基于最大隶属度的网络安全模型

张 铮,赵荣彩,单 征
(中国人民解放军信息工程大学 信息工程学院,河南 郑州 450002)

摘 要:结合免疫系统中多层防御机制和入侵行为中不确定性的特性,利用了改进的模糊识别的方法,提出了一种基于最大隶属度算法的网络安全管理模型。应用数学方法对该模型进行了描述,论证了应用该模型进行网络安全管理的可行性,实现了该系统的原型,并对该模型做了测试和评价,对其应用范围进行了展望。

关键词:网络管理;模糊识别;网络入侵;最大隶属度
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1673-629X(2007)06-0141-04

A Network Security Model Based on Maximum Degree of Membership

ZHANG Zheng, ZHAO Rong-cai, SHAN Zheng
(Sch. of Information Engineering, Information Engineering University of PLA, Zhengzhou 450002, China)

Abstract: Inspired by the multi-level defense mechanism in immune system and uncertainty in intrusion behavior, this paper uses an improved fuzzy recognition method to present a network security management model based on maximum degree of membership algorithm. The model is described in mathematical language. The feasibility of applying it to network security management has been discussed. Implement, test and evaluate the prototype of the model, and finally make estimation to its applicable range.

Key words: network management; fuzzy recognition; network intrusion; maximum degree of membership

0 引 言

随着网络技术的迅猛发展,网络已经成为现代社会重要的信息基础设施,但同时也成为攻击主机或系统的主要途径。同时网络规模不断扩大,网络流量的日益增长,网络拓扑结构逐渐复杂化,网络中应用服务多样化,造成目前的网络管理工作常要管理由不同厂家、不同型号、运行不同操作系统的设备组成的跨地区大规模网络,这使得目前的网络管理大多依靠网络管理系统来实现。

网络管理系统是实现网络管理自动化的主要途径和工具,在网络管理中主要需要实现安全管理、配置管理、故障管理、性能管理、业务管理等几个方面的功能,其中安全管理对保障网络的正常运行有着非常重要的意义。由于近年来网络安全问题日益突出起来,网络入侵事件频频发生,其增长速度异常惊人,破坏范围不断扩大,网络用户对网络安全管理的需求也更加迫切。但是目前的网管系统中的安全管理大多只是简单的基于访问控制的权限管理,这已经不能保障网管系统自身和被管理网络、设备的安全。即使是整合了第三方的安全软件(防火墙、入侵检测等),目前的网络安全管理仍然面临着功能单一、系统开销大、运行效率低、体系结构不合理等方面上的诸多问题。

文中讨论的模型,利用了免疫系统中多层防御机制,对网络中的各种设备划分了几个层次并在每个层次上设定一些检测参数,结合移动代理技术获取这些参数;并应用模糊识别的最大隶属度原则建立了这些参数与网络设备和整个网络安全状态的隶属度函数,从而实现网络的状态识别和安全管理。

1 基于最大隶属度的网络安全管理模型

1.1 改进最大隶属度模型简介

安全管理是网络管理系统的重要组成部分和主要功能之一,它主要包括:与安全措施有关的信息分发(权限设置、密钥分发等);与安全有关的时间通知(非法入侵、非法访问企图等);安全服务措施的创建、控制和删除;与安全有关的网络操作时间的记录、维护和查

收稿日期:2006-08-28
基金项目:河南省杰出人才创新基金(0521000200)
作者简介:张 铮(1976-),男,湖北黄梅人,讲师,博士研究生,研究方向为计算机网络;赵荣彩,教授,博士生导师,研究方向为信息安全、并行算法、先进编译技术。

阅日志等工作[1]。因此本模型根据各种网元(网络设备)的特点对其进行了层次划分和各层的参数设置(如图 1 所示),通过对这些参数进行采用来分析网络及其设备、系统的运行状态,从而实现以上功能。

