基于复合免疫算法的入侵检测系统

冯 翔 马美怡 赵天玲 虞慧群
(华东理工大学信息科学与工程学院 上海 200237)

摘 要 计算机安全系统与生物免疫系统具有很多的相似性,它们都需要在不断变化的环境中维持自身的稳定性。提出复合免疫算法,并应用到入侵检测系统中,以保护网络安全。针对经典的人工免疫算法在性能上存在的缺陷进行了改进,完善了其核心算法——否定选择算法,在否定选择算法中加入了分段技术和关键位,避免了恒定的匹配概率导致的匹配漏洞,降低了系统漏检率。并将遗传算法中的克隆选择算法和改进的否定选择算法结合为复合免疫算法,提高了检测器生成的动态性和多样性。最后,通过数学理论分析与仿真实验模拟,验证了改进算法的有效性和可行性,并且与其它经典算法进行了比较,结果证明,改进算法可以提高系统性能。

关键词 人工免疫算法,入侵检测,否定选择算法,生物免疫系统,克隆选择算法
中图法分类号 TP309,TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.12.010

Intrusion Detection System Based on Hybrid Immune Algorithm

FENG Xiang MA Mei-yi ZHAO Tian-ling YU Hui-qun
(School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract Computer security system and biological immune system have much comparability, so the artificial immune algorithm can be applied in intrusion detection system to solve various problems in the field of computer security. After studying the classical algorithm named negative selection algorithm, it was discovered that the matching algorithm would cause the examination black hole. A novel hybrid immune algorithm was proposed to solve the intrusion detection problem. The effectiveness and feasibility of the improved algorithm were verified. This paper partitioned the match string and set different coefficient for each section, thus to eliminate the problem that the r-continual position match algorithm has the constant match probability in the reverse choice algorithm, and to reduce the missing rate of intrusion detection system. This paper also combined the negative selection algorithm with the clonal selection algorithm. This will increase the reproduction, selection and intersection into the produce of detection. Thus the missing rate will be reduced. At last, we compared and analyzed the different parameters, including the section number, threshold value and r-continual parameter.

Keywords Artificial immune, Intrusion detection, Negative selection algorithm, Biological immune system, Clonal selection algorithm

1 引言

生物免疫系统通过抵御各种异己细胞来保护人体免受外界侵害,相似地,计算机通过入侵检测系统来监测和抵御不安全信息[1]。所以,生物免疫系统抗体抵御抗原的机制被应用于计算机网络安全中,进行错误诊断、病毒检测、入侵检测等[2,3],可以改善检测系统的响应时间和检测效率与正确性。

基于人工免疫算法(AIS)的入侵检测系统(IDS)受到更多的关注,许多新技术[4,5]被加入到这一模型中。在分布式智能电网入侵检测系统中,应用AIS算法和支持向量机来检测和分类恶意数据和可能的网络攻击,识别恶意网络流量,提高了系统的安全性[6]。基于Agent人工免疫的主机入侵检测系统[7]与多个嵌入系统的Agent之间相互协作可以有效地检测输入信号和临时的输出信号,输出危险或安全信号。智能计算具有适应性、容错性、高运算速度和恢复嘈杂的信息错误的特点,因此人工神经网络、模糊系统、进化计算、人工免疫系统、群体智能、软计算等智能计算方法能在入侵检测系统中得到广泛应用[8]。

鉴于人工免疫算法与入侵检测的关联,J. Timmis 和 A. Hone[9]综述了现有的AIS理论,分析了3个主要类型的AIS算法:克隆选择、免疫网络的选择和阴性选择算法。文献[10]提出了一种基于AIS的分布式代理入侵检测模型。该

到稿日期:2013-12-19 返修日期:2014-01-28 本文受国家自然科学基金(60905043,61073107,61173048),上海市教育委员会科研创新项目,中央高校基本科研业务费资助。

冯 翔(1977-),女,博士,教授,博士生导师,CCF会员,主要研究方向为分布并行计算、人工智能、网络通信,E-mail:xfeng@ecust.edu.cn;马美怡(1989-),女,硕士生,CCF学生会员,主要研究方向为分布并行计算、计算机网络;赵天玲(1987-),女,硕士生,主要研究方向为分布并行计算、计算机网络;虞慧群(1967-),男,博士,教授,博士生导师,CCF会员,主要研究方向为软件工程、可信计算、云计算。

