计算机病毒演化模型及分析

张 瑜[1,2]，李 涛[1]，吴丽华[2]，彭小宁[3]，覃仁超[1]

(1. 四川大学计算机学院 成都 610065; 2. 海南师范大学信息学院 海口 571158;
3. 湖南怀化学院计算机系 湖南 怀化 418000)

【摘要】计算机病毒的演化特性可为反病毒技术的改进与提高提供研究思路，使其防毒于未然。该文借鉴人工生命的思想，从计算机病毒的生命特征层面分析计算机病毒的演化特性，建立了一种基于免疫遗传算法的计算机病毒演化模型，给出了计算机病毒的形式化定义以及计算机病毒演化算子的数学模型，模拟了计算机病毒的繁殖演化过程。仿真实验结果表明，计算机病毒及其演化将长期存在。最后，从计算机病毒演化的角度讨论了计算机病毒的防御策略。

关 键 词 人工生命; 计算机病毒; 演化; 遗传算法
中图分类号 TP393 **文献标识码** A doi:10.3969/j.issn.1001-0548.2009.03.024

Computer Virus Evolution Model and Its Analysis

ZHANG Yu[1,2], LI Tao[1], WU Li-hua[2], PENG Xiao-ning[3], and QIN Ren-chao[1]

(1. College of Computer Science, Sichuan University Chengdu 610065;
2. College of Information Science and Technology, Hainan Normal University Haikou 571158;
3. Department of Computer Science, Huaihua University Huaihua Hunan 418000)

Abstract Computer viruses play extremely important roles in the anti-virus industry, because their existence makes software developers pay more attention to security and develop anti-virus technology. Therefore, computer viruses and their evolutions are worthy of thoroughly studying in scientific research sense. To further investigate the evolution of computer viruses, an immune genetic algorithm based model for computer viruses evolution is proposed, which draws inspirations from artificial life. The formal definition of computer virus is introduced, and the evolution operators including selection, crossover, inversion, and immune operator are presented. The simulation experiments indicates that computer viruses have enormously potential capability of self-propagation and self-evolution. Some defense strategies are discussed focusing on preventing unknown computer viruses.

Key words artificial life; computer virus; evolution; genetic algorithm

Internet信息传输的便捷性以及计算机软硬件系统的设计漏洞，为计算机病毒演化提供了物质基础[1-2]。计算机病毒的泛滥已经造成了巨大的经济损失。在病毒与反病毒的长期斗争中，反病毒技术通常滞后于病毒技术。为提高反病毒技术对病毒变种或未知病毒的检测率，从而有效地保障信息系统安全，就必须对计算机病毒的演化发展规律进行详细的研究[3]。

计算机病毒[4]作为一种计算机程序，具有算法特征。此外，计算机病毒还是一种可能的人工生命体[5]，具有生命特征。文献[6]从图灵机的可自我复制性的角度讨论了计算机病毒进行自我复制、自我演化的特性。文献[7]和[8]从人工生命体的角度讨论了计算机病毒的演化特性。

本文为验证计算机病毒是否具有自然生物病毒自我复制、自我构造和自我进化的生物学特征，建立了一种基于免疫遗传算法的计算机病毒演化模型。仿真实验表明，既具有算法特征又具有生命特征的计算机病毒具有巨大的演化潜力。因此计算机病毒及其演化还将长期存在，而对于计算机病毒的防御、检测以及清除仍将是网络安全领域的挑战性问题。

1 模型理论

1.1 计算机病毒的生命特征

计算机病毒修改宿主程序，并将自身的精确拷贝或者其演化的拷贝插入其中，从而感染该宿主程序。由于这种感染特性，病毒可随信息流的扩散而

收稿日期：2008－03－10；修回日期：2008－10－25
基金项目：国家自然科学基金(60573130、66873246)；国家863计划(2006AA01Z435)；教育部博士点基金(20070610032)
作者简介：张 瑜(1975－)，男，博士生，讲师，主要从事网络安全和计算智能等方面的研究.

传播，从而破坏信息的完整性。

计算机病毒的定义表明，计算机病毒与生物病毒是两个不同范畴的概念。前者是人工制造，后者是自然产物；前者是机器编码，后者是核酸编码；前者是物理存储指令，后者以化学存储方式为主[8]。尽管如此，两者在功能上以及危害和感染的本质上是一致的。因此，计算机病毒几乎具有生物病毒全部的生物学特征。从这个意义上而言，计算机病毒是一种可能的人工生命体，即人工病毒，它具有自我繁殖、自我构造、自我进化等生命特征。

