

# 基于免疫遗传算法的网格入侵检测模型

李 钦, 余 谅

(四川大学 计算机学院, 四川 成都 610064)

**摘 要:** 分析研究了网格安全的特点以及网格环境下的入侵检测技术, 并针对传统入侵检测技术难以适应动态的网格计算环境等问题, 根据生物免疫原理提出了一种基于免疫遗传算法的网格入侵检测模型。该模型采用了一种将免疫算法和遗传算法结合使用的混合算法, 与标准遗传算法相比, 该算法既保留了遗传算法随机全局并行搜索的特点, 又在很大程度上避免了未成熟收敛, 确保快速收敛于全局最优解; 算法具有较好的效率和收敛性, 提高了该模型的多样性和自适应性。

**关键词:** 网格安全; 入侵检测; 虚拟组织; 免疫遗传算法

**中图分类号:** TP393.08

**文献标识码:** A

**文章编号:** 1673-629X(2009)05-0162-04

## Grid Intrusion Detection Model Based on Immune Genetic Algorithm

LI Qin, YU Liang

(College of Computer Science, Sichuan University, Chengdu 610064, China)

**Abstract:** The characteristic of grid security and intrusion detection technology are analyzed in this article. Being that conventional intrusion detection systems can not adapt to the dynamic grid environment, a grid intrusion detection model based on application of biology immunity theory and immune genetic algorithm is proposed, which adopts an algorithm blending by immunity algorithm and genetic algorithm. Compared with the standard genetic algorithm, this algorithm reserves the capability of random global parallel search of genetic algorithm, avoids immaturity convergence in great sense, and assures convergence to the global optimized solution. The algorithm has better efficiency and constringency, and improves the diversity and self-adaptability of the model.

**Key words:** grid security; intrusion detection; virtual organization; immune genetic algorithm

### 0 引言

网格是近年来兴起的一种重要信息技术, 而网络安全是网格技术的核心问题<sup>[1]</sup>。网格环境中的安全问题比通常的分布式环境中的安全问题更加复杂, 将入侵检测技术应用到网格领域中可以大大提高网格的安全性, 在网格安全管理中有着重要的作用。入侵检测技术是动态安全技术的最核心技术之一, 其目标是通过检查操作系统的审计数据或网络数据包信息来检测系统中违背安全策略或危及系统安全的行为或活动, 从而保护信息系统的资源不受攻击, 防止系统数据的泄漏、篡改和破坏。入侵检测系统在发现入侵后, 应及时响应, 包括切断网络连接、记录事件和报警等。但是传统的入侵检测方法都是从定义入侵模式开始, 然后把采集的数据与事先定义的模式进行匹配, 使系统失

去了多样性和自适应性。将免疫遗传算法引入网格入侵检测系统中, 使它的自适应性和鲁棒性在网格入侵检测系统中发挥作用, 可以取得更好的效果。

### 1 问题分析

#### 1.1 网格及网格安全的特点

网格把分散在互联网各节点上的计算资源、存储资源、数据资源、信息资源、知识资源、专家资源等整合为一台巨大的虚拟超级计算机。在此计算机中, 以网格计算为基础的各种网格应用与使用的资源类型、数量、节点等不断变化, 所以网格入侵检测模型的设计必须考虑网格环境的如下特性:

- (1) 用户数量庞大, 且在不断变化, 参与者变化的频率较高。
- (2) 资源数量庞大种类繁多, 且动态可变。
- (3) 一个服务可能要求在它的执行期间动态地请求、启用或释放资源。
- (4) 构成服务的进程可以用不同的机制进行通

收稿日期: 2008-09-03

基金项目: 四川省科技计划项目资助(2006j13-101)

作者简介: 李 钦(1979-), 男, 河南洛阳人, 硕士研究生, 研究方向为网络应用与网络安全; 余 谅, 副教授, 研究方向为计算机应用。

信,包括单播和多播。程序执行期间,低级别的通信连接(例如 TCP/IP 套接字)可能被动态地创建或撤销。

(5) 资源可支持不同的认证和授权机制,包括 Kerberos、明文口令、全套接协议 (SSL)、SecureShell (SSH)。

(6) 出于计帐和访问控制的考虑,成员在不同的资源上可能要映射到对应不同的本地名字空间或本地帐号。

(7) 资源和用户可属于多个组织。

网格环境不可避免地会遇到一些意外的和蓄意的安全威胁,包括对网格环境中的资源、数据以及基础设施的完整性、保密性和可利用性的安全威胁。根据其后果,网格入侵行为<sup>[2]</sup>大致可以分为:非授权访问、信息泄漏或丢失、破坏数据完整性和拒绝服务攻击。