模型动态免疫监视期间模型以提高的自我学习的能力，提供了良好的网络监控解决方案。文献[11]研究了不同的 AIS 理论以及如何结合不同的思路来解决网络安全领域的问题。混合动力系统结合了异常检测和误用检测[12]，其初始网络连接基于 AIS。更进一步，文献[13]解释了生物学和计算机网络之间的关系，全面概述了受生物启发的相关研究。Visconti 提出了一个基于间隔类型 2 的模糊集的人工免疫系统，来模仿自组织免疫系统的运作模式[14]。文献[15]则成功应用 AIS 机器学习技术来解决入侵检测问题。

本文提出的复合免疫算法针对经典的人工免疫算法的缺陷进行了改进，完善了其核心算法——否定选择算法，在否定选择算法中加入了分段技术和关键位，避免了恒定的匹配概率导致的匹配漏洞，降低了系统漏检率；并将遗传算法中的克隆选择算法和改进的否定选择算法相结合，提高了检测器生成的动态性和多样性。最后，通过仿真实验模拟，验证了改进算法的有效性和可行性，并且与遗传算法和人工免疫算法的系统进行了比较，通过分析证明了改进算法可以提高系统性能。

2 入侵检测问题描述

定义 1 设入侵检测的问题域为一个有限长的符号序列 $X, X\in\{0,1\}$ 或 $X\in\{0\cdots255\}^{(l/8)}$。$X$ 包含两个子集，即自我集 S 和非我集 N，其中 $S\cup N=X, S\cap N=\emptyset$。对于一个给定入侵字符串 $I: I\in\{0,1\}^l$，一个检测器集 $D: D=\{\alpha_1,\alpha_2,\cdots,\alpha_l\}, \alpha\in\{0,1\}^k, k\leqslant l$ 匹配模式为 $f: f(I,\alpha)\rightarrow\{p:\mathcal{R}\mid p\geqslant0\wedge p\leqslant1\}$，匹配阈值如式(1)所示：

$$match(f,\lambda,I,D)=\begin{cases}nonself, & f(I,a)\geqslant1-\lambda\\ self, & \text{other}\end{cases}\tag{1}$$

对于病毒检测，非我集代表恶意病毒代码，自我集代表合法良性程序。对于入侵检测，非我集代表网络攻击中的 IP 包，自我集为正常的认可的网络交易服务。检测算法的任务是将一个入侵模式 $I\in X$ 分为自我和非我。

这种检测方法可能产生两种错误：假阳性错误和假阴性错误。一个假阳性错误 δ^+ 表示一个自我集 S 被误判为非我的。相反地，一个假阴性 δ^- 表示非我集 N 被误认为是自我的，$(I\in S\cap match(f,\lambda,I,D)=nonself)\rightarrow\delta^+$，$(I\in N\cap match(f,\lambda,I,D)=self)\rightarrow\delta^-$。

定义 2(检测器的生成效率) 候选检测器 N_R 与自体集 N_S 的关系如式(2)所示：

$$N_R=\frac{\ln P_f}{\ln P_M*(1-P_M)^{N_S}}\tag{2}$$

其中，P_f 为漏检率，P_M 为检测器与一个随机字符串的匹配概率，定义如式(3)所示：

$$P_M=m^{-r}[(l-r)(m-1)/m+1]\tag{3}$$

式中，m 表示字符串的字幕数目，r 为匹配阈值，l 为字符串长度。在本系统中，检测器由于是系统随机生成的，大多数的候选检测器被随机抛弃，因此检测效率低下。

定义 3(漏检率) 漏检率的产生过程如下所示：

$s_1=a_1\cdots a_k b_{k+1}\cdots b_{k+r-1}c_{k+r}\cdots c_l$

$s_2=a_1'\cdots a_k' b_{k+1}\cdots b_{k+r-1}c_{k+r}'\cdots c_l'$

⇓

$h_1=a_1\cdots a_k b_{k+1}\cdots b_{k+r-1}c_{k+r}'\cdots c_l'$

$h_2=a_1'\cdots a_k' b_{k+1}\cdots b_{k+r-1}c_{k+r}\cdots c_l$

假设自体集中 s_1 和 s_2 有相同的连续位 $b_{k+1}\cdots b_{k+r-1}$，与 h_1 和 h_2 匹配的字符串也必与 s_1 和 s_2 匹配，则 h_1 和 h_2 即称为“黑洞”。

