第32卷 第20期 计 算 机 工 程 2006年10月
Vol.32 № 20 Computer Engineering October 2006
·人工智能及识别技术· 文章编号：1000—3428(2006)20—0188—03 文献标识码：A 中图分类号：TP393.08

基于免疫的入侵检测系统研究

肖 毅[1]，胡伟雄[1]，肖 明[2]，赵 慧[3]
（1. 华中师范大学信息管理系，武汉 430079；2. 华中师范大学网络中心，武汉 430079；3. 华中师范大学数学与统计学院，武汉 430079）

摘 要：将生物免疫机制引入入侵检测系统，设计了一个基于免疫代理的入侵检测系统。系统探测和响应采用层次结构，各Agent既相互独立又相互协作，游走于各网络节点间，检测分布式的攻击。该文介绍了免疫算法的寻优机理，抗体扩增和抑制、记忆单元更新、亲和度和浓度计算等关键技术，确保了抗体的多样性，改善了未成熟性的收敛。

关键词：免疫代理；进化；入侵检测系统

Study on Intrusion Detection System Based on Immune Agent

XIAO Yi[1], HU Weixiong[1], XIAO Ming[2], ZHAO Hui[3]
(1. Department of Information Management, Central China Normal University, Wuhan 430079; 2. Network Center, Central China Normal University, Wuhan 430079; 3. College of Mathematics and Statistics, Central China Normal University, Wuhan 430079)

【Abstract】Inspired by the biological immune mechanism, this paper designs an intrusion detection model based on immune agents, which uses the level construction of sensor and response scheme. The immunity-based agents are independent and collaborate each other and roames among network nodes to detect the distributed intrusions. It introduces the artificial immunization algorithm to seek the superior mechanism, the antibodies promotion and suppression technology, the update of memory, the affinity and the density. The experiment shows that this system guarantes the antibodies multiplicity, improves premature convergence.

【Key words】Immune agent; Evolution; Intrusion detection system

将生物免疫原理引入计算机安全领域，通过模拟生物免疫系统来构建计算安全免疫系统已成为一个新的研究热点[1,2]。

入侵检测系统(Intrusion Detection System, IDS)作为系统防御的重要手段，它在计算机安全系统中的作用与免疫系统在生物体中的作用非常类似[3]。入侵检测系统的作用在于检测并阻止系统内外部非法用户的攻击，而免疫系统的作用在于保护生物体免受外部病原体(如病毒、细菌等)的攻击。二者的行为本质上可以归结为对危险“非我”的识别和清除，这种相似性为借鉴免疫机制研究入侵检测系统提供了一个新的思路。

智能Agent所具有的自治、适应、学习、移动等特性为入侵检测免疫系统的实现提供了技术支持。本文以此为基础提出了一个基于免疫学原理的网络入侵检测系统模型。免疫机制和Agent的引入使得本系统具有分布式、多样、适应和学习等特性。

1 自然免疫系统

人体免疫系统[4]是一个复杂的动态系统，主要由免疫器官、免疫组织及多种免疫细胞组成，通过遍布全身的免疫细胞(即抗体)来抵御外部病原体(即抗原)的入侵。

抗体(包括B细胞和T细胞)通过随机选取胸腺中候选基因库里的基因片段组合产生，经过负选择(negative selection)过程后成为成熟抗体，负选择的作用在于防止发生自免疫。成熟抗体从胸腺释放出来后在各个淋巴器官间游走于机体执行抗原检测任务。如果某种抗体识别了抗原，便发生初次应答反应(激活)，T细胞和B细胞对病原体的识别是通过结合病原体表面的抗原及淋巴细胞表面的抗原识别受体来实现的。当抗原与抗体结合后，发生一系列复杂的反应，导致抗原被巨噬细胞所消灭。随后抗体进入克隆选择(Clonal Selection)阶段，那些和抗原亲和度高的抗体通过自我克隆扩增产生多个自身的复制，称为记忆细胞。这些记忆细胞可以加速对该抗原的识别和消除，另外当该抗原或变种再次入侵时，免疫系统能据此产生更快速有效的两次应答。

