# 基于GEP的web服务器安全防护技术研究

龙 珑[1,2],宁 葵[1,2]
(1.广西大学 计算机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)

**摘 要**:目前网络安全问题日益严重,由于互联网开放性和通信协议的安全缺陷等原因使web服务器面临着越来越多的安全威胁。基因表达式编程(GEP)融合了遗传算法和遗传编程的优点,基于GEP的web服务器安全防护技术开发小组将GEP人工智能技术引入到web服务器,建立了一个多层次的安全防御模型。利用GEP算法的动态学习功能不断地提高安全防御能力,基于GEP的web服务器安全防护技术小组最终目标就是希望把web服务器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;web服务器;访问控制;多库协同
**中图分类号**:TP393 **文献标识码**:A **文章编号**:1673-629X(2011)10-0241-05

# Research of Web Server Security Technology Based on GEP

LONG Long[1,2], NING Kui[1,2]
(1. College of Computer Science and Electronic Engineering, Guangxi University, Nanning 530004, China;
2. Department of Computer Science and Information Technology, Guangxi Normal College, Nanning 530003, China)

**Abstract**: Nowadays there is a growing concern about the problem of internet security. As the internet and communication protocols open security holes and other reasons, web server is more and more confronted with security threats. Gene Expression Programming (GEP) combines genetic algorithms and genetic programming advantages, GEP artificial intelligence technology will be used for the web server to establish a multi-layered security defense model. Using dynanic learning function in GEP algorithm can improve the security and defense capabilities, the way can better improve the web server security.

**Key words**: GEP; web server; access control; multi-library collaborative

## 0 引 言

随着计算机应用的普及和网络技术的发展,企事业、学校等单位都相继建立了内部信息网络并设立了自己的web网站,并且越来越多的应用系统开始向internet平台转移,基于web的应用在全球被越来越多的公司和机构使用。很多企业在享受电子商务、CRM、ERP、EAI等带来的快捷便利的同时,却又被紧随其后的web服务器的非法访问入侵或安全问题所困扰。

国内外现行的web服务器安全技术主要包括防火墙、身份认证、授权控制、数据加密、日志审计、数据备份和恢复等安全管理技术,大部分有关的工作都以美国1985年的TCSEC标准为主要参照系。相关应用型产品包括有防火墙、安全路由器、安全网关、黑客入侵检测系统等,主要集中在系统应用环境的较高层次上,对于web服务器安全而言在完善性、规范性、实用性上还存在许多不足,特别是在多平台的兼容性、多协议的适应性、多接口的满足性方面存在很大距离,其理论基础和自主的技术手段也有待于发展和强化。我国在这些方面都比国外落后5~10年,到1999年10月发布了"计算机信息系统安全保护等级划分准则",该准则为安全产品的研制提供了技术支持,也为安全系统的建设和管理提供了技术指导[1]。

安全访问控制是对web服务器进行保护的重要措施,而传统的安全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。文中提出将基因表达式编程(Gene Expression Programming, GEP)[2-8]的人工智能技术引入到web服务器安全防护体系中,实现对计算机及其网络上复杂环境下的web服务器进行多角色、多途径、多等级、多条件的保护和应用,为用户提供了一种高度安全的互访途径。

**收稿日期**:2011-03-24;**修回日期**:2011-06-27
**基金项目**:国家级火炬计划重点项目(2007GH010246)
**作者简介**:龙 珑(1980-),男,硕士,高级工程师,主要研究方向为网络安全、数据挖掘。

## 1 系统中基因表达式模型

在 GEP 中，基于 GEP 的 web 服务器安全防护系统（下文简称系统）GEP 基因是一个线性的字符串，系统中一般为 K-表达式，用于表示预测函数 $f(x_1, x_2, \cdots, x_m)$。系统中 K-表达式由函数符及终止符组成，其语义等价于表达式树（expression tree，ET）[9~12]，系统中二者都对应了一个唯一的函数表达式。

