蜜罐系统模型的有限自动机*)

马新新　陈　伟　秦志光
(电子科技大学计算机科学与工程学院　成都 610054)

摘　要　蜜罐系统作为一种主动响应的安全技术，利用诱骗技术吸引入侵，对网络数据进行捕捉和控制，及时获取攻击信息并记录入侵过程，研究攻击手段和过程。对该技术的研究已成为目前信息安全领域的研究热点。本文通过确定性有限自动机理论对蜜罐系统模型进行描述，根据确定性有限自动机状态转换图模拟捕捉攻击行为的全过程，为蜜罐系统的设计和部署提供有力的理论依据和论证。

关键词　蜜罐系统，蜜罐，网络攻击诱骗，自动机

Definite State Machine of Honey Pot System Model

MA Xin-Xin　CHEN Wei　QIN Zhi-Guang
(College of Computer Science and Engineering, UESTC of China, Chengdu 610054)

Abstract　Honey Pot System, one of active response security technologies, makes use of trap technologies attracting attacks from rogues and at the same time captures and controls data, records the intrusion information as well as investigates the attacking methods and process. The Honey Pot System has recently attracted many research works. In this paper, we present theory of the finite state machine to describe the Honey Pot System model and simulate the process of capturing the attack action upon finite state machine state transition diagram to provide the theoretical method and reasoning for designing and deploying of it.

Keywords　Honey pot system, Honey pot, Network attack trap, State machine

传统安全防护侧重被动防御，储如访问控制、防火墙、入侵检测(IDS)等都是采用预先设定的安全策略响应入侵的静态方法。但无法应对越来越多的未知攻击，不能适应动态变化的网络安全需求。因此，网络安全的主动防御技术应运而生。与传统的被动防御不同的是主动防御能预测攻击，动态响应入侵行为，并能通过网络陷阱记录新的攻击方式、手段和攻击过程，提取新的入侵模式。蜜罐是一种具有主动性的入侵响应技术。在检测到攻击的情况下通过诱骗技术，将攻击行为吸收到一个预先设置的网络陷阱—蜜罐中，对网络入侵数据进行捕捉和控制及时获取攻击信息，记录入侵过程，研究攻击者的攻击策略、攻击手段和过程。同时，利用蜜罐系统拖延了攻击者对真正攻击目标的攻击，消耗了攻击者的时间和资源。蜜罐作为一种新的主动响应技术在攻击的检测、分析、研究，尤其是对未知攻击的捕捉、分析、研究方面日益显示了其优越性。对该技术的研究成为目前信息安全领域的研究热点。目前国外在蜜罐技术的研究领域有影响的机构主要有 Honeynet Project 和 Honeynet Research Alliance。

1　蜜罐系统基础

1.1　蜜罐系统

蜜罐(HoneyPot)作为一种网络攻击诱骗工具，通过模拟一个或多个没有经过安全加固的操作系统，给黑客提供一个包含漏洞并容易被攻破的系统作为他们攻击的目标。从而吸引入侵者对该网络攻击诱骗系统进行攻击，同时他们的活动不被察觉地记录下来。攻击行为在蜜罐中持续的时间越长，攻击者攻击系统所使用的技术就暴露的越多。同时，蜜罐拖延了攻击者对真正攻击目标的攻击，消耗了攻击者的时间和资源，使需保护的系统得到防护。

1.2　蜜罐的分类与区别

蜜罐的种类很多，通常分为研究性蜜罐和业务性蜜罐。但任何蜜罐总要与攻击者进行交互，因此以蜜罐和攻击者交互程度的高低为分类依据能更好地区分不同蜜罐，根据蜜罐使用的操作系统和提供的服务是否真实本文将蜜罐分为二类：低交互蜜罐(low-interaction)和高交互蜜罐(high-interaction)。低交互蜜罐(Low-interaction honeypot)即有限的交互。该蜜罐的操作系统和提供的服务都是模拟的，攻击者的行为被局限于该模拟层次。例如，模拟的 FTP 服务能监听 21 端口并能模拟 FTP 登录或支持相关的 FTP 命令。低交互蜜罐的优点在于部署和维护的方便。由于蜜罐中的一切都是模拟的而不是真实的，因此被破坏的风险最小。同时，攻击者的行为被限于这些模拟的服务中而无法跨出这个“牢笼”去攻击或破坏其它的系统。构建低交互蜜罐只须安装软件，选择要模拟和监控的操作系统和能提供的服务。与低交互蜜罐不同高交互蜜罐提供给恶意攻击者的是真实的操作系统和应用服务。因此构建高交互蜜罐系统考虑的因素要比低交互蜜罐复杂。例如，要建立一个提供 FTP 服务的 Linux 系统蜜罐，相应的要建立一个运行真实 FTP 服务的 Linux 系统。高交互蜜罐的优点在于：第一，通过建立能和攻击交互的真实系统可以获取更多的攻击信息，更全面地掌握攻击行为诸如攻击者所使用的工具、攻击策略和方法等；第二，高交互蜜罐提供真实的开放环境，因此无须对攻击者的行为做出预先的假设同时还可以了解未知的攻击行为，进而分析攻击行为以提取未知入侵模式。低交互蜜罐和高交互蜜罐的比较如图 1 所示。