1.2 计算机病毒的生命周期

计算机病毒是一种可能的人工生命体，其生命周期可分为4个阶段，即新病毒的产生、病毒传播及潜伏、病毒触发运行破坏和病毒被反病毒程序查杀，如图1所示。

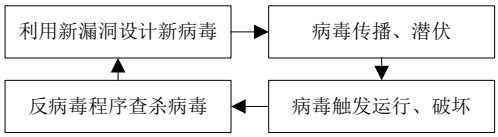

图1 计算机病毒的生命周期

从计算机病毒的生命周期来看，病毒变种一般产生于病毒传播、潜伏过程中。本文提出的基于免疫遗传算法的计算机病毒演化模型，是根据计算机病毒的生命特征和生命周期，研究计算机病毒的演化特性，揭示其演化发展规律。

1.3 计算机病毒的逻辑结构

计算机程序的自我复制性、计算机网络的共享性以及计算机软硬件系统设计上的漏洞，为计算机病毒的产生与发展提供了物质基础，也决定了计算机病毒的结构。计算机病毒的这种结构也是其充分利用系统资源进行破坏活动的最合理体现[9-10]。计算机病毒一般由感染标记、初始化模块、感染模块和表现模块组成，如图2所示。

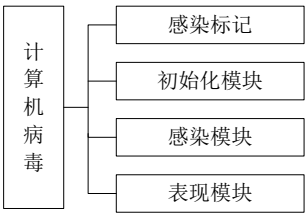

图2 计算机病毒的逻辑结构

由此，计算机病毒可形式化定义为一个四元组：

$$V=(V_1,V_2,V_3,V_4)$$

式中 V_1表示感染标记集合；V_2表示初始化模块集合；V_3表示感染模块集合；V_4表示表现模块集合。而每个集合又可用n个指标的特征向量来表示，即：

$$\boldsymbol{V}_i=(v_{1i},v_{2i},\cdots,v_{ni})^{\mathrm{T}}$$

故计算机病毒可表示为$n\times4$阶矩阵，即：

$$\boldsymbol{V}=\begin{pmatrix} v_{11} & v_{12} & v_{13} & v_{14} \\ v_{21} & v_{22} & v_{23} & v_{24} \\ \vdots & \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} & v_{n3} & v_{n4} \end{pmatrix}$$

1.4 计算机病毒演化模型

基于计算机病毒的算法特征和生命特征，病毒变种或未知病毒一般诞生于已知病毒的演化之中。病毒编写者在制造新病毒时，通常采用如下方式：

(1) 对已知病毒进行编码分析；

(2) 提取病毒的各种模块；

(3) 运用不同的算法对已知病毒的模块进行组合，得到新的病毒。

借鉴生物学中染色体的概念，将计算机病毒结构表示为基因模式。计算机病毒结构为一长度为4的染色体：

$$\mathrm{VC}=(\mathrm{VC}_1,\mathrm{VC}_2,\mathrm{VC}_3,\mathrm{VC}_4)$$

第i位基因上的等位基因为：

$$\mathrm{VC}_i=(\mathrm{vc}_{1i},\mathrm{vc}_{2i},\cdots,\mathrm{vc}_{n_i}) \quad i=1,2,3,4$$

式中 n表示第i位基因的等位基因数量。设计算机病毒$V=(V_1,V_2,V_3,V_4)$，则其演化变异集合V_{mutation}为一笛卡尔集，即：

$$V_{\mathrm{mutation}}=\prod_{i=1}^{4}V_i=V_1\times V_2\times V_3\times V_4=$$
$$\{(v_1,v_2,v_3,v_4)\,|\,\forall i\leqslant 4, v_i\in V_i\}$$

本文将病毒演化过程中所呈现的算法抽象为选择算子、交叉算子、逆转算子和免疫算子等4个算子，且分别进行建模。

1.4.1 选择算子模型

病毒群体通过选择算子的演化过程描述如下：

$$V_{\mathrm{select}}(t)=\begin{cases} V_{\mathrm{initial}} & t=0 \\ V_{\mathrm{select}}(t-1)-V_{\mathrm{del}}(t) & t\geqslant 1 \end{cases} \tag{1}$$

$$V_{\mathrm{del}}(t)=\{v\,|\,v\in V_{\mathrm{select}}(t-1)\wedge f_{\mathrm{adaption}}(v)=0\} \tag{2}$$

$$f_{\mathrm{adaption}}(v)=\begin{cases} 0 & \text{被检测} \\ 1 & \text{否则} \end{cases} \tag{3}$$

式中 V_{initial}是病毒的初始化集合；V_{del}是为反病毒软件检测出而被删除的病毒集合。病毒的适应度f_{adaption}定义为是否被反病毒软件检测到的能力，如被检测到则其适应度为0，反之则为1。本文模型模拟了遗传算法的选择过程，即从当前病毒群体中选择适应度高的病毒个体，构成下一代病毒群体的交配池。