当自体集数量变大时，漏检率会随之增大，究其原因，一是检测器是随机产生的，存在一定的重复率，二是与自体的分布和匹配阈值有关。

定义 4(匹配阈值) P_f 和 P_M 的关系如式(4)所示：

$$N_R=\frac{\ln P_f}{\ln(1-P_M)}\tag{4}$$

一个性能优良的入侵检测系统要求以尽量少的检测器覆盖尽量大的“入侵”空间，即 P_f 和 N_R 都尽可能小。可以看出，N_R 与 P_M 成正比，而 P_M 与匹配阈值 r 成正比，所以，N_R 会随着匹配阈值 r 的增加而增加，N_R 与匹配阈值 r 成正比。同时，P_f 与自体的分布和匹配阈值 r 有关，若 r 越大，则两个字符串有连续 r 位相同位的机率就越低，P_f 也就越小，因此 P_f 与匹配阈值 r 成正比，N_R 与匹配阈值 r 成正比。匹配阈值 r 的选择是矛盾的，r 选值偏大，会造成检测器集 N_R 偏大，r 选值偏小，则会造成漏检率偏大。因此，选择一个合适的匹配阈值 r 对入侵检测性能有很大的影响。

3 复合免疫算法

本文将克隆选择算法与改进的否定选择算法相结合，提出了一个基于改进的否定选择操作克隆选择算法的复合的人工免疫算法，以更加准确快速地检测入侵。

3.1 改进的否定选择算法

传统免疫算法采用否定选择算法，r 连续位匹配规则生成检测器集，采用该算法生成检测器集时，r 连续位恒定的匹配概率将导致匹配过程中不可避免地出现漏检或误检，即“黑洞”。系统中某些非法字符串，找不到有效的检测器检测到它。

为了消除“黑洞”，本文通过将字符串分段，并根据不同段的重要程度设置不同的匹配权值来降低“黑洞”发生的概率。算法描述如算法 1 所示。

算法 1

(1)将待匹配的两个字符串初始化为 L_1，L_2 两个二进制字符串；

(2)将 L_1，L_2 两个字符串分成 N 段，每段长为 l，设置两个字符串匹配阈值为 λ；

(3)对一些关键的字段，设置关键位 $Key_{i,j}$(i 表示字符串的第 i 段，j 表示第 i 段的第 j 个字符)；

(4)为字符串的每个字段设置匹配阈值 φ_i，φ_i 的大小根据每段的重要程度而定。其中($\varphi_1+\varphi_2+\cdots+\varphi_N=1$)；

(5)设置循环 i=0；

(6)i++。如果片段 i 上有关键位 $Key_{i,j}$，则转步骤(7)。否则，在片段 i 上采取 r 连续位匹配规则，若匹配，则 Match(i)=1；否则，Match(i)=0；

(7)如果关键位 $Key_{i,j}$ 与片段 i 上关键位相同，则 Match(i)=1；否则，Match(i)=0；

(8)如果 i<N，则转到步骤(5)；否则，转到步骤(9)；

(9)计算 φ_1Match(1)+φ_2Match(2)+…+φ_NMatch(N)=f；

(10)若 f<r，则认为两个字符串 L_1，L_2 不匹配，否则，二者匹配。

3.2 复制和否定选择

在人工免疫系统中，抗原被分为两组：“自我”抗原(人体自身细胞)和“非我”抗原(入侵细胞)。基于否定选择的克隆

选择算法，使匹配非自体抗原的抗体复制，而匹配自体细胞的抗体则被抛弃。对初始数据进行聚类分析，两个点 x 和 y 之间的距离采用欧几里德距离。若两个点之间的距离 $d(x,y)<\varepsilon$，则认为 x 和 y 为一类。最后将初始数据分为两类：自体抗原集和非我抗原集。

在评估检测器生成过程中产生的检测器后，人工免疫系统选择父类检测器来复制检测器后代。所有检测器中最坏的W%检测器将被新产生的检测器后代中最好的B%检测器所替代，然后被送到人工免疫系统中以供使用。新产生的B%的检测器后代要经过否定选择算法来确保数据有效性。整个复制流程图如图1所示。

