基于人工免疫的入侵检测系统的研究*

姚国祥，官全龙，冯伟伦
（暨南大学信息科学技术学院，广东 广州 510632）

摘 要： 探讨了入侵检测系统的发展现状，研究生物免疫系统的特点和基本理论。在分析传统的人工免疫系统及LISYS系统的基础上，提出了一种改进的基于人工免疫的入侵检测系统。该系统在研究现有检测器生成算法的基础上，提出了位变异的初始检测器生成算法，对检测攻击的变异更为有效。该系统引入生物免疫学的协同刺激机制，并用LRU算法取代随机淘汰策略。实验结果证明这些方法能降低误报率，保证检测器的检测效率。

关键词： 人工免疫；入侵检测；检测器；LISYS；位变异
中图分类号： TP393.08 **文献标识码：** A **文章编号：** 0529-6579（2008）06-0082-05

入侵检测系统（IDS）与人工免疫系统有很大程度的相似性，因此可将人工免疫原理与知识应用在入侵检测的技术中。免疫原理在计算机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、自适应性、自动应答和自我修复等特点，而这些特点都是生物免疫系统特有的。入侵检测技术综合免疫原理对不完备信息进行检测，具备检测异常现象的能力。

目前入侵检测技术的研究主要集中在大规模分布式和智能化检测这两个方向。在智能化的研究方面，主要有专家系统、神经网络、数据挖掘以及人工免疫等。其中，具有代表性的Homfmeyr和Forrest[1-2]等人认为基于免疫学的计算机系统应有如下组织原则：分布式、多层、多样化、自治、自适应、自安全、动态、相互识别、异常检测以及不完全匹配检测，他们提出了一个轻量级的入侵检测系统LISYS（Lightweight intrusion detection system）。Kim和Bently[3-4]小组提出了“分布式、自组织、轻量级”的入侵检测模型，该模型在物理结构上由主IDS及从IDS两部分组成；而逻辑结构是由三部分组成的层次结构：基因进化（gene revolution）、阴性选择（negative selection）和克隆选择（clonal selection）。在他们的模型中，考虑了阴性选择机制、克隆选择机制、记忆机制与基因库机制等，但未对分布式环境加以考虑。Dasgupta[5]小组提出一种基于移动agent的系统，agent可以在节点和监督网络情况的路由器周围巡游。

在充分研究了人工免疫的相关原理[6-9]及入侵检测的相关知识后，并且对传统的基于人工免疫的入侵检测模型进行研究分析后，提出了一个改进的基于人工免疫的入侵检测系统模型，提高入侵检测系统的检测率，降低误警率。

1 入侵检测模型设计

LISYS系统作为一个简单的免疫激励入侵检测系统原形，在一定程度上实现了对于免疫系统行为的模拟具备了一定的分布性、可扩展性和自适应性，但仍然存在不少缺陷：检测器是非移动的agent，当agent被激活时不能复制；非完全自治的，存在人工操作，需要添加一个实现自动反应系统；只分析TCP数据流，不少的攻击行为采用ICMP，UDP数据报对系统进行攻击，对这些攻击LISYS将束手无策，因而需要对LISYS的检测数据对象进行一定的扩充；存在异常检测系统的通病，即系统是运行于“正常状态远大于异常状态”的前提下的。

模型设计基于LISYS系统，对LISYS系统的算法进行改进。模型还设计了检测器的BV算法以及对于正常连接过滤的自我过滤器，图1是本系统的模型设计。

1.1 合格检测器

合格检测器（包括成熟检测器和记忆检测器）是IDS的核心组成部分。但是，目前对于基于人工免役的入侵检测的研究焦点主要集中在合格检测器上，产生了一些常见的合格检测器的生成算法，如阴性选择算法、线形算法和贪心算法，这些算法各有特点，各有优劣。而对于初始检测器的生成的讨

* 收稿日期：2008-09-04
基金项目：国家自然科学基金资助项目（60773083）；广东省科技计划资助项目（2006B15401002）
作者简介：姚国祥（1959年生），男，教授，博士生导师；通讯联系人：官全龙；E-mail：gql@jnu.edu.cn