下文给出 K-表达式、ET 和函数表达式的相互转化算法，以及在 GEP 技术 web 安全防护系统中 GEP 方法在系统中预测算法的地位。

### 1.1 系统中 K-表达式到 ET 算法

（1）GEP 系统中按从左到右的顺序遍历 K-表达式每个字符，如果还没遍历完转（2），如果遍历完成转（5）；

（2）GEP 系统中读取 K-表达式当前字符 c，如果树 T 为空转（3），否则转（4）；

（3）GEP 系统中用 c 生成树根节点，转（1）；

（4）GEP 系统中用 c 生成节点 N，在树 T 上按照从上而下，GEP 技术 web 安全防护系统中自左向右的顺序寻找第一个未挂满子树的节点 N '（即如果节点是 $n$ 元函数，那么节点需要 $n$ 个子树，如果是终止符，则不能挂子树），GEP 系统中将节点 N 生成为节点 N ' 的左数第一个子树，转（1），如果没找到 N '，则 exit success；

（5）GEP 系统中如果树 T 还存在没有挂满子树的节点，exit error；否则 exit success。

### 1.2 GEP 系统中 ET 转化为函数表达式

GEP 系统中子树的计算结果 Q 是其父节点（$N$ 元函数）的操作数，并且 Q 在函数式中的顺序对应于其作为子树在父节点下的顺序。GEP 中函数的运算顺序是从上而下，从左至右。下面一个例子说明 K-表达式、ET、函数表达式三者之间的转化。

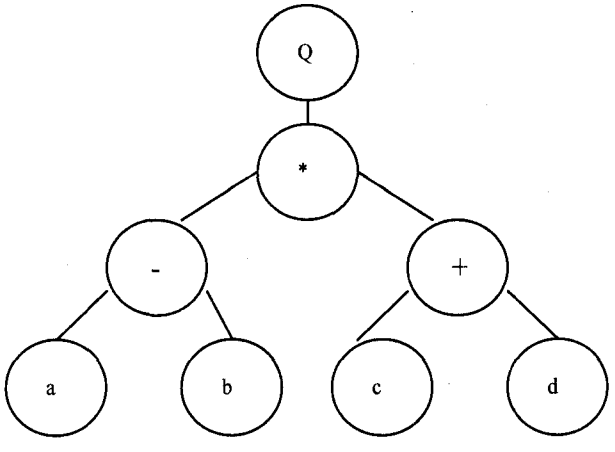


图 1 函数演示图

GEP 系统中 K-表达式：

Q * -+abcd

公式（1）转化为 ET，如图 1 所示。

$$\sqrt{(a-b)*(c+d)} \tag{1}$$

GEP 为保证 K-表达式的有效性（使其转化为 ET 的算法总返回 success），特作以下规定：

（1）GEP K-表达式由头部（head）、尾部（tail）组成，设其长度分别为 head1、tail1。

（2）GEP 头部既可以包含函数符又包含终止符，尾部只能由终止符组成。

（3）GEP head，body1 长度 head1 由实际问题确定，tail 必须满足以下条件：

$$\text{Tail} \geqslant \text{head1} * (n-1) + 1 \tag{2}$$

### 1.3 GEP 系统中 GEP 方法与系统预测算法

在 GEP 系统中预测函数的演化算法中，用 K-表达式表示系统预测函数，GEP 预测函数的演化在算法中实际表现为字符串的不断变化，系统可以做出相应变化。GEP 只要系统中的字符串（K-表达式）按照上文提出的 3 条规定来变化，GEP 中利用 K-表达式到 ET 的转化算法转化为确定预测函数。

## 2 GEP 基于群体搜索技术的演化算法

在 GEP 数据流预测 GEP 的演化中，GEP 采用函数优化算法（郭涛算法）[4] 对演化得到的预测函数做参数优化。

郭涛算法是一种基于子空间搜索（多父类重组）和群体爬山法结合的群体随机搜索算法，用来求解带不等式约束条件的函数优化问题。该算法有以下特点：

（1）采用了演化计算的群体搜索策略，保证了搜索空间的全局性，因此，特别有利于系统在全局范围内寻找最优的预测函数解集。

（2）系统算法采用了“劣汰策略”，每次只把群体中适应度最差的个体淘汰出局，淘汰压力最小，既保证了群体的多样性，又保证了适应性最好的预测函数个体可以保留。GEP 技术 web 安全防护系统中这种“群体爬山策略”保证了整个群体最后集体登上最高峰。