自然免疫系统具有的分布性、多样性、自适应性、动态性为构建入侵检测免疫系统提供了多种组织形式。

2 入侵检测免疫系统设计原则

人体免疫系统所表现出的良好特性正是入侵检测系统所期望的。下面将利用人体免疫系统中一些思想来指导入侵检测系统的设计，主要体现在：

(1)分布性：免疫细胞遍布全身，各自并行工作自主对抗原入侵作出响应，彼此之间大多独立，不需要集中控制。这种方式将载荷分散到各个单元有利于提供系统效率，同时局部免疫功能缺陷不会导致整个机体的免疫崩溃，保证了系统的健壮性。网络攻击同样具有分布式特点，这为我们设计入侵检测系统提供了有益的启示。

(2)多样性：对个体而言，免疫细胞功能各异，能够消除不同种类的抗原；同种细胞能力有别，可以防止个体脆弱性。对种群而言，个体的免疫机制相同但免疫能力不同，个体免疫系统抗体集合具有唯一性、相异性，当有个体对某抗原出现免疫缺陷不至于威胁整个种群的安全。

(3)自我应答和记忆性：免疫系统具有识别新抗原并通过

作者简介：肖 毅(1975－)，男，讲师，主研方向：智能信息处理，信息安全；胡伟雄，副教授；肖 明，博士生；赵 慧，博士后
收稿日期：2006-01-23 **E-mail：**yxiao@mail.ccnu.edu.cn

记忆细胞快速复制的能力。入侵检测中借鉴这种近似匹配和记忆机制，在以后遇到相似攻击时能够快速作出 2 次响应。

由此可见，借鉴自然免疫系统的分布性、多样性和自适应性的特性为构建入侵检测免疫系统的设计提供了现实依据。

3 基于免疫代理的入侵检测系统

3.1 免疫代理工作机制

免疫代理作为入侵检测的基本组件分布于网络各节点，模拟免疫细胞执行入侵检测任务，其实现机制如图 1 所示。

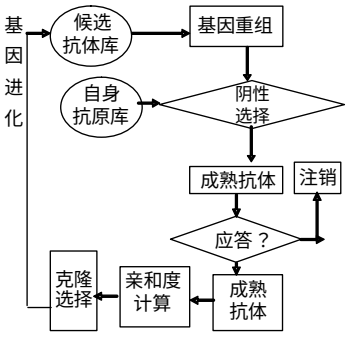

图 1 免疫代理工作机制

基于 Agent 技术的免疫代理是建立在体细胞理论和免疫网络理论的基础之上的。算法过程如下：

(1)系统从候选抗体库中随机抽取一组基因，通过交叉、变异产生新生免疫代理，为了防止因自免疫而产生虚警，即将正常的网络行为误报为异常，因此需要将其与特征提取过程所产生的自身行为或网络特征进行比较，所有匹配自身特征模式的免疫代理必须丢弃，这就是模拟生物免疫的负选择过程。经过负选择过程后的成熟免疫代理被分发到各检测主机上，根据各自规则自主识别异常数据完成实时的检测任务。

(2)一个新的免疫代理如果匹配到某种网络异常，便发生应答，及时切断这些异常数据包的网络连接，并保存异常特征(基因)到候选规则库以供其它检测子共享。它自身也成为一个记忆免疫代理，通过与同类型免疫代理比较亲和度来竞争扩增机率，亲和度越大则复制机率越大，进而又会增加其亲和度，形成正向激励。记忆免疫代理通过扩增复制自身到其它节点使其具有快速检测能力。

(3)为了降低网络负荷，免疫代理必须根据检测异常数据的能力(抗原结合亲和度)及有效时间(生命周期)决定何时注销。通过模拟生物界优胜劣汰机制可以实现候选基因库的进化，提高系统性能，由于只需保存有限基因从而降低了系统负荷。通过设置合适的基因库容量和生存期可以保证那些低亲和度和未达到激活阈值的免疫代理也有足够的进化时间。

