

高速网络入侵防御系统中免疫算法的改进

盛云雷, 李永忠, 荆春伟

(江苏科技大学 计算机科学与工程学院, 江苏 镇江 212003)

摘 要: 将人工免疫原理与高速网络分流技术结合, 提出了一种基于免疫原理的高速网络入侵防御算法. 在对阴性选择算法改进基础上引入疫苗算子和选择算子, 降低了系统二次免疫应答时间, 增强了系统抗体库自我学习能力, 进而改善了系统的实时性和高效性. 在分析了高速网络分流技术和改进阴性选择算法的基础上, 建立了基于免疫原理的高速网络入侵防御系统. 仿真结果证明了该算法的有效性.

关键词: 入侵防御; 人工免疫理论; 疫苗算子; 选择算子

中图分类号: TP301.6

文献标识码: A

文章编号: 1000-7180(2011)05-0109-04

Improve of Immune Algorithm in the Intrusion Prevension System of High-Speed Network

SHENG Yun-lei, LI Yong-zhong, JING Chun-wei

(School of Computer Science and Engineering, Jiangsu University of Science and Technolog, Zhenjiang 212003, China)

Abstract: A new method is proposed against the limited capacity in making real-time intrusion prevention efficiently for high-speed network, that is the intrusion prevension algorithm based on immune principle, combining of the artificial immune principle and data-distribution technology. In the research on negative selection algorithm, the vaccine operator and positive selection operator are introduced to the algorithm. Owing to the new algorithm, self-learning ability of antibody bank gets enhanced and the time needed for secondary response is reduced, so that the efficiency of system gets greatly improved. After the analysis of data-distribution technology and negative selection algorithm, the intrusion prevension system based on immune principle is established. Simulation results show the effectiveness of the algorithm.

Key words: intrusion prevention; artificial immune theory; vaccine operator; selection operator

1 引言

入侵防御系统(IPS)是在入侵检测系统(IDS)基础上发展起来的网络新生安全产品. 与入侵检测系统相比,入侵防御系统设备串联在网路中,能够检测到网络攻击并及时进行阻断,更加有效的保护网络的安全. 随着人工免疫系统的发展,将免疫原理运用到网络入侵防御中已经成为近年来的研究热点^[1].

文中将免疫原理引入高速网络中进行分析,利用对高速网络分流后网路流量大大减轻的改善结

果,在免疫算法的研究基础上引入免疫算子和选择算子,使系统免疫响应时间缩短、检测效率提高.

2 高速网络入侵防御免疫系统

该系统主要由分流器、事件采集机、事件分析器、规则库、响应单元以及策略库组成. 具体结构如图 1 所示.

分流器按照一定的负载均衡算法与链路完整性原则完成对数据的转发分流. 事件采集机功能类似于抗原提呈,负责从系统外部的网络环境中收集事件,将其转化为系统可处理的模式. 事件分析器根据

收稿日期: 2010-07-30; 收稿日期: 2010-08-28

基金项目: 江苏省教育厅、江苏科技大学课题(2005DX006J)

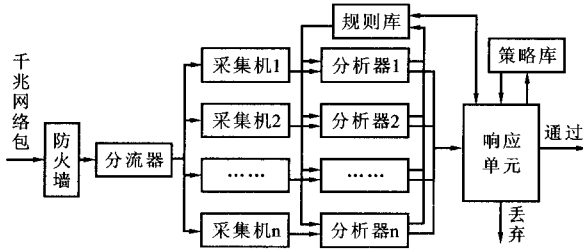


图1 基于免疫原理的高速网络入侵防御系统结构图

人工免疫原理,计算亲和力,区分事件采集机所采集到得网络事件是自体的还是非自体的. 规则库是入侵防御系统中用于辅助事件分析器对网络事件进行分析的依据. 规则库负责组织现有的规则,通过接受外界提供的信号完成规则的激活等操作. 响应单元根据检测器对网络事件的亲和力积累结果,以及策略库中的响应方式和协同刺激信号,对激活的网络事件作出准确的响应,同时更新规则库中的检测器集、自体集等信息. 策略库为响应单元执行动作提供依据,即为响应单元提供协同刺激信号^[2].

