

基于 GEP-UCON 的 Web 数据库安全技术研究

宁 葵¹, 龙 珑², 宁德鹏¹, 杨武英³

(1. 广西大学 计算机学院与电子信息工程学院, 广西 南宁 530004;

2. 广西师范学院 信息技术系, 广西 南宁 530003;

3. 广西浦北 龙门中学, 广西 浦北 535300)

摘 要:目前网络安全问题日益严重,各种各样的攻击手段严重威胁网络的安全,而 Web 数据库为众多用户直接共享,安全性问题更为突出。基因表达式编程(GEP)融合了遗传算法和遗传编程的优点,笔者将 GEP 人工智能技术引入到 UCON 模型为 Web 数据库构建了一个多层次的安全防御模型,以多库协同方式建立起授权规则和分类模型,实现用户访问行为模式分析,从而识别和访问主体的角色类型和行为特征,达到安全访问控制的目的。系统使用结果表明拦截对 Web 数据库非法、越权访问的准确较高,GEP-UCON 算法的动态学习功能不断地提高安全防御能力,从而把 Web 数据库的安全性提升到一个较为理想的状态。

关键词:基因表达式编程;UCON;Web 数据库;访问控制

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2011)03-0181-04

GEP-UCON's Web Database Security Technology

NING Kui¹, LONG Long², NING De-peng¹, YANG Wu-ying³

(1. College of Computer Science, Guangxi University, Nanning 530004, China;

2. Department of Computer Science and Information Technology, Guangxi Normal College, Nanning 530003, China;

3. Longmen Middle School of Guangxi Pubei, Pubei 535300, China)

Abstract: Nowadays there is a growing concern about the problem of Internet security. Various means serious threat to attack the network security. Because Web database for many users to directly share, more prominent security issues. Gene Expression Programming (GEP) combines genetic algorithms and genetic programming advantages. GEP artificial intelligence technology will be used to the UCON model for the Web database to establish a multi-layered security defense model. Establish a multi-library collaborative manner authorized by the rules and classification models, analysis of user access patterns to achieve. In order to identify and access the main features of the role and behavior of the type, achieve the purpose of security access control. The results show that the interception system is effective to forbid invade database of illegal, GEP-UCON algorithm can dynamic improve the security and defense capabilities, the way can better improve the Web database security.

Key words: GEP; UCON; Web database; access control

0 引 言

Web 数据库就是把 Web 技术与数据库结合,使用户可以在 Internet/Intranet 上方便地存取、修改、检索数据库中的内容。然而,由于 Internet 的连通性和开放性使网络上的信息安全得不到保障,而 Web 数据库集中存放着大量的数据,且为众多用户直接共享,安全性问题更为突出。Web 数据库通常包含最为敏感、机密的

数据(例如金融数据、信用卡信息、订单资料等),世界上多家政府、银行和企事业单位都在不同程度上遭到非法入侵者的袭击。攻击者可充分利用各种配置和应用程序级漏洞,以多种方式将数据库服务器作为目标来实施危害,Web 数据库服务器所面临的主要威胁包括:SQL 注入、网络窃听、服务器越权访问、数据库密码破解等。

访问控制是对 Web 数据库及服务器进行保护的重要措施,而传统的访问控制模型难以满足用户规模和数据规模激增、访问方式和访问需求呈现动态变化的数字信息访问的需要。近几年提出的使用控制模型(Usage Control, UCON),包含并超越了传统的访问控制模型,为满足现今的数字资源访问需要指明了一个

收稿日期:2010-07-30;修回日期:2010-10-09

基金项目:广西壮族自治区科学研究与技术开发计划项目(桂科合0815007-1-15)

