ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.19, No.10, October 2008, pp.2746−2753
DOI: 10.3724/SP.J.1001.2008.02746

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

一种基于攻击图的入侵响应方法*

石 进[1+], 郭山清[2], 陆 音[1], 谢 立[1]

[1](南京大学 计算机科学与技术系,江苏 南京 210093)

[2](山东大学 计算机科学与技术学院,山东 济南 250101)

An Intrusion Response Method Based on Attack Graph

SHI Jin[1+], GUO Shan-Qing[2], LU Yin[1], XIE Li[1]

[1](Department of Computer Science and Technology, Nanjing University, Nanjing 210093, China)

[2](School of Computer Science and Technology, Shandong University, Ji'nan 250101, China)

+ Corresponding author: E-mail: zgnjack@gmail.com

Shi J, Guo SQ, Lu Y, Xie L. An intrusion response method based on attack graph. *Journal of Software*, 2008, 19(10):2746−2753. http://www.jos.org.cn/1000-9825/19/2746.htm

Abstract: System incentive and alternation of attacker's strategies are not taken into full consideration in current intrusion response research. An intrusion response model (intrusion response based on attack graph, IRAG) based on the attack graph is proposed to solve this problem. IRAG model deals well with the attack's intent and alternation of strategies, and takes account of incentives of system and attacker across-the-board. The experimental results show that the IRAG model can effectively improve the accuracy and effectiveness of alert response.

Key words: attack graph; information security; intrusion response; security metric; game theory

摘 要: 针对当前入侵响应工作中存在的不能充分考虑系统的收益,以及不能充分考虑攻击者策略变化因素等问题,提出了一种基于攻击图的入侵响应 IRAG(intrusion response based on attack graph)模型.该模型较好地解决了攻击意图及策略变化的问题,并全面考虑了系统、攻击者的收益等因素.实验结果表明,IRAG 模型有效地提高了响应的准确性和效果.

关键词: 攻击图;信息安全;入侵响应;安全尺度;博弈理论

中图法分类号: TP393 **文献标识码**: A

自从20世纪80年代Anderson[1]首次提出入侵检测概念以来,入侵检测系统(intrusion detection system,简称IDS)作为网络安全的一个组件获得了极大的发展.但与防火墙、VPN 等安全组件发挥着越来越重要的作用相比,IDS 的作用还未能真正体现出来,主要原因是报警响应问题未能得到很好的解决.因为随着攻击手段的改进,攻击越来越朝向自动化、复杂化的方向发展,而目前的响应则主要以人工为主,这种不对称性使得入侵检测和响应领域的工作陷入了被动的局面,为了解决这个问题,人们开始了自动或半自动响应方式的研究.

这类研究首先是从静态映射型[2,3]响应方式开始的,即按一定的原则对攻击进行分类,并用人工的方式将每

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2003AA144010 (国家高技术研究发展计划(863)); the High-Tech Research Plan of Jiangsu Province of China under Grant No.BG2005029 (江苏省高技术研究计划)

Received 2007-07-30; Accepted 2008-02-25

一报警映射到一个预先定义好的响应措施上,目前的很多入侵响应系统(intrusion response system,简称 IRS)正是基于这种响应方式.静态映射型入侵响应很大程度上解决了人工响应时间过长、负担过重的问题,但是它也有一些很明显的缺点:一方面,易于被攻击所利用;另一方面,它没有充分考虑入侵响应的适应性,响应措施的选择应该随着网络环境的变化而变化.于是人们开始了自适应映射型入侵响应的研究,文献[4,5]提出通过考虑IDS 自身的误报率和以往响应方式的成功率来自动调整入侵和响应措施的映射.

自适应方法解决了响应措施的适应性问题,但是因为没有考虑响应的代价,使得有时响应的效果会得不偿失.针对这个问题,Lee 在文献[6]中提出了一种基于成本分析的入侵检测和响应的参考模型,它先就入侵检测和响应提出了 3 种代价:操作代价、响应代价和损失代价,并在综合考虑这 3 种代价的基础上选择适当的应对措施.文献[7]提出另一种基于成本分析的攻击容忍模型,它使用 IGraph 图对系统可能的入侵路径进行描述,然后根据入侵可能造成的损失和响应代价选择相应的应对措施.但是它们仍然存在一些不足:① 未能充分考虑攻击者和系统的收益偏好,造成所计算的成本不够准确;② 未将攻击者的因素纳入到模型中,仅从系统单方面的角度考虑响应措施对系统利益的最大化,在攻击者改变其攻击策略的情况下,系统原先预计的收益模型就可能不再适用,因而这种仅从系统单方面考虑的最优响应决策是不稳定的.

