摘要

近几年来，随着经济与科技的不断发展，校园网的规模得到了迅速的增长，但同时校园网的安全问题也变得越来越突出。本文针对山西大学商务学院的网络环境和面临外网的安全威胁；同时，校园网中一些学习网络知识的学生，尤其是信息安全专业的学生，他们具有较高的网络知识水平，并且人数众多，求知欲强，将网络攻击作为他们检验自己所学知识的一个途径，因而他们也成为校园网中主要潜在的不安全因素。通过使用序列比对蜜罐系统解决受到来自校内外的网络攻击。其目标是：1）用蜜罐系统收集对校园网攻击的数据，并且进行分析和提取；根据整理好的数据对校园网攻击进行针对性的防范，减少和消除网络中出现的问题，解决了现实中山西大学商务学院校园网中不安全因素。2）在蜜罐系统中提取的宝贵数据，可以不断提高网络防御技术，同时对今后的教学和科研工作都有很大的帮助。

大学校园网安全是一项系统工程，基于核心交换机的安全策略是其中比较重要的一项，只要长期有效地坚持，合理地配置好网络设备，再将防火墙、入侵检测系统、网络杀毒、蜜罐等系统纳入，就可以构筑立体动态的有效、安全、灵便的安全网络空间，为教学和科研提供安全高效的保障。网络管理方案要充分考虑各种情况，根据不同情况采取相应的措施。针对山西大学商务学院网络安全的需要，为保护商院网络制定了相关的管理制度和目标。

本文在商院网络安全管理体系模型的基础上，设计了一种基于序列比对算法的蜜罐系统，进而协助商院校园网站服务器抵御来黑客的攻击。蜜罐系统（Honeyd）是一款用于创建网络上虚拟主机的蜜罐后台程序，通过配置可以为虚拟的主机提供任意服务。针对过去 Honeyd 插件接口是手工编写脚本程序为其模拟网络服务。但由于各种操作会使脚本的编写比较复杂，可用的脚本非常有限。为了增强 Honeyd 的仿真能力，针对该问题本文设计了一个脚本自动产生模型。模型的实施过程中，不依赖于已知有关服务或协议的任何信息。该模型有自学习的阶段，通过自学习阶段得到了各种参数，从而提高网络防御过程的准确性。通过对重排蜜罐系统获取的会话信息进行提取，将有用的信息作为模拟真实服务器的回应信息。这样省去了为模拟特定服务而编写特定脚本的过程。模型实现过程

中，使用了两个算法 1）序列比对算法，通过该算法更新和提高了蜜罐系统中服务器的信息的能力，其思想是获取协议会话过程的语义信息，从而化简会话状态机；2）本文提出区域分析算法，利用单模式匹配算法 RK，只针对固定区域，而把变异的区域看作一些未知字符的集合。考虑到会话信息中频繁出现的相同字节区域及有可能含有特殊语义信息的因素，利用该算法计算模型中一个新到来的消息和状态机中某个转换消息之间的相似度值，选择相似度值最大的转换消息指向的状态作为一个新的回应消息状态，提高模型提取信息的有效性。

最后，分别对模型参数和系统有效性在商院校园网中进行了实验验证，测试结果表明，该系统具有一定的有效性。

关键词：校园网络安全；蜜罐系统；自动脚本产生；序列比对

Abstract

With the further development of science and economy, there has been a rapid growth of campus networks, and correspondently the security of campus network is becoming a major problem and is drawing increasing attention. The dissertation focuses on the network environment and external threats facing the campus network of Business College of Shanxi University. Apart from the external threats, it also faces some threats from internal factors such as some of college students, who acquired some professional knowledge about network, are eager to do some kinds of attack the campus network as a way testing what they have learned. In my dissertation, a way of a honeypot system based on the sequence alignment algorithm, is utilized to fight back attacks internal and external. By doing this the following purposes can be achieve:1) collect and analyze data on campus attacks so as to diminish and eliminate networks problems and safeguard the security of campus network 2) The collected data can be used in the future teaching process.

Security of campus network is a systematic task, where security strategy based on core exchange plays an important role, only by allocating network devices, and Fire walls、intrusion detection systems、network antivirus、honeypot systems, can we set up an effective, secure, and dynamic net space, and therefore provide secure teaching and studying environment. Based on the specific situation of Business College of Shanxi University, specific objectives and related management principles have been formulated.

This dissertation, based on the model of network security management system, designs a honeyd system based on the sequence alignment algorithm to defend the attacks from hackers. Honeyd is a daemon which can create

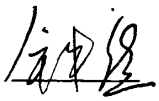
virtual hosts on Internet. We can provide any service for the virtual host through configuring the Honeyd. In order to enhance the simulate capacity of the Honey, people always write script by hand for simulating the network services using the plug in interface provided by Honeyd. However, due to the diversity of the fingerprint for different operate systems, the script can be used are very limited. We designed an automated script generation model. In the model we don't depend on any known information about the daemon implementing the service, nor about the protocol, simply simulate the real server's response through replay and obtain the useful information from protocol session. So we needn't write the script for the service we prepare to simulate. During the implementation of the model, we use two algorithms: 1) sequence alignment algorithm, we use it to obtain the semantic information from the protocol session and then simplify the conversation state machine; 2) region analysis algorithm, which is a new algorithm we provide in this paper. We also consider the conversation bytes which are frequently appear always have the special semantic information, so we use the region analysis algorithm to simplify the conversation state machine again. This can improve the effectiveness of the model.

At the end of my dissertation effectiveness of the system and model parameter have been tested in the campus network and the system has been proved effective.

Keywords: campus network security; honeypot system; automatic script generation; sequence alignment

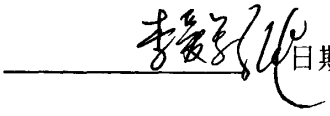
独 创 性 声 明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京工业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签名：  日期：_____

关于论文使用授权的说明

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

签名：  导师签名：  日期：_____

第1章 绪论

1.1 论文研究背景和意义

1.1.1 校园网面临的安全威胁

随着网络在校园中普及，校园网已经成为我国高校基础建设的重要组成部分。充分利用和开发校园内各类信息资源,实现系院之间、大学之间的资源共享,科学计算和科研合作,促进大学对外交流等方面都起到了不可估量的作用。对信息的快速处理、教育资源的配置、利用网络资源提高教学效率、减轻教师的工作负担等提供了较多便利。但是在网络中资源共享和网络安全一直处于矛盾的对立面。在高校中网络信息安全问题日益突出。安全问题的定义，信息安全的含义主要指：信息的完整性、实用性、安全性、可靠性、不可否认性和可控性[1]。

①完整性：保证信息的完整性是信息安全的基本要求，对信息在存储或传输过程中保持不被修改、不被破坏和不丢失的特性，破坏信息的完整性则是对信息安全发动的目的之一。

②实用性：指授权实体对信息资源的正常请求能够及时、准确、安全地得到服务和响应。对实用性的攻击则是阻断信息的可用性。例如在网络环境下破坏网络上有关系统的正常运行就属于这种类型的攻击。

③安全性：指对非授权的用户、实体或进程的信息不泄露，获取和访问只能是授权者。这是信息安全最重要的要求。

④可靠性：指保证信息系统能以被人们所接受的质量水准待续地运行。

⑤不可否认性：在信息系统的信息交互过程中，确信参与者的真实同一性，即所有参与者都不可能否认或抵赖曾经完成的操作和承诺。利用信息源证据可以防止发信方不真实地否认已发送信息，利用递交接收证据可以防止收信方事后否认已经接收的信息。

⑥可控性：是对信息及信息系统实施安全监控。

对于校园网络中经常遇见的问题：1）非法盗版资源众多，容易使恶意代码侵入和利用隐藏在网页中；2）一些用户利用校园网免费资源下载媒体、游戏、软件等，占用了大量的网络带宽，影响校园网的正常应用；3）互联网中的非法

内容对在校学生造成不良影响，许多问题摆在面前，对校园网的计算机系统管理非常困难。

1.1.2 网络安全的技术

通过分析网络的现状，结合山西大学商务学院的实际情况，对面临校园网络安全问题提出了以下几点：校园网接通 Internet 后，对网关需要进行防护，防止网络内部进行攻击；网络病毒容易对校园网络安全造成巨大影响，可能造成操作系统崩溃、数据丢失损坏、网络瘫痪等严重后果；出于校园网具有独立的信息发布系统，所以需要对这些系统进行集中的安全防护，因此重点保证这些系统正常的运行；其次，对各种先进技术应积极采用，如虚拟交换网络(VLAN)、防火墙技术、加密技术、虚拟专用网络(VPN)技术、PKI 技术等，并实现集中统一的配置、监控、管理；最后，应加强有关网络安全保密的各项制度和规范的制定，并予以严格实行。为了便于分析网络安全风险和设计网络安全解决方案，我们采取对网络制度分层的方法，并且在每个层面上进行细致的分析，根据风险分析的结果设计出符合具体实际的、可行的网络安全整体解决方案。

在一个网络中防火墙是使用最多的安全设备，是网络安全的重要基石。防火墙厂商为了占领市场，对防火墙的宣传越来越多，越宣传越强，让一些人无形的形成依赖，市场出现了很多错误的东西。其中一个典型的错误，是把防火墙万能化。但防火墙的攻破率已经超过 47%。正确认识和使用防火墙，确保网络的安全使用，研究防火墙的局限性和脆弱性已经十分必要。同时这种安全都是被动防御，而蜜罐技术就是采取主动的方式。顾名思义，就是用特有的特征吸引攻击者，同时对攻击者的各种攻击行为进行分析并找到有效的对付办法。

1.1.3 校园网利用蜜罐技术的现实意义

基于以上情形，山大商务学院曾专门立项进行网络安全管理系统化设计和实施，本文在此基础上针对服务器安全配置中的不足，设计并实现一种序列比对蜜罐系统，从而弥补了服务器的一些漏洞，更加有效的保障了服务器的安全。针对该问题，本文设计了一个脚本自动产生模型。模型的实施过程中，不依赖于已知有关服务或协议的任何信息，只需通过重排蜜罐系统获取的会话信息并从中获取有用的信息即可模拟真实服务器的回应信息。这样省去了为模拟特定服务而编写

特定脚本的过程。

1.2 国内外研究现状

在1990年出版的一本小说《The Cuckoo's Egg》中 “蜜罐”一词是最早出现，在这本小说中讲述了作者是一个位网络公司的管理员，在一起商业间谍的事件中是如何追查和发现的故事。“蜜罐”一词从出现到今定义有许多种：1）Lance Spitzner 是“蜜罐项目组”（The Honeynet Project）的创始人，他权威给出的蜜罐定义：蜜罐是一种安全资源，其价值在于被扫描、攻击和攻陷[2]。从定义中可以看出蜜罐并没有什么实际作用，只是对经过蜜罐网络数据的流量都进行了扫描、攻击和攻陷。而对这些攻击活动进行监视、检测和分析的作用是蜜罐的核心价值。2）同样蜜罐也可以这样定义：它是一种其价值在于被探测、攻击、破坏的系统[2]。也就是说蜜罐是一种可以用来监视和观察攻击者行为的系统，所以它的设计目是为了使从更价值的系统避免攻击而将攻击者的注意引开，或者说是对网络入侵提供的一种及时预警系统[10]。

蜜罐概念的从出现到今天，“蜜罐”还只是停留在一种思想层面上，通常网络管理人员才进行应用，为达到追踪的目的对黑客欺骗一种手段。此阶段的蜜罐实质上是一些真正被黑客所攻击的主机和系统。

1998 年开始蜜罐技术在实际的网络中得到了应用，网络安全研究人员针对蜜罐思想开发出一些开源工具，其专门用于欺骗黑客，如 Fred Cohen 所研究的DTK（欺骗工具包）、Niels Provos 研究的 Honeyd[3][4][5]等等，同时也出现了一些用于商业像 KFSensor、Specter 等的蜜罐软件产品。本阶段的开发的这些蜜罐工具可以称为是虚拟蜜罐，即它能够将现实的操作系统和网络服务模拟成虚拟，并对黑客做出的攻击动作模拟成真实的回应，从而对黑客进行欺骗。这样对部属蜜罐也变得比较方便。

由于虚拟蜜罐工具是存在交互程度低、容易被黑客识别等问题，安全研究人员在 2000 年之后对蜜罐更倾向于实用真实的主机、操作系统和应用程序进行搭建，与虚拟蜜罐不同之处是，将蜜罐纳入到一个完整的蜜网体系[6][7][8] 中，并且对此工具融入了更强大的数据捕获、数据分析和数据控制，对追踪侵入到蜜罐中的黑客并对他们的攻击行为，研究人员能够方便地进行分析。