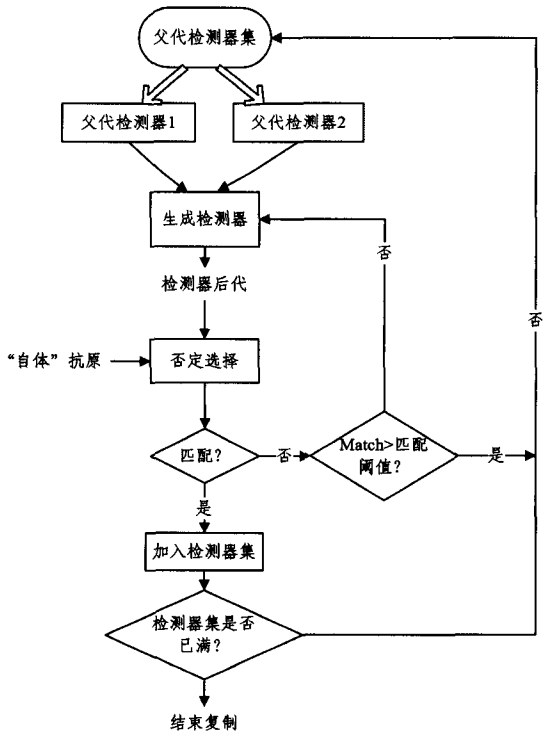

图1 复制流程图

3.3 变异和交叉

本文中，克隆选择算法用到了遗传算法中的两种操作：变异和交叉。

变异实际上是对检测器进行小概率的变动。通过对二进制字符串个别字符进行改变，实现对检测器的变异。

交叉就是结合两个父代检测器的遗传信息来产生新的检测器后代。本文中检测器表示为二进制字符串，故交叉采用单点式交叉。单点交叉从父代检测器集中随机选择两个检测器A和B，对A和B进行交换遗传信息。如上面所示，对A和B随机选择交换点，对交换点以后的部分进行交换，形成交叉后的子代检测器A和B。

算法2 复合免疫算法

(1)初始化随机产生二进制字符串集。

(2)使用聚类算法将字符串集分为"自我"集和"非我"集。

(3)根据"非我"集生成检测器模板集。

(4)将"非我"集与检测器模板集进行亲和力计算，采用改进的否定选择算法，使用r连续位匹配规则，亲和力大于指定阈值的被放入检测器集，否则被抛弃。

(5)将检测器集中的父代检测器进行克隆选择，并与"自我"集进行否定选择，删除与"自我"集匹配的检测器。

(6)检测器集大小是否大于指定大小。若大于，则转(7)；否则，转(4)。

(7)实时监测网络数据，将实际数据与检测器集进行否定选择，如匹配，则认为入侵，报警。

4 算法性能与数学分析

4.1 检测范围分析

设检测器检测范围即覆盖空间为 X，则 X 与检测器长度 l、字符串长度 n 和匹配阈值 r 的关系函数如式(5)所示：

$$X(l)=n^l-Y(l) \tag{5}$$

其中，$Y(l)$ 表示检测器不与之匹配的字符串个数，如式(6)所示。

$$Y(l)=\begin{cases} n^l-n^{l-r}-(l-r)n^{l-r-1}, & r\leqslant l\leqslant 2r \\ 2Y(l-1)-Y(l-r-1), & l>2r \end{cases} \tag{6}$$

从式(5)和式(6)可以看出，检测器覆盖范围随着 r 的增加而减少，因此，可以通过减少 r 来增加检测器覆盖范围，同时，减少 r 值又会导致检测器生成个数变少。所以，选择一个适合的 r 值对于人工免疫系统性能的提高有着决定性的作用。

定义5(两个字符串的相似程度) 给定字符串 x 和字符串集 R，$R=\{x_1,x_2,\cdots,x_k\}$，字符串 x 与 R 中字符串的相似度定义为集合 $\{r_{x,x_1},r_{x,x_2},\cdots,r_{x,x_k}\}$，则 x 与 R 的相似程度如式(7)所示：

$$q_{xR}=\max(\{r_{x,x_1},r_{x,x_2},\cdots,r_{x,x_k}\}) \tag{7}$$

定义6(检测器与自体集的相似程度) 检测器集 d 与自体集 S 的相似程度定义为 r_{dS}。对于任何一个检测器集 d，如果与自体集的相似程度为 r_{dS}，则对于任何的字符串 a，如果其与检测器集 d 的最大连续位匹配为 r_a，若 $r_a>r_{dS}$，则字符串 a 不属于自体集 S，即可以判定字符串 a 为非自体。

综上所述，只要满足 $r_a>r_{dS}$，就可以正确地识别出非自体，因此，当 $r=r_{dS}$ 时，检测器的覆盖范围为最大。如果取 $r>r_{dS}$，则浪费了检测器的检测能力，r 增加1，则检测器的覆盖范围可能减半。

本文以一个半径为 l 的圆来代表字符串全集 U，字符串的个数 N 表示全集 U 的体积，即 $V(U)=2^l$。自体集 S 为包含在全集 U 中的一个任意形状体，$V(S)=N_S$，N_S 为自我集的字符串个数，检测器集为包含在全集 U 中的圆状体，每个检测器集半径为 rd，$rd=l-r$，则 r 越大，rd 越小。当 $r=0$ 时，$rd=l$，此时，检测器集的覆盖空间为整个全集 U。当 $r=l$ 时，$rd=0$，此时，检测器的覆盖空间为0，如图2所示。