同时,为了提高报警响应的准确性而开展的攻击关系研究有些也已取得了比较显著的成果,例如报警关联[8,9]、攻击图[10,11]等.报警关联、攻击图均为通过分析各攻击动作之间的关系给出描述攻击次序的报警关联图或攻击图,它们一方面能够降低 IDS 误报并增加报警信息的可读性,同时也为管理员理解整个网络以及做出响应提供了帮助.不过在直接使用报警关联和攻击图进行自动响应决策方面,到目前为止并没有很大的进展,主要原因在于,未能在使用报警关联或攻击图进行响应时很好地处理攻防双方的偏好、收益及策略变化等问题.

本文受攻击关系的启发,提出了一种基于攻击图的动态入侵响应 IRAG(intrusion response based on attack graph)模型,该模型基于“任意一个攻击都会有某些攻击作为后继”[12]这样的特点,使用了攻击图来描述攻击者的攻击意图和策略,并提出一种改进的博弈理论算法来进行攻防双方的策略的推理.这种算法充分发挥了攻击图在进行攻击意图描述方面以及博弈理论在处理攻击者和系统的收益、偏好和策略变化方面的优势,因而达到了比较好的响应效果和准确性.

本文第 1 节详细阐述 IRAG 模型所需要的基本元素,对相关工作进行介绍.第 2 节给出报警响应的具体算法.第 3 节是实验和分析.第 4 节是结论.

1 IRAG 模型中的基本元素

1.1 参与方

在报警响应中,参与方有系统管理员、安全员、各安全机制、合法用户、攻击者等等.为了降低复杂度,因为一般情况下合法用户对博弈的影响很小,因此在本文报警响应模型中合法用户不作为参与方出现.另外,系统管理员、安全员、安全机制等利益几乎是一致的而且他们的策略可以统一起来,因此本文将它们归结到一个参与方,统称系统.故本文报警响应博弈的参与方为攻击者 U_a 和系统 U_s.

1.2 参与方的类型空间

攻击者发出任何攻击都具有特定的目的,本文用攻击者在各安全尺度上的偏好来定义攻击者的类型,并以此来描述攻击者的攻击目的.设攻击者类型 $\theta_a \in \Theta_a$,其中 Θ_a 表示攻击者的类型空间,一般攻击者发起攻击的目的性很强,在安全尺度上常表现为对某种尺度的关心,如修改主页、删除系统文件等主要是针对完整性,窃取文件、破解密码等主要是针对机密性,而 DOS 攻击、大多数蠕虫攻击等主要针对可用性.因此,攻击者典型的类型空间可设为{(1,0,0),(0,1,0),(0,0,1)},表示攻击者分别在机密性、完整性和可用性方面的攻击偏好.

由于系统必须同时拥有机密性、完整性和可用性才能正常工作,因此其类型不可能只针对某一尺度,但是不同的系统对不同的尺度是有偏向的,比如主要从事网络服务提供的网络系统,其偏向可用性要多些;机要部门的网络系统对机密性偏向多一些;而电子商务类的网络系统几乎在完整性、可用性、机密性上是同等重要的.

因此可设系统的典型类型为{(0.1,0.1,0.8),(0.1,0.8,0.1),(0.8,0.1,0.1),(0.45,0.45,0.1),(0.45,0.1,0.45),(0.1,0.45,0.45),(1/3,1/3,1/3)}.

为了进行系统和攻击者类型的计算,本文建立了两个信息集:攻击者的信息集和系统的信息集.攻击者的信息集包含的信息主要来自于攻击者的踩点、嗅探、扫描以及根据系统的响应信息,而系统的信息集包含的信息则来自于系统内包括 IDS、防火墙、主机等各组件的报警、日志信息.信息集对对方类型的概率推理过程可以用专家系统、神经网络或模糊数学等方式进行,本文使用的是产生式规则的方式,即对攻击者或系统可能发现的每一条信息,设置其对对方在各类型上先验概率的增加值作为一条产生式知识,当所有信息推理完之后,对得出的概率进行归一化处理即得出对方各类型的先验概率.

1.3 节点价值

节点价值是一个用来表示节点重要性的量化的一个值.设节点价值度量的范围从 1~N,其中 1 表示的是最低的重要性,而 N 表示最高的重要性,N 的大小是根据系统环境确定的.例如,一些诸如不重要的匿名 FTP 服务器或者密罐系统等节点的节点价值可能被赋予 2,像生产服务器这样非常关键的节点可能被赋予一个比较高的节点价值,比如 15.网络中节点价值的度量范围和取值是根据整个系统的具体情况由系统管理员指定的.

系统的类型和节点价值共同表达了系统在安全防护时的偏向性.

