文章编号:1673-9469(2008)02-0052-03

基于生物免疫原理的过滤器算法研究

黄光球,张干涛
(西安建筑科技大学 管理学院,陕西 西安 710055)

摘要:从实际应用的角度,分析了负检测方法的局限性,针对网络入侵检测的效率问题,构建了一个具有动态自适应性的过滤器系统,对过滤器检测元的生成算法进行了深入的研究。根据正选择算法识别自我的独特性,借鉴动态科隆选择算法的思想,以正选择算法为核心并结合数据挖掘技术来确保检测器的生成效率及检测效率。

关键词:免疫原理;过滤器;正选择;动态克隆选择;检测元
中图分类号: TP31 **文献标识码**:A

Research on a filter algorithm based on biological immune theory

HUANG Guang-qiu, ZHANG Gan-tao
(College of Economics & Management, Xi'an University of Architecture & Technology, Xi'an 710055, China)

Abstract: The paper analyzes the localization of the negation detection approach. A dynamic and auto-adapted filter system is established in view of the question of the network intrusion detecting efficiency. The generation and detecting efficiency of detection are ensured by the specificity of the positive selection algorithm that can discern self; the algorithm uses the dynamic clone selection algorithm as the frame, the positive selection algorithm as the core and data mining technique.

Key words: immune theory; filter, positive selection; dynamic clone selection; detector

生物免疫系统(Biological Immune System , BIS)具有良好的多样性、耐受性、免疫记忆、分布式并行处理、自组织、自学习、自适应和鲁棒性等特点。BIS的这些诱人特性,引起了研究人员的普遍关注。随着计算机安全技术研究的不断深入,人们发现,计算机安全技术与生物免疫系统所遇到的问题具有惊人的相似性,两者都要在不断变化的环境中维持系统的稳定性[1]。

计算机安全系统对计算机系统的保护正如同生物免疫系统对于生物体的保护。免疫学者传统上把免疫系统抵御病菌入侵描述为区分"自我"和"非我"。"自我"是人体内正常的细胞;而"非我"是指那些外来的病菌。类似的,保护计算机系统不受恶意入侵也可以被视为区分"自我"和"非我"。最早提出用于"自我"和"非我"辨识的仿生物免疫检测算法的是美国新墨西哥大学的Forrest教授。她提出了一种负选择的思想,这成为仿生物免疫检测算法的基本思路。但单纯的负选择算法所带来的最大的弊病就是庞大的抗体(检测器)数目所带来的时间和空间的巨大代价[2]。在前人研究的基础上,张雅静以负选择算法为基础,结合生物学中的小生境策略及适应性免疫机制,提出了一种负选择模式匹配检测算法(简称NSMA),该算法具有较高的检测效率及不占用过多的时间和空间的优点。本文以正选择算法为基础提出一种过滤器系统模型并给出检测器生成算法,进一步提高检测效率。

1 免疫原理在入侵检测中的应用

当前,利用生物免疫学原理进行计算机网络入侵检测的技术主要有两种:正检测方法和负检测方法。正检测方法(positive detection)也是最直接和最早使用的方法,它是直接将正常的、合法的

收稿日期:2008-03-24
基金项目:陕西省自然科学基金资助项目(2004F14);西安建筑科技大学基础研究基金项目(JC0616)
作者简介:黄光球(1964-),男,陕西西安人,教授,博导,从事电子商务与网络安全的研究。

或可接受的操作模式定义为自我集,以此来构建正常行为模式库,再对操作模式进行监视,通过在正常模式库中搜索,若未找到相关记录就标记为不匹配,若找到相关记录就说明是合法操作或正常操作模式[3],与其对应的免疫学算法是正选择算法。

在生物免疫系统中,那些能识别出自体 MHC (self - MHC)分子的 T 细胞能被激活,是依赖于 T 细胞自身的正选择。这个过程产生一种 T 细胞属性,通过它,只有与自体 MHC 有联系的 T 细胞才能识别抗原[1]。

正选择算法如图 1 所示。具体步骤如下:

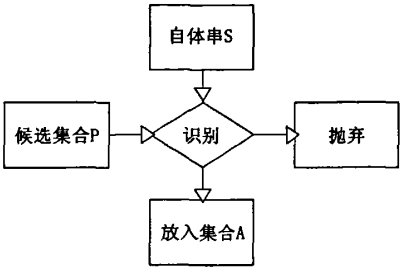

图1 正选择算法

Fig.1 the positive selection algorithm

初始化:产生一个 T 细胞候选集合 P。假设所有的分子都用长度为 l 的二进制串表示,则可能产生 2^l 个不同的细胞。

亲和力计算:通过集合 S,计算 P 中所有元素与自体 S 的亲和力。

可用集合的产生:如果 P 中某个元素与 S 中某个元素的亲和力大于或等于一个给定阀值 ϵ,即这个 T 细胞能识别这个 MHC 分子,则它肯定被系统选用,放入集合 A,否则删除它。