论却比较少见。初始检测器对于生成合格检测器的质量有比较直接的影响，所以讨论初始检测器的生成对生成高质量的合格检测器具有重大意义。

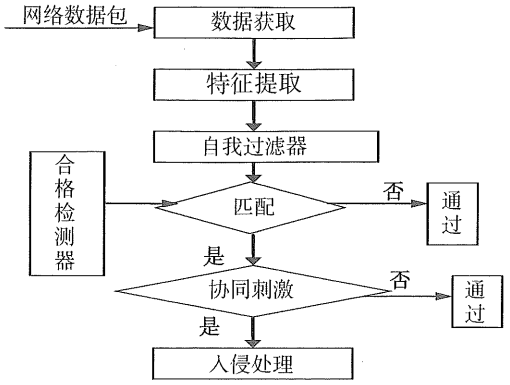

图 1　基于人工免疫的入侵检测模型

Fig. 1　Intrusion detection model based on artificial immunity

本文提出用一种生成初始检测器的算法，位变异 BV（Bit Variation）的算法。在对网络攻击的特征的大量研究上，发现很多攻击都是已有攻击的一种变异，就类似病毒的变种一样，这些攻击的变异可能只是对原攻击端口的简单修改，在基于规则的入侵检测中，对于这种攻击的变异是无法彻底解决的，因为变异就要求不断更新规则库，但是攻击的变体却可以是无限的，这样会令到规则的编写者疲于奔命。而基于异常的入侵检测正是在这种情况下产生的，但也存在问题。其中在基于人工免疫的入侵检测中，由于初始检测器都是随机生成的，因此可能检测字符串间没有太大关联，对于这种只是稍微改变了形式的攻击的变体，通常也会无能为力。

正是基于以上原因，便可以利用 BV 算法，可以针对这种比较多的变异的攻击进行有限抑制。下面说明 BV 算法的原理：

While（检测器数量 < 默认值）{

随机选择记忆检测器复制为新检测器；

对新检测器的字段进行随机变异（0 - >1；1 - >0）；

将新检测器放入初始检测器集合；

}

在进行完 BV 算法后，就可以进行否定选择，去掉与自我模式匹配的检测器。当然，最开始的时候，没有合格检测器，所以无法应用 BV 算法，还要依靠随机生成，而且我们也应看到，除了变异的攻击外，攻击还有很多种类，所以 LISYS 系统原有的随机生成初始检测器的算法不能丢弃，应该与 BV 算法共同使用，提高检测效率。

1.2　自我过滤器

从阴性选择算法的实质来看，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。而在通常情况下，网络中绝大部分数据都是正常数据，异常或入侵行为只占少数。另外，从阴性选择算法中可以看出，检测器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器才会成为成熟检测器，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器进行匹配检验是毫无意义的，因为这些自我模式根本就不可能与检测器集合中的某个检测集匹配。因此，让大量的正常模式与检测元进行匹配检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。

针对上述问题，本文提出统计出现频率高的正常模式自我集，设置一个过滤器将大部分正常模式预先过滤掉，就可大大减少系统的整体匹配比较次数，因为要将所有的正常模式在检测器之前完全过滤掉是不可能的，也是不合实际的，正常模式间的匹配的判断一般只能使用完全匹配规则，故要过滤所有的正常模式就需要整个自我集空间，这要远大于检测器集合。然后在由成熟检测器对剩下的模式进行匹配，从而达到提升系统效率的目的。

2　关键技术

2.1　检测器生成算法的改进

根据阴性选择算法的思想，基于免疫机制的 IDS 是借助阴性选择过程产生的检测器与非自我模式的匹配来达到对非自我模式的识别。检测器的检测能力表现在它对非自我模式空间的覆盖范围上。检测器对非自我模式的覆盖空间越大，则检测器识别或检测的非自我模式越多，即检测器的检测能力越大。由于产生过程的随机性以及 r—连续位匹配规则固有的特性，使得随机产生的检测器可能存在互相匹配的现象，从而导致检测器的覆盖空间无法达到最大化。