迄今为止国内只有北京大学计算机研究所信息安全工程研究中心在做蜜罐方面的研究，在2004年9月该中心启动了一个称作狩猎女神的项目[9]，2004年12月发布狩猎女神项目网站，2004年12月，部署了一个Gen II蜜网，并连入因特网，2005年1月26日，提交加入蜜网研究联盟申请，2005年2月12日，Mr Lance Spitaner，蜜网项目组和蜜网研究联盟的创始人，宣布联盟接受狩猎女神项目。2005年12月20日，发布Walleye Attack and Vulnerability Information Patch，2007年11月，项目组发布开放课题，广招研究人员参与并开源发布研究成果。狩猎女神项目（Chinese Honeynet Project）是北京大学计算机研究所信息安全工程研究中心推进的蜜网研究项目。网络与信息安全技术的核心问题是对计算机系统和网络进行有效的防护。网络安全防护涉及面很广，从技术层面上讲主要包括防火墙技术、入侵检测技术,病毒防护技术,数据加密和认证技术等。在这些安全技术中，大多数技术都是在攻击者对网络进行攻击时对系统进行被动的防护。而蜜罐技术可以采取主动的方式。顾名思义，就是用特有的特征吸引攻击者，同时对攻击者的各种攻击行为进行分析并找到有效的对付办法。狩猎女神项目的研究目标包括如下三个方面：

1、通过部署蜜网，对恶意代码及黑客攻击行为进行捕获和分析，给入侵检测与关联研究提供知识和数据基础。

2、为学生提供一个网络攻防对抗的实验环境，使他们在部署蜜网以及利用蜜网对攻击活动进行分析的过程中，提高实践动手能力以及加深对网络攻防对抗技术的理解。

3、在掌握现有的蜜网技术的基础上，能够在相关的一些研究方向提出自己的观点和看法，并加以实现，促进蜜网技术的发展。目前，狩猎女神项目部署的蜜网融合了“蜜网项目组”提出的第二代蜜网框架与蜜罐虚拟蜜罐系统，此外，狩猎女神项目提供了虚拟蜜网作为学生的网络攻防对抗技术实验平台。

1.3 本文主要工作

本文由蜜罐、蜜罐技术和蜜罐系统的概念入手，分重点分层次的提出自己的思想，并针对山大商务学院校园网安全系统中服务器的安全而设计，研究基于序列对比技术的蜜罐系统的设计与实现。论文重点做了以下几个方面的工作：

1、给出了山大商务学院校园网安全方案

2、对山大商务学院校园网安全蜜罐系统的设计

3、对 PI（Protocol Informatics 信息协议）中应用的生物信息学中的序列比对算法进行了分析；

4、提出了 Honeyd（蜜罐系统）脚本的自动生成模型；

5、通过实验证明了模型的参数变化及在真实环境中模型的有效性。

1.4 论文组织结构

论文结构如下：

第 1 章　主要综述了本文的研究背景和意义，蜜罐技术的国内外研究现状及本文的主要研究内容；

第 2 章　介绍商院网络环境以及管理方案；

第 3 章　校园网安全蜜罐系统的设计并分析 Honeyd 的工作原理；

第 4 章　基于序列比对方法的蜜罐脚本产生模型的设计，该模型由四个模块组成，旨在为 Honeyd 提供使攻击者可信的脚本信息；

第 5 章　实验分析，分别验证了模型参数的变化情况和在商院校园网环境中模型的有效性。

第 6 章　总结和展望

研究的过程也是不断地思考、实践验证的过程。通过对山西大学商务学院校园网分析，结合蜜罐的研究，为今后在该领域提出了新的目标。在论文的最后列举了一些下一步想要做的工作和要解决的问题。

第 2 章 商院网络安全管理方案

2.1 商院网络环境

山大商务学院的网络结构如图 2-1 所示：

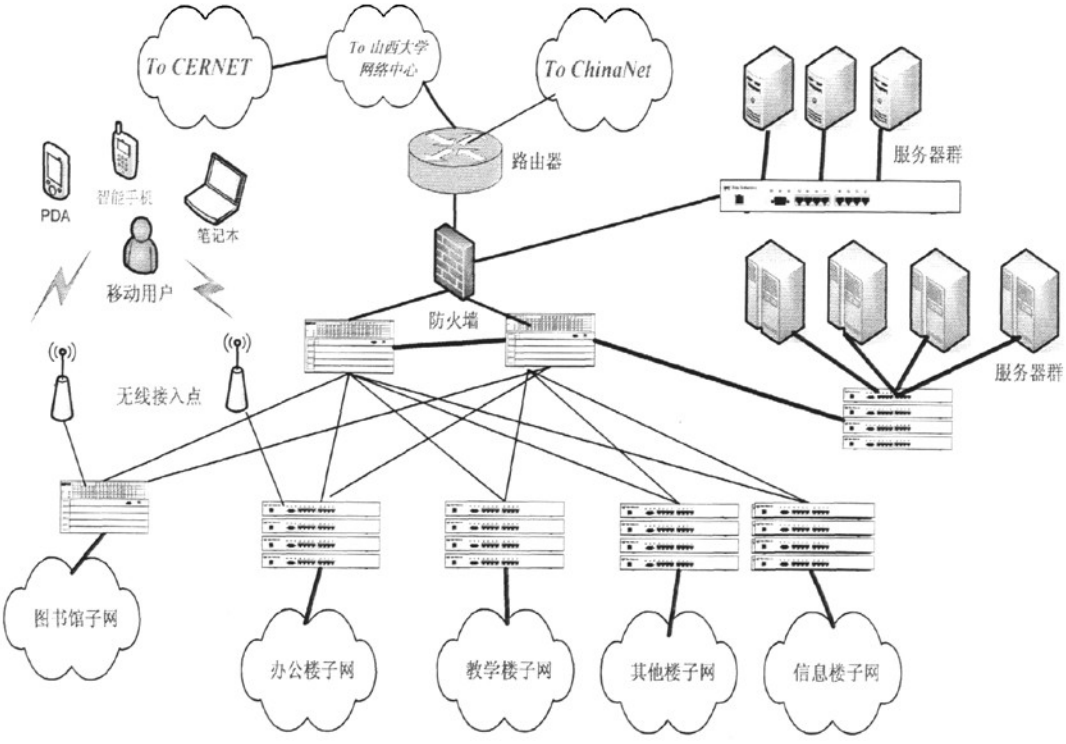


图 2-1 山大商务学院网络结构图

Figure 2-1 Business College's network structure

由结构图可知，该校园网由以下几部分组成：图书馆子网、办公楼子网、教学楼子网、信息楼子网、其他楼子网。图书馆、办公楼、教学楼、信息楼、其他楼组成了校内网，他们之间的相互访问通过设在网络中心的服务器群控制，这些校内网用户可以访问因特网，在子网用户和路由器之间设置了防火墙。下面介绍三种需要保护的网络：

1、无线接入网

无线接入网是全部或部分替有线本地环路，应具备通话质量高、可靠性高、保密性强、容量大，而成本要低、维护方便等，为支持灵活的工作方法，山大商

务学院提供了小范围的群移动用户，使用 NetFlow/IOS 作为无线接入网的主要设备。

此种访问方式的主要隐患在于：非授权用户有可能通过无线接入网接入服务器访问学校内部网络。破坏学院相关数据，因此必须限制无线接入网。

2、Internet 访问

山大商务学院为校外居住的教师提供了一个 ISP（Internet Service Provider，互联网服务提供商）电路连接到校园网上的一台路由器。校外的教师和出差人员就可以通过连接访问学院。这个主机叫做堡垒主机，它是一台完全暴露给外网攻击的主机。它没有任何防火墙或者包过虑路由器设备保护。堡垒主机执行的任务对于整个网络安全系统至关重要。事实上，防火墙和包过虑路由器也可以被看作堡垒主机。由于堡垒主机完全暴露在外网安全威胁之下，需要做许多工作来设计和配置堡垒主机，使它遭到外网攻击成功的风险性减至最低。其他类型的堡垒主机包括：Web、Mail、DNS、FTP 服务器。一些网络管理员会用堡垒主机做牺牲品来换取网络的安全[11]。这些主机吸引入侵者的注意力，耗费攻击真正网络主机的时间并且使追踪入侵企图变得更加容易。 该主机运行着含有山大商务学院信息的几种软件系统。同时内、外部用户可以相互转发电子邮件。

3、校园网访问

校园网访问所涉及的部门有：电教中心、办公楼的各个学院、各系实验室、图书馆、信息安全实验室。他们都有自己单独的服务器，其中信息安全实验室有一台集中网络管理的 Windows Server 2003 互联网服务器，负责管理、监督网络基础设施、服务器和工作站之间的连通性，它可以通过 Telnet 协议访问所有的网络设备。对于校园网访问，主要隐患在于校园内的用户操作失误、存心捣乱的关键数据泄密以及不良影响。

2.2 山大商务学院网络的安全目标和设计原则

通过对以上情况的分析，山大商务学院制定了相应的网络安全目标：

（1）网络中心应配置功能较强的网络管理软件，可实现对校园网主要设备的远程管理。有权限设置、虚网 VLAN 划分、流量检测、故障报警等功能；

（2）防火墙＋入侵检测和防范系统：可实现对来自外网和内部攻击的防范，

防止校园内部敏感数据被窃取和篡改；

（3）防病毒软件（网络版）：防止病毒入侵校园网，可自动在线升级病毒库定义，最好能在线更新工作站端口的防病毒软件（在线分发）；

（4）电子邮件管理软件，可进行帐户管理、费用管理、内容过滤、反垃圾邮件、系统日志等功能

（5）用户管理系统：对校内用户进行身份识别、权限分配和设置，

（6）各系统应能提供完善的日志记录功能：对用户的对外访问进行控制和记录，用户记录应能保存60天以上。

（7）做好数据的备份工作。

山西大学商务学院校园网设计方案的原则为：

（1）先进性：校园网应由先进的网络体系结构和先进的网络设备构成，能够满足校园网教学、科研和管理应用的需求。

（2）标准化：校园网建设应采用国际标准或事实上的工业标准，支持网络互联的开放标准，可实现多协议转换，便于不同厂家产品互连。

（3）实用化：采用先进而成熟的技术，不落后也不奢侈，够用并留有发展余地，不追求最新的或未经实践的技术。

（4）高的可靠性：网络系统和设备选型应该能够提供接近电信级的安全性能，即保证校园网安全运行率在99.9%以上。

（5）安全性：网络系统应具有较高的安全性，提供高性能的用户身份认证和访问授权管理，有良好的防病毒、防黑客、防垃圾邮件、内容过滤功能。

（6）较高的性能价格比：少花钱、多办事，尽可能减少不必要的投资。良好的可扩充性：可保护前期设备投资和支持将来的发展。

2.3 商院网络管理方案

通过以上的安全目标和设计原则，为山大商务学院具体安全管理措施如下：

2.3.1 根据用户的特性和需求划分无线接入网

无线接入技术(也称空中接口)是无线通信的关键问题。它是指通过无线介质将用户终端与网络节点连接起来,以实现用户与网络间的信息传递。无线信道

传输的信号应遵循一定的协议,这些协议即构成无线接入技术的主要内容。无线接入技术与有线接入技术的一个重要区别在于可以向用户提供移动接入业务。

无线接入网是指部分或全部采用无线电波这一传输媒质连接用户与交换中心的一种接入技术。在通信网中,无线接入系统的定位:是本地通信网的一部分,是本地有线通信网的延伸、补充和临时应急系统。

将商院无线接入网络,由于商院的行政办公机构分布在同一办公楼中的不同楼层,因为 WLAN 易安装、易扩展、易管理、易维护,作为有线局域网络地延伸而存在的,各个系、办公单位在构建其办公网络后,使无线网络用户具有高移动性、保密性强、抗干扰等特点。

2.3.2 在校园网出口设置防火墙

防火墙是校园网信息安全保障的核心点,它负责校园网中最基本的信息服务系统的安全,一旦被非法进入,存在着内容被窃取、泄密、篡改、损坏等巨大风险,属于安全等级中最严重的事件。通过部署防火墙,把这些信息系统集中隔离到一个逻辑安全区中,在防火墙集中控制点处定制严格的访问控制策略,实施严格的数据流监控。因为防火墙的安全性源于其优秀的访问控制能力,可做到基于IP、协议、用户的访问控制,能灵活地对服务对象、操纵权限、服务范围进行控制。 同时也对各具体的服务系统从底层操作系统到应用系统做了针对性的安全优化和加强,如停用无关服务、启用系统审计、精简定制系统等。而且对安全性有特殊要求的服务系统,在防火墙和服务系统上也做了较为详细的日志记录,为安全事件的事后取证工作提供依据,当然取证是多因素的综合,也包括其它手段如 IDS、蜜罐等手段的补充。