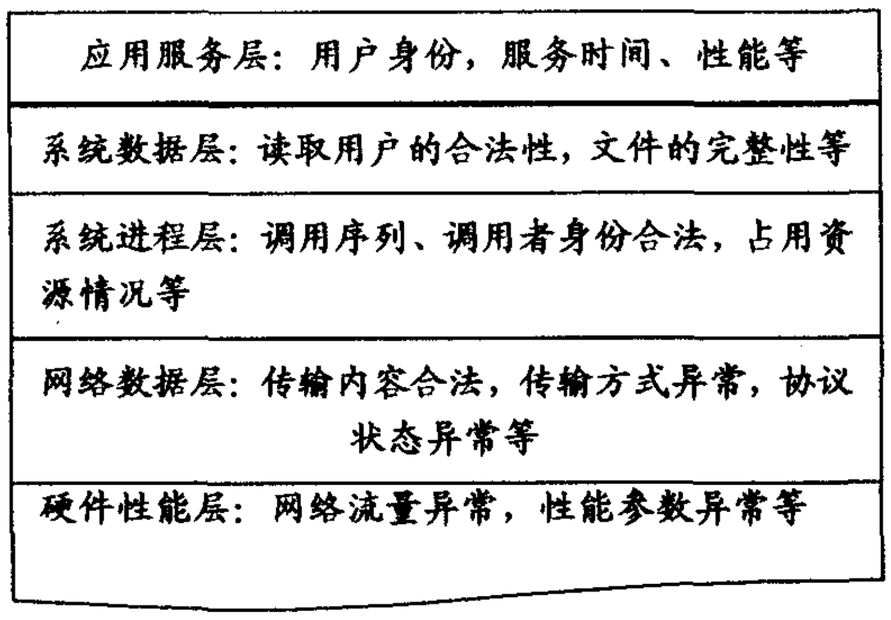

图 1 层次划分与检测参数

不难看出这些参数是该设备的运行状况的信息载体,具体分析这些参数可以得到该设备的运行状况。

根据模式识别中的最大隶属度原则:设论域 $U=\{x_1,x_2,\cdots,x_n\}$ 上有 m 个模糊子集 $A_1,A_2,\cdots,A_m$(即 m 个模型),构成了一个标准模型库。若对任一 $x_0\in U$,有 $i_0\in\{1,2,\cdots,m\}$,使得 $A_{i_0}(x_0)=\bigvee_{k=1}^{m}A_k(x_0)$,这认为 x_0 相对隶属于 A_{i_0}($\bigvee_{k=1}^{m}$ 的含义是从 $1\sim m$ 中的最大值)。在本模型中 x_i 就是各层上设置的检测参数在时间(i)的采样值集合,A_k 则是该设备的运行状态(k)的隶属度;当一个检测参数采用样本集合 x_i 带入到模型的隶属度函数库中计算求得最大值 A_{i_0},则可以相对地认为目前该设备的状态是 i_0。

但是简单的隶属度模型并不适用于检测网络设备的运行状态,因为网络设备的各层参数之间往往有着很大的相关性,如果简单地利用最大隶属度模型将使得这些具有相关性的参数对结果的影响过大,影响判断的准确性。例如:设备负载与进程的资源占用情况有着很强的相关性,当某一进程的资源占用率很高时,该设备的负载也会很高,这就使得设备性能的因素在最终结果的判断中起不到合理的作用。因此文中的模型对最大隶属度模型进行了改进。

改进后的模型描述如下:设论域 $U=\{x_1,x_2,\cdots,x_n\}$ 上有 m 个模糊子集 $A_1,A_2,\cdots,A_m$(即 m 个模型),构成了一个标准模型库;另有 n 个模糊子集 $B_1,B_2,\cdots,B_n$(即 n 个属性聚合),构成了一个中间隐层库。若对任一 $x_0\in U$,有 $i_0\in\{1,2,\cdots,m\}$,使得 $A_{i_0}(x_0)=\bigvee_{k=1}^{m}A\left(\sum_{j=1}^{n}B_j(x_0)\right)$,则认为 x_0 相对隶属于 A_{i_0}。

