

融合危险理论和否定选择的入侵防御系统

魏 萌, 贾遂民, 胡明生

(郑州师范学院 信息科学与技术学院, 河南 郑州 450044)

摘 要: 为了提高入侵防御系统的检测性能, 提出了一种结合免疫危险理论和否定选择机制的入侵防御模型. 模型中使用危险理论进行异常检测, 使用否定选择机制进行误用检测, 二者结合使用提高了检测性能. 给出了模型的实现思想和关键实现技术, 并进行了仿真实验. 结果表明, 在真实收集的网络数据流和 KDDCup 检测评估数据上, 本模型都具有较好的检测性能, 提高了入侵防御能力.

关键词: 入侵防御模型; 危险理论; 否定选择; 人工免疫

中图分类号: TP18; TP393

文献标识码: A

文章编号: 1000-7180(2014)07-0109-04

An intrusion Prevention Model Based on Combination of Danger Theory and Negative Selection

WEI Meng, JIA Sui-min, HU Ming-sheng

(School of Information Science and Technology, Zhengzhou Normal University, Zhengzhou 450044, China)

Abstract: In order to improve the detection performance of intrusion prevention systems, an intrusion prevention model which combines immune danger theory and negative selection mechanism was proposed. It uses the danger theory for anomaly detection and negative selection mechanism for misuse detection, which improves the detection performance. The implementation and key technologies of the model are given, and the simulation experiments are done. The experimental data included the data collected from the actual network and KDDCup intrusion detection evaluation data sets. The results show that the model has a good detection performance.

Key words: intrusion prevention model; danger theory; negative selection; artificial immune

1 引言

入侵防御是目前网络安全防范的主要技术之一, 其主要特点在于对网络流量进行检测和分析, 进而采取主动防御措施. 生物免疫系统的自我防御与网络入侵防御有着惊人的相似性^[1], 二者都需要在不断变化的环境中保持自身免受侵害. 同时, 生物免疫的防御功能也为网络入侵防御的研究提供了一条新的思路和途径^[2]. 已有的基于免疫原理的入侵检测和防御主要有两种方法: 否定选择机制和免疫危险模型. 这两种方法借鉴不同的免疫机理, 在入侵检测中取得了很好的求解效果, 但在应用方面还存在一定的局限性^[2]; 否定选择机制可用性较差, 尤其是

数据维数较高时; 而危险模型机制检测性能偏低. 此外, 已有的研究中, 多是将两种免疫机制分开进行讨论, 对两种免疫机制的协同研究报道尚不多见. 基于此, 本文分析了免疫否定选择机制和免疫危险理论之间的异同点, 设计了一种两种机制进行协作的入侵防御机制, 提高了检测性能.

2 设计思想和算法模型

2.1 设计思想

生物体是一个复杂的自适应智能体系, 其免疫系统可以分为固有免疫系统与自适应免疫系统. 其中, 固有免疫系统是抵抗抗原入侵的第一道防线, 可以阻止大部分入侵抗原; 自适应免疫系统能够记忆

收稿日期: 2013-09-13; 修回日期: 2013-11-08

基金项目: 国家自然科学基金项目(U1204703); 河南省重点科技攻关项目(122102310004)

入侵的抗原特征,可以进行下一次快速响应.

否定选择机制模拟生物体识别自体和非自体的免疫耐受过程,其认为:生物体对所有的自体进行耐受(即不产生排斥反应),而对所有的非自体进行应答.不同于否定选择机制,免疫危险理论的基本思想是:免疫系统并不是对所有的非自体都产生应答,而是只对危险的非我进行应答^[3].

在生物免疫系统中,固有免疫系统和适应性免疫系统并不是孤立存在的,而是协同工作,为生物体提供保护.免疫系统不但能快速检测已知入侵抗原,而且能检测未知抗原,这对基于免疫的误用检测和异常检测进行协同工作提供了重要的借鉴意义^[4].

本文设计了一个双层机制,首先使用基于危险理论的异常检测,根据行为特征进行检测,识别危险的网络行为;然后对判别为正常的网络行为,使用否定选择机制进行误用检测,根据结构特征进行检测.实验结果表明,通过危险理论和否定选择机制的协同,提高了检测的准确率.