作者简介:宁 葵(1970-),男,硕士研究生,高级工程师,研究方向为网络安全、数据挖掘。

方向。但目前该模型只提供了一个理论框架,需要进行深入研究,丰富完善其具体内涵并加以应用。文中涉及的《Web 数据库安全中间件》研究项目将基因表达式编程(Gene Expression Programming, GEP)的人工智能技术,引入到 UCON 模型中,实现对计算机及其网络上复杂环境下的 Web 数据库进行多角色、多途径、多等级、多条件的保护和应,为用户提供了一种高度安全的互访途径。

1 GEP 简介

基因表达式编程(Gene Expression Programming, GEP)融合了遗传算法和遗传编程的优点,比传统进化计算快了 2~4 个数量级,在挖掘关联、聚类、分类规则,时间序列预测,太阳黑子预测中有出色表现。和传统的 GP 以及 GA 算法比较, GEP 方法最大的特点对于一个二类的分类问题,只需要一条 GEP 规则就可以表达个体的基因型^[1-4]。GEP 的个体由定长的头部、尾部两部分组成,头部中的元素包括函数集与终止符集两个集合,尾部的元素由自终止符集组成,其头部长度(h)与尾部长度(t)的关系如下:

$$t = h \times (n - 1) + 1 \quad (1)$$

GEP 中个体的表现形式一般被称为 ET 树, ET 树是按照顺序扫描基因个体的各个字符,再通过层次顺序而构成。比如,如果定义函数集 $F = \{Q, *, -, +\}$ (依次为开方运算、乘、加、减),终止符集 $T = (a, b, c, d)$, 若 $n = 2$, 取头部长 $h = 7$, 又式(1)得到尾部长 $t = 8$ 。假定有基因型个体: $Q * - + abcd$, 它可以转化为如图 1 所示的 ET 树^[5]。

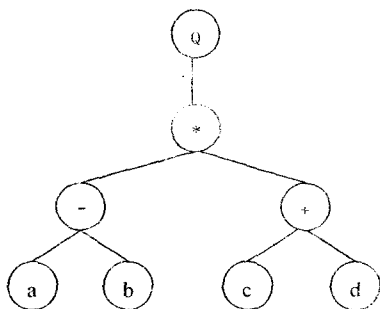


图 1 GEP 的表达式树

2 基于 GEP-UCON 模型的研究

本系统将 GEP 人工智能技术引入到 UCON 模型中,实现对计算机及其网络上复杂环境下的 Web 数据库进行多角色、多途径、多等级、多条件的保护和应,为用户提供了一种高度安全的互访途径。UCON 由主体 S (Subject)、主体属性 ATT(S) (Subject attributes)、客体 O (Object)、客体属性 ATT(O) (Object attributes)、权限 R (Rights)、授权 A (authorizations)、证书 B

(obligations)、条件 C (Conditions) 八个部分组成,如图 2 所示。与传统访问控制模型不同,使用控制模型增加了主体属性、客体属性、证书和条件,这些新增加的成分用于解决传统访问模型中存在的问题^[6]。

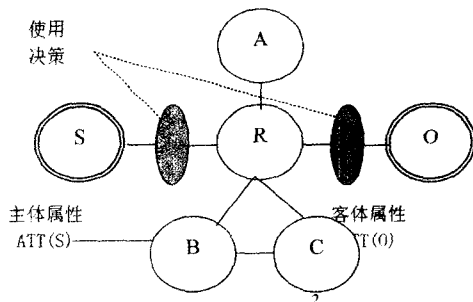


图 2 UCON 模型授权过程

UCON 模型授权过程中需要考虑较多的因素:授权、证书和条件以及主体和客体属性在访问期间的延续性和可变性。从而,基于这些因素的不同状态会得到相关众多的使用控制模型。授权过程涉及到基本核心模型的模式、主体属性、客体属性以及条件等。

在 UCON 模型实现过程中,存在两个问题:

(1)传统的方法在用户群庞大和客体属性过多的情况下,这种复杂的授权过程难以实现;

(2)授权过程没有完全摆脱传统的访问控制模式,一旦“骇客”破解了授权密码,取得授权权限时,整个访问控制过程就形同虚设^[7,8]。

为此,本课题提出基于 GEP 的 UCON 模型:GEP-UCON,其主要思想如下:

