

基于免疫克隆与差分进化的入侵检测方法

边根庆, 赵 宏, 张维琪, 孙继武

(西安建筑科技大学 信息与控制工程学院, 陕西 西安 710055)

摘 要: 随着网络技术的飞速发展, 信息安全变得越来越重要。入侵检测方法已经成为信息安全领域的热门研究方向之一。入侵检测实质上是一个分类的问题, 对于提高分类精度是十分关键的。免疫克隆算法和差分进化算法都是用于解决分类问题的比较有效的方法。结合先前免疫克隆算法和差分进化算法分别在入侵检测中所取得的良好效果, 在对差分进化算法改进的基础上, 提出了一种基于免疫克隆和差分进化的混合入侵检测方法。仿真实验结果表明, 该方法比文献中提出的方法具有更高的检测率和更低的误检率, 能使检测系统在一定程度上有效的识别未知入侵。

关键词: 免疫克隆算法; 差分进化算法; 入侵检测; 检测率; 误检率

中图分类号: TP393.08

文献标识码: A

文章编号: 1000-7180(2012)05-0124-05

Intrusion Detection Method Based on Immune Cloning and Differential Evolution

BIAN Gen-qing, ZHAO Hong, ZHANG Wei-qi, SUN Ji-wu

(College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: With the rapid development of network, information security has become more and more important. Intrusion detection method has become one of research direction in the field of information security. Intrusion detection is essentially a classification problem. It is very critical to increase the classification accuracy. Immune cloning algorithm and differential evolution algorithm are both effective methods to solve classification problems. Combined with previous good effect of intrusion detection in which immune cloning algorithm and differential evolution algorithm were used respectively. This paper proposes one intrusion detection method based on the improvement of differential evolution algorithm which includes immune cloning algorithm and differential evolution algorithm. Simulation experiments show that this method has higher detecting rate and lower error detecting rate than the method of literature, it can make the intrusion detection system recognition unknown intrusion effectively in certain extent.

Key words: immune cloning algorithm; differential evolution algorithm; intrusion detection; detection rate; error detecting rate

1 引言

随着计算机技术和互联网技术的高速发展, 计算机系统遭受的入侵和攻击的概率也越来越多, 信息安全成为人们必须面对和解决的一个重要问题。入侵检测系统作为一种能够自动、实时地保障网络

安全的动态安全设备, 已成为信息安全主要的研究方向之一。目前有很多智能算法被应用于入侵检测系统, 这些算法来源于模拟自然界生物的决策过程, 在一定范围内取得了良好的效果, 保障了网络的安全。

鉴于入侵检测系统与生物的自然免疫系统在抵

收稿日期: 2011-08-30; **修回日期:** 2011-10-23

基金项目: 国家自然科学基金(61073196); 陕西省自然科学基金(2011JM8026); 陕西省教育厅自然科学基金专项基金(11JK0982)

抗病毒等病原体的入侵方面具有类似的特征^[1].最早由 Forrest 小组将自然免疫的思想引入到信息安全领域,应用免疫耐受机制来检测程序和受保护数据的异常改动,其实验结果显示,这种方法能够很容易地发现病毒感染引起的数据文件的改变,并且能够检测未知攻击^[2];之后的人工免疫系统^[3]实现的轻量级入侵检测系统功能过于简单,检测效率不是很高;针对网络入侵检测的人工免疫系统方面,文献^[4]提出了一个基于概念的框架,但没有具体的实现.差分进化计算^[5-6]产生于 1995 年,是由美国的 Rainer Storn 和 Kenneth Price 在遗传算法等进化思想的基础上,提出的一种新颖的进化算法,该算法同样是模拟自然界生物进化机制的一种仿生智能算法,它最主要的操作思想是基于种群内的个体差异度生成临时个体,采用一对一的贪婪选择竞争机制,通过随机重组实现种群进化.差分进化具有收敛速度快、鲁棒性好、实现简单和参数少等优点,然而有时差分进化算法也会面临陷入早熟或者收敛速度下降等问题^[7].

针对以上算法存在的不足,本文对差分进化算法做出了改进,并结合免疫克隆算法的优点,提出了一种基于免疫克隆和差分进化的混合算法的人侵检测方法.仿真试验证明,该方法在入侵检测方面具有更好的效果.