### 1.2 人工免疫原理在 GIDS 上的应用

生物体免疫系统与入侵检测系统有着功能上的相似之处<sup>[3]</sup>。免疫系统的分布性、多样性、自成体系、完备性和精简性使它精确有效地保护着生物个体。这使得人们希望借助生物免疫原理<sup>[4]</sup>更好地实现入侵检测的功能,在合法的“自己”行为中判别出非法的“异己”行为。在免疫系统中起检测作用的是抗体,抗体的生成演化和工作过程以及如何模拟基因库更新、否定选择、克隆选择等抗体生成过程建立入侵检测器是建立基于免疫的入侵检测系统的关键。合格的入侵检测系统 (Intrusion Detection System, IDS) 要具有准确性、完整性、可扩展性、可适应性和自身的健壮性等特点<sup>[5]</sup>。而网格入侵检测系统不仅要满足一般入侵检测系统的特性,还必须能够兼容于网格环境,并且能够充分利用网格环境中的资源。这就要求它至少达到以下几个设计目标:(1) 完备的;(2) 分布式的;(3) 自组织的和精简的。

为了便于表达和理解,根据功能比较将生物体免疫系统的有关术语和网格入侵检测系统的有关概念对照见表 1。

## 2 一种基于免疫遗传算法的网格入侵检测模型

### 2.1 模型设计

网格环境中的入侵检测模型同样应该包括一般入侵检测模型所必需的组成部分,但是由于网格环境的

特殊性,它们又与一般的入侵检测模型有所区别:

表 1 生物体免疫系统概念和网格入侵检测系统概念对比

缩氨酸/抗原决定基	被检测的行为模式串
抗体	检测模式串
单克隆淋巴细胞 (T-细胞、B-细胞)	检测器
抗原	异己模式串
绑定	检测模式串和异己模式串的匹配
耐受性(阴性选择)	否定选择
淋巴细胞克隆	检测器复制
抗原检测	入侵检测系统的检测
抗原清除	检测器响应

① 网格中存在着大量独立的计算资源,必然会产生大量的数据流,为了充分发挥网格资源共享的优势,减轻数据检测的负担,应设置多个检测服务器,并由目录服务器进行资源分配,而不是采用集中式检测结构。

② 一些网格漏洞攻击的踪迹可能分布于多个主机,只有联合分析多个主机的日志才能发现此攻击,另外网格用户允许在同一时间使用不同域内的多个资源,只分析某一个主机上的日志不足以发现误用或滥用的攻击行为,因而入侵检测必须广泛地分布于网格范围内。

