

一种改进的基于混沌理论的木马检测器生成算法

秦 杰, 司 群, 颜会娟, 闫付亮

(河南工业大学 信息科学与工程学院, 河南 郑州 450001)

摘 要: 在分析现有应用于入侵检测系统中的检测器生成算法的基础上, 提出了一种基于混沌理论的木马检测器生成算法, 通过混沌的初值敏感性实现了广泛的搜索空间. 结合加权欧式距离, 并利用混沌的遍历特性产生了均匀的检测器集合. 采用 Microsoft visual 2008 和 Matlab 进行仿真实验, 结果表明, 改进的算法能有效保持抗体的多样性, 并具有较快的收敛速度.

关键词: 混沌; 木马; 加权欧氏距离; 检测器

中图分类号: TP309.2

文献标识码: A

文章编号: 1000-7180(2012)04-0101-03

An Improved Trojan Detector Generating Algorithm Based on Chaos Theory

QIN Jie, SI Qun, YAN Hui-juan, YAN Fu-liang

(College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: After studying the existing detector generating algorithms used in the intrusion detection systems, a new approach of detector generating was presented which was based on chaos theory. The over-spread character of chaos sequence combined the concept of weighted Euclidean distance was used to generate set of detector with better distribution, and chaos initial value sensitivity was used to enlarge the searching space. Simulation experiment and the results indicate that the arithmetic not only remains the diversity of population but also has good fast astringency.

Key words: chaos; Trojan; weighted Euclidean distance; detector

1 引言

基于混沌理论^[1]的检测器生成算法克服了随机检测器生成算法难度随自体集规模增大呈指数增加的限制, 另外, 该算法适用于任何一种匹配规则, 而不像线性时间检测器生成算法和贪婪检测器生成算法仅仅适用于 r -连续位匹配规则^[2]. 当前, 基于混沌理论的检测器生成算法存在搜索盲区大和收敛速度慢的不足, 导致这些不足的原因是该类算法产生初始检测器大多数采用的是 Logistic 模型^[3]生成混沌序列和添加混沌扰动^[4]. 有学者提出混沌力学和信息熵^[5]相结合, 该算法调整和控制了初始检测器的分布情况, 但仍存在缺陷, 在优化算法中, 对应由抗体 $m = \{0111111111\}$ 和抗体 $n = \{1000000000\}$ 构成

的抗体群, 其平均信息熵 $H(2) = 0.693\ 147$, 表明这两个抗体是两个很不相同的抗体, 但是在优化算法中它们是两个很相近的解(通过解码后).

本文将混沌理论引入到木马检测器生成算法中, 并针对以上缺陷, 提出了一种基于混沌理论, 结合数学中的加权欧氏距离产生均匀检测器的方法, 该方法能够较好地解决搜索盲区大、收敛速度慢、误删抗体等问题.

2 免疫系统的混沌特征

混沌是由连续或离散的动力系统产生的, 是一种没有固定周期的循环行为, 具有初值敏感性、遍历性和随机性. 混沌理论被广泛应用在了工程学、物理学、数学和生物学等领域^[2,5]. 同时混沌具有对初值

敏感性遍历性和随机性等特性^[6].

免疫系统属于耗散系统,需要不间断的吸收能量和物质,并不断的耗费这些能量与物质.通过考察抗原细胞浓度变化,可得出抗原细胞增殖方式,由方程(1)给出:

$$A_g(t+1) = k \left(1 - \frac{A_g(t)}{A_{g\max}} \right) A_g(t) - k' A_g(t) \quad (1)$$

式中,抗原细胞在 t 时刻的浓度用 $A_g(t)$ 表示,要得到 $t+1$ 时刻的抗原浓度,可 t 时刻计算得出.抗原的增长率用 k 表示,细胞的死亡率用 k' 表示.初值设置为 $A_g(0)=0.3$,当 $\mu=k-k'$ 的值小于 3.0 时候, $A_g(t)$ 趋于稳定,当 $3.1 < \mu \leq 3.4$ 时,函数趋于振荡,当 $3.4 \leq \mu < 4.0$ 时,函数呈现混沌现象^[6].

