

# 基于模糊集的免疫克隆选择算法

张 葵,袁细国

(武汉科技大学 计算机科学与技术学院,湖北 武汉 430081)

**摘 要:**分析了将人工免疫原理应用到入侵检测系统中存在的不足。为了克服传统的基于精确数学模型免疫算法的局限性,提出了一种基于模糊集理论的免疫克隆选择算法,引入了模糊集合和隶属度的概念,采用了一种动态的智能优化策略,有效地改善了检测元的特性,提高了检测元在复杂网络环境下的适应能力,从而增强了网络的安全性。分析了该算法能在一定程度上弥补反向选择算法的不足。

**关键词:**入侵检测;人工免疫;克隆选择算法;模糊集合

**中图分类号:**TP393.08

**文献标识码:**A

**文章编号:**1673-629X(2007)12-0024-03

## An Immune Colon Selective Algorithm Based on Fuzzy - Set

ZHANG Kui, YUAN Xi-guo

(Sch. of Computer Science & Technology, Wuhan University of Science & Technology, Wuhan 430081, China)

**Abstract:** After analyzing the inadequacy occurring in the application of artificial immune theory to intrusion detection systems, in order to overcome the limitation lied at the traditional immune algorithm based on exact mathematic model, proposes an immune colon selective algorithm based on fuzzy - set. Introduces the concept of fuzzy - set and degree of membership, effectively optimizes the detectors' characteristics. The detector adaptive is improved to gear to the complex network environment and the computer security is enhanced. Finally, analyze the new algorithm which can supplement the shortages of negative selection at a certain degree.

**Key words:** intrusion detection; artificial immune; colon selective algorithm; fuzzy - set

## 0 引 言

自然免疫系统和入侵检测系统(Intrusion Detection Systems, IDSs)有着惊人的类似,建立一种较为智能化的检测模型是引入免疫机制原理。Forrest<sup>[1]</sup>、Kim等免疫专家将人工免疫系统(Artificial Immune System)应用于IDSs中,并进行了客观的分析和深入的研究。Forrest认为,免疫系统最重要的任务是如何准确识别本体和异物,生物免疫系统使用诸如蛋白质片断(肽)等特征来完成本体的识别,而计算机系统同样具有多个特征可供选择,如主机上资源访问的模式、主机网络流量信息等。因此,采用生物免疫系统的特征可以有效地提高IDSs的检测效果。

在日新月异的网络环境下,入侵检测系统应不仅能够监测多参数,还应该较好地解决大数据量的问题,Forrest<sup>[2]</sup>提出用否定选择算法(Negative Selection Algorithm)来产生成熟检测元,并利用R连续位匹配算

法来判定模式间的匹配程度。但是在Kim和Bentley的研究中表明,否定选择算法只能对网络通信的一个小的子集起效,在处理真正大量网络通信数据时会产生较大的缩放问题,而且产生的大部分成熟检测元的适应度较低,检测结果不太理想。针对上述问题提出了一种基于模糊集的免疫克隆选择算法,该算法建立了一种新的检测元适应度的计算方法,力图将模糊数学理论<sup>[3]</sup>和生命科学的免疫概念结合起来并引入到工程实际领域,借助其中的有关知识和理论并将其与已有的一些智能算法有机地结合起来,提高了算法的整体性能,有效地改进了人工免疫模型。

## 1 相关研究及问题分析

Forrest, Perleson<sup>[2]</sup>等人在否定选择算法(NSA)中提出采用穷举检测元生成算法建立合乎标准的检测元集合。这种算法定义Self集为正常的网络行为,Non-self集为异常的网络行为。随机选择生成一个候选检测元集合,把每个检测元和自体集合中的元素做匹配比较。如果匹配成功,则放弃该检测元;如果不匹配,则把它放入合格的检测元集合中。

收稿日期:2007-02-26

基金项目:湖北省教育重点科研项目基金(2004D006)