对于给定字符串长度 L 和匹配参数 r 之后，带有固定匹配率的 r 连续对应位匹配算法不可避免的会出现这样的检测黑洞。对于固定不变的自我集合，不同“形状”的检测器可以覆盖不同的非自我集合，产生不同的检测黑洞，如果能将不同“形状”的检测器共同作用，那么必然可以减少黑洞的存在。

因此对检测器生成算法进行了改进，改进的算法如图 2 所示。对于阴性选择匹配规则，除了 r 连续位匹配函数外，也可以用“海明距离”（Hamming Distance）来描述两个字符串的相似程度。对于字符集 m 上长度为 L 的位串 x 和 y，用海明距离

计算两个位串不同取值的位数，取补就得到 Hamming 相似性度量。在负选择过程中，对于不断随机产生的字符串，让入侵检测系统随机选择用 r—连续位匹配规则或海明匹配规则来产生各种不同形式的检测器。让这些检测器形成检测器集合，共同作用于非自我空间，进而减少黑洞的存在。

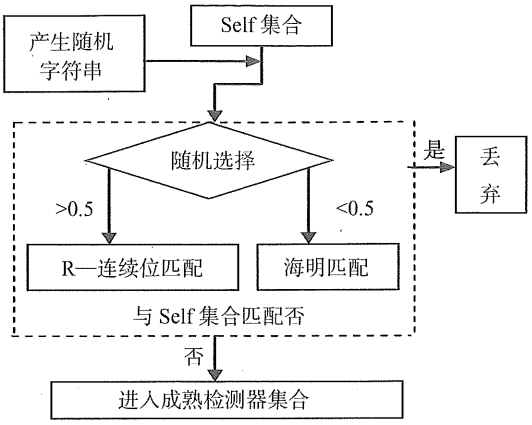

图 2 改进阴性检测器生成算法

Fig. 2 Generating algorithm of improved cathode detector

2.2 LRU 算法的引入

在传统的基于免疫的入侵检测系统中，对记忆检测器的处理机制，多采用的是长期器长期存在，最终将使记忆检测器集合达到饱和状态，从而导致系统误报率提高和不能检测到新的入侵。而在 LISYS 中，对记忆检测器引入了淘汰机制，但因为用的是随机淘汰的策略，则很容易导致一个问题：随机删除的检测器有可能是亲和力系数比较高，并且是刚刚升级为记忆检测器的成熟检测器，这样则会导致检测效率不高，当有类似入侵发生时，不能即时检测出来。

本文引入“最久未使用”算法（LRU）的机制，即淘汰最久未被激活的记忆检测器，这个淘汰策略需要一个记录记忆检测器最近一次被激活的时间属性，在记忆检测器检测收集到的数据时，若出现绑定抗原的时候，由该属性记录时间。

2.3 协同刺激机制的改进

在机体免疫系统中，B 细胞激活除了需要满足激活阀值的要求外，还必须得到 T 细胞发出的信号。协同刺激就是模仿这种双重条件。当一个检测字符串检测到异常时，必须有另外的检测字符串也认为是异常，系统才认为确实发生了入侵。传统的基于免疫的检测模型中，一个检测器只包含一个检测字符串，当检测器发现异常时，必须与其它检测器通信，才能确定是否有其它检测器也与其匹配。

在 LISYS 系统中，一个检测器包含一组检测字符串，但是 LISYS 采用的协同机制是通过人工响应。这种机制虽然能够在一定程度上避免在短期内由于错误地发送激励信号导致的误报现象，但同时，如果系统中真的出现了入侵行为，也无疑会导致系统没有及时报警对系统造成破坏。

本系统模型中改进了这种协同刺激的机制。如果检测器内有两个以上检测字符串对其匹配成功，就可以确认为异常，如果检测器内只有一个检测字符串能够匹配，那么再按照 LISYS 系统的响应方式，交给系统管理员来发出响应判断是否为入侵攻击，再作出响应。利用这种方式，对于可能性比较大的入侵行为，可以减少响应时间，因为不需要系统管理员给出信号，系统就会自动判断出是攻击。这种协同刺激方式可以大大降低网络通信的开销，提高速度，其准确率也更高。当然，也有可能会产生误报。