③ 对于网格环境中特有的协同攻击,需要对数据进行全局综合分析,因此需要中心分析服务,以及中心分析服务之间更广域范围内(整个网格环境中)的交互。

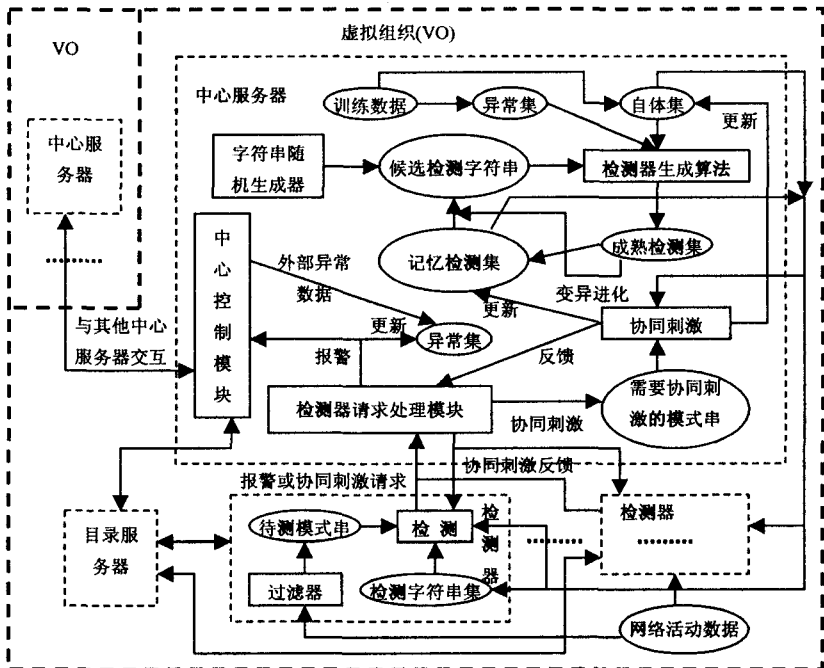


图 1 基于免疫原理的网格入侵检测模型

文中提出的网格入侵检测模型以虚拟组织<sup>[6]</sup>为单位,由检测器、中心分析控制器以及抗体库组成,虚拟

组织环境下的基于免疫原理的网格入侵检测模型的总体框架如图1所示。虚拟组织是一个动态环境,它包含多个基于若干资源共享规则和限定条件定义的个体和(或)机构。因此,可以将虚拟组织看作是网格整体环境的一个细胞,完成了细胞的免疫功能以及细胞间的交互,就完成了网格整体的入侵检测功能。在该模型中,检测器采用结合基于主机和基于网络的IDS技术收集各个主机节点的日志以及网络数据包;设置一个资源目录服务器,虚拟组织内的资源在目录服务器端进行注册,由目录服务器根据注册资源的状态,通过最优算法的计算来发现最佳的可用资源并完成资源分配;在及时响应用户数据检测需求之后,各个检测器将过滤后的数据传送给虚拟组织内惟一的中心分析控制器(中心服务器),以综合分析整个虚拟组织范围内的数据,发现网格特有的攻击,同时中心服务器与网格中的其他虚拟组织中的中心服务器交互,进行综合处理。

## 2.2 模型工作流程

系统的运行可分为训练期(Training Phase)和检测期(Detecting Phase),训练期中,系统在安全状态下收集网络活动的模式字符串来构造自体集,同时根据已知的入侵特征模式构造异常集。当自体集和异常集达到一定规模后,就可以进入检测阶段。

检测阶段中,系统生成候选检测字符串并通过检测器生成算法生成检测字符串集,将其发送到各个检测器,同时更新记忆检测集(抗体库),并保证各个检测器获得的检测集都是抗体库的一个有一定覆盖面的不同子集。

当系统收到用户的入侵检测请求后,由目录服务器为用户分配一个或数个可用的检测器,检测器将网络活动数据处理成模式字符串,并计算抗体与抗原亲和度与检测集进行匹配。如果匹配则向中心服务器报警并把该模式串发送过去;否则发出协同刺激请求。中心服务器综合各个检测器的信息判断是否是入侵行为,若确认入侵则发出报警信号,并将该模式串送到异常集中,而被确认为自体的模式串则发往自体集;然后向发出请求的检测器发送确认信号,同时删除错误检测的检测串并更新抗体库。

## 2.3 模型的改进策略

为了进一步提高系统的检测速度和准确率,本模型采用了多层次的协同刺激策略:①检测器内部协同:与两个以上检测串匹配的模式串才被判定为入侵,未与检测集匹配的模式串再与自体集比较,都无法匹配的才向中心服务器发出协同刺激请求;②中心服务器内部协同:中心服务器先用自身数据库中的数据进行判定,仍无法判定的由外部进行人工协同;③中心服

务器间协同:通过各中心服务器之间的交互实现了整个网格环境下的协同。采用这种协同刺激策略大大减少了外部人工干预,同时也降低了网络通信的开销。

网格中的活动是在不断变化的,为了提高系统对网格环境的自适应性,本模型采取了以下策略:①抗体库的进化同时考虑了抗体的检测能力和抗体的浓度,并在候选抗体中加入了成熟检测集的变种以及随机抗体,保证了抗体的检测效率和多样性;②检测器周期性检查检测集,计算出各检测串与抗原的匹配次数和死亡时间,并同步更新抗体库中相应记忆串的匹配次数,若检测串超过死亡时间或匹配次数过少,就申请一组新的检测串将其替换;③按照②中的方法周期性检查抗体库中的记忆串,对于超过死亡时间或匹配次数过少的记忆串,则从抗体库中删除并重新加入到候选抗体中;④定期更新和存储抗体库、自体集以及异常集,并通过与其它中心服务器的交互以及人工协同操作,进一步提高系统的可靠性。

