一种改进的动态克隆选择免疫算法在入侵检测中的应用研究

许　春，　李　涛，　刘孙俊，　杨　频，　刘　念，　张建华

(四川大学计算机学院，四川成都 610064)

摘要：给出自体、非自体、抗原、抗体、免疫细胞的定义，改进亲和力计算公式，提出可控变异和随机变异方法并以此改进动态克隆选择算法。设计并实现基于该改进免疫算法的入侵检测系统(IDS)模型，仿真实验表明，改进后的算法有效提高入侵检测系统的自适应性。

关键词：动态克隆选择算法；入侵检测；人工免疫；亲和力

中图分类号：TP309　　文献标识码：A　文章编号：1009-3516(2006)03-0050-05

基于免疫计算的计算机网络入侵检测克服了传统网络入侵检测系统的缺陷，被认为使一条非常重要且有巨大实际应用前景的研究方向[1]。在免疫计算的研究中，Kim 和 Bentley 进而提出了动态克隆选择算法(DynamiCS)[2]。基于 DynamiCS 的免疫系统在计算机网络入侵检测(NIDS)中能够识别新的入侵。针对当前网络入侵具有某特定时间的特定趋势的特点，把 DynamiCS 做适当的改进，引入可控变异和随机变异方法，能够收到更好的免疫效果。

1　基本概念定义

在基于免疫计算的入侵检测系统中，所有的网络行为被定义为问题空间 P，正常的网络服务被定义为自体集合 S，网络攻击被定义为非自体 N，则 $S\cap N=\Phi$，且 $S\cup N=P$。能代表一个确定意义的字符串定义为基因 Gene，比如协议类型、端口号、目的 IP 地址、源 IP 地址等，有：Gene $=\{0,1\}^k$，k 为 Gene 的长度。抗原 A_g 定义为：$A_g=\{(0,1)^{k^e}\mid\ k^e$ 为第 e 个基因的长度，$e=1,2,\cdots,n$，n 为抗原基因的个数. 抗体是与抗原等长的字符串，二者具有相同的数据结构。另外，抗体具有属性 A_b. age，表示抗体年龄在计算机免疫学中，抗体也称为检测器(Detector)检测器有产生、成长、进化等生理过程，相应阶段分别称为未成熟检测器(Non - Maturation Detector)、成熟检测器(Maturation Detector)和记忆检测器(Memory Detector)成熟检测器识别抗原(即检测到网络入侵)，在计算机免疫学上叫做初次免疫应答(Primary Immune Response)，记忆检测器识别抗原，叫做二次免疫应答(Secondary Immune Response). 检测器识别抗原可以通过计算二者的亲和力 $f(x,y)$ 来实现，亲和力的计算公式为：$f(x,y)=r/N$，其中，x 为抗体、y 为抗原、r 为二者字符串之间相同位置上相同字符的个数、N 为二者字符串的字符个数。识别函数 dect(ag)值可由下式计算。

$$\mathrm{dect(ag)}=\begin{cases}1 & \text{if}\quad f(x,y)\geqslant A\_\mathrm{match}\\ 0 & \text{otherwise}\end{cases}$$

其中，A_match 为激活阈值，即当 $f(x,y)\geqslant A$_match 时，系统认为抗体识别了抗原。A_match 是一个经验值。

收稿日期：2005 - 12 - 20
基金项目：国家自然科学基金(60373110)教育部博士点基金资助项目(20030610003)
作者简介：许　春(1971 -)，男，河北秦皇岛人，讲师，博士生，主要从事计算机网络安全、人工免疫研究；
李　涛(1965 -)，男，四川岳池人，教授，博士导师，主要从事计算机网络安全、人工神经网络、人工免疫研究

2　改进动态克隆选择免疫算法

2.1　动态克隆选择算法

动态克隆选择算法有如下几个步骤:

1)初始化:随机生成一个属性串(免疫细胞)的群体。

2)群体循环:对每一个抗原,①选择那些与抗原具有更高亲和力的细胞进行检测;②变异产生的新免疫细胞,遗传上代免疫细胞的特性,同时经自体耐受后,成长为成熟免疫细胞,参与免疫计算,如不能经过自体耐受,则将该免疫细胞去掉,并把其特性作为自体加入自体集合;③计算每个变异后的细胞与抗原的亲和力。

