基于角色访问的非冗余数据库容侵结构

邓 伟 吴中福 叶春晓 钟 将
(重庆大学计算机学院 重庆 400044)

摘 要 在数据库受到攻击的情况下,保证关键业务的持续服务和动态恢复入侵造成的破坏是非常重要的问题。当前很多中小型网络数据库并无冗余机作为后备,但又必须采用容侵机制来保证其可用性和安全性,所以本文提出了一个基于角色访问的非冗余容侵数据库结构。该结构能通过隔离执行关键数据来提供不间断的关键业务服务;对于入侵能够在线进行自动恢复;根据用户的角色、历史记录和容侵要求,自动进行状态迁移和系统参数调节等。

关键词 容侵,控制访问,角色,数据库

Architectures for Intrusion Tolerant Databases Based on Role-Based Accesss Control

DENG Wei WU Zhong-Fu YE Chun-Xiao ZHONG Jiang
(Department of Computer Science and Technology, Chongqing University, Chongqing 400044)

Abstract In face of attacks, it's very important for database system to protect the key data from intrusions and dynamically locate and repair the damage caused by the intrusions. Nowadays, lots of medium and mini-type network databases have no redundancy to guarantee their availability and survivability. Under the circumstances, we propose a intrusion tolerant database architecture which is of no hardware redundancy revolved and base on the role-based access control. The architecture can provide uninterrupted important services by executing key data in isolated places, can automatically recover the damage on the fly, and can self-transfer the system states and also self-regulate the system parameters base on the users' roles, history logs and the requirements of intrusion tolerance etc.

Keywords Intrusion tolerant, Access control, Role, Database

1 前言

容侵概念早在1985年由J. Fraga和D. Powell提出[1],Deswarte,Blain和Fabre在1991年开发了一个具有容侵功能的分布式计算系统[2],但相关研究工作的兴起则是在最近几年才开始的。目前,美国国防高级研究项目署(DARPA)启动了一个新的研究和开发方向,名为"The 3 rd Generation Security(3GS)"[3],主要研究容侵技术,包括系统在面临攻击的情况下保持系统幸存性和弹性(自动恢复)的能力,以及对这些能力进行评估的手段。欧洲启动了MAFTIA研究项目[4],以期系统地研究容侵模型,建立大规模的可靠分布式应用。另外,我国的一些科研单位也在开展这方面的研究工作[5,6]。

一个容侵系统ITS(Intrusion Tolerant System)是这样的信息系统[1],它能够在面向攻击的情况下,仍然连续地为预期的用户提供及时的服务。容侵系统能够检测一些用攻击避免和预防手段无法检测的信息攻击(这些攻击可能透过外层防御,即用攻击避免和预防手段设置的防御,如防火墙系统,认证和加密系统等),并采取一些必要的措施保证关键应用的功能连续正确。容侵技术从本质上讲是一种使系统保持幸存性(Survivability)的技术。根据安全需求,一个入侵容忍系统应达到以下目标:(1)能够阻止和预防攻击的发生;(2)能够检测攻击和评估攻击造成的破坏;(3)在遭受到攻击后,能够维护和恢复关键数据、关键服务或完全服务。

2 容侵数据库和角色访问控制

2.1 容侵数据库

美国国防部DARPA[7]和MITRE[8]公司从20世纪90年代后期开始对网络数据库安全研究这一新兴领域的资助。数据库管理系统的容侵技术[9]正是在这一背景下提出来的。它研究的主要内容是如何使数据库管理系统在遭受到成功的恶意攻击时,能够在对受到破坏的数据进行恢复的同时保留尽可能多的未受恶意攻击影响的其它授权用户在系统受攻击期间的工作结果,能够保持一定的及时提供可靠数据服务的能力。具有容侵服务能力的数据库管理系统的理想情况是:数据库管理系统在遭受到恶意攻击破坏的情况下,能够发现系统被入侵,并及时对受到入侵影响的数据进行隔离和恢复,在系统恢复的同时能够保留尽可能多的未受恶意攻击影响的用户在系统受恶意攻击期间的工作结果;不间断地提供尽可能多的可靠、及时的数据服务;对系统自动进行重新配置,消除攻击者在此次攻击中所利用的入侵途径,避免同样问题再次发生。国际上对数据库管理系统容侵技术的研究起源于美国,主要是GMU的Sushil Jajodia, Paul Ammann[10~12]和UMBC的Peng Liu[13,14]以及North Dakota University的Branjendra Panda[15~17]等几个研究组。国内浙江大学计算机系蔡亮等[18]对信息战语义下数据库管理系统的恶意行为作了研究。

