doi:10.3969/j.issn.1000-436x.2013.01.015

基于阴性选择算法的异常检测系统黑洞覆盖优化

芦天亮[1]，郑康锋[1]，傅蓉蓉[2]，杨义先[1]，武斌[1]，郭世泽[1]
(1. 北京邮电大学 信息安全中心，北京 100876；2. 北京交通大学 计算机科学与技术系，北京 100044)

摘 要：针对阴性选择算法存在大量无法检测的黑洞，提出了一种基于黑洞集合和自我集合定向生成匹配阈值可变的 r 块黑洞检测器的算法。对阴性选择算法进行改进，提出了采用双重检测器的阴性选择算法 DLD-NSA，该算法在保证较快的检测速度的前提下，通过提高黑洞元素检测率，实现更大范围的非我空间覆盖。仿真结果表明，该算法与变长 r 连续位阴性选择算法相比，具有更高的非我空间覆盖率，尤其是在黑洞覆盖方面效果更好。
关键词：人工免疫系统；阴性选择算法；黑洞；r 连续位
中图分类号：TP311 **文献标识码**：B **文章编号**：1000-436X(2013)01-0128-08

Anomaly detection system with hole coverage optimization based on negative selection algorithm

LU Tian-liang[1], ZHENG Kang-feng[1], FU Rong-rong[2], YANG Yi-xian[1], WU Bin[1], GUO Shi-ze[1]
(1. Information Security Centre, Beijing University of Posts and Telecommunications, Beijing 100876, China;
2. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: With the problem that a large number of undetectable holes existed in negative selection algorithm, an algorithm of directional generating holes' detectors using *r*-chunk matching rule with variable matching threshold based on hole-set and self-set was proposed. Improvement was made to negative selection algorithm that NSA with double layers detectors was proposed. This algorithm achieved a wider range of non-self space coverage with the precondition of ensuring fast detection speed by increasing the detection of holes. Simulation result shows that comparing with NSA using variable matching length of *r*-contiguous bits this algorithm can achieve higher non-self space coverage, especially better performance in hole-space coverage.
Key words: artificial immune systems; negative selection algorithm; hole; *r*-contiguous bits

1 引言

生物科学作为计算机科学和工程问题创新解决方案灵感的来源已有很多年。该领域最新引起广泛关注的研究是人工免疫系统(AIS, artificial immune system)[1]。AIS 是受生物免疫学启发，通过模拟免疫系统的功能、原理和模型来解决复杂的实际问题的自适应系统[2]。人工免疫系统已被应用到多个不同的领域，包括：机器学习、模式识别和分类、计算机病毒检测、异常入侵检测、最优化等[3~5]。

阴性选择算法[6]作为人工免疫系统的重要分支之一，由 Forrest 等首次提出，通过模拟淋巴细胞的产生过程，生成检测器用于区分“自我(self)/非我(non-self)”，并成功应用于异常检测系统。

基于阴性选择算法的异常检测系统中存在大量的黑洞，这些黑洞是不能被所有可能的检测器识别的非我个体[7]。产生黑洞的原因主要是匹配准则和自我集合内部的模式关系[8]。黑洞同样存在于生物免疫系统中，Hofmeyr[9]借鉴了生物免疫系统的 MHC 机制，提出利用多重表示的方法，减少黑洞的数目。

张衡等[10]在 r 连续位匹配规则的基础上，引入

收稿日期：2011-12-12；**修回日期**：2012-03-16
基金项目：国家科技重大专项课题基金资助项目（2011ZX03002-005-01）；国家自然科学基金资助项目（61101108）
Foundation Items: The National S&T Major Program (2011ZX03002-005-01); The National Natural Science Foundation of China (61101108)

r 可变检测算法，通过调整匹配阈值来降低黑洞数目。但是该方法在生成检测器时采用了随机生成的方法，具有盲目性，对黑洞的覆盖效果不够理想。

采用实数值编码的阴性选择算法同样存在黑洞的问题。文献[11]对 Zhou J 提出的 V-detector 算法进行了分析和改进，通过在生成检测器过程中识别和记录自我集合边界元素，可更加有效地区分自我和非我边界，达到了减少黑洞数目的目的。

在黑洞探测方面，Stibor 等[12]利用自我集分层交叉方法给出了黑洞探测的过程，但该方法只能探测到交叉黑洞，无法探测限长黑洞。刘星宝等[13]提出能找出给定系统全部黑洞的 EHANDP 算法，并证明了算法的完备性。

本文提出了一种基于黑洞集合和自我集合定向生成匹配阈值可变的 r 块匹配检测器的方法，并对阴性选择算法进行改进，引入双重检测器，构建了 DLD-NSA 模型。最后，通过仿真证明算法具有较高的非我空间覆盖率，尤其在黑洞空间覆盖方面效果显著。

2　阴性选择算法

1994 年，Forrest 等通过模拟生物免疫系统识别自我和非我的免疫耐受过程提出了阴性选择算法。信息安全领域的众多问题（如病毒检测、入侵检测等）可转化为免疫系统中区分自我和非我的过程。生物免疫系统中，T 淋巴细胞表面的受体可检测到外来抗原。T 淋巴细胞通过遗传重组形成，在胸腺中经历阴性选择成熟后，才会释放到血液，保护身体不受外来抗原的损害。阴性选择过程中，对自身蛋白有反应的未成熟 T 细胞被销毁，只有那些不与自身蛋白结合的 T 细胞可以成熟和离开胸腺。