## 3 基于 GEP 的 web 服务器安全防护原理

GEP 技术 web 安全防护系统中要针对当前的几种主流 Web 服务器建立一种可靠的、适用性强的安全机制，必须要解决好以下问题：

（1）GEP 技术 web 安全防护系统中保证 Web 服务器用户身份验证安全：确保网络用户登录安全，防止对信息的非授权存取。很多安全技术（如防火墙技术），只是根据机器地址和端口来阻截非法访问。这种检查技术是很粗糙的，很容易被伪装，且它不能限制某台机

器的一部分用户进入网络,web 安全防护系统中而另一部分用户却可以登录到网络。要做到这一点,必须尽可能使用更为细致的用户身份认证技术。用户身份认证是用户对特定信息进行存取前所必须进行的屏障,对于非匿名的请求,访问用户首先需要向服务器证明其身份,GEP 技术 web 安全防护系统中以便让服务器决定是否授权用户做所请求的事。

(2)web 安全防护系统中保证 web 服务器对话期安全:确保 web 服务器的数据在 Internet/Intranet 上传输时不会被截获。当用户提供了正确的标志并被赋予了对数据的存取权限后,对话期安全保证了私有数据在整个对话期间不被截获和干扰。计算机网络和电话网不一样,GEP 技术 web 安全防护系统中它的基本协议中并没有建立一个点到点的连接,信息是在网络上进行广播并由被请求的机器进行应答。

(3)web 安全防护系统中访问 web 服务器的安全审计机制:安全审计机制是指 web 服务器主动防护系统设置相应的日志记录,特别是对数据更新、删除、修改的记录,以便日后查证。日志记录的内容可以包括操作人员的名称、使用的密码、用户的 IP 地址、登录时间、操作内容等。若发现系统的数据遭到破坏,可以根据日志记录追究责任,或者从日志记录中判断密码是否被盗,以便修改密码,重新分配权限,确保系统的安全。

针对以上这些问题建立了一个 GEP 技术 web 安全防护系统,该系统是位于操作系统和应用系统之间具有标准协议和接口的通用软件,依据国家保密技术规范的有关规定,以机密级涉密信息系统为对象,运用人工智能的分析手段对 web 服务器的身份鉴别、访问控制和审计跟踪等三个方面进行了控制和管理,从而提高 web 服务器在网络环境下的安全性。具体地说,就是利用 GEP 强大的自学习功能和函数发现功能以及过滤规则生成功能,实现以下功能:一是 GEP 技术 web 安全防护系统利用用户访问行为模式分析,复杂的过滤规则生成,从而识别和预防恶意攻击;二是利用 GEP 函数发现技术进行用户访问模式分类,提高用户类型识别的实时性,从而达到利用 GEP 技术在线防范的目的。GEP 技术 web 安全防护系统中该防护系统是构建在 Linux/Windows 平台上的主动防护系统,GEP 技术 web 安全防护系统应用时只需在服务器上安装即可。系统由身份鉴别、访问控制、安全审计和智能分析四部分组成,其网络结构如图 2 所示。

基于 GEP 的 web 服务器安全防护系统的应用流程是这样的:客户端用户在每次登陆或者访问远程服务器前首先要进行身份鉴别,鉴别信息的传输和存储采用密码处理,确保其安全保密性和完整性。用户身份鉴别通过后,在访问具体的信息前,采用 GEP 的智能方法来进行综合分析和权限控制,只有经过授权允许访问该信息的用户方可访问。GEP 技术 web 安全防护系统用户在整个会话过程中与安全相关的信息将记录为审计日志。