3)循环:重复步骤2),直到一个给定的收敛标准被满足。

2.2　对动态克隆选择算法的改进

根据网络入侵的一般特点,当前流行的入侵可能会持续一段时间,而按照动态克隆选择算法,变异产生的免疫细胞与当前正在流行的网络入侵的亲和力很小。为此,在遗传变异中引入可控变异方法,引入随机变异方法,提高识别不同于当前流行入侵的效率。另外,把亲和力计算公式做适当改进,更加适合网络入侵检测的特点。

2.2.1　改进 r 连续位算法计算亲和力

抗体与抗原的匹配,是通过计算亲和力来实现的。可以用改进的 r 连续位算法(r - Contiguous Bits)来计算亲和力 $f(x,y)$。

在传统的 r 连续位规则中[4],如果有如下两个符号串 α、β,α = 10100011011110110011001111、β = 11000111011110111000110000,二者最多有10个连续位符号相同,则 $r=10$。$f(\alpha,\beta)=r/N=10/25=0.4$。

但是单纯比较两个符号串的连续位符号相同的个数,在基于免疫计算的入侵检测中作为抗体/抗原匹配,是不合适的。传统 r 连续位匹配规则计算出来亲和力很大,但不一定代表是抗体的基因(比如 Port、Protocle Type 等)正好与抗原的相应基因匹配。比如有如下两个符号串 γ 与 θ:

γ = 0111010001101 11101100 0111000 001101000100, θ = 1011010001101 11101100 0111000 001101000111

gene1　gene2　gene3　gene4,　gene1　gene2　gene3　gene4

二者相同位置相同字符数 $r=36$,则 $f(\gamma,\theta)=r/N=36/40=0.9$。如果 $A\_match=0.8$,则系统认为抗体 γ 识别了抗原 θ,即 θ 为入侵。但从二者的基因序列来看,只有 gene2 和 gene3 完全相同,代表完全相同的物理意义,而 gene1 和 gene4 尽管有相同字符,但不是完全相同,代表完全不同的物理意义,比如 γ 的 gene1 代表协议 UDP 协议类型,而 θ 的 gene1 代表 ICMP 协议类型。

因此,亲和力计算公式要做如下改进:$f(x,y)=r'/N'$。其中,x、y 仍分别为抗体和抗原,N'为抗体或抗原的基因个数,r'为二者相同的基因个数。则在上述 $f(\gamma,\theta)=0.5$,$f(\gamma,\theta)<A\_match$。θ 是入侵。

2.2.2　可控变异(Controllable Aberrance)和随机变异(Stochastic Aberrance)

动态克隆选择算法,用于检测器遗传变异,取得较好的实验效果。引入可控变异方法改进动态克隆选择算法,能更好地把亲和力高的变体选中进入了记忆检测器库,并使它们支配免疫应答。引入随机变异方法能够消除亲和力频谱中局部极值[1],抗体能在亲和力变化较宽的范围内搜索,使之变异能越过局部极值点,朝亲和力最大点变异。

可控变异发生在记忆检测器库中,具体方法:新加入的记忆检测器具有代表当前网络攻击趋势的特性,选择 Memory. age = 0 的记忆检测器进行变异,原 r 个相同基因保持不变,其余($N-r$)个基因按照随机函数 Random(ab)在相应基因库中随机变化,子代抗体 Memory. age = 1。这样就确保了子代抗体至少具有与父代抗体相同的抗原亲和力,而更高亲和力的子代抗体进入记忆检测器库中,原父代抗体进化为死亡。

定理1:在可控变异中,子代抗体与相应抗原的亲和力满足关系:$f(x_{子},y)\geq f(x_{父},y)$。

证明:$f(x_{子},y)=r_{子}/N$,$f(x_{父},y)=r_{父}/N$,在可控变异中,$r_{子}\geq r_{父}$。

随机变异发生在成熟检测器库中,具体方法:$\forall$ ma - dete $\in \{x \mid x.\ age = T_{ma}-1 \cap x \in \{成熟检测器\}\}$ 的 N 个基因都在相应基因库(Gene set)中按照 Random(ab)随机选取,其中,T_{ma}表示成熟检测器生命周期。这样变异后的子代抗体与父代抗体有较大差异,确保免疫系统的多样性,扩大了抗体在亲和力变化较宽的范围内搜索,越过局部价值点,可能寻找到更高的亲和力点。

