基于人工免疫的记忆检测器研究

王 涛,张凤斌,李鑫鑫
(哈尔滨理工大学 计算机科学与技术学院,黑龙江 哈尔滨 150080)

摘 要:传统的手段已不能充分地解决计算机网络的安全问题。为了确保计算机网络系统安全,建立一个有效的入侵检测系统IDS,针对IDS中成熟检测器检测率低和错误肯定率高的问题,根据人工免疫记忆原理,研究了免疫检测器集中成熟检测器激活,记忆检测器生成与变异机制以及演化,给出了记忆检测器生成算法,研究了记忆检测器变异和淘汰机制。实验结果证明记忆检测器为主的检测器集合实现了检测器自学习和联想记忆的功能,提高了入侵检测系统的自适应能力和检测率,减少了错误肯定率。

关键词:错误肯定率;网络入侵;人工免疫原理;记忆检测器;联想记忆
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1673-629X(2008)06-0148-03

Memory Detector Research Based on Artificial Immune Theory

WANG Tao, ZHANG Feng-bin, LI Xin-xin
(Computer Science & Technology College, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: Conventional means haven't already been able to resolve the security problem of computer network, for ensuring computer network system security, and founding an effective intrusion detection system, aiming at mature detector inefficiencies and the high rate of errors sure problem. Under artificial immune theory, research immunoassay mature detector activation, memory detectors generation and variation mechanism and evolution, put forward the memory detectors generation algorithm, study the memory detectors variation and elimination mechanism. Experimental results show the main memory detectors of the detector set achieve a self-learning detectors and associative memory function, enhanced intrusion detection system and the adaptive ability of detection rate and reduce the rate of errors sure.

Key words: rate of errors sure; network intrusion; artificial immune theory; memory detector; associative memory

0 引 言

随着计算机网络应用的逐渐普及,计算机网络的安全问题越来越为人们关注。但是,传统的安全手段已不能充分地解决计算机入侵。为了确保计算机网络系统安全,入侵检测显得尤为迫切和重要。建立一个有效的入侵检测系统IDS,需要达到3个目标:分布、自组织和轻负担。自然界中,生物的免疫系统成功地保护生物自身免受病原体的侵害,免疫系统具有良好的多样性、耐受性、记忆性、自学习和自适应等特点,基于机体免疫系统的特性来建立一个鲁棒的、适应型的计算机入侵检测系统(IDS)是研究的重点。

收稿日期:2007-09-15
基金项目:黑龙江省博士后基金(LBH-Z05092)
作者简介:王 涛(1981-),男,河南西平人,硕士研究生,研究方向为信息安全;张凤斌,教授,博士生导师,研究方向为信息安全技术、网络安全及防范技术。

在IDS中,检测计算机网络入侵和攻击的检测器十分关键。免疫检测器的特性主要包括为特异性、多样性、自学习、自组织、记忆性和克隆选择等特性。自然免疫系统利用不同类型的防御细胞的共同努力,能够高效地、用最短的响应时间、最大限度地利用有限的资源来区分"自身"与"有害的非自身",保证生物体的存活和正常生理活动的进行。自然免疫系统的作用是保护生物体免受病菌的侵害,而IDS的作用是保护网络系统免遭黑客的入侵,两者具有非常类似的功能。在计算机免疫学上成熟检测器识别抗原(即检测到网络入侵),叫做初次免疫应答,记忆检测器识别抗原,叫做二次免疫应答,免疫应答的时间很短。检测器识别抗原是通过计算二者的亲和力 fitness (x, y) 来实现的。现在有很多基于自然免疫系统的成熟检测器生成算法,比如否定选择算法[1]及改进、克隆选择算法[2]及改进、基于基因库的检测器生成算法[3]等,但是使用成熟检测器检测抗原,检测效率低,错误肯定率高。

文中重点根据生物系统的免疫记忆原理,研究记忆检测器生成及演化策略,然后与成熟检测器比较来验证记忆检测器的检测效率和错误肯定率。

1　相关研究

克隆选择和否定选择是生物系统中抗体生成和演化过程中两个重要过程,是现代免疫学中比较完善的两个理论学说,也是入侵检测系统中成熟检测器生成的主要算法。

Forrest 在 1994 年成功地模拟细胞的免疫耐受过程提出否定选择算法(NSA)来处理各种异常检测问题[1]。这种算法定义"Self"为一个网络监测系统中的正常行为模式。随机产生许多模式,同每一个已定义的 Self 模式进行比较,如果随机产生的模式与其中一个 Self 模式匹配,那么这个模式被丢弃。否则,它将成为一个成熟检测器并且用于匹配系统中随后出现的异型模式"NonSelf",然后报警。