IDS(Intrusion Detection Systems,入侵检测系统)系统是重要的网络安全诊断工具,可实时监控网络中的异常情况、跟踪安全事件的新动向、统计分析安全历史记录等。IDS 系统通常把探测引擎分布式地部署在网络的各关键点,然后汇总到控制中心进行分析、统计、显示、报警等动作。由于外网的攻击,如网络扫描、蠕虫、DOS 等攻击,只能是被动地阻断或向相关组织反馈,而对攻击源无法处罚,内网则由于可利用的监控手段多,攻击源的定位和控制也相对容易[12]。

校园网所用的防火墙是 FortiGate-300A,FortiGate-300A 有两个 10/100/1000M 自适应以太网接口,可以升级到千兆网络。它有四个用户定义的

10/100M 接口，可以提供冗余的 WAN 连接，高可靠性和多区域的特性，允许管理员在划分其网络的区域时有更高的灵活性和在不同区域之间设置策略[24]。它能有效隔离校园网与外部互联网的连接，使之间的访问连接得到有效控制。针对重要的网段(如院长办公室、教务、财务、人事、科研中心、重要实验室等)可以设置相应的隔离网关，提供最基本的网络层的访问控制。

2.3.3 合理运用入侵检测技术

入侵检测技术是主动保护自己免受攻击的一种网络安全技术。目前，实现入侵检测和防火墙之间的联动有两种方式。一种是实现紧密结合，即把入侵检测系统嵌入到防火墙中，即入侵检测系统的数据来源不再来源于抓包，而是流经防火墙的数据流。但由于入侵检测系统本身也是一个很庞大的系统，从目前的软硬件处理能力来看，这种联动难于达到预期效果。第二种方式是通过开放接口来实现联动，即防火墙或者入侵检测系统开放一个接口供对方调用，按照一定的协议进行通信、警报和传输，这种方式比较灵活，不影响防火墙和入侵检测系统的性能。防火墙与入侵检测系统联动，可以对网络进行动静结合的保护，对网络行为进行细颗粒的检查，并对网络内外两个部分都进行可靠管理。

2.3.4 设置访问控制管理系统和智能信息过滤系统

校园网采用扁平结构设计，主干网采用千兆交换网，网络中心 1#楼宇交换机 1#桌面三级互联，支持 VLAN（Virtual Local Area Network，虚拟局域网）。整个网络系统采用了二/三层交换的混合架构：核心层、汇聚和接入层。

核心层采用双核心冗余备份结构，采用二台高性能的可扩展万兆模块的模块化路由交换机（未来可以平滑的升级到万兆），在交换机间采用互为冗余备份和负载均衡的技术。核心交换机安装在校网络中心，可为汇聚/接入层的交换机提供大量的千兆接口，为出口的网络安全设备防火墙提供连接，为商务学院的校园网提供路由和交换的骨干。当一台核心交换机出现故障时，二台交换机之间会自动进行切换，校园网的用户根本感觉不到故障的存在，使得视频教学、给上级单位的多媒体课程演示等所有的网络活动能够不间断的进行，为校园网的可靠运行提供了强有力的保障。

同时，核心交换机上应配置双电源、双引擎，二个引擎上均有相同的软件配置信息，如果一个引擎出现问题，系统会自动检测到并在毫秒的时间内完成自动的切换，不会导致数据的丢失，从硬件设备上保证了核心设备的稳定可靠性。

在整个网络系统中，建议采用安全网络管理平台，保证商务学院校园网的安全和有效管理，采用支持 802.1x＋Radius＋SNMP 的网络管理平台，实现所有网络设备的集中统一管理、用户授权统一管理以及网络流量的统一监控。其 SAMS 校园网安全管理系统由于其管理严谨、科学合理、功能强大，特别符合高校复杂网络应用环境和高校学生对网络过度应用的情况，能够防止私接代理、带宽无限占用，并能提供无线网络的有效安全管理，大大减轻了网络中心管理人员的工作强度，受到各用户单位的好评。

2.3.5 加强服务器安全

服务器是网络上一种为客户端计算机提供各种服务的高性能的计算机，它在网络操作系统的控制下，将与其相连的硬盘、磁带、打印机、Modem 及各种专用通讯设备提供给网络上的客户站点共享，也能为网络用户提供集中计算、信息发表及数据管理等服务。它的高性能主要体现在高速度的运算能力、长时间的可靠运行、强大的外部数据吞吐能力等方面。是校园网的核心设备，因此要有提高的安全性。windows server2003 是目前最为成熟的网络服务器平台，安全性相对于 windows 2000 有大大的提高，但是 2003 默认的安全配置不一定适合我们的需要，所以，我们要根据实际情况来对 windows server2003 进行全面安全配置。本文则在此基础上从主动防御的角度考虑，设计了一种序列比对蜜罐系统来提前发现攻击，从而及时采取相应的措施来保护服务器。

2.3.6 加强防范，防止病毒泛滥

网络中心为全校教师和学生提供一个统一域名的邮箱账号，目前用户数规模高达 2 万多。根据日常的监控统计，垃圾邮件和高度疑似的垃圾邮件就占了总量的 90%之多，消耗了大量的系统资源，包括 CPU 性能、磁盘空间、内存、响应速度等。垃圾邮件主要分两大类，一类是病毒造成，另一类是广告行为。

对于大量泛滥的广告邮件，采用基于行为分析的垃圾邮件过滤系统，即通过识别垃圾邮件的行为（如分析邮件路由逻辑，反向验证 IP 地址域，辨别仿冒邮件地址等行为特征）进行过滤。事实证明通过这种基于行为方式的过滤，垃圾邮件会大幅度地降低，根据对校园网邮件系统的统计，最后能到达用户的邮件只占邮件总量的 5-10%，而且即使是过滤后的邮件，垃圾邮件也能占到一定的比例[21]。这部分被漏掉的垃圾邮件，一是由于系统不对内容做判断，另一方面是因为有一些行为没能被识别出来造成的。

病毒邮件同样是垃圾邮件，而且占相当大的比例，由于病毒极强的传染力，导致邮件系统成了散播病毒的源头，严重威胁着邮件接收者的终端系统，因此在校园网邮件系统的前端我们部署了邮件防病毒网关，凡是进入邮件系统的信件都要经防病毒网关的过滤后，才最终被送到用户邮箱中。

2.3.7 在防火墙内口上捆绑 IP 和 MAC 地址

IP 地址的修改非常容易，而 MAC（Media Access Control，介质访问控制）地址存储在网卡中，而且网卡的 MAC 地址是唯一确定的。因此，为了防止内部人员进行非法 IP 盗用(例如盗用权限更高人员的 IP 地址，以获得权限外的信息)，可以将内部网络的 IP 地址与 MAC 地址绑定，盗用者即使修改了 IP 地址，也因 MAC 地址不匹配而盗用失败；而且由于网卡 MAC 地址的唯一确定性，可以根据 MAC 地址查出使用该 MAC 地址的网卡，进而查出非法盗用者。

目前，学校校园网都采用了 MAC 地址与 IP 地址的绑定技术。许多防火墙（硬件防火墙和软件防火墙）为了防止网络内部的 IP 地址被盗用，也都内置了 MAC 地址与 IP 地址的绑定功能。这样绑定 MAC 地址和 IP 地址可以防止内部 IP 地址被盗用。

2.3.8 容错和备份

备份是容错的基础，是指为防止系统出现操作失误或系统故障导致数据丢失，而将全部或部分数据集合从应用主机的硬盘或阵列复制到其它的存储介质的过程。传统的数据备份主要是采用内置或外置的磁带机进行冷备份。但是这种方式只能防止操作失误等人为故障，而且其恢复时间也很长。随着技术的不断发展，数据的海量增加，不少的企业开始采用网络备份。网络备份一般通过专业的数据

存储管理软件结合相应的硬件和存储设备来实现。

2.3.9 日志的记录

对校园网安全中网络管理人员注意日志的收集，在出现安全事故时，如果有交换机等设备的日志信息，就可以通过查看和分析日志，来确定攻击来源，进而对漏洞采取措施，使损失降到最低。所以，要建立一个可以运行在 windows 或 Linux 环境下的日志服务器软件可以使用 3cdemon 等。在核心交换机上还要启用日志记录，并指定 syslog 服务器的 IP 地址。

2.3.10 加强内部安全管理

对管理人员安全思想应进行提高，对网络和系统一定要保证运转正常，在此基础上建立严格的校园网管理制度和实验室上机管理制度，对人为因素造成的不安全事情应进行杜绝。整个校园网安全的日常管理及维护应该配备相应的专业管理人员负责。

制订并实施系统备份计划，建立和完善网络故障紧急处理预案，对系统完整的数据做好备份。对网络监控值班和技术值班应实施 24 小时，每一个环节事故处理都要确保，以达到有备无患，保证网络的稳定运行。

第一阶段：预备——盛食厉兵

第二阶段：确定——检查应全面细致

第三阶段：封闭——对失控的局面应冷静处理

第四阶段：删除——彻底的补救措施

第五阶段：复原——恢复备份

第六阶段：追踪——对特定的攻击是否还会有第二次

2.4 本章小结

本章主要介绍了商院网络的总体管理方案，对三种需要保护的网络分别采取了相应的措施来加以保护。网络操作系统是校园网服务器系统中最重要的组成部分，针对 2.3.5 中对服务器的安全设置的不足，本文设计了一种序列比对蜜罐系

统，将该系统布置在靠近服务器的地方，能够提前发现攻击，从而及时采取相应的措施加以阻止，使该方案成为更加适合商院实际情况的，具有可扩展性的、唯一的、最优的方案。下一章将详细介绍蜜罐技术。

第3章 校园网安全蜜罐系统的设计

3.1 蜜罐技术概述

3.1.1 蜜罐的概念

蜜罐的定义是："蜜罐是一个安全资源，它的价值在于被探测、攻击和损害。"此定义包括以下两层意思：

1、蜜罐出现的目的就是要诱骗攻击者的攻击，相反不被攻击者攻击的蜜罐是没有意义的。

2、蜜罐是不用修补造成的任何损伤，它的作用就是尽最大能力捕获攻击者的手段，这些手段中包括受攻击造成的损伤。

蜜罐的设计初衷是通过黑客入侵，以此收集数据，同时让真实的服务器避开攻击，因此一台合格的蜜罐要求拥有一下功能：发现攻击、产生报警、记录能力强大、诱骗、调查。剩下的功能由机房管理员去操作，那就是根据蜜罐收集的证据在必要的时候来起诉入侵者。

蜜罐在网络安全资源中的位置如图3-1所示：

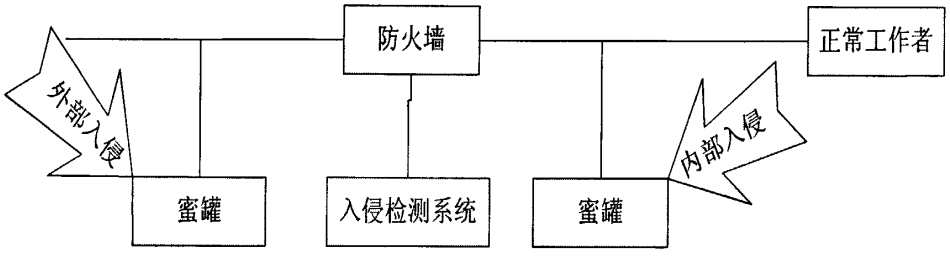


图3-1 蜜罐在网络安全资源中的位置

Figure 3-1 honeypot network security resources in the location of

3.1.2 蜜罐的分类

（一）按照部署目的分类

按照部署目的，蜜罐可以分为产品型蜜罐和研究型蜜罐两类，研究型蜜罐专门用于对黑客攻击的捕获和分析，通过部署研究型蜜罐，研究人员可以对黑客攻

击进行追踪和分析，捕获黑客的键击记录，了解到黑客所使用的攻击工具及攻击方法，甚至能够监听到黑客之间的交谈，从而掌握他们的心理状态等信息。研究型蜜罐需要研究人员投入大量的时间和精力进行攻击监视和分析工作[14]。具有代表性的工具是“蜜网项目组”所推出的第二代蜜网技术[41]。而产品型蜜罐的目的在于为一个组织的网络提供安全保护，包括检测攻击、防止攻击造成破坏及帮助管理员对攻击做出及时正确的响应等功能。一般产品型蜜罐较容易部署，而且不需要管理员投入大量的工作。较具代表性的产品型蜜罐包括前面提到的DTK、Honeyd[45]等开源工具和 KFSensor、ManTrap 等一系列的商业产品[15]。

