基于人工免疫的新型入侵检测系统研究

黄均才[1]，王凤碧[2]，罗 讯[3]，佘 堃[3]，周明天[3]

(1. 东莞理工学院计算机系 广东 东莞 523106; 2. 东莞理工学院电子工程系 广东 东莞 523106;
3. 电子科技大学计算机科学与工程学院 成都 610054)

【摘要】分析研究了人工免疫原理在网络入侵检测中应用的可行性，结合人工免疫模型和数据挖掘技术建立了一个网络入侵检测系统模型。对抗体生成过程中的关键算法进行了描述。为克服在抗体生成阶段由于采用遗传算子导致时空效率不佳的缺陷，将数据流分割成字符串集合，根据数理统计原理，讨论了分割参数和检测器数目的选定，使它在通用性、鲁棒性上具有优势。

关 键 词 人工免疫理论； 入侵检测系统； 参数选定； 抗体生成
中图分类号 TP399 文献标识码 A

Research on a Novel Intrusion Detection System Based on Artificial Immune Theory

HUANG Jun-cai[1]，WANG Feng-bi[2]，LUO Xun[3]，SHE Kun[3]，ZHOU Ming-tian[3]

(1. Department of Computer Science and Technology, Dongguan University of Technology Guangdong Dongguan 523106;
2. Department of Electronic Engineering, Dongguan University of Technology Guangdong Dongguan 523106;
3. School of Computer Science and Engineering, UEST of China Chengdu 610054)

Abstract The feasibility of applying artificial immune theory in intrusion detection system is Analyzed, establishes a model combining artificial immune theory and data mining technique is estoblished. Based on the statistical theory, the amount of information that is lost by splitting a data stream into unordered strings can be estimated, and this estimate can be used to guide the choice of string length. Based on information- theory, a lower bound on the size of the detector set is derived. Detector Generating algorithm is described. The performance of Artificial Immune Intrusion Detection System (AIIDS) is better than the normal intrusion detection system based on knowledge engineering.

Key words artificial immune theory; intrusion detection system; choice of parameter; antibody generating

自然界中，生物的免疫系统成功地保护了生物自身免受外来病原体的侵害，入侵检测是计算机安全研究中的一个新领域，目前已有许多网络的入侵检测系统被开发出来，但大部分是基于知识工程的方法[1]，这使系统的灵活性和准确性不够，不能有效地识别新型攻击。免疫原理在入侵检测领域的应用是一个刚刚兴起的研究，它的目的是使检测系统具有分布性、多样性、自适应性、自动应答和自我修复的特点，具有检测异常现象、利用不完备信息进行检测的能力，这是原有系统达不到的。这方面已进行的工作参见文献[2-4]。

1 免疫系统的生物原理

生物免疫系统是多层免疫系统，如图1所示。最外层的是物理屏障的免疫，第2层是生理屏障的保护，第3层是先天性免疫系统。若病原体突破了前3层，则由淋巴细胞-T细胞和B细胞等构成的自适应免疫系统来处理。

根据免疫功能的不同，淋巴细胞可分为T细胞和B细胞两类。B细胞经

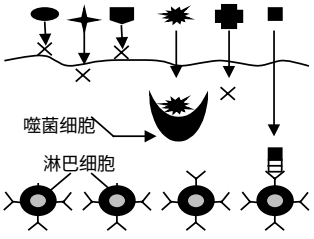

图1 免疫系统防御层次抽象图

收稿日期：2003 – 12 – 18
基金项目：电子产业发展基金资助项目(51415010101DZ02)
作者简介：黄均才(1973 –)，男，硕士，助教，主要从事生物信息学、网络计算、信息安全方面的研究.

过分化，一部分成为能产生抗体的浆细胞。抗体分子分为可变区和恒定区，可变区决定了抗体的特异性，产生的抗体和相应的抗原会发生特异性结合，将抗原杀死。另一部分发展为记忆细胞，记忆细胞对抗原十分敏感，能记住入侵过的抗原，当有同样抗原再次入侵时，记忆细胞能更快地做出反应。这就是适应性免疫反应。接种疫苗也可以建立适应性免疫。T细胞能够专门识别并直接破坏外来组织，发生免疫移植排斥反应，称为细胞免疫。T细胞抗原受体是能直接进入细胞膜内部和特异性抗原结合的蛋白质，以此进行识别，分裂产生大量的新的细胞，这些T细胞分泌细胞毒素，使移植器官的细胞溶解死亡[5]。

2 人工免疫原理在IDS上的应用

合格的入侵检测系统(Intrusion Detection System，IDS)要具有准确性、完整性、可扩展性、可适应性和自身的健壮性等特点[6]。这就要求它至少达到以下的几个设计目标：(1) 完备的；(2) 分布式的；(3) 自组织的和精简的。其中“精简”保障了系统的高效灵活[2]。