1)将 GEP 超强的模式发现功能运用到对主体属性和客体属性的关系模式发现中;

2)将基于 GEP 的决策树分类方法应用到根据授权、证书、条件以及主体访问模式等要素进行授权的过程中;

3)将基于多层染色体基因表达式编程的方法应用到主体权限等级划分中。

从而实现构建智能化的访问控制模型:GEP-UCON 模型,实现对计算机及其网络上复杂环境下的关键用户数据进行多角色、多途径、多等级、多条件的保护和应。

在 GEP-UCON 模型中,利用 GEP 的智能化特性,解决问题(1);在 GEP-UCON 模型中,即使授权密码被破解,使得主体得到充分授权,但从主体访问行为模式的匹配度,系统也有可能对其进行识别、控制,从而解决问题(2)。

3 基于 GEP-UCON 的 Web 数据库访问控制模型

本课题中基于 GEP-UCON 模型的关键用户数据

访问控制方法为:利用 GEP 的智能分析技术强大的函数发现功能以及授权规则生成功能,建立一个智能型的专家系统。系统建立了一个多库协同的模型,建立复杂的授权规则和分类模型,实现用户访问行为模式分析,提高用户类型识别的实时性,从而识别和访问主体的角色类型和行为特征,达到利用 GEP 技术实现安全访问控制的目的。

授权流程如图 3 所示。

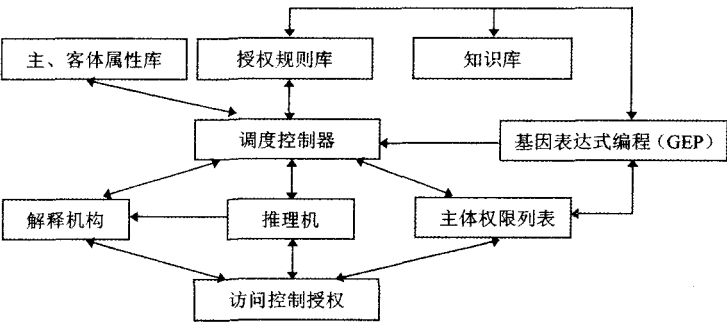


图 3 基于 GEP-UCON 模型的访问控制授权流程

该方法采用多库协同的模式,建立了主客体属性库、授权规则库、知识库等,主客体属性库存储了各种类型和等级的访问主体和客体的相关属性,授权规则库存放着一些有关的数学计算模型、方法和程序以及经过不断学习而积累起来的大量规则,知识库存放着的是计算机领域的专业知识。

调度控制器是在本模型的核心模块,它建立起一种各个库之间、库与推理机之间的协同机制。对于某个具体的推理任务来说,输入、输出的方式都是固定的,推理过程也是基本类同的,相区别的是推理过程中的内容细节,因此每一个具体类型的项目都有着自己的调度控制方式。

系统中的推理过程可以描述为首先利用 GEP 技术进行用户访问模式分类,建立相关主体分类模型,然后利用 GEP 生成的授权规则、主体类型及其权限列表,进行推理,提高 GEP 技术应用的实时性,达到访问控制的目的^[9-12]。

本项目的推理过程采取不确定性的、正向的推理策略,推理的本质是把分散的知识规则有序地链接起来,形成若干条推理链。系统再经过文本、媒体等信息的处理后,将主、客体的属性特征以数据库的方式保存,把输入的事实提供给推理机使用。

同时,本系统基于 GEP 的智能分析过程是一个动态学习的自适应过程。GEP 智能分析模块具有学习功能,多次成功的分析结果可以作为新的分析规则存在,并可以根据系统积累的分析经验来自动弃用长期无效的规则,从而达到优化推理规则库,提高分析准确度的目的。