二进制字符串间的匹配规则有很多,如统计规则(Statistical)、Euclidean 距离、海明距离(Hamming Distance)和 R 连续匹配规则(R Contiguous)等.Harmer 在实验中对比了这几种常用规则,结果证明 R 连续规则具有很高的信号噪声比(Signal to Noise),在本文中采取该匹配规则,计算抗体抗原间亲和力用 r 连续位匹配算法。在两个二进制串 x 和 y 的相应位置上,至少有连续 r 位相同,则 x 和 y 就是 r 连续位匹配的。具体计算见式(1)。

$$g_{affinity}(x,y)_r=\begin{cases}1 & iff\ \exists\, m,n,r(x_k=y_k,k\in[m,n]\\ & r\in N,r\le n-m,0<m<n<l)\\ 0 & otherwise\end{cases} \tag{1}$$

式中 l—抗体和抗原的长度。

随机产生的抗体需要经过自体耐受过程才能生成可用集合,根据正选择算法的原理,在耐受期内与自体集任一元素都不匹配的抗体被删除,有匹配的抗体放入可用集合 A 中,见式 2。

$$h_{tolerance}(x)=\begin{cases}1 & iff\ \exists\, y\in S,\\ & g_{affinity}(x,y)_r=1\\ 0 & otherwise\end{cases} \tag{2}$$

其中 $x\in P$。

负检测方法的核心思想是定义一个自我集作为训练集来产生不与自我集模式匹配的检测元集,使用这些检测元来进行入侵检测。与其对应的免疫学算法是负选择算法。负选择算法是对生物免疫细胞的成熟过程的模拟,其与正选择算法非常类似,但作用刚好相反。对负选择算法来说,与自体匹配的元素必须清除,而与所有自体元素都不匹配的元素保留为可用元素。

2 网络入侵检测系统的改进

2.1 局限性

在计算机网络系统中,待检模式中的绝大多数都是正常模式,异常攻击模式只属个别极少数,而从以负选择算法为核心的负检测方法的实质来看,系统中的检测元只是为了检测异常模式,而不是为了检测正常模式。另外,从负选择算法中可以看出,检测元在生成过程中要经历一个类似于免疫耐受的审查过程,只有与自我集的任何模式都不匹配的检测元才会成为有效检测元,由此可见,在检测过程中将代表正常连接的自我模式与检测元进行匹配试验是毫无意义的,因为这些自我模式根本就不可能与检测元集中的某个检测元匹配。因此,让大量的正常模式与检测元进行匹配试验就势必会导致系统在性能上存在一定的局限性,降低系统的检测效率[3]。

2.2 检测方法的改进

在实际应用中,计算机网络是一个实时的持续变化的系统,Kim 和 Bentley 于 2002 年提出的动态克隆选择算法(DynamiCS),主要是针对系统实时变化的情况而设计的,用于网络的入侵检测,该算法是基于 Hofmeyr 提出的 CIS 的基本概念,主要是想精简出对系统适应性有关键作用的部分,其适应性有 3 个不同的检测器群体相互协调实现,

它们分别是未成熟,成熟和记忆检测器群体[1]。同时,对单一的网络系统来说,由于其使用用户相对固定,待检模式中会有大量的相同或相近的模式高频率出现,我们称之为高频模式。若能在待检模式进入检测器之前就已将其中的高频正常模式过滤掉,就可大大减少系统的整体匹配比较次数,提高检测效率。这里借鉴动态克隆选择算法在适应性方面的优势和正选择算法,对其改进来生成具有自适应性的高频模式过滤器,来达到过滤高频待检模式的目的。

2.3 高频模式检测器生成算法

检测器定义:系统使用检测器模拟免疫细胞和抗体的功能,定义监测器及检测器集合如下:

$D=\{d|<s,h,g,t,p>,s\in U,h,g,t,p\in N\}$

每个检测器 d 是一个五元组,其中 s 是抗体,也是长为 n 的01串,h 是检测器年龄即时间计数器初始值为0,g 是检测器的代数,刚出生为一代,t 为检测器的生存期即最高代数,p 是检测器与抗原的单代匹配次数,可表示匹配效率。U 为字符表为0和1字符串长为 l 的全集空间即 $U=\{0,1\}^l$, N 为自然数集。