生物体免疫系统与IDS有着功能上的相似之处如表1所示。免疫系统的分布性、多样性、自成体系、完备性和精简性使它精确有效地保护着生物个体。如何模拟基因库更新、阴性选择、克隆选择等抗体生成过程建立入侵检测器是建立基于人工免疫原理的入侵检测系统的关键。

表1 生物体免疫系统概念和网络入侵检测系统概念对比

缩氨酸/抗原决定基	被检测的行为模式串
受体	检测模式串
单克隆淋巴细胞(T-细胞、B-细胞)	检测器
抗原	异己模式串
绑定	检测模式串和异己模式串的匹配
耐受性阴性选择)	阴性选择
淋巴细胞克隆	检测器复制
抗原检测	入侵检测系统的检测
抗原清除	检测器响应

为了便于表达和理解，根据功能比较将生物体免疫系统的有关术语和网络入侵检测系统的有关概念对照如表1所示。

3 一种基于免疫原理的入侵检测系统模型

3.1 模型描述

依据生物免疫的基本思想，将正常的数据访问当成是自己的正常行为，将异常访问当成异己行为，区别自己和异己从而判别出入侵行为。原理可表示为：

$$X_{non\_self,v}(\bar{x})=\begin{cases}1 & 若D(\bar{x},non\_self)<v\\ 0 & 其他\end{cases} \tag{1}$$

式中 v为预先设定的系统阈值；$D(\bar{x},non\_self)=\min\{d(\bar{x},\bar{s}):\bar{s}\in non\_self\}$。

式(1)说明，当一个行为模式和某一“异己”模式很相近或相同时可以判别为非法行为。因此从宏观的角度来看，基于免疫原理的检测系统如图2所示。

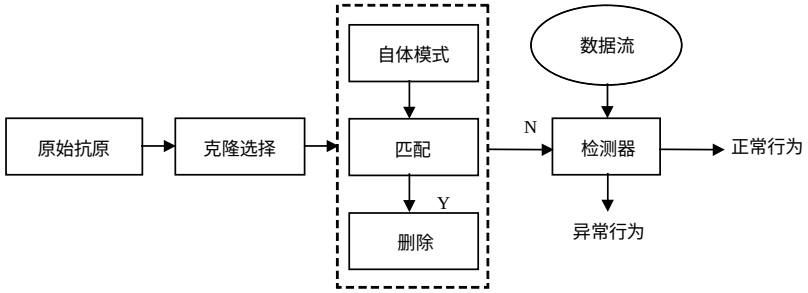

图2 基于免疫原理的入侵检测系统原理

在生物体中，抗体对抗原物质的识别是依靠抗体表面的受体与特定抗原的抗原决定基间化学健的“绑定”，安全系统中的检测是指检测模式和被检测模式间的匹配。

3.2 遗传算子运算

在初始父代抗原生成后，需要以它们为种子大量繁殖新的异己模式。这种演化采用简单遗传算法-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。这里采用单点交叉的方法，交叉点选取在属性边界，因此长度为K的两个模式会有个$2(K-1)$不同的交叉结果， 配种个体随机选择。变异的目的在于维持模式空间的多样性，扩大异己模式空间。在本应用中变异指的是单个属性值的突变。但为了防止算法趋于纯粹的随机过程，只对初始异己模式实施基本单值变异，变异限定在系统基因链中已有值。

3.3 重要参数的确定

对于任何一个实用的计算机系统，对应的数据量是相当大的，这里的数据包括存储设备上的文件，某过程的行为模式，以及程序的系统调用序列等。而抗体生成阶段采用了遗传算子，因此继承了遗传算法的特点，时空效率不佳，为克服其缺陷，可将数据流分割成字符串集合，根据数理统计原理确定其分割参数和检测器数目，因此，为提高系统的效率，必须事先将这些数据分割成由m种字符构成的长度为l的字符串，但当把连续的数据流分解成无序的字符串后会丢失一些信息，而这些信息对提高检测精度是很有帮助的，所以要确定一个合适的分割参数。另外，检测器的数目也必须加以考虑，检测器数目越多越有利于检测，但相应地检测器的产生时间会加长，检测器数目越少其产生时间越短，但又不利于检测。所以分割参 数和检测器数目都要慎重选择。

给定参数l(l表示字符串长度)，一个长为L bits的流 $\hat{S}$ 被分割成包含N_s个长为l bits的字符串集合S，因为这N_s个字符串中可能有相同的，现假定有k个唯一串，每个字符串s_i出现N_i次，即：

$$\sum_{i=1}^{k} N_i = N_s \tag{2}$$

给定字符串集合S(该字符串集合满足上述假设)，原始数据流可能是该字符串集合的N种可能的重新排列中的一种。其中：

$$N=\binom{N_s}{N_1, N_2, \cdots, N_k} = \frac{N_s!}{N_1!N_2!\cdots N_k!} \tag{3}$$