## 3 相关算法

### 3.1 算法描述

本模型的检测器生成算法采用了免疫遗传算法,它是一种将免疫算法和遗传算法结合使用的混合算法。基于免疫原理的遗传算法<sup>[7]</sup>与标准遗传算法相比,具有如下显著特点:①具有免疫记忆功能,该功能可以加快搜索速度,提高遗传算法的总体搜索能力;②具有抗体的多样性保持功能,利用该功能可以提高遗传算法的局部搜索能力;③具有自我调节功能,这种功能可用于提高遗传算法的全局搜索能力,避免陷入局部解。

文中提出的免疫遗传算法的程序框图如图2所示。

### 3.2 算法关键步骤

为了提高算法的效率和适应性,算法中采用了否定选择<sup>[8]</sup>运算和遗传算子运算。无论是随机产生的新抗体,还是通过遗传算子运算得到的新抗体,都有可能与自体匹配。否定选择就是先用自我库来排除与自体匹配的候选抗体,从而提高算法的时间性能。为了能对未知抗原产生免疫应答,需要通过遗传算子运算来产生新抗体,采用的方法包括多点交叉和变异。为了进一步提高遗传算子运算的效率,将遗传算子的算法步骤设计如下:(1)对新产生的抗体按交叉概率 $P_c$ 进行多点交叉操作,然后重新进行否定选择及亲和度计算以决定是否接受新个体。若接受,则令新抗体为新种群的个体;否则,令旧抗体为新种群的个体;(2)对上一步产生的抗体按变异概率 $P_a$ 进行变异操作,同样按(1)

中的方法决定是否接受新抗体;(3)生成新一代抗体。当进化代数达到预先设定的数值或者抗体平均浓度达到一个稳定的范围时,算法结束,生成的记忆抗体集合即为最优成熟检测集。

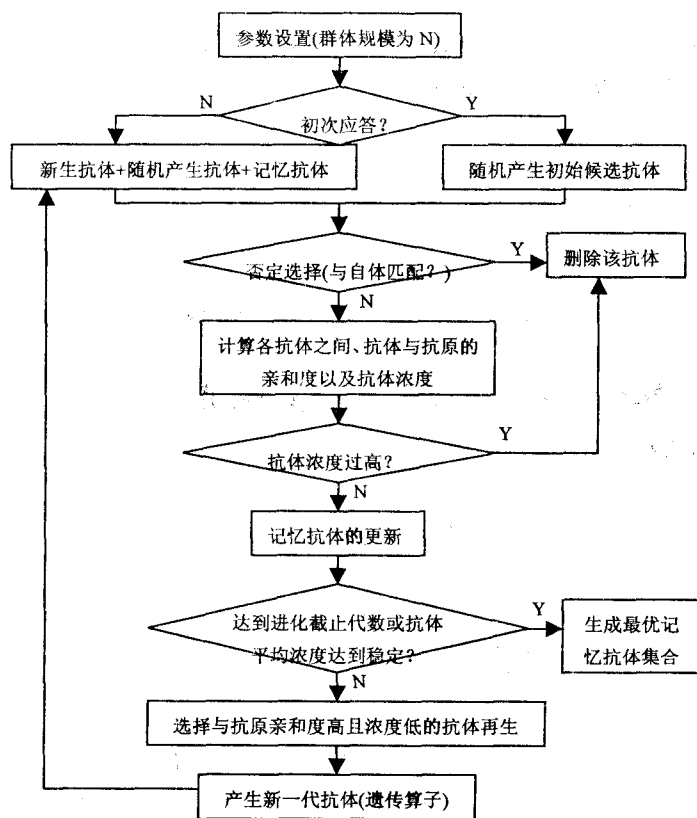


图2 检测器生成算法程序框图

## 4 算法性能分析

### 4.1 实验环境及参数设置

(1) 开发环境:CPU: Intel Pentium 4. 2.4 G; 内存:512 M;硬盘:80 G;操作系统:Windows XP。

(2) 开发工具:采用 C 语言。

(3) 参数设置:参数的取值是否恰当对于算法的性能有着相当重要的影响:种群进化代数和种群规模太大会降低执行速度,太小又可能得不到最优解或仅得到局部最优解;交叉概率和变异概率决定了进化的效率;而前面公式中的各个阈值的取值则关系到抗体的检测效率以及抗体的多样性。此外匹配阈值、检测串和记忆串的死亡时间以及周期性检测的时间周期的设置也与整个系统的运行效率有很大的关系。