针对以上的检测参数可以用图 2 来简单描述本模型。

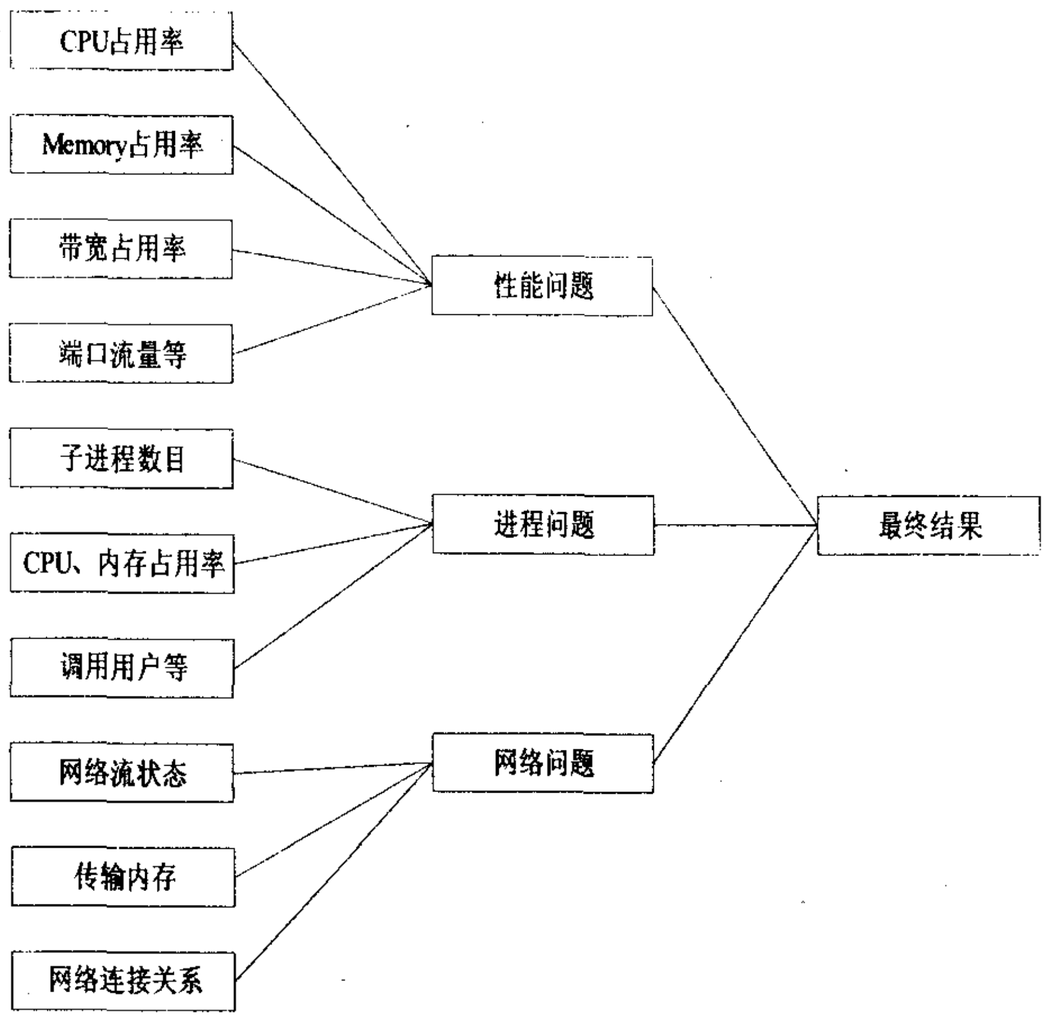

图 2 改进后的最大隶属度模型在安全管理中的应用

1.2 基于最大隶属度的网络安全管理模型的可行性分析

所谓模糊识别就是指在模型识别中,模型是模糊的。也就是说标准模型库中提供的模型是模糊的[2]。由于入侵是指危害资源的完整性、保密性和可用性的活动集合[3],而目前的网络设备并没有确定的完整性、保密性和可用性标准,因此从定义的角度来看网络的安全管理存在着不确定性[4],所以很难用二值逻辑解决,而模糊逻辑则能圆满地解决以上问题。因此文中应用模糊识别来解决网络安全管理的问题。