2.1　问题定义

模式空间，在阴性选择领域，一般采用二进制编码[6~10,12,13]的方式。$U=\{0,1\}^L$ 表示所有由元素 0 和 1 组成的长度为 L 的字符串模式空间。S 表示所有自我模式，N 表示所有非我模式，满足：$U=N\cup S$，$N\cap S=\varnothing$。

匹配规则，目前使用较多的匹配规则[14]有汉明距离(Hamming distance)匹配、r 连续位(r-contiguous bits)匹配、r 块(r-chunk)匹配等。

r 连续位匹配，长度为 L 的字符串 $a=a_1a_2\cdots a_L$ 和 $b=b_1b_2\cdots b_L$ 满足 r 连续位匹配，当且仅当 $\exists i\le L-r+1$，使得 $a_j=b_j,\ j=i,i+1,\cdots,i+r-1$。

2.2　算法概述

阴性选择算法包含如下 2 个阶段[6]。

1) 生成检测器集合，如图 1 所示。检测器是随机生成的不匹配任何自我元素的字符串，检测器集合记为 D。

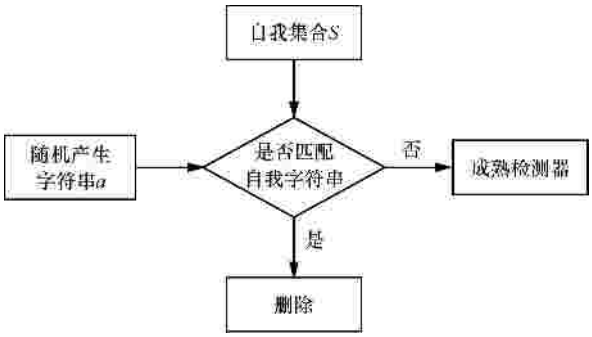

图 1　生成检测器集合

2) 用检测器集 D 测试待检数据，若匹配成功，则为非我有害数据，如图 2 所示。

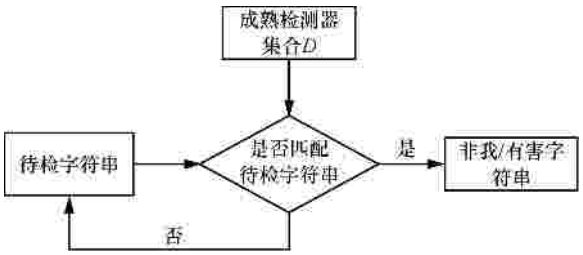

图 2　检测待检数据集

2.3　黑洞分析

基于 r 连续位匹配规则的阴性选择算法存在以下两类黑洞[15]：交叉黑洞和限长黑洞。

交叉黑洞，黑洞字符串 h 不在自我集合 S 中，但 h 的所有长度为 r 的窗口都与 S 相邻窗口交叉。例如：设 S={1010,0001}，字符串长度 $L=4$，匹配长度 $r=2$，利用有向非循环图表示如图 3 所示。

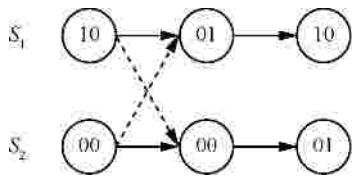

图 3　交叉黑洞示意

自我元素 S_1 的窗口为{10,01,10}，S_2 的窗口为{00,00,01}。从左侧沿着箭头方向出发，可生成字符串{1010,1001,0001,0010}，其中，{1001,0010}是

交叉黑洞，无法生成对应的检测器。

限长黑洞，黑洞字符串 h 至少有一个 r 长的窗口不在 S 中。例如：设 S={010,011}，$L=3$，$r=2$，字符串 h=110 就是一个限长黑洞。h 的检测器必须由模板 11*或*10 生成，*代表可填充 0 或 1，但这样的检测器都会因为与自我元素匹配而无法生成。

3 黑洞覆盖优化

3.1 问题描述

产生黑洞的主要原因是匹配规则和自我集合内部的模式关系。阴性选择算法采用的匹配规则需要定义匹配阈值，这体现了部分匹配(partial matching)规则的特点：2 个字符串不需完全一致，只要相似程度大于阈值即认为匹配成功。部分匹配规则可以被认为是一种近似或泛化。自我集合模式关系是影响黑洞数目的另一个主要原因，一般随机的自我集合存在的黑洞要比人工自我集合的黑洞数目要少。

在生物免疫系统中，黑洞代表那些不能被免疫系统识别的病原体，并且病原体总是朝着黑洞的方向进化，以躲避免疫系统的检测。对于计算机异常检测系统，入侵行为和病毒程序也是朝着正常行为和正常程序方向进化。利用交叉黑洞的原理，入侵行为通过对一系列正常操作片段拼接形成黑洞，增加异常检测系统的难度。对于异常检测系统来说，如何提高黑洞覆盖率是迫切需要解决的问题。