容侵系统大致可分为两种实现方式:一是攻击响应的容

邓 伟 博士生,从事故障诊断、容侵等方面的研究。吴中福 教授,博士生导师,主要研究领域为网络信息安全和多机系统。叶春晓 博士主要研究数据库安全。钟 将 博士,主要研究入侵检测。

侵方法，通过检测到局部系统的失效或估计到系统被攻击，而加快反应时间，调整系统结构，重新分配资源，使信息保障上升到一种在攻击发生的情况下能够继续工作的系统。这种实现方法对“入侵检测系统”的要求比较高，但是由于不改变原系统大的结构，只是增加和修改一些模块，因此实现较为简单，花费也较小。二是攻击遮蔽的入侵容忍方法，就是待攻击发生之后，整个系统通过冗余，多数表决，拜占廷协议等机制来屏蔽攻击造成的影响。这种方法需要重新设计整个系统，增加冗余硬件，并通过冗余、容错技术，门限密码学技术等来实现，所以系统的实现代价很大。

2.2 角色的访问控制

同时，基于角色的访问控制 RBAC(Role Based Access Control)技术[19]是近年来安全访问控制领域的研究热点，越来越受到人们的重视。特别是在互联网中安全管理机制由于具有很大的复杂性，而且用户对服务器资源的需求是动态变化的，RBAC 就提供了系统资源访问权限的有效控制，它提供角色的概念和基于角色的灵活的管理策略，适应于保证系统的安全性。RBAC 的主要目标就是阻止非授权用户对机密信息的访问，防止非法用户的侵入或合法用户的不慎操作所造成的破坏。RBAC 模型主要由三部分组成，即用户、角色和权限，其中用户是指信息系统的合法使用者，包括普通用户和系统管理员；角色是指权限的集合，包括普通角色和管理角色；权限是指用户对客体(信息系统)的操作功能，包括普通用户权限和管理权限。RBAC 的基本思想[20]是：如图 1，由用户在一个组织中担当的角色来确定用户在系统中的访问权限，也就是用角色来充当用户行使权限的中介，安全的管理就可以根据需要定义各种各样的角色，并设置合适的访问权限，而用户根据其责任和权利被指派为不同的角色。这样整个访问控制过程就被分成了两部分，即访问权限与角色相关联，角色再与用户相关联，从而实现了用户与访问权限的逻辑分离。RBAC 技术用角色实施对系统资源访问权限的统一管理，具有减少授权管理复杂性，降低系统开销的特点。

图 1 RBAC 示意图

彭文灵等[21]提出了一种基于 RBAC 的攻击屏蔽容侵机制，该文提出了攻击遮蔽下的角色访问容侵机制，并重点阐述了攻击发生时各个主从角色管理服务器之间的相互协调和转换策略。

目前攻击响应的容侵数据库是国内外的一个研究热点，可以分成以下两类：一类是对可疑入侵行为进行入侵隔离[10,11]的事前预防；另一类是对受到攻击破坏后的系统自动进行破坏范围评估和恢复[12~17]的事后补救。但是基于 RBAC 和攻击响应的容侵数据库结构在国内外均无相关文献报道。为此，本文提出了一个基于 RBAC 和攻击响应的自恢复容侵数据库结构，该结构通过对关键数据的访问隔离进而实现关键业务的不间断服务，并且结合容侵的事前预防和事后恢复，能够对攻击造成的破坏进行快速自恢复。