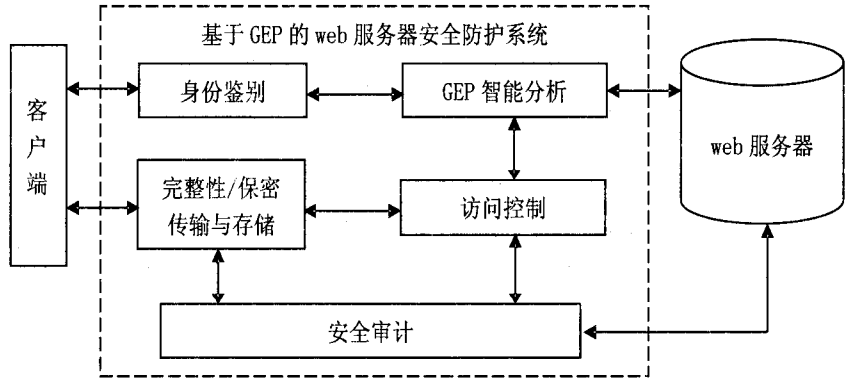


图 2　系统网络结构

## 4　基于 GEP 的 web 服务器安全防护流程

本系统中访问行为智能化分析功能的主要实现的目标是:针对访问 web 服务器的行为提供一种报警和控制的措施,以便规范用户行为。本功能是利用 GEP 强大的函数发现功能以及 IF-THEN 规则生成功能,GEP 技术 web 安全防护系统建立一个智能化、集成化、协调化的专家系统。系统建立了一个多库协同的模型,建立复杂的 IF-THEN 规则和分类模型,GEP 技术 web 安全防护系统实现用户访问行为模式分析,提高用户类型识别的实时性,GEP 技术 web 安全防护系统从而识别和预防恶意攻击,达到利用 GEP 技术在线防范的目的。基本功能流程如图 3 所示。

系统采用多库协同的模式,建立了数据库、方法库、知识库等,数据库存贮了有关的数据和结果,系统方法库存放着一些有关的数学计算模型、方法和程序,GEP 技术 web 安全防护系统知识库存放着一些有关计算机网络领域中专家性、规律性的知识。

系统调度控制器是在网络模型的基础上,建立起的一种多库之间、知识库与推理机之间的协同策略。因为,对于每一个具体的项目来说,输入输出的形式都是固定的,推理关系也是相同的,不同的是推理过程中具体的内容,所以,GEP 技术 web 安全防护系统每一个行为基本信息项目都有自己的调度控制模式。调度控制器是连接各信息库和功能模块的枢纽,主要是依靠编程手段来实现的;在 GEP 技术 web 安全防护系统的

调度、运作流程中，GEP 技术 web 安全防护系统知识表达和推理过程是最核心的内容。下面详细介绍一 DY] 们在本系统的实现原理：

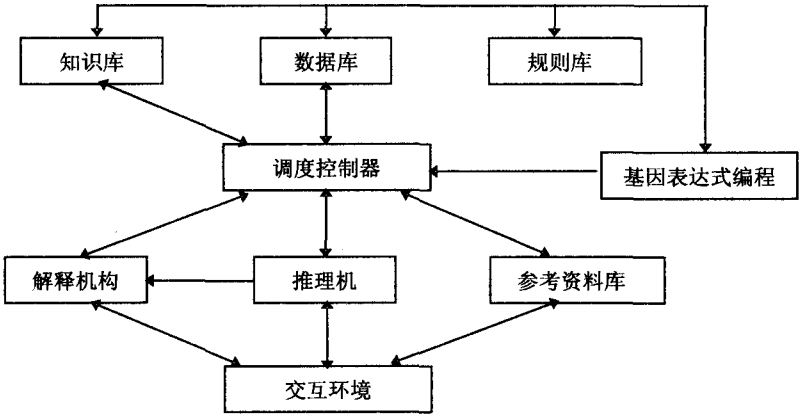


图 3 访问行为智能化分析功能流程图

(1)GEP 技术 web 安全防护系统知识表达。

专家系统强调人工智能活动是以知识为中心，知识库是本系统中重要的组成部分。本 GEP 技术 web 安全防护系统的知识库是一个由许多评定规则组成的集合，GEP 技术 web 安全防护系统中每条规则抽象为前提(P)、结论(C)和可信度(CF)赋值，GEP 技术 web 安全防护系统中编程实现的形式为：