3.2 变长 r 块检测器

本文在 Forrest 提出的基于 r 连续位匹配规则的阴性选择算法基础上，引入变长的 r 块匹配规则的检测器，可有效地覆盖异常检测系统中的黑洞。

r 块匹配规则，给定字符串 $a=a_1a_2\cdots a_n$ 和检测器 $d=(i,d_1d_2\cdots d_m)$，$m\le n$，满足第 i 位 r 块匹配，当且仅当 $a_j=d_j, j=i,i+1,\cdots,i+r-1$，$i\le m-r+1$。

r 块匹配表示，当检测器 d 与抗原 a 从第 i 位开始至少存在 r 个连续相同位时，两者匹配。r 块匹配通过限定匹配的开始位置和长度，缩小了检测器非我空间的覆盖范围，解决了限长黑洞的问题。对于 2.3 节中限长黑洞的示例，匹配长度 r=2 的检测器 d={1,11}，可以成功检测到黑洞 h=110。

同时，本文提出了匹配阈值可调的变长 r 块匹配检测规则，通过增加 r 的长度，进一步缩小检测器的检测范围，可以覆盖和自我元素更加接近的黑洞。对于 2.3 节的交叉黑洞的示例，当匹配长度变为 3 时，检测器 d={1,100}和{1,001}可成功的检测到黑洞{1001,0010}。

基于变长 r 块匹配的检测器黑洞覆盖的直观表示如图 4 所示，变长 r 块检测器可以覆盖到 r 连续位检测器无法覆盖的区域。

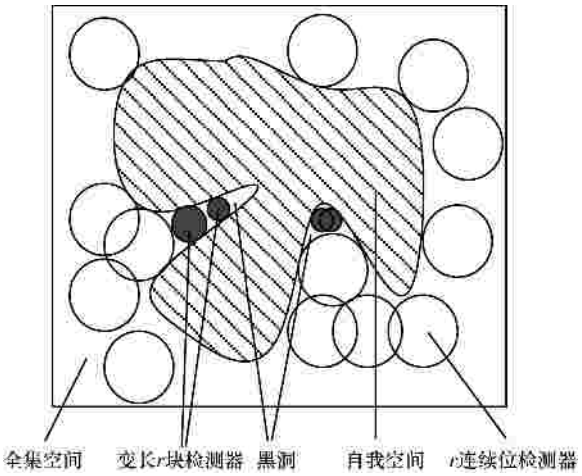

图 4 变长 r 块检测器黑洞覆盖的直观表示

3.3 算法描述

多位学者曾提出黑洞探测的方法[12,13]，本文在已知黑洞集合的基础上，结合自我集合，提出了基于变长 r 块匹配算法的黑洞检测器定向生成算法。

设自我集合为 S，$S=\{s_1,s_2,\cdots,s_N\}$，个体数目为 N，黑洞集合为 H，$H=\{h_1,h_2,\cdots,h_M\}$，黑洞数目为 M，字符串长度为 L。自我集合表示如下

$$S=\begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,i} & \cdots & s_{1,L} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,i} & \cdots & s_{2,L} \\ \vdots & & & & & \\ s_{N,1} & s_{N,2} & \cdots & s_{N,i} & \cdots & s_{N,L} \end{bmatrix} \quad (1)$$

依据连续 r 位匹配规则，将自我集合 S 分为 L-r+1 层模式集合。如下所示

$$S=[S_r[1],S_r[2],\cdots,S_r[i],\cdots,S_r[L-r+1]] \quad (2)$$

$$S_r[i]=\begin{Bmatrix} s_{1,i} & \cdots & s_{1,i+r-1} \\ s_{2,i} & \cdots & s_{2,i+r-1} \\ \vdots & s_{j,i} & \vdots \\ s_{N,i} & \cdots & s_{N,i+r-1} \end{Bmatrix}, i=1,2,\cdots,L-r+1 \quad (3)$$

同理，黑洞集合 H 也可分为 L-r+1 层模式集合。

$$H=[H_r[1],H_r[2],\cdots,H_r[i],\cdots,H_r[L-r+1]] \quad (4)$$

$$H_r[i]=\begin{Bmatrix} h_{1,i} & \cdots & h_{1,i+r-1} \\ h_{2,i} & \cdots & h_{2,i+r-1} \\ \vdots & h_{j,i} & \\ h_{M,i} & \cdots & h_{M,i+r-1} \end{Bmatrix}, i=1,2,\cdots,L-r+1 \quad (5)$$

计算集合差集 $H_r[i]-S_r[i]$，即如果存在模式 $H_r[i](j)=h_{j,i}h_{j,i+1}\cdots h_{j,i+r-1}$ 不与自我模式集合 $S_r[i]$ 中的任何模式匹配，则可由模式 $H_r[i](j)$ 生成 r 块匹配规则的检测器 $d_{r,i}=\{i,h_j\}$ 用于检测黑洞 h_j。

可变 r 块黑洞覆盖优化算法描述如下。

1) 定义长度为 L 的字符串自我集合 S。

2) 确定需要覆盖的黑洞集合 H。

3) 依据 r 长度窗口，将 S 和 H 分为 L- r+1 层，初始匹配阈值 $r=r_0$。

4) 计算 $D_r[i]=H_r[i]-S_r[i],\ i=1,2\cdots L-r+1$，对于非空集合 $D_r[i]$，生成检测器集合 $D_r=\{d_{r,i}\}$，并将 D_r 可覆盖到的黑洞元素从 H 中删除。