3 改进的基于混沌理论的检测器生成算法

在免疫系统中,初始状态微小的改变就导致系统状态发生巨大差异,这就是混沌初值敏感的特点,这个特点增大了空间的范围.利用混沌初值敏感的特点来产生的个体均匀的分布在个体解空间且不会产生聚集.这样可能出现的数据冗余就得到了有效避免,而利用混沌扰动对不适应的个体进行“变异”得到更优个体^[6].

本文中提出的改进的基于混沌理论的检测器生成算法,主要是通过混沌序列产生候选检测器集合,然后初始检测器集合的生成是根据数学中加权欧氏距离原理.基于加权欧氏距离的计算检测器平均亲和力的算法与基于信息熵的算法相比较,计算产生加权欧氏距离中的权值会花费一定时间,但是它比计算信息熵的速度大大提高,对初始检测器集合的处理能力也提高了.

检测器生成算法流程如图 1 所示.

3.1 初始检测器的初始化

检测器的初始化成功与否使得成熟检测器的生成效率和质量受到直接影响.因为混沌序列的遍历性和随机性,以及个体均匀分布在解空间中这一特性,所以在本文算法中,对于初始检测器的产生采用了混沌序列的算法.

检测器集合的第一个个体 x_0 随机产生,并映射到 $[0,1]^n$ 区间中得到 y_0 ,将 y_0 作为混沌序列的初始值,继而通过 Feigenbaum 迭代方程产生混沌序列 y_i ,该迭代方程是由 Logistic 映射形成的,最后将 y_i 映射到解空间生成初始检测器集合 y_i , ($i=1,2,\dots,n$),并对每个检测器随机产生一个代表其重要性的权值 $w_1, w_2, \dots, w_n$ ($0 < w_i < 1$).

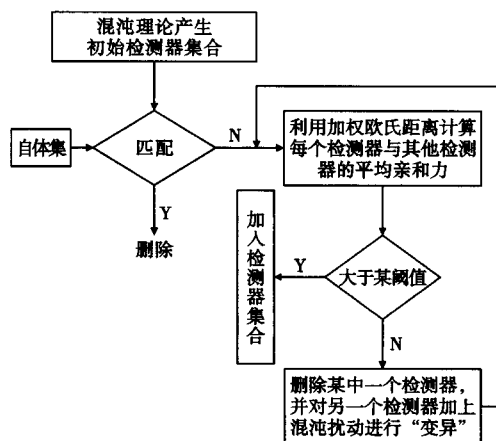


图 1 基于混沌理论和加权欧氏距离的检测器生成算法流程图

3.2 适应度函数的构造

本文中引入加权的欧氏距离来计算某检测器与其他检测器的平均亲和力.一般的计算采用欧氏距离,我们分别将抗原和抗体定义为 N 维实体空间中的不同的坐标点,则亲和力表述为两点间的距离,假如给定抗体坐标 $\langle x_1, x_2, \dots, x_N \rangle$,抗原坐标 $\langle y_1, y_2, \dots, y_N \rangle$,欧氏距离计算抗体和抗原间的亲和力

公式为 $D = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$.文中使用了加权欧氏

距离,即为 $D = \sqrt{\sum_{i=1}^N w_i (x_i - y_i)^2}$, w_i 为定义的权值.文中抗体和抗原采用二进制表示,计算公式做了一些改进,下面会提到具体计算公式.

在基于信息熵的人工免疫算法中,由于需要计算每代抗体群中各个抗体的信息熵,所以算法的速度大大减缓^[8].用检测器 x_i 与其他检测器的平均亲和力来表示检测器 x_i 的适应度,而适应度高的检测器在选择过程中会保留下来,改进的适应度函数构造如公式(2):

$$a(\vec{x}_i) = \frac{\sum_{j=1, j \neq i}^M w_j \sqrt{\sum_{k=1}^N (y_i^k - y_j^k)^2}}{M-1} \quad (2)$$

式中 y_i^k 表示第 i 个检测器第 k 位上的值, M 表示所有初始检测器的个数, w_j 位第 j 个检测器的重要性等级即权值,并且每个检测器都是 N 位的二进制编码.

3.3 混沌扰动

对于不大于规定阈值的检测器采取混沌扰动,具体过程参照公式(3):

$$y'_r = (1-\eta)\hat{y} + \eta y_r \quad (3)$$

$\hat{y}$ 为当前系统的最优值, η 随着代数 r 的变化而自适应改变, 可以取 $[0, 1]$ 之间的参数, y_r 为迭代 r 次后的混沌向量, y'_r 为添加了扰动的混沌向量. 根据公式(3)得出, 因为混沌具有初值敏感性, y_r 的巨大差异是由混沌初值的微小变化所引起的, 从而使生成检测器的速度比较快, 同时完成了变异个体范围增加、个体优化、种群丰富这些目标.