### 4.2 实验结果及对比分析

通过理论分析和多次试验,发现选取下面一组参数值时,算法的性能较为理想:种群进化代数  $M = 500$ ,种群规模  $N = 100$ ,交叉概率  $P_c = 0.7$ ,变异概率  $P_a = 0.003$ ,匹配数阈值  $\alpha = 4$ ,适应度阈值  $\delta_1 = 0.1$ ,

相似度阈值  $\delta_2 = 0.9$ ,抗体浓度控制阈值  $\delta_3 = 0.6$ 。

图3是本算法在采取上面的参数取值时与标准遗传算法(相对应的参数设置相同)的抗体适应度变化情况对比图,从图中可以看出当群体收敛的时候,本算法的平均适应度更接近最佳情况,而且本算法的收敛速度也更快一些。

## 5 结束语

目前网格入侵检测技术还处于理论研究阶段,依然是一个热点问题。文中结合网格环境中的安全特点以及人工免疫原理在入侵检测技术中的应用,讨论了网格环境中入侵检测模型的建立,提出了一种基于免疫遗传算法的网格入侵检测模型,通过免疫原理提高了系统的自适应性、高效性和可靠性,能够满足网格环境下的入侵检测需求。然而,在广泛的网格环境中,各种基于不同策略的入侵检测模型都有其自身的局限性,如何处理在具有不同安全策略的虚拟组织之间的信息交互,以及如何在检测速度与检测准确率之间达到平衡,使系统发挥出最优性能,都将是今后工作的重点,有待进一步研究。

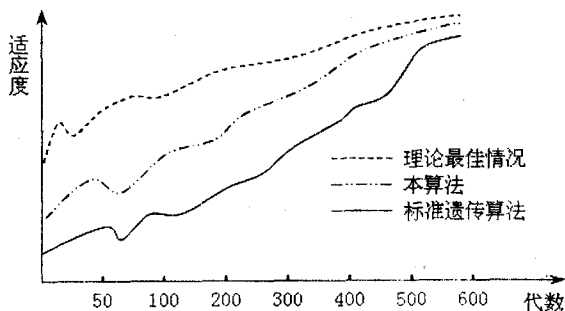


图3 本算法与标准遗传算法的适应度变化情况对比图

### 参考文献:

- [1] Humphrey M, Thompson M, Jackson K R. Security for grids [J]. IEEE Grid Computing, 2005, 93(3): 644-652.
- [2] Schuster A, Reis J A, Koch F, et al. A Grid-based Intrusion Detection System [C] // the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies. Mauritius: [s. n.], 2006: 187-187.
- [3] 杨向荣, 沈军毅, 罗 浩. 人工免疫原理在网络入侵检测中的应用[J]. 计算机工程, 2003, 29(6): 27-30.
- [4] 李 涛. 计算机免疫学[M]. 北京: 电子工业出版社, 2004.
- [5] Forrest S, Hofmeyr S A, Somayaji A. Computer immunology

(下转第169页)

中,硬件采用 Intel X86(Pentium IV 2.7G,1GB 内存)计算机作为平台。软件采用 Red Hat 9.0 操作系统,GT-Nets 仿真平台,底层 RTI 实现采用 libSynk,gc++ 编译环境。实际拓扑结构如图 3 所示。

