状态等作为数据源进行正常与异常的界定。入侵检测的算法模式也是多种多样的,通常采取的方法有基于专家系统的、基于神经网络的、基于用户行为概率统计的以及基于模型推理的四大类方法。在IDS系统的部署上,随着网络化水平的提高和计算机成本的降低,也逐步由单一的集中(centralized)模式向多代理的分布(distributed)模式过渡。

通过对现有的IDS算法模型以及IDS系统功能的分析,我们认为目前在入侵检测系统中比较重要的问题是:IDS算法模型的效率、准确性、适应性以及IDS系统对于整个网络系统的性能影响。本文以入侵检测为主题,针对上述问题中的IDS模型算法适应性以及IDS对于网络系统的性能影响问题进行了一些深入研究工作。首先探讨IDS系统在网络中不同部署方法及其对于入侵检测系统效率-安全平衡性和IDS本身安全性问题的


图1 具有单一安全连接的网络安全系统模式

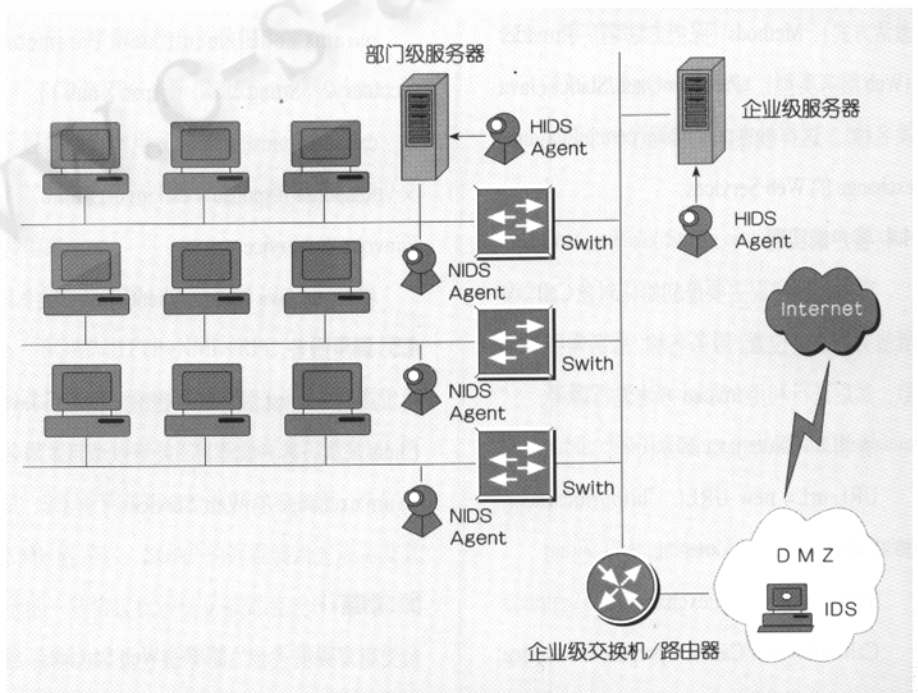


图2 IDS Agent及其分析主机分布逻辑示意图

改进；其次，在分析了大量的典型入侵行为过程后，从入侵行为所具有的阶段性特征出发，提出一种入侵模式的网络化结构，文献调研表明这种想法与Bro的思想相仿；随后，文章给出了一种根据上述入侵模式网络结构判断入侵行为的算法流程；最后，总结全文并提出今后的工作方向。

## 2 IDS系统的部署——旁路IDS

所谓网络，是由多个节点构成的一个复杂互联系统。网络安全问题是一个系统性的问题，而非单一个体（服务器、路由器等）的安全问题。当我们在考虑一个系统的安全问题的时候，往往不能孤立地看待网络中每一个节点的安全性，因为网络所具有的互联特性使得网络系统的安全性往往取决于系统中安全性最差的节点。针对这种特点，在设计网络安全系统的时候通常采用单一遏制点的方法，如图1所示。对于网络系统只开放唯一的安全通路，在此连接上部署各种所需要的安全策略，使得网络系统的安全性完全取决于这个连接的安全性。

在这样的一个系统中，对于入侵检测系统的部署必须要考虑两个问题：第一，如何使入侵检测系统能够方便地获得所需要的分析数据；第二，如何使入侵检测在不影响系统运行效率的情况下对入侵行为做出实时反应。

分布式智能代理（Distributed Intelligent Agents）可以用来解决第一个问题，即通过分布在网络系统中关键位置（如服务器）上的多个智能代理（见图2），收集整理系统日志以及到达该点的数据流，向IDS主机做出汇报，并由IDS分析主机做出实时的分析响应。