（二）根据蜜罐与攻击者的交互程度分类

蜜罐也可以按照其交互度的等级划分为低交互蜜罐和高交互蜜罐，交互度反应了黑客在蜜罐上进行攻击活动的自由度。高交互蜜罐提供完全真实的操作系统和网络服务，没有任何的模拟，从黑客角度上看，高交互蜜罐完全是其垂涎已久的“活靶子”，因此在高交互蜜罐中，我们能够获得许多黑客攻击的信息[41]。高交互蜜罐在提升黑客活动自由度的同时，自然地加大了部署和维护的复杂度及风险的扩大。研究型蜜罐一般都属于高交互蜜罐，也有部分蜜罐产品如 ManTrap，属于高交互蜜罐。低交互蜜罐一般仅仅模拟操作系统和网络服务，较容易部署且风险较小，但黑客在低交互蜜罐中能够进行的攻击活动较为有限，因此通过低交互蜜罐能够收集的信息也比较有限，同时由于低交互蜜罐通常是模拟的虚拟蜜罐，或多或少存在一些容易被黑客所识别的指纹（Fingerprinting）信息。产品型蜜罐一般属于低交互蜜罐[37]。

（三）根据实现方法分类

蜜罐还可以按照其实现方法区分成物理蜜罐与虚拟蜜罐。物理蜜罐是真实的网络上存在的主机，运行着真实的操作系统，提供真实的服务，拥有自己的 IP 地址；虚拟蜜罐则是由一台机器模拟的，这台机器会响应发送到虚拟蜜罐的网络数据流，提供模拟的网络服务等。

3.1.3 蜜罐的作用

蜜罐的作用主要表现在下列三个方面：

1、蜜罐可以在抵御攻击上化被动为主动

传统的防火墙和 IDS 只限于对已知攻击的响应，不能预防未知攻击，主动权

掌握在黑客手中，而蜜罐能够诱惑和欺骗攻击者，让他们优先攻击蜜罐而不是世纪的工作系统。蜜罐从捕获的数据中学习攻击者的动机、方法和工具，这样就能更好地理解面临的威胁，赢得了研究对策定时间，掌握了主动权；

2、蜜罐可以有效地收集攻击信息

由于蜜罐不提供真实的系统并将任何对自己的访问都视为入侵，所以收集到的信息都是与攻击有关的，虽然信息量不大，但从中可以迅速找到黑客攻击的证据；

3、蜜罐可以保护周边网络的安全

由于蜜罐转移了攻击者的注意力，让入侵者在蜜罐中逗留了较长的时间，网络管理员就有足够的时间对入侵做出反应，这在一定程度上减小了周边网络被攻击的风险。

3.1.4 蜜罐的优势和缺陷

（一） 蜜罐的优势

1、数据价值

当前，众多的安全组织所面临的一个就是如何从收集到的数据中获得由价值的信息，这些组织每天都收集到大量的数据，包括防火墙日志、系统日志和入侵检测系统所发的告警信息。文献这些数据的量是庞大的，大多都是几兆，甚至以吉比特为单位，从中提出有价值的信息非常困难。而蜜罐收集的数据量很少，但是含金量高，蜜罐的概念决定了可以将噪音降到最低（蜜罐没有任何产品型的功能，任何对他的访问都是非法的、可疑的），这些数据记录都是扫描、探测、攻击，价值非常高。

蜜罐可以迅速、简单、易懂的格式提供所需的信息，这样反应时间大大缩短。例如：Honeynet Project 是研究蜜罐的非官方组织，每天收集到的数据不到 1M。虽然数据量很小，但是包含的信息主要都是可疑的行为。这些数据可以用来做统计模型、趋势分析、检测攻击，甚至研究攻击者。这个过程如同在显微镜下做显微分析，将捕获的数据放在显微镜下做进一步、详细的审查研究[13]。

2、资源

绝大多数安全组织所面临的另一个难题就是资源的限制，有的甚至是资源的

耗尽。例如：当防火墙的状态检测表满的时候，它就会强迫阻断所有的连接。入侵检测系统会因为网络流量太大而丢失数据包。这些都是网络资源耗尽的情况。因为蜜罐需要捕获和监视的网络行为很少，一般情况不会出现资源耗尽的情况。蜜罐只需监视其关心的连接，不存在网络流量太大的问题，所以不需要最新的技术——高容量的 RAM、高速 CPU，这样就意味着配置蜜罐不需要消耗太多的资源。

3、简单性

蜜罐不需要像 IDS 一样维护特征数据库，配置规则库，只要配置好蜜罐放在网络中就可以了。对于一个系统而言，越简单，其工作的可靠性越好。

4、投资回报

当防火墙成功的将攻击者拒之门外的时候，它自己就成了成功的牺牲品。如果投资者在安装了防火墙的几年内，并没有遭受到任何攻击，那么，他们会怀疑安装防火墙的必要性。他们很可能不去想没有遭受攻击的原因正是因为防火墙所致。开发者耗费了时间和资源，而他们都成了成功的受害者。

但是，蜜罐却能够快速而又重复的证明自己的价值。通过捕获未授权行为，不但会让人们知道攻击者存在，而且也证明其他安全资源的投资也是正确的。如果投资者认为不存在威胁，蜜罐会证明大量风险的存在。

（二） 蜜罐的缺陷

1、脚本难以统一

蜜罐可以通过模拟网络上的某个服务来吸引攻击者的注意，当攻击者向其发出攻击时，从中获取攻击者的信息。要达到欺骗攻击者的目的，蜜罐系统必需同时发给攻击者可信的回应信息，而这些回应信息的来源，大多来自于手工编写的服务脚本。而且不同系统中的服务脚本又有差异，很难得到一个标准的服务脚本。而且对于一些专用的协议，也缺少合适的脚本。针对该问题，本文中设计了一种自动生成脚本的模型。

2、视野狭窄

由于蜜罐系统只能部署在个别的机器上，攻击者闯入网络后，可以绕过蜜罐系统，去攻击其它系统，这种情况下蜜罐系统无法发现攻击者。

3、指纹

由于蜜罐会有一些专业特征和行为，使得攻击者能鉴别出蜜罐的存在。这些

特征和行为就是指纹。例如：当攻击者连接到一个模拟 Web 服务器的蜜罐时，正常的 Web 服务器应该回送一个用标准 HTML 语言写的错误标记。但蜜罐模拟的 Web 服务器将一个 HTML 的标记拼错了，例如：将“length”评成“legnth”。这个错误对蜜罐来说就是一个指纹。还有一种就是蜜罐的错误配置也能导致自身被暴露。例如：一个蜜罐本身是想模拟 NT IIS Web 服务的，但是蜜罐却有 Unix Solaris 服务的特征，这些错误的特征就成了蜜罐存在的标记。

4、风险

一旦蜜罐被攻击，就有可能被攻击者利用，作为攻击、渗透其他系统或组织的跳板。低交互性的蜜罐系统，仅仅模拟几个服务，很难被利用作为攻击其他系统的跳板。而高交互性的蜜罐系统，可能就是一个缺省的操作系统，攻击者有可能取得系统的 root 权限，利用蜜罐向其他的系统或组织发动攻击。

通过以上的分析可以看出蜜罐，一方面 Honey 是一台存在多种漏洞的计算机，而且管理员清楚它身上有多少个漏洞，这就像狙击手为了试探敌方狙击手的实力而用枪支撑起的钢盔，蜜罐被入侵而记录下入侵者的一举一动，是为了管理员能更好的分析广大入侵者都喜欢往哪个洞里钻，今后才能加强防御。另一方面是因为防火墙的局限性和脆弱性，因为防火墙必须建立在基于已知危险的规则体系上进行防御，如果入侵者发动新形式的攻击，防火墙没有相对应的规则去处理，这个防火墙就形同虚设了，防火墙保护的系统也会遭到破坏，因此技术员需要蜜罐来记录入侵者的行动和入侵数据，必要时给防火墙添加新规则或者手工防御。

3.2 蜜罐系统的设计

3.2.1 Honeyd 系统简介

Honeyd 是一个小后台程序主要用于创建虚拟的网络上的主机，这些虚拟主机可以随意配置让它们提供特定的服务，利用这个特性可以使得主机显示为在某特定版本的操作系统上运行[17]。

GNU（General Public License）下发布开源软件的 Honeyd 软件。最初面向的是 Linux 操作系统，可以运行在 BSD(Berkeley Software Distribution，伯克利软件套件)系统, Solaris，GNU/Linux 等操作系统上，由 Niels Provos 开发

和维护。最新版本是2007年5月27日发布的Honeyd 1.5c. 应用于windows系统的Honeyd程序也已经出现，其开发者为Mike Davis. 最新版本为windows ports for Honeyd 0.5. 这里主要介绍面向类linux系统的Honeyd程序[18]。

Honeyd可以通过提供威胁检测与评估机制来提高计算机系统的安全性，也可以通过将真实系统隐藏在虚拟系统中来阻止外来的攻击者。因为Honeyd只能进行网络级的模拟，不能提供真实的交互环境，能获取的有价值的攻击者的信息比较有限，所以Honeyd所模拟的蜜罐系统常常是作为真实应用的网络中转移攻击者目标的设施，或者是与其他高交互的蜜罐系统一起部署，组成功能强大但花费又相对较少的网络攻击信息收集系统[44]。

3.2.2 山大商务学院网络系统的设计

山大商务学院的网络安全体系中重点设计运用了蜜罐系统。为了能够有效地实施防御，本蜜罐系统的设计基于Honeyd系统。本章讨论蜜罐系统的设计：

（一）蜜罐系统的设计开发重点解决的问题：

1）如何将发送到虚拟蜜罐的数据引入到Honeyd主机；

2）如何使得模拟主机对攻击者看上去真实可信同时还保证 honeyd 主机的安全；

3）如何模拟任意的网络拓扑；

4）如何支持Honeyd主机利用网络隧道与其他系统中的主机通信；

5）如何记录网络连接和恶意的攻击行为；

6）如何使用尽量简单的配置语法来配置Honeyd。

（二）蜜罐系统体系结构

山大商务学院网络安全蜜罐系统体系由几个组件构成，这些组件是配置数据库、中央包分发器、协议处理器、个性引擎和可选路由构件[43]。如下图3-2所示：

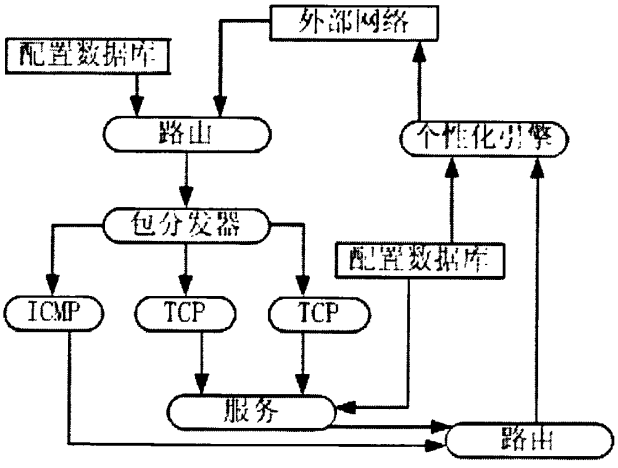


图 3-2 Honeyd 体系结构图

Figure 3-2 Honeyd system structure

系统接受到的数据会由中央包分发器进行处理，首先中央包分发器进行处理会检查 IP 包的长度，修改包的校验和。Honeyd 体系结构响应的是：ICMP、TCP 和 UDP 这最主要的 3 种互联网协议，在记入日志后其他协议包会被悄悄丢弃。

在数据包处理之前，分发器会对配置数据库进行查询，以查找到一个符合目标地址的蜜罐配置。如果对特定的配置没有查询到，系统会采用默认配置模板。在配置确定后，相应的配置和数据包会被转交给相应的协议处理器处理。

多数的查询都能受 ICMP 协议处理器支持。默认情况下，所有的蜜罐配置都会响应 echo 请求，并且处理“destination unreachable”消息。其他请求的处理主要依赖于个性引擎的配置。

对于 TCP 和 UDP 包，Honeyd 可以建立到任意服务的连接。这些服务是外部的应用程序，可以通过标准输入输出来接收和输出数据。不同于为每个连接创建一个新进程，Honeyd 支持子系统(subsystem)和内部服务(internal service)，子系统是一个运行在某个虚拟蜜罐的名称空间下的应用程序，子系统的特定应用是在相应的虚拟蜜罐实例化的时候创建的。一个子系统可以绑定端口、接收连接和创建网络连接。子系统是作为外部进程运行的，而内部服务则是一个 Honeyd 内部运行的 python 的脚本。内部服务要求的资源比子系统更少，但只能接收连接，不能创建网络连接[42]。

UDP 数据报文会直接传递到应用程序，当接收到一个发送到关闭的端口的

数据报文的时候，如果个性化配置中没有设置禁止的话，系统会发送一个端口不可达消息。在发送端口不可达消息的时候，系统允许网络映射工具如 traceroute 来查探网络路由。