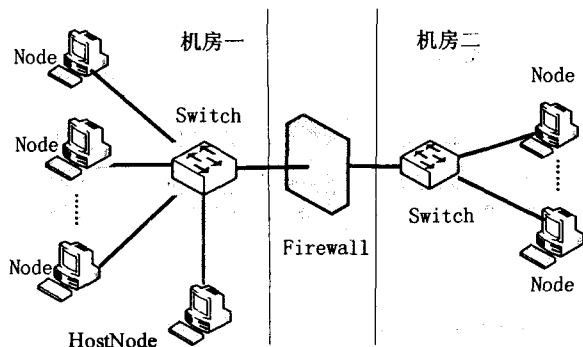


图3 蠕虫自动扩散的实验拓扑

拓扑中具有 HostNode(扩散发起者)结点,Firewall 结点和 Switch 结点。

实验假设如下:①未进行扩散时,实验中的网络物理链路和通信链路正常工作;②每次扩散成功的状态都能被捕获到。

实验时仿真网络参数配置如下:仿真执行时间为 10s;基本带宽 100Mb。

文中研究中针对采用加权算法和没有采用加权算法的两种策略各进行了 20 次扩散仿真实验,实验结果平均值如图 4 所示。

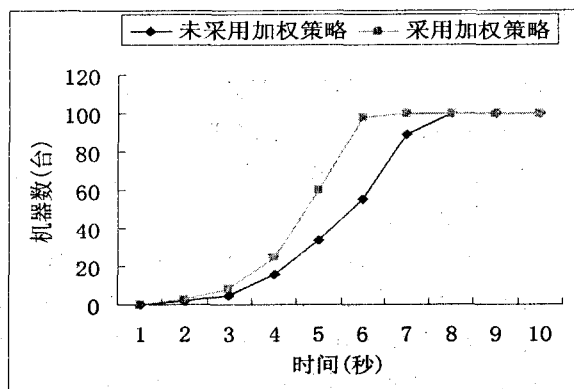


图4 两种蠕虫扩散算法的效率对比

实验结果表明,两组数据之间存在着较大差异。

①没有采用加权算法的蠕虫在第 5s 仅感染了 37 台主机;而此时采用加权算法的蠕虫已经感染了 61 台主

机,5s 内扩散的速度效率是前者的 1.65 倍。②没有采用加权算法的蠕虫在第 8s 感染所有主机;而采用加权算法后在第 6s 就完成所有任务,完全扩散的时间效率比前者提高 25%。

由此可见,采用加权策略算法的蠕虫,其扩散速度、扩散时间都得到较大改善。

#### 4 结束语

通过引入结点之间的顺序关系以及权值的概念对扩散树进行改进,较好解决了蠕虫扩散策略和环境之间如何相适应的问题。在此基础之上,设计一种基于加权策略树的自动扩散模型,并给出相应的队列遍历算法来实现。

仿真结果说明该模型既能使蠕虫在相同或相近网段中快速扩散,又能使蠕虫有较好的环境适应能力,能根据环境的变化动态调整扩散策略。

#### 参考文献:

- [1] Tidwell T, Larson R, Fitch K, et al. Modeling Internet Attacks[C]//Proceedings of the 2001 IEEE Workshop on Information Assurance and Security. Oakland, CA:[s. n.],2001.
- [2] 周伟,王丽娜,张焕国.一种基于攻击树的网络攻击系统[J].计算机工程与应用,2006(24):38-40.
- [3] Kienle D M, Elder M C. Recent Worms: A Survey and Trends[C]//Proc. of the ACM CCS Workshop on Rapid Malcode. Washington D. C.:[s. n.],2003.
- [4] Gebhart G. Worm Propagation and Countermeasures[R].[s. l.]:SANS Institute,2004.
- [5] Bishop M, Bailey A. Critical Analysis of Vulnerability Taxonomies[D].Davis:Department of Computer Science, University of California,1996.
- [6] 黄家林,姚景周,周婷.网络扫描原理的研究[J].计算机技术与发展,2007,17(6):147-150.
- [7] 丁常福,方敏,徐亮.端口扫描技术及防御分析[J].微机发展(现更名:计算机技术与发展),2003,13(6):7-12.
- [8] 陈峰,罗养霞,陈晓江.网络攻击技术研究进展[J].西北大学学报:自然科学版,2007,37(2):208-211.
- [9] Sedgewick R, Flajolet.算法分析导论[M].冯舜玺,等译.北京:机械工业出版社,2006.

(上接第 165 页)

- [J].Communications of the ACM,1997,40(10):88-96.
- [6] Foster I,Kesselman C,Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations[J]. International Journal of Supercomputer Applications,2001,15(3):1-10.
- [7] 王煦法,张显俊,曹先彬,等.一种基于免疫原理的遗传

算法[J].小型微型计算机系统,1999,20(2):117-120.

- [8] Forrest S, Perelson A S, Allen L, et al. Self-Nonself Discrimination in a Computer[J]. Proceedings of IEEE Symposium on Research in Security and Privacy. Oakland:[s. n.],1994.