关于入侵检测系统的效率问题，我们先进行一个简单的计算：假设攻击特征字符串的平均长度为20个字节，数据包的平均大小为576个字节，那么对于一个数据包为了匹配一个攻击特征所需要进行的比较次数平均为：[20+20x(576-19)]/2=5580次，如果每次比较计算需要15个时钟周期，则匹配一个攻击特征字符串就需要大约$8.37x10^4$个时钟周期。对于主频为2GHz的CPU来说每秒钟可以匹配约24000次，根据CERT组织的统计报告预计2002年系统脆弱性问题的报告将超过4000种，如果针对所有这些系统漏洞进行特征匹配计算，那么使用2GHz的CPU在1秒钟可以完成6个数据包的匹配检测工作。而以线速度运行的快速以太网每秒钟达到的数据包数量为：$(100x10^6)/(8x576)=21701$个，在这种情况下检测1秒钟达到的所有数据包需要大约21701/6=3617秒，即大约1个小时左右的时间。尽管我们可以通过一些方法（比如协议匹配）减少一些不必要的计算，但是随着网络运行速度的增加、攻击种类的增加（需比对的攻击特征增加），匹配攻击特征的计算时间比将以几何级数递增，IDS的分析主机必然不堪实时计算的重负，从而成为拒绝服务攻击的首要攻击对象。如果将入侵检测系统部署在网络唯一的安全连接的遏制点上，那么所有的网络流量都需要经过入侵检测系统的检验，一旦发生拒绝服务的攻击，那么网络系统内的所有主机都将面临失去网络连接的威胁。因此，我们不能将入侵检测系统部署在网络连接的遏制点上。基于上述对计算效率的考虑，参考了一般网络安全系统的配置部署方法，这里我们考虑采取一种旁路配置入侵检测系统的设计方案，见图3。

图中，IDS系统不放在安全连接的遏制点上，而是处于旁路的位置。这是因为：首先网络性能对于网络系统应用极其重要，所以应保证向大量正常访问提供较高的吞吐量和较快的响应速度；其次，大部分网络入侵都不是在瞬间完成的，而是要经过一定的步骤、方法才能取得预期效果，所以可以将所有经过两级防火墙的流量数据（位置A、B）在发往其

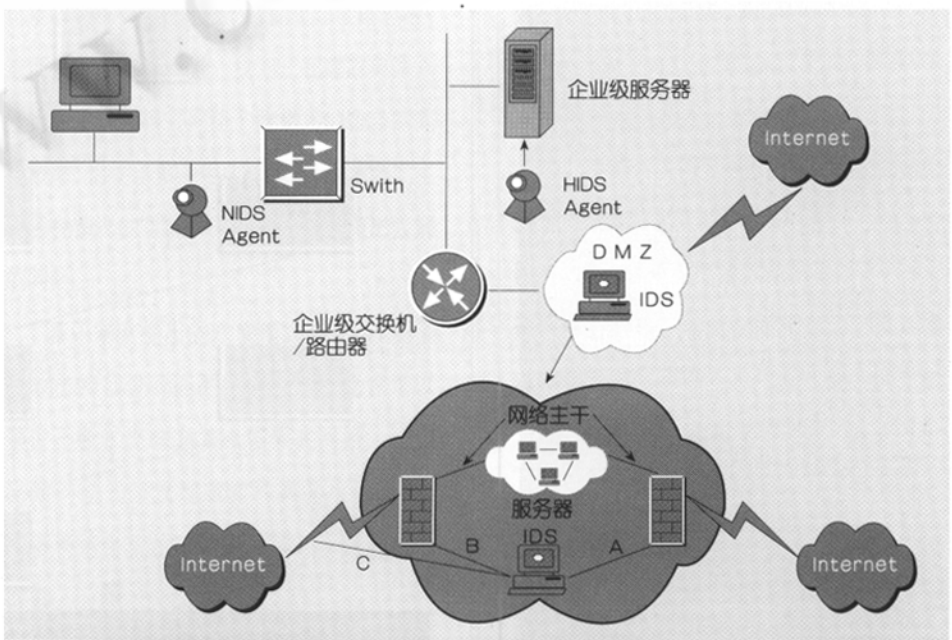


图3 含有旁路入侵检测系统的网络安全体系逻辑结构

目的地的同时，传送到IDS主机，此时，分布在网络中各关键点上的智能代理将各自收集到的日志及数据（位置C）也同时汇总到IDS主机，由IDS主机统一做出分析判断。这样，用户数据包不必经过IDS系统的判断即可直接转发到指定的目的地，所以不存在上述的计算效率问题。其次位于旁路位置的IDS系统对于整个网络来说更像一个处身局外的旁观者，对于入侵者来说旁路的IDS系统透明度更高，IDS系统本身受到的安全威胁更小。为了使网络瘫痪，直接以IDS系统为目标的攻击几乎无法完成。另外，也可根据当前IDS系统的负载情况以及整个网络系统所要求的安全级别来动态调整IDS系统的计算实时性要求，即在最高要求下可达到同配置在安全连接遏制点上相同的实时计算；而在较低的安全及实时性要求下，可以在数据转发的同时由IDS主机做出分析判断，甚至是采取抽样检测的办法检验数据流的安全性，从而提高了IDS的配置灵活性，扩大了其适用范围。