检测器生成算法具体步骤如下:

Step1:产生随机的未成熟机检测器;根据系统保留数据挖掘出频繁模式补充入未成熟检测器集。

Step2:用正选择算法比较未成熟检测器与所给的自体集,发生匹配的检测器放入成熟检测器集,删除那些与自体集不发生匹配的监测器,再补充入新的未成熟检测器。

Step3:网络数据作为测试集来对成熟监测器进行免疫耐受,成熟检测器与网络数据比较,匹配则p值加1;网络数据为通过异己模式集被删除后的网络数据流。

Step4:判断是否 g > t,是则删除成熟检测器。

Step5:判断是否 p > = v, v 是匹配效率阀值,是,生成高频模式检测器,否则,检查 h 是否大于或等于 δ, δ 是成熟检测器单代生存期阀值,是,h 清0,并且 g 加1,返回 Step3。

Step6:删除检测效率低于阀值的高频模式检测器,删除重复的高频模式检测器,补充入新生成的高频模式检测器;生成未成熟检测器,使未成熟检测器、成熟检测器和高频模式检测器的总数等于预设的最大值。

3 算法性能

比较原检测方法与改进后的检测方法。事实上,自我集 S 中的正常模式在出现频率上存在大的差异,即正常模式的出现频率呈现出分散分布的点。鉴于此,我们使用高频模式检测器生成算法获得自我集中高频率出现的模式来构建过滤器,这样即可保证过滤器远小于异己模式检测器集又可起到过滤掉占较大比例的属于正常模式的高频模式的作用。在文献[3]中,魏春英、刘培玉已经证明了改进后的检测方法比原检测方法具有更高的检测效率。

采用数据挖掘算法,根据保留数据提取频繁模式,这样做以现有自我模式为基础,而非一种不可能存在的模式,这样做有利于保障检测器生成效率。

算法采用动态科隆选择算法的核心思想,针对实时持续变化的系统设计过滤器,使过滤器系统具有自组织、自学习自适应和鲁棒性的特点。

4 结束语

基于免疫原理的新过滤器系统具有自组织、自学习自适应和鲁棒性的特点,能够实时有效的检测高频网络连接模式并过滤掉,从而有效的提高了网络入侵检测系统的检测效率。

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] 张雅静,单国东.基于生物免疫原理的负选择模式匹配检测算法[J].计算机工程与应用,2004,(16):15-17.
[3] 魏春英,刘培玉.一种基于生物免疫学原理的入侵检测方法[J].计算机技术与发展,2006,(2):226-228.
[4] Kasthurirangan Parthasarathy , Clonal Selection Method for Immunity Based Intrusion Detection [N/OL]. http://web.umr. edu/~ tauritzd/courses/cs447/project/Parthasarathy.pdf 2/1/2002~3/21/2003.
[5] FORREST S, PERELSON A, ALLEN L, *et al*. Self-nonself discrimination in a computer [A]. In proceedings of the 1994 IEEE symposium on research in security and privacy [C]. Los. Alamitos., CA: IEEE Computer Society Press, 1994: 202-212.
[6] 赵林惠,戴亚平,徐立新. 免疫学原理在入侵检测中的应用研究[J].计算机应用,2005,(8): 1726-1729.

(责任编辑 刘存英)

基于生物免疫原理的过滤器算法研究

作者： 黄光球， 张于涛， HUANG Guang-qiu， ZHANG Gan-tao
作者单位： 西安建筑科技大学,管理学院,陕西,西安,710055
刊名： 河北工程大学学报（自然科学版） ISTIC
英文刊名： JOURNAL OF HEBEI UNIVERSITY OF ENGINEERING(NATURAL SCIENCE EDITION)
年，卷(期)： 2008,25(2)
被引用次数： 1次

参考文献(6条)

1. 李涛 计算机免疫学 2004
2. 张雅静;单国东 基于生物免疫原理的负选择模式匹配检测算法[期刊论文]-计算机工程与应用 2004(16)
3. 魏春英;刘培玉 一种基于生物免疫学原理的入侵检测方法[期刊论文]-计算机技术与发展 2006(02)
4. Kasthurirangan Parthasarathy Clonal Selection Method for Immunity Based Intrusion Detection 2003
5. FORREST S;PERELSON A;ALLEN L Self-nonself discrimination in a computer 1994
6. 赵林惠;戴亚平;徐立新 免疫学原理在入侵检测中的应用研究[期刊论文]-计算机应用 2005(08)

引证文献(1条)

1. 黎明.范震宇.张岩.宋广军 生物免疫机理在网络防盗链中的应用[期刊论文]-网络安全技术与应用 2013(11)