1.4 攻击图

攻击图的定义:攻击图是一个四元组 $G=(S,\tau,S_0,S_s)$,其中,S 表示状态集,表示系统中存在的一个个可能遭受的攻击,可由系统中存在的漏洞推导出来;$\tau\subseteq S\times S$ 是传递关系,指的是攻击之间的关系,由攻击间的因果联系及网络连接状态决定;$S_0\subseteq S$ 是开始状态集,表示攻击者首先发起的攻击,如扫描等;$S_s\subseteq S$ 是成功状态集,表示满足攻击者攻击目的的攻击状态,同时也是攻击者最后一步攻击.攻击图的形成过程如图 1 所示.

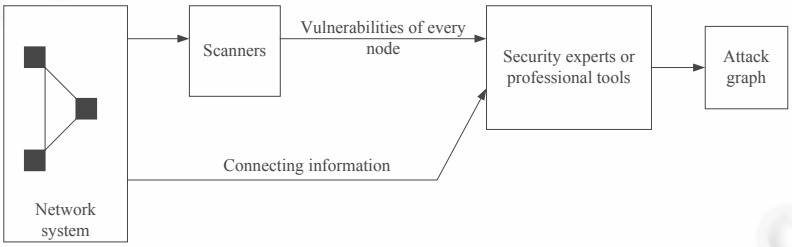

Fig.1 Constructing process of attack graph

图 1 攻击图的形成过程

为了将攻击图应用到本文的模型中,需要对攻击图进行以下扩展:① 孩子集 $C_i=\{\forall j, j.parent=i\}$,表示状态 i 子节点的集合;② 兄弟集 $B_i\subseteq\{\forall j\in s, j.parent=i.parent\}$,表示与状态 i 有相同父节点的状态集;③ 朋友集 $F_i=\{\forall j, j.sucees=i.sucees\in S_s\}$,其中,$j.sucees$ 表示节点 j 连接到的成功状态,因此朋友集表示与状态 i 连接到同一个成功状态的状态集.因为建立攻击图不是本文的主要工作,所以我们选用一个已有的工作[10,11]来进行分析.

1.5 参与方的行动空间

本文有两处需要确定行动空间,一是系统选择响应措施时的行动空间,记为 A_s,另一个是攻击者发出下一步动作时的行动空间,记为 A_a.

系统在进行响应时,实际上会根据攻击类型的不同,确定一个响应集,由于通常系统的响应措施并不多,因此,这项工作管理员或领域专家通过经验是容易完成的.

而攻击者一般不会只发出一个单独的攻击,他们的攻击之间是有联系的,也就是说,每个攻击之后的下一步攻击动作是有规律可循的.本文根据攻击图来确定攻击者下一步的攻击动作.

攻击 p 的后继攻击集定义为

$$S(p)=\{\forall a,a\in C_p\}\cup\{\forall a,j\in B_p\cap F_p\}\cup\{DOS,Null\} \tag{1}$$

表示攻击 p 的后继攻击集包括攻击图中 p 的子节点、与 p 具有相同目的的兄弟节点以及报复攻击的 DOS 和不再攻击的 *Null*.因为 DOS 攻击不一定需要特定的漏洞,所以攻击图中不完全能包括这个攻击,并且 DOS 攻击只针对攻击者已获取访问权限的服务器,如公共的 Web 服务器,因此本文暂只计算针对 Web 服务器的 DOS 攻击,即 DOS 攻击的目的节点为攻击者能够访问到的 Web 服务器.

实际上,在某些响应措施实施后,攻击者的有些后继动作将不会发生,因此在不同的响应措施后的后继动作集较式(1)会有所减少,本文为了简便起见,对各响应措施还是使用同样的后继动作集,而后继动作的削减则体现在第 2.6 节的响应措施对攻击及后继动作的阻止率上.

1.6 响应措施对攻击的阻止率

响应措施阻止攻击的效果,本文用它成功阻止攻击的概率来体现,如针对攻击 p 采用了响应措施 r,设其成功阻止的概率为 r_p.r_p 的值是根据以往的统计结果[3,4]或领域专家、管理员的经验知识配置的.因为与报警所报攻击有直接因果关系的后继攻击,如果报警所报攻击被阻止,它也不可能成功,因此,计算响应措施阻止率的时候需要继承对前一攻击的阻止率.例如,p 的后继动作 a 与 p 之间有直接因果关系,而响应措施 r 针对攻击 p 的阻止率为 r_p,那么响应措施 r 针对 a 的阻止率为 $\overline{\overline{r_p}\times\overline{r_a}}$,其中 r_a 为 r 对 a 元阻止率.另外,因为针对某些攻击,系统可能会同时使用多种响应措施,对于多种响应措施的阻止率,本文用 $r_{am}=\overline{\prod_{i=1}^{n}\overline{r_{ai}}}$ 来计算,其中,r_{am} 表示合并后的响应措施对攻击 a 的阻止率,r_{ai} (i=1,…,n)表示被合并的响应措施单独使用时各自对攻击 a 的阻止率.