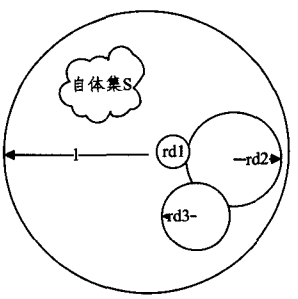

图2 检测器的覆盖空间

根据上面所述，两个字符串 s_1 和 s_2 的距离 dS 可以表示如下：

$$dS(s_1,s_2)=l-rmatch(s_1,s_2) \tag{8}$$

其中，l 为字符串长度。

4.2 预期漏检率的数学计算与分析

在该系统中，成熟检测器集中每个检测器都与自体集有一个相似度，即 r_s，每个检测器的自体相似度都各不相同，各自发挥其作用。

在否定选择算法中，假设检测器集中检测器与自体集匹配概率 P_M 是不变的，设自体集大小为 N_S，期待漏检率 P_f 与自体集大小 N_S 的关系如式(9)所示：

$$P_f=(1-P_M)^{N_S} \tag{9}$$

若检测器集中检测器与字体集匹配概率 P_M 是可变的，则 P_f 与自体集大小 N_S 的关系如式(10)所示：

$$P_f=\prod_{i=1}^{N_S}(1-P_{Mi}) \tag{10}$$

其中，N_S 为自体集大小，P_{Mi} 为检测器集中每个检测器与自体集的匹配概率。

在算法运行中，否定选择算法以 P_f 达到预期标准为准则，当达到标准时，循环结束。

5 仿真实验

为了验证改进后的算法的有效性与可行性，通过使用 r 连续位匹配规则，采用改进的否定选择算法和克隆选择算法，产生检测器集，将随机生成的入侵集与产生的检测器集进行匹配，若匹配，则视为入侵，加入检测器集；否则，视为正常，将其抛弃。最后分析基于改进的算法的入侵检测系统的性能。初始参数设置如表 1 所列。

表 1 初始参数设置

参数名称	数值
r 值(r 连续位匹配规则)	3
match(匹配阈值)	0.6
L(字符串长度)	10
自我集大小	50
入侵集大小	50
分段数	2
Match(1)	0.337
Match(2)	0.663
Key(i,j)	Key(1,3)

5.1 实验结果分析

(1)r 连续位与检测器生成时间、生成个数之间的关系如表 2 所列。

r 连续位匹配规则中，r 位数的大小直接影响该检测器的生成时间与个数，在本文中，通过采用不同的 r 位数来分析其对检测器生成性能的影响。

从表 2 可以看出，当 $L=20$，自我值和入侵值分别为 10 和 50，匹配阈值为 0.6 时，随着 r 值的增大，生成的检测器个数增加，但是相应消耗的时间也增大。因此，选择适宜的 r 值对于入侵检测系统的精确性与时间性能有着很大的影响。同时，由于采用带有关键位分段匹配算法，使得检测器生成时间有所增加。

表 2 r 值与检测器个数与生成时间的关系

r	生成时间	检测器个数
4	30765	76589
5	30987	76890
6	40865	304700
7	145613	593457
8	151627	900087
9	159469	925176
10	161473	1055543
11	163097	1345569

(2)连续位与漏检率的关系如图 3 所示。

从图 3 可以看出，漏检率与连续位数 r 成正比。r 越大，漏检率越高；r 越小，漏检率则越低。r 越大，则检测器匹配的精度越高，当 $r=L$ 时，则代表为完全匹配，此时，生成的检测器个数将减少，检测器之间的冗余就越小，同时与检测器匹配的入侵就越可能被漏掉。r 越小，生成的检测器就越多，但是漏检率则会降低。

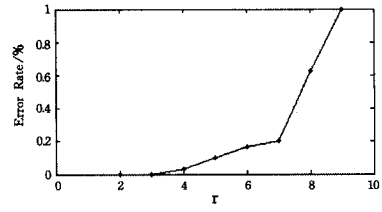

图 3 r 连续位与漏检率的关系

(3) 自我集的大小同样影响系统的性能。

自我集越大，检测效果则越高，但是同时，自我集过大，将增加系统时间开销，增长检测时间等。图 4 为自我集大小与系统 CUP 利用率的关系。

从图 4 可以看出，随着自我集的增大，系统 CPU 利用率增大。随着自我集的逐渐增大，系统 CPU 利用率趋于不变。

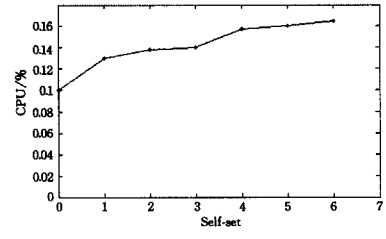

图 4 自我集大小与 CPU 利用率的关系

5.2 与其它算法比较与分析

将复合免疫算法与经典人工免疫算法、遗传算法的入侵检测系统进行对比分析，并通过设置几个主要的参数来对比 3 个算法的实现效果，主要通过检测器生成时间、生成个数、r 值与匹配阈值之间的关系来进行分析。