除了可以建立到本地服务的连接外，Honeyd 还支持网络连接的重定向。这种重定向可以是静态的，也可以是与网络连接的四个参数相关（源地址与源端口，目标地址与目标端口）。重定向使得我们可以将一个到虚拟蜜罐上的服务的连接请求转发到一台真实服务器运行的服务进程。

在发送数据到外部网络之前，数据包会经由个性引擎处理。个性引擎会修改数据包的内容，使得数据包看上去和从指定配置的操作系统的网络栈中发出的一样。

不同操作系统的网络栈处理各不相同，这导致他们所发送的数据包具有各自不同的特点。网络攻击者常常会使用一些网络指纹识别工具，如 Xprobe、Nmap 等来分析接收到的数据包的特点，从而达到收集目标系统信息的目的。

对蜜罐系统来说，在被指纹识别的时候不要暴露出来是非常重要的。为了使得虚拟蜜罐在被探测得时候显得像真实主机一样，蜜罐模拟给定的操作系统的网络栈行为，我们称之为虚拟蜜罐的“个性”。不同的虚拟蜜罐可以被赋予不同的“个性”。在每个发送出去的数据包的协议头中引入适当的修改，使得数据包符合指纹识别软件预期的操作系统的特征。

蜜罐系统通过路由部件实现虚拟网络拓扑结构。山大商务学院的蜜罐系统给予 Honeyd 技术，只支持有根树网络拓扑结构。当蜜罐接收到一个数据包时，就从根结点开始传输数据包，直到找到拥有目的 IP 的网络，而在传输过程中，路由部件会计算包丢失和等待时间以解决是否丢弃该数据包，且对包传输中的每一次路由转换路由部件都对数据包中的 TTL 值做减 1 操作。当 TTL 值减为 0 时，路由部件会发送一个 ICMP 超时数据包。

本文设计的蜜罐系统借助了 Honeyd 软件框架支持多种记录网络活动日志的方法，可以记录并报告所有协议的尝试连接与完成连接的日志，也可以配置成以人工可读的方式来存储蜜罐系统所接受到所有数据包。同时，服务程序也可以通过标准错误输出向 Honeyd 报告它们收集到的网络信息

3.2.3 自动脚本产生模型

本文的设计框架中，蜜罐系统在规划模拟服务器时，需要手动编写相应的服务脚本，而且不同系统中的服务脚本又有差异，很难得到一个标准的服务脚本，而且对于一些专用的协议，也缺少合适的脚本，为此本文提出了一种能够自动产生服务脚本的模型作为 Honeyd 的子系统。图 3-3，通过对服务器中的数据进行分析和脚本模型对比，并进行剪辑使之产生服务脚本，产生好的服务脚本存入蜜罐系统中。对下一次的入侵形成有效地防御。

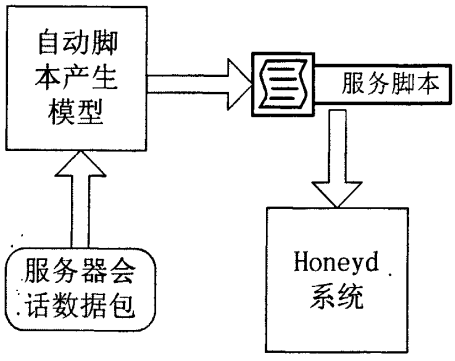


图 3-3 Honeyd 与自动脚本产生模型的关系

Figure 3-3 Honeyd scripts and auto-generated model of the relationship between

3.3 本章小结

本章简要阐述了蜜罐及蜜罐系统的定义、分类、作用以及优缺点，在针对此类情况之后具体介绍了改进 Honeyd 技术的山大商务学院网络安全蜜罐系统的设计思想、体系结构、等内容。

下一章详细阐述模型的设计思路及功能模块，在模型中，重排蜜罐系统获取的会话信息，并以状态机的形式来表示不同会话消息之间的次序，之后使用生物学上的序列比对方法来获取协议会话过程中的语义信息，合并相同语义信息的状态，使得状态机得到简化，最后再使用本人提出的一个区域分析算法再次化简提合并语义信息之后的状态机，区域分析算法中考虑了频繁出现的相同字节区域及有可能含有特殊语义信息的因素，再次对状态机进行化简。在第四章将详细介绍模型的设计方法。

第 4 章 序列算法蜜罐脚本模型的设计

4.1 模型描述

4.1.1 模型的提出

序列比对蜜罐系统模型基于以下的原因提出：Honeyd 只提供一些虚假的服务，这些服务通过监听特殊的端口实现，进入系统的数据流很容易被识别和存储[30][31][32]。一直以来，这种蜜罐系统对攻击者请求的回应是通过手工编写脚本实现的。所以，这种系统的脚本很少，特别是针对一些专用的协议，可用的脚本就更少。基于以上原因，本文设计了一种能够为 Honeyd 自动产生脚本的蜜罐系统模型。

4.1.2 模型功能描述

序列比对蜜罐系统模型的设计是在既不知道任何有关协议，也不知道正在执行的服务的前提下，自动产生脚本。本模型主要的思想是设定一个适中的目标：为攻击蜜罐系统的攻击者的请求提供一个可信的回复。由于攻击者总是以一种很确定的控制操作来访问蜜罐系统，所以，我们只需要关心协议数据单元中的个别字段即可，而且攻击者的执行路径只是我们所模拟服务的执行树上的非常有限的几条。这样我们就有可能为攻击者提供一个可信的回复。

序列比对蜜罐系统模型产生脚本的步骤分为以下三步：

（1）在所测试的网络上配置一台 PC 作为蜜罐（Honeypot），并且截获经过这台 PC 的所有数据包，并以 tcpdump 文件的形式保存。如果这台 PC 被攻击则停止实验，并将其截获的数据包清空。

（2）分析截获的包中服务器与客户的会话信息，并用一个状态机来表示这些会话信息。我们在每个监听端口上建立一个状态机。

（3）从状态机中得到脚本，通过这些脚本可以识别客户发出的请求，并给出可信的回复消息。

当然，这种方法只能获得对真实服务的一个近似模拟。我们得到的服务器与攻击者会话的信息越多，就能识别更多种类的攻击。在该系统模型中，使用了序

列比对算法来识别消息序列中的语义信息，对于序列中频繁变化，但内容相同的字节区域则使用区域分析算法来发掘其中的特殊语义信息，从而使构建的状态机更加高效。

4.2 序列比对算法

序列比对蜜罐系统模型中使用序列比对算法来获取蜜罐截获的数据包中的语义信息，将含有相同语义信息的消息合并，从而简化会话状态机，这一步对脚本的产生至关重要，所以在介绍模型功能模块之前，先对该算法做一个详细的描述。

4.2.1 序列相关概念

序列比对问题是生物信息学中最基本、最常见的问题，也是生物信息学最基本的分析方法。常用的序列比对方法有两两序列比对和多序列比对。两两序列比对又分为全局序列比对和局部序列比对[34]。

下面介绍其相关概念：

1、编辑操作和编辑距离

通常用距离来表示两个序列之间的相似程度，距离越大，相似性越小，距离越小相似性越好。汉明距离即是最简单的一种，从图 4-1 看出用汉明距离表示两序列之间的相似性不灵活。例如：当两序列的长度不等的时候，就不能用它来表示，或者两序列中有个别字符错位了（尽管两序列种的大部分字符是相同的），可导致它们之间的距离很大如图 4-1 所示：

*X*1	:	*A*	*C*	*G*	*T*	*A*				
*Y*1	:	*C*	*C*	*G*	*A*	*A*				
汉明距离： 2										
*X*2	:	*C*	*G*	*T*	*A*	*T*	*A*	*T*	*A*	*T*	*A*
*Y*2	:	*C*	*T*	*A*	*T*	*A*	*T*	*A*	*T*	*D*	*A*
汉明距离： 8										

图 4-1 汉明距离示意图

Figure 4-1 Hamming distance

如果在序列里面引入一个“-”字符，通过比对算法在序列插入“-”来表示：与之进行比较相应序列中对相应字符的删除；序列比对过程中引入“-”后，X、Y 两个序列相应的字符会出现以下 4 中情况：

(α,β)：称为“匹配（Match）”，代表相应字符匹配；

$(\alpha,-)$：称为“删除（Delete）”，代表从序列 X 中删除α，或代表在序列 Y 中插入α；

(α,β)：称为“替换（Replacement）”，代表序列 Y 中的β替换序列 X 中的α，也就是α不等于β；

$(-,\beta)$：称为“插入（insertion）”，代表在序列 X 中插入β；

由以上的可以看出在 X 序列中的删除（插入）和在 Y 序列中的插入（删除）是等价，由此可以得到：

$X2$ ： $C\ G\ T\ A\ T\ A\ T\ A\ T\ -\ A$
$Y2$ ： $C\ -\ T\ A\ T\ A\ T\ A\ T\ D\ A$

其中，对于$X2$ 来说，有一个删除即$delete(G,-)$和一个插入即$insertion(-,D)$和 9 个匹配。可以看出，在引入“-”，定义了 4 中编辑操作后，避免了在汉明距离中出现的问题。如果给每个操作赋予一定的权值ω_{OP}，由此计算两序列所有相应字符操作的权重和叫做编辑距离。[33]

例如：假设$w(\alpha,\beta)=0$； $w(\alpha,\beta)=1,\alpha\neq\beta$； $w(\alpha,-)=w(-,\beta)=1$；

$X2$， $Y2$的编辑距离$D_{edit}=1+1+9\times0=2$

2、序列比对

实际上进行序列比对就是利用特定的算法计算字符序列之间的差别，也就是求在允许插入、删除和替换字符的情况下序列之间的比对分值，以显示序列之间的相似性。相似性包括定量和定性两个方面：相似性定量通过一定的度量单位来表现两个序列定量方面的内容；而两序列的比对排列是两序列相互之间的排列位置，表现了两序列之间在哪些地方是相似的或在哪些地方是不同的，它表达了相似性的定性方面的内容。最优比对就是在两序列之间找到对应相同字符最多、差异最小的两序列排列比对[47]。

4.2.2 序列比对算法思想

序列比对算法是根据给定的编辑操作权重函数ω_{OP}，计算得到两个或多个字符串序列的最优比对，即对两个或多个字符串序列通过匹配相对应的字符或通过插入“-”来表示插入或删除而得到序列之间的最大相似性排列。

在该算法中，不是直接算出确定 X 和 Y 的整体相似性，而是建立在确定的两

个序列的任意相似整体上，从最短的前缀开始，利用先前的计算结果求的最大的前缀的相似性。

假设两序列$x(0\wedge m)$、$y(0\wedge n)$，并且x_i、y_i分别表示序列X、Y中的第i个字符。假设现在已经知道了两序列$x(0\wedge i)$、$y(0\wedge j)$（$i,j\geq 1$）之间所有前缀之间的最优比对，可以有以下三种方式来获得$x(0\wedge i)$、$y(0\wedge j)$之间的比对：

1、比对$x(0\wedge i)$、$y(0\wedge j-1)$，在y_j处匹配一个空格

2、比对$x(0\wedge i-1)$、$y(0\wedge j)$，在x_i处匹配一个空格

3、比对$x(0\wedge i-1)$、$y(0\wedge j-1)$，x_i与y_j匹配

说明：前两种方式的比对，相当于插入了一个空格，加大了两序列之间的距离，所以将其权重函数ω_{OP}定义为一个罚分值g，而第三种方式时则为匹配或替换，所以将权重函数ω_{OP}定义为一个函数$p(i,j)$。

两序列之间的相似度距离用公式表示如下：

公式 $$sim(x(0\wedge i),y(0\wedge j))=\max\begin{Bmatrix} sim(x(0\wedge i),y(0\wedge j-1))+g \\ sim(x(0\wedge i-1),y(0\wedge j-1))+p(i,j) \\ sim(x(0\wedge i-1),y(0\wedge j))+g \end{Bmatrix}$$

公式说明：

1、g为空格的罚分，当x_i与y_j相等时，$p(i,j)$为匹配的分数，不相等时取相互替换的分数；

2、如果是局部比对：那么当$sim(x(0\wedge i),y(0\wedge j))$小于0时，将$sim(x(0\wedge i),y(0\wedge j))$的值改为0；

3、当$i,j\geq 1$时，递归计算$sim(x(0\wedge i),y(0\wedge j))$；

4、当$i=0$或$j=0$或$i=j=0$时，为其递归出口，即

$$sim(x_0,y_0)=sim(0,0)=0$$

$$sim(x(0\wedge i),y_0)=sim(x(0\wedge i-1),y_0)+\omega(i,-)=sim(x(0\wedge i-1),y_0)+g \quad i=1,2,\cdots,m$$

$$sim(x_0,y(0\wedge j))=sim(x_0,y(0\wedge j-1))+\omega(-,j)=sim(x_0,y(0\wedge j-1))+g \quad j=1,2,\cdots,n$$

