文章编号:1001 - 2486(2005)06 - 0130 - 05

基于遗传算法的信息系统安全技术方案*

陈　光,匡兴华

(国防科技大学 信息系统与管理学院,湖南 长沙　410073)

摘　要:组织如何制定优化的信息安全技术方案以降低脆弱性对其信息系统的威胁,是信息安全管理领域的关键问题。在描述信息系统安全技术方案决策模型的基础上,提出了一种求解信息系统安全技术方案优化问题的自适应遗传算法,使得组织能以最少的方案实施费用最大限度地处置脆弱性,并以实例说明了该算法的有效性。

关键词:信息系统;安全技术方案;遗传算法;脆弱性

中图分类号:TP309;G203　　文献标识码:B

The Information System Security Technology Scheme Based on Genetic Algorithm

CHEN Guang, KUANG Xing-hua

(College of Information System and Management, National Univ. of Defense Technology, Changsha 410073, China)

Abstract: How can the organizations form optimum system security technology schemes to reduce the threat caused by vulnerabilities of their information systems is the key problem in information security management field. On the basis of describing the multi-goal decision-making model of the information system security technology scheme, this paper presented a self-adaptive genetic algorithm for security technology scheme of the information system enabling the organizations to choose the minimal-cost security technology scheme that can address the maximum vulnerabilities. And examples are given to demonstrate the validity of the algorithm.

Key words: information system; security technology scheme; genetic algorithms; vulnerability

随着计算机与通信技术的不断发展,信息系统已经成为组织赖以生存的重要战略资源。然而,信息系统本身存在大量脆弱性——由于系统硬件、软件或者安全策略上的错误而引起的缺陷或安全隐患。这些脆弱性一旦被恶意利用,将会对组织的信息安全产生严重的威胁[1]。因此,制定优化的安全技术方案,以减轻脆弱性对信息安全的威胁,对控制组织信息安全风险具有重要意义。

以往对信息系统安全技术方案的研究大多是从技术角度展开的。当前,有研究指出,应将经济学思想引入到信息系统安全管理实践中,追求安全投入 ROI(投资回报率)最大化[2]。因而,在制定安全技术方案时,必须综合考虑脆弱性、安全技术及其脆弱性处置能力,以及安全技术费用等多种因素,以最少的方案实施总费用,最大限度地降低脆弱性对信息安全的威胁。这样,信息系统安全技术方案问题不再局限于单纯的技术层面,而是成为一个非常复杂的管理决策问题。人们往往难于直接解决,而必须依赖于有效的决策工具求解。为此,本文在描述信息系统安全技术方案决策模型的基础上,提出了一种求解安全技术方案问题的遗传算法,并验证了该算法的有效性。

1　信息系统安全技术方案问题的决策模型

信息系统安全技术方案问题可以表述如下:

设信息系统的脆弱性分布情况为:$V=(v_1,v_2,\cdots,v_m)$,m 为信息系统可能具有的脆弱性的种类数,$v_i(i=1,\cdots,m)$表示第 i 种特定脆弱性,可以取两个值 1 或 0。值 1 表示组织的信息系统存在第 i 种脆

* 收稿日期:2005 - 06 - 09
基金项目:国家自然科学基金资助项目(70272002);国防科技大学校预研基金资助项目(JC02 - 00 - 024)
作者简介:陈光(1978—),女,博士生。

弱性,值 0 表示不存在。各种脆弱性会在不同程度上对信息系统安全造成威胁。信息安全管理人员可通过分析信息系统的脆弱性权图或攻击图[3],采用专家评分法、层次分析法(AHP)、熵值法等方法,为每种脆弱性 v_i 赋予一个权重 a_i($\forall\ i=1,\cdots,m$),表示其被利用而对信息安全产生威胁的严重程度,满足:$a_1+a_2+\cdots+a_m=1$。

表 1 给出了大多数信息系统可能具有的一组脆弱性[4],该分类方法的 m 值等于 20。m 值可随着新的信息系统脆弱性种类的发现而增加。如果采取不同的脆弱性分类与描述标准,m 的取值也会不一样。

组织可选择安全技术来处置(address)脆弱性[4]。假设可供组织选择的安全技术有 n 种,组织采用的安全技术方案可以表示为:$S=(s_1,s_2,\cdots,s_n)$,其中 s_j($j=1,\cdots,n$)表示第 j 种安全技术,可以取两个值:1 或 0。1 表示安全技术方案纳入了第 j 种安全技术,0 表示没有纳入;实施安全技术 s_j 所需费用为 c_j。

表 2 给出了一组能用以处置脆弱性的通用信息安全技术[4]。按照这种分类,则 n 等于 13。同样,随着新开发的安全技术的出现以及采取不同的技术分类标准,n 的取值也可以作相应变化。