(1)不同的匹配阈值与检测器生成个数的关系的结果统计分析如图 5 所示。从图 5 可以看出，在相同的匹配阈值下，复合免疫算法中，检测器生成个数明显多于遗传算法和人工免疫算法。匹配阈值越大，检测器生成个数越小。复合免疫算法将否定选择算法与克隆选择算法相结合，父代检测器根据适应度计算生成后代检测器，后代检测器又有经过否定选择算法的检验，匹配的后代检测器被加入到检测器集中，如此循环直到检测器达到一定数量。因此，改进后的检测器数量在相同的匹配阈值下比其它算法产生的检测器数量更多。

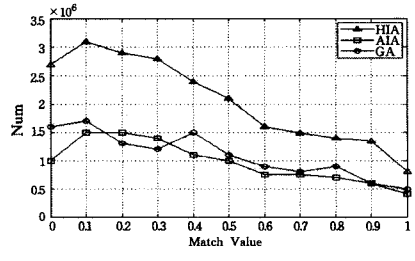

图 5 匹配阈值关系对比

(2)复合免疫入侵检测系统与遗传算法、人工免疫算法在

检测器生成时间与个数之间具有很大的差异，它们的性能均与 r 值的选取具有直接关系。不同的 r 值会导致不同的检测器性能。3 种算法的入侵检测系统 r 值与检测器生成时间、生成个数之间的关系如表 3 所列。

表 3　AIA、GA 和 HIA 3 种算法 r 值与检测器生成时间、生成个数之间的关系

R	AIA		GA		HIA	
	Time(ms)	Num	Time(ms)	Num	Time(ms)	Num
4	14265	68699	32126	52459	30765	76589
5	14187	68699	33927	68232	30987	76890
6	51844	304633	41263	282341	40865	304700
7	101418	591821	154212	435328	145613	593457
8	131627	800733	160463	770729	151627	900087
9	149469	925176	168234	893199	159469	925176
10	161297	988854	170765	832933	161473	1055543
11	178735	1020758	172263	920352	163097	1345569
12	185000	1035830	196235	936738	187005	1367894
13	189516	1042828	209234	1037531	190033	1347989
14	189536	1046016	1208232	1183581	199876	1378906
15	190003	1047456	2320213	1179516	2000003	1456976
16	190022	1048096	2435432	1200892	2045674	1501256

复合免疫算法在检测器生成时间上与人工免疫算法相比略有增加，但优于人工免疫算法。这是因为在检测器生成过程中，匹配算法采取分段的关键位匹配算法，在匹配时，需要判断当前字符在哪个字段上，并且要判断是否为关键位，因此，增加了系统的计算时间。同时，复合免疫算法生成的检测器个数比其它两种算法明显增加，这是因为算法把否定选择算法融入到克隆选择算法当中，优化了检测器的整体质量，加入了交叉、变异和复制过程，使生成的检测器更具有适应性。

(3)3 种算法基于不同的连续位数 r 下的系统漏检率的对比如图 6 所示。

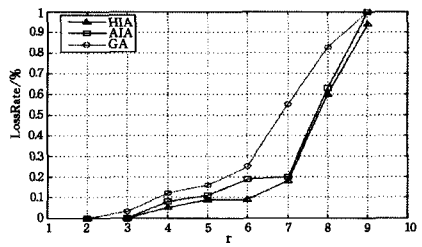

图 6　漏检率对比

从图 6 可以看出，复合免疫算法的漏检率明显低于遗传算法和人工免疫算法，这就证明了新算法的有效性。由于复合免疫算法的匹配规则增加了关键位 Key 与分段数，并将否定选择算法与克隆选择算法相结合，因此增强了检测器的鲁棒性，同时也降低了检测器的重复性，从而降低了黑洞发生的概率，即漏检率。