这样就得到了序列s,t之间的最优化比对。

当给定一个 g 和 $p(i,j)$ 的值和两个序列

X ： $C\ G\ T\ A\ T\ A\ T\ A\ T\ A$
Y ： $C\ T\ A\ T\ A\ T\ A\ T\ G\ A$

可以构造距离矩阵 $D[i,j]_{m+1,n+1}$ 来表示其计算过程。

表 4-2 是 $g=1$， $p(i,j)=\begin{cases}0 \wedge \wedge x_i = y_j \\ 1 \wedge \wedge x_i \neq y_j\end{cases}$ 时的距离矩阵 $D[i,j]_{11,11}$：

表 4-1 序列 x,y 之间的距离矩阵

Table 4-1 The distance matrix between x and y

	j=	0	1	2	3	4	5	6	7	8	9	10
i=	--	--	C	T	A	T	A	T	A	T	G	A
0	--	(0)	1	2	3	4	5	6	7	8	9	10
1	C	1	(0)	1	2	3	4	5	6	7	8	9
2	G	2	(1)	1	2	3	4	5	6	7	7	8
3	T	3	2	(1)	2	2	3	4	5	6	7	8
4	A	4	3	2	(1)	2	2	3	4	5	6	7
5	T	5	4	3	2	(1)	2	2	3	4	5	6
6	A	6	5	4	3	2	(1)	2	2	3	4	5
7	T	7	6	5	4	3	2	(1)	2	2	3	4
8	A	8	7	6	5	4	3	2	(1)	2	3	3
9	T	9	8	7	6	5	4	3	2	(1)	(2)	2
10	A	10	9	8	7	6	5	4	3	2	2	(2)

通过以上序列比对算法，我们能得到序列 $x(0 \wedge m)$、$y(0 \wedge n)$ 之间的最优比对为

X ： $C\ G\ T\ A\ T\ A\ T\ A\ T\ -\ A$
Y ： $C\ -\ T\ A\ T\ A\ T\ A\ T\ G\ A$

距离值为 2。

可以看出，该算法的结果是在两序列之间找到对应相同字符最多，差异最小的两序列排列比对。本文将该特性用在蜜罐系统获取的客户发回的请求消息序列中，屏蔽了相同类型、不同消息之间的内容的差异性，从而获取消息序列中的差异最小的共性序列，即得到消息中的语义信息，将会话状态机中，在同一条状态

转换线上的有共性语义信息的请求消息合并，从而化简会话状态机。 下面详细介绍本系统模型各功能模块的设计。

4.3 系统功能模块设计

序列比对蜜罐系统模型的设计分为四个功能模块：消息队列重建模块、状态机建立模块、状态机简化模块、脚本产生模块。如图 4-2 所示：

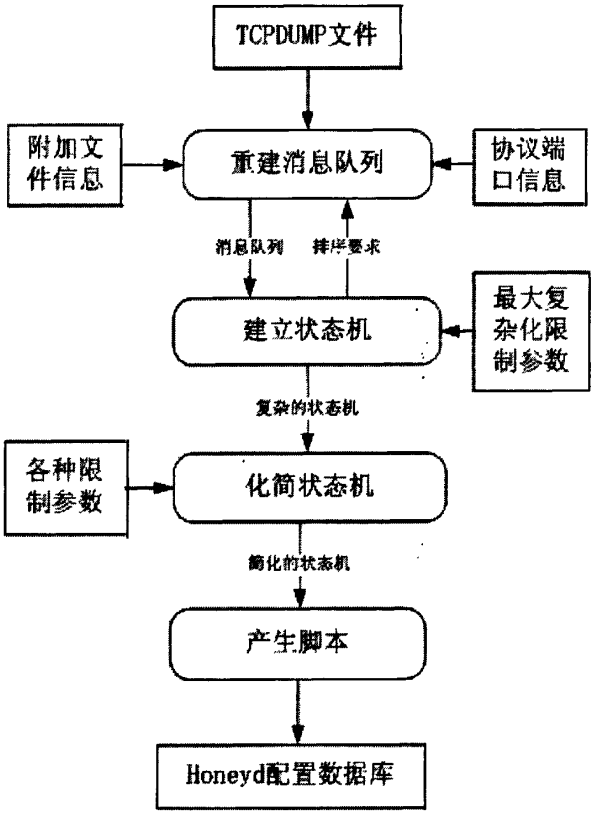


图 4-2 功能模块关系

Figure 4-2 The relationship between functional modules

（1）消息队列重建模块

该模块从 tcpdump 文件中提取会话信息，按照时间、协议分为不同的消息队列，试图重建这些会话信息的原型。在这个过程中，需要结合协议端口对应关系、会话状态等相关信息来完成。

（2）状态机建立模块

该模块在已被重建的消息队列的基础上，构建会话状态机。所谓会话状态机，

目的在于将客户、服务器的会话过程以图的形式来表现，以便于对其提取产生相应脚本。

（3）状态机化简模块

该模块对已构建的状态机进行化简，一方面结合序列比对算法来去掉部分重复的语义信息，一方面通过区域分析算法识别状态机中的一些隐含语义信息。

（4）脚本产生模块

该模块根据化简的状态机，在 honeyd 工作过程中，产生实时的脚本。

4.3.1 消息队列重建模块

这个模块主要是从 tcpdump 文件中提取客户、服务器之间的会话信息，并按照协议将这些信息分成不同的消息队列。本文中只提取 TCP 协议会话信息，对 TCP 协议的数据流进行重建，对其进行转发并重排。

消息队列是一个有序的消息列表。一个消息可以看作服务器和客户交互过程中的一个片断。一个消息可以定义为流向同一个方向的连续的字节集合（可以是服务器到客户，也可以是客户到服务器）。一个 TCP 会话能被分解为一个一个消息组成的有序消息列表。这个消息列表代表我们观测到的客户和服务器之间的会话。一个消息队列的长度以客户或服务器发出的消息的数量来定义。

下面介绍重建 TCP 消息队列的过程：

首先需要定义一个优化的算法来分析 tcpdump 文件，并重建客户、服务器之间的会话。考虑到攻击者为了躲避 IDS，往往不会改变 TCP 的正常状态值。所以做以下假设：

（1）仅仅考虑包含净荷数据的数据包，比如单纯的 ACK(ACKnowledge Character，确认字符）数据包我们将不予考虑，转发的数据包也不予考虑。

（2）一个 TCP 会话以第一个 SYN（synchronize，握手信号）包开始。为每个新的 SYN 包分配必要的数据结构，作为新的数据流的开始标志。

（3）一个 TCP 会话以第一个 FIN（Finish 终结链接）/RST（Reset the connection 错误链接）包结束。当会话双方中的一方决定结束会话时，我们认为会话就结束了。

（4）TCP 包中的序列号将作为我们存储净荷数据的数组的索引号。这样有

助于处理由于转发等原因而乱序的数据包，这种情况下，把最先的接受到的一个包接受下来，后面的包就可以丢弃了。

当然，基于这些假设会引起一些错误，比如未将检验和字段考虑在内，将不能发现错误的发送包，还有错误的序列号也将导致存储区分配过大的问题。尽管这样，经过几个月的测试，这四个假设基本可以满足需要。

4.3.2 状态机建立模块

本模块中，用在消息队列重建模块中重建的消息队列，建立一个状态机。状态机由边和状态组成，边表示一个状态可能转向另一个状态的条件。我们以服务器回应给客户的一条消息为一个状态，以客户在服务器回应后发出的一个请求为边，这样将服务器和客户的会话转化为一个一个状态之间的转换。在这个模块中建立的状态机，非常庞大，并且有冗余，效率也很低。我们必须定义一些阈值来限制状态机中从某个状态出发的边数目的膨胀，当然，设定阈值，会使得所产生脚本的执行到达的状态跟真实服务器上的服务不一致。

在状态机中，每条边都有一个权重，表示取样的实验数据中经过这个转换过程的次数。另外，如果服务器的一个回应消息不仅包含应有的交互信息，还包含额外的信息（如：时间），这种情况下，一个状态将包含多个标签。也就是说，客户的一个请求将得到服务器的多条回应消息，对应状态机中，则是一条边可能到达的状态含有多个字节区域（见 4.3.3）。所以，服务器的状态标签存放在一个数组中。并同时保存每个标签的出现次数。出现次数最多的一个作为默认的回应消息。

为了减少状态机的复杂性，我们采用了两个阈值：一个状态连接的最多的边数和状态的最多数目。图 4-3 所示为一个简单的状态机，第一个状态 S_0 为一个服务器的消息，当然，如果在连接建立时，协议（服务器）没有发出一个欢迎消息，则这个状态就为空。S_0 有三条边 C1、C2、C3 分别表示三种不同的客户端请求消息。每条边将与不同的到达状态连接，并且，这些状态可能包含一个或多条服务器消息。

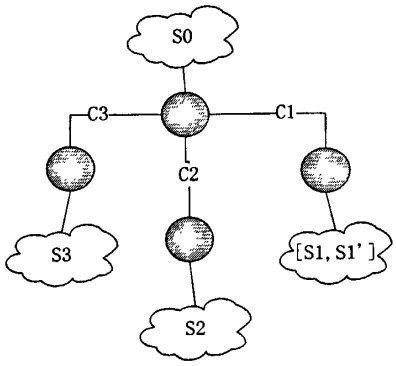


图 4-3 状态机的简例

Figure 4-3 Example of the state machine

4.3.3 状态机化简模块

在状态机建立模块中建立的状态机中，没有考虑任何的语义上的因素，这样的状态机是只是针对给定的 tcpdump 文件，缺少一般性，不能处理任何在样本数据包中未出现过的情况，所以，我们需要简化状态机并使其一般化。

这个模块是本文的核心部分。我们分析状态机建立模块中建立的粗糙的状态机，并且引入一些语义学的概念，使用了两个算法，一个是序列比对算法，一个是本文中提出的局部分析算法。通过使用这两个算法，将得到一个简化的状态机。在本模块中，我们将一个到来的消息，不仅看作简单的字节队列，而且看作满足某些属性的字节域。

下面用一个例子来说明：

一个简单的即时消息协议（Instant Messaging Protocol），其简化的消息如表 1 所示。我们可以看到，一旦与服务器连接，客户发出了 12 条消息。每一条消息在状态机中体现为一条从初始状态发出的边。这样，从初始状态发出的边数与用户名是相当的。这样的状态机太特殊化，不能处理一个新的用户名。所以有必要从这些转换的条件中得到一些更一般的模式。

出现这种问题的原因在于我们忽视了消息中的语义信息。其实，上面的例子中，我们可以只用两条边作为初始状态的发出边，一条为：“GET MSG FROM <username>”，一条为：“SEND MSG TO <username> DATA”。因为我们想在不知道任何有关协议的情况下自动获取脚本，所以，需要一种技术能够得到消息队列

中的语义信息。这就是状态机简化模块的重点，这里我们通过一个两层的聚类（序列比对算法和局部分析算法）过程来实现。

表 4-2 即时消息协议

Table 4-2 A sample of the Instant Messaging Protocol

编号	消息
1	GET MSG FROM <bob>
2	SEND MSG TO <john> DATA：“Hi”
3	GET MSG FROM <marty>
4	SEND MSG TO <ken> DATA：“I’m coming”
5	GET MSG FROM <corrado>
6	GET MSG FROM <liz>
7	SEND MSG TO <bill> DATA：“Be patient”
8	GET MSG FROM <robert>
9	SEND MAG TO <diego> DATA：“Sorry”
10	SEND MSG TO <miki> DATA：“It’s beautiful”
11	SEND MSG TO <dan> DATA：“See you”
12	GET MSG FROM <rei>

（1）序列比对算法

首先宽度优先遍历初始状态机，找在语义上相似的边，逐渐合并一些在语义上相似的边，这意味着我们可以从消息流中发现语义信息，并进行比较。这个问题也是协议信息工程（Protocol Informatics Project）[28]所面临的一个问题。他们借鉴了生物信息学上处理 DNA 序列和蛋白质结构的序列比对算法来进行协议的逆向工程。PI 是用序列比对算法来减轻手工分析协议的工作量，而我们则是用来识别语义上相似的消息，从而简化初始状态机。

采用在 4.2 节介绍的序列比对算法，可以对从同一个状态发出的边代表的消息进行比对，并得到其中的最长的匹配前缀字节序列，从而识别消息中的共性信息（比如：表 4-2 中的 GET 消息和 SEND 消息）。对表 4-2 中 GET 消息的比对结果如表 4-3 所示。这样就等到了共性信息：GET、MSG、FROM 我们把序列比对的结果输入下一步局部分析算法。