2.2.3 检测器自体耐受与生命周期

一般,检测器有两种产生方式:系统初始化时候,受病毒样本刺激产生和父代检测器变异产生。新产生检测器首先经历自体耐受:

$$\forall x, x = \begin{cases} \text{Dead} & \text{if} \quad if(x,y) \geqslant A\_\text{tolerance} \\ \text{Successed} & \text{otherwise} \end{cases}$$

其中,x 为新产生的检测器,y 为任意自体抗原,A_tolerance 为耐受阈值。

新检测器自体耐受成功,即成长为成熟检测器. 一般来说,检测器的生命周期如图 1 所示。其中,used 表示检测器识别抗原的次数,M_{max} 表示成熟检测器进化为记忆检测器识别抗原的次数阈值,T_{ma} 表示成熟检测器生命周期,T_{evol} 表示记忆检测器不被使用而进化为成熟检测器的时间阈值,如果记忆检测器不被使用的时间超过该阈值,表明相应的攻击抗原长时间不活跃,记忆检测器将进化为成熟检测器,进入成熟检测器的生命过程,T_{memo} 表示记忆检测器进化为死亡的时间阈值。

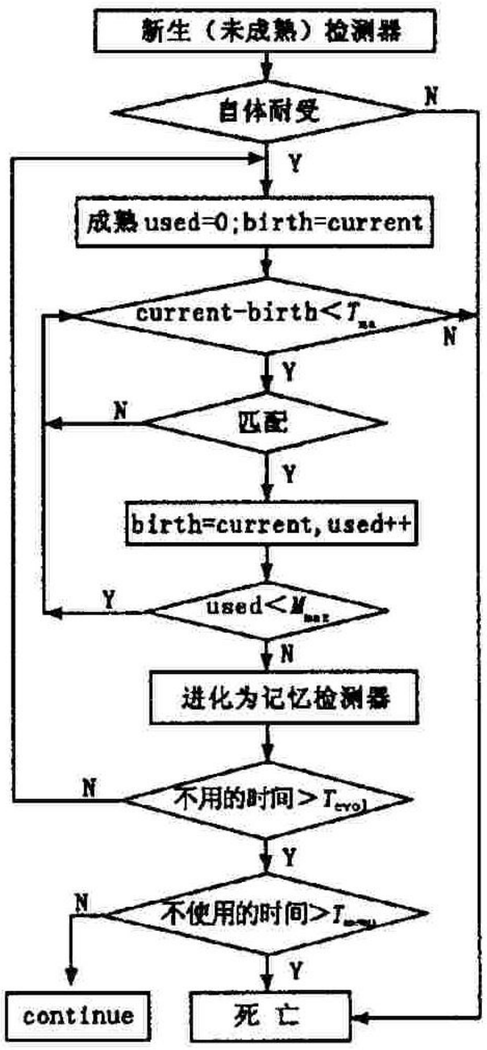

图 1 检测器的生命周期图

2.2.4 改进后的动态克隆选择算法

可控变异与随机变异被加入动态克隆选择算法,可以用如下的伪代码表示。

```
procedure Dynamic immune - algorithm
Begin
    Memory_detect(ag);//记忆检测器检测
    If (Memory_detect(ag) =1)
    Begin
       * * * * * *();//二次应答,阻断攻击
   memory. age =0;//记忆检测器年龄置 0
end;
else
begin
    if(memory. age =0)
    controllable_aberrance(ab);//可控变异
    stochastic_aberrance(ab);//随机变异
end;
custom_detect(ag);//成熟检测器检测
if(custom_detect(ag) =1)
begin
    rejust _
sever();//初次应答,拒绝服务
    if(custom. count > =阈值 q)
    active(custom);//成熟检测器被激活并被添
                   加到记忆检测器库中
    else
    custom. count + +;//成熟检测器次数加 1
  End
  Else
  Begin
    Insert_self(ag);//非入侵,将 ag 加入自体库
    ab. age + +;//抗体年龄加 1
    check(ab_age);//检查抗体年龄
    if(check(ab_age) > =阈值 T)
      dead(ab);//达到生命年龄的抗体死亡
      continue;
  end;
End
```