如果用ΔI 表示将流 $\hat{S}$ 分割成字符串集合S所丢失的信息，进一步根据信息熵理论，可以计算得信息的丢失量为：

$$\Delta I = O(L_s H(N_i / N_s)/l) = \ln\binom{N_s}{N_1, N_2, \cdots, N_k} \tag{4}$$

因此为了使丢失的信息量最少，可以选择对应的每比特的信息熵(平均信息量)最少的l值。

如果检测器也是一些m种字符组成的长为l的的字符串，类似地，可得到检测器数目N_R应满足

$$N_R \geq \frac{N_s \ln(1/P_f)}{l \ln(m)} \tag{5}$$

式中　P_f为异己字符串被漏报的概率。

3.4 适应度计算和阴性选择[7]

如前所述，为保证克隆选择的需要，必须对新生模式进行适应度测定。即计算新模式的基因型与已有异己模式基因型的“距离”，如与某一基因型的距离小于预定阈值σ 且不为零者，接纳为候选检测器。适应度的计算按式(6)进行：

$$\min(\sum_{i=0}^{3} | code_1(i) - code_2(i) |) \tag{6}$$

式中　$code_1$表示待测模式的编码；$code_2$为原有抗原群体中某一模式的编码。这里取最小值，事实上，只要有小于σ 的结果出现即可停止该模式的计算。必须对候选检测器进行阴性选择。同样，依据式(6)计算候选模式和自体集中模式的距离，距离不为零者保留为检测器。

(下转第99页)

4　结 论

FRAM-FD是一个相对飞行安全风险指数的计算模型，不仅可以对一个航段、一条航线、一个机组或一个飞行单位的安全进行定量的评价，而且能为安全形势的发展趋势分析、预测提供依据。进一步的研究包括在考虑飞行参数相关性的前提下，参数同时和依次超限时的风险指数的计算方法及基于知识的规则的优化和对其它飞行事故和事故征候的分析等。

参 考 文 献

[1] Roelne A L C. The development of aviation safety performance indicators: An exploratory study [M]. Amsterdam, Netherlands: National Aerospace Laboratory, 1998:12-16.

[2] Warren D. Guidelines and methods for conducting the safety assessment process on civil airborne systems and equipments[S]. SAE ARP4761, U S: Aerospace Recommended Practice, 1996: 46~55.

[3] Robert D. CFIT checklist: Evaluate the risk and take action[M]. Alexandria, VA, U S: The Flight Safety Foundation. 1994:43-49.

[4] Biggs D, Hamilton G. Risk indicators and their link with air carrier safety[J]. Flight Safety Foundation Digest, 2001(10):1-6.

[5] Jones S. An overview of the NASA aviation safety program assessment process[C] // Proceeding of AIAA's 3rd Annual Aviation Technology, Integration, and Operations. Denver, Colorado, U.S: AIAA. 2003: 17-22.

编　辑　徐安玉

(上接第95页)

4　结 束 语

本文分析了生物体免疫的基本原理，说明入侵检测和生物体免疫系统的相似性，给出了两者相关概念的比较。提出一种基于免疫原理的网络入侵检测系统模型，对模型中的关键部分给出了较详细的说明。可以看出，基于免疫原理的入侵检测系统建立的开销主要集中于抗体生成阶段，本文该阶段采用了遗传算子，因此继承了遗传算法的特点，为克服遗传算法的时空效率不佳的缺陷，将数据流分割成字符串集合，根据数理统计原理确定其分割参数和检测器数目，使它在通用性、鲁棒性上具有优势。一旦有了合适的抗体集，这种方法的检测效率是很高的。

参 考 文 献

[1] Mukherjee B, Heberlein L T, Levitt K N, Network intrusion detection[J]. IEEE Network, 1994,8(3):26-4l.

[2] Kim J, Bentley P. The artificial immune model for network intrusion detection[C]. 7th European Conference on Intelligent Techniques and Soft Computing, Aachen, Germany, 1999:35-40.

[3]Dasgupta D. Immunity-based intrusion detection system: a general framework[C]. In Proc. of the 22 nd NISSC, Arlington, Virginia, USA, 1999:113-127.

[4] Hofmeyr S. An immunological model of distributed detection and its application to computer security[D]. Santa Fe USA: University of New Mexico, 1999.

[5] 何球藻, 吴厚生. 医学免疫学[M]. 上海: 上海医科大学出版社, 2000.

[6] Forrest S, Hofmeyr S, Somayaji A. Computer immunology[J]. Communications of the ACM, USA, 1997, 40(10): 88-96.

[7] 杨向荣, 沈钧毅, 罗　浩. 人工免疫原理在网络入侵检测中的应用[J]. 计算机工程, 2003, 26(6): 27-29.

编　辑　孙晓丹