图1中,网络事件分析器需要利用规则库中的各类检测器进行判断,同时分析的结果也反馈给规则库,对其进行动态更新. 规则库主要由各类检测器集和自体集库等组成,自体集即正常网络事件,通过保存正常网络事件,为未成熟的检测器的免疫耐受提供依据. 响应单元综合各模块的信息,对网络事件进行最终的处理. 策略库同时也是系统与管理进行交互的主要途径,提供系统动态信息给管理员,以便管理员及时对系统进行调整.

3 阴性选择算法及其改进

在入侵免疫防御系统中,抗原是指待解决的网络行为,自体抗原是指正常的网络行为,非自体抗原是指异常网络行为或是入侵行为. 抗体则是能够识别入侵行为的检测器. 入侵防御系统中的网络行为正常与否在生物免疫系统中就表现为识别抗体是自体还是非自体的过程.

3.1 阴性选择算法

1994年Forrest提出了传统的阴性选择算法,该算法采用部分匹配规则(连续 r 为匹配规则)^[3],算法描述如下:

(1) 定义自体为长度等于 l 的字符串集合,自体集 S (S 为高速网络环境中正常的长度为 l 的数据串的集合),成熟检测器集 R ,确定匹配规则、匹配阈值 r 以及期望达到的系统匹配失败率 P_f 等参数;

(2) 提取从高速网络环境中分流采集的长度为 l 的二进制字符串 a ;

(3) 字符串 a 依次与自体集 S 中的自我元素相匹配;

(4) 依据步骤(1)中定义的匹配规则与阈值判断 a 是否与自我集中的元素匹配,若匹配则返回步骤(2)和(3),若不匹配则成为有效检测器,将其加入成熟检测器集 R ;

(5) 持续将 R 中的元素与自体集 S 比较达到对 S 的监控.

然而,传统算法存在这样两个问题:① 连续匹配规则的匹配概率相对较小,以至系统所需有效检测器的数量很大,检测器的生成时间过长,这在高速网络背景下是个致命的缺陷;② 传统算法中存在着不可避免的漏检区域,即“黑洞”问题. 针对这两个问题,分别在匹配规则及匹配阈值上进行改进,采用基于海明距离的变阈值阴性选择算法^[4]. 算法对步骤(4)进行了改进:

(4') 确定初始阈值 r_0 、最大匹配阈值 r_m ,根据海明匹配规则,确定 a 是否与自我集中的元素相匹配,若匹配则令阈值 $r = r + 1$,再若 $r \leq r_m$ 则返回步骤(3),否则返回步骤(2);

(5') 如果字符串 a 通过步骤(4')并且其阈值 $r \leq r_m$,则字符串 a 成为匹配阈值为 r 的有效检测器,加入成熟检测器集. 算法流程如图2所示.

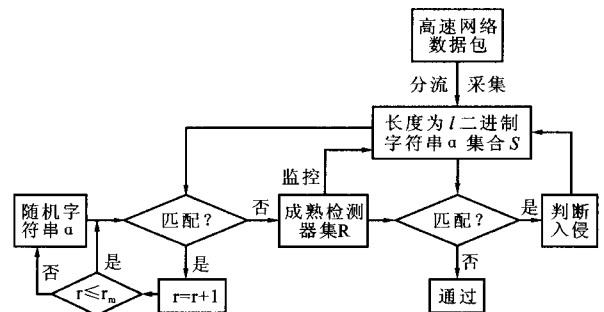


图2 基于海明距离的变阈值阴性选择算法

3.2 疫苗算子和正选择算子

在自体集相对稳定的情况下,成熟检测器集只需生成一次即可,并不需要频繁的生成有效检测器,在第二次入侵时完全不必要像第一次检测那样复杂,因此在基于海明距离的变阈值阴性选择算法中加入疫苗算子,形成含有疫苗的抗体库,当同样的抗原入侵时,可以首先用疫苗检测到,大大缩短了二次应答的时间^[5].

疫苗算子主要通过接种疫苗来实现,在检测过程中由算法自己产生疫苗,将初次应答的抗原作为

抗体疫苗加入抗体库,再利用正选择的方法进行待检数据,若匹配则视为异常.正选择是相逆于阴性选择的一个成熟抗体的过程.算法中引入正选择算子检测的目的就是将抗体库中的疫苗提取出来对数据进行检测,以便高效地获取抗原信息,提高检测的效率和准确性.正选择算子是疫苗算子完成其功能的具体实施途径.