3 自恢复的容侵数据库结构

3.1 容侵数据库结构

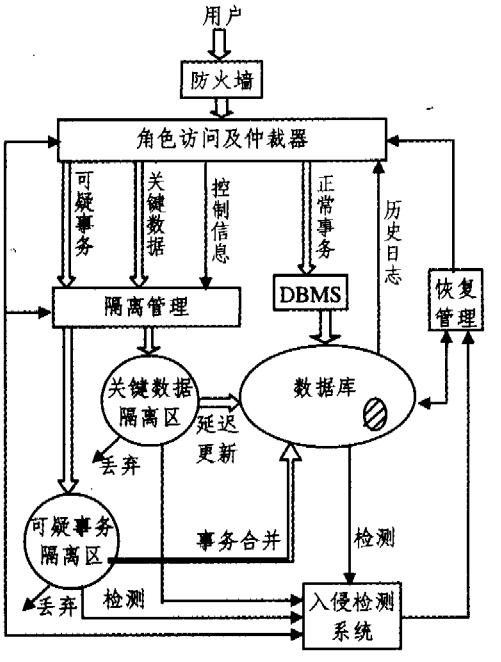

图 2 容侵数据库结构

3.2 工作机制

3.2.1 两种隔离操作 一种理想的角色访问容侵数据库的实现就是根据不同角色采取不同的容侵数据库策略，但在实现上，由于数据库中的数据并不是按照不同角色分块存储的，因此很难解决对同一数据对象不同角色操作而产生的不一致问题。为此我们针对数据库中的不同数据采取不同的容侵策略，这也是符合实际的，因为具体应用中划分角色和关键业务的一个重要依据就在于他们处理的数据的重要程度。为此，我们将重要角色和关键业务相关的数据定义为关键数据，凡对关键数据的写操作均隔离执行，经过入侵检测延迟时间后，再将这些隔离操作体现在数据库中。这种做法消除了入侵检测延迟造成的“时间窗”，避免了恶意事务在此时间窗内造成的破坏扩散和恢复破坏时对这些扩散数据的锁定，从而保证了攻击时重要角色和关键业务的继续执行。由于关键数据的定义将使系统产生额外的开销，因此必须对容侵程度、系统速度等因素进行权衡。

另外，系统还采用隔离执行嫌疑用户事务[10]方法。对可疑入侵行为进行入侵隔离的基本想法是在一个可疑入侵行为被确认之前，采取预防措施，限制该行为可能对系统造成的破坏，同时又必须能够在判定该行为不是恶意攻击时能尽可能多地保留它的操作结果，节省资源，提高系统性能。对可疑用户的后续数据库事务操作隔离执行，这样如果数据库系统在后续操作中发现该数据库用户不是恶意攻击者时，数据库系统能够以较少的资源消耗，达到保留该用户尽可能多的事务操作的目的。该方法把数据库分成主版本和嫌疑数据库版本。当数据库系统发现某个用户具有恶意攻击嫌疑时，它就透明地把该用户和主数据库版本隔离开来，防止(可能造成的)系统破坏的进一步蔓延。同时生成一个对应的嫌疑数据库版本，并把该嫌疑用户的所有后续操作转换为对此嫌疑数据库版本的操作。当发现该嫌疑用户不是恶意攻击者时，数据库管理系统将对应于该用户的嫌疑数据库版本和主数据库版本合并，从而实现既减轻恶意攻击可能造成的危害，又尽可能多保留非恶意用户工作的目的。由于主数据库版本和嫌疑数据库版本之间可能存在着不一致，在进行版本合并时需要

找到并消除冲突。它以优先图[10](Precedence graph)作为工具,描述主数据库版本和嫌疑数据库版本之间是否存在合并时冲突的关系。

3.2.2 主要模块 在图2中,入侵检测系统实时分析数据库的事务操作,从中发现恶意事务,并将其立即送往恢复管理模块。入侵检测系统根据用户的角色信息将提供两种警告信息:当某个事务的评价值超过阈值 TH_m,则该事务为恶意事务;如果评价小于 TH_m 但大于阈值 TH_s,则该事务为嫌疑事务。由于不同的用户和角色发生恶意事务的几率是不同的,因此其 TH_m 和 TH_s 也是不同的。可以采用的一种做法是根据历史数据,对不同角色首先分别训练出 TH_m 和 TH_s,然后建立不同用户审计数据库,根据其历史纪录调高或降低其 TH_m 和 TH_s。目前,有大量的入侵检测算法[23~26]。但是文中出现的入侵检测系统与它们却有以下不同:1. 本文的入侵检测系统的行为往往与容侵系统的其它部分是联动的、紧密耦合的。如本系统对不同角色访问的警告阈值是不同的,本系统当一个事务确认恶意时将通知恢复管理模块定位和修复破坏数据,当与关键数据相关的事务为正常时通知数据库更新等等。2. 本文的入侵检测系统是与具体应用的语义相关的。比如一个职员的薪金为3000元是正常的,而10000元就很异常了。又由于针对不同的应用,不同的入侵检测算法效果是不一样的,因此必须训练系统使其适应不同的应用。3. 对于不同的层次(如应用层、会话层、事务层等)有不同入侵检测算法,为此必须对这些入侵检测算法进行集成,从而提高入侵检测系统的准确率。