3 仿真实验及结果分析

3.1 仿真实验设计及主要参数设定

基于改进的免疫算法,可以设计如图 2 所示的入侵检测模型。

在特定构造的蕴涵 Syn Flood、Smurf、Pscan、Land、RPC Locator、Win Nuke、Teardrop、IP Watcher 等 15 种常见入侵攻击免疫主机进行系统初始化,使免疫系统具有一定的记忆检测器集合(针对该 20 种入侵)和一定的成熟检测器集合。

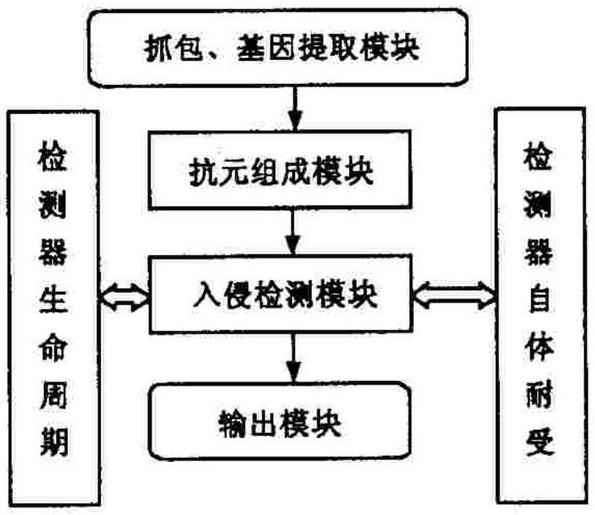

图 2 AIS - IDS 模型图

仿真实验把该系统放置在网络攻防实验室局域网环境中,使用网络工具 Iris 抓包,取 32 位源/目的 IP 地址、16 位源/目的端口、16 位协议类型以及 16 位包长度等 6 类基因构成的二进制串提呈抗原,其中每类基

因值集合构成相应基因库。检测针对各种网络服务(用端口号进行区分)的攻击。成熟检测器的激活阈值 $A\_match = 0.83$,自体耐受阈值 $A\_tolerance = 0.83$,成熟检测器生命周期 $T_{ma} = 550$,取迭代次数 550 次,最大变异代数 $N_{variation} = 10\ 000$,记忆检测器生命周期 $T_{mem} = 1\ 000\ 000$,抗体随机变异 ab. age 的阈值 $T_{maturation-detector} = 150$;抗体可控变异 ab. age 的阈值 $T_{control-detector} = 0$。

3.2　仿真结果及分析

为对比改进系统与传统系统的工作效率,在网络攻防实验室局域网环境中,分别设置两台主机,主机 A 安装改进的免疫系统,记为 new_DCS_ids,主机 B 安装按照 Kim 和 Bentley 算法设计的传统免疫系统,记为 old_DCS_ids. 每次向局域网中检测系统送 100 个抗原,非自体的串和自体串的比例为 9:1,也就是说攻击程序所发的 10 个包中夹杂一个自体的包。

3.2.1　免疫学习效率

免疫学习效率是人工免疫系统的最重要的指标之一。可控变异的引入,使得记忆检测器中与当前入侵趋势比较亲和力大的子代被保留下来,大大增加了免疫系统二次应答的效率。如图 3 所示,改进后的系统,只需很少变异,就迅速识别抗原,效率比改进前有明显提高。从图 3 可以看到,有部分亲和力值达到 1,可以肯定的是发动攻击的源 IP 地址正好与可控变异产生的抗体基因中源 IP 地址相同,这种情况在实验取值点中是少数,同时由于局域网规模比较小,源 IP 地址的基因库比较小,这样恰巧的现象有重复出现。从实验来看,没有重复出现规律。

