学校编码：10384　　分类号＿＿＿密级＿＿＿
学号：B200324007　　UDC ＿＿＿

厦门大学

＿＿博士＿学　位　论　文

基于模糊集的自适应伪装入侵检测算法及其在 ASP 服务安全中的应用研究

Adaptive Masquerade Intrusion Detection Algorithm Based on Fuzzy Set and Its Application in the Security of ASP Service

曾　剑　平

指导教师姓名：　黄美纯教授
　　郭东辉教授

专　业　名　称：　凝聚态物理

论文提交日期：　2006 年 5 月

论文答辩时间：　2006 年　月

学位授予日期：　2006 年　月

答辩委员会主席：＿＿＿＿＿＿

评　　阅　　人：＿＿＿＿＿＿

2006 年　月

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密（　），在　　年解密后适用本授权书。

2、不保密（　）

（请在以上相应括号内打“√”）

作者签名：　　日期：　　年　月　日

导师签名：　　日期：　　年　月　日

摘 要

伪装入侵是网络信息系统中普遍存在且难于克服的安全隐患。由于用户行为的可变性，以及伪装用户行为与正常用户行为的相似性使得伪装入侵检测算法的性能提高很有限，并且对不同用户的适应能力也不好。本论文针对这些问题，在总结前人研究成果的基础上，对伪装入侵检测中的滑动窗口设置、决策量的变换、不确定性处理和自适应模型更新等方面做了深入研究。所研究的这些问题有利于提高伪装入侵检测算法对不同用户的适应能力，对不同应用场合的适应性，并提高检测算法的性能。因此，本文的研究工作具有一定的理论意义和实际应用价值。

论文研究的创新之处在于：

(1) 给出了确定滑动窗口的方法。采用条件熵的判定方法，提出了一种确定滑动窗口大小的算法。证明了给定模型的滑动窗口长度与由该模型生成的一定长度的序列对应的滑动窗口长度是近似相等的。

(2) 提出了序列相对模型的似然值的有效转换方法。对该似然值进行规范化处理，采用遗传算法计算规范化过程中所需要的最大、最小似然值。计算复杂度低，并且能求得正确的结果。

(3) 提出了基于区间值模糊集的伪装入侵检测算法。对每个用户特征分别给出相应的可信度定义及计算方法，采用区间值模糊集对这些可信度值进行模糊融合计算，从而提高了检测算法对不确定信息的处理能力。

(4) 提出了一种快速聚类分析算法。采用基于模型的方法，定义一种新的序列相似性计算方法，不需要在聚类过程中对模型进行重复的合并更新。算法具有较小的计算复杂度，并且聚类性能比普通的基于模型的聚类算法有所提高。

除了上述创新性研究工作之外，用 Java 语言实现了论文中的主要算法，并提出了一种可集成的基于 Agent 的检测框架，设计了三层结构的 WEB EDA 平台应用软件，并实现了两者的集成应用。

本论文所提出的新方法都是以提高伪装入侵检测算法的性能、以便于检测为重点，注重提出的算法的可行性和可实现性。所提出的检测模型及算法具有创新性和较强的实用性。

关键词： 伪装入侵; 隐 Markov 模型; 区间值模糊集

ABSTRACT

Masquerade intrusion is a security problem that is difficult to be dealt with and it exists in many network-based information systems. Theoretically, all kinds of the anomaly detection method could be applied to detect masquerade intrusion. However, due to the variation of user action and the similarity between masquerade user and normal user, ability of the detection algorithm to adapt to different users is lower and the performance improvement is very limit. In our work, we focus on the selection of sliding window size, decision variable, detection algorithm and adaptive model update. All these work are related to the key issues of the intrusion detection algorithm, for example, uncertainty information process and parameters setting, etc. The research work has the important sense both in theoretic and practical factor.

The innovation of this thesis is as follow:

(1) A method for deciding sliding window size. By applying conditional entropy, a new algorithm for deciding sliding window size is proposed. It is proved that the window size of the HMM is approximately equal to the window size of sequence that is generated by the model.

(2) A method to transform the likelihood of sequence to models. Applying genetic algorithm to compute the maximal and minimal value of the likelihood, the sequence's likelihood can be normalized to a reasonable decision value.

(3) A detection algorithm based on interval type-2 fuzzy set. In the algorithm, three features of the user are selected and the corresponding user trustee computation methods are defined, and the final user trustee is got by applying interval type-2 fuzzy set data fusion. And it can improve the ability to deal with uncertainty information in the detection process.

(4) A fast clustering algorithm. A new similarity measurement is proposed and model-based clustering is applied to the clustering algorithm. Model update is not necessary in the clustering algorithm, and its computation complexity is lower than that of conventional model-based clustering algorithms while the clustering

performance is still kept in a good state.

The new proposed detection algorithm is focused on the algorithm performance improvement and detection ability. Furthermore, the feasibility and ability for practical implementation are also mainly concerned in the thesis. The proposed algorithm is not only creative but also practical.