(4)自我集大小与 CPU 利用率的关系如图 7 所示。

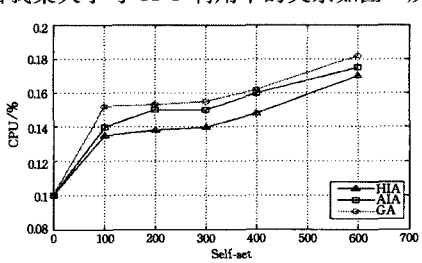

图 7　CPU 利用率对比

从图 7 可以得出，复合免疫算法的入侵检测系统在 CPU 利用率上明显低于遗传算法和人工免疫算法，这证明了改进算法可以提高系统性能。

从比较结果可以看出，复合人工免疫算法在漏检率上有所下降，降低了黑洞出现的概率，提高了系统的性能，对实时网络环境中入侵的检测具有极高的价值；在检测器生成个数上较原算法在相同参数情况下有所增加，检测器生成个数增多，意味着可以检测更多的入侵类型，从侧面也说明了漏检率降低的原因；同时，复合人工免疫算法在检测器生成时间上跟其它算法相比有所增加，但是不影响系统性能，在实时环境中，应根据系统性价比进行权衡。同时，将否定选择算法嵌入到克隆选择算法过程当中，加入了选择、复制、变异等遗传学过程，复制那些适应度好的父代细胞，生成新的子代细胞，又通过变异来减少检测器中重复的数量。

结束语　本文主要研究了复合免疫算法在入侵检测中的应用。根据经典人工免疫算法的不足，提出了基于复合免疫算法的入侵检测系统：(1)在否定选择算法中加入了分段技术和关键位，避免了恒定的匹配概率导致的匹配漏洞，降低了系统漏检率；(2)将遗传算法中的克隆选择算法和改进的否定选择算法相结合，提高了检测器生成的动态性和多样性。同时，对算法进行了数学分析与证明，并通过仿真模拟了系统的实现，并对实验结果进行了分析；最后，将经典人工免疫算法、遗传算法与改进算法进行了分析与比较，从而证明了改进算法的有效性。

生物免疫系统与计算机网络安全之间有着惊人的相似性，如何深度挖掘它们之间的这种相似性，以建立更加完善有效的入侵检测系统还有待深入研究。

参 考 文 献

[1] anonymity. NET NEWS: Policing the Computer Underworld [J]. Science, 1998, 282(11): 1223-1224

[2] El-Khatib K. Impact of feature reduction on the efficiency of wireless intrusion detection systems[J]. IEEE Transactions on Parallel and Distributed Systems, 2010, 21(8): 1143-1149

[3] Wang Y, Fu W, Agrawal D. Gaussian Versus Uniform Distribution for Intrusion Detection in Wireless Sensor Networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2013 (2): 342- 355

[4] 方贤进，李龙澍，钱海. 基于人工免疫的网络入侵检测中疫苗算子的作用研究 [J]. 计算机科学，2010，37(1)：239-242

[5] 黄建忠，裴灿浩，谢长生，等. 一种基于人工免疫技术的存储异常检测系统[J]. 计算机科学，2010，37(1)：42-46

[6] Zhang Y, Wang L, Sun W, et al. Distributed intrusion detection system in a multi-layer network architecture of smart grids[J]. IEEE Transactions on Smart Grid, 2011, 2(4): 796-808

[7] Ou C M. Host-based intrusion detection systems adapted from agent-based artificial immune systems [J]. Neurocomputing, 2012(7): 78-86

[8] Wu S X, Banzhaf W. The use of computational intelligence in intrusion detection systems: A review[J]. Applied Soft Computing, 2010, 10(1): 1-35

(下转第 77 页)

算法还存在不足，通过分析不确定数据特点，基于 x-tuple 规则元组模型，采用簇内与簇间的两阶段数据查询处理机制，提出基于 Poisson 分布的分布式不确定数据 PT-Top k 查询处理近似算法 TPQP，以达到减少查询相应时间、降低网络开销的目的。仿真实验从总体通信开销、与概率阈值 p 相关分析、与排序数 k 相关分析以及数据敏感度分析等方面，验证了 TPQP 算法在通信消耗、查询响应时间上的有效性。

另一方面，TPQP 算法也有不足之处，其仅适合 x-tuple 规则元组来自同一数据源的情况，当 x-tuple 规则元组来自不同数据源时，TPQP 算法并不适用。同时，概率阈值 p 对查询的影响分析缺乏相应的理论分析，这将在后续工作中展开。

参 考 文 献

[1] Hu C A, Fan L W, Mao Y M. HPDBSCAN: Efficient clustering algorithm for processing uncertain data[J]. Computer Engineering and Design, 2013, 34(3): 1044-1049

[2] Liu X, Yang D N, Ye M, et al. U-skyline: A new skyline query for uncertain databases [J]. IEEE Transactions on nowledge and Data Engineering, 2013, 25(4): 945-960

[3] Ye M, Lee W C, Lee D L, et al. Distributed processing of probabilistic top-k queries in wireless sensor networks [J]. IEEE Transactions onKnowledge and Data Engineering, 2013, 25(1): 76-91

[4] Wang Y, Yu J. A Top-k Query Algorithm on Uncertain Streaming Data [J]. Journal of Computational Information Systems, 2013, 9(13): 5273-5279