5) 判断 H 中是否还有未被覆盖到的黑洞，如果存在，则将匹配阈值 r 加 1，依次取值为 $r_1,r_2\cdots$ 直到最大值 L，转到 3)；否则，已生成完整黑洞覆盖的检测器集合 $D_H=D_{r_0}\cup D_{r_1}\cup\cdots$，算法结束。

上述算法，示例如下。

自我集合 S={01001,00001,10100,11101}，连续位匹配长度 3，根据 EHANDP 黑洞探测算法得到黑洞集合 H={00000,01000,10101,11100}，令 r_0=3，将 S 和 H 分为 $L-r_0+1$ 层模式集合，如下所示

$$S=\begin{bmatrix} 010 & 100 & 001 \\ 000 & 000 & 001 \\ 101 & 010 & 100 \\ 111 & 110 & 101 \end{bmatrix},\quad H=\begin{bmatrix} 000 & 000 & 000 \\ 010 & 100 & 000 \\ 101 & 010 & 101 \\ 111 & 110 & 100 \end{bmatrix}$$

$$H_3[1]-S_3[1]=\varnothing$$

$$H_3[2]-S_3[2]=\varnothing$$

$H_3[3]-S_3[3]=\{000\}$，该集合可生成 r 块检测器 d={3,00000}，可检测到黑洞 00000 和 01000。

在 r_0=3 时，仍然存在黑洞 H={10101,11100}。为了进一步检测剩余黑洞，令 $r_1=r_0+1=4$，将 S 和 H 分为 $L-r_1+1$ 层。

$$S=\begin{bmatrix} 0100 & 1001 \\ 0000 & 0001 \\ 1010 & 0100 \\ 1110 & 1101 \end{bmatrix},\quad H=\begin{bmatrix} 1010 & 0101 \\ 1110 & 1100 \end{bmatrix}$$

$$H_4[1]-S_4[1]=\varnothing$$

$H_4[2]-S_4[2]=\{0101,1100\}$，该模式可生成匹配长度 $r_1=4$ 的检测器 d={2,10101} 和检测器 d={2,11100}，可分别检测黑洞 10101 和 11100。

3.4 算法分析

设自我集合为 S，个体数目为 N_S，黑洞集合为 H，黑洞数目为 N_H，字符串长度为 L，初始匹配长度为 r_0，将 S 和 H 分为 $L-r_0+1$ 层模式集合。对于 $i=1,\cdots,L-r_0+1$，计算 $H_r[i]-S_r[i]$ 的时间复杂度为 $N_SN_H(L-r_0+1)r_0$。

最好的情况下，匹配长度取值为 r_0，$i=1$ 时，即第一轮第一层匹配后，即可完全覆盖 H 中的黑洞，时间复杂度为 $N_SN_Hr_0$。

最差的情况下，匹配长度取值为 $r_0, r_0+1,\cdots,L$-1 时，H 中的模式均包含在 S 的模式集合中，无法生成有效检测器。当匹配长度取值为 L 时，S 和 H 只有 1 层，每个模式长度为 L，此时可生成 N_H 个检测器，时间复杂度为

$$N_SN_H[(L-r_0+1)r_0+(L-r_0)(r_0+1)+\cdots+1\times L]$$
$$=N_SN_H(L-r_0+1)(L-r_0+2)(L+2r_0)/6 \quad (6)$$

当 r_0=1 时上式取最大值，时间复杂度最高为 $N_HN_SL(L+1)(L+2)/6$。

下面计算空间复杂度，内存空间主要消耗于存储 S 和 H 分层后的模式集合。当初始匹配长度为 r_0，将 S 和 H 分为 $L-r_0+1$ 层模式集合占用的空间大小为 $(N_S+N_H)(L-r_0+1)r_0$。最差的情况下，匹配长度需依次取值为 $r_0, r_0+1,\cdots,L$- 1,L，空间复杂度为

$$(N_S+N_H)[(L-r_0+1)r_0+(L-r_0)(r_0+1)+\cdots+1\times L]$$
$$=(N_S+N_H)(L-r_0+1)(L-r_0+2)(L+2r_0)/6 \quad (7)$$