3.3 算法的改进

在算法的异常监测部分加入疫苗算子和正选择算子,通过对抗原形成免疫记忆功能,可以缩短系统检测时间.而算法关键在于疫苗的提取,改进后算法过程如下:

- (1) 由基于海明距离的变阈值否定选择算法生成有效检测器集 R ;
- (2) 初始化抗体库为有效检测器集 R ;
- (3) 对自我集进行疫苗正选择检测(抗体库中无疫苗则跳过),若匹配则告警丢失该数据并记录时间开销,若不匹配则进入步骤(4);
- (4) 对自我集进行 R 检测,若不匹配则进入步骤(5),否则生成疫苗算子,更新抗体库,进入步骤(5);
- (5) 若检测完成则数据通过,若未完成则回到步骤(3).算法流程图如图 3 所示.

4 仿真实验

4.1 海明匹配变阈值阴性选择算法检测率分析

为得到正常的高速网络数据流,采用一台千兆计算机作为流量发生器,用来产生高速数据流.测试数据集采用 MIT 美国林肯实验室的 DARPA 1999 IDS 入侵检测评估数据集 kddcup99 第 1 周和第 3

周的数据.第 1 周作为训练数据(无攻击行为),用于产生成熟检测器集,第 3 周的数据作为测试数据(包含部分标识的入侵行为).实验数据如表 1 所示.

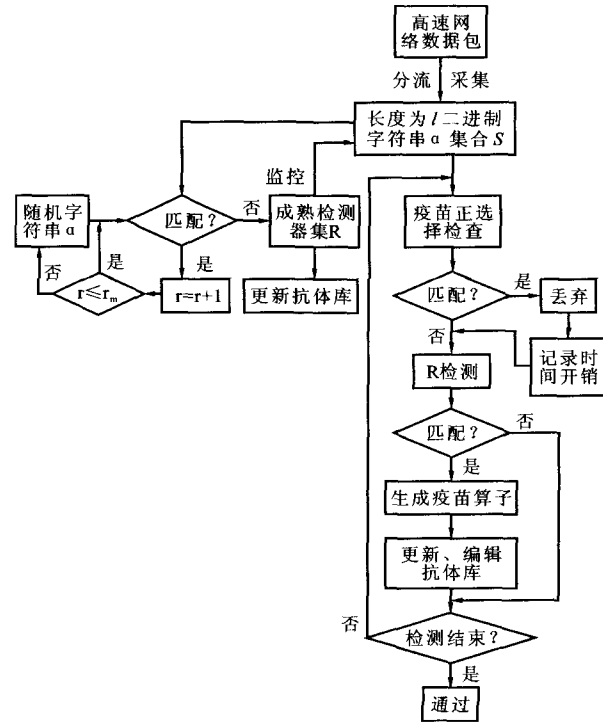


图 3 改进算法基础上加入疫苗算子流程图

4.2 疫苗加入后检测效果分析

疫苗提取采用下面的方法:在海明规则变阈值负选择算法产生检测器集 R 后,改变自我集一部分(改变字、片段和源文件,可以重复改变多次),模拟抗原入侵,并用 R 来进行检测,若发生正选择匹配则提取为疫苗并加入抗体库,此时抗体库包含了疫苗和检测器集 R ,然后参照图 3 算法流程对系统进行监测.

表 1 海明规则变阈值负选择算法 / 传统算法的系统入侵检测对比

检测时刻 $t(s)$	带有人侵行为的 数据包(个)	$N_s = 50$				$N_s = 100$			
		检测数(个)		检测率		检测数(个)		检测率	
		改进后	改进前	改进后	改进前	改进后	改进前	改进后	改进前
10	9	2	1	0.21	0.08	3	1	0.3	0.1
20	17	4	1	0.23	0.09	7	2	0.4	0.13
30	29	10	5	0.34	0.12	14	5	0.48	0.17
40	87	36	17	0.41	0.13	52	17	0.6	0.18
50	134	75	27	0.56	0.15	88	25	0.66	0.19
60	356	224	64	0.63	0.18	278	71	0.78	0.2
70	658	513	118	0.78	0.18	559	161	0.85	0.2
80	865	709	173	0.82	0.2	770	190	0.89	0.22
90	1139	1036	251	0.91	0.22	1050	342	0.96	0.3
100	1676	1626	536	0.97	0.32	1659	670	0.99	0.4