恢复管理模块将定位攻击造成的数据破坏并通过日志修复这些破坏。在入侵检测和攻击修复过程中,数据库继续执行新的操作。由于恶意事务对数据的修改将影响后续的与这些污染数据相关的事务,进而涟漪般地扩大破坏影响范围,这叫做破坏扩散。所以恢复管理模块要首先定位恶意事务写的数据集合,并将写之后所有对这些数据读写的事务包含,然后根据日志和检查点等进行数据恢复。当然,能够进行数据恢复的前提是修理的速度快于破坏扩散的速度。在攻击修复期间,整个数据库的完整性是随着时间变化的。如果攻击发生得很频繁,破坏扩散可能非常严重,那么数据库的完整性将得不到有效的保证。为此,角色控制及策略调节器必须根据不同的角色降低新的事务执行速度,具体的策略采用与数据库完整性要求、历史统计数据和数据库的应用领域等有关。

3.2.3 简要流程 假设该事务的用户之前没有过恶意攻击的纪录,该事务与关键数据无关且用户不是重要角色,那么该事务将以正常事务进入数据库运行。在入侵检测的时间窗内,入侵检测系统发现该事务为恶意攻击,立刻通知角色访问及仲裁器,由它根据事务相关性[12~14]确定后续被影响的事务和破坏数据的范围,暂时拒绝这些数据的访问,然后将后继工作交给恢复管理模块,由其根据日志等进行恢复。因为往往恶意用户会连续发出恶意事务进行攻击,所以当入侵检测系统发现恶意事务时,就会将该用户设为嫌疑用户,并根据其恶意事务的破坏程度和数量调节其角色的重要性和其它参数(如THm、THs和事务处理速度等)。在一定时限和一定事务数量内,降低该用户的后续事务接收速度,并将他的事务放入嫌疑数据库隔离执行。如后续事务经入侵检测后正常,就将该事务合并入主数据库,否则丢弃该事务,并修改该用户的相关参数。

如果一个事务为重要角色发出的关键业务事务且与关键数据写有关,那么该事务自动进入关键数据隔离区执行。在入侵检测时间窗后,如果该事务为恶意事务,就丢弃该事务,并标识该用户为嫌疑用户,降低该用户角色等级并修改其相关参数;如果该事务正常,就在主数据库中进行更新。又因为关键数据的主数据库更新是独占的且严格按照事务接受顺序执行,所以在主数据库更新时不存在数据不一致的情况。

结论 当前很多中小型网络数据库并无冗余机作为后备,但又必须采用容侵机制来保证其可用性和安全性,为此本文基于角色访问提出了一个非冗余的容侵数据库结构。该结构能通过隔离执行关键数据来提供不间断的关键业务服务;能对于入侵能够在线进行自动恢复;能根据用户的角色、历史纪录和容侵要求,自动进行状态迁移和系统参数调节等。

下一步工作为:由于入侵检测系统在该体系结构中占据非常重要的地位,因此必须重点研究一个检测延迟短,检测率高,误检测率低的入侵检测系统;将操作系统级的容侵机制与数据库应用级的容侵机制进行无缝连接,从而形成多层容侵数据库系统;将文中思想引入冗余硬件的数据库系统,增强其可生存性。

参考文献

1 Fraga J, Powell D. A fault and intruslon tolerant file system. In: Proc. of the 3' Intl. Conf. on Computer Security, 1985. 203~218

2 Deswarte Y, Blain L, Fabre J-C. Intrusion tolerance in distributed computing systems. In: Proc. of the 1991 IEEE Symposium on Research in Security and Privacy. 1991. 110~121

3 http://www.Darpa.mil/ipto/programs/oasis/techprogram.htm, 2003.

4 Powell D, Stroud R. Conceptual Model and Architecture of MAFTIA. MAFTIA Deliverable D21: [Research Report, RZ 3377]. IBM Zurich Research Laboratory. Jan. 2003. http://www/newcastle.research.ec.org/maftia/deliverables/D21.pdf

5 荆继武,周天阳. Internet上的入侵容忍服务技术. 中国科学院研究生院学报, 2001,19(2):119~123

6 荆继武,冯登国. 一种入侵容忍的CA方案. 软件学报, 2002,13(8): 1417~1422

7 Anderson R H. Research and Development Initiatives Focused on Preventing, Detecting and Responding to Insider Misuse of Critical Defense Information Systems. Results of a Three-Day Workshop, RAND CF-151-OSD, 1999

8 Graubart R, Schlipper L, McCollum C. Defending Database Management Systems Against Information Warfare Attacks: [Technical Report.] The MITRE Corporation, 1996

