doi:10.3772/j.issn.1002-0470.2012.01.004

基于免疫优化原理的网络安全态势预测方法①

石元泉②[*,**] 刘晓洁[*] 李 涛[*] 彭小宁[**] 陈 文[*] 张瑞瑞[*]

([*]四川大学计算机学院 成都 610065)

([**]怀化学院计算机系 怀化 418000)

摘 要 为有效监视网络安全态势变化和预防网络遭受大规模安全攻击,受人工免疫系统启发,提出了一种基于免疫优化原理的网络安全态势预测方法(NSSPAI)。该预测方法首先给出网络安全态势预测环境下抗原、抗体和亲和力的定义,以及用于挖掘态势预测模型的抗体优化算子的抽象数学模型;然后利用相空间重构理论分析网络安全态势时间序列,利用重构后的样本空间和免疫优化建模方法挖掘态势预测模型;最后利用该预测模型来预测未来的网络安全态势。实验结果表明,与基于遗传算法的预测方法相比,NSSPAI预测方法能够更为准确地预测网络安全态势,是一种有效预测网络安全态势的新方法。

关键词 网络安全,态势感知,时间序列,免疫优化,预测

0 引 言

随着计算机网络的迅速发展和人们对网络应用的极度依赖,网络安全问题已经成为人们日益关注的焦点。网络攻击、黑客入侵和病毒感染等大量安全事件使计算机网络面临着拒绝访问、网络崩溃、重要信息被窃取或破坏等一系列安全性问题,传统的单一防御或检测系统已无法预防这些安全问题的发生,已不能适应网络安全需求。近年来,许多研究人员已将网络安全态势感知作为网络安全领域中的一项重要研究内容,旨在通过对复杂网络环境进行监测,对能够引起网络安全态势发生变化的各种安全要素进行获取、评估和预测来保证网络的安全运行。本文针对目前网络态势预测研究的不足,借鉴人工免疫系统解决非线性问题的理论,提出了一种新颖的基于免疫优化原理的网络安全态势预测方法(network security situation prediction approach based on immune optimization theory,NSSPAI),并对这一方法进行了较详尽的论述,同时也介绍了为验证该方法对网络安全态势的预测性能而进行的仿真实验的结果。

1 相关研究

目前,有关网络安全态势感知的研究,国内外尚处于未成熟阶段,主要集中于网络安全态势框架的理论研究。通常,国内外研究人员根据不同的网络应用环境和特定的安全需求,提出并设计不同的网络安全态势感知模型。1999 年,Bass 等在文献[1]中首次提出了网络安全态势感知概念,并且在文献[2]中提出了网络安全态势感知概念模型。Stephen Lau[3]等人利用三维空间中的点来表示网络流量,开发了三维网络流量检测(the spinning cube of potential doom)系统,提高了网络安全态势的感知能力。卡内基梅隆大学[4]开发了互联网报文流量分析系统(system for internet-level knowledge,SILK)。该系统可以提供整个网络的安全态势感知,适应对大规模网络进行安全分析。在国内,陈秀真等[5]提出了基于统计分析的层次化安全态势量化评估模型,李涛[6]提出了基于免疫的网络安全风险检测模型,韦勇等[7]提出了基于日志审计与性能修正算法的网络安全态势评估模型等。

尽管上述各种模型为网络安全态势感知研究提供了良好的理论基础,但大多数模型只考虑态势要

① 国家自然科学基金(60873246),教育部博士点基金(20070610032)和教育部重大项目培育基金(708075)资助项目。

② 男,1977 年生,博士;研究方向:网络安全,人工免疫系统和智能计算;联系人,E-mail:syuanquan@163.com

(收稿日期:2010-05-14)

素的获取和评估这二个方面,而对态势预测方面的研究相对较少,而且一般采用传统的预测方法。由于网络系统的复杂性以及安全因素的多样性,网络安全态势预测通常被认为是一种非线性时间序列预测[8]。因此,利用传统的预测方法不能有效地揭示网络安全态势的变化规律,所得到的态势预测值与目标值存在较大的偏离。

由于人工智能方法对非线性时间序列具有很强的逼近和拟合能力,许多研究人员利用人工智能方法来解决非线性时间序列预测问题,如遗传算法[9]、神经网络[10]和支持向量机[11]等智能预测方法。但是遗传算法的进化学习机制较为简单,神经网络存在易陷入局部优化、隐层数难以确定等缺陷,而支持向量机的算法性能易受惩罚参数、不敏感损失参数等关键参数的影响。与上述智能方法相比,人工免疫系统具有全局优化、收敛速度快等优点[12]。由于人工免疫系统继承了生物免疫系统的自学习、自适应、自组织和免疫记忆等优化学习机理,目前,许多研究人员更青睐于利用免疫优化机理来解决一些非线性问题[12,13]。