当 r_0=1 时上式取最大值，空间复杂度最高为 $(N_S+N_H)L(L+1)(L+2)/6$。

4 阴性选择算法改进

本文提出了基于双层检测器的阴性选择算法 (DLD-NSA)，模型如图 5 所示。

模型中包含 2 个匹配过程，分别用到成熟检测器集合 D 与黑洞检测器集合 D_H。

1) 匹配过程 1

D 由图 1 中的经历阴性选择过程的成熟检测器组成，主要用于识别非我字符串，该过程的目标是以尽量少的检测器覆盖更多的非我空间。

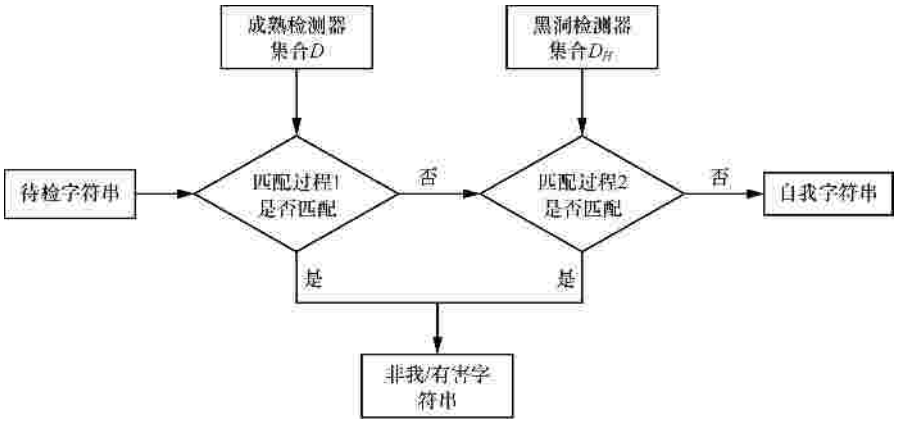

图 5 DLD-NSA 模型

2) 匹配过程 2

D_H 由 3.3 节介绍的算法生成的变长 r 块检测器组成，主要用于覆盖黑洞，提高检测的准确性。

该模型提出了一种黑洞覆盖优化的通用框架。在匹配过程 1 中，可以选择任意的匹配规则，并生成适合数目的检测器集合 D，以达到期望的非我空间覆盖率。匹配过程 2 中，利用黑洞检测器 D_H 检测黑洞元素。黑洞集合在自我集合和过程 1 的匹配规则确定后，利用黑洞探测算法生成，或者由用户输入危险黑洞字符串。黑洞检测器的数目也可根据用户设定的黑洞覆盖率动态调整。

5 仿真结果

5.1 数据集采集

由于数据集内部模式结构对黑洞的规模有直接影响，所以实验中选取两类数据集：人工数据集和随机数据集。人工数据集采用 UCI 标准数据集 Pima Indians Diabetes 数据集，选取其中 500 个糖尿病检测结果为阴性的样本生成自我集。将每个样本的 8 个属性值二进制编码后生成长度为 63bit 的符串。为了缩短个体元素字符串长度，将 63bit 均分为 3 个字符串，去冗余后，最终选取其中的 1 200 个字符串，生成 4 组自我集合，每组 300 个元素，每个元素 21bit。随机数据集利用 MATLAB 生成 4 组自我集合，每组集合为随机生成的 300 行 21bit 的二进制字符串。

5.2 黑洞规模比较

在连续 r 位匹配规则下，定义 2 个长度为 L 的字符串 s_1、s_2 的亲和度如下

$$affinity(s_1,s_2)=\frac{r(s_1,s_2)}{L} \tag{8}$$

其中，$r(s_1,s_2)$ 为 s_1、s_2 连续匹配的最大长度，当 s_1、s_2 完全相同时，亲和度为 1。

定义字符串集 S 中元素的平均亲和度如下

$$affinity(S)=\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}affinity(s_i,s_j) \tag{9}$$

经计算，人工数据集平均亲和度为 0.258，随机数据集平均亲和度为 0.181。

在 r 连续位匹配规则下，设匹配阈值为 r_c，利用 EHANDP 算法探测得到两类数据集中的黑洞，黑洞的数目如表 1 所示。

表 1 数据集黑洞数目比较

r_c	人工数据集				随机数据集			
	组 1	组 2	组 3	组 4	组 1	组 2	组 3	组 4
10	3 164	2 611	3 404	3 582	2 210	2 271	1 961	2 015
11	1 137	1 867	1 384	2 359	555	651	586	565
12	552	507	732	1 929	243	233	209	165
13	346	287	454	1 457	101	98	94	71
14	246	175	246	852	36	32	47	33
15	149	107	105	413	8	18	16	16

从表 1 可以看出，人工数据集的黑洞数目普遍高于随机数据集的黑洞数目。随 r_c 的增加，黑洞数目逐渐减少，但是当 r_c 取值较大时，需要更多的检测器才能覆盖足够的非我空间，会增加匹配时间。所以在确定 r_c 时，需权衡黑洞数目和检测器规模。

5.3 黑洞检测器分布仿真

对两类数据集中的各组数据，选取连续 r 位匹配的阈值 $r_c=10,11,12$，生成对应的黑洞集合，根据 3.3 节提出的算法，生成变长 r 块匹配规则的黑洞检测器。黑洞检测器匹配长度分布如图 6 所示。