2 算法思想描述

2.1 免疫克隆算法

克隆选择学说^[8]是由澳大利亚免疫学家 Burnet 等人提出的抗体形成理论.克隆算法是根据克隆选择原理设计的免疫算法,克隆算法的实质是在进化过程中,在每一代最优解的附近,根据亲和度的大小进行克隆,产生一个变异解的群体,增加抗体的多样性,从而扩大了搜索范围.克隆就是将一个低维空间的问题转化到更高维的空间中去解决,然后将结果投影到低维空间中,从而获得全局优化^[9].

免疫克隆算法思想是借鉴了生物体的免疫系统在对病毒入侵时做出的反应策略,当抗原侵入生物体时,其免疫系统在机体内选择出能够识别和消灭相应抗原的抗体,主要借助克隆使之激活、分化和繁殖,以增加该抗体的数量,最终清除抗原^[10].从信息处理的角度来看,免疫系统具备强大的识别、记忆和学习的能力.

克隆选择算法的设计是人工免疫系统中的重要组成部分,对完成系统的进化以及特定问题的收敛

具有重要作用^[11].克隆选择是由亲和度诱导的抗体随机映射,抗体群的状态转移情况可以表示成如下的随机过程: $Cs; A(K) \xrightarrow{\text{clone}} A'(K) \xrightarrow{\text{mutation}} A''(K) \xrightarrow{\text{selection}} A(K+1)$.

(1)编码方案.生产初始种群时,随机产生 Np 个抗体,每个抗体表示一种特征组合,可以采用二进制或十进制编码形式.在这里我们采用十进制编码形式,若基因串长度为 L ,编码为 $(A_1, A_2, \dots, A_L)$.

(2)亲和度函数的构造.抗原和抗体的亲和度可以用它们之间的距离来计算,本文采用 Euclidean 公式来计算距离:

$$D = \sqrt{\sum_{i=1}^L (Ab_i - Ag_i)^2} \quad (1)$$

式中, L 为编码的长度.

在计算出每个个体的亲和度后,可以设定一个阈值来终止克隆变异过程,则在实验中可如下设定阈值 T :

$$T = \left[\frac{f_{\max} - f_{\min}}{2} \right] \quad (2)$$

式中, $[\cdot]$ 表示四舍五入算子, $f_{\max}$ 、 $f_{\min}$ 分别表示最大亲和度和最小亲和度.

2.2 差分进化算法

2.2.1 基本的差分进化算法

差分进化是基于实数编码的进化算法,它的整体结构类似于遗传算法,与遗传算法的主要区别在变异操作上,差分进化的变异操作是基于染色体的差异向量进行的^[12].

差分进化算法的执行过程如下:

(1)初始化:确定种群规模 M ,交叉概率 $P_c \in (0, 1]$;随机选取初始种群 $X(t) = (X_1(t), X_2(t), \dots, X_i(t))$; $X_i(t)$ 表示第 t 代第 i 个个体,置 $t=0$, $X_i(0)$ 为 n 维向量.

(2)种群演化:对 $X(t)$ 中的每个个体 $X_i(t)$ 作如下操作:

① 变异操作:变异是 DE 算法产生个体的关键步骤.从种群中随机选取两个个体 X_{p1}, X_{p2} 作如下操作:

$$V_{ij}(t+1) = x_{\text{best},j}(t) + \eta(x_{p1,j}(t) - x_{p2,j}(t)) \quad (3)$$

式中 $x_{p1,j}(t) - x_{p2,j}(t)$ 为差异化向量, i, p_1, p_2 互不相等; $V_{ij}(t+1)$ 代表第 i 个个体的第 j 个分量, $x_{\text{best},j}(t)$ 是当前代最好的个体向量, η 为变异因子, $\eta \in [0, 2]$; $i=1, 2, \dots, B$, $j=1, 2, \dots, n$.

② 交叉操作:

$$u_{ij}(t+1) = \begin{cases} V_{ij}(t+1) & \text{randl}_{ij} \leq p_c \text{ 或 } j = \text{rand}(i) \\ x_{ij}(t) & \text{randl}_{ij} > p_c \text{ 或 } j \neq \text{rand}(i) \end{cases} \quad (4)$$

式中, p_c 为交叉概率且 $p_c \in (0, 1]$, randl_{ij} 为 $[0, 1]$ 间随机小数, $\text{rand}(i)$ 为 $[0, n]$ 间随机整数.

③ 选择操作: 将向量 U_i 和目标向量 X_i 的适应度进行比较, 只有当 U_i 的适应度比 X_i 更好的时候才被选作子代; 否则, 直接将 X_i 作为子代. 选择操作的方程为:

$$X_i(t+1) = \begin{cases} U_i(t+1) & f(U_i(t+1)) < f(X_i(t)) \\ x_{ij}(t) & \text{otherwise} \end{cases} \quad (5)$$

式中, $f(X_i(t))$ 为第 t 代第 i 个个体的适应度.