IF < P > THEN < C > WITH < CF = ? >

GEP 技术 web 安全防护系统中：

$$P = \overset{n}{\underset{i=1}{\theta}} P_j$$

$n$ 为每条规则的前提数，GEP 技术 web 安全防护系统中 $\theta = \{$ AND, OR $\}$，GEP 技术 web 安全防护系统中 $P_i$ 为规则前提集合 $\{ P_1, P_2, \cdots, P_n \}$ 中的元素，GEP 技术 web 安全防护系统中 CF = ? 为每条规则的可信度赋予某一数值。

可信度的引入是为了更好地表达知识的模糊性和不确定性，GEP 技术 web 安全防护系统中的可信度区间定义为[0,1]，GEP 技术 web 安全防护系统中可信度的大小由多位心理专家研究和系统反复实验相结合的途径来获取。

GEP 技术 web 安全防护系统中的规则生产，采用 GEP 技术进行自动生产，具有人工干预少，自动化程度高，系统精度高的特点。

(2)GEP 技术 web 安全防护系统中推理过程。

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，建立相关分类模型(可以离线保存供后面调用)，GEP 技术 web 安全防护系统然后利用 GEP 生成的 IF-THEN 规则和用户类型进行推理，提高 GEP 技术应用的实时性，达到在线防范的目的。

系统推理过程采用数据驱动、正向的不确定性推理策略，推理的实质是把知识规则链接起来，GEP 技术 web 安全防护系统形成一条或多条推理链。系统根据用户的测量结果，经过文字、图像的处理后，将相应的信息特征以数据库的形式保存，作为输入的事实供推理机使用。

具体过程是：

①调度控制器根据用户的请求，在数据库获取对应的特征事实，在方法库中获取对应的计算方法，GEP 技术 web 安全防护系统经过计算和处理后得到新的事实特征，然后一方面保存到数据库，另一方面送往推理机。

②推理机读入特征事实，用这些特征事实与知识库中规则的前提条件进行匹配，GEP 技术 web 安全防护系统将匹配成功的规则的结论返回调度控制器，再存入数据库。

③将上一步骤生成的、保存在数据库的中间结论又作为新的特征事实，然后再重复以上的步骤。

④当无新的特征事实生成时，GEP 技术 web 安全防护系统调度控制器就达到一种稳定的状态，这时推理过程结束，并输出评定结果。

同时，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，GEP 技术 web 安全防护系统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。GEP 技术 web 安全防护系统具有免疫特性和自适应性。

## 5 GEP 技术 web 安全防护系统测试实验与分析

GEP 技术 web 安全防护系统主要构建在 windows 系统平台上，利用 C 和 Delphi 开发 GEP 算法。本次实际测试用了一台内存 4G、CUP2.40GHz 的 IBM 服务器，测试 GEP 在 web 服务器的预测效果。

GEP 技术 web 安全防护系统 GEP 方法的主要运行参数如下：

(1)GEP 技术 web 安全防护系统种群参数(见表 1)；

表 1 GEP 技术 web 安全防护系统中种群参数

| | | | |
|---|---|---|---|
| 系统种群规模 | 100 | 系统进化代数 | 1000 |
| 系统基于重组概率 | 0.1 | 系统 IS 变换 | 0.1 |
| 系统单点重组概率 | 0.4 | 系统两点重组概率 | 0.4 |
| 系统 RIS 变化概率 | 0.1 | 系统基因变换概率 | 0.1 |

(2)系统函数模型中的函数集，系统函数集合：+, −, ×, /, $ex$, $In$, $sqrt$, cos, sin, $pow(10,x)$，系统终止

符集合:$x_{n-9+1},\cdots,x$;

(3)系统基因头长 headl 为可变参数,根据式(2)使 taill=headl+1 即可保证 GEP 技术 web 安全防护系统基因表达式的有效性;