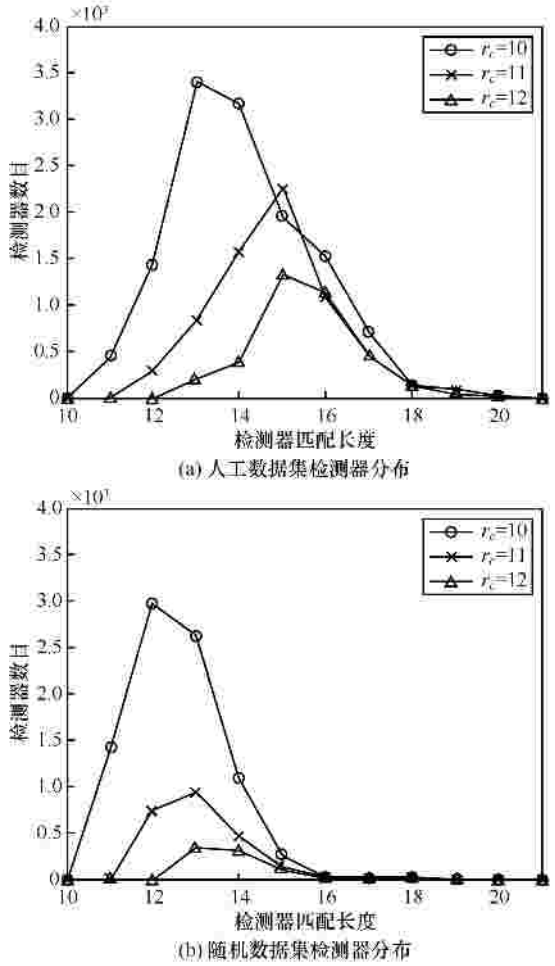

图 6　检测器匹配长度分布

从图 6 中可以看出，当 r 连续位匹配阈值 r_c 相等时，人工数据集的黑洞检测器总数多于随机数据集黑洞检测器总数。检测器数目随匹配长度的增加分布呈现先上升后下降的趋势。对于相同数据集，随着 r_c 的增大，黑洞检测器的总数逐渐减少，并且检测器的平均匹配长度增加（折线中心向右偏移）。

5.4　非我空间覆盖率仿真

对比 Forrest 提出的 NSA 算法，以及文献[10]提出的 r 可变阴性选择算法(RA-NSA, r-adjustable negative selection algorithm) 与本文提出的 DLD-NSA 算法在非我空间覆盖率的性能。

根据非我空间中的元素能否在定长 r 连续位匹配规则下被检测到，将非我空间分成两部分：可检空间和黑洞空间。非我空间覆盖率包含 2 个方面：可检空间覆盖率和黑洞空间覆盖率。

实验过程中，自我数据集分别选取人工数据集和随机数据集中的第一组数据。

构造测试数据集，数据集中共包含 2 000 个元素。其中，1 000 个随机生成的可检空间元素，1 000 个黑洞集合中随机选取的黑洞空间元素。

对于 NSA 算法，设定检测器的匹配长度为 10，生成检测器 N 个。

对于 RA-NSA 算法，设定检测器初始匹配长度为 10，匹配长度上限为 21，生成检测器 N 个。

对于 DLD-NSA 算法，生成的检测器的总数目也为 N 个，定义黑洞检测器所占比例为 p。匹配过程 1 采用定长的 r 连续位匹配规则，匹配阈值为 10，检测器个数为 $N_1 = N(1-p)$；匹配过程 2 采用变长 r 块黑洞检测器，初始匹配长度为 10，检测器个数为 $N_2 = Np$。当 N_2 大于实际的全部黑洞检测器数目 N_H 时，取 $N_2 = N_H$，$N_1 = N - N_H$。其中，人工数据集黑洞检测器的总数 $N_H = 3\,164$，随机数据集黑洞检测器的总数 $N_H = 2\,210$。

首先，分析 p 的取值对可检空间和黑洞空间覆盖率的影响，仿真结果如图 7 和图 8 所示。

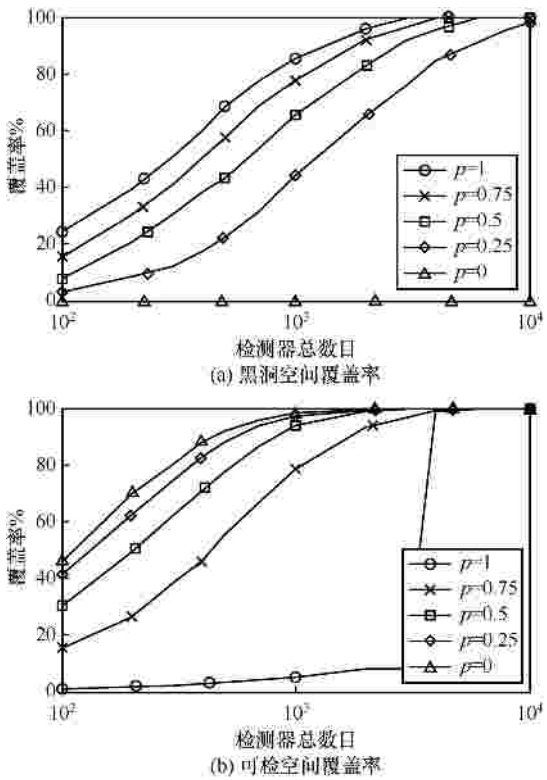

图 7　人工数据集非我空间覆盖率

两类数据集仿真结果基本一致，随着检测器数目的增加，黑洞空间和可检空间的覆盖率都逐渐升高。当检测器的总数固定时，p 的取值越大，黑洞空间覆盖率越大，可检空间覆盖率越小。