3.2 基于多免疫代理的系统构建

传统的 BDI 模型简单地把系统结构分成信念(Belief)、目标(Desire)、意图(Intension) 3 个部分。

在具体实现中这种方式只适用于功能简单的系统，而免疫系统所具有的分布式和多样性难以在一个 Agent 中实现，采用多 Agent 是必然的选择。对于多 Agent 系统而言，必须考虑 Agent 之间的通信与合作问题，因此将系统按功能划分层次结构，分别由不同的 Agent 完成，通过 Agent 间的协作实现系统功能。该结构具有较强的层次性和高度模块化，有助于发挥多 Agent 系统的优势。

分布式 IDS 体系结构如图 2 所示，它由决策代理(Decision Agent，DA)、检测代理(Intrusion Agent，IA)、巨噬细胞代理(Macrophage Agent，MA) 3 个部分组成。

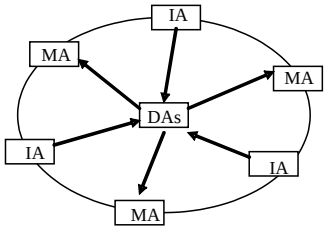

图 2 多 Agent 系统框架

(1)检测代理游走于网络节点之间负责数据信息、过滤数据和传送消息给决策代理。对基于主机的检测，采用长度为 L 的滑动窗口序列分析算法所获取的系统调用短序列可以构造一个完全的自我集，这些自我模式被证明是稳定的正常行为特征，当异常发生时，该特征会发生变化；对网络数据可通过 IP 包的特征属性(如协议类型、源 IP 地址、目的 IP 地址、连接时间、端口、包的数量、连接终止状态、特定错误等)来分析和处理。本文中用一个 61bits 来表示一个自我，其中前 8 位表示局域广播网的内部主机，第 9 位~第 40 位表示外部计算机或内部客户机，第 41 位表示 IP 包方向，第 42 位~第 49 位为端口号，最后 12 位表示状态。每一特征属性表示一种基因，各基因段的编码形式依不同的服务类型决定。

(2)决策代理处理所收集的数据。核心功能是区分“自我”和“非我”，即正常数据和异常数据。决策是通过模式匹配实现的，可以记作$M=\{M_i\}$,每个M_i就是一个模式(包括一个条件c_i 和一个动作a_i)，记为$M_i=[c_i|a_i]$。基于主机和网络的异常检测的复杂性，必须对上述定义进行扩充，即M_i由一系列条件 $\{c_i1,c_i2,\ldots,c_in\}$ 和动作 $\{a_i1,a_i2,\ldots,a_in\}$ 组成，记为 $M_i=\{c_i1,c_i2,\ldots,c_in|a_i1,a_i2,\ldots,a_in\}$。

(3)吞噬细胞代理模拟免疫系统吞噬细胞的功能，杀死抗原。记为$MA_i=[m_i|k_i]$,其中m_i表示第i个异常，k_i为对第i个异常的反应。一旦决策代理发生异常匹配，立即激活相应MA采取行动。对于主机系统表现为注销一个异常进程；对于网络则表现为切断异常数据包的网络连接。

3.3 Agent 间通信

Agent 通信是 Agent 相互协作共同完成任务的基础，Agent 间的通信与协作机制的设计对于整个系统的设计而言有着非常重要的作用，直接影响着整个系统的效率、健壮性和兼容性。