表 4-3 GET 消息序列比对结果

Table 4-3 Message sequence alignment results

0012	x47 x45 x54	X4d x53 x47	X46 x52 x4f x4d	x3c x77 x71 x72 x61 x66 ___ ___ ___ x3e
0001	x47 x45 x54	X4d x53 x47	x46 x52 x4f x4d	x3c ___ x75 x73 x65 x72 ___ ___ x61 x3e
0005	x47 x45 x54	X4d x53 x47	X46 x52 x4f x4d	x3c ___ x64 x73 x61 x66 ___ ___ x61 x3e
0006	x47 x45 x54	X4d x53 x47	X46 x52 x4f x4d	x3c ___ ___ ___ x68 x66 x67 x68 x66 x3e
0003	x47 x45 x54	X4d x53 x47	x46 x52 x4f x4d	x3c ___ ___ ___ x61 x62 x63 ___ ___ x3e
0008	x47 x45 x54	X4d x53 x47	X46 x52 x4f x4d	x3c x65 x71 x74 x73 x64 x67 ___ ___ x3e
ASCII	G E T _	M S G _	F R O M _	< ? ? ? ? ? ? ? ? >

（2）局部分析算法

模型中第二步执行的局部分析算法是一种常用的单模式匹配算法 RK，它利用 HASH 函数和素数理论，首先定义一个 HASH 函数，然后将模式串 P 和文本串 T 中长度为 M 的子串函数转换成数值。显然只需要比较那些与模式串具有相同 HASH 函数值的子串而提高效率。当然因为 HASN 冲突的存在，还要进一步进行字符串比较，但只要选择适当的素数 HASH 冲突的概率就会很小，计算时间为 O(M+N)。

图 4-4 表示了序列比对算法和局部分析算法的关系。序列比对算法合并了一部分消息序列，相当于进行了第一层的聚类。下一步，我们定义了一种局部分析算法，利用序列比对算法的输出数据进行第二层的聚类。模型中第二步执行的算法是一种常用的单模式匹配算法 RK，只考虑固定区域，而把变异区域看作一些未知字符的集合。利用 RK 算法，模型计算一个新到来的消息和状态机中某个转换消息之间的相似度值，选择相似度值最大的转换消息指向的状态作为一个新的回应消息状态。

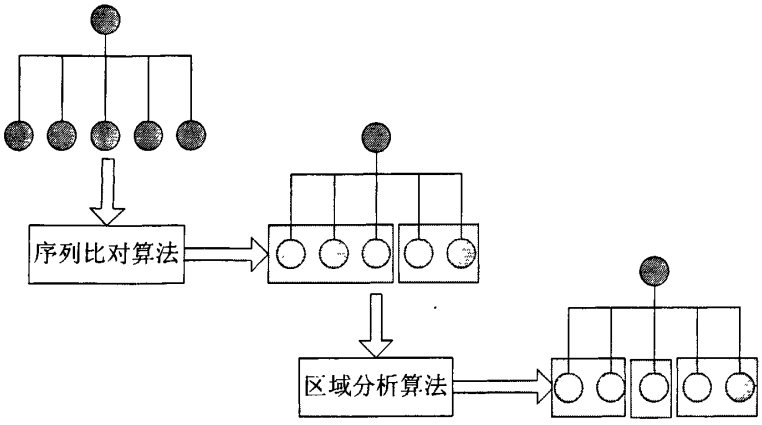


图 4-4 序列比对算法和局部分析算法关系

Figure 4-4 Sequence alignment algorithm and Local analysis algorithm

先看表 4-3 中序列比对的结果，我们还可以为每个排序的字节计算如下值：

1）出现次数最多的数据类型（二进制、ASCII、0 值）；

2）出现次数最多的值；

3）每个值的突变率（可变性）；

4）在某个字节上存在的间隔（表中用短横线“-”表示）。

基于以上的计算，局部分析的范围规定为满足以下条件的字节序列：

1）有相同的数据类型；

2）有相似的突变率；

3）包含相同的数据；

4）同时有或同时没有间隔。

局部分析的字节序列有共同的特性，所以可能含有相同的语义信息。

序列比对中定义的两个序列之间的距离只是依赖于不同字节的数量。当然，可能两个序列之间只有一个比特位的不同，但这个不同的比特位却是非常重要的。所以，作为对第一层聚类的补充，局部分析算法利用局部范围的突变率作为另一个参考因素，所谓的突变率是一个随着字节序列中的区域的不同而可变的值。本算法基于这样一个假设：如果一些字节频繁的出现，则判定可能包含一定的语义信息。

在图 4-5 所示的样例中，第一层聚类不能区分一个 HTTP GET 是获取了一个图像文件还是获取了一个 HTM 文件，而在局部分析时找到了频繁出现，且字

节内容均相同的区域，从而区分出了图像文件和 HTM 文件。在局部分析算法中，利用了一个有趣的特性：协议中频繁出现的功能部分，由极高的可能性将被再一次聚类。所以，使得该模型更适合于协议中最普通的功能部分。

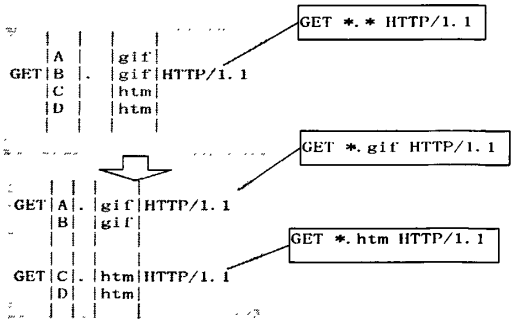


图 4-5 局部分析算法简例

Figure 4-5 A sample of the local analysis algorithm

局部分析算法达到了目的：可以将一般性加入到复杂的特定的状态机中，通过语义信息识别不同的字节区域。

另外，一些被识别的区域在会话过程中起着重要的作用。有时客户发出一些序列，并希望服务器将这些序列原路发送回去，比如：会话 ID，依据这些来回发送的字节区域之间的依赖性很重要。可以将这些区域称作随机区域，这些字节区域有几乎 100%的突变率。在每个样本序列中寻找这些随机区域的值，并且努力在服务器的回应中发现与其匹对的值，将这些消息的连接存储起来，可以用在模拟的状态机中。

4.3.4 脚本产生模块

这个模块是从简化的状态机中得到与 Honeyd 协调的脚本。

一旦状态机被简化，则要将状态机保存起来，并使其能够被 Python script 模拟。为了使状态机能够方便的被利用，为每一个新到来的客户连接都建立一个新的处理程序的实例。为每一个状态列出其标签（如果存在），并且从局部分析算法的输出中得到将对客户采取的回应。这种办法对于一个重量级的蜜罐系统是不够用的，但是对于验证一个方法还是可以的。

下一步选择一个算法来匹对新到来的消息和各种可能的转换消息（状态机中的边代表的消息）。重新利用序列比对算法是不可行的，因为需要的资源超出

了所部署的蜜罐系统可被利用的资源。考虑一下原因：脚本产生器着重于语义上重要的那些字节区域，称作固定区域，相反，那些突变率高的区域则称为变异区。

4.4 本章小结

本章是论文的核心，介绍了序列比对蜜罐系统模型的提出缘由及功能描述。

在该蜜罐系统中，重排 Honeyd 获取的会话信息，并以状态机的形式来表示不同会话消息之间的次序，之后使用生物学上的序列比对方法来获取协议会话过程中的语义信息，合并相同语义信息的状态，使得状态机得到简化，最后再使用本人提出的一个区域分析算法再次化简提合并语义信息之后的状态机，区域分析算法中考虑了频繁出现的相同字节区域及有可能含有特殊语义信息的因素，再次对状态机进行化简，将简化之后的状态机中的状态保存起来，供新到来到的客户请求匹对所用。本系统模型中第三模块为核心部分，通过序列比对算法和区域分析算法使系统模型的有效性得到了提高。

第 5 章 实验与分析

5.1 实验环境

序列比对蜜罐系统模型在布置实验时，有以下一些局限性：

1、由前文中介绍的模型中状态机的结构可知，该系统模型的实验只局限于 TCP 会话或 UDP 的请求/回复会话组，即以会话形式完成的服务；

2、如果服务器的响应不是来自于会话，则不能进行模拟；

3、对于使用加密隧道技术的会话信息不能处理。

实验由多个模拟 NETBIOS（NetBIOS Services Protocols，网络基本输入/输出系统协议）会话的虚拟蜜罐主机和多个向模拟服务器发送请求的客户机组成。实验数据则使用 Leurrecom.org Honeypot project[21]提供的 tcpdump 文件，包含一个蜜罐 5 个月期间收集的来自于 1107 个客户的请求消息。用这些消息序列来建立状态机，产生脚本。另外，TCPopera[35]是一个 TCP 流量操纵工具，它可以重现 TCP 流的会话。我们用 TCPopera 来协助重排 TCP 流。

5.2 实验思路

1、系统模型参数验证

从 Leurrecom.org Honeypot project 得到的 tcpdump 文件中的消息序列中包含了客户发出的请求和服务器的响应信息，分析之后，提取出客户发出的请求信息，利用客户机将其重新发向参与实验的模拟服务器[36][37]（即 Honeyd），将模拟服务器发回的响应信息和原来 tcpdump 文件中的服务器信息进行字节相似度比较，这样我们的模拟服务器不会受到外界未知攻击的干扰。将模拟响应信息和真实服务器发出的响应信息的相似度作为衡量模型有效性的一个因素。实验模拟的质量主要决定于以下几个参数：状态机中节点的出度、状态机中节点的数目、状态机化简时的阈值（序列比对算法中的相似度上 S 和区域分析算法中的频率值 P）。

实验中，考虑了以下两个主要参数：

（1）两个消息序列之间的最小相似度距离 sim(x,y)（以下简称为 X）。X 值决定着两个客户请求是否被合并，即状态机中从一个状态发出的边的数目。X 值越大，则第一层聚类越是粗略，更有可能在第二层聚类中将其分开。X 值越小，则第一层聚类越是精细，将使得状态机更为复杂、详细。

（2）作为区域分析算法中决定是否将某个突变率的区域取出，从而产生不同的聚类的频率值 P。P 值越大则，第二层聚类将起到很小的作用，P 值越小，则第二层聚类会将第一层的聚类结果分的更细，第二层聚类起的作用更大。

2、系统有效性验证

下面，我们讨论另一个问题，就是该系统在真实网络环境中能否有效的欺骗攻击者。

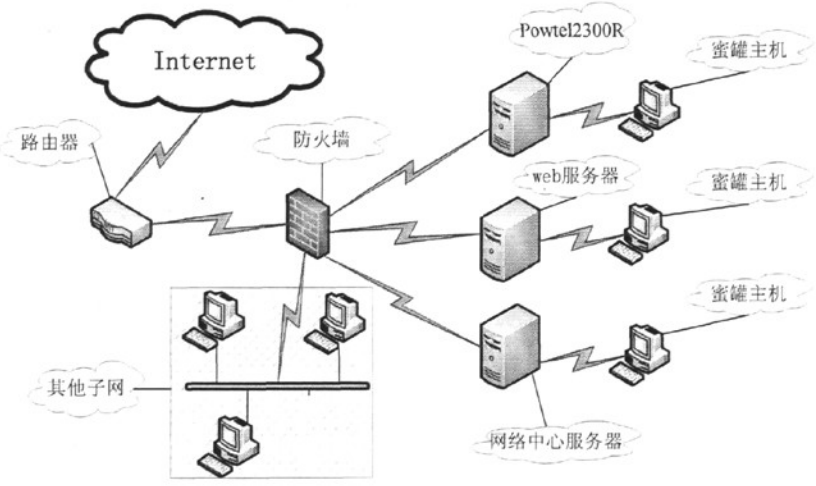


图 5-1 有效性验证网络结构图

Figure 5-1 The network structure of validation

我们在商院校园网的拨号访问服务器 Powtel2300R、Internet 访问的堡垒主机和网络中心的服务器旁分别部署一个真实的序列比对蜜罐系统，如图 5-1 所示。用该蜜罐系统模拟了以下一些端口：