## 3 入侵模式识别

在部署好IDS并获得相应的访问数据后，IDS需要从大量访问数据中找出可疑的恶意访问，并定量给出入侵警告信息。所以，对于入侵模式的形式化描述就显得尤为重要。

经过对绝大部分可能发生的网络入侵的分析，我们提出从入侵后果来看，可以把网络入侵分为破坏网络资源与非法控制网络资源两大类。其中属于破坏网络资源的入侵后果有：主机堆栈溢出、网络线路阻塞、主机瘫痪等，属于非法控制网络资源的入侵后果有：取得合法、超级用户权限、窜改系统数据（用户数据、服务数据、日志数据）、修改系统应用等。而每一种入侵行为的后果又可以归结为几种具体的网络攻击手段，比如Land、SYN Flood、网络监听、CGI攻击等。进一步分析这些具体的攻击形式，不难发现，几乎没有一种入侵手段是一蹴而就的，入侵者往往在大量正常网络应用行为中埋藏少量的攻击手段，渐渐导致入侵结果的出现。因此，许多正常的网络应用都可能同入侵产生的最终结果发生直接或者间接的因果关系，图4说明，一个试图进入系统的行为结果可能由多种前提所产生，其中，第一阶段由五种不同的行为构成，并且只要有一种行为能成功地施行即可进入第二个阶段。而这两个阶段又可能是构成一个完整的入侵行为中的一个部分，那么第二个阶段即成为进入下一阶段的前提。综合上面的讨论可以得出一个从最初的试图入侵开始，到入侵最终结果的一个多层次的网络。

在这个网络中，可以根据以往的系统应用中得到的经验对于不同的中间过程给予不同的初始权重，例如，假设由于配置的失误导致的ROOT权限被窃取的可能发生概率比较小，则可以对于配置失误设定较低的权重，对于通过窃取密码取得ROOT权限的发生概率比较高，则可以对于窃取密码环节设定较高的权重。由于在预测分析