4 Web 数据库安全访问控制技术中 GEP-UCON 算法的应用

根据上节提出的基于 GEP-UCON 的 Web 数据库访问控制模型,以实际项目中的应用为例介绍“Web 数据库安全访问控制技术中 GEP-UCON 算法”。

GEP-UCON 算法的基本思想是:在 Web 数据库安全访问控制过程中,通过对 Web 数据库中“主、客体属性库”,“主体权限”等加以采集和分析,运用基于 GEP 的智能选取资源算法对 UCON 的“主、客体属性库”数据进行挖掘,在数据自治区域通过在线排队函数对本域节点的任务完成服务进行排序、过滤,选择对该次访问拒绝或是放行,从而达到对 Web 数据库安全访问控制的目的。因此在系统实际应用之前,必须进行特征训练,即是根据实际网络环境构造针对特定的主、客体属性库,这些属性数据首先经过系统后台服务进行

预处理,而在不同的服务器上构建的用户特征库可以重用和共享,这样更增强了系统的准确性和健壮性。

在本系统以 GEP-UCON 算法控制 Web 数据库安全访问的过程中,定义了主、客体属性分析参数:令 $Lwcorrect_k$ 表示分析方法中的方向 k (放行或拒绝)标注正确的权限关系, $Lwalgorm_k$ 表示分析方法中在方法 $k(k=1$ 表示访问放行, $k=2$ 表示访问拒绝)上所有的权限关系; $Lwmall_k$ 标准中标注的权限关系,则准确率(Correction)、召回率(recall)分别定义如下:

$$Correction_k = \frac{Lwcorrect_k}{Lwalgorm_k} \times 100\%$$
 (2)

$$recall_k = \frac{Lwcorrect_k}{Lwmall_k} \times 100\%$$
 (3)

这里一共给出三组实际应用的数据:第一组包括 1000 次的合法与非法的外部 Web 访问,第二组包括大概 3000 次的合法与非法的外部 Web 访问,最后一组是在没有安装本系统的情况下的 1000 次的合法与非法的外部 Web 访问,如表 1、2 所示。

表 1 对 Web 数据库合法、正常访问的系统响应结果

组	合法、正常访问 的次数	放行的 准确率/%	召回率/%
1	1000	95.20	4.05
2	3000	98.12	5.08
3	1000	94.33	3.92

表 2 对 Web 数据库非法、越权访问的系统响应结果

组	非法、越权访问 的次数	拒绝的 准确率/%	召回率/%
1	1000	80.05	19.21
2	3000	89.21	23.02
3	1000	59.52	11.31

表 1 说明了对 Web 数据库正常访问,本系统对其响应结果影响不大,仅比普通的情形稍好;表 2 说明了对 Web 数据库非法、越权访问时,应用本系统的基于 GEP-UCON 的安全访问控制技术准确率、安全性比普通的情形高很多,而主、客体属性库的规模直接影响分析结果,主、客体属性库随着访问次数的增多而增多,准确率越来越高。文中的 GEP-UCON 算法方法的优势在于无须手工添加主、客体属性参数,所以易于扩展,有利于软件的升级,更适合实际软件设计需要。同时保证了过滤结果基本达到了无知道学习的一般水平。

5 结束语

在计算机安全性体系中,任何优秀的加密技术和密钥管理都可能存在意外的时候,换句话说,访问控制技术是用来抵御攻击的最后屏障,这对于 Web 数据库而言也不例外。基于 GEP-UCON 模型的安全访问控制是一种先进的访问控制技术,它在保证安全的同时把访问控制的复杂性以智能化、自学习的方式解决,使访问决策变得更容易。我们利用基于 GEP-UCON 模型的访问控制技术为 Web 数据库建立了一个多层次安全防御模型,其原型系统已在实际应用,对解决 Web 数据库及服务器的安全问题具有十分现实的意义。

参考文献:

- [1] Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems[J]. *Complex Systems*, 2001, 13(2): 87-129.
- [2] Ferreira C. Gene expression programming: mathematical mod-

eling by an artificial intelligence[M]. New York: Springer-Verlag, 2002.