作者简介:张 葵(1970-),女,湖北武汉人,硕士研究生,研究方向为模糊信息处理、计算机网络安全。

Hofmeyr<sup>[4]</sup>提出了一种分布式的免疫检测模型,通过反向选择算法产生检测元,建立系统正常行为模式定义为“自体”,随机生成的大量模式和“自体”进行模式匹配,摒弃匹配的随机模式,剩余的作为“检测元”,模仿抗体功能,这些检测元用于对所有进入系统的新的行为模式进行检测,匹配者认为是“非己”的非法操作。

以上成熟检测元的构造方法中,在一定的程度上都存在以下不足:

#### (1) 没有动态覆盖性。

成熟检测元往往是在系统相对静止的状态下产生的,并且只适合于在当前的网络环境下进行检测。而在实际应用中,计算机网络系统是一个实时的持续变化的系统,今天被认为是正常的行为到了明天就可能成为极度危险的行为,这样势必会造成较大的错误否定率。

#### (2) 缺乏优化能力。

由于候选检测元的随机生成以及自体集的不够完善,会不可避免地产生过多的冗余成熟检测元,大大降低了检测系统的效率。

针对上述的问题提出了一种新的基于模糊集的免疫克隆选择算法,建立了一个模糊检测元集合,用于综合评判克隆选择的精英个体。引入了隶属度,作为检测元的属性,并给予一个动态可调整的阈值,这样不仅控制了成熟检测元集合的容量,即具有较高的覆盖性,而且还能实现成熟检测元集的自我调整和优化算法。

## 2 知识储备

### 2.1 模糊数学理论基础

#### (1) 模糊集合。

模糊集合是模糊数学的理论基础,描述了一种模糊的现象,反映了一类“亦此亦彼”的模糊性,是内涵和外延都不明确的集合。

#### (2) 隶属函数、隶属度。

在给定的论域(也可说问题域)  $U$  上规定一个映射  $A: U \rightarrow [0, 1]$   $u \mapsto A(u)$ 。

则称  $A$  为  $U$  上的模糊(Fuzzy)集,  $A(u)$  称为  $A$  的隶属函数,也称为  $u$  对  $A$  的隶属度。

### 2.2 相关定义

#### (1) 状态空间。

在克隆选择算法中,用二进制编码来表现个体的基因型,它使用的编码符号集由二进制符号 0 和 1 组成,其优点在于编码、解码操作简单,而且便于利用模式定理进行理论分析等。状态空间用集合  $S = \{0, 1\}^l$  表示,设问题域  $U \subseteq S$ 。

#### (2) 模糊检测元集合。

优秀检测元用一个模糊集  $A$  来表示,  $A = A_a = \{u \mid A(u) \geq \alpha\}$ , 其中  $\alpha$  为阈值,用于控制优秀个体。根据每个个体当前所在模糊集中隶属度的等级,可以确定它们各自的克隆及存活率。

## 3 免疫克隆选择算法

在一个实时的持续变化的计算网络系统中,针对现有 IDSs 中存在的不足,提出了一种基于模糊集的免疫克隆选择算法,构造了一种新的检测元适应度的计算方法,采用了精英选择的克隆规则,有效地改善了 IDSs。

### 3.1 检测元的表示及适应度(fitness)的计算

#### 3.1.1 检测元的表示

使用一个一维的结构体数组来表示检测元的集合。结构体的构造如下:

```
typedef structure
{int count; /* 表示该检测元识别入侵行为的次数 */
int score; /* 表示检测元与入侵行为的匹配程度 */
float fitness; /* 表示该检测元的适应度值 */
}Detector[Num]
```

其中 Num 是检测元种群的个数。

#### 3.1.2 适应度的计算

适应度函数的选取至关重要,直接影响到克隆选择算法的收敛速度以及能否找到最优解。一个好的适应度函数是能够尽可能地使得每个检测元都有同等的竞争机会并且能够取得优胜劣汰的效果。

隶属函数:

$$A(\text{det}) = \frac{(\omega_1 B(\text{det}) + \omega_2 C(\text{det})) * 2}{B(\text{det}) + C(\text{det})} \quad (1)$$