前述"处置"一词反映了安全技术与脆弱性之间的复杂关系,即采用某种安全技术可能同时带来正面和负面的效果:一方面全部或部分消除了某些脆弱性,另一方面又可能导致或部分导致某些脆弱性。所谓部分消除,指安全技术能减少某种脆弱性的传播范围并限制其破坏程度,但是安全技术不能消除导致脆弱性产生的根本因素。根据文献[4]以及实践经验,表 3 给出了各种安全技术处置各脆弱性的能力(安全技术和脆弱性分类同表 1 和表 2):其中 1 表示全部消除脆弱性;0.5 表示可部分消除脆弱性;-1 表示可导致脆弱性;-0.5 表示可部分导致脆弱性。这里,忽略安全技术部分消除或部分导致各脆弱性的程度的细微差别,并假设它们所产生的效果均为直接效果的一半。

表 1　脆弱性

Tab.1　Vulnerabilities

分　类	脆弱性	表示
固有设计/结构	唯一性	v_1
	单一性	v_2
	集中性	v_3
	可分离性	v_4
	同质性	v_5
行为复杂度	敏感性	v_6
	可预见性	v_7
适应性与处理	刻板,僵化	v_8
	顺应性	v_9
	易受骗性	v_{10}
操作/配置	容量限制	v_{11}
	不可恢复性	v_{12}
	缺乏自我认识性	v_{13}
	难于管理	v_{14}
	互操作性	v_{15}
非直接/非物理风险	电子可访问性	v_{16}
	透明性	v_{17}
直接/物理风险	物理可访问性	v_{18}
	电磁敏感性	v_{19}
支撑设备/基础设施	相关性	v_{20}

当实施所制定的安全技术方案后,组织信息系统仍具有的脆弱性称为残留脆弱性,可表示为:$R=(r_1,r_2,\cdots,r_m)$。R 来自于没有被安全技术消除或被部分消除的脆弱性,以及由安全技术导致或部分导致的脆弱性。其中,r_i 是与脆弱性向量 V 中 v_i 相同种类的脆弱性。r_i 可以取三个值:1,0.5 和 0,1 表示存在该脆弱性,0 表示不存在,0.5 表示该脆弱性是被部分消除的或部分导致的。

安全技术对脆弱性的处置情况可以用矩阵 $\boldsymbol{T}=\{t_{ij}\}$($\forall\ i=1\cdots m,\forall\ j=1\cdots n$)表示,$t_{ij}$ 对应于表 3 中 s_j 对 v_i 的处置能力的值。则确定 R 的规则如下(规则与应用顺序无关):

(1)如果 $v_i=0$ 或 1,存在 $j\in\{j\mid s_j=1\}$ 使得 $t_{ij}=1$,则 $r_i=0$;

(2)如果 $v_i=0,\forall\ j\in\{j\mid s_j=1\}$ 有 $t_{ij}=0$,则 $r_i=0$;

(3)如果 $v_i=1,\forall\ j\in\{j\mid s_j=1\}$ 有 $t_{ij}=0$,则 $r_i=1$;

(4)如果 $v_i=0,\forall\ j\in\{j\mid s_j=1\}$ 有 $t_{ij}\neq 1$,存在 $j\in\{j\mid s_j=1\}$ 使得 $t_{ij}=-1$,则 $r_i=1$;

(5)如果 $v_i=0,\forall\ j\in\{j\mid s_j=1\}$ 有 $t_{ij}\neq 1$,存在 $j\in\{j\mid s_j=1\}$ 使得 $t_{ij}=-0.5$,则 $r_i=0.5$;

(6)如果 $v_i = 1, \forall j \in \{j \mid s_j = 1\}$ 有 $t_{ij} \neq 1$,存在 $j \in \{j \mid s_j = 1\}$ 使得 $t_{ij} = 0.5$,则 $r_i = 0.5$。

组织所选择的信息安全技术方案最终必须实现两个目标:

目标一,信息系统残留脆弱性被利用对信息安全总的威胁最小,即:

$$\min_s R, R = \sum_{i=1}^{m} a_i r_i$$

目标二,实施该安全技术方案所需的总费用最少,即:

$$\min_s E, E = \sum_{j=1}^{n} c_j s_j$$

这是一个多目标优化问题,可以使用目标的加权和的方法,构造合成的优化目标函数:$\min Z = \alpha R + \beta E'$,其中:$E' = \sum_{j=1}^{n} \frac{c_j}{c_{\max}} s_j$。

$c_{\max}$是各种安全技术费用中的最大值。E'是对 E 的无量纲修正,可避免 R 与 E 的决策值在数量级上不一致。α 和 β 是权重,表示组织对目标一与目标二的偏好($\alpha + \beta = 1$ 且 $0 \leqslant \alpha, \beta \leqslant 1$)。若组织认为消除脆弱性的危害比控制安全技术费用更加重要,可令 $\alpha > \beta$;反之,若组织认为在有限的安全预算下,控制安全技术的费用更重要,则可令 $\alpha < \beta$。由前述可知值 $0 \leqslant Z \leqslant 1$。