Key words：Masquerade Intrusion; Hidden Markov Model; Interval Fuzzy Set

目 录

CONTENTS

第一章　　绪　论

本论文的研究目的是针对网络应用系统中的伪装入侵问题，设计相应的检测模型及算法。为了强调本文工作的必要性和可行性，本章首先分析网络信息系统的安全问题以及各种解决方法的缺陷，指出了伪装入侵是一个普遍存在、难于克服的安全问题；接着对各种伪装入侵检测模型及相关算法进行了回顾和总结，说明了伪装检测研究的可行性，并指出该研究中需要解决的一些重要技术问题，同时针对这些问题，我们提出了本论文要重点解决的问题。最后，简要介绍论文中其他各章节的内容安排。

1.1 网络信息系统及其安全问题

随着计算机网络技术及应用的飞速发展，人们将企业管理、企业发展研究与网络的应用紧密地结合起来,推动了生产方式的革新,促进了社会发展[1-5]。特别是上世纪末期互联网的出现，不仅改变信息产业的运作方式，而且对世界上其它大多数行业的生产、经营、管理等过程产生了更深远的影响[6-10]。

在这个过程中,网络及信息系统一直起着的核心作用,它们完成了用户接口、数据处理、数据管理分析、存储等各种工作。从体系结构看，网络及信息系统具有一定的分层结构，它具有与 ISO 规定的相类似的七层结构[11]，即物理层、链路层、网络层、传输层、会话层、表示层、应用层。其中物理层负责物理连接的建立与管理等，链路层负责差错控制、流量控制和异步处理等，网络层则处理连接的建立保持和终止，传输层负责数据传输、优化网络服务等，会话层是面向应用处理的核心，表示层提供数据的表现形式、语法及用户在通信中传递或引用的信息的形式等，而应用层为应用进程提供服务，应用层的主要功能有操作系统命令响应、虚拟终端、作业的传输和操纵等。

由于软硬件设施中可能存在安全漏洞,使得信息安全问题存在于这个分层结构的各个层次中。例如，发生在网络层上的端口扫描入侵就是利用 TCP 协议中的 SYN、FIN 等标志位来实现的，使得系统管理员无法从日志中发现这种入侵[12,13]。网络监听[14,15]是入侵者经常使用的一种方法，由于大多数网络是广播型的，网关、路由器或者以太网中的任何一台主机上都有可能发生网络监听。发

生在应用层上的安全问题更多，例如密码强度不足、代码漏洞、或安全规则不完善等引起的安全问题等[16,17,18]。

在分布式服务系统中，安全问题就变得更加严峻和难于处理。这主要是由以下几方面原因引起的：（1）分布式服务面向各种各样的用户，网络规模大，受到攻击的可能性大；（2）攻击手段更加多样化，使得分布式服务系统受到的安全威胁大大增加；（3）由于用户节点动态变化、存在多个访问入口，需要保护的对象多，每个用户需要不同的访问机制等，安全管理是一项很困难的工作，因此，产生安全漏洞的可能性大大增加。例如，基于 WEB 平台的电子设计自动化(W-EDA)系统[19]，这是一个典型的分布式服务系统，系统中存在数量较多的用户，而且用户的类型较多，有设计人员、测试人员、管理人员等，每个用户的访问接入方式差别大；另一方面，由于 W-EDA 是一个开放系统，其主机及资源必须在互联网上开放。因此，必须解决分布式服务系统中的安全问题，以便使得应用系统能够安全、稳定地为用户提供服务。

为了保证信息系统的安全，应该提供合适的安全措施。根据 OSI（ISO 7498-2）的定义，一个网络安全体系结构应该提供五类安全服务[20]，即鉴别、访问控制、数据保密性、数据完整性、抗抵赖性。各种安全服务可以通过相应的手段来实现，例如，身份认证可以通过简单的用户名及密码验证方式，也可以通过 IC 卡，甚至是生物识别的方式来实现。而用户具有什么权限则需要通过访问控制来决定，由访问控制通过某种途径显式地允许或限制主体对客体的访问能力及范围，保证合法授权主体访问客体，阻止非授权的访问。

但是，目前不同的安全服务或产品各自为政，它们之间的衔接与协调存在许多问题[21]，因此，这些安全服务不能完全解决一个网络信息系统的所有安全问题。一旦入侵者突破某些安全服务的防线而进入了系统，那么系统将面临着严重的安全威胁。而入侵检测作为一种安全辅助手段，其主要任务是发现网络信息系统中的入侵行为，为网络安全管理员提供有效的证据，以便进行系统安全配置上的优化，有效地制止入侵行为。因此入侵检测的研究是非常必要的。

在研究入侵检测之前有必要对入侵行为作一个完整的分类，根据入侵行为的发生点，可以分为[22]：基于网络的入侵（即网络入侵）、基于主机的入侵（主机入侵）和针对应用软件的入侵（应用入侵）。网络入侵是发生在网络层和传输