其中  $B(\text{det}) = \text{count}$ , 表示检测元 det 识别入侵行为的次数,  $C(\text{det}) = \text{score}$ , 表示检测元与入侵行为的匹配程度,  $\text{score} = \text{检测元} \oplus \text{入侵行为}$ ,  $\oplus$  为异或运算,  $\omega_1$  和  $\omega_2$  为权值 ( $\omega_1 + \omega_2 = 1$ ,  $\omega_1 \geq \omega_2$ ), 隶属度  $A(\text{det})$  表示检测元的适应度值 fitness, 取值范围为 0 ~ 1。

具体实现过程如下:

第一步 随机产生一定大小的网络入侵行为种群, 设为  $M$ ;

第二步 从检测元种群中选择一个检测元;

第三步 采用  $r$ -连续位匹配规则, 计算检测元 det 识别网络入侵行为的个数, 若为 count, 同时计算检测元 det 与  $M$  个网络入侵行为之间的一致程度, 把它作为检测元 det 的得分 score;

第四步 利用公式(1) 综合计算检测元 det 的适应度 fitness;

第五步 重复以上步骤, 直至检测元种群适应度计

算完毕。

图 1 中只是对种群样本计算的结果,令  $r = 3$ ,从图中可以看出,  $\text{det.count} = 2$ ,  $\text{det.score} = 16$ , 那么  $\text{det.fitness} = A(\text{det}) = \frac{(\omega_1 B(\text{det}) + \omega_2 C(\text{det})) * 2}{B(\text{det}) + C(\text{det})} = \frac{(\omega_1 * 2 + \omega_2 * 16)}{9}$ , 若取  $\omega_1 = \omega_2 = 0.5$ , 则  $\text{det.fitness} = 1$ 。在实际中,通常取  $\omega_1 > \omega_2$  为宜,  $\omega_1 = 0.8, \omega_2 = 0.2$ , 那么,  $\text{det.fitness} = 0.533$ 。图 2 为网络入侵行为群样本。

count	score	
2	16	01101 (det)
2	16	01011
1	13	10010
2	16	01100
2	14	10101
		...

图 1 检测元种群样本 图 2 网络入侵行为群样本

### 3.2 模糊检测元集合的建立

模糊检测元集合  $A$  表示如下:

$$A = A_\alpha = \{u \mid A(u) \geq \alpha\}$$

其中  $\alpha$  为阈值,由公式(1)可计算出检测元  $u$  的适应度,即  $u$  对集合  $A$  的隶属度。那么  $A$  则是这样一个水平集合,由在整个状态空间  $U$  中的一切隶属度大于  $\alpha$  的元素组成。

实际上,集合  $A$  中总是保留了优秀的检测元,在采用精英选择的克隆选择算法中,只有集合  $A$  中的个体才有再生(克隆)的机会。

### 3.3 算法的描述

通过否定选择算法初始化检测元种群;利用上文中适应度的计算方法对种群进行 fitness 的计算;从中选择优秀的个体建立模糊集合  $A$ ,即具有克隆机会的精英检测元的集合;利用克隆选择繁殖产生子检测元;通过否定选择算法过滤掉不成熟的检测元;得到新的成熟检测元集后,再进行克隆选择过程。算法如图 3 所示。

### 3.4 算法分析

通过隶属函数(式(1))来计算检测元种群的 fitness 值,进而建立模糊检测元集合  $A$ 。在 fitness 值的计算过程中,综合考虑了检测元与 Nonself 集的匹配频率以及匹配程度,可以更好地评价检测元的可适应度。克隆选择算法在新的 fitness 计算方法的基础上,采用了精英选择。即优秀检测元总是保留在模糊集  $A$  中,根据每个检测元对  $A$  的隶属值来赋予它们各自的繁殖程度。其中  $\alpha$  值的设置较为关键,它直接影响到检测元是否有再生(克隆)的机会。