1.4.2　交叉算子模型

病毒群体通过交叉算子的演化描述如下：

$$V_{\text{crossover}}(t)=\begin{cases}V_{\text{select}} & t=0\\ V_{\text{crossover}}(t-1)-V_{\text{del}}(t)+V_{\text{new}}(t) & t\geqslant 1\end{cases} \quad (4)$$

$$V_{\text{del}}(t)=\{v\mid \exists v\in V_{\text{crossover}}(t-1)\wedge |V_{\text{del}}(t)|=2\} \quad (5)$$

$$V_{\text{new}}(t)=\{v\mid \exists\, x,y\in V_{\text{crossover}}(t-1)\wedge (v\in \text{Crossover}(x,y))\wedge |V_{\text{new}}(t)|=2\} \quad (6)$$

$$\text{Crossover}(x,y)=\{(x,y)\mid x.v_i\leftrightarrow y.v_j, 1\leqslant i,j\leqslant 4\} \quad (7)$$

式中　V_{select} 是选择算子操作后产生的交配池；V_{del} 是选作交叉的一对个体；V_{new} 是交叉操作后生成的一对个体。式(7)模拟了交叉算子生成一对新病毒的过程，即在两个病毒中随机选择相应的基因座上的等位基因相互交换，进行基因重组，便产生了包含更复杂基因结构的新病毒。

1.4.3　逆转算子模型

病毒群体通过逆转算子的演化过程描述如下：

$$V_{\text{inversion}}(t)=\begin{cases}V_{\text{crossover}} & t=0\\ V_{\text{inversion}}(t-1)+V_{\text{new}}(t) & t\geqslant 1\end{cases} \quad (8)$$

$$V_{\text{new}}(t)=\{v\mid \exists x\in V_{\text{inversion}}(t-1)\wedge (v\in \text{Inversion}(x))\} \quad (9)$$

$$\text{Inversion}(x)=\{x\mid x.v_i\leftrightarrow x.v_j \quad 1\leqslant i;j\leqslant 4\} \quad (10)$$

式中　V_{new} 是病毒进行逆转操作后生成的新病毒。式(10)模拟了逆转算子生成新病毒的过程，即随机选择任一病毒，通过两个基因座上的基因进行交换(变换了其特征码)，便生成了新的病毒。

1.4.4　免疫算子模型

病毒群体的自我保护机制描述如下：

$$V_{\text{immune}}(t)=\begin{cases}V_{\text{inversion}} & t=0\\ V_{\text{immune}}(t-1)-V_{\text{del}}(t) & t\geqslant 1\end{cases} \quad (11)$$

$$V_{\text{del}}(t)=\{v\mid \exists v\in V_{\text{immune}}(t-1), \exists x\in V_{\text{sig}}, \text{Match}(x,v)=1\} \quad (12)$$

$$\text{Match}(x,y)=\begin{cases}1 & \exists i;1\leqslant i\leqslant 4;\ x.v_i=y.v_i\\ 0 & \text{否则}\end{cases} \quad (13)$$

式中　V_{del} 是病毒群体经过自体耐受过程而被删除的病毒集合，即如一个病毒的特征码与反病毒软件相匹配，则会被删除被淘汰。而存活下来的病毒会继续繁殖、进化，以提高病毒后代的生存能力。

2　仿真实验及分析

本文通过模拟计算机病毒可能的演化，验证了计算机病毒具有生物病毒的演化特性以及其存在的长期性。

2.1　实验数据集与实验环境

本文所用病毒是从WildList中选取了10种Windows脚本病毒。为简化实验设计，每种病毒只提取其感染模块或表现模块的基因，最后的病毒基因库由10个病毒基因组成。

实验分为10组进行，第一组从病毒基因库随机选取1个基因，第二组从病毒基因库随机选取2个基因，依此类推，第十组选取全部10个基因。每组都生成100个病毒样本进行实验。实验环境如表1所示。

表1　病毒演化实验平台

系统平台	反病毒软件	病毒基因库规模	Windows脚本病毒样本
Windows XP (2.4 GHz、512 MB)	瑞星2008	10	Soraci, Areses!ITW#48, Redlof.C, Gedza, Solow!ITW#1, Solow!ITW#10, Solow!ITW#20, Yosenio!ITW#1, Z0X!ITW#1, Redlof.A-m

2.2　实验结果及分析

从图3所示结果可知，当病毒基因规模变大时，经过演化而生成的未知病毒或病毒变种的存活数目也随之增多。从理论上分析，随着病毒样本数的增加，病毒基因库的增大，则病毒染色体的特征码空间随之增大，应用选择算子、交叉算子、逆转算子和免疫算子而新生成病毒的存活概率也随之增大。假设病毒基因库由n个基因组成，随机选择m个基因进行重组，则会产生 P_n^m 种不同的未知病毒或病毒变种。