在上述实验基础上,取 $t = 80(s)$ 和 $t = 100(s)$ 时刻数据进行对比分析(初始抗体数 N_{R0} 分别为 709 和 1626),初始阈值 $r_0 = 8$,最大阈值 $r_m = 12$,采用初始阈值 $r_0 = 8$ 间隔迭次选取方式对自我集进行扫描,匹配阈值取 9(保证在自我集未改变时不会被检测器检测到).分别改变自我集一个字、片段和源文件,统计算法抗原数、疫苗数、初次应答时间 t_1 和

检测失败率 P_{f1} 、二次应答时间 t_2 和检测失败率 P_{f2} ,每种方式实验重复 100 次,取平均值作为最终结果.实验结果如表 2 所示. Matlab 仿真结果如图 4、图 5 所示.

实验结果表明,加入疫苗算子后,二次应答的时间明显减少,检测效率大大提高,检测失败率 P_f 也明显有所改善.

	初始抗体	实际抗体	抗 原	疫 苗	初次应答		二次应答	
	N_{R0}	N_R			t_1/ms	P_{f1}	t_2/ms	P_{f2}
改变一个字	709	24	7	3	11.80	0.163	1.68	0.121
改变一个字	1626	86	10	3	12.32	0.038	2.06	0.021
改变一个片段	709	48	10	5	12.95	0.112	2.31	0.098
改变一个片段	1626	93	13	6	13.40	0.152	2.86	0.066
改变源文件	709	58	9	7	14.65	0.035	4.01	0.006
改变源文件	1626	103	27	12	15.71	0.009	4.93	0.001

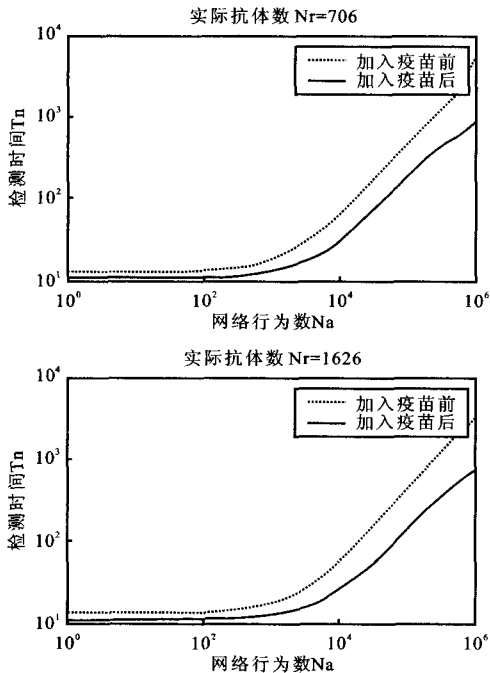


图 4 疫苗加入前后系统应答时间对比

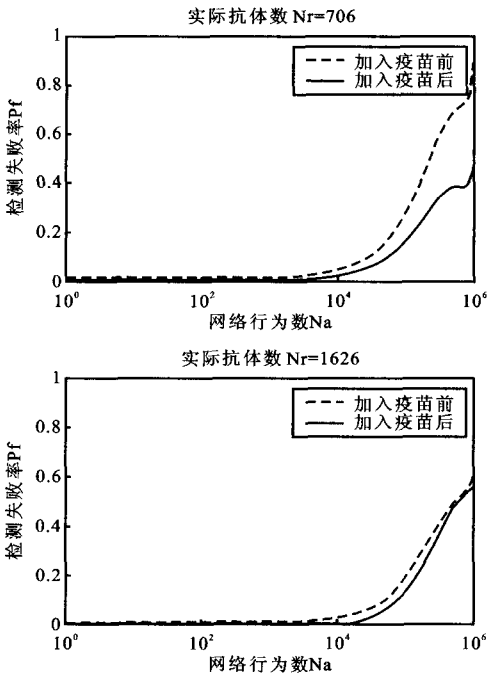


图 5 疫苗加入前后系统检测失败率对比

5 结束语

文中研究了高速网络环境下的免疫原理的应用,将网络分流技术与免疫算法结合,在对阴性选择算法的研究基础上,引入免疫算子和正选择算子,缩短系统二次免疫应答时间,同时在抗体库中添加疫苗,增加了抗体多样性和自适应性.最后通过仿真实验得出算法改进后在检测效率上得到大大提升,为

高速网络数据的处理提供了有力的保障.然而在实际运用中还会出现一些问题,比如网络分流存在的负载均衡问题,这里没有具体考虑.