- [3] Ferreira C. Function finding and the creation of numerical constants in gene expression programming[C]//The 7th Online World Conference on Soft Computing in Industrial Applications. England:[s. n.], 2002.
- [4] Yuan Chang-an, Tang Chang-jie, Wen Yuan-guang, et al. Intelligent Function Model Discovery System Based upon Gene Expression Programming[J]. *Journal of Computational Information Systems*, 2006, 2(4): 1299-1304.
- [5] Yuan C, Tang C, Wen Y, et al. Convergency of Genetic Regression in Data Mining Based in Gene Expression Programming and Optimized Solution[J]. *International Journal of Computers and Applications*, 2006, 28(4): 359-366.
- [6] Lopesh S, Weinert R. Egipsys: an enhanced gene expression programming approach for symbolic regression problems[J]. *International Journal of Applied Mathematics and Computer Science*, 2004, 14(3): 375-384.
- [7] 邓集波,洪帆.基于任务的访问控制模型[J]. *软件学报*, 2003, 14(1): 76-81.
- [8] 徐锋,吕建. Web 安全中的信任管理研究与进展[J]. *软件学报*, 2002, 13(11): 57-64.
- [9] 张欢. 基因表达式编程中的转基因关键技术研究[D]. 成都: 四川大学, 2006.
- [10] 彭京,唐常杰,元昌安,等. 基于重叠表达的多基因进化算法[J]. *计算机学报*, 2007(5): 1778-1781.
- [11] 谢方军,唐常杰,元昌安,等. 基于基因表达式的演化硬件进化和优化算法[J]. *计算机辅助设计与图形学学报*, 2005, 17(7): 1415-1420.
- [12] 王东,吴湘滨. 遗传编程运行期个体多样性分析方法及应用[J]. *计算机技术与发展*, 2006, 16(9): 59-61.

(上接第 185 页)

参考文献:

- [1] 彭友,王延章. 信息系统内部安全审计机制[J]. *北京交通大学学报*, 2009(2): 112-116.
- [2] 王希忠. 安全审计在信息安全策略中的作用[J]. *信息技术*, 2010(3): 171-172.
- [3] 周洪昊. 安全审计系统的设计与实现[J]. *计算机应用*, 2004(7): 105-107.
- [4] 黄晨. 分布式安全审计系统设计与实现[J]. *计算机工程与设计*, 2007(4): 811-813.
- [5] 郭建东,秦志光,郑敏. 信息系统的安全模型[J]. *电子科技大学学报*, 2008(2): 285-288.
- [6] 邓小榕,陈龙,王国胤. 安全审计数据的综合审计分析方法[J]. *重庆邮电学院学报(自然科学版)*, 2005(5): 604-607.
- [7] 毕晓冬. 电子商务安全审计系统的研究与设计[J]. *工会论坛*, 2007(2): 94-96.
- [8] Agrawal R, Strikard R. Fast Algorithms for Mining Association Rules[C]//Proceedings of the 20th VLDB Conference. Santiago, Chile:[s. n.], 1994.
- [9] Klemettinen M, Mannila H, Ronkainen P, et al. Finding Interesting Rules from Large Sets of Discovered Association Rules[C]//Proceedings of the 3rd International Conference on Information and Knowledge Management (CIKM'94). Gainthersburg, MD:[s. n.], 1994: 401-407.
- [10] Stolfo S L, Promidis A L, Tselepis S, et al. JAM: Java Agents for Meta - Learning Overdistributed Databases[C]//Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. Newport Beach, CA: AAAI Press, 1997: 74-81.
- [11] 张世永. 信息安全审计技术的发展与应用[J]. *网络与信息安全*, 2003(2): 29-33.
- [12] 周琪锋. 基于网络日志的安全审计系统的研究与设计[J]. *计算机技术与发展*, 2009, 19(11): 139-142.