2.2 算法模型

本文设计的入侵防御模型如图 1 所示.模型主要由数据采集系统、入侵分析和入侵防御部分组成.

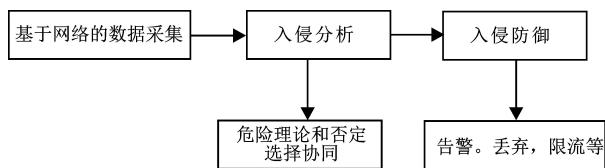


图 1 入侵防御系统模型

模型首先需要对网络上的数据流进行采集,数据采集系统部分通过 Windows 平台下的 Winpcap 抓包工具来实现,提取出网络的通信特征;入侵分析部分通过免疫危险理论和否定选择机制的协同来对网络行为进行分析;入侵防御部分对检测到的入侵进行相应的处理,如模拟吞噬细胞的功能,对数据包进行丢弃处理或者告警后交给网络管理员进行人工处理.

为了便于后面的说明,本文基于免疫的入侵防御与生物免疫系统的映射关系如表 1 所示.

表 1 生物免疫系统与入侵防御系统的映射关系

生物免疫系统	入侵防御系统
生物体	网络主机
抗原	网络数据流
自体抗原	正常网络行为
非自体抗原	非法网络行为
抗体	检测器
抗体识别抗原	检测器检测网络行为

本模型的根本目标描述为:根据进入的网络流,输出检测到的异常入侵.基本步骤如下:

(1)对输入的数据流进行判断(根据危险信号),如果不是异常抗原,则将其加入自体集合;否则,输出抗原的动态异常指标;

(2)对进入自体的抗原进行自体耐受,然后通过否定选择机制对网络流进行检测,输出异常行为.

3 算法关键技术

本入侵防御模型中,入侵数据检测是关键.其中,主要关键技术有基于危险理论的异常检测和基于特征代码的误用检测,具体描述如下:

3.1 基于危险理论的异常检测

危险理论是人工免疫的主要算法之一,以危险作为识别对象,可有效避免自体/非自体识别中难以解释的问题.危险信号的生成是危险理论的关键技术之一.危险信号决定哪些抗原可以参与否定选择过程.当有危险信号的时候,相关抗原被禁止进入否定选择过程.本文基本思路为:当抗体浓度超过一定阈值时,则为异常情况,否则,为正常.

依据危险理论,抗原的死亡有两种不同方式^[5]:凋亡与坏死,不同的死亡方式对抗体有不同的作用.本文中,抗原凋亡表示网络行为正常结束,抗原坏死表示网络行为非正常结束.相应地,抗原凋亡产生抗体抑制信号,而抗原坏死会产生抗体激励信号.因此,可以通过抗体的浓度表示危险信号的强弱.

抗体浓度定义为

$$\rho(t) = \sum_{i=1}^{n_1} \text{new}(t) + \sum_{i=1}^{n_2} \text{necro}(t) - \sum_{i=1}^{n_3} \text{apopt}(t),$$

式中, $\sum_{i=1}^{n_1} \text{new}(t)$ 、 $\sum_{i=1}^{n_2} \text{necro}(t)$ 、 $\sum_{i=1}^{n_3} \text{apopt}(t)$ 分别表示 t 时刻新加入的 n_1 个抗原、以坏死方式死亡的 n_2 个抗原(坏死的抗原增强抗体浓度)、以凋亡方式死亡的 n_3 个抗原(凋亡方式抑制抗体浓度).

当 $\rho(t) \geq \theta$ 时候,认为有异常情况,否则认为正常.

3.2 基于否定选择的误用检测

经过危险理论的检测后,被认为安全的抗原进入自体集合.否定选择机制认为生物体对自体进行耐受,而对非自体进行反应.否定选择分为两个过程:产生阶段和检测阶段.在检测器产生阶段,给定一个自体样本集合,然后随机产生候选检测器.根据

定义的匹配规则,如果检测器与自体样本匹配,则丢弃它;否则,增加到检测器集合.在异常检测阶段,将待检测的数据与检测器集合进行匹配,如果匹配成功,则是正常数据;否则为异常数据.