针对上述各种网络安全态势感知模型在态势预测方面存在的不足以及人工免疫系统解决非线性问题的优点,本文提出了基于免疫优化原理的网络安全态势预测方法(NSSPAI),试图利用该方法对网络安全态势进行相应的预测,实验结果已证明该方法的预测性能优于 GA 预测方法。

2 网络安全态势预测原理

网络安全态势预测是利用过去、当前的网络安全态势值的变化规律来挖掘合理的态势预测模型,随后利用该模型对网络安全的未来态势进行相应的预测。对于一个具有非线性特性的网络安全态势时间序列,其预测过程包括态势时间序列分析、预测建模和未来态势预测。Packard 等人在文献[14]中提出了相空间重构思想并将混沌理论引入到非线性时间序列分析中,其思想是用系统中某一变量的延迟坐标来重构相空间,随后 Takens[15] 对相空间重构理论给出了相应的数学证明,结论表示,通过一个合理的嵌入维度重构一个高维相空间可以把系统中有规律的轨迹恢复出来。因此,利用相空间重构理论对网络安全态势时间序列进行分析和重构,能够获取合理的用于挖掘网络安全态势预测模型的相空间(样本空间)。假设网络安全态势时间序列 $X = \{x(t) \mid t = 1,2,\cdots,n\}$,则利用相空间重构方法构造的样本空间 S 可表示为

$$\{X(t),Y(t)\} = \{[x(t-\tau), x(t-2\tau),\cdots, x(t-d\tau)],[x(t)]\} \quad (1)$$

其中 $t = n, n-\tau,\cdots,(d+1)\tau,\tau$ 为时间延迟,d 为嵌入维度,$X(t)$ 为预测模型的输入样本,$Y(t)$ 为预测模型的目标值,样本空间 S 中存在 $n-(d-1)\tau$ 个样本点。

在网络安全态势预测中,预测值受与其相邻的时间序列点影响较大,而受那些距离较远的时间序列点影响较小,因此,合理选择与预测点邻近的训练样本数(即样本窗口)可以提高挖掘预测模型的预测性能。假设样本空间 S 拥有 l 个样本,样本窗口长度为 w,则当预测 $Y(t)$ 时,先取 t 之前的 w 个样本用于挖掘预测模型 $f: R^d \to R$,然后利用该预测模型对 $Y(t)$ 进行预测,预测值 $\hat{Y}(t)$ 与预测模型的映射关系为

$$\hat{Y}(t) = f(x(t-\tau), x(t-2\tau), \cdots, x(t-d\tau)) \quad (2)$$

本文采用 C-C 方法[16]计算时间延迟 τ、嵌入维度 d 和样本窗口 w(延迟时间窗口)。

3 基于免疫优化的网络安全态势预测

在网络安全态势感知系统中,态势感知包括态势要素获取、态势评估和态势预测。在进行态势预测时,首先需要选取合理的态势评估方法来获取反映网络安全状况的态势评估值,然后利用态势预测方法对网络安全态势进行相应的预测,以便为网络决策者维护网络的安全运行提供有效的参考依据。

3.1 态势评估建模

态势评估是态势预测的基础,态势评估通过态势察觉,对态势要素进行态势提取,随后利用态势评估模型评估当前网络安全状况以便反映网络的安全态势情况。本文采用文献[6]提出的基于免疫的网络安全风险检测模型对网络安全态势进行量化评估。在该模型中,利用记忆细胞的抗体浓度值来定量评估网络遭受安全攻击的强弱程度。当记忆细胞检测到抗原时,如果检测到非自体抗原,则记忆细胞抗体浓度增加,反之,如果在给定周期内没有检测到非自体抗原,则记忆细胞抗体浓度逐渐衰减,当抗体浓度衰减至零时,则表示某类网络威胁已经被消除。

根据目标网络所提供的各类服务所面临的网络

安全态势情况，建立运合于服务、主机或网络的多种态势评估模型。假设 φ_j 为第 j 类攻击的危害性，ω_i 为主机 i 的重要性，p_{ij} 为主机 i 检测到第 j 类攻击的记忆细胞抗体浓度，则对于主机 i，在 t 时刻受到第 j 类攻击的安全态势评估值由式

$$St_{Host_{ij}}(t) = 1 - \frac{1}{1 + \ln(\varphi_j \cdot p_{ij} + 1)} \tag{3}$$

求得，而面临所有攻击的整体安全态势评估值由式

$$St_{Host_i}(t) = 1 - \frac{1}{1 + \ln(\sum_j (\varphi_j \cdot p_{ij}) + 1)} \tag{4}$$

求得。

同理，该网络在 t 时刻受到第 j 类攻击的安全态势评估值由式

$$St_{Net_j}(t) = 1 - \frac{1}{1 + \ln(\varphi_j \cdot \sum_i (\omega_i \cdot p_{ij}) + 1)} \tag{5}$$