*)基金项目：国家信息关防与网络安全保障持续发展计划项目(2002-研 1-B-007)。马新新　博士研究生；陈　伟　讲师，博士研究生；秦志光　教授，博士生导师，主要研究方向开放系统、中间件、信息安全。

低交互蜜罐(Low-interaction) 模拟的操作系统和服务	高交互蜜罐(High-interaction) 真实操作系统和服务
·易于安装和部署 ·模拟服务能控制攻击者的行为,风险最小 ·只能捕捉有限的攻击信息	·安装和部署复杂 ·提供真实的系统和服务相应增大了风险。 ·能捕捉更多的攻击信息。如攻击者所使用的工具,攻击策略和方法,击键等。

图 1 低交互蜜罐与高交互蜜罐的比较

2 蜜罐系统的模型

针对上述蜜罐的类别我们建立蜜罐系统模型。图 2 所示为蜜罐系统的应用模型,蜜罐系统由以下各组件构成:IDS 组件,firewall 组件,路由器控制组件,log 日志组件。图中所示的各组件可根据具体需求决定是否部署或部署在一台或若干台主机上。蜜罐系统从功能上主要分数据控制和数据捕捉两部分。其中数据控制分为连接控制和路由控制。用来对入侵者的攻击行为进行监控以防止蜜罐系统遭攻陷后被入侵者利用做为攻击其他系统的跳板;数据捕捉由 firewall 组件、IDS 组件、蜜罐主机配合协同完成。Firewall(第一层次的访问控制)组件负责控制进入和外出蜜罐系统的连接控制。该组件不限制外来的连接,但只允许有限数目的外出连接,当外出的连接数量达到预先设定的阀值上限时,阻断信息包。路由控制充当第二层次的访问控制,位于网络陷阱和连接控制之间。对外出的数据包加以控制,禁止非蜜罐系统的 IP 包路由,并对 ICMP 外出包的流量进行限制。IDS 组件对蜜罐系统内的所有网络数据进行捕捉,并记录保存。日志组件记录系统内部所有进程的活动和用户的操作,并对蜜罐主机日志进行远程服务器存放,保证攻击信息的安全存储。多级分层的安全机制模型确保蜜罐系统的有效部署和实施。

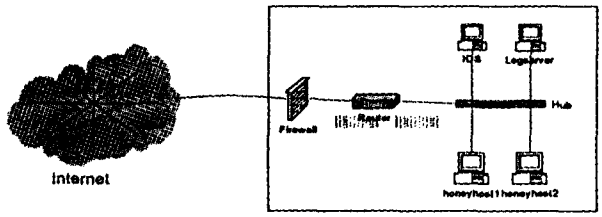

图 2 蜜罐系统模型

3 蜜罐系统的有限自动机

确定型有限自动机 M 是一个五元组:(K,Σ,δ,q_0,F)

其中 K 是状态的有限集合,Σ 是有限的输入字母,δ 是 $K\times\Sigma$ 到 K 一种映射,q_0 是初始状态且 $q_0\in K$;$F\subseteq K$ 是结束状态集合。当自动机 M 处于状态 q,注视输入符号 a 时,根据指令 M 将从状态 q 转到状态 p,记为 $\delta(q,a)=p$ 或简记为 $qa|p$。有限自动机的工作情况可用状态转换图表示。图 3 中每个结点对应于一个状态,若输入符号 a 把状态 q 变成状态 p,则有一条从结点 q 到结点 p 的有向弧,弧上标记字母 a,结束状态用双圈表示。

蜜罐系统的有限自动机 $M=(K,\Sigma,\delta,q_0,F)$,其中 $K=\{q_0,q_1,q_2,q_3,q_4,q_5,q_6,q\}$;$\Sigma=\{0,1\}$;$F=\{q\}$。