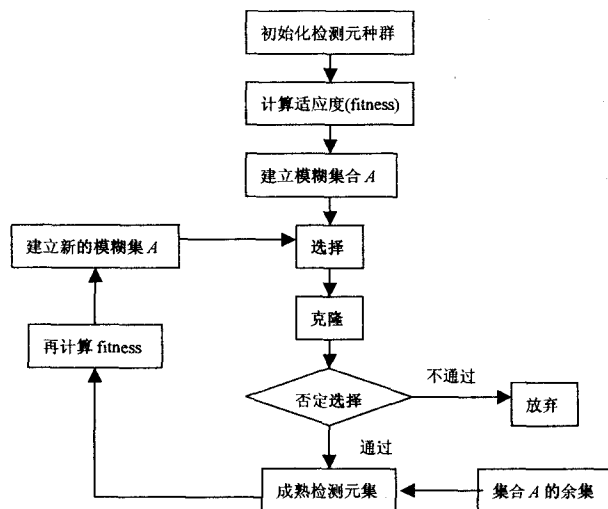


图 3 免疫克隆选择算法

在克隆选择算法中,通过高频变异和受体编辑来实现检测元的多样性。高频变异机制是为了使免疫应答能够快速成熟,适应度低的检测元可能会经历更加深度的变异,如果它没有出现更好的情况,则会在否定选择中被淘汰。但是,对于高适应度的检测元,其变异可能会被抑制。受体编辑提供了一种消除局部极值的能力,变异会导致局部极值,受体编辑进行更大范围的搜索,有可能找到更好的检测元。

由于变异可能会导致更加糟糕或者没有任何功能的检测元,如果一个刚经历了有益变异的检测元在下一次的免疫应答中继续保持相同的频率进行变异,那么不良变化的聚集可能抵消有利的变异。为了有效地降低自免疫反应情况的出现,算法在经过克隆后还引入了反向选择算子对无效的检测元进行过滤,从而减少了错误肯定率。

## 4 结论

提出了一种基于模糊集的免疫克隆选择算法,该算法能够产生较为完善的成熟检测元集合,并在以下几个方面作了改进:(1)建立了一种新的检测元适应度的计算方法,平滑了将检测元一分为二划分的严格界限;(2)建立了优秀检测元的模糊集合,降低了检测元的冗余度并减少了优秀检测元的流失。同时,在基于精英选择的克隆过程中,总保持了检测元的择优准则。以上特点不仅在一定程度上弥补了反向选择算法的不足之处,而且提高了克隆选择算法的局部搜索能力和总体搜索能力。