(4)系统基因数 5;

(5)系统连接函数+;

(6)系统单点杂交概率 0.4;

(7)系统 2 点杂交概率 0.4;

(8)系统变异概率 0.044;

(9)系统郭涛算法参数 $N=20, M=7$。

本次测试 GEP 技术 web 安全防护系统主要为了观测相对误差△与系统的预测函数模型的基因头部长度 headl 的关系。

图 4 给出了测试实验结果。

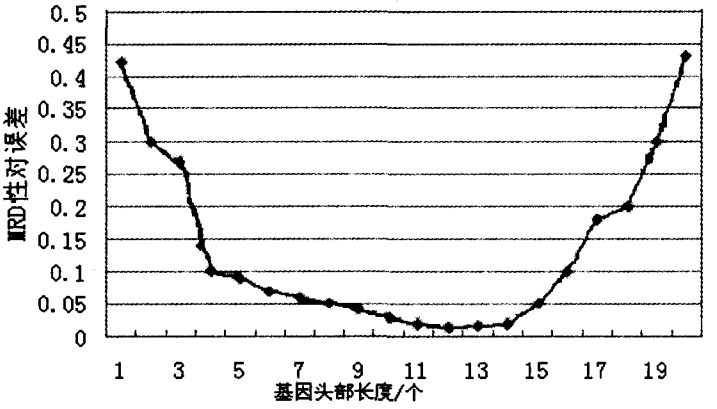


图 4　GEP 技术 web 安全防护系统基因头部长度的选取对相对误差的影响

图 4 表明 GEP 技术 web 安全防护系统在使用基因表达式头部的长度必须要取得适当的数值。如果数值太小,会导致函数复杂度降低,从而使演化在一个局部范围内收敛;如果太多也不好,因为 GEP 算法的时间复杂度与基因的长度成正比,GEP 技术 web 安全防护系统表达式太长会使演化跟不上数据流的变化,从而导致 GEP 技术 web 安全防护系统预测失败。

## 6　结束语

研究了传统安全网关系统、防火墙、入侵检测系统、防病毒系统等在安全应用上存在的缺陷和侧重点的不同,得出结论: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,而安全访问控制技术是用来抵御攻击的最后屏障,这对于 web 服务器而言也不例外。基于 GEP 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。

利用基于 GEP 的访问控制技术为 web 服务器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 服务器的安全问题具有十分现实的意义。

**参考文献:**

[1] 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):2057-2064.

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

[3] Ferreira C. Gene expression programming:mathematical modeling by an artificial intelligence[M]. New York:Springer Verlag,2002.

[4] Zuo Jie, Tang Changjie, Li Chuan, et al. Time Series Prediction Based on Gene Expression Programming[C]// WAIM04 (International Conference for Web Information Age 2004). [s. l.]:[s, n.],2004:55-64.

[5] Yuan Changan, Tang Changjie, Zuo Jie, et al. Attribute Reduction Function Mining Algorithm Based on Gene Expression Programming[C]// Proceedings of 2006 International Conference on Machine Learning and Cybernetics (ICMLC2006). [s. l.]:[s. n.], 2007: 1007-1012.

[6] Yuan Changan. Markov Convergence of Gene Expression Programming[C]//Proceedings of the 4th International Conference on Impulsive and Hybrid Dynamical Systems. [s. l.]:[s. n.],2007:2725-2728.

[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] 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.

[9] 陆昕为. 一种改进的 GEP 方法及其在演化建模预测中的应用[J]. 计算机应用,2005,25(12):102-104.

[10] 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.

[11] Ferrira C. Mutation, transposition, and recombination: An analysis of the evolutionary dynamics[C]// Proceeding of the 6th Joint Conference on Information Science, 4th International Workshop on Frontiers in Evolutionary Algorithms. USA:[s. n.],2002:389-408.

[12] Ferrira C. Genetic representation and genetic neutrality in gene expression programming[J]. Advanced in Complex System, 2002,5(4):389-408.