(3) 终止判断

令由(2)产生的新种群为

$$X(t+1) = (X_1(t+1), X_2(t+1), \dots, X_N(t+1)) \quad (6)$$

记 $X(t+1)$ 中的最优个体为 $X_{best}(t+1)$. 如果满足精度要求或整个进化已达到时限, 则终止并输出 $X_{best}(t+1)$ 作为最优解; 否则, 置 $t = t+1$, 转到(2).

2.2.2 算法的改进

(1) 适应度函数的阈值.

我们可以用类之间和类内各个体的距离表示它们的适应度. 类之间的距离越大, 类内各个体之间的距离越小, 则适应度越好. 用于计算类间距离的常用公式是欧式距离, 但是在差分进化算法中使用欧式距离计算过于复杂^[12], 所以本文采用如下的方法设置适应度函数.

$$f_{\text{fitness}} = S_b - S_w \quad (7)$$

$$S_b = \sum_{c_i \neq c_j} \|S_i - S_j\| \quad (8)$$

$$S_w = \sum_{c_i = c_j = k} \|X_i - X_j\| \quad (9)$$

$$S_i = \frac{1}{n} \sum_{i=1}^{n_i} X_i \quad (10)$$

上述公式中, S_b 表示两个不同类之间的欧式距离, S_w 表示同一个类样本之间的距离. 对于其中某一个类来讲, S_i 表示某类的类中心, 用均值表示. 在这里, 我们可以在试验中设置平均适应度作为阈值:

$$H_{DE} = \frac{\sum_{i=1}^n f_{\text{fitness}}(x_i)}{n} \quad (11)$$

(2) 改进对每个个体的变异方法. 首先从种群中

按照适应度大小选取种群一半的个体, 随机组成一条 Markov 链, 并按照以适应度概率应用 Powell 法, 找出最优解个体, 避免过早收敛而产生早熟现象.

$$V_{ij}(t+1) = x_{best,j}(t) + \eta H_j \quad (12)$$

式中, $H_j = x_{p1,j}(t) - x_{p2,j}(t)$ 为加速方向, 进行一维搜索, 直到找到最优个体. 在整个差分进化算法中 $x_{p1,j}(t) - x_{p2,j}(t)$ 表示为差异化向量, i, p_1, p_2 互不相等; $V_{ij}(t+1)$ 代表第 i 个个体的第 j 个分量, $x_{best,j}(t)$ 是当前代最好的个体向量, η 为变异因子, $\eta \in [0, 2]$; $i=1, 2, \dots, B$, $j=1, 2, \dots, n$.

3 基于免疫克隆和差分进化的入侵检测方法

本文提出的是一种基于免疫克隆和差分进化的混合算法(ICDE)检测器, ICDE 算法的流程(图 1)如下:

将一个种群分为 n 个子种群, 分别对每一个子

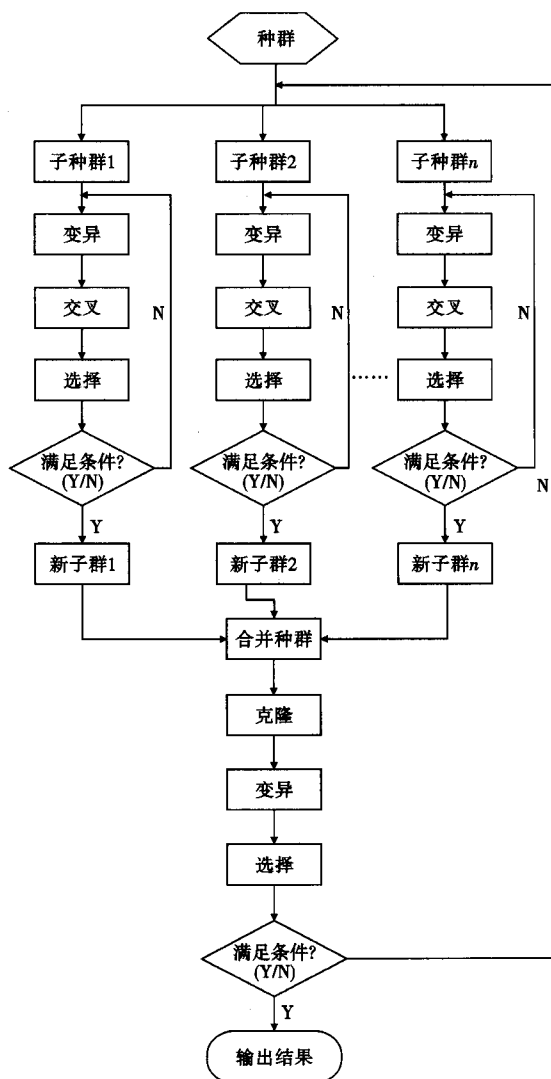


图 1 ICDE 检测器

种群作差分进化的操作;然后,合并所有的子种群,再对合并后的子种群运用免疫克隆算法。

步骤1 将种群随机的分为 n 个子种群,当前代种群 $t=0$;

步骤2 对每个子种群作差分进化处理

步骤2.1 计算每个个体的适应度,确定当前种群中的最佳个体。

步骤2.2 从当前的子种群中按照适应度大小选取其中一半的个体,进行变异操作。

步骤2.3 对当前的子种群进行交叉操作。

步骤2.4 对当前的子种群进行选择操作。

步骤2.5 判断个体是否满足终止条件,是否大于等于阈值 H ;若是输出当前新的子种群,跳到步骤3。否则,返回到步骤2.2。

步骤3 合并所有的 n 个新子种群,组成合并的种群。

步骤4 根据亲和度(式(1))计算每个个体的亲和度。

步骤5 对当前的第 t 代种群进行克隆,每个抗体的规模根据亲合度大小按比例分配。

步骤6 对克隆后的当前第 t 代种群,以一定的变异概率进行变异操作。

步骤7 计算变异后的当前第 t 代种群个体的亲和度,以一定的比例选择亲和度较大的个体作为下一代种群。

步骤8 满足迭代次数,或阈值时,输出结果;否则返回步骤2, $t=t+1$ 。

4 ICDE 入侵检测系统

将改进的免疫克隆和差分进化算法应用到入侵检测系统当中去,主要有数据预处理器、异常模式库、ICDE 检测器等组成,如图2所示。

数据预处理器用于对输入数据进行标准化,以消去干扰数据对分类的影响;在很多情况下,一些入侵者攻击的方式存在着很多的相似之处。因此,使用异常数据模式库,记录一些异常入侵的攻击模式,若下次入侵者使用相同的方法异常入侵,系统将会报警,拒绝其访问。ICDE 检测器使用 ICDE 算法来计算入侵者是否为异常入侵,若是异常入侵,则报告给系统管理中心,并记录到异常数据模式库;若非异常入侵,则允许访问系统。

5 实验结果与分析

为评价本文提出的入侵检测算法,我们以文献

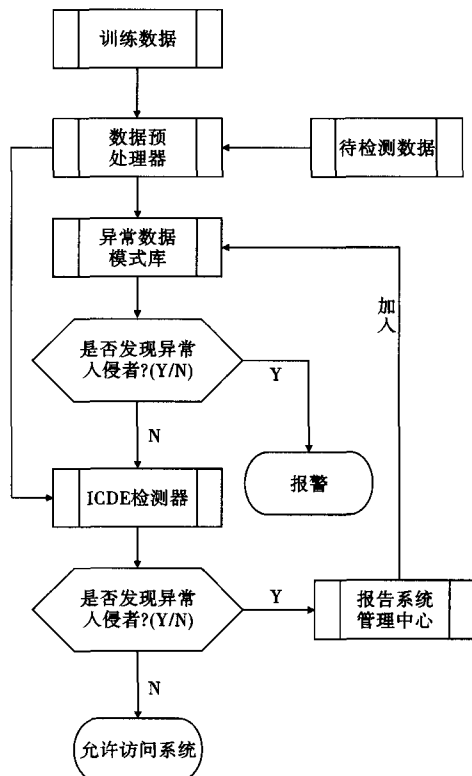


图2 ICDE 入侵检测系统

[11]为参照,选用的测试数据集是 KDD cup99 数据^[12]中的数据。为了便于进行仿真实验,采用随机抽取的5组13 000个样本数据集,其中的1组为训练集,其余4组为测试集。每个数据子集中的正常数据记录个数占本子集所有数据记录个数的比例约为98.5%,同时各种入侵活动数据记录也相应被均匀分配到这4个数据子集中,符合了检测算法中正常数据的数目要远多于入侵数据数目的假设。

本文所选用的数据集中,正常入侵的数据远远大于异常入侵的数据,用误检率和检测率来检验算法的性能,并与文献[11]作比较。误检率=被误判为入侵的正常记录数/总测试记录中的正常记录数;检测率=检测出来的入侵记录数/总测试中的入侵记录数。

