基于人工免疫的网络入侵检测模型的研究

张玉芳[1] 熊忠阳[1,2] 孙桂华[1] 赖 苏[1] 赵 鹰[1]
(重庆大学计算机学院 重庆 400030)[1] (重庆大学信息与网络管理中心 重庆 430030)[2]

摘 要 针对现有的应用于网络入侵检测中的人工免疫系统存在的缺陷,在Kim小组的动态克隆选择算法的基础上,提出了改进的网络入侵检测模型。在该模型中,提出产生少量的自体模式类对正常访问数据进行处理,加快其访问速度;通过动态增减自体集合来适应网络环境的变化,并且解决传统AIS中自体集合庞大的问题;采用基于约束的检测器表示抗体,采取任意R位间隔匹配规则来判定抗体与抗原之间的匹配,使用分割算法来解决抗体与自体抗原的匹配情况。最后,对该模型进行了网络入侵检测仿真实验,并与相同实验条件下的动态克隆选择算法的实验结果进行了对比,验证了所提模型的有效性和可行性。

关键词 入侵检测,人工免疫,抗体,抗原
中图法分类号 TP309 **文献标识码** A

Research of Network Intrusion Detection Model Based on Artificial Immune

ZHANG Yu-fang[1] XIONG Zhong-yang[1,2] SUN Gui-hua[1] LAI Su[1] ZHAO Ying[1]
(Department of Computer Science, Chongqing University, Chongqing 400030, China)[1]
(Network and Information Management Center, Chongqing University, Chongqing 430030, China)[2]

Abstract Aiming at limitation of existent network intrusion detection model with artificial immune idea, an improved network intrusion detection model based on dynamic clonal selection algorithm was presented. For accelerating normal IP packets access, the self-pattern class was proposed and most of self-antigens were filtered and amended the self-antigen set dynamically in detection process. The constraint-based detectors were adopted as antibody, any-r intervals matching rule was used to determinant antibody and antigen, and split-detector method were settled to self-antigen matching. The experimental results show that the proposed model can achieve a faster running speed, the better detecting rates, and adapt to dynamically changing environments.

Keywords Intrusion detection, Artificial immune, Antibody, Antigen

1 引言

现有的应用于网络入侵检测中的AIS还存在着许多缺陷。其一是自体集合非常庞大,而未成熟检测器只有在这些自体集合中通过自体耐受后,才能成为成熟检测器。因此,未成熟检测器进行自体耐受的时间代价太大,可行性存在着问题;另外,系统中自体、非自体的描述是一种静态方式,一旦定义后就极少变化。同时抗体与抗原的匹配规则多采用RCMF规则[1,2],虽然匹配规则实现容易,并能够以少数的抗体识别较大范围内的抗原,得到较高的检测率。但是该方法计算量大,效率低;同时,该匹配规则尚存在着大量漏洞,导致检测器无法检查一些异常数据。为了能够检测到尽可能多的入侵访问数据,并且尽可能地降低误报率,则需要保证检测器的多样性。因此,入侵检测系统将会产生大量的检测器来检测多种类型的入侵IP包。然而,在实际网络中,大部分的访问数据都是正常的,如果都要经过大量的检测器一一进行检测,那么对于正常访问数据来说,访问速度太慢。

针对上述存在的问题,本文在Kim小组提出的动态克隆选择算法DynamiCS[3]的基础上进行改进,提出了改进的网络入侵检测模型。第2节给出相关定义,整个模型框架在第3节介绍,网络入侵检测仿真实验安排在第4节,最后是小结。

2 相关定义

相关术语[4]的定义如下:

定义1(自体与非自体) 抗原集合分为自体集合和非自体集合。非自体为来自网络攻击的IP数据包;自体为来自正常网络活动的IP数据包。

定义2(抗体) 抗体用于检测非自体抗原。抗体又包括未成熟抗体、成熟抗体和记忆抗体。

定义3(自体模式类) 采用聚类算法得出的自体模式类。

自体模式类表示大量正常数据的共有特征,因此,自体模式类能检测出大部分自体抗原。

3 基于人工免疫的网络入侵检测模型框架

基于人工免疫的网络入侵检测模型主要包括6个重要模块:自体模式类检测模块、记忆抗体检测模块、成熟抗体检测

到稿日期:2009-05-09 返修日期:2009-07-01

模块、自体集合动态演化模块、未成熟抗体自体耐受模块、分割算法模块。

模型流程如图 1 所示，系统主要分为 3 个流程：一是系统运行前的预处理，得到自体模式类；二是抗原处理流程（图中实线箭头）；三是抗体演化过程（图中虚线箭头）。

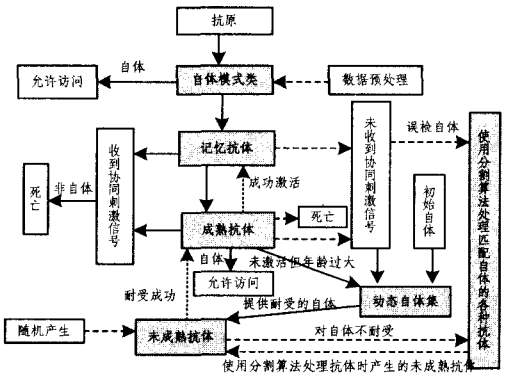

图 1　基于人工免疫的网络入侵检测模型

3.1　自体模式类检测模块

在实际网络中，大部分是正常的访问数据，如果都要经过大量的检测器一一检测，则访问速度太慢。在系统运行前，为了避免自体模式类误将非自体抗原归类为自体而允许访问，从而造成系统的入侵，因此在进行聚类分析时，需保证自体模式类的精确率，使其所检测到的抗原均为自体。

通过 k-means 算法[5]进行聚类分析，得到少量的表示大量正常数据共有特征的自体模式类。将这些自体模式类放入系统运行，首先与少量的自体模式类进行匹配，若匹配，则表示为自体，允许访问；否则表示该 IP 包有可能是入侵，需要经过检测器集合进行检测判别，这样提高了正常访问数据的访问速度。通过自体模式类识别后的自体抗原，不用再经过抗体的检测，也就不存在抗体对这些自体抗原误检的问题，因此这些自体抗原就不需要放入自体集合中提供未成熟抗体的耐受，大大减少了自体集合的规模，从而提高了自体耐受的效率。

3.2　记忆抗体检测模块

记忆抗体检测模块是对免疫机制中的二次应答机制进行模拟，当遇到相同或相似的抗原，将快速地检测出来，不用再进行学习。若遇到以前未曾出现过的自体抗原，并且与一些非自体抗原非常相象，则有可能导致记忆检测器误将该自体检测成非自体抗原，因此，需要采用协同刺激机制来确定所检测到的抗原是否为非自体。

3.3　成熟抗体检测模块

成熟抗体检测模块模拟了成熟免疫细胞的免疫机理，在检测模型中起着检测抗原的作用，是模型的主要检测部分。成熟抗体为已经通过自体耐受，但还未被抗原激活的抗体。成熟抗体是系统的发展力量，通过不断学习进化以及死亡机制，确保抗体的多样性，保证了抗原空间的持续搜索能力，并能保留那些最好的抗体。

成熟抗体检测模块与记忆抗体的检测流程非常相似，相同点表现为：一是均采用高亲和力优先检查的原则对抗原进行检测；二是抗原与抗体匹配后，均要进行协同刺激：若收到协同刺激信号则杀死该非自体抗原，并增加该抗体的亲和力；若在一定时间内未收到协同刺激信号，则将误检的抗体进行分割算法处理，并将该自体放入自体集合中。

二者也存在区别，区别具体体现在：输入、输出、生命周期以及抗体来源的不同。

3.4　动态自体集合

由于系统环境的不断变化，原来是正常的访问数据在下一时刻有可能成为异常数据，因此为了适应该需求，本文采用动态的自体集合。自体的来源为在线收集来自网络上某一段时间内，经过抗体与自体模式类过滤后的正常访问数据，并形成相应的、供未成熟抗体进行自体耐受的自体集合。在自体动态变化过程中，每隔一段时间清空自体集合，加入最新自体，保证数量不会很大，这样大大降低了在自体耐受过程中时空复杂度上的系统开销。