求得，面临所有攻击的综合安全态势评估值由式

$$St_{Net}(t) = 1 - \frac{1}{1 + \ln(\sum_j (\varphi_j \cdot \sum_i (\omega_i \cdot p_{ij})) + 1)} \tag{6}$$

求得。

3.2 态势预测建模

态势预测是态势感知的最高级别，也是网络安全预警的主要参考依据。为挖掘态势预测模型，本文利用免疫优化机理实现态势预测模型建模。与挖掘态势预测模型有关的抗原、抗体和亲和力的定义如下：

定义 1 抗原是指用于挖掘态势预测模型的训练样本集。

定义 2 抗体是指用于预测的态势预测模型候选解。

定义 3 亲和力是指态势预测模型对态势时间序列曲线的拟合强度。

图 1 给出了态势预测模型建模流程。在预测模型建模过程中，其建模效率主要取决于抗体编码方法、亲和力评估方法和抗体优化算子。克隆选择、抗体变异、记忆抗体更新和基因更新等抗体优化算子的合理选取直接影响到预测模型的预测性能。

3.2.1 抗体编码

通常，预测模型是一个数学表达式，其构成元素主要涉及到算术运算符、数学函数、变量和常量等。对于数学表达式，除了线型表示形式外，其另一种表示形式为表达式树[17]。根据预测模型的表达式树

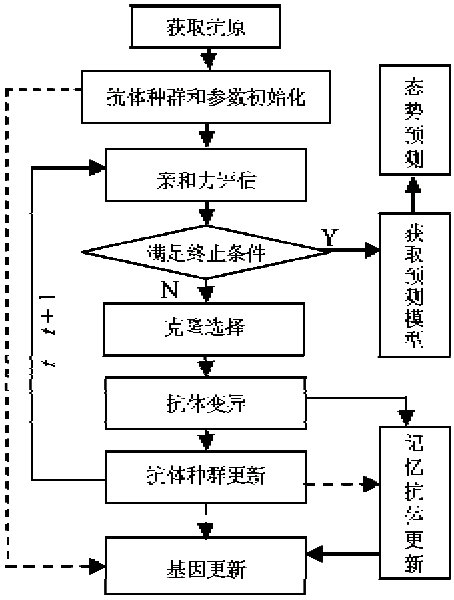

图 1 态势预测模型建模流程

表示形式和生物免疫系统中的抗体结构特点[13]，本文提出的抗体结构由易变区和恒定区组成，易变区和恒定区分别表示预测模型和该模型的优化系数，抗体编码采用字符编码机制。抗体编码的形式化描述为：抗体 $atb = < VR, CR >$，其中 $VR \in \{Funs, Vars\}^{L_v}$ 和 $CR \in \mathbb{N}^{L_c}$ 分别表示抗体的易变区和恒定区，L_v 和 L_c 分别表示抗体的易变区和恒定区长度；$Funs \subseteq \{+, -, *, /, Q, X, N, O, \cdots\}$ 为用户选定的运算符和函数集，Q 表示平方根函数，X 为指数函数，N 为正弦函数，O 为余弦函数；$Vars \subseteq \{?, a, b, c, \cdots, x, y, z\}$ 为根据给定问题的参数需求由用户指定的变量集，? 指定系数在预测模型中的位置。由此，一个预测模型 $0.5 * (a + b) - c/sin(d)$ 可以表示为 $? * (a + b) - c/N(d)$。图 2 给出了预测模型、表达式树与抗体编码三者之间的转换关系。图 2(a)表示一个预测模型，图 2(b)为对应于图 2(a)的一棵最大出度为2的表达式树。根据抗

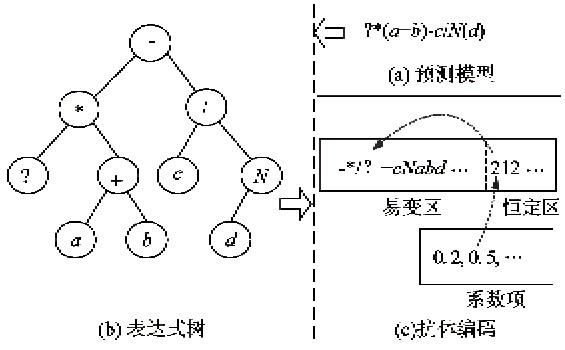

图 2 预测模型、表达式树与抗体编码

体结构特点和树的层次遍历法,通过图2(b)可得到图2(c)所示的抗体编码。由图2可知,抗体解码过程与抗体编码过程互为逆过程。

由抗体结构可知,易变区是确定抗体编码长度的决定性因素。依据抗体编码与表达式树的映射关系,本文借鉴表达式树的一些特性来确定易变区长度。表达式树包括非终端结点和终端结点,其中非终端结点包括运算符和函数集等,终端结点为变量集等。