协同刺激可以有效降低误报的出现，但是也会增加漏报的几率，如何在两者之间取得平衡，也是今后研究的方向之一。

3 模型实验测试

为了测试模型实现程序对 LISYS 原有算法进行改进后的效果，程序没有实时的捕抓数据包进行检测分析，而是将捕获的数据包保存下来，由改进前和改进后的程序进行对比分析，从而能够比较明显的得出算法改进的结果。

为了比较算法的改进是否有效，用两个指标衡量算法优劣：程序的运行时间和检测出异常连接的数量。

图 3 所示的结果是没有对 LISYS 原有算法进行修改的程序的运行结果。可以看到，程序运行共用了 13 s 左右，而并没有检测到有异常连接。

图 4 是对检测算法以及记忆检测器的淘汰机制进行改进了后的程序结果。由程序的前后结果对比可以看出，因为在检测器生成算法中，多了对匹配规则的选择，并且在对记忆检测器淘汰过程中，也加入了激活时间的判断，所以程序在改进后的执行时间明显增长，但是检测出的异常连接也增多，这是因为对原由算法进行改进，使得检测器的覆盖空间变大，对于一些原来不认为是异常的字符串，和原来一些不能检测出的字符串，由于检测器覆盖空间变大后，检测范围也变大，所以也会判断为异常的，可以有效覆盖检测黑洞。当然，上面检测到的异常其实不属于误报，这是为了在有限的时间明显看到效果，而将检测器的耐受期缩短，所以初始检测器就成为成熟检测器，从而检测到异常。

但是，我们也应该看到，这样的检测到的异常的集合会变大，实际上误报率也有增加，而且检测时间太长，效率低下，所以引入了自我过滤器。

图 5 所示在加入了自我过滤器后，我们可以明显看到检测时间比原来减少，提高了效率，而且检测出的异常连接也大幅度减少，这是因为很多的正

```
 连接信息：202.116.9 16:11:01.958041 IP 202.116.9.92.1946 > 210.51.186.72.80:
[01111100.00010100.11000110.10001001.00010110.11001011.1]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
[10111101.11110000.00001101.00101001.00010100.10001000.1]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
 连接信息：202.116.9 16:11:02.107132 IP 210.51.186.72.80 > 202.116.9.92.1946:
[00001001.01010100.01001100.00010110.11010100.00100111.0]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
[11010001.00001100.01111000.11111010.10001110.00000010.1]: died of old age
detection.DetectionNode@69b332：检测器死亡
[01011011.10110100.01001010.10000111.00100000.01101001.1]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
 执行检测共耗时 13438 ms
 共有 4530 个连接，其中有  0 个异常连接！
 生成检测器  100 个
 其中记忆检测器  50 个
 其中成熟检测器  99 个
```

图 3　改进前程序结果

Fig. 3　Program result before improvement

```
 连接信息：202.116.9 16:11:01.958041 IP 202.116.9.92.1946 > 210.51.186.72.80:
 连接信息：202.116.9 16:11:02.107132 IP 210.51.186.72.80 > 202.116.9.92.1946:
[01010001.01000010.00000100.11011101.01000111.11010011.1]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
 执行检测共耗时 17234 毫秒
 共有 4530 个连接，其中有  908 个异常连接！
 生成检测器  100 个
 其中记忆检测器  50 个
 其中成熟检测器  97 个
```