实际上入侵行为往往不是某个孤立的行为,而是一个有机的活动序列。入侵的前阶段对目标造成的危害比较小,因此可以在此时收集某些信息,待收集到足够的证据判定入侵行为后,再采取相应措施。同时当网络中某台设备遭到入侵后必然会在其运行状态中显示出某些异常现象,否则其面临的入侵方式则是无效的。由于以上因素,入侵的检测可以视为一个多模糊证据综合识别(判定)的问题。这样做一方面可以降低仅仅从某个证据(如,系统日志、网络传输内容等)检测入侵带来的不确定程度;另一方面也降低了漏报和误报,提高了判断的准确性[4]。因此文中的模型在网络安全管理中是可行的。

1.3 原型系统的实现和测试结果

基于以上模型,笔者实现了原型系统,并进行了测试。该系统的体系结构如图 3 所示。

在程序设计方法中采用了中间件(DCOM)技术,将对设备状态的隶属度函数库作为中间件应用在系统中,如有新的状态函数需要添加,则将其代码填写到隶属度函数库中重新编译并在网上发布即可。

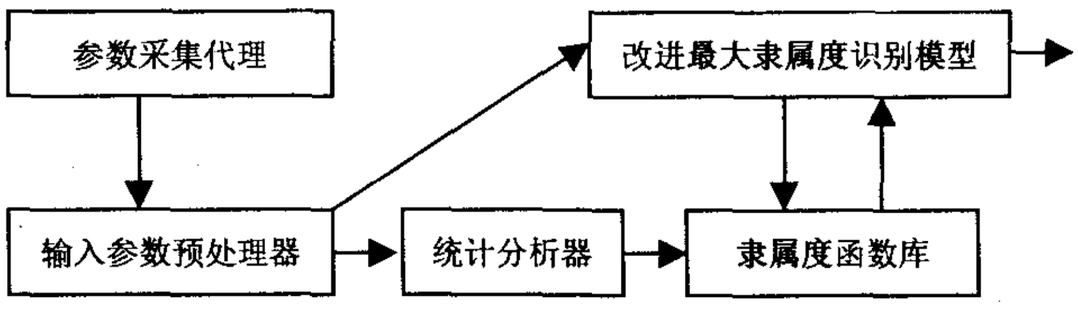

图 3　原型系统体系结构

该系统在拥有仅 714 台主机、36 台路由器、33 台交换机等设备的网络进行了测试。由于网络规模较大,采用了参数采集机 6 台,处理服务器 1 台,数据库服务器 1 台;参数采集程序 5 分钟采集一次性能、网络状态、进程信息和系统调用等参数,这给网络带来的额外流量小于 40MB/天;并对端口扫描、暴力破解密码, Syn_ flood, land, Nimda, Smurf, Arp 攻击等入侵、网络风暴和病毒等不安全状态做了隶属度分析。

其处理流程如下:

首先通过各种方式采集被监控设备的各层检测参数,如表 1 所示。

表 1　被监控设备各层参数设置情况

项目	参数	备注
性能	Cpu, Memory, Disk 占用率; Interface IO 流量; Interface 丢包、误包率; Interface 状态等	利用 SNMP 协议实现
网络状态	流个数;流详细信息(ipadress, port);状态(established, wait, …);统计信息等	利用 SNMP 协议、netstat 命令、shell 脚本与 agent 技术等实现
网络传输信息	网络包内容;标志;窗口等	利用 libpcap 等实现
进程信息	Cpu、Memory 占用率; User;线程数目等	利用 SNMP 协议、top 脚本等实现
系统调用	重要进程的调用序列等	利用 strace 实现