[5] Soliman M A, Ilyas I F, Chang K C C. Top-k query processing in uncertain databases[C]// Proceedings of the 23nd International Conference on Data Engineering, Istanbul, Turkey. IEEE, New York, 2007

[6] Nasridinov A, Park Y H. Optimal Aggregator Node Selection in Wireless Sensor Networks [J]. ICCA 2013, ASTL, 2013, 24: 37-39

[7] Sharaf A, Beaver J, Labrinidis A, et al. Balancing energy efficiency and quality of aggregate data in sensor networks[J]. VLDB Journal, 2004, 13(4): 384-403

[8] Silberstein A S, Braynard R, Ellis C, et al. A sampling-based approach to optimizing top-k queries in sensor networks [C]// Proc. International Council for Open and Distance Education. 2006: 68

[9] Zeinalipour-Yazti D, Vagena Z, Gunopulos D, et al. The Threshold Join Algorithm for Top-k Queries in Distributed Sensor Networks[C]// DMSN'05 Proceedings of the 2nd International Workshop on Data Management for Sensor Networks. 2005: 121-1

[10] Fagin R, Lotem A, Naor M. Optimal aggregation algorithms for middleware [C]// Proceedings of Special Interest Group on Management of Data. 2001: 23-33

[11] Bast H, Majumdar D, Schenkel R, et al. Io-top-k: Index-aceess optimized top-k query processing[C]// Very Large Data Base. 2006: 475-486

[12] Das G, Gunopulos D, Koudas N, et al. Answering top-k queries using views[C]// Very Large Data Base. 2006: 451-462

[13] Theobald M, Weikum G, Schenkel R. Top-k query evaluation with probabilistic guarantees[C]// Very Large Data Base. 2004: 648-659

[14] Han Q, Mehrotra S, Venkatasubramanian N. Energy efficient data collection in distributed sensor environments[C]// Proc. Institute of Electrical and Electronics Engineers. 2004: 590-597

[15] http://berkeley.intel-research.net/labdata

[16] Ye M, Lee W, Lee D, et al. Distributed Processing of Probabilistic Top-k Queries in Wireless Sensor Networks [J]. IEEE Transactons on Knowledge and Data Engineering, 2013, 25(1): 76-91

[17] Li J, Saha B, Deshpande A. A unified approach to ranking in probabilistic databases[J]. Pro. Very Large Data Base, 2009, 2(1): 502-513

[18] Ye M, Liu X, Lee W C, et al. Probabilistic Top-k Query Processing in Distributed Sensor Networks[C]// Proc. International Council for Open and Distance Education. 2010

[19] Li J, Saha B, Deshpande A. A unified approach to ranking in probabilistic databases[J]. Pro. Very Large Data Base, 2009, 2(1): 502-513

[20] Sun Yong-jiao, Yuan Ye, Wang Guo-ren. Top-k query processing over uncertain data in distributed environments [C]// Proc. Springer Science Business Media. 2011

[21] Manjeshwar A, Agrawal D P. TEEN: A protocol for enhanced efficiency in wireless sensor network[C]// The 15th Parallel and Distributed Procesing Symp. Sam Francisco: Institute of Electrical and Electronics Engineers Computer Society, USA, 2001

[22] Hua M, Pei J, Zhang W, et al. Ranking queries on uncertain data: a probabilistic threshold approach[C]// Proc. Special Interest Group on Management of Data. 2008

(上接第 47 页)

[9] Timmins J, Hone A, Stibor T, et al. Theoretical advances in artificial immune systems[J]. Theoretical Computer Science, 2008, 403: 11-32

[10] Yang J, Liu X J, Li T, et al. Distrubited agents model for intrusion detection based on AIS [J]. Knowledge-Based Systems, 2009, 22: 115-119

[11] Sobh T S, Mostafa W M. A cooperative immunological approach for detecting network anomaly [J]. Applied Soft Computing, 2011, 11: 1275-1283

[12] Powers S T, He J. A hybrid artificial immune system and self organising map for network intrusion detection [J]. Information Sciences, 2008, 178: 3024-3042

[13] Meisel M, Pappas V, Zhang L. A taxonomy of biologically inspired research in computer networking [J]. Computer Networks, 2010, 54: 901-916

[14] Visconti A, Tahayori H. Artificial immune system based on interval type-2 fuzzy set paradigm [J]. Applied Soft Computing, 2011, 11: 4055-4063

[15] Tsai C F, Hsu Y F, Lin C Y, et al. Intrusion detection by machine learning: a review [J]. Expert Systems with Applications, 2009(36): 11994-12000