4 实验比较

将本文算法与基于混沌理论和信息熵的检测器算法进行对比实验. 实验的平台是 Windows XP, CPU 2.83 GHz, 2 GB 内存, 150 GB 硬盘. 开发工具采用 Microsoft Visual 2008, 并用 Matlab 进行仿真实验.

由于 Feigenbaum 迭代方程使用极其方便并且有较小的计算量, 所以实验中混沌序列的产生采用公式(4):

$$x_{n+1} = \mu x_n (1 - x_n) \quad (4)$$

在实验过程, μ 取 3.7, x 为混沌变量, $0 < x_n < 1$, 其混沌空间为 $[0, 1]$. 实验中随机产生个体 x_0 , 将 x_0 映射到 $[0, 1]^n$ 即 $y_1 = 0.3$, 以迭代次数为横坐标, 以每次的 y 值为纵坐标, 如图 2 所示.

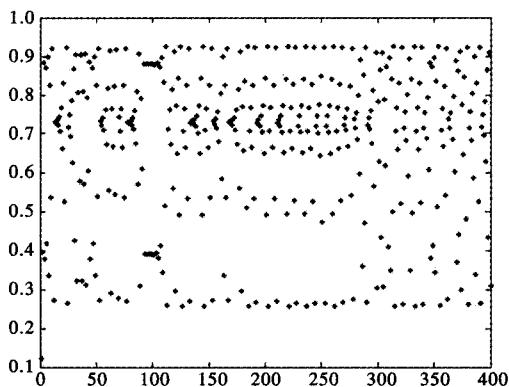


图2 代数与混沌序列初值关系图

由图 2 可看出混沌的遍历性, 基本上遍历了大部分空间, 此特性保持了抗体的多样性, 取自体集大小为 40, 通过计算检测器适应度函数值, 选择其中适应度函数值大于规定阈值的检测器生成后代检测器集合, 比基于混沌理论和信息熵的检测器生成算法的时间开销少. 在同等条件下, 用这两种检测器生成算法来进行对比实验, 图 3 所示为进化代数与所生成检测器个数的关系.

由图 3 分析可知, 本文算法在产生相同检测数

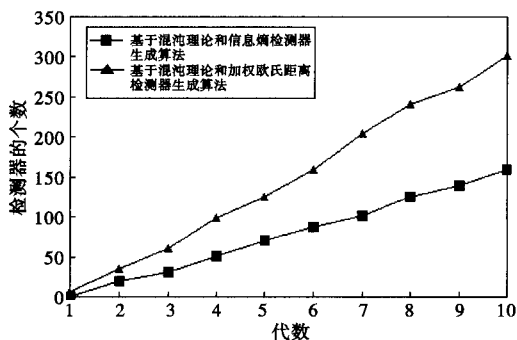


图3 检测器代数和检测器个数的关系

时, 比基于信息熵的迭代次数低即该算法运行效率高, 另外, 在相同的迭代次数下产生可用检测器个数多, 产生的抗体具有更好的多样性. 表 1 中给出了两种算法性能的对比较结果.

表1 两种算法性能比较结果

性能 算法	算法运行 的次数	每次进 化代数	所耗 时间	每次收敛到全 局的最优次数
基于混沌理论和信息熵的检测器生成算法	50	400	25.6	44
本文算法	50	400	4.0	45

实验中算法重复运行的次数为 50 次, 所耗时间是算法运行 50 次所需时间之和. 由表 1 可以看出, 在基于混沌理论和信息熵的检测器生成算法中, 因为需要计算每个抗体的信息熵, 使得算法执行时间较长. 同时在实验中发现随着种群数量的增加, 信息熵的计算量成爆炸式增长, 这大大阻碍了基于混沌理论和信息熵的检测器生成算法的效率, 相比之下, 本文提出的算法的优势就显示出来.