层的一种入侵行为，它通常是利用这两层中的相关网络数据传输协议的漏洞，进入到网络设备，如路由器、交换机等，并获得一定的控制权，从而可以获得各种网络传输数据；主机入侵则是针对主机软硬件设备的入侵，通常是由于主机中的系统软件存在某些缺陷，例如操作系统的访问控制机制不完善或系统软件存在的bug 等，使得入侵者可以很容易地进入主机系统，这种入侵所能得到的数据是主机信息、对主机资源有一定的控制权；而应用入侵则是利用某个应用系统的漏洞，如安全规则过于简单或不完备等，它发生在应用层，入侵者可以与正常用户一样自如地在应用系统中查看、修改各种资料。与前两种入侵相比，应用入侵是一种直观的入侵方式，是用户被冒充的入侵，因此又称为伪装入侵[22]。

一般而言,网络入侵和主机入侵可以通过构建更严格的安全设施来预防，例如可信计算可以作为解决这个问题的一种技术[23,24,25]，它是从计算机的硬件结构的底层重新定义严格的安全规则，实现强安全访问。但是，这类安全措施并不能有效阻止伪装入侵的发生。一个典型的伪装入侵过程是：入侵者通过密码猜测或其他途径获得应用系统中的某个用户的身份及认证信息后，直接利用该信息进入系统，从而可以读取用户本人的保密资料或进行其他操作。因此，采取高强度的身份认证方式，使得用户身份被他人获取的可能性就比较小，发生应用入侵的可能性就小些。但是，高强度的认证方式往往对用户的正常操作带来很多的不方便，从而使得在 Internet 上的应用范围非常有限。另一方面，如果用户没有及时退出应用系统，即使采用高强度认证，也不能阻止伪装入侵的发生，它仍然是一个严重的问题。因此，伪装入侵的检测是很有必要深入研究的。

1.2 WEB EDA 设计平台中的安全问题与伪装入侵

网络信息系统中可能存在很多的安全问题，伪装入侵是一种危害性较大而且很难于克服的潜在攻击行为。为了更好地理解及设计检测算法，这里以一个实例直观地分析说明这种入侵的发生、危害以及检测的必要性和可行性。

这个例子是基于 WEB 的电子设计自动化设计平台（W-EDA：WEB-based Electric Design Automation Platform）[19]，为该平台提供安全保障是本研究的一个总体目标。电子设计自动化（EDA）是利用设计工具进行集成电路（IC）设计的方法[26]，能够有效提高设计效率、降低设计过程中产生错误的可能。目

前存在多种 EDA 工具软件[27]，如 Cadence、Mentor , Spice 等，这些不同的工具主要用于进行协议、系统结构仿真、线路功能等的分析与设计。

而 W-EDA 实际上是 Internet 上的一种 ASP（Application Service Provider）应用。工具厂家将各种 EDA 工具放在 WEB 站点上，以服务的方式提供给用户使用。这样，从事电子设计的企业就不必购买各种昂贵的 EDA 工具，也不必安排专门人员维护工具软件的版本、配置各种复杂的工具运行环境。所有这些工作都交给了 ASP 的提供商，企业就可以专心地去做他们真正需要完成的工作。并且可以实现电子设计中的资料共享、设计工具共享以及方便设计人员在更大的范围内进行协作设计，提高设计工作效率、缩短电子产品的设计周期。

W-EDA 是部署在互联网上的一个应用系统，基本的网络结构如图 1-1 所示。工具服务器保存了用户的设计资料，是系统的核心，也是安全保障的重点。

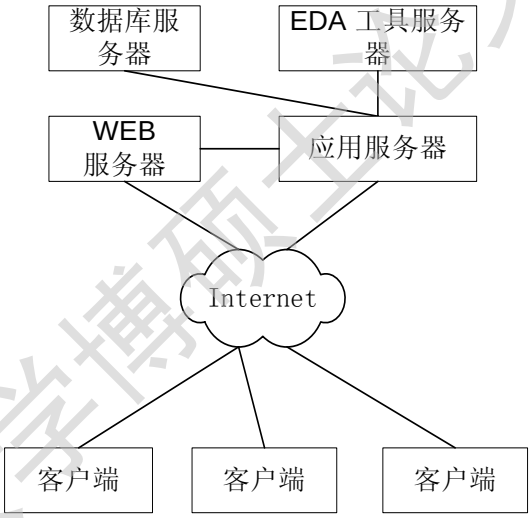

图 1-1　W-EDA 的网络结构

而客户端（Designer）需要通过 X-window 应用软件连接到工具服务器进行设计操作，而 X-window 通信的双方都需要打开通信端口，并接受对方的连接请求，并在建立连接的基础上进行数据通信。因此，把这种以 client/server 工作方式直接应用到 Internet 上，容易产生以下几方面的安全问题[28]：

（1）　主机端口问题

客户端和服务器直接暴露在 Internet 上，受到针对 UDP 或 TCP 端口 177、6000 的攻击的可能性增大。例如，用户可以创建一个 socket，直接与运行 Xserver