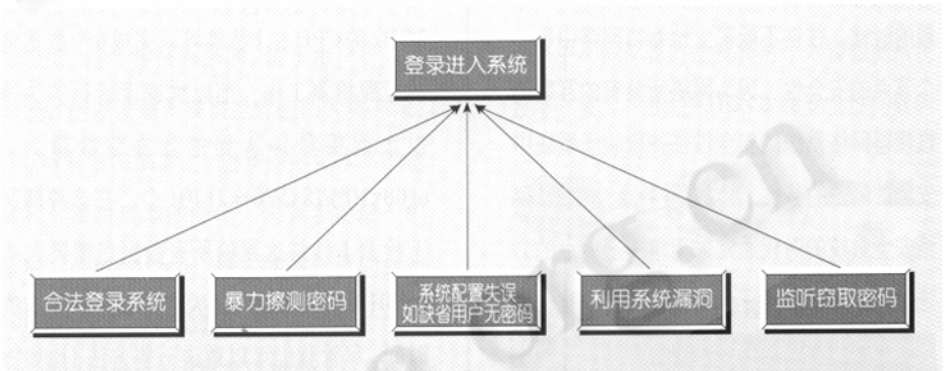


图4 分阶段的入侵行为

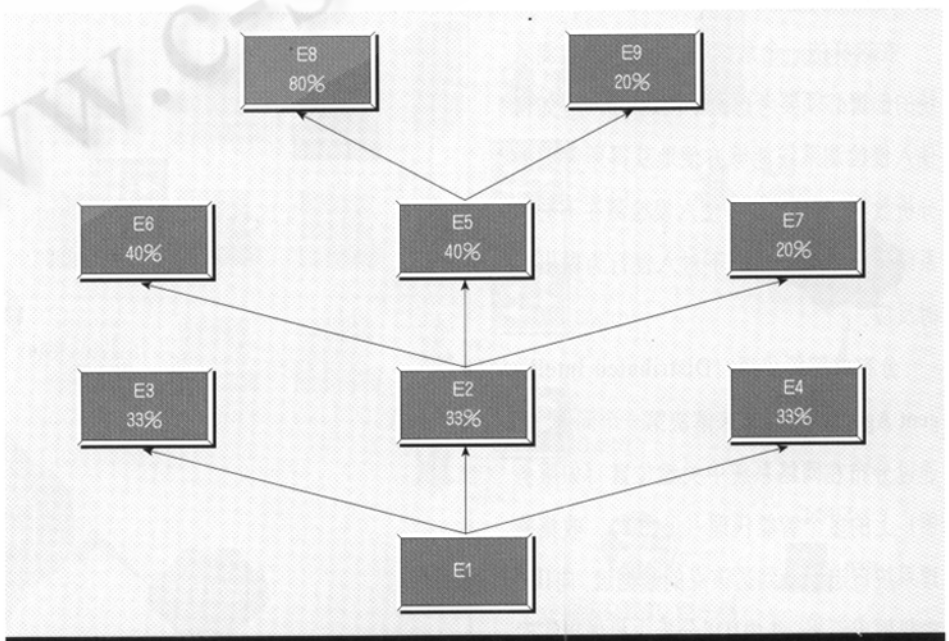


图5 SYN Flood入侵检测实例

| 入侵行为/步骤 | IDS系统响应 | 实验过程 |
|---|---|---|
| 使用nmap工具确认目标系统的操作系统。 | 1. 确认数据包使用协议为TCP<br>2. 在TCP行为中寻找确认操作系统行为<br>3. 如果在确认操作系统行为不存在前趋节点，则建立入侵序列，确认操作系统作为该序列的头节点。<br>4. 寻找确认当前行为的所有后继节点<br>5. 取得所有后继节点发生的可能 | ES={NULL};<br>C=E1;<br>Add C to ES<br>ES={E1};<br>P(E2\|ES)= 0.33<br>P(E3\|ES)= 0.33<br>P(E4\|ES)= 0.33 |
| 使用端口扫描工具进行TCP端口扫描，寻找开放的端口及服务。 | 6. 重复确认协议，在入侵行为模式中寻找、匹配节点的过程<br>7. 端口扫描行为加入到入侵序列中，作为确认操作系统行为的后继节点<br>8. 重复4、5的过程 | C = E2;<br>Add C to ES<br>ES = {E1,E2};<br>P(E5\|ES)= 0.4<br>P(E6\|ES)= 0.4<br>P(E7\|ES)= 0.2 |
| 对开放端口发动SYN Flood Dos攻击 | 9. 重复6的过程<br>10. 将SYN请求连接行为作为端口扫描的后续节点加入到入侵序列中<br>11. 重复4、5的计算过程<br>12. 后继节点E8（SYN Flood）发生的概率超过阈值（0.75），产生警报 | C = E5;<br>Add C to ES<br>ES = {E1,E2,E5};<br>P(E8\|ES)= 0.8<br>P(E9\|ES)= 0.2<br>MAX(P(E8\|ES),P(E9\|ES))= 0.8》0.75<br>Alert begin |

表1

过程中信息是不断的变化，则需要根据系统受到入侵手段的变化动态的调整不同的权重。我们以一个完整的SYN Flood入侵为例，说明IDS系统的工作流程。

## 4 结论

针对危害网络安全的入侵行为进行检测是当今网络安全领域的一个热点问题，本文提出了一种兼顾效率与性能的旁路IDS结构，可适用于安全性要求从低到高的各种应用场合。对于入侵检测系统中的核心问题——入侵模式的识别，作者分析了多种实际入侵行为，提出入侵行为是有步骤发生的，其攻击过程首先潜伏在大量正常访问中，而且操作和攻击结果构成了一个分层网络化结构。基于该分层网络结构，可以在发生已知行为/事件A的前提下，将需要预测的下一行为/事件的可能界定在有限的范围内，简化了匹配过程，进而大大减少计算量。我们正在建立典型入侵行为的入侵模式网络，收集它们各自的先验概率，借助于预测生成模型，计算将要发生的各种情况的可能，来对入侵行为做出较好的判断。

基于本文提出的入侵模式网络，改变了与攻击特征进行匹配的思路，因此在可以通过不断扩充与完善入侵模式网络，使用已知的模式来发现未知种类的入侵行为，作者将结合实际数据，继续进行这方面的探索。


参考文献

1 吴秋峰编，自动化系统计算机网络，第8章，机械工业出版社，2001年2月。

2 项子著，防患于未然，中国计算机报2000第8期。

3 Rebeca Bace, Peter Mell, NIST Special Publication on Intrusion Detection Systems。

4 Char Sample, Mike Nickle, Lan Poynter, Firewall And IDS Shortcomings, SANS Network Security, Monterey, CA, October 2000。

5 金波，吴咏炜，邹淳，入侵检测技术综述，http://www.nsfocus.com/。

6 周夕崇，探究安全的利器攻击检测技术，2000.7.31，http://xexploit.css.com.cn/aqjs/content/rqjcyrzjs/15tjaqlq.htm。