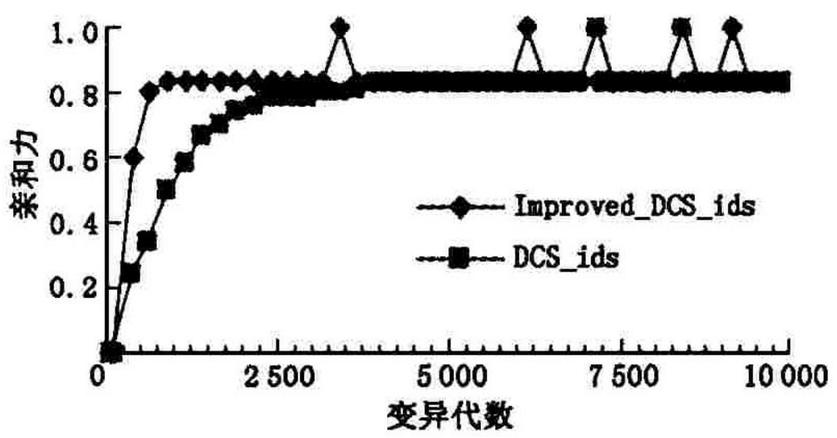

图 3　系统对 DDOS 二次应答效率图

免疫初次应答体现了系统对新攻击的自适应识别,随机变异使系统成熟检测器具有较大多样性,这种多样性对初次应答具有直接贡献。系统识别新的攻击的效率,如图 4 所示。在图 4 中,改进前后系统都随变异代数增加,亲和力增大,显然是克隆选择原理起作用。但改进算法的系统(如图 4 中菱形曲线)有两个明显的阶跃点:a、b。a、b 点的产生,主要是由于随机变异,产生了与抗原亲和力更高的子代抗体,该子代抗体后经克隆选择遗传变异,能够明显提高成熟检测器的亲和力,特别是 b 点,明显提高曲线的纵坐标。c 点表示系统识别了攻击。

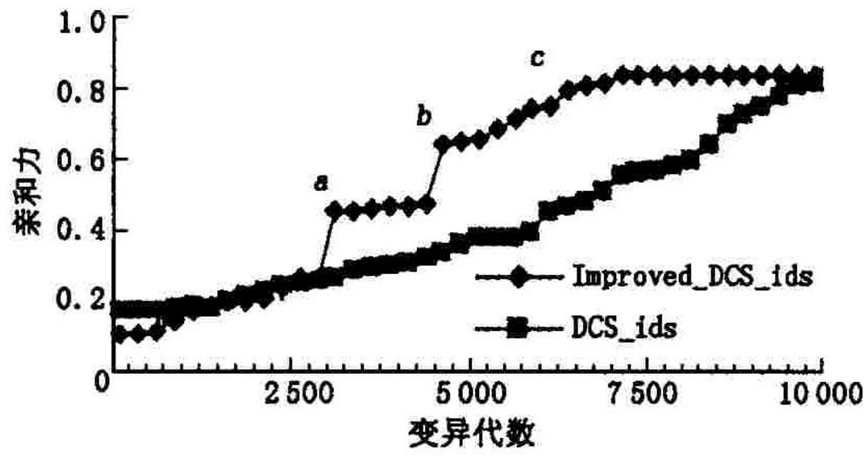

图 4　系统初次应答效率图

3.2.2　降低 FPR

免疫计算可能出现的错误有两种,一种是错误肯定(FP),把网络攻击当成正常的网络行为加入自体库;另一种是错误否定,把正常网络行为当成入侵。往往通过协同刺激来减少该两类错误。在入侵检测系统中,FP 是不能容忍的,要尽量减少。设 Ncon_active 表示程序员的协同刺激次数,Mfp 表示经程序员判断为系统 FP 的次数,FPR(rate of FP)是衡量系统发生 FP 的指标,则 FPR = Mfp/Ncon - active。

可控变异的引入,使更多代表当前入侵趋势的攻击(流行入侵)在记忆检测器中被识别,即发生二次应答,成功降低 FPR。如图 5 所示,系统 FPR 整体上呈下降趋势,主要是克隆选择算法在起作用。算法改进后,在菱形曲线 a 点,有较明显下降,应该是随机变异产生了亲和力较大的抗体,识别攻击能力增强。注意到,在 b 点,几乎是两支曲线都大幅度上扬,主要是当前攻击趋势骤变,新的攻击种类出现,系统 FP 迅速增加。不过,慢慢都呈下降趋势,只是由于最大变异代数的限制,该跟踪数据没有降到更低,可以预见,如果最大变异代数更大时,FPR 将还要继续下降。FPR 下降趋势,从另一个侧面反映了免疫系统具有较好的自适应性。