9 Lee P A, Anderson T. Fault Tolerance: Principles and Practice, Springer-Verlag, Wien, Austria, second edition, 1990

10 Jajodia S, Liu P, McCollum C D. Application-Level Isolation to Cope With Malicious Database Users. In: Proc. 14th Annual Computer Security Applications Conference, Phoenix, AZ, 1998. 73~82

11 Fayad A, Jajodia S, McCollum C D. Application-Level Isolation Using Data Inconsistency Detection. In: 15th Annual Computer Security Applications Conf. Phoenix, Arizona, 1999. 119~126

12 Ammann P, Jajodia S, Liu P. Recovery from malicious transactions. IEEE Transactions on Knowledge and Data Engineering, 2002,14(5): 1167~1185

13 Liu P, Ammann P, Jajodia S. Rewriting histories: recovering from malicious transactions. Distributed and Parallel Databases, 2000, 8(1): 7~40

14 Liu P. Trusted Recovery from Malicious Attacks: [Ph. D. Dissertation]. George Masson University, USA, 1999

15 Panda B, Giordano J. An Overview of Post Information Warfare Data Recovery. In: Proc. of the 1998 ACM Symposium on Applied Computing, Atlanta, GA, Feb. 1998

16 Panda B, Giordano J. Reconstructing the Database after Electronic Attacks. In Database Security XII: Status and Prospect, Jajodia S, ed. Kluwer Academic Publishers, 1999. 143~156

(下转第134页)

来确定。

(2)采用最优保存策略[8]实现跨代保留

反复地执行交叉与变异操作,有可能导致删除最佳应用实例,影响应用的优化求解。设 P_G 是 GSF 应用的第 G 代种群,$AI_B^G \in P_G$ 是当前的最优应用实例(即 $F(AI_B^G) \geqslant F(AI)$, $\forall AI \in P_G$),P_{G+1}是下一代种群,最优保存策略的目的是强制跨代保留最优应用实例:

$$P'_{G+1}=\begin{cases} P_{G+1}, & AI_B^{G+1} \geqslant AI_B^G; \\ P_{G+1}-AI \cup AI_B^G, & AI_B^{G+1} < AI_B^G, \end{cases}$$

其中,$AI \in P_{G+1}$是随机删除的个体。最优保存策略有可能导致算法的未成熟收敛。

(3)通过适应度定标[9]防止早熟收敛

传统的染色体选择方法存在下列问题:

1)在算法执行的初期,往往有可能出现几个超级个体主宰了后代的产生,并且导致算法的未成熟收敛;

2)在算法执行的后期,种群的平均适应度经常接近最佳适应度,在这种情况下,个体间竞争力减弱,最佳个体和其它大多数个体在选择过程中有几乎相等的选择机会,从而使有目标的优化过程趋于无目标的随机漫游过程。

设 $\bar{F}$ 为种群的平均适应度,线性定标定义了 GSF 应用实例的新的定标函数:

$$F'=a \cdot F+b$$

其中,通过求解下列方程可以得出 a 和 b:

$$\begin{cases} a \cdot \bar{F}+b=\bar{F}' \\ a \cdot F_{\max}+b=C_{mult} \cdot \bar{F} \end{cases}$$

这 2 个方程保证了实现遗传算法正确收敛的两个关键特点:

1)定标后的平均适应值 $\bar{F}'$等于原来的平均适应值,这是因为每个平均应用实例都要为下一代产生子孙;

2)适应值为 $F_{\max}$ 的最佳应用实例将为下一代产生 C_{mult} 个子孙,这样可以减少先辈中超级个体和平均个体间的差距,从而避免未成熟收敛;增加后代中这种差距以确保必要的竞争。

4 实验结果

为验证 GSFGA 算法,我们模拟了由 5～30 个网格服务组成的工作流,其中每个服务都有 2～5 个网格节点可以执行该服务。初始种群数为 100,变异概率是 0.04,交叉概率是 0.8,本文算法与 SJLF[10]和 LJLF[10]两个常用算法比较,实验结果如图 2 所示,实验表明,采用 GSFGA 算法所产生的所有网格服务的全部执行时间均小于 SJLF 和 LJLF 算法产生的执行时间。