本文将整个多 Agent 系统通信从逻辑上划分成 3 个层次，自顶向下依次为：

(1)传输层。通信数据在物理线路上传输的协议和有关机制。这里采用了点对点方式，它是建立在 TCP/IP 协议基础之上的，有很好的网络兼容性。考虑到在系统运行过程中,可能会因为各个 Agent 间有大量的消息相互传递而导致系统崩溃,为了保证系统的正常运行，必须对各个 Agent 间的通信进行规划，以保障一些主要的 Agent 间的消息传递能正常进行。这种规划可以通过给各个 Agent 设置不同的优先级来实现。本系统中 DA 和 IA 间通信最频繁，可以将它们之间的通信优先级设为最高。本系统的通信机制采用 WinSock 实现的。WinSock 是一种基于 Windows 的网络通讯应用程序标准。它是一组封装好的对象，使得复杂的硬件实现细节对用户透明。

(2)表示层。Agent 将通信内容或动作传输给会话层的一种媒介或工具，即通信语言。Agent 可以就某特定问题提出请求、查询、作出响应等。目前国际上流行的 Agent 通信语

言有 KQML 和 FIPA-ACL2 种，它们都是基于言语动作理论(Speech-Act)，通过 Agent 发送消息完成某一特定目的或动作。它们能够较好地支持 Agent 协作和知识共享，但目前仍缺乏有关的标准化规范。本文结合具体应用，设计了一个基于 KQML 的扩展协议。

(3)传输层。将 Agent 间发送的信息分为 3 个部分：信息头，动作和行为，其中常用的动作如表 1 所示。最后，会话层用来实现多 Agent 间会话过程的相关规则和策略。

表 1 常用动作表

Ask 请求命令	Ack 应答命令
Inform 通知命令	Retrieve 检索命令
Order 执行命令	Export 结果命令

4 检测代理的生成和进化

检测代理是整个入侵检测系统的核心部件，其对未知异常的识别和处理能力直接决定了系统的质量。

4.1 检测代理的生成

在主机模式下，虽然每台主机提供的服务不尽相同，但却是确定的，因此自身行为模式也是确定的，可以采用基于肯定的误用检测方式。而网络的复杂性使得从训练样本生成的候选基因不完全，如何扩大非自我集的范围，保证抗体的多样性对于检测未知攻击至关重要。本文采用Holland J H遗传算法(Genetic Algorithm, GA)来模拟自然选择和遗传变异机制，进而实现候选基因的进化[5]。

遗传算法通过选择(Reproduction)、交叉(Cross－over)和变异(Mutation) 3 种算子实现一种全局概率优化搜索算法。

不同抗体的繁殖机率是根据其与抗原的亲和度来确定的，亲和度高的抗体竞争到更大的繁殖机会。抗体 v 的选择概率可以表示为

$$p_r = \frac{ax_v}{\sum_{w=1}^{N} ax_w}$$

交叉指将 2 个候选个体按某一概率 Pc 从某一位置起互换，形成 2 个新的后代，设 2 个个体为

$$X_p = (X_0^p,...,X_{a-1}^p,X_a^p,...,X_{n-1}^p)$$

$$X_q = (X_0^q,...,X_{a-1}^q,X_a^q,...,X_{n-1}^q)$$

其中，交叉点的位置为 a（0≤a≤n-1），X_p'，X_q'为交叉产生的后代。则单点交叉算子 $\otimes$ 表示为

$$\otimes(X_p',X_q') = \begin{cases} X_0^p,...,X_{a-1}^p,X_a^p,X_{a+1}^q,...,X_{n-1}^q \\ X_0^q,...,X_{a-1}^q,X_a^q,X_{a+1}^p,...,X_{n-1}^p \end{cases}$$

通过选择更多的交叉点并交替交换交叉点之间的位段可以将单点交叉算子扩展到多点交叉算子。

变异指对某一个体中任一位按某一概率 Pm 进行取反，它只在一个个体中发生，变异算子定义如下：

$$\Theta(X) = \Theta(X_0,...,X_a,...,X_b,...,X_{n-1}) = (X_0,...,\neg X_a,...,\neg X_b,X_{n-1})$$

个体$(X_1,X_2,X_3,...,X_n)$经变异后，可表示为$(X_1'X_2'X_3',...,X_n')$。

其中，$X_i' = \begin{cases} X_i, \theta_i > p_m \\ 1-X_i, \theta_i \le p_m, \forall i \in \{1,...,n\} \end{cases}$，$\theta_i$是（0，n）之间的一致随机数。