表2 信息安全技术

Tab.2 Security technologies

安全技术	表示
异质性技术	s_1
静态资源分配	s_2
动态资源分配	s_3
冗余技术	s_4
弹性和鲁棒性技术	s_5
快速恢复与重组	s_6
欺骗技术	s_7
分段、分散与隔离	s_8
免疫学识别技术	s_9
自组织与集体行为技术	s_{10}
人员管理	s_{11}
信息资源的集中管理技术	s_{12}
威胁/告警响应组织技术	s_{13}

表3 安全技术对脆弱性的处置能力

Tab.3 The ability of security technologies

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}
v_1			1		1	1			1	-1			
v_2		1	1	1	0.5	1	1	1		1			
v_3		-0.5			0.5	1	1	1		1		-1	
v_4		-1	1	1		1		-1		1			
v_5	1			-1	0.5			1				-0.5	
v_6	-1		-1		0.5	1	-1	1	-1	-1			
v_7	1		1				1	-1	1			-0.5	
v_8		-1						-1	1	1			0.5
v_9	1						1		1		0.5		0.5
v_{10}			-1						1				0.5
v_{11}			-1										0.5
v_{12}			1	1	0.5	1		1		1		0.5	0.5
v_{13}									1		0.5	0.5	
v_{14}	-1		-1				-1	1	-1	1	1	1	-0.5
v_{15}		0.5			-0.5		-1	1	1	-0.5	1		1
v_{16}		1	1				1	1	1		0.5		1
v_{17}						1						1	
v_{18}		1	1	1	1	1	1	1			0.5	1	
v_{19}	1	1		1	1	1	1	1					0.5
v_{20}	1		1	1	1			1		1		-0.5	0.5

2　信息安全技术方案的遗传算法

组织的安全技术方案是其可选的 n 种安全技术集合的子集,具有 2^n 个可行解,且问题解空间随 n 增加呈几何级数增长。针对选取安全技术方案这一多目标优化问题,必须采用高效的优化策略来搜索最优解。本文提出一种自适应的遗传算法来求解安全技术方案问题。该遗传算法给出一种让遗传过程中每一代各个个体能按其适应度大小自动选择不同的交叉概率和变异概率的自适应规则。这样,群体中每个个体对环境的变化就具有自适应调节能力[5]。

2.1　问题编码

根据遗传算法,每条染色体可以表示为一个 n 位(bit)二进制编码符号串,表示一种安全技术方案。染色体中的每一位对应于一种备选安全技术的状态:1 表示安全技术方案采用了该种安全技术,0 表示不采用。这样,安全技术方案优化问题就转换成利用遗传算法求解最优染色体编码的问题。

2.2　适应度函数设计

如下设置适应度函数,可将安全技术方案 S^k 的目标函数最小化问题转化为最大化问题,其中 γ 为大于 1 的足够大的常数:$F=\gamma-Z=\gamma-(\alpha\sum_{i=1}^{m}\alpha_i r_i+\beta\sum_{j=1}^{n}\frac{c_j}{c_{\max}}s_j)$。

2.3　遗传算子设计

(1)选择算子

首先采用优异个体选择策略,设种群规模为 N,优异个体选择比率为 P_e。求当前代中所有个体的适应度值,将个体按适应度值由大到小排序,将排在前面 $N\times P_e$ 个个体直接复制作为下一代的个体。然后运用赌轮选择方法在当前代剩下的个体中选择一对个体作为下一代的个体,各个体的选择概率与其适应度值成比例。该选择过程不断重复,直至下一代个体数目达到 N 时结束。

(2)交叉算子

对于前述用赌轮选择方法选取的每对个体,依照交叉概率 P_c 进行交叉,交叉位固定选择为个体符号串的中点。设 P_{c0}为较差个体的交叉概率,P_{c1}为最优个体的交叉概率($P_{c1}<P_{c0}$),则自适应交叉概率 P_c 的表示式如下:

$$P_c=\begin{cases}P_{c0}, & f\leqslant f_{\text{aver}}\\ P_{c1}\left[\dfrac{P_{c0}}{P_{c1}}\right]^{\left[\frac{f_{\max}-f}{f_{\max}-f_{\text{aver}}}\right]}, & f>f_{\text{aver}}\end{cases}\tag{1}$$