克隆选择学说认为机体免疫系统事先就存在能识别各种抗原的细胞克隆,每个克隆细胞表面都有针对不同特定抗原的受体,不同抗原选择与之相适应的受体结合,从而刺激该细胞克隆的增殖分化,产生免疫应答而生成多样性的各种抗体。克隆选择算法[2]如下:

```
Begin
  随机生成一个属性(免疫细胞)的群体
  while 收敛标准没有满足时/* 初始化 */
    Begin
      While not 所有抗原搜索完毕时
      Begin
       选择那些与抗原具有更高亲和力的细胞;
       生成免疫细胞的副本:越亲和力的细胞越有更多的副本;/* 再生 */
       根据它们的亲和力进行变异:亲和力越高,变异越小;/* 遗传变异 */
       End;
    End;
End;
```

2　记忆检测器的生成及演化

2.1　记忆检测器生成算法

在自然免疫系统中记忆细胞能够自动提取病菌抗原的签名,当再次检测到相同或相似的病菌抗原时能迅速活化,从而发起快速的免疫应答。记忆细胞是长寿细胞,其二次免疫应答的能力可持续相当长的时间甚至终身[4]。基于此原理,亲和力成熟的检测器升级成记忆检测器 Dm(类比于记忆细胞)。记忆检测器能够直接确认入侵。

抗原首次对成熟检测器刺激,成熟检测器被激活,当累计足够的亲和力(检测匹配次数达到一定阈值)时,激活成记忆检测器,然后进行克隆扩增,同时进行变异。若激活的成熟检测器在其生命周期内未能累计足够的亲和力,则走向死亡,并被新的成熟检测器代替。该过程和激活检测器的死亡机制确保了检测器的多样性,保证了其对入侵的持续搜索能力,并能保留最好的检测器。同时,对一个激活检测器,协调刺激也是需要的,这样可以降低错误肯定率。当亲和力成熟的激活检测器只有在协调刺激下才能变成记忆检测器。算法如下:

步骤 1:检测抗原,如果匹配,验证激活检测器亲和力是否达到成熟,亲和力成熟的转到第 4 步,不成熟的转到第 3 步;

步骤 2:对进入激活检测器集合的检测器设定生命周期,生命周期中亲和力不成熟的死亡;

步骤 3:验证激活检测器生命周期是否到期,到期的检测器死亡;没有到期的转回第 1 步;

步骤 4:验证是否受到协同刺激,刺激成功的则从激活检测器集合进入记忆检测器集合,变成记忆检测器,否则激活检测器死亡;

步骤 5:根据淘汰机制对记忆检测器进行淘汰,淘汰出的记忆检测器进入激活检测器集;

步骤 6:根据亲和力大小和变异规则,记忆检测器克隆扩增;

步骤 7:对变异生成的个体否定选择,成功地计算亲和力,子个体中亲和力最大的子个体亲和力大于父个体的,替换父个体,否则子个体全部死亡。

2.2　记忆检测器的变异

记忆检测器检测到的抗原是已经被确认的入侵,当遇到相同抗原时,能更快的反应。新加入的记忆检测器具有代表当前网络攻击趋势的特性,由于网络中新入侵一般是旧入侵的变种,对记忆检测器进行遗传变异,可以生成检测当前入侵变种的检测器;而对记忆检测器进行高频变异生成新的检测器,可以检测未知的入侵,保持对未知入侵的检测能力,增加检测器的适应能力。

具体方法:使用 r－连续位算法计算亲和力,当大于阈值时,认为是高亲和力,小于阈值时,认为是低亲和力。选择刚刚从激活检测器转变的记忆检测器进行变异,变异依靠检测器的亲和力,低亲和力的免疫检测器高频变异[4],高亲和力的遗传变异。更高亲和力的子代抗体进入记忆检测器库中,原父代抗体死亡。这样就确保了子代抗体至少具有与父代个体相同的亲

和力,确保个体的多样性和分布均衡。

高频变异时使用单点交叉和单点变异操作完成,交叉算子的使用可以扩大新的搜索区域,开拓新的模式,交叉点选取在属性边界。可以维持检测器的多样性,增大亲和力。

2.3 记忆检测器的淘汰

随着时间的推移,记忆检测器的数目可能变得很大,以至于超过机器的限制。当达到设定的阈值时,借鉴高速缓存中的替换算法进行替换。替换出的检测器进入激活检测器集合中而不是直接死亡,这样的话可以保持免疫检测器的完备性。主要淘汰算法有:

(1)最近最少使用 LRU(Least Recently Used)算法:将最近最少使用的记忆检测器替换出。缺点是把一些亲和力高的检测器替换。

(2)先进现出 FIFO(First In First Out)算法:将最先进入记忆检测器集的检测器替换出。缺点是把常用的检测器替换出。

(3)代价替换算法,该类算法使用一个代价函数对检测器进行评估,最后根据代价值的大小决定替换对象。

结合(1)和(3)进行替换:和 LRU 算法一致,只是满足一定条件的检测器才被替换。这样既能保持记忆检测器活性,又能使得亲和力满足一定条件的记忆检测器不被淘汰。其主要思想是:在所有记忆检测器中计算它们之间的亲和力,如果记忆检测器亲和力满足公式(1),删除此检测器。

$$E(A(x)) > E(A(Dm)) \tag{1}$$

其中

$$A(Dm) = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathrm{dist}(Dmi, Dmj) \tag{2}$$

$$A(x) = \sum_{i=1}^{n}\mathrm{dist}(x, Dmi) \tag{3}$$

$$E(A(x)) = A(x)/n \tag{4}$$

$$E(A(Dm)) = \frac{A(Dm)}{n(n-1)/2} \tag{5}$$

其中 $x, Dmi, Dmj \in Dm$,$\mathrm{dist}(x, Dmi)$ 为汉明距离,Dm 为记忆检测器,n 为记忆检测器阈值数。

3 实验及结果分析

3.1 实验设计及结果分析

实验数据使用 1999 年 DARPA 的入侵检测评估数据,该数据是美国 DARPA 资助的第二次入侵检测评估数据集合[5],它共包括五周的数据,我们只利用了第一周和第二周的数据,其中第一周数据是不包括攻击的训练数据,第二周数据包括一些攻击,在实验中用作测试数据。检测长度 $L = 96$(32 位源 IP 地址、32 位目的 IP 地址、16 位端口号和 16 位协议标志),选取初始自体集合 $S = 40$,取 $r = 10$ 进行匹配计算亲和力。

为了验证系统在引入记忆检测器后的性能,使用克隆选择算法生成成熟检测器集的系统表示 OLDIDS,然后在原系统上引入激活检测器和记忆检测器,新系统表示为 NEWIDS,分别做实验。对比实验在两个方面进行:检测效率和错误肯定率。

3.2 检测效率

检测效率是人工免疫系统的最重要的指标之一。记忆检测器变异的引入,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 1 所示,改进后的系统,只需很少变异,就迅速识别抗原,检测效率比改进前有明显提高。

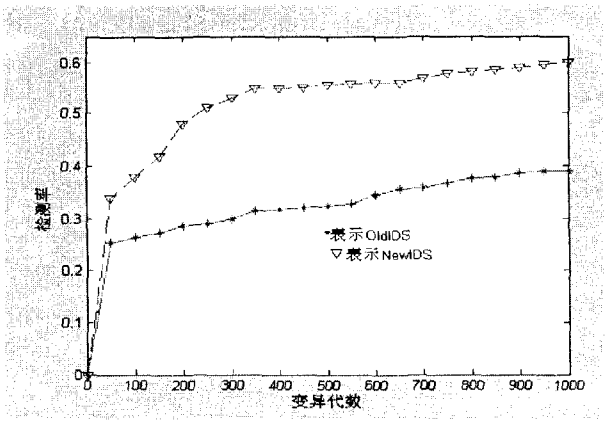

图 1 系统对 DDOS 的检测效率对比图

3.3 错误肯定率

入侵检测中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;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。FPR = 判断系统 FP 数/系统报警数。记忆检测器的变异,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。

在图 2 曲线 A 点,有较明显下降,应该是记忆检测器的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。

(下转第 155 页)

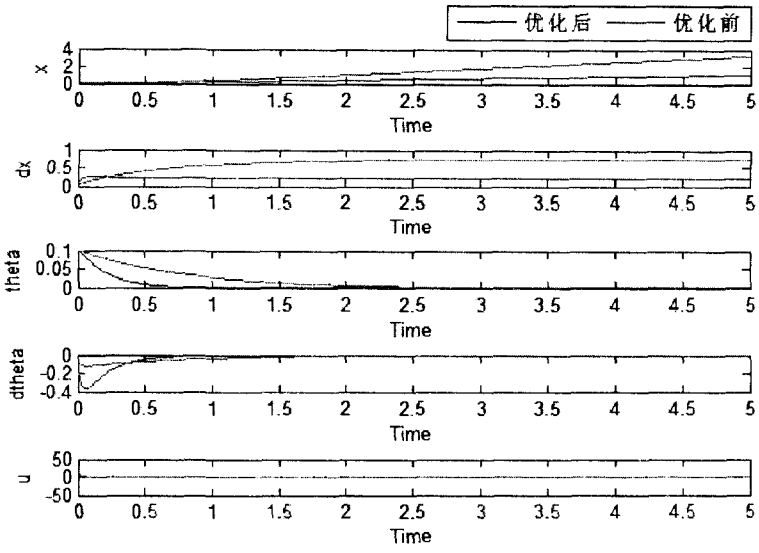

图 5　优化前后倒立摆系统仿真图

参考文献:

[1] Kennedy J, Eberhart R. Particle swarm optimization[C]// Proc. of the IEEE Conf. on Neural Networks, IV. Perth: IEEE Press,1995:1942－1948.

[2] Peng X, Venayagamoorthy G K, Corzine K A. Combined training of recurrent neural networks with particles swarm optimization and backpropagation algorithms for impendance identification[C]// Proc. of the IEEE Swarm Intelligence Symposium (SIS 2007). Honolulu, HI: IEEE Press, 2007:9－15.

[3] Carlisle A, Dozier G. An off－the－shelf PSO [C]// Proc. of the Workshop on Particle Swarm Optimization. Indiana, USA: IEEE Press,2001:1－6.

[4] Shi Y, Eberhart R. A modified particle swarm optimizer [C] // Evolutionary Computation Proc. of the IEEE World Congress on Computational Intelligence. New York: IEEE Press, 1998:69－73.

[5] Clerc M. The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization[C]//Proc. of the ICEC. Washington: IEEE Press, 1999:1951－1957.

[6] Kennedy M P. Three Steps to Chaos－Part I: Evolution[J]. IEEE Transactions on Circuits and System－I: Fundamental Theory and Applications,1993,40(10):640－656.

[7] 洪　旭.倒立摆系统模糊控制算法研究[D].西安:西安电子科技大学,2005.

(上接第 150 页)

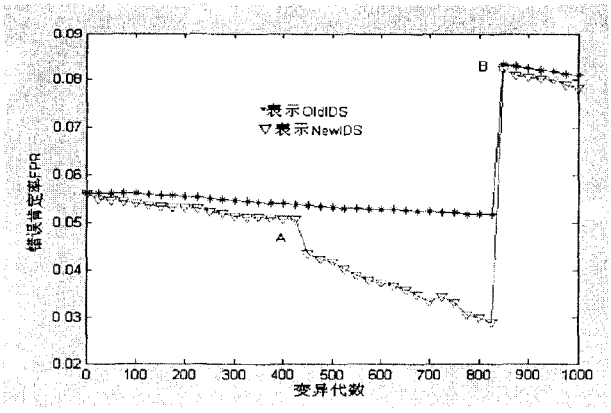

图 2　系统 FPR 对比图

4　结束语

根据生物系统的免疫记忆原理,由免疫检测器集中的成熟检测器生成记忆检测器,形成了记忆检测器为主的多层次检测器集合。实验证明有很好的检测率和学习能力,并能降低错误的肯定率,而对记忆检测器变异,进一步提升检测效率、增强自适应性和降低伪肯定率,然后再借鉴 Cache 中的替换算法给出记忆检测器的淘汰策略,使整个系统的性能进一步优化。但实验中的一些参数的设定对结果影响很大,下一步的工作是研究实验参数的影响程度。以提高入侵检测系统的自适应性和有效性。

参考文献:

[1] Forrest S, Perelson A S, Allen L, et al. Self－Nonself discrimination in a computer[C]// In Proceedings of IEEE Symposium on Research in Security and Privacy. Oakland, CA:[s. n.],1994:202－212.

[2] Kim J, Bentley P J. Towards an artificial immune system for network intrusion detection: An investigation of clonal selection with a negative selection operator[C]//Kim J H, Zhang B T, Fogel G, et al(Eds.). in The Congress on Evolutionary Computation (CEC－2001). Seoul, Korea: [s. n.], 2001: 1244－1252.

[3] Michaud S R, Lamont G B, Zydallis J B, et al. Protein Structure Prediction with Immunological EA Computation[C]//In: Proceeding of Genetic and Evolutionary Computation Conference (GECCO－2001). San Francisco, California: Morgan Kaufmann, 2001:1367－1374.

[4] 李　涛.计算机免疫学[M].北京:电子工业出版社,2004.

[5] Haines J W, Lippmann R P, Fried D J, et al. DARPA intrusion detection system evaluation: Design and procedures [R]. Technical Report 1062. Lexington: MIT Lincoln Laboratory, 1999.