3.5　未成熟抗体自体耐受模块

未成熟抗体主要有 3 个来源：一是通过系统随机生成；二是当记忆抗体或者成熟抗体误检自体时采用分割算法产生，这部分新的未成熟抗体含有以前曾经检测到非自体抗原的信息，带有记忆信息，以便下次碰到同样的抗原，能快速检测；三是当未成熟抗体在进行自体耐受误匹配自体时采用分割算法产生，这部分未成熟抗体含有以前的耐受信息，以便后面以更快的速度进行耐受。

未成熟抗体为新生成的、尚未完成自体耐受的抗体。它是整个系统的新生力量，通过不断产生未成熟抗体，进行不断的学习与耐受，来检测新出现的非自体抗原，使系统具有更好的适应性。

3.6　分割算法

当抗体误将自体检测成非自体抗原时，需要采用分割算法对该抗体进行处理。分割算法的核心操作是对匹配自体抗原的抗体进行分割，删除匹配自体抗原信息，保留其余抗体信息生成新的未成熟抗体，即分割抗体。对于采用基于约束的检测器与任意 r 位间隔匹配规则[6]结合来表示抗体检测机制，使用分割算法能够删除该抗体中误检自体的信息，保留其中大量已经匹配非自体抗原的信息或者已经耐受的信息。

4　仿真实验与分析

4.1　实验环境

实验数据来自 KDD1999[6]。该数据集包含 24 种攻击类型，共有 41 个特征。

首先从 KDDCUP99 数据集中随机选取 5000 个记录，分为 3 个子群体，每个子群体都含有不同的入侵类型。在实验中，每隔 N 代选择一个新的子群体；每一代随机选择该子群体中 80%的抗原进入系统，实现了系统环境的不断变化（此处每一代代表一天；每 N 代表示一周。在整个系统中，相当于每一天都面临着不同的网络数据包；一周之后，重新采用一个子群体。）。

基于以上的系统环境，在每一代中动态地将经自体模式类、抗体过滤后的自体，以及抗体误检的自体加入到自体集合当中。同时在每 N 代结束后，定时清空自体集合，最终实现自体集合的动态变化。

4.2　实验结果

本模型在 matlab7.0 中直接使用 k-means 工具包，提取实验数据的 3500 个记录进行聚类得到自体模式类。本文具体做了 3 个实验。

(1) 实验 1:本文提出的改进方法。

(2) 实验 2:未采纳自体模式类,其余同实验 1。

(3) 实验 3:传统的动态克隆选择算法,采用 RCMF 匹配规则。

在实验过程中,依据经验值,采用试探法进行参数调整,具体参数设置如表 1 所列。

表 1 参数设置表

参数名称	取值
r(任意 r 位间隔匹配规则阈值)	6
N(选择子群体代数阈值)	5 代
自体模式类个数	2 类
未成熟抗体耐受期	8 代
成熟抗体激活阈值	18 个
成熟抗体寿命	2 代
非记忆抗体的总个数	50 个
总迭代次数	300 代

实验结果图中,实验 1 用红实线表示;实验 2 为蓝虚线;带 * 的黑虚线代表实验 3。

从图 2 可以看出:

• 由于实验 1 与实验 2 每隔 N 代定期地将自体集合进行清空,一些未成熟抗体对清空的自体可能不耐受,最后导致误检;然而,实验 3 采用传统的动态克隆选择算法,自体集合规模不断增加,虽然否定选择需要花费大量的时间,但是却保证了低误检率。

• 实验 1 通过自体的动态变化,花费少量的时间进行自体耐受,误检率只比实验 3 高 0.2%~0.3%左右,但是 TP 检测率却高出 5%~6%左右。因此,本模型采用的策略是通过牺牲误检率来加快检测速度与提高检测率的。

从图 3 可以看出:实验 1 能够更快地提高检测率,同时还能在自体耐受和协同刺激失败时保留过去检测的信息,使系统更快地适应动态网络环境的变化;自体模式类的采用,使后面的抗体检测部分面对的网络环境相对稳定些,也使模型更快趋向稳定。