可见,对于否定选择过程来说,检测器的生成是其关键技术.这里,采用改进的 V-detector 算法^[6],主要步骤描述如下:

步骤 1:随机产生候选检测器 x ;

步骤 2:计算 x 与已有检测器集合 D 中检测器 $d_i (i = 1, 2, \dots, n)$ 的欧氏距离 $L_{d_i} = \text{Euclidean}(d_i, x)$;

步骤 3:如果 L_{d_i} 小于任一检测器 d_i 的半径 $r(d_i)$,说明点 x 已经被检测器覆盖,则放弃该点,同时计数器 $t = t + 1$ (t 记录任一随机点被已有检测器覆盖的次数);否则转步骤 5;

步骤 4:如果 $t \geq 1/(1-c)$ (c 为期望覆盖率),则说明覆盖率已经足够,结束算法;否则,转步骤 1;

步骤 5:计算 x 与自体样本集合 S 中的自体样本点 s_i 的欧氏距离 $L_i = \text{Euclidean}(s_i, x) (i = 1, 2, \dots, n)$,记录距离最近的两个自体点及其距离,分别记作 $(s_1, L_1), (s_2, L_2)$,且 $L_1 \leq L_2$,并计算 $r_1 = L_1 - r_s, r_2 = L_2 - r_s, r_s$ 为自体半径.

步骤 6:如果 $r_1 > 0$,此时不再使用 (x, r_1) 作为新的检测器,而是移动检测器的位置到 $p = x + \frac{r_2 - r_1}{2} \times \frac{x - s_1}{\|x - s_1\|}$ ($\| \cdot \|$ 表示 n 维向量范数),同时更改检测器的半径为 $r_p = \frac{r_1 + r_2}{2}$,则新的检测器为 (p, r_p) ,将 (p, r_p) 作为新的检测器加入到检测器集合 D ;

步骤 7:如果 $r_1 \leq 0$,说明点 x 落在自体区域,则 $T = T + 1$. 如果 $T > 1/(1-c_s)$ (c_s 为最大自体覆盖率),则算法结束;

步骤 8:如果检测器集合 $|D|$ 达到最大检测器数量 $N_{\max}$,则结束;否则,转步骤 1.

这里,使用改进的 V-Detector 算法训练检测器集,用来检测网络通信特征,进而判定当前网络是否存在异常.这种改进的检测器生成算法扩大了检测器的覆盖空间,使用更少的检测器可以检测到更大的异常,提高了检测性能.后面的实验也证明了这一点.

此外,考虑到现实中,自体集合和非自体集合都是动态变化的,相应的自体耐受过程也是动态的,所以检测器要经过未成熟状态、成熟状态、激活状态、

记忆状态和死亡状态,每个状态对应各自的生命周期.在该算法中新检测器不断生成,过期的检测器不断被清除,保证整个过程适应实际的网络变化情况(具体参考文献[1,7]).

综上,通过理论上的创新,本模型有效提高了入侵防御系统的检测性能.

4 系统仿真实验与分析

为了验证本模型性能,分别采用了在网络实验室捕获的实际数据和 KDDCup 1999 入侵检测评估数据数据进行了实验.实验中采用的是 Windows XP 操作系统,开发工具为 Visual C++ 6.0. 实验中参数的设置:最大自体覆盖率 $c_s = 99.99\%$,期望覆盖率 $c_s = 99\%$. 自体半径和阈值对算法检测结果影响较大,经过实验,设置为自体半径 $r_1 = 0.1$,浓度阈值 $\theta = 0.3$.

实时的网络数据通过 Windows 平台下的 WinPcap 抓包工具来实现并对数据进行特征分析.自体和非自体采用实值表示^[4]:将网络信息流映射为对应的自然数,用最小最大化的方法将数据集进行归一化处理,规范到超立方体 $[0, 1]$ 中.具体过程参考文献[5].收集实验室网络中一周的正常网络流量数据作为训练数据,利用模拟工具对网络进行 syn flood, teardrop 等攻击,将攻击后产生的网络流量与正常的网络流量一起作为测试数据.

为了验证算法效果,将本文算法与采用否定选择的异常检测(文献[8])和基于危险理论的异常检测(文献[9])进行比较.实验结果如表 2 所示.