TCP Port 80（HTTP），用脚本模拟

TCP Port 135(DCE)，打开

UDP Port 137(NetBios Name Service)，用脚本模拟

TCP Port 139(NetBios Session Service),用脚本模拟

5.3 结果及分析

1、系统模型参数验证

图 5-2 为按第一种实验思路得出的状态机节点数目与参数 S、P 的关系。由图可知在以下两种情况下，使得状态机的节点数目最多：

（1）S=0，即消息序列中的每条消息构成一个一层聚类；

（2）P=0，即一层聚类结果中的每条消息作为了一个二层聚类。

显然，在这两种情况下，区域分析算法未起到任何作用。

相反，当 S=1 并且 P=1 时，得到的状态机节点数目最少。这时，只有最基本的会话状态被保存下来。实验结果基本与原先设想的 S、P 的变化情况一致。

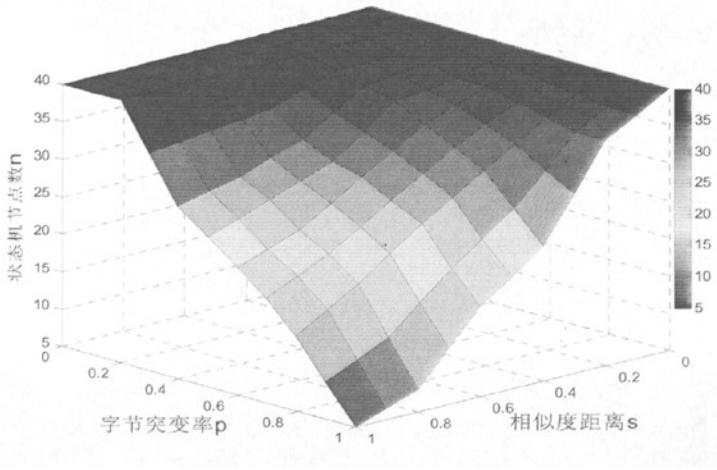


图 5-2 状态机的节点数随相似度距离 s 和字节突变率 p 的变化

Figure5-2 The number of state machine's note as the change of s and p

图 5-2 中横轴一个表示根据序列比对算法得到的两个客户端消息之间的最小距离 S，一个表示区域分析算法中决定是否将某个突变区取出的频率值 P，纵轴则表示得到的状态机的节点数目。两个横轴分别对应了序列比对算法和区域分析算法的粒度大小，该图是随着这两个粒度大小的变化，而引起的状态机节点数目的变化情况。也即在状态机化简过程中，两种算法的影响力、有效性问题。

2、系统有效性验证

实验过程中，在我们的蜜罐主机上运行不同的 NetBIOS scanners[38][39][40]，

所有的扫描器利用 NetBIOS Name Service 正确地识别蜜罐主机，并且正确地获取蜜罐主机的消息。所以，序列比对蜜罐系统可以在这种网络环境中运行。

为了比较本文中序列比对蜜罐系统与高交互式蜜罐系统的有效性，我们将本文蜜罐运行在与数据源 Leurrecom.org Honeypot project 中的高交互式蜜罐相似的网络环境中。考虑到相同协议地址可能会在一些时间后，一般 24 小时，被分配给不同的机器，所以从获取的 tcpdump 文件中选取时间间隔在 24 小时之内的客户机消息序列作为一个源（即实验中的一个模拟攻击者）。本实验中只考虑所构建的状态机最复杂的 NetBIOS TCP 139 端口。

表 5-1 本文蜜罐与高交互式蜜罐的比较

Table 5-1 The results of the comparison

条件	发现的攻击者数
Powtel2300R 的蜜罐主机	327
Internet 访问的堡垒主机的蜜罐主机	329
网络中心服务器的蜜罐主机	333
基于虚拟机的高交互式蜜罐系统	325
同时被发现的攻击者数	45
两者均未发现的攻击者数	30

表 5-1 给出了本文中序列比对蜜罐系统与高交互式蜜罐发现的攻击者的数量对比，实验数据中包含了 300 个攻击源，但是结果只有 45 个攻击源同时被两个蜜罐系统发现。

实验中发现，本文中序列比对蜜罐系统之所以不能发现一些攻击，有部分原因是由于 SMB(服务器信息块)协议头部中还包含一个服务进程 id(逻辑运算)字段。该字段能够唯一识别建立虚联接的客户进程，服务器的响应信息必须包相同 id 值的进程 id 才能更好地欺骗攻击者。由于实验中的服务器没有很好的模拟这个特点，使得部分客户端拒绝了与蜜罐系统的会话，还有待在以后的研究中继续发现、更新本文的方法。

5.4 本章小结

本章从两方面验证第四章中提出的序列比对蜜罐系统模型：系统模型参数验证、系统有效性验证。实验中发现可以通过设置模型参数来调整其敏感度。本文中序列比对蜜罐系统能够发现的攻击者数目可以与高交互式蜜罐系统相媲美。

总结

本文主要阐述了山西大学商务学院的校园网环境以及校园网管理方案，并针对其对服务器的安全设置的不足，设计了一个序列比对蜜罐系统，并在商院校园网环境中进行了简单的测试，结果表明，该蜜罐系统能够有效地工作。

本文成果有以下几点：

1、一直以来，蜜罐系统脚本生成依赖于人工编辑，而本文设计的序列比对蜜罐系统能够自动生成脚本；

2、序列比对蜜罐系统模型理论上能够提供任何协议的模拟脚本，不需要事先知道任何有关协议的信息，所以它能被广泛的应用在不知道其任何行为特征的协议上；

3、序列比对蜜罐系统对状态机简化过程中的参数有较强的敏感性，而且所构建的状态机越复杂，其模拟结果越准确；

4、从实验得到的结果看出该蜜罐系统有一定的前景，蜜罐的使用者可以动态的改变的蜜罐的模拟能力，当然，这有待做更多的实验验证。由于时间仓储，我们没有验证有未知攻击数据时的状态机变化情况。

虽然蜜罐本身的概念及其简单，但涉及的知识面很广，例如分析从蜜罐采集到的数据就需要对网络协议有深入的了解。加上本文初次接触这方面的工作，限于知识水平和经验的不足，难免会出现一些不够完善的地方。恳请各位老师、同行批评指正。

参考文献

1 诸葛建伟. 蜜罐与蜜罐系统技术简介. 北大狩猎女神项目组技术报告，2004：39-43 页

2 Honey D.Retrieved 9 October 2007，from：http://www.honeyd.org

3 Niels Provos. A Virtual Honeypot Framework[R]. CITI Technical Report，03-1：1-13 页

4 els Provos. A Virtual Honeypot Daemon(Extended Abstract)[R]. Center for Information Technology Integration University of Michgan，2003：2-3 页

5 尹曙明，严曲，聂琨坤，高坚. 基于序列比对算法的伪装入侵检测技术 [J]. 计算机工程，2007：7 页

6 Honeynet Project，The (2007a) know your enemy Honeynets. Retrieved on 7 October 2007，from：http://www.honeynet.org/papers/honeynet/index.html

7 Honeynet Project，The. (2007b) Know Your Enemy GenII Honeynets Retrieved 2 September 2007，from：http://www.honeynet.org/papers/gen2/

8 Border C (2007). The development and deployment of a multi-user，remote access virtualization system for networking，security, and system administration classes. Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education，Covington, Kentucky，USA：9-11 页

9 www.honeynet.org.cn

10 Honeynet Project，The.Know your enemy：Learning about security threats (2nd ed.). Addison-Wesley Professional，2005：3-4 页

11 Collins D (2006). Using VMWare and live CD's to configure a secure，flexible，easy to manage computer lab environment. Journal of Computing in Small Colleges，21(4)：273-277 页

12 Georg Wicherski. Medium Interaction Honeypots, April 7, 2006

13 P.Diebold, A.Hess, G.Schafer, A Honeypot Architecture for Detecting and Analyzing Unknown Network Attacks. In Proc, Of 14th Kommunikation in Verteilten Systemen 2005(KiVS05), Kaiserslautern, Germany, February 2005: 44-45 页

14 Master Thesis. A Practical Comparison of Low and High Interactivity Honeypots. Senior Research Fellow, ISI Professor Marc Dacier, Enrecom Institute, September 2005: 12-14 页

15 王雪松. 基于 IPS 和蜜罐技术的安全防御系统的研究. 南京理工大学硕士学位论文, 2008: 23-36 页

16 倪永玮. 基于主动安全策略的蜜罐系统的设计与实现. 贵州大学硕士学位论文, 2008: 14-16 页

17 官凌青. 蜜罐 Honeyd 的扩展设计与实现. 西安电子科技大学硕士学位论文, 2007: 31-32 页

18 李磊. 基于蜜罐技术的分布式入侵防御模型研究. 西安理工大学硕士学位论文, 2008: 24-25 页

19 诸葛建伟, 梁知音. 虚拟蜜罐软件 Honeyd(v1.0)简介、安装与使用文档 The Artemis Project/狩猎女神项目组, 2006: 10-11 页

20 Spitzner L. Honeypots Tracking Hackers. Addison-Wesley Professional, 1st edition, September 2002, ISBN: 0321108957

21 G. Kessler, S.Shepard. RFC 1739: A Primer On Internet and TCP/IP tools and Utilities, June 1997: 11-12 页

22 Darpa. Internet program protocol specification RFC 793: tranmission control protocol

23 J Postel. ISI, RFC 768, User Datagram Protocol, 28 August 1980

24 T.Narten IBM，E. Nordmark Sun Microsystems，W. Simpson Daydreamer．RFC 2461：Neighbor Discovery for IP Version 6．December 1998

25 Ami，D. Lakhani，Dr. Kenneth，G. Paterson．Deception Techniques Using Honeypots．Information Security Group Royal Holloway，University of London UK，2003：5-7 页

26 Jan Gobel，Jens Hektor，Thorsten Holz．Advanced honeypot-based instruction detection．Login，December 2006：8-11 页

27 Provos N，Holz T．Virtual Honeypots：From Botnet Tracking to Intrusion Detection．Addison-Wesley Professional，1st edition，July 2007，ISBN-13：978-0321336323：2-3 页

28 Jon Oberheide，Manish Karir．Honeyd Detection via Packet Fragmentation. Networking Research and Development Merit Network Inc，2005：4-5 页

29 M. Dacier，F. Pouget，H. Debar．Honeynets：foundations for the development of early warning information systems．In J Kowalik，J. Gorski，A. Sachenko，editors，Proceed-ings of the Cyberspace Security and Defense：Research Is-sues，2005：21-25 页

30 梁兴柱．网络安全——“蜜罐”技术研究与实现．大庆石油学院硕士学位论文，2006：27-29 页

31 郭文举．反蜜罐技术的研究与实践．重庆大学硕士学位论文，2006：12-14 页

32 尚立．基于 Linux 平台的蜜罐识别系统的研究与实现．北京邮电大学硕士学位论文，2006：41 页

33 张福祥．序列比对算法 CLUSTAL W 并行化的探索与研究[J]．潍坊学院学报，2007 4：112-113 页

34 Zhiqiang Lin，Xuxian Jiang，Dongyan Xu，Xiangyu Zhang．Antomatic Protocol Format Reverse Engineering through Context-Aware Monitored

Execution. National Science Foundation，2007：15-16 页

35 G H Hong，S F Wu. On interactive internet traffic replay. In 8th Symposium on Recent. Advanced Intrusion Detection (RAID)，LNCS，Seattle，September 2005，Springer

36 John R，Lange，Peter A. Dinda，Fabian E.Bustamante. Vortex：Enabling Cooperative Selective Wormholing for Network Security Systems. Kruegel，R. Lippmann，A.Clark(Eds)：RAID 2007：317-336 页

37 王晓东.一种新型诱骗蜜罐系统的设计与实现.四川大学硕士学位论文，2005：25-26 页

38 www.softperfect.com .SoftPerfect Network Scanner.

39 www.radmin.com .Advanced IP Scanner.

40 www.angryziber.com .Angry IP Scanner.

41 熊华等. 网络安全—取证与蜜罐[M]. 人民邮电出版社，2003：45-46 页

42 巨乃岐等. 信息安全—网络世界的保护神[M]. 军事科学出版社，2003：36-38 页

43 www.leurrecom.org

44 www.honeyd.org

45 熊华等. 网络安全—取证与蜜罐[M]. 人民邮电出版社，2003：29-31 页

46 周仲义等. 网络安全与黑客攻击[M]. 贵州科技出版社，2004：54-56 页

47 诸葛建伟，张芳芳，吴智发. 斗志斗勇战黑客—最新蜜罐与蜜罐系统技术及应用[J]. 电脑安全专家，2005：1 页

致谢

在本课题的研究和论文的写作期间，得到了很多老师和同学的热心帮助，在此向他们表示诚挚的谢意。

首先要感谢我的导师张建教授、李爱军教授和王建珍教授，本论文的写作工作是在他们的悉心指导下完成的。他们以自己深厚的理论造诣、丰富的实践经验和对前沿科学敏锐的洞察能力，为我的研究工作提供了有力的指导和帮助；在我硕士阶段的学习过程中，他们严谨的治学方针，稳重的工作作风，积极乐观的人生态度对我的影响会令我终身受益。

感谢众多老师对我的支持和帮助。对我的研究工作长期以来的关心和帮助令我感动不已。

感谢实验组的所有成员和教研室中的其他同学，与大家一起工作、探讨和解决问题是一个很愉快的回忆。

感谢我的家人父母、姐夫、姐姐等给我的鼓励与支持。

最后，衷心感谢为评阅本论文而付出辛勤劳动的专家和教授们!