定理1 对于一棵非终端结点数为 Fnc 的表达式树,如果树的最大出度为 m,则终端结点数最多不超过 $Tnc = (m-1) * Fnc + 1$。

证明:假设一棵表达式树具有 n 个结点,结点的最大出度为 m,n_i 表示出度为 i 的结点数,则该树的结点总数可表示为

$$n = n_0 + n_1 + n_2 + \cdots + n_m \tag{7}$$

在表达式树中,除根结点外,其余结点有且仅有一个分支进入,令 s 表示分支总数,则 $n = s + 1$。由于这些分支由出度为 1,2,3,…,m 的结点射出,可得 $s = n_1 + 2n_2 + 3n_3 + \cdots + mn_m$,所以该树的结点数也可表示为

$$n = n_1 + 2n_2 + 3n_3 + + mn_m + 1 \tag{8}$$

由式(7)(8)可知,出度为0的终端结点数 $n_0 = n_2 + 2n_3 + 3n_4 + \cdots + (m-1)n_m + 1$,所以,在最坏情况下,当非终端结点的出度都为 $m(n_2 = n_3 = \cdots = n_{m-1} = 0)$ 时,可得

$$n_0 = (m-1)n_m + 1 \tag{9}$$

式(9)即为所证结论,证毕。

由定理1可知,易变区长度为在最坏情况下表达式树的非终端结点数和终端结点数之和,即 $Fnc * m + 1$;对于恒定区,主要涉及到索引项和系数项两个概念,其长度可以根据问题的复杂性进行设定。

3.2.2 亲和力评估

依据定义1,抗原 $atg = \{ag_i \mid ag_i \in \mathbb{R}^d, i \in m\}$ 为样本空间中的训练样本集,抗原分量 $ag = < x_1, x_2, \cdots, x_d >$ 为样本点,其中 d 为样本空间的嵌入维度,m 为样本窗口长度。抗体 atb 与抗原 atg 的亲和力为

$$aff(atb) = \frac{1}{1 + \sqrt{\frac{1}{m}\sum_{j=1}^{m}(p_j(atb, ag_j) - t_j)^2}} \tag{10}$$

其中,p_j 为将第 j 个抗原分量 ag_j 所对应的训练样本数据代入到抗体 atb 所对应的预测模型时所得到的时序预测值,而 t_j 为第 j 个抗原分量所对应的时序目标值;$\sqrt{(\cdot)}$ 表示抗体与抗原的均方根误差。当 $\sqrt{(\cdot)}$ 越小时,抗体与抗原的亲和力就越大,表明抗体对抗原的适应度越好,即预测模型对网络安全态势时间序列的拟合度越强。

3.2.3 克隆选择

在克隆选择过程中,根据抗体亲和力大小,首先对抗体种群 Pop 进行降序排列,然后按克隆比率 R_{clone} 从 Pop 中选择 $\delta = \lceil N * R_{\text{clone}} \rceil$ 个高亲和力抗体进行克隆操作。克隆抗体总数为

$$N_{cln} = \sum_{i=1}^{\delta} \lceil (N-i)/i \rceil \tag{11}$$

克隆抗体子群为

$$Pop_{cln} = \{e_i \times atb_i \mid i = 1, 2, \cdots, \delta\} \tag{12}$$

其中 N 为抗体种群规模;i 为被选择克隆的抗体在有序抗体种群中的序号,i 越小,抗体亲和力越大,被克隆的抗体就越多;e_i 为一个具有 $\lceil (N-i)/i \rceil$ 维的单位列向量。

3.2.4 抗体变异

抗体变异可以发生在抗体的任意位置,其目的是通过变异操作来提高抗体的多样性和高亲和力抗体的产生几率。在抗体变异过程中,克隆抗体子群 Pop_{cln} 中的抗体 atb 根据其亲和力按式

$$p_m(atb) = e^{-aff(atb)} \tag{13}$$

定义的可变概率进行变异,并得到变异抗体子群 Pop_{mut}。

定理2 对于抗体变异操作,Pop_{cln} 变异为 Pop_{mut} 的概率 $P(Pop_{cln} \rightarrow Pop_{mut})$ 大于0。

证明:假设 $Pop_{cln} = \{a_i \mid i \in m\}$,$Pop_{mut} = \{b_i \mid i \in m\}$,由式(10)和式(13)可知:$P_m(a_i) \equiv P_m(aff(a_i)) > 0$。依据定义2,令任意 a_i 和 b_i 的二进制编码串长度为 l,且 a_i 和 b_i 的海明距离为 d,由文献[18]可知,a_i 变异为 b_i 的概率为 $P(a_i \rightarrow b_i) \equiv P_m(a_i)^d (1-P_m)^{l-d} > 0$,因此,必存在一个常数 $\zeta (0 < \zeta < 1)$ 使得

$$P(Pop_{cln} \rightarrow Pop_{mut}) = \prod_{i=1}^{m} P\{a_i \rightarrow b_i\} \geqslant \zeta \tag{14}$$