接着对采集的参数进行预处理(选择和规范化),争取排除时空不一致性等因素造成的噪音,并做出一定的分析处理,同时根据统计结果修改最大隶属度函数库。如对系统负载进行统计,图 4 是某数据库服务器(Solaris)的系统负载(uptime 命令得到的 load 值)三个月的统计图。由统计可知,当该 load 值超过 2 时,该系统的非正常状态的可信度约是 99.8%;又如统计得到的连接状态和数目并用基于网络状态的入侵检测模型分析[5],某端口(23)的的 tcp 连接处于 synsend 状态的数目超过 128 synflood 入侵的可信度是 100%。最后将规范化后的检测参数带入隶属度函数库进行计算得到系统当前状态(当然以上的值可以通过修改隶属度函数库进行客户定制)。

最后将采集样本带入改进最大隶属度模型做分析判断。例如,本系统中操作系统是 Solaris 2.6 的主机 Syn_Flood 攻击的隶属度函数:

* $A_{syn\_flood} = B_{性能} \times 0.2 + B_{流量} \times 0.2 + B_{连接} \times 0.6$;

* $B_{性能}$ 值通过图 4 的统计值得到;

* $B_{流量}$ 值通过统计值得到;

* $B_{连接} =$

$$\begin{cases} 1 & \text{系统核心队列半连接数} > 128 \\ \left(\dfrac{x-80}{128-80}\right)^e & \text{系统核心队列半连接数} \leqslant 128 \\ 0 & \text{系统核心队列半连接数} < 80 \end{cases}$$