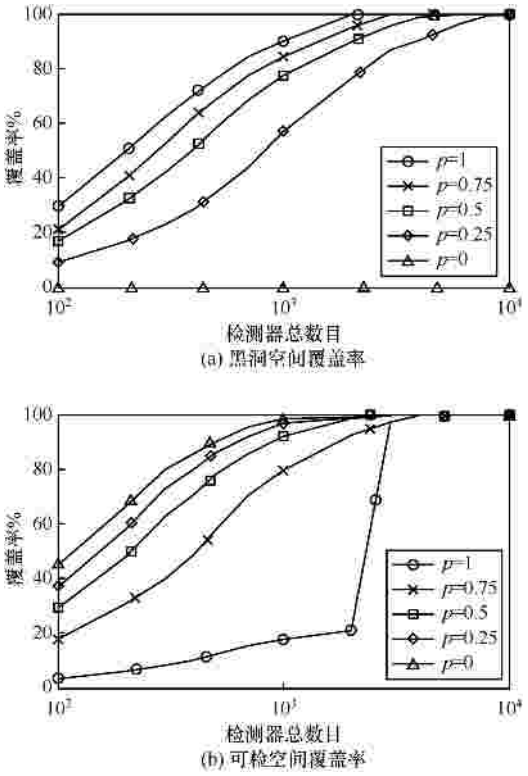

图 8 随机数据集非我空间覆盖率

当 p=0 时，DLD-NSA 简化为图 2 所示的原始 NSA 算法，此时系统具有较高的可检空间覆盖率，但黑洞空间覆盖率为 0。

当 p=1 时，只有当黑洞检测器数目达到 N_H 后才会生成 DLD-NSA 第 1 层检测器。从图 7(b)和图 8(b)中可以看出，当检测器总数 $N<N_H$ 时，可检空间覆盖率并不为 0，而是缓慢上升，这是因为第 2 层的黑洞检测器可以覆盖少量的可检空间。当检测器数目 $N>N_H$ 时，可检空间覆盖率上升较快。

下面，通过仿真对比 NSA 算法、RA-NSA 算法与 DLD-NSA 算法非我空间覆盖的性能。在检测器的总数目 N 相同时，三类算法的检测速度相当。对于 DLD-NSA 算法，为保证可检空间和黑洞空间都具有较高的性能，取 p=0.2，仿真结果如图 9 和图 10 所示。

从图 9 和图 10 可以看出，对于两类数据集，DLD-NSA 算法在可检空间覆盖率和黑洞空间覆盖率方面都明显优于 RA-NSA 算法，NSA 算法的可检空间覆盖率略高于 DLD-NSA。

三类算法在检测器数目达到 5 000 左右时，可检空间覆盖基本都接近了 100%。检测器数目在 100~5 000 之间时，DLD-NSA 算法明显优于 RA-NSA。

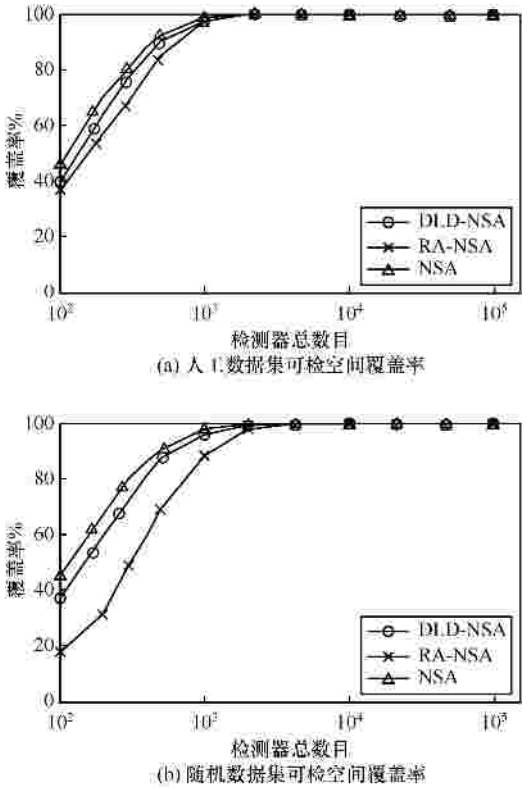

图 9 可检空间覆盖率对比

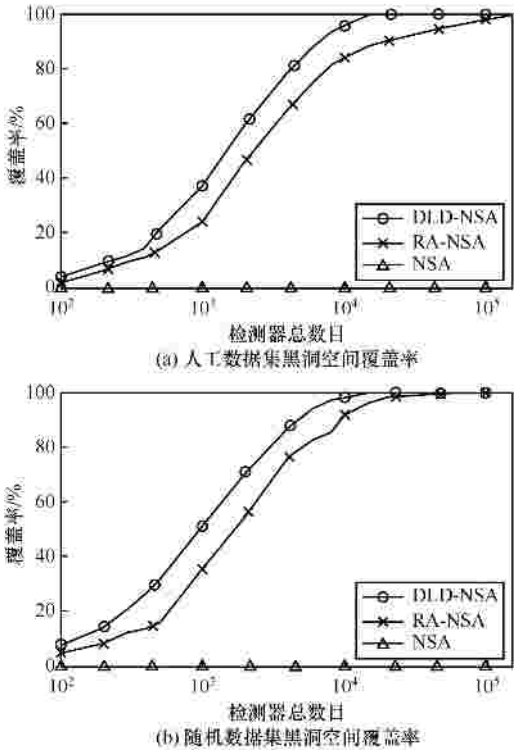

图 10 黑洞空间覆盖率对比

黑洞覆盖方面，NSA 算法的黑洞空间覆盖率为 0，DLD-NSA 算法由于采用了定向生成黑洞检测器的方法，相比随机盲目生成检测器的 RA-NSA 算法

优势更加明显，DLD-NSA 算法在检测器数目 1×10^4 左右时，已基本达到 100%的黑洞覆盖，而 RA-NSA 算法检测器数目达 10×10^4 时，黑洞覆盖率才接近 100%。