即式(14)满足所证结论,证毕。

3.2.5 记忆抗体更新

对于记忆抗体更新,每一代进化抗体中高亲和力抗体会对记忆抗体库 ML 中低亲和力抗体进行更新或替换操作,以保证库中优势抗体的逐渐增加。假设 t 为抗体进化代数,mn 为记忆抗体库的抗体

数，ml 为记忆抗体库容量，则式

$$ML(t) = Pop_{exc}(t) \cup ML(t-1) \tag{15}$$

和式

$$ML(t) = Pop_{exc}(t) \cup \{atb_i \mid atb_i \in ML(t-1), i = 1, 2, \cdots, ml - m\} \tag{16}$$

分别表示当 $mn < ml$ 和 $mn > ml$ 时记忆抗体库 ML 中抗体的更新途径，其中 Pop_{exc} 表示由 Pop_{mut} 中 m 个高亲和力抗体构成的优势抗体子群；当 $t = 0$ 时，ML 为 $\varnothing$。

3.2.6 基因更新

在抗体进化中引入基因的优势在于基因具备遗传特性，优秀基因在抗体种群中迅速扩散，可提高进化抗体的收敛速度。假设基因 $gen = < g, c >$，其中 $g \in \{Funs, Vars\}$ 为基因分子，$c \in \mathbb{N}$ 用于记录基因的使用频率，则基因在抗体基因库 GL 中的演化方式为

$$GL(t) = \begin{cases} \{gen \mid gen.g \in \{Funs, Vars\}, \\ \qquad gen.c = 0\}, t = 0 \\ TL - GL_{dead}(t) \cup GL_{exc}(t), t \geqslant 1 \end{cases} \tag{17}$$

在式(17)中，t 为抗体进化代数，当 $t = 0$ 时，$GL(0)$ 为由系统随机生成的初始抗体基因库。TL 为临时抗体基因库，用于统计当前代抗体种群中抗体对上一代基因库中基因的使用频率，如果基因被发现在抗体种群中的 p 个抗体内，则该基因的 c 加 p，反之，则 c 减 1。GL_{exc} 为从当前代记忆抗体库 ML 中提取的优势抗体基因群，并将 c 设置为上一代基因库中基因使用频率最高频率值，保证最新加入的基因不会过早被淘汰。GL_{dead} 为从上一代基因库中淘汰使用频率较低的基因群。

3.3 基于免疫优化的网络安全态势预测算法

基于免疫优化的网络安全态势预测算法描述如表 1 所示。该算法的特点是利用预测点之前的具有时间序列特性的有限个样本点和本文提出的态势预测模型建模方法来挖掘网络安全态势预测模型，随之，对网络安全态势进行相应的预测。

3.4 算法理论分析

由表 1 算法可知，态势预测建模是决定预测性能优劣的关键，下面对影响算法性能稳定、用于预测建模的抗体进化策略进行收敛性分析。

表 1 网络安全态势预测算法

步骤 1	利用态势评估方法获取网络安全态势时间序列数据
步骤 2	利用相空间重构方法分析和重构态势时间序列，获取用于挖掘预测模型的样本空间
步骤 2.1	分析态势时间序列，通过 C-C 方法获取相关参数：时间延迟、嵌入维度和样本窗口
步骤 2.2	由式(1)构造态势时间序列的样本空间
步骤 3	网络安全态势预测模型建模
步骤 3.1	获取预测样本点之前的多个样本点，样本数等于样本窗口长度
步骤 3.2	利用基于免疫优化的态势预测模型建模方法挖掘满足条件的预测模型
步骤 4	利用该预测模型进行网络安全态势预测

设抗体种群空间为 S^N，其中 N 表示种群规模；$s_n \in S^N$ 为 S^N 中的第 n 个状态；$AP(n) \subset S^N$ 为处于状态 s_n 的抗体种群；$AP*$ 为 S^N 上的最优抗体集；$O_{cg} = O_{cs} \circ O_{hm} \circ O_{mu} \circ O_{gu}$ 为复合算子，其中 O_{cs}、O_{hm}、O_{mu} 和 O_{gu} 分别表示克隆选择算子、高频变异算子、记忆抗体更新算子和基因更新算子；$P(AP(n))$ 表示 $AP(n)$ 处于状态 s_n 时的概率，$P(AP(n+1) \mid AP(n))$ 表示 $AP(n)$ 经过一步转移为 $AP(n+1)$ 的概率。因此，抗体种群的状态转移形式可表示为 $AP(n) \xrightarrow{O_{cg}} AP(n+1)$。