4.2 检测代理的进化

检测代理的进化能力直接决定着系统对于未知入侵的检测能力。假设免疫系统由具有 *M* 个基因的 *N* 个抗体组成，则第 *j* 个基因的信息熵 Hj(N)为

$$H_j(N) = \sum_{i=1}^{N} p_{i,j} \lg p_{i,j}$$

抗体群差异的平均信息熵为

$$H(N) = \frac{1}{K}\sum_{j=1}^{K} H_j(N)$$

在此基础上可以定义 2 个抗体 v，w 之间的亲和度为

$$ay_{v\,w} = 1/(1+H(2))$$

其中，*H(2)*表示抗体v和w的信息熵，$ay_{v,w}$介于(0，1)之间，$ay_{v,w}$越大，表示v和w2 个抗体越亲和或越类似。

抗体 v 的浓度 Cv 表示群体中相似抗体所占的比例为

$$C_v = \frac{1}{N}\sum_{w=1}^{N} ac_{v,w}$$

其中，$ac_{v,w} = \begin{cases} 1 & ay_{v,w} \ge T \\ 0 \end{cases}$，$ay_{v,w}$为抗体当v和w的亲合度，若$ay_{v,w}$>T则表示抗体v和w相似，其中T为设定的阈值。

抗体 v 与抗原的亲和度可定义为

$$ax_v = 1/(1+opt_v)$$

其中，opt_v表示抗体v和抗原之间的差异度，介于(0，1)之间。当opt_v＝0 时，ax_v＝1，说明抗体v和抗原非常匹配，即该抗体为最优解。当抗原与抗体x的连接强度opt_v达到最小值 0，则完全匹配。实际中选择亲和力高于T的抗体进入激活状态。

抗体 v 的预期扩增率可以表示为

$$e_v = ax_v / c_v$$

其中，ax_v为抗体v的亲和度。

浓度与亲和度决定了抗体的扩增与抑制水平，抗体的亲和度越大，则扩增率越高； 抗体浓度越大，则扩增率越小。这样既可保留亲和度高的抗体，又可确保抗体的多样性，改善未成熟收敛。

5 结束语

除了算法的选择，在实际应用中，还要解决如群体规模限制，交叉、变异概率选择等问题，这些参数的组合对于防止未成熟收敛，提高系统优化质量具有重要意义。本文设群体大小 *N*=120，进化代数 *T*=20，交叉概率 *P*c=0.6，变异概率 *P*m=0.005，经过 20 代进化，模拟显示该方法可以得到较理想的优化个体。

本文借助 Agent 技术通过模拟生物免疫机制实现入侵检测。该模型采用分布式结构，能够很好地满足入侵检测系统对自适应性、多样性和动态性的要求。尚需改进之处：如何合理地控制群体间的相互作用，来模拟并行执行过程，如何利用粗糙集理论进行数据约简以提高系统效率以及如何提高匹配算法的模糊识别能力等。

参考文献

1 Forrest S, Perelson A, Allen L. Self-nonself Discrimination in a Computer[C]. Proceedings of the IEEE Symposium on Research in Security and Privacy. Los Alamitos, CA: IEEE Computer Society Press, 1994: 202-212.

2 Forrest S, Hofmeyr S A , Somayaji A. Computer Immunology[J]. Communications of the ACM, 1997, 40(10): 88-96.

3 Dipankar D. Immunity-based Intrusion Detection Systems: A General Framework[C]. Proceedings of the 22th National Information Systems Security Conference, 1999-10: 18-21.

4 于善谦，王洪海，朱乃硕等. 免疫学导论[M]. 北京: 高等教育出版社, 1999.

5 Haifeng D, Maoguo G, Licheng J, et al. A Novel Algorithm of Artificial Immune System for High-dimensional Function Numerical Optimization[J]. Progress in Natural Science, 2005, 15(5).