映射 $\delta:K\times\Sigma\to K$ 为

$\delta(q_0,0)=q$ $\delta(q_0,1)=q_1$

$\delta(q_1,0)=q_0$ $\delta(q_1,1)=q_2$

$\delta(q_2,0)=q_1$ $\delta(q_2,1)=q_3$

$\delta(q_3,0)=q_3$ $\delta(q_3,1)=q_1$

$\delta(q_4,0)=q_3$ $\delta(q_4,1)=q_5$

$\delta(q_5,0)=q_5$ $\delta(q_5,1)=q_6$

$\delta(q_6,0)=q_0$ $\delta(q_6,1)=q_2$

以上映射关系中若 $\delta(q,a)=p$,且 $p\in F$,则认为 a 被自动机接受。而由自动机 M 接受的所有字的集合记为 $T(M)$:

$$T(M)=\{x|\delta(q,a)\in F\} \quad (1)$$

自动机 M 的状态转换图如图 3 表示。图中各状态的含义为:q_0 为初始状态;q 为结束状态;q_1 为 firewall 连接控制;q_2 为 IDS 网络抓包和报警;q_3 为路由控制;q_4 为蜜罐主机;q_5 远程日志记录;q_6 为入侵分析。蜜罐系统最初处于状态 q_0,当有外来连接时自动机 M 从 q_0 状态转到 q_1 状态,若中止蜜罐系统则自动机 M 从 q_0 状态转到 q 状态;在 q_1 状态,若检测是正常的流量自动机 M 将从 q_1 状态转到 q_0 状态;当检测到是异常的连接时将会将异常转入蜜罐系统,自动机 M 将从 q_1 状态转到 q_2 状态;在 q_2 状态监测网络流量,抓取网络数据包进行分析并将入侵活动引入蜜罐主机,自动机 M 从 q_2 状态转到 q_4 状态;在 q_4 状态将攻击行为局限在"牢笼"中并将其中的所有活动远程备份到 Log 服务器安全存储,自动机 M 从 q_4 状态转到 q_5 状态,若攻击行为突破"牢笼"时,自动机 M 从 q_4 状态转到 q_3 状态,在 q_3 状态禁止非蜜罐系统的 IP 包路由,限制外出数据包,避免蜜罐主机被作为攻击跳板,自动机 M 从 q_3 状态转到 q_2 状态;在 q_6 状态系统管理员对所获取的未知攻击进行分析,提取新的入侵模式并加入到 IDS 组件中,自动机从状态 q_6 转到 q_4 状态;若管理员结束分析,自动机 M 从状态 q_6 转到 q_0 状态。完成对一次异常行为的捕获、分析。

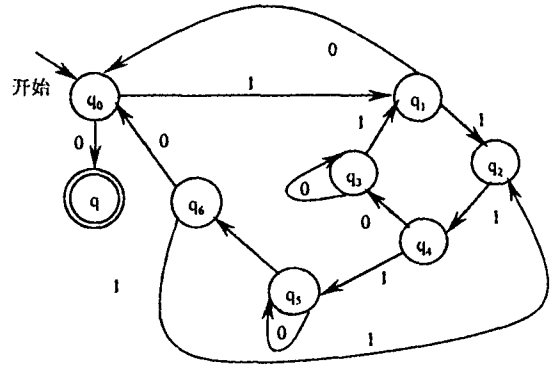

图 3 蜜罐系统有限自动机状态图

由蜜罐系统自动机 M 的状态转换图,利用上述的映射关系并根据公式(1)可以得出蜜罐系统对一次异常行为捕捉完整过程的二进制序列 111110。即此二进制序列在 T(M)中。

总结 蜜罐系统作为一种主动响应的安全技术是网络安全中的一个新兴研究领域,是对现有安全体系的重要扩充和延伸。适应了网络防御技术的主动防御结合被动防御并且建立多层次动态网络安全机制的总体发展趋势。本文通过确定性有限自动机理论对蜜罐系统模型进行描述,根据确定性有限自动机状态转换图模拟捕捉攻击行为的全过程,为蜜罐系统的设计和部署提供有力的理论依据和论证。

基于模糊支持向量机的网络入侵检测研究

李 华 张简政
(重庆大学计算机学院 重庆 400044)

摘 要 模糊支持向量机理论属于统计学习理论,是支持向量机理论的推广,使支持向量机更好地运用到实际工作中。我们将其运用到网络入侵检测中,实验证明是可行的、高效的,有其特点和优势的。

关键词 网络入侵检测,支持向量机,模糊支持向量机,Libsvm

Network Intrusion Detection Based on Fuzzy Support Vector Machine

LI Hua ZHANG Jian-Zheng
(School of Copmputer, Chongqing University, Chongqing 400044)

Abstract FSVM belongs to Statiscal Learning Theory, it exends SVM's application. We introduce it to network intrusion detection. The experiment results proves it is effective, efficient and preponderant.