定义 4[19] 若随机序列 $\{AP(n), n \geqslant 0\}$ 满足 $\lim\limits_{n\to\infty} P(AP(n) \cap AP* \neq \varnothing) = 1$，则称此序列概率弱收敛。

定理 3 对任意初始分布，抗体种群序列 $\{AP(n), n \geqslant 0\}$ 是概率弱收敛的。

证明：设 $P_0(n) = P(AP(n) \cap AP* = \varnothing)$，则由 Bayes 公式有

$$\begin{aligned} P_0(n+1) &= P(AP(n+1) \cap AP* = \varnothing) \\ &= P(AP(n+1) \cap AP* \\ &= \varnothing \mid AP(n) \cap AP* \neq \varnothing) \\ &\quad \cdot P(AP(n) \cap AP* \neq \varnothing) \\ &\quad + P(AP(n+1) \cap AP* \\ &= \varnothing \mid AP(n) \cap AP* = \varnothing) \\ &\quad \cdot P(AP(n) \cap AP* = \varnothing) \end{aligned} \tag{18}$$

依据复合算子 O_{cg}，当存在最优抗体 $atb_1 \in AP(n)$，$atb_2 \in AP(n+1)$ 且 $aff(atb_1) \leqslant aff(atb_2)$ 时，可推出 $P(AP(n+1) \cap AP* = \varnothing \mid AP(n) \cap AP* \neq \varnothing) = 0$，即得

$$P_0(n+1) = P(AP(n+1) \cap AP* = \varnothing \mid AP(n) \cap AP* = \varnothing) \cdot P_0(n) \tag{19}$$

令 $\varphi(n) = |AP(n) \cap AP*|$ 为 $AP(n)$ 中包含的最优抗体数,则存在

$$\gamma = \min_x P(\varphi(n+1) = 1 \mid \varphi(n) = 0)_{\min} > 0 \tag{20}$$

由此得

$$P(AP(n+1) \cap AP* = \varnothing \mid AP(n) \cap AP* = \varnothing) = 1 - P(AP(n+1) \cap AP* \neq \varnothing \mid AP(n) \cap AP* = \varnothing) \leqslant 1 - \gamma < 1 \tag{21}$$

依据式(19)(21)可得

$$\begin{aligned} 0 &\leqslant P_0(n+1) \leqslant (1-\gamma) \cdot P_0(n) \\ &\leqslant (1-\gamma)^2 \cdot P_0(n-1) \\ &\leqslant \cdots \leqslant (1-\gamma)^{n+1} \cdot P_0(0) \end{aligned} \tag{22}$$

由于 $\lim_{n\to\infty}(1-\gamma)^{n+1} = 0$, 可得: $0 \leqslant \lim_{n\to\infty} P_0(n) \leqslant \lim_{n\to\infty}(1-\gamma)^{n+1} \cdot P_0(0) = 0$, 即

$$\lim_{n\to\infty} P_0(n) = 0 \tag{23}$$

所以

$$\begin{aligned} &\lim_{n\to\infty} P(AP(n) \cap AP* \neq \varnothing) \\ &= 1 - \lim_{n\to\infty} P(AP(n) \cap AP* = \varnothing) \\ &= 1 - \lim_{n\to\infty} P_0(n) = 1 \end{aligned} \tag{24}$$

故 $\{AP(n), n \geqslant 0\}$ 是概率弱收敛的,证毕。

4 仿真实验与结果分析

为验证 NSSPAI 方法对网络安全态势的预测性能,我们对该方法进行了仿真实验。实验环境为对外提供 WWW、FTP 和 E-mail 等服务(按端口号区分)的局域网络,实验数据为 MIT Lincoln 实验室提供的 DARPA 1999 数据集的部分数据[20]。利用该数据集提供的 DoS、R2L、U2R 和 Probe 四类攻击数据对 WWW、FTP 和 E-mail 等服务器进行模拟攻击。DoS、R2L、U2R 和 Probe 四类攻击的危害性分别设为0.7、0.6、0.4和0.2, WWW、FTP 和 E-mail 服务器的重要性分别设为0.8、0.4和0.6,安全态势采样周期为 1min。我们首先利用网络安全态势评估方法从实验环境中获取具有时间序列特性的网络安全态势值,图 3 为网络遭受 DoS 攻击的网络安全态势值及其相应的态势归一化值。然后利用NSSPAI方法对网络安全态势进行短期预测,以验证 NSSPAI 方法的有效性。