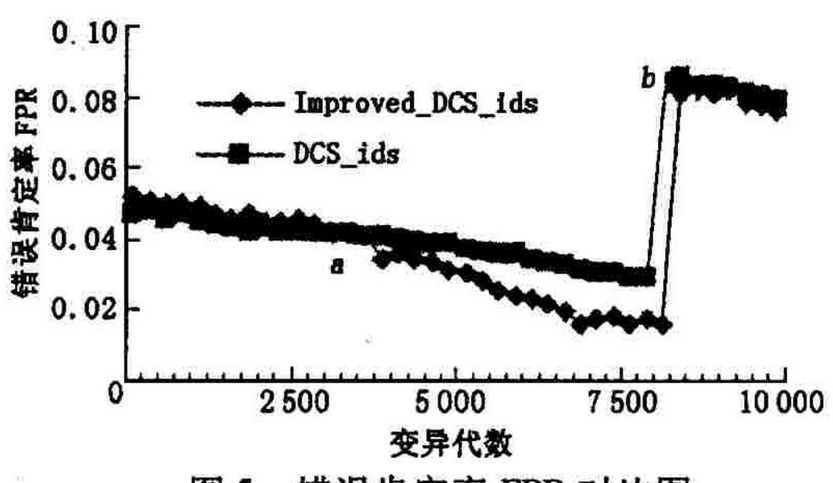

图 5　错误肯定率 FPR 对比图

4 结束语

实验结果表明,引入可控变异和随机变异的改进克隆选择算法在入侵检测中的应用中,能够发挥免疫计算的优势,同时又增加系统自适应能力,并降低系统误判率。免疫算法是免疫计算的灵魂,改进算法中,一些阈值的设定还值得深入研究,比如,可控变异 ab. age 的阈值 以及随机变异 ab. age 的阈值 等等,以便取得最佳效果。

参考文献:

[1] 李 涛. 计算机免疫学[M]. 北京:电子工业出版社,2004.

[2] Kim J Bentley P J. Immune Memory and Gene Library Evolution in the Dynamical Clonal Selection Algorithm[J]. Journal of Genetic Programming and Evolvable Machines,2004,5(4):361 -391.

[3] 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]. Oakland, 1994,202 -212.

[4] Hunt J,Timmis J,Cooke D,et al. The Development of an Artificial Immune System for Real World Applications[A]. Artificial Immune Systems and Their Apptications[C]. Springer - Verlag,1999,157 -186.

[5] De Castro L N,Von Zuben F J. The Clonal Selection Algorithm with Engineering Applications[A]. Proceedings of GECCO00 [C]. Las Vegas,2000,36 -37.

(编辑:门向生)

The Research for an Improved Dynamic Clonal Selection Algorithm Applied to Intrusion Detection

XU Chun, LI Tao, LIU Sun - jun, YANG Pin, LIU Nian, ZHANG Jian - hua

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

Abstract:Sets of self, non -self, antigen, antibody and immune cell are defined. Method of the appetency calculation is improved. An idiographic dynamic clonal selection algorithm is put forward based on controilable - aberrance and random - aberrance. An intrusion detection system (IDS) model based on the idiographic immune algorithm is designed and realized. Emulative experiment shows that the idiographic immune algorithm is effective in improving the self - adaptability of IDS.

Key words:dynamic clonal selection algorithm; intrusion detection; artificial immune i appetency

(上接第 49 页)

Research on One Type or Key Exchange Techniques in tntormation Hiding Techniques Based on Key

XU Run - ping, WANG Pan - qing

(The Ordnance Engineering Institute, Shijiazhuang, Hebei 050003, China)

Abstract: To improve the safety of key in information transmitting system, an individuation encryption approach based on existing information hiding system is designed and implementing steps are presented. The experiment result proves that both the security of key system and the performance of information hiding system have been improved after using this technology.

Key words: network security ; information hiding; individuation encryption; key exchange