5 结束语

本文将混沌理论应用在木马检测领域, 提出了一种基于混沌理论和加权欧氏距离的木马检测器生成算法, 并通过实验验证了本文方法的有效性. 本文方法比传统的基于信息熵的木马检测器生成算法效率高, 由于结合了混沌理论能够生成均匀的木马检测器, 从而保证了种群的多样性. 本文算法利用加权欧氏距离提高了计算能力, 对初始检测器的处理能力也大大提高. 下一步的工作是将该算法应用于木马检测系统中, 进一步设置加权欧氏距离的合适的权值来验证该算法的有效性.

(下转第 109 页)

的划分,利用符合人眼视觉特性的高斯加权系数对图像子块进行加权,并进一步考虑到颜色分布在方向上的差异性,增加了纵横两个方向的投影,以增强图像局部颜色特征的表征能力,再通过二值化得到颜色位图作为局部颜色特征。最后,合理地融合了全局颜色分布特征和局部颜色分布特征,综合两者进行图像检索。实验结果表明,本文算法具有良好的检索效果。

参考文献:

- [1] 汪孔桥,沈兰荪,刑昕. 一种基于视觉兴趣性的图像质量评价方法[J]. 中国图像图形学报, 2000, 5(4): 300-303.
- [2] Datta R, Li J, Wang J Z. Content-based image retrieval: approaches and trends of the new age [J]. ACM SIGMM int'l Workshop Multimedia Information Retrieval, 2005: 253-262.
- [3] Smeulders A W M, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2000, 22(12): 1349-1380.
- [4] Swain M J, Ballard D H. Color indexing [J]. International Journal of Computer Vision, 1991, 7(1): 11-32.
- [5] 薛少娟,左万利,赫枫龄. 基于颜色分块全局直方图的图像检索方法及系统实现[J]. 吉林大学学报:理学版, 2006, 44(4): 606-610.
- [6] 宋琳琳,王相海. 一种基于 HVS 加权颜色特征的图像检索算法[J]. 微电子学与计算机, 2008, 25(6): 60-62.
- [7] Yoon J, Kweon I. Color image segmentation considering of human sensitivity for color pattern variations [J]. SPIE, 2001, 4572: 269-278.
- [8] Jin L, Li D. A switching vector median filter based on the CIELAB color space for color image restoration [J]. Signal Processing, 2007, 87(6): 1345-1354.
- [8] Lu T, Chang C. Color image retrieval techniques based on color features and image bitmap [J]. Information Processing and Management, 2007, 43(2): 461-472.
- [9] 何姗,郭宝龙,洪俊标. 基于区域熵的图像检索[J]. 计算机工程, 2006, 32(18): 214-216.
- [10] 高燕,胡学龙,李玲,等. 基于小波多尺度分析的综合特征图像检索 [J]. 微电子学与计算机. 2010(04): 21-24.
- [11] Chan Y, Ho Y, Liu Y, et al. A ROI image retrieval method based on CVAOA [J]. Image and Vision Computing, 2008, 26(11): 1540-1549.

作者简介:

- 林晨航** 男,(1987—),本科生.研究方向为图像检索。
潘志斌 男,(1963—),教授.研究方向为图像编码、图像识别。
邹彬 男,(1986—),硕士研究生.研究方向为图像处理与识别。

(上接第 103 页)

参考文献:

- [1] 张楠,李志蜀,张建华,等. 基于混沌理论的免疫遗传算法[J]. 计算机应用, 2006.
- [2] 王宏霞,何晨,丁科. 基于混沌映射的鲁棒性公开水印[J]. 软件学报, 2004, 15(8): 1245-1251.
- [3] Tavazoei M S, Haeri M. An optimization algorithm based on chaotic behavior and fractal nature[J]. Comput Applied Math, 2007(206): 1070-1081.
- [4] Zuo X Q, Fan Y S. A chaos search immune algorithm with its application to neuro-fuzzy controller design [J]. Chaos Solitons and Fractals, 2006(30): 94-109.
- [5] 舒红平,游志胜,蒋建民. 基于信息熵的决策属性分类挖掘算法及应用[J]. 计算机工程与应用, 2004(1): 186-189.
- [6] Hrycej T. Identifying chaotic attractore with neural networks[C]// Proceedings of IEEE Interational Conference on Neural Network. [s. l.]: IEEE, 1995: 2664-2668.

作者简介:

- 秦杰** 男,(1968—),博士,副教授,硕士研究生导师.研究方向为网络信息安全。
司群 女,(1985—),硕士研究生.研究方向为网络安全。