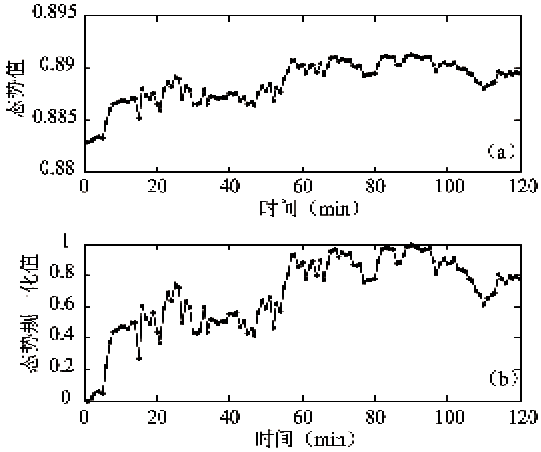

图 3 网络安全态势与态势归一化曲线

在 NSSPAI 模型建模过程中,采用 C-C 方法获取用于网络安全态势时间序列分析的相关参数值:时间延迟 τ 为 1、嵌入维度 d 为 10、样本窗口 w 为 9。而用于挖掘预测模型的抗体进化所涉及到的相关参数如表 2 所示。为避免预测结果的无偏性,预测值取 T 次实验结果的平均值。

表 2 抗体进化参数

参数名称	参数值
实验次数 T	100
抗体种群规模 N	50
最大进化代数 η	5000
抗体易变区长度 L_v	37
抗体恒定区长度 L_c	10
亲和力阈值 γ	0.98
克隆比率 R_{clone}	0.2

依据表 2 的抗体进化参数,图 4 给出了抗体进化时其亲和力的演化过程。当进化代数t =247时,

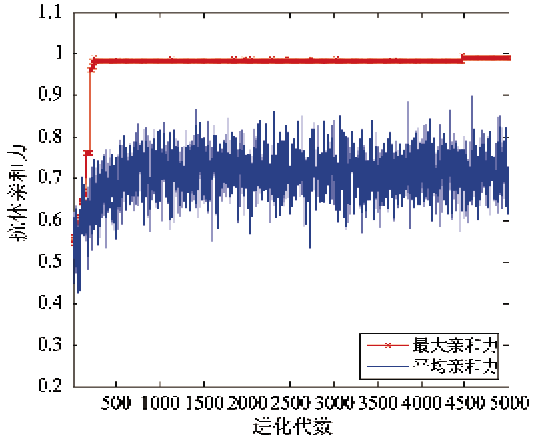

图 4 抗体亲和力演化

抗体亲和力值为 0.9835，达到亲和力阈值 γ，即得到符合系统要求的预测模型；当 $t \geqslant 4479$ 时，得到最大亲和力值为 0.9897。其结果表明 NSSPAI 预测模型对网络安全态势时间序列具有较好的拟合性和收敛速度。

为验证 NSSPAI 预测模型对网络安全态势的预测效果，我们将该模型与遗传算法[9]建立的预测模型（即 GA 预测模型）进行比较与分析。图 5 为网络遭受 DoS 攻击的态势目标值和预测值的变化曲线，与GA预测模型比较，当网络安全态势的随机波动

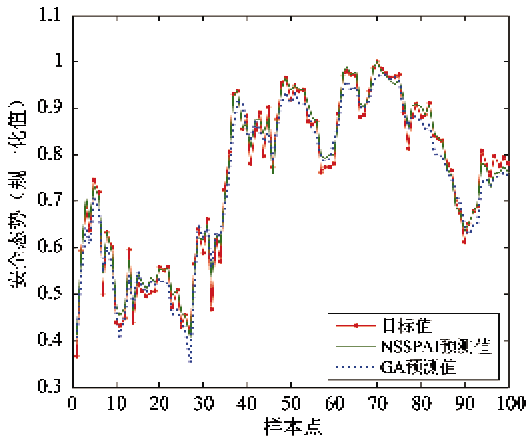

图 5 网络遭受 DoS 攻击的安全态势预测曲线

性比较强时，NSSPAI 预测模型的预测值仍能较准确地反映未来态势的发展趋势并对态势目标值具有较好的逼近能力。图 6 为对应于图 5 的态势目标值和预测值的偏离值（即绝对误差）对比曲线。由图 6 可以看出，GA 预测模型在样本点上的预测偏离值普遍高于 NSSPAI 预测模型的偏离值，最大预测偏离值达到0.0888，而NSSPAI预测模型的偏离值最