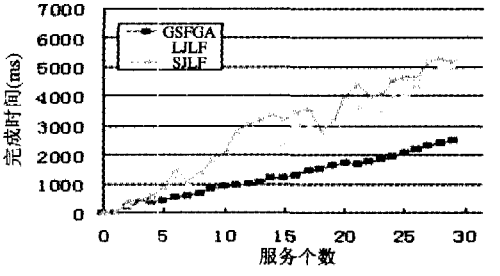

图 2 GSFGA 与 LJLF、SJLF 算法比较

结语 本文针对网格服务工作流调度问题,提出基于遗传算法的网格服务工作流调度算法 GSFGA,并采用灵活的收敛判据、最优保存策略和适应度定标来克服传统遗传算法收敛性差和早熟收敛的缺陷,并通过 GSF 应用实例验证了该算法的有效性。为简化研究,本文定义的网格服务工作流假设每个服务仅提供一种活动,而实际的网格服务应用却复杂得多;此外,网格环境的动态性将很大程度地影响 GSF 调度算法地实施,因此,在下一步工作中,我们将针对复杂的 GSF 应用,研究 GSF 的动态调度问题。

参 考 文 献

1 Kesselman F I, Nick C J, Tuecke S. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. Globus Project, 2002. http:// www. globus. org/research/papers/ogsa. pdf.

2 van der Aalst W, van Hee K. Workflow Management Models, Methods, and Systems. The MIT Press. Mar. 2004

3 Buyya R, Abramson D, Giddy J. An economy driven resource management architecture for global computational power grids1 Int'l Conf on Parallel and Distributed Processing Techniques and Applications, Las Vegas, 2000

4 Frey J, Tannenbaum T, Foster I, Livny M, Tuecke S. Condor-G: A computation management agent for multi institutional grids. Cluster Computing, 2002, 5: 237～246

5 Chapin S, Karpovich J, Grimshaw A. The Legion resource management system. In: 5th Workshop on Job Scheduling Strategies for Parallel Processing, Apr. 1999

6 Goldberg D E. Genetic Algorithms in Search, Optimization 6 Machine Learning. Reading. Addison-Wesley, Massachusetts, 1989.

7 Geist G A, Heath M T, Peyton B W, Worley P H. A user's guide to PICL: a portable instrumented communications library. Technical Report ORNL/TM-11616, Oak Ridge National Laboratory, Oak Ridge, Tennessee, Jan. 1992

8 Bhandari D, Murthy C A, Pal S K. Genetic Algorithm with elitist model and its convergence. Int. J. Pattern Recognition Artif. Intell. 10, 1996

9 Kreinovich V, Quintana C, Fuentes O. Genetic algorithms: What finess scaling is optimal?. Cybernetics and Systems, 1993, 24(1): 9～26

10 Di Martino V, Mililotti M. Sub-optimal scheduling in a grid using genetic algorithms. Parallel Computing, 2004, 30(5-6): 553～565

(上接第 114 页)

17 Tripathy S, Panda B. Post-Intrusion Recovery Using Data Dependency Approach. In: Proc. of the 2001 IEEE Workshop on Information Assurance and Security, United States Military Academy, West Point, NY, June, 2001

18 蔡亮,杨小虎,董金祥. 信息战下的数据库安全-我国的持殊需求分析和对策. 计算机研究与发展, 2002, 39(5): 568～573

19 Sandhu R, Samarati P. Access control: principles and practise [J]. IEEE Communications, 1994, 32(9): 40～48

20 Lebkicher M. Role Based Access Control [EB/OL]. http:// www. giac. org/practical/GSEC/Michael. Lebkicher. GSEC, pdf, 2000.

21 彭文灵,王丽娜,等. 基于角色访问控制的入侵容忍机制研究. 电子学报, 2005, 33(1): 91～95

22 Leonard J. Lapadula State of the Art in Anomaly Detection and Reaction: [Technical report]. MITRE, Bedford, Massachusetts, 1999

23 Javitz H S, Valdes A. The sri ides statistical anomaly detector. In Proceedings IEEE Computer Society Symposium on Security and Privacy, Oakland, CA, May 1991

24 Lee W, Xiang D. Information-theoretic measures for anomaly detection. In: Proc. 2001 IEEE Symposium on Security and Privacy, Oakland, CA, May 2001

25 Samfat D, Molva R. Idamn: An intrusion detection architecture for mobile networks. IEEE Journal of Selected Areas in Communications, 1997, 15(7): 1373～1380

26 Sekar S, Bendre M, Bollineni P. A fast automation-based method for detecting anomalous program behaviors. In: Proc. 2001 IEEE Symposium on Security and Privacy, Oakland, CA, May 2001