其中,通过命令 ndd/dev/tcp tcp_ conn_ req_ max_ q 得到 solaris 的核心以连接队列参数默认是 128。

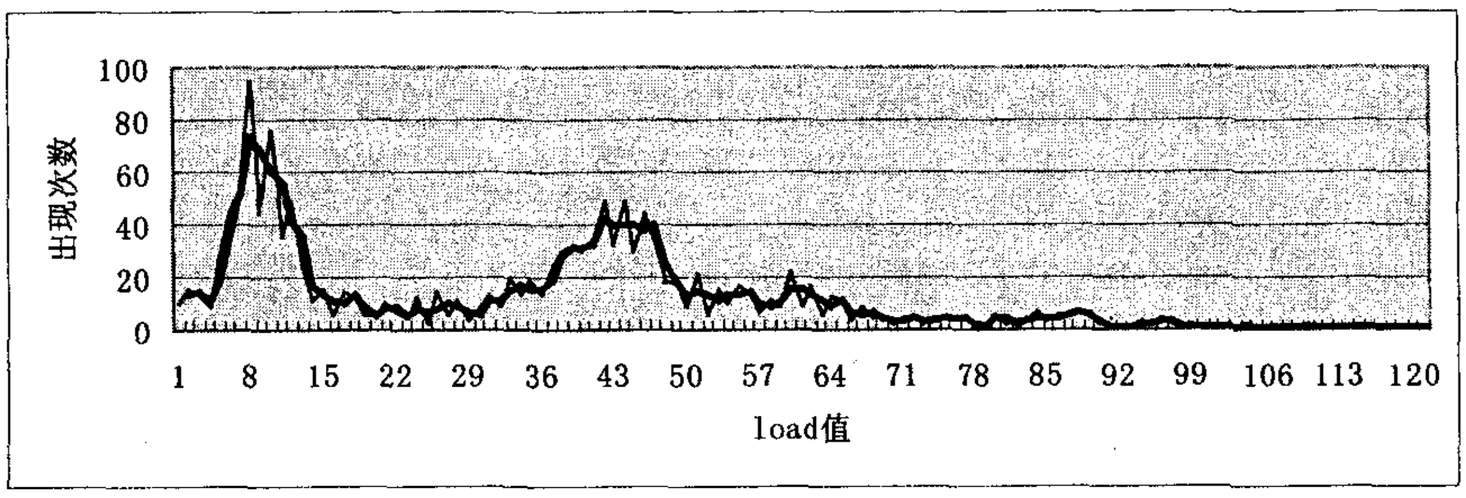

图 4　某 solaris 服务器的系统负载(load)值的统计图

因此当某 solaris 服务器参数采样样本(load = 2.25, i_o = 57265 个(包)/秒, ……,半连接数 134)传送到隶属度函数中的时候,将输出 $A_{i_0}(x_0) = \bigvee_{k=1}^{m} A_k\left(\sum_{j=1}^{n} B_j(x_0)\right) = A_{syn\_flood} = 1 \times 0.2 + 0.89 \times 0.2 + 1 \times 0.6 = 0.978$,故该服务器遭受了 syn_ flood 攻击的可能性为 0.978。

在运用模糊数学方法进行模式识别时,首先要确定隶属度函数。当隶属度函数选取不合适,离实际情况较远时,将影响故障诊断的精确性[6]。本系统中确定隶属度函数的主要方法有:

1)指派法。

指派隶属函数的方法普遍被认为是种主观的方法,它可以把人们的实践经验考虑进去。所谓指派法,就是根据问题的性质套用线程的某些形式的模糊分布,然后根据测量数据确定分布中所含的参数[2]。例如:上面例子中隶属度函数 $B_{连接}$ 的确定就是应用了指派的方法采用了偏大型的 k 次抛物线分布,如图 5 所示,参数确定则应用了 Solaris 中系统默认的核心连接队列参数值 128。

2)统计法。

通过一定时间的统计得到某参数 x_i 的 n 个样本,设 m 为 $x_i = j$ 的频数,则 $f = m/n$ 为 $x_i = j$ 的隶属频率。例如:上面例子中隶属度函数 $B_{性能}$ 是通过对 Solaris 服务器的系统负载(load)值 40 天(每分钟采样一次)的统计确定的。因为期间服务器运行正常,因此可以认为该图是 load 值与正常状态的隶属度(隶属频率)。

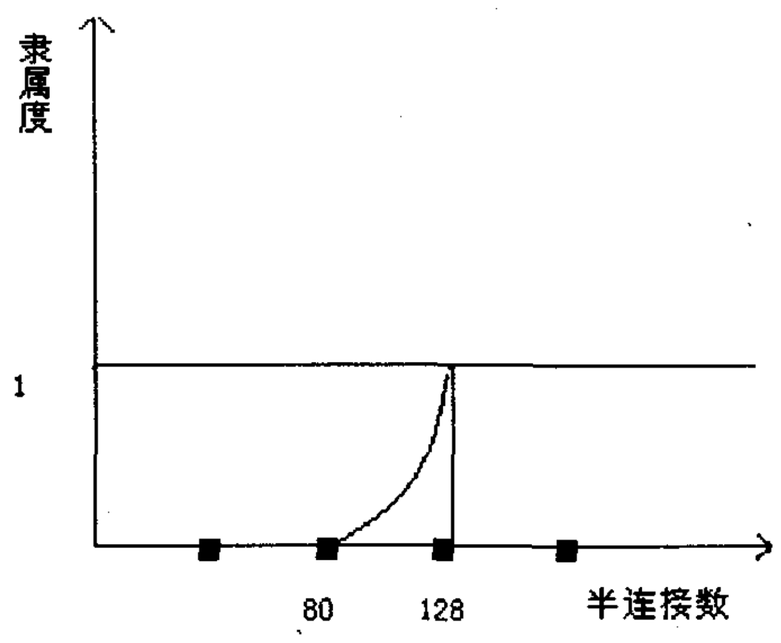

图 5 $B_{连接}$ 的分布图

3)经验法。

就是指根据网络管理员的经验确定隶属度函数。例如:如有采用 UDP 协议,端口是 7626 的网络流,则极有可能是该主机感染了冰河木马,根据经验可以认为这种可能性有 95% 以上,故该情形下该主机感染冰河木马的隶属度是 0.95。

其它的如二元对比法等在该系统中应用的较少。通过以上描述,可以看出该模型的缺点之一是需要采集一段时间内网络正常运转的情况下的参数的数据,同时对使用者的素质要求较高。