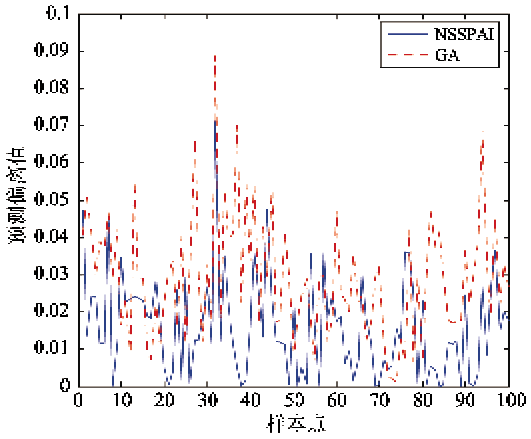

图 6 态势预测偏离值

大不超过 0.0712。实验结果表明，NSSPAI 方法的预测精度要优于 GA 方法，并且其预测结果能够较为准确地反映网络安全态势的未来变化趋势。

5 结 论

受生物免疫系统启发，本文结合网络安全态势预测原理和免疫优化学习方法，提出了一种基于免疫优化原理的网络安全态势预测方法（NSSPAI）用于解决网络安全态势预测问题。通过实验结果对比，NSSPAI 方法具有较好的预测效果，其拟合和预测结果都能较好地克服网络安全态势时间序列大幅度变化的影响。因此，在网络安全态势预测中，NSSPAI 方法为网络决策者提供了一种有效的网络安全监管方法，使网络决策者能够较为准确地判断和处理网络异常行为，提高了网络系统的稳定性和安全性。

参考文献：

[1] Bass T, Gruber D. A glimpse into the future of id. http://www.cyberstrategics.com/papers/pdf/, 2006

[2] Bass T. Intrusion detection systems and multisensor data fusion: creating cyberspace situational awareness. *Communications of the ACM*, 2000, 43(4):99-105

[3] Lau S. The spinning cube of potential doom. *Communications of the ACM*, 2004, 47(6): 25-26

[4] Carnegie Mellon's SEI. System for Internet Level Knowledge (SILK). http://tools.netsa.cert.org/silk/,2006

[5] 陈秀真，郑庆华，管晓宏等. 网络化系统安全态势评估的研究. 西安交通大学学报，2004, 38(4): 404-408

[6] Li T. An immunity based network security risk estimation. *Science in China Ser F Information Sciences*, 2005, 48(5):557-578

[7] 韦勇，连一峰. 基于日志审计与性能修正算法的网络安全态势评估模型. 计算机学报. 2009,32(4):763-772

[8] Lai J B, Wang H Q, Liu X W, et al. WNN-based network security situation quantitative prediction method and its optimization. *Journal of computer science and technology*, 2008, 23(2):222-230

[9] Szpiro G G. Forecasting chaotic time series with genetic algorithms. *Physical Review E*, 1997, 55(1): 2557-2568

[10] Lee C M, Ko C N. Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm. *Neurocomputing*, 2009, 73(3):449-460

[11] Keerthi S S. Efficient tuning of SVM hyper parameters u-

sing radius/margin bound and iterative algorithms. *IEEE Transactions on Neural Networks*, 2002, 13(5):1225-229

[12] Gong M G, Jiao L C, Zhang L N, et al. Immune secondary response and clonal selection inspired optimizers. *Progress in Natural Science*, 2009, 19(2):237-253

[13] Gan Z H, Chow T W S, Chau W N. Clone selection programming and its application to symbolic regression. *Expert Systems with Applications*, 2009, 36(2):3996-4005

[14] Packard N H, Crutchfietd J P, Farmer J D, et al. Geometry from a time series. *Physical Review Letters*, 1980, 45(9):712-716

[15] Takens F. Detecting strange attractors in turbulence. *Lecture notes in Mathematics*, 1981, 898:366-381

[16] Kim H S, Eykholt R, Salas J D. Nonlinear dynamics, delay times, and embedding windows. *Physica D*, 1999, 127(1):48-60

[17] Puntambekar A A. Data Structures and Algorithms. India: Technical Publications Pune, 2008. 220-281

[18] Rudolph G. Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 1994, 5(1):96-101

[19] 张文修, 梁怡. 遗传算法的数学基础. 第二版. 西安: 西安交通大学出版社, 2001. 100-101

[20] MIT Lincoln Laboratory. 1999 DARPA intrusion detection evaluation data set. ftp://ftp.ll.mit.edu/pub/ideval/, 2003

A new approach for network security situation predictionbased on the immune optimization theory

Shi Yuanquan[*,**], Liu Xiaojie[*], Li Tao[*], Peng Xiaoning[**], Chen Wen[*], Zhang Ruirui[*]

([*] College of Computer Science, Sichuan University, Chengdu 610065)

([**] College of Computer Science, Huaihua University, Huaihua 418000)

Abstract

To effectively monitor networks' security situation and prevent large-scale networks from being attacked, a novel network security situation prediction approach based on the immune optimization theory (NSSPAI) is proposed. The NSSPAI works according to the steps described belows: firstly, it gives the definitions of antigen, antibody and affinity in the environment of network security situation prediction, and gives the mathematical models of some antibody optimizing operators for constructing a prediction model; secondly, the time series of network security situation is analyzed by using the phase space reconstruction theory, and a prediction model is established by using the reconstructed sample space and the immune optimizing modeling method; lastly, the future network security situation is predicted by this model. The experimental results show that the NSSPAI can forecast the future network security situation more exactly than the genetic algorithm based prediction approach, and it is a new method for effective prediction of network security situation.

Key words: network security, situation awareness, time series, immune optimization, prediction