式中,f 是两个执行交叉操作个体适应度中较小的一个,$f_{\max}$是群体中的最大适应度,f_{aver}是群体的平均适应度。

(3)变异算子

前面每对个体经过交叉操作后,接着进行基本位概率变异。在个体的 1 到 n 位的范围内依变异概率 P_m,随机产生若干变异位进行变异。设较差个体的变异概率为 P_{m0},最优个体的变异概率为 P_{m1}($P_{m1}<P_{m0}$),自适应变异概率 P_m 可以由下式表示:

$$P_m=\begin{cases}P_{m0}, & f'\leqslant f_{\text{aver}}\\ P_{m1}\left[\dfrac{P_{m0}}{P_{m1}}\right]^{\left[\frac{f_{\max}-f'}{f_{\max}-f_{\text{aver}}}\right]}, & f'>f_{\text{aver}}\end{cases}\tag{2}$$

式中,f' 是执行变异操作个体的适应度。

由公式知,交叉概率的自适应调整,就是对适应度小的个体采用较大的交叉概率,以加速其更新速度;对适应度大的个体采用较小的交叉概率,使这些较优的解不易丢失。而由公式可知,若群体中的个体趋于相同时,个体的适应度就接近群体的平均适应度,每个个体的变异概率也较大,可以很快得到新个体,从而保证了群体的多样性,并克服了算法的早熟现象。P_{c0}和 P_{c1}的取值范围一般在 0.2 ~ 0.7 之

间,而 P_{m0}和 P_{m1}的取值范围一般在 0.1~0.5 之间。

2.4 遗传算法求解过程

将随机产生的一定规模的安全技术方案个体作为初始种群,即初始代。自初始代开始,重复进行上述遗传操作产生新的子代,直到产生若干连续稳定的子代,这些子代的最优个体的适应度没有显著的改善。所谓最优个体即当前代中适应度值最优的个体。

3 算 例

采用表 1 和表 2 分类方法,则 $m = 20$, $n = 13$。假设信息系统脆弱性分布情况为:V = (11010100110111000110),且由专家给出的脆弱性权重如表 4 所示,各安全技术费用权重值($c_j / c_{\max}$)如表 5 所示。且系统管理人员认为处置脆弱性与控制安全技术方案的费用具有同等重要性,即 $\alpha = \beta = 0.5$。

表 4 脆弱性权重

Tab.4 The weights of vulnerabilities

v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}	v_{15}	v_{16}	v_{17}	v_{18}	v_{19}	v_{20}
0.05	0.09	0.005	0.06	0.11	0.027	0.014	0.08	0.04	0.03	0.1	0.015	0.04	0.06	0.02	0.14	0.026	0.013	0.07	0.01

运用本文所提出的自适应遗传算法求解相应的安全技术方案。种群规模 $N = 100$;优异个体选择比率 $P_e = 40\%$;$P_{c0} = 0.5$, $P_{c1} = 0.3$;$P_{m0} = 0.4$, $P_{m1} = 0.2$;最大迭代次数为 300 代,最大稳定代数为 50 代。求解结果为 S = (1000010000010),求解结果表明:当组织的信息系统存在的脆弱性为 v_1、v_2、v_4、v_6、v_9、v_{10}、v_{12}、v_{13}、v_{14}、v_{18}和 v_{19}时,组织应制定包括安全技术 s_1、s_6 和 s_{12}的安全技术方案。

表 5 安全技术的费用权重

Tab.5 The weights of security technologies' cost

s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}
0.08	0.04	0.1	0.03	0.13	0.06	0.17	0.05	0.12	0.07	0.04	0.09	0.02

此外,为了验证该算法解决此类问题的有效性,分别给出不同的初始数据进行了计算(因篇幅所限,不作列示)。结果表明,当脆弱性种类和候选安全技术种类数增加时,该算法的优越性更明显。

4 结 论

实践证明,本文所提出的求解信息系统安全技术方案的自适应遗传算法,对于解决信息系统安全技术方案决策问题具有很好的效果。在未来的研究中,可以将这种基于遗传算法的方法合成到决策支持工具中去,为信息安全管理人员制定信息安全技术方案提供决策支持,帮助组织在可接受的费用范围内降低信息系统脆弱性所导致的信息安全威胁。

参 考 文 献:

[1] Peltier T R. Information Security Risk Analysis[M]. CRC Press LLC,2001.

[2] Mercuri R T. Analyzing Security Costs[J]. Communications of the ACM, 2003, 46(6):15-18.

[3] Ritchey R O, Berry B, Noel S. Representing TCP/IP Connectivity for Topological Analysis of Network Security[A]. Proceedings of the 18th Annual Computer Security Applications Conference[C], California, 2002:25-31.

[4] Anderson R H, Feldman P M, Gerwehr S, et al. Securing the U.S. Defense Information Infrastructure: A Proposed Approach[A]. RAND Document MR-993-OSD/NSA/DARPA[C],1999.

[5] Coello C A C. An Updated Survey of GA-based Multi-objective Optimization Techniques [J]. ACM Computing Surveys,2002,32(2):109-143.