2 评价与展望

文中提出的改进模糊识别的网络安全管理模型判断准确,能够很好地降低漏报率和误报率。一方面,它直接围绕需要管理的网络的安全状态本身构造模型,因此具有很强的目的性,从而可以提高安全管理中对被管对象的状态识别的准确性;另一方面,本模型从可能性和入侵行为的不确定性角度出发与基于统计的异常状态检测相比,更具人性化,能够更准确地确定网络安全状态。同时,本模型吸取了免疫系统中"多层防御"机制的特点,不同于其它方法仅从一两项参数判断管理目标状态,并改进了最大隶属度模型利用分组机制消除了某些参数的相关性影响,这些特点使得本模型的漏报和误报率大大降低。

但是,本模型也存在着需要参数的统计数据;对使用者要求较高;某些隶属度函数的确定主要依靠专家经验不能从理论上保证入侵定义的完备性等缺点。未来的工作应集中于识别中不确定性传播的推理;隶属度函数库的学习和改进算法研究等方向。

同时这种改进模糊识别模型具有普遍性和推广价值,不仅仅可以用于网络安全状态管理,也可以应用于设备故障判断定位、物种识别、疾病判断、决策判断等等众多领域中。

参考文献:

[1] 杨家海,任宪坤,王沛瑜. 网络管理原理与实现技术[M]. 北京:清华大学出版社,2000.

[2] 谢季坚,刘承平 . 模糊数学方法及其应用[M]. 武汉:华中科技大学出版社,2000.

[3] Monji A. Languages and Tools for Rule－Based Distributed Intrusion Detection[D]. Namur, Belgium: Faculte´s Universitaires Notre Dame de la Paix,1997.

[4] 李之棠,杨红云 . 模糊入侵检测模型[J]. 计算机工程与科学,2000,22(2):49－53.

[5] Shan Z,Chen P,Xu Y,et al. A Network State Based Intrusion Detection Model[C]//In Proceedings IEEE 2001 International Conference on Computer Networks and Mobile Computing. Beijing,China:[s.n.],2001:481－486.

[6] 虞和济,陈长征,张 省,等 . 基于神经网络的智能诊断[M]. 北京:冶金工业出版社,2000.

(上接第 140 页)

统的更深入研究或升级带来了方便。

文中在分析网格安全特点的基础上,提出了基于虚拟组织的网格安全体系,可以满足网格的各种安全需求。在该体系的认证模型中,用户只需在提交作业前提出认证需求,当登录成功之后便可以重复调用网格中提供的资源,即实现了"单一登录"。此外采用桥接 CA,可兼容不同的本地安全方案,其可扩展性也很强。进一步采用 UML 对虚拟组织的认证进行建模,给出了系统的总体分析图即用例图、静态类图及描述系统动态行为的序列图,阐述了用 UML 进行系统设计的优点,有助于开发人员对系统有清晰的认识,从而提高了开发效率和质量,为进一步完善虚拟组织的安全认证打下了基础,同时也对探讨网格计算的安全实现具有很好的参考价值。

参考文献:

[1] 徐志伟,冯百明,李 伟. 网格计算技术[M]. 北京:电子工业出版社,2004.

[2] Foster I,Kesselman C,Tuecke S. The Anatomy of the Grid－Enabling Scalable Virtual Organizations[J]. International J. Supercomputer Applications, 2001,15(3):200－220.

[3] Pearlman L, Welch V, Foster I, et al. A Community Authorization Service for Group Collaboration[C]// In: Werner B. Proceedings of IEEE Workshop on Policies for Distributed Systems and Networks. Los Alamitos: IEE Computer Society, 2002:50－59.

[4] Pender T. UML Bible[M]. 北京:电子工业出版社,2004.

[5] Selic B. A Generic Framework for Modeling Resources with UML[J]. Computer: Innovative for Computer Professionals, 2000,33(6):64－69.