为了得到准确的数据,进行了多次仿真,出现的整体结果要优于文献[11]取得的效果。对每一组数据的检测率和误检率与文献[11]作比较。

在四组数据中,除了其中一组数据的检测结果和文献[11]的结果差不多以外,其余3组的测试结果均要好于文献[11]。为了更清楚的比较检测率-误检率的整体性能,通过图3能直观看到结果:

从图3所示曲线可以发现,与文献[11]的入侵检测方法相比,本文算法的整体性能较高,拥有较高

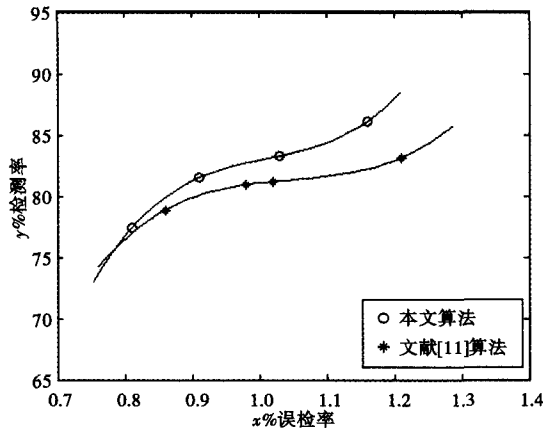


图3 两种算法检测率的比较

的检测率和更低的误检率. 这是由于本文提出的算法结合了免疫克隆和差分进化算法的优点, 较好地克服了差分进化算法过早收敛的缺陷, 提高了入侵检测准确率.

6 结束语

为了提高网络入侵的检测率, 本文将一个大的种群分 n 个子种群, 进行差分进化后, 再将所有子种群的优秀个体合并, 再对其进行免疫克隆, 最后产生下一代种群. 为针对差分进化算法的过早收敛而产生早熟现象, 在变异操作这一步作出了改进. 对每个个体进行变异操作时, 按照适应度大小选取一定的个体, 随机组成一条 Markov 链, 应用 Powell 法进行计算, 找到最优解, 避免了过早收敛而产生早熟的现象.

在这些基础上, 本文提出了一种新的入侵检测方法(ICDE), 通过利用 KDD CUP99 数据集对该方法进行了仿真试验, 表明其有效地提高了检测效率, 对未知入侵有较好的检测效果.

参考文献:

[1] 莫宏伟. 人工免疫系统原理与应用[M]. 哈尔滨: 哈尔

滨工业大学出版社, 2003.

- [2] Forrest S, Perelson A, Allen L. Self-noself discrimination in a computer[C] // Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy. [S. l.]: IEEE Press, 1994.
- [3] Hofmeyr S A, Forrest S. Immunity by design: an artificial immune system [EB/OL]. [2007-02-10]. <http://www.cs.unm.edu/~immsec/publications/geccosteve.pdf>.
- [4] KIM J, BENTLEY J P. An artificial immune model for network intrusion detection[C] // Proceedings of the 7th European Conference on Intelligent Techniques and Soft Computing (EUFIT'99). Aachen: [s. n.], 1999.
- [5] Storn R, Price K. Differential evolution-a simple and efficient heuristic for global optimization over continuous Spaces[J]. Journal of Global Optimization, 1997 (11): 341-359.
- [6] Storn R, Price K. Differential evolution for multi-objective optimization [J]. Evolutionary Computation, 2003(4): 8-12.
- [7] 赵昕, 李剑. 应用于负荷经济分配的改进差分进化算法[J]. 微电子学与计算机, 2010, 27(2): 159-166.
- [8] 周光炎. 免疫学原理[M]. 上海: 上海科学技术出版社, 2002.
- [9] 白琳. 基于克隆选择聚类的入侵检测[J]. 微电子学与计算机, 2007, 24(3): 135-141.
- [10] 焦李成, 刘芳, 侯水平, 等. 智能数据挖掘与知识发现[M]. 西安: 西安电子科技大学出版社, 2006.
- [11] 任健, 邵定宏. 基于克隆选择算法的最大模糊熵图像分割[J]. 微电子学与计算机, 2010, 27(9): 114-116.
- [12] 王培崇. 基于群智能计算技术的网络入侵检测算法研究[D]. 北京: 中国矿业大学, 2010.

作者简介:

边根庆 男, (1968—), 硕士, 副教授. 研究方向为海量信息存储及信息安全技术的研究.