参考文献:

[1] Xue Hua, Li Xueying, Chen Yu, et al. Research on the intrusion detection and its implementation through data streaming[J]. Computer Applications, 2004(1):112-114.

表产生出最终的产生式规则.如表4可得到值约简后的产生式规则如下:

R_1 : if $b = 0$ and $c = -1$ then $D = 0$

R_2 : if $a = 0$ and $b = -1$ then $D = -1$

R_3 : if $b = -1$ and $c = -1$ then $D = -1$

R_4 : if $c = 1$ then $D = 1$

令 R 表示产生式规则, $[x]_R$ 表示满足该规则条件属性的元组集合, D 表示满足该规则决策属性的元组集合, $\text{card}([x]_R)$ 表示元组集合中元组总数.令规则 R 的可信度为 Cf_k , 定义如下:

$$Cf_k = \frac{\text{card}([x]_R \cap D)}{\text{card}([x]_R)} \quad (10)$$

由于规则前件中每一个条件属性对得出该决策结论的贡献大小不同,也就是该前提属性对决策属性的依赖度不同,因此通过值约简后的决策表,重新计算没有删除前的条件属性对决策属性的依赖度 β , 通过式(9)计算出 β_{kj} ; 然后再将规则中简化的数值通过反模糊化处理,从而得到语言描述式的产生式知识规则.

至此已完成了规则的自动化抽取,从而得到可用的产生式规则^[7].

$$R_k : \text{AND}_{j=1}^m P_{kj}(\beta_{kj}) \rightarrow Q_k(Cf_k) \quad (11)$$

式中, P_{kj} 表示产生式规则的前提条件, Q_k 表示产生式规则的结论, k 表示新产生的规则序号, j 表示第 k 条产生式规则的第 j 个前提条件, m 表示该条规则的前提总数.

4 结束语

文中提出了基于粗集运算的产生式关联规则自

动获取知识的原理及过程.并提出了以多维代表值的求解以完成信息表的值约简;对获取知识过程中的关键步骤均给出了具体的算法.所获取的规则具有完整、准确、形式简洁、冗余度低,不必进一步化简等特点,知识获取的自动化程度高.系统可作为“知识自动获取机”嵌入到各种类型的专家系统中,具有较高的实用价值.

参考文献:

- [1] 韩敏,孙燕楠.地理信息知识获取 Rough—NN 模型研究[J].信息与控制,2005,34(1):104—108.
- [2] 刘同明.数据挖掘技术及其应用[M].北京:国防工业出版社,2001.
- [3] 王萍,王学峰.基于遗传算法的粗糙集属性约简算法[J].计算机应用与软件,2008,25(5):42—44.
- [4] 张文修,仇国芳.基于粗糙集的不确定决策[M].北京:清华大学出版社,2005:32—35.
- [5] 程玉胜,张佑生.粗集理论原型系统的集成与实现[J].微电子学与计算机,2006,23(3):119—121.
- [6] 谢孟军,黄国兴.基于 Rough Set 的规则自动抽取设计方案[J].计算机工程,2002,28(3):167—168.
- [7] 刘道华,原思聪.基于带区间值的产生式规则及其模糊推理机制研究[J].计算机应用与软件,2006,23(3):86—89.

作者简介:

陈功平 男,(1972—),硕士,讲师.研究方向为计算机网络、人工智能.

秦乐阳 男,(1982—),助理实验师.研究方向为计算机应用技术.

(上接第112页)

- [2] 高丹,彭新光.基于 DET 曲线的入侵检测评估方法[J].微电子学与计算机,2008(8):36—40.
- [3] Forrest S, Hofmeyr S, Somayia A. Computer immunology[J]. Communications of the ACM, 1997, 40(10): 88—96.
- [4] 余燕芳,陆军.基于改进遗传算法的服务器端负载均衡算法[J].微电子学与计算机,2007(7):41—45.
- [5] Hammad M, Conor R. A less destructive, context-aware crossover operator for GP[M]. Berlin, Heidelberg: Springer—Verlag, 2006:36—48.

作者简介:

盛云雷 男,(1985—),硕士研究生.研究方向为网络安全.

李永忠 男,(1961—),教授,硕士生导师.研究方向为计算机通信、信息安全、计算机应用.

荆春伟 男,(1984—),硕士研究生.研究方向为网络安全.