由于病毒染色体的随机组合概率演化，在演化过程中会生成一些无作用病毒，但计算机病毒总演化趋势仍是病毒基因库不断增大、病毒染色体空间不断扩展以及新病毒或已知病毒变种不断增加。计算机病毒的这种演化特性以及计算机病毒存在的长期性，将是反病毒技术的严峻挑战。

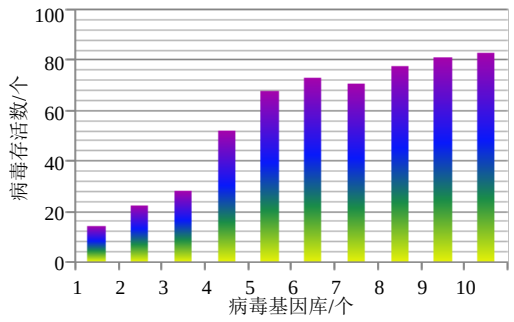

图3　病毒基因库规模与存活病毒数目的关系

2.3　计算机病毒的防御策略

通过以上模型的分析，不难发现计算机病毒具有代码变异的特点。为此，本文从计算机病毒代码变异的角度提出相应的防御策略。

当前传统的反病毒技术是基于病毒特征码的，它对于已知病毒的检测率高，但对于未知病毒或已知病毒的变种的检测率低。由于计算机病毒基因片段长度短小且是计算机病毒的原子单位，所以，可采取只检测现有计算机病毒基因片段而非每个病毒特征码的方式。另外，对每个计算机病毒基因片段赋予相应的权重。由此推断，任何由现有的计算机病毒变异而成的新病毒都能被检测出，从而达到启发式的智能化查杀未知病毒的目的。

3 结 论

本文在研究计算机病毒的生命特征及逻辑结构的基础上，建立了一种基于免疫遗传算法的计算机病毒的演化模型，揭示了计算机病毒一种可能的演化特性。仿真实验表明了计算机病毒及其演化存在的长期性。该模型为反病毒技术的改进与提高提供了一种研究思路，使其未雨绸缪，研究反制之道，防毒于未然，更有效地保障信息系统安全。反病毒技术可运用免疫遗传原理去检测病毒基因片段，以达到智能化查杀未知病毒的目的。

本文的研究工作得到了海南师范大学引进教授(博士)科研启动项目的支持，在此表示感谢。

参 考 文 献

[1] FORD R, EUGENE H S. Happy birthday, dear viruses[J]. Science, 2007, 317: 210-211.
[2] BALTHROP J, FORREST S, NEWMAN M E J, et al. Technological networks and the spread of computer viruses[J]. Science, 2004, 304: 527-529.
[3] NACHENBERG C. Computer virus-antivirus coevolution[J]. Communications of the ACM, 1997, 40(1): 46-51.
[4] COHEN F. Computational aspects of computer viruses[J]. Computers & Security, 1989, 8: 325-344.
[5] EUGENE H S. Computer viruses—a form of artificial life?[R]. Indiana, USA: Purdue University, 1990.
[6] COHEN F. Computer viruses: theory and experiments[J]. Computers & Security, 1987, 6: 22-35.
[7] MARK A L. Computer viruses, artificial life and evolution[M]. Tucson, Arizona: American Eagle PublicationsInc, 1993.
[8] 郝宁湘. 计算机病毒: 一种可能的生命形式[EB/OL]. [2008-02-25]. http://cyborg.bokee.com/2586799.html.
HAO Ning-xiang. Computer virus: a possible form of life [EB/OL]. [2008-02-25]. http://cyborg.bokee.com/ 2586799. html.
[9] SZOR P. The art of computer virus research and defense [M]. California: Symantec Press, 2005.
[10] 张仁斌, 李 钢, 侯整风. 计算机病毒与反病毒技术[M]. 北京: 清华大学出版社, 2006.
ZHANG Ren-bin, LI Gang, HOU Zheng-feng. Computer virus and antivirus technology[M]. Beijing: Tsinghua University Press, 2006.

编 辑 熊思亮

·我校科研成果介绍·

相干式CO_2激光跟踪雷达系统

相干式CO_2激光跟踪雷达系统具有以下特点：

(1) 采用CO_2激光管PZT线性调频及峰值稳频兼容技术，简化了光学头的设计，省去了腔外调制器及附属电路。

(2) 采用迈克逊干涉仪自差光路，用四象限探测器提取信号，达到四路均匀输出并产生误差信号，保证了作用距离。

(3) 采用硬件快速傅里叶变换及多次平均测距、测速技术。

(4) 采用窄带跟踪滤波误差信号提取技术。

(5) 采用小型化光学天线结构设计。

(6) 采用“边缘跟踪”法，实现了四象限探测器大目标跟踪。