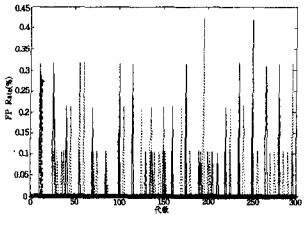

图 2 3 个实验的 FP 检测率比较图

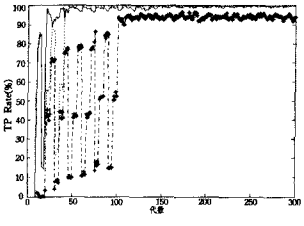

图 3 3 个实验的 TP 趋势比较图

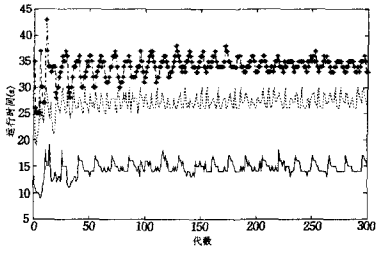

图 4 3 个实验在检测速度上的比较

图 4 表明:实验 1 的检测速度较快。这是由于实验 1 不仅采用了基于约束的检测器、任意 r 位间隔匹配规则结合的机制,加快了检测速度,而且采用了自体模式类,加快了正常访问数据的访问速度,同时,自体模式类过滤了大量的自体,大大加快了每代的自体耐受时间,因此每代的运行时间减小了许多。另外,自体耐受的时间减短,也加快了系统对动态网络的适应时间,增强了系统的健壮性。

表 2 列出了 3 个实验在比较稳定时的平均 TP,FP 以及运行时间。结果表明,改进的模型能够极大地提高 TP 检测率,加快检测速度,更快地适应网络环境的动态变化。

表 2 3 个实验的 TP,FP 及检测时间

实验 \ 指标	TP 检测率	FP 误检率	每代检测时间
实验 1	99.47%	0.211%	14
实验 2	99.39%	0.233%	27
实验 3	93.94%	0%	34

结束语 引入自体模式类不仅过滤了大量的自体,减小了自体集合的规模,提高了自体耐受的效率,而且加快了正常数据的访问。采用新型的抗原-抗体编码与任意 r 位匹配规则加快了检测速度及提高了检测率;采用分割算法删除误检信息,保留已检测抗原信息或者已耐受信息。采用动态的自体集合,每隔一段时间对自体集合清空,同时在线收集来自网络上经过抗体与自体模式类过滤后的正常访问数据,保证自体集合的及时更新,并且数量不大。由于自体模式类过滤掉大量的自体,使自体集合比其他系统要小很多,因此,提高了自体耐受的效率。采用抗体与生命周期集合机制,不但实现了各种抗体的动态更新,也实现了自体的动态变化,构成了未成熟抗体动态的耐受机制,从而不断更新抗体使得系统的运行处于一个动态环境中,因此系统具有很好的自适应性。

参 考 文 献

[1] Hofmeyr, Forrest. Immunity by design: An artificial immune system[C]// Proceeding of the Genetic and Evolutionary Computation Conference(GECCO'99). 1999:1289-1296

[2] Esponda F, Forrest S, Helman P. A Formal Framework for Positive and Negative Detection Schemes[J]. IEEE Transactions on Systems, Man and Cybernetics-part B: Cybernetics, 2004, 34(1):357-373

[3] Kim J, Bentley P J. Towards an Artificial Immune System for Network Intrusion Detection: An Investigation of Dynamic Clonal Selection[C]// Evolution Computation CEC'02. May 2002:1015-1020

[4] Zhang Yu-fang, Sun Gui-hua, Xiong Zhong-yang. A Novel Method of Intrusion Detection Based on Artificial Immune System[C]// Proceedings of the 2006 International Conference on Machine Learning and Cybernetics, ICMLC 2006. 2006:1602-1608

[5] Han Jiawei, Kamber M. Data Mining Concepts and Techniques. 2nd edition[M]. Beijing: China Machine Press, 2006

[6] http://kdd.ics.uci.edu/databases/kddcup99/task.html

[7] Zhang Ya-jing. A Novel Immune Detection Algorithm for Anomaly Detection[C]// Proceedings of the 2005 IEEE International Symposium on Intelligent Control Limassol, Cyprus. June 2005:1441-1446

[8] Li Rui-fan, Wang Cong, Tu Xu-yan. A New Immunity-based Model for Network Intrusion Detection[J]. Networking, Sending and Control, March 2005:106-109

[9] 王益丰,李涛,胡晓勤,等. 一种基于人工免疫的网络安全实时风险检测方法[J]. 电子学报,2005,33(5):945-949