Keywords Network intrusion detection, Support vector machine, Fuzzy support vector machine, Libsvm

1 引言

随着互联网络的飞速发展,网络安全正日益得到人们的关注,网络安全概念主要涵盖四个方面:保密性、完整性、可用性和认证。由于互联网络在设计实现和使用时候的种种安全问题,以上四个方面的安全无法得到根本保证,这使得入侵事件不断发生,也使入侵检测系统成为当前研究的热点。

根据入侵检测系统的分析数据的来源可以将入侵检测系统分为两大类:基于网络的入侵检测系统(Network-based IDS,简称 NIDS)和基于主机的入侵检测系统(Host-based IDS,简称 HIDS)。根据检测所使用的方法,IDS 还可以分为两种:误用检测和异常检测。误用检测需要建立入侵者的行为模式,异常检测需要建立正常用户的行为模式。由于各有各的优缺点,它们在不同的安全政策中有不同的应用。

入侵检测可以看作是一个分类问题,也就是对给定的审计数据进行分类:什么样的数据是正常的,什么样的数据是异常的。在有关文献中,Forrest[6] 等人把入侵检测看作是区分"自我"(也就是"正常")和"非自我"(也就是"异常")的过程,提出了基于免疫模型的入侵检测技术。Ghosh[7] 利用神经网络来提取特征和分类. W. Lee[8] 从数据挖掘技术的角度探讨了入侵检测的实现问题。以上方法都需要大量或者是完备的数据集才能达到比较理想的检测性能,并且训练时间长,对数据的规律性要求比较高。那么,如何在小样本,入侵特征呈现高维性的入侵检测中,取得更好的检测性能呢?

支持向量机(support vector machines, SVM)是一种建立在统计学习理论基础之上的机器学习方法。其主要的特点是根据 Vapnik[4] 结构风险最小化原则,尽量提高学习机的泛化能力,即由有限的训练集样本得到小的误差仍然能够保证对独立的测试集保持小的误差。另外,由于支持向量算法是一个凸优化问题,所以局部最优解一定是全局最优解。这是其他学习算法所不及的。

为了将支持向量机理论的进一步推广运用到实际中去,Chun-Fu Lin[1] 将模糊关系引入到支持向量机中,在分类问题上,更加注重异常数据的分类情况,进一步提高了检测异常数据的准确率。在理论上得到了验证,但在实际运用中还不够。本文首次将该算法引入到网络入侵检测中来,实验结果表明,该算法可行,有效,特点突出。

本文首先介绍了支持向量机,模糊支持向量机原理,在此基础上,将模糊支持向量机引入到网络入侵检测中,最后利用构造的数据集对模糊支持向量机算法进行实验,给出实验结果和分析。

2 支持向量机

支持向量机(SVM)是从线性可分情况下的最优分类面发展而来的,基本思想可用图 1 的二维情况说明。

图中,白色点和黑色点代表两类样本,H 为分类线,$H1$,$H2$ 分别为过各类中离分类线最近的样本且平行于分类线的直线,它们之间的距离叫做分类间隔(margin)。所谓最优分类线就是要求分类线不但能将两类正确分开(训练错误率为 0),而且使分类间隔最大。分类线方程为 $x \cdot w+b=0$,我们可以对它进行归一化,使得对线性可分得样本集:

$$(x_i, y_i), i=1,\cdots,n, y\in\{+1,-1\} \quad (1)$$

李 华 副教授,硕士生导师,主要研究方向:网络和信息安全、综合网络信息技术;张简政 硕士,主要研究方向:网络安全。

参 考 文 献

1 何成武. 自动机理论及其应用. 北京:科学出版社,1990
2 秦志光,刘锦德. 安全系统的有限自动机. 电子科技大学学报,1996,25(1):71~75
3 宋荆汉,朱伟,等. HoneyPot 系统研究. Net Security Technologies and Application,2002. 1
4 赵伟峰,曾启铭. 一种了解黑客的有效手段——蜜罐(Honeypot). 计算机应用,2003
5 锁廷锋,马士尧. 网络欺骗技术. 信息网络安全,2003
6 Honeynet Project. http://www. honeynet. org