表 2 在真实数据集上的性能比较

所用方法	检测率/%	虚警率/%
文献[8]	96.1	3.35
文献[9]	76.4	1.22
本文算法	97.1	2.19

KDDCup1999 入侵检测评估数据数据本身是高维的数据,而实值否定选择算法本身对高维数据处理能力较差^[4]. 本文在实验中,引入信息增益对数据进行降维处理^[5]. 首先对数据进行归一化处理.在样本库的选择上,采用 KDDCUP99 10% 的数据集作为基准数据.实验中,采用不包括任何攻击的数据作为训练数据,测试数据包含有正常数据和攻击数据.用 5 个不同的数据集(其中包含攻击类型不同)对系统进行测试,同样的数据重复进行 10 次实验,并与文献[8]、文献[9]进行比较,实验结果如表 3

所示.

从实验结果来看,本模型具有较高的检测率.不论是在真实的网络数据流还是 KDDCup99 入侵检测评估数据上,均取得了较好的实验结果.本模型的检测率平均可以达到 97% 左右,虚警率可以降低到 4% 以内.与文献[8]相比较,本文的检测率和虚警率都有所降低.相比于文献[9],本模型的检测率有明显提高,但虚警率也有所增加,但是幅度不大(检测率和虚警率是相互制约的),本模型将危险理论和否定选择机制结合使用,提高了检测性能,也验证了思路的可行性.

表 3 Kddcup99 数据集上的性能比较

方法	检测率/%	虚警率/%
文献[8]	94.2	2.51
文献[9]	79.8	1.73
本文算法	96.3	2.44

5 结束语

本文提出了一种集成免疫否定选择机制和危险理论的入侵防御模型.实验结果表明,本模型有较高的检测性能.如何进一步优化危险信号的提取和检测器的生成及入侵防御手段的完善是下一步研究的方向.

参考文献:

- [1] 柴争义,刘芳,朱思峰.新型智能入侵防御模型[J].华中科技大学学报:自然科学版,2010,31(1):28-30.

- [2] 张然,刘洋,尹毅锋,等.基于 Multi-Agent 的入侵检测动态协同机制研究[J].微电子学与计算机,2013,50(8):57-62.
- [3] 傅蓉蓉,郑康锋,芦天亮,等.基于危险理论的无线传感器网络入侵检测模型[J].通信学报,2012,33(9):31-37.
- [4] 陈岳兵,冯超,张权,等.一种面向入侵检测的集成人工免疫系统[J].通信学报,2012,33(2):125-131.
- [5] 王彦涛.基于免疫计算的实时入侵检测系统的设计与实现[D].西安:西安电子科技大学,2012.
- [6] 柴争义,王献荣,王亮.用于异常检测的实值否定选择算法[J].吉林大学学报:工学版,2012,23(1):61-65.
- [7] 柴争义,刘芳.应用危险理论的网络安全风险感知模型[J].北京邮电大学学报学报,2010,33(3):40-43.
- [8] Gonzalez F A, Dasgupta D. Anomaly detection using real-valued negative selection[J]. Genetic Programming and Evolvable Machine, 2003, 4(4): 383-403.
- [9] Gu F, Greensmith J, Aickelin U. Further exploration of the dendritic cell algorithm[C]//International Conference on Artificial Immune Systems. Phuket, Thailand:Springer,2008:142-153.

作者简介:

魏 萌 女,(1981-),硕士,讲师.研究方向为网络与信息安全,人工智能,数据挖掘.

贾遂民 男,(1968-),副教授.研究方向为网络异常检测、智能计算.

胡明生 男,(1973-),博士,副教授.研究方向为人工智能、数据挖掘等.

(上接第 108 页)

- [8] 蒲凤平,陈建政.基于 ZeroMQ 的分布式系统.电子测试,2012(7):24-29.
- [9] 袁培森.内存数据库的设计与实现[D].南京:南京航空航天大学,2009.

作者简介:

许登元 男,(1971-),博士,教授.研究方向为计算机网络、无线传感器网络.

王 鹏 男,(1988-),硕士研究生.研究方向为计算机网络、嵌入式.