6　结束语

阴性选择算法作为人工免疫系统重要分支在异常检测等方面被广泛应用。本文对阴性选择算法的黑洞问题进行了深入研究，提出了基于变长 r 块匹配规则的定向黑洞检测器生成算法。对 Forrest 提出的阴性选择算法模型进行改进，提出了双层检测器阴性选择算法 DLD-NSA，该算法在保证较高检测速度的前提下，通过增加黑洞检测器匹配层提高了算法的黑洞覆盖率。仿真结果表明，DLD-NSA 算法在可检空间覆盖率和黑洞空间覆盖率方面均有显著的改善。

参考文献：

[1] GAO X Z, CHOW M Y , PELTA D. Theory and applications of artificial immune systems[J]. Neural Computing and Applications, 2010,19(8):1101-1102.

[2] CASTRO L N D, TIMMIS J. Arti?cial Immune Systems: a New Computational Intelligence Approach[M]. Berlin: Springer, 2002.

[3] 安辉耀, 吴泽俊, 王新安等. 用于网络入侵检测的群体协同人工淋巴细胞模型[J]. 通信学报, 2010,31(9): 122-130.
AN Y H, WU Z J, WANG X A, *et al*. Population-based cooperative artificial lymphocyte model for network intrusion detection[J]. Journal on Communications, 2010,31(9):122-130.

[4] ZHENG J Q, CHEN Y F, ZHANG W. A survey of artificial immune applications[J]. Artificial Intelligence Review, 2010,34(1):19-34.

[5] DASGUPTA D, YU S H, NINO F. Recent advances in artificial immune systems-models and applications[J]. Applied Soft Computing, 2011,11(2):1574-1587.

[6] FORREST S, PERELSON A S, ALLEN L, *et al*. Self-nonself discrimination in a computer[A]. Proceedings of IEEE Symposium on Research in Security and Privacy[C]. Los Alamitos, CA, USA, 1994. 202-212.

[7] DHAESELEER P. An immunological approach to change detection: theoretical results[A]. Proceedings of the 9th IEEE Computer Security Foundations Workshop[C]. Kenmare, Ireland, 1996. 132-143.

[8] ZHOU J, DASGUPTA D. Revisiting negative selection algorithms[J]. Evolutionary Computation, 2007, 5(2): 223-251.

[9] HOFMEYR S A. An Immunological Model of Distributed Detection and Its Application to Computer Security[D]. Albuquerque: Department of Computer Sciences, University of New Mexico,1999.

[10] 张衡, 吴礼发, 张毓森等. 一种 r 可变负选择算法及其仿真分析[J]. 计算机学报, 2005, 28(10): 1614-1619.
ZHANG H, WU L F, ZHANG R S, *et al*. An algorithm of *r*-adjustable negative selection algorithm and its simulation analysis[J]. Chinese Journal of Computers, 2005,28(10): 1614-1619.

[11] LI G Y, LI T, ZENG J, *et al*. An improved V-detector algorithm of identifying boundary self[A]. Proceedings of the 2009 International Conference on Machine Learning and Cybernetics[C]. Baoding, China, 2009. 3209-3214.

[12] STIBOR T, MOHR P, TIMMIS J. Is negative selection appropriate for anomaly detection[A]. Proceedings of Genetic and Evolutionary Computation Conference[C]. New York, NY, USA, 2005. 321-328.

[13] 刘星宝, 蔡自兴. 异常检测系统的漏洞分析[J]. 中南大学学报, 2009, 40(4): 986-992.
LIU X B, CAI Z X. Properties assessments of holes in anomaly detection systems[J]. Journal of Central South University(Science and Technology), 2009, 40(4): 986-992.

[14] GONZALEZ F, DASGUPTA D, GOMEZ J. The effect of binary matching rules in negative selection[A]. Proceedings of Genetic and Evolutionary Computation Conference[C]. Lecture Notes in Computer Science, 2003. 195-206.

[15] ESPONDA F, FORREST S, HELMAN P. A formal framework for positive and negative detection schemes[J]. IEEE Transactions on Systems Man and Cybernetics, 2004, 34(1):357-373.

作者简介：

芦天亮（1985-），男，河北保定人，北京邮电大学博士生，主要研究方向为信息安全、人工智能与恶意代码检测。

郑康锋（1975-），男，山东烟台人，博士，北京邮电大学副教授，主要研究方向为网络与信息安全。

傅蓉蓉（1987-），女，江苏盐城人，北京交通大学博士生，主要研究方向为自组织网络安全与人工智能算法。

杨义先（1961-），男，四川盐亭人，北京邮电大学教授、博士生导师，主要研究方向为信息安全与密码学。

武斌（1981-），男，山东泰安人，博士，北京邮电大学讲师，主要研究方向为网络安全。

郭世泽（1969-），男，河北石家庄人，北京邮电大学教授、博士生导师，主要研究方向为网络安全。