### 参考文献:

- [1] Hofmeyr S A, Forrest S. Architecture for an Artificial Im-

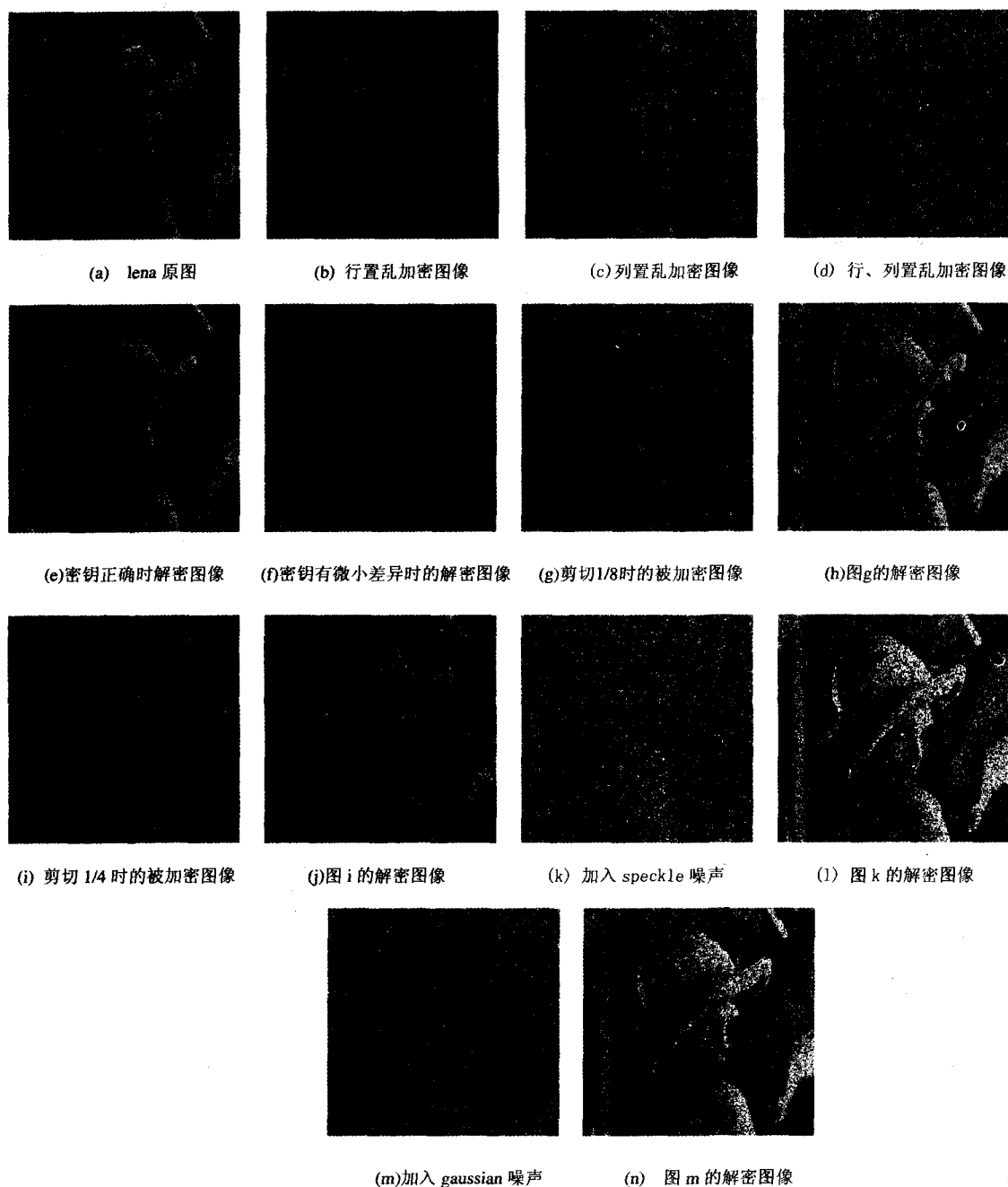


图 1 仿真结果

- [J]. 计算机工程, 2005, 31(9): 34-36.
- [10] 王 毅. 基于混沌序列的图像加密研究[J]. 计算机工程与应用, 2002(20): 99-102.
- [11] 易开祥, 孙 鑫, 石教英. 一种基于混沌序列的图像加密算法[J]. 计算机辅助设计与图形学学报, 2000, 12(9): 672-676.
- [12] 齐东旭, 邹建成. 一类新的置乱变换及其在图像信息隐蔽中的应用[J]. 中国科学(E辑), 2000, 30(5): 440-447.
- [13] 孙 鑫, 易开祥, 孙优贤. 基于混沌系统的图像加密算法[J]. 计算机辅助设计与图形学学报, 2002, 14(2): 136-139.
- [14] 刘向东, 焉德军, 朱志良, 等. 基于排序变换的混沌图像置乱算法[J]. 中国图像图形学, 2005, 10(5): 656-660.

(上接第 26 页)

- immune System[J]. Evolutionary Computation, 2000, 8(4): 443-473.
- [2] Forrest S, Perleson A, Allen L, et al. Self-Nonself discrimination in a computer[C]// In: Proceedings of IEEE Symposium on Research in Security and Privacy. Oakland, USA: [s. n.], 2002: 202-212.
- [3] Zhang Jian. An Anomaly Detection Method Based on Fuzzy Judgment[J]. Journal of Computer Research and Development, 2003, 40(6): 776-783.
- [4] Hofmeyr S. An immunological model of distributed detection and its application to computer security[D]. New Mexico: Dept of Computer Science, University of New Mexico, 1999.