图 4　改进后程序结果一

Fig. 4　The NO. 1 program result of improvement

```
 连接信息：202.116.9 16:11:01.958041 IP 202.116.9.92.1946 > 210.51.186.72.80:
[01010100.00011001.10110000.10100001.01000001.10111000.1]：死亡于未成熟的匹配
detection.DetectionNode@69b332：检测器死亡
[00101001.01001110.11011010.11110100.01001101.00010111.0]：死亡于未成熟的匹配
detection.DetectionNode@69b332：检测器死亡
[10000101.00000011.00101010.11101000.00000110.11101011.1]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
 连接信息：202.116.9 16:11:02.107132 IP 210.51.186.72.80 > 202.116.9.92.1946:
[01110101.11001110.10111000.00100101.00010100.01100111.1]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
[01001000.01101100.00101100.11111000.01110110.00100100.0]:  检测器被激活
detection.DetectionNode@69b332：成熟检测器成功匹配
[10101011.01001101.10101100.11111001.11010111.11000001.1]：死亡于未成熟的匹配
detection.DetectionNode@69b332：检测器死亡
 执行检测共耗时 13734 毫秒
 共有 4530 个连接，其中有  383 个异常连接！
 生成检测器  100 个
 其中记忆检测器  43 个
 其中成熟检测器  97 个
```

图 5　改进后程序结果二

Fig. 5　The NO. 2 program result of improvement

常模式已经进入了自我过滤器，不用再进行检测，当然也就不会产生误报了，证明自我过滤器有效。

4 总 结

基于人工免疫原理的入侵检测模型和系统是近年来生物领域中研究的重点和难点，它的研究对于解决当前网络安全所面临的严重危机具有十分重要的意义。本文根据人工免疫基本原理建立入侵检测模型，该模型提高入侵检测的检测效率、降低误报率与减少人工干预做了仔细地研究，下一步将对人工免疫各算法进行并行化研究，探索更有效的检测匹配算法。

参考文献：

[1] HOFMEYR S FORREST S. Architecture for an artificial immune system [J]. Evolutionary Computation, 1999, 7 (1): 1289 - 1296.

[2] FORREST S, HOFMEYR S A, Immunology as information processing[C]//SEGEL L A, COHEN I R. Design Principles for the Immune System and Other Distributed Autonomous Systems [C]. USA: Oxford University Press, 2000.

[3] 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.

[4] KIM J, BENTLEY P. Towards an artificial immune system for network intrusion detection: An investigation of clonal selection with a negative selection operator [C]// Proc Congress on Evolutionary Computation. Korea: Seoul, 2001: 27 - 30.

[5] DASGUPTA D. Artificial Immune Systems and Their Applications[M]. Berlin: Springer-Verlag, 1999.

[6] 肖人彬；王磊. 人工免疫系统：原理、模型、分析及展望[J]. 计算机学报，2002，25(12)：1281 - 1293.
XIAO Renbin, WANG Lei. Artificial immune system: Principle, models, analysis and perspectives [J]. Chinese Journal of Computers, 2002, 25(12): 1281 - 1293.

[7] 葛红. 免疫算法综述[J]. 华南师范大学学报：自然科学版，2002(03)：120 - 126.
GE Hong. Immune algorithm: A survey [J]. Journal of South China Normal University: Natural Science, 2002 (03): 120 - 126.

[8] 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 (29)：27 - 29.
YANG Xiangrong, SHEN Junyi, LUO Hao. Application of artificial immune theory in network-based IDS[J]. Computer Engineering, 2003(29): 27 - 29.

[9] 葛丽娜，钟诚. 基于人工免疫入侵检测检测器生成算法[J]. 计算机工程与应用，2005(23)：149 - 152.
GE Lina, ZHONG Cheng. The algorithm of detectors generating in intrusion detection system based on artificial immure theory [J]. Computer Engineering and Applications, 2005(23): 149 - 152.

Research and Improvement of the Intrusion Detection System Based on Artificial Immunity

YAO Guo-xiang, GUAN Quan-long, FENG Wei-lun

(Information Science Technology College of Jinan University, Guangzhou 510632, China)

Abstract: The developing trend of intrusion detection technology is explained and the characteristic and basic theory of biological immune system is researched in this paper. An improved intrusion detection system based on artificial immune are proposed after Analyzing traditional artificial immune system and LISYS. Researching the generating algorithm of existing detectors, the improved system proposes initial detector's generating algorithm of bit variation. It makes more efficient for detect the variation of attacks. The co-stimulation mechanism of biological immunology is adopted and the strategy of random death is replaced by LRU algorithm in the system. The experiment shows that these methods can decrease the false positive rate and improve detection efficiency of detector。

Key words: artificial immune; intrusion detection; detector; LISYS; bit variation