1.7 参与方的收益

博弈局势 s=(s_s,s_a)后,系统的获利为

$$G_s(s,\theta_s)=\sum_{i=1}^{3}w_i(\theta_s)(-l(C^i,s)) \tag{4}$$

其中,$w_i(\theta_s)$ 为系统在类型 θ_s 时各安全尺度的权值,$l(C^i,s)$表示经过博弈局势 s 后,在 C^i 安全尺度上的损失.同样,攻击者的获利为

$$G_a(s,\theta_a)=\sum_{i=1}^{3}v_i(\theta_a)l(C^i,s) \tag{5}$$

其中,$v_i(\theta_s)$ 是攻击者在类型 θ_a 时各安全尺度的权值.设博弈局势 s 包含报警所报攻击 p、系统预测的下一步攻击 a 和响应 r,则

$$l(C^i,s)=\lambda_p(L_{N_p}(1-r_p)l(C^i,p)+L_{N_a}(1-r_a)l(C^i,a)) \tag{6}$$

其中,λ_p 表示报警 p 真实发生的可能性,由报警评价系统[13]获得.$l(C^i,a)$,$l(C^i,p)$ 是指攻击 a,p 所造成的系统在各安全尺度上的损失,是以节点价值等于 1 的节点为基础估算出来并存放在知识库中的,估算标准可参照文献[6].L_{N_p},L_{N_a} 分别指攻击 p,a 发生的节点 N_p,N_a 处的节点价值,r_p,r_a 指的是响应措施 r 阻止攻击 p,a 的概率.

2 基于贝叶斯博弈的报警响应

根据贝叶斯理性原则[14],如果人们进行决策时对与之相关的某种客体没有确定性的了解,而且也不知道其发生的客观概率,那么人们将对其做出主观概率判断,并在决策中如同应用客观概率一样应用这种主观概率判断.因此,可以使用先验概率进行收益函数的计算.在本文的入侵响应博弈中,由于攻击者在博弈过程中可以观察到一次系统的响应,因此可以根据观察的结果来改变其对系统类型的判断,而系统在博弈过程中不改变对攻击者类型的判断.在实际场景中,系统的响应动作攻击者通过探测几乎都可以知道,因此本文假定系统所实施的响应动作对攻击者来说是了解的,即博弈过程是完美的[14].因此攻击者对系统类型的信念就可以根据观察到的响应动作而做出调整,即形成后验信念.设攻击者和系统对对方类型的先验信念为 $p_a(\theta_s)$和 $p_s(\theta_a)$,而 r 为系统的

响应动作,攻击者对系统类型的后验信念 $\tilde{p}_a(\theta_s \mid r)$ 由下式求得:

$$\tilde{p}_a(\theta_{s_j} \mid r) = \frac{p_a(\theta_{s_j}) \cdot p_a(r \mid \theta_{s_j})}{\sum_{i=1}^{n}(p_a(\theta_{s_i}) \cdot p_a(r \mid \theta_{s_i}))} \tag{7}$$

其中, $p_a(r \mid \theta_{s_i})$ 表示在系统类型为 θ_{s_i} 的情况下,针对攻击 a 实施响应措施 r 的概率,可由统计获得.在计算出双方类型的信念后,本文的不完全信息博弈可转换成完美贝叶斯均衡 $s^*(\theta) = (s_a^*(\theta_a), s_s^*(\theta_s))$.

根据式(4)~式(9)可得,在系统使用 r 响应动作,攻击者使用 a 作为下一步攻击动作时,攻击者和系统的收益分别为

$$u_a(r,a) = \sum_{\theta_a}\left(\tilde{p}_s(\theta_a \mid \theta_s)\left(\sum_{i=1}^{3} w_i(\theta_s)(-l(C^i,s)) - L_{N_a} cost(r_a) - L_{N_p} cost(r_p)\right)\right) \tag{8}$$

$$u_s(r,a) = \sum_{\theta_s}\left(\tilde{p}_a(\theta_s \mid \theta_a)\left(\sum_{i=1}^{3} v_i(\theta_a) l(C^i,s) - \lambda_p(cost(p) + cost(a))\right)\right) \tag{9}$$

其中,*cost*(*p*),*cost*(*a*)是指攻击 *p*,*a* 的攻击代价,攻击者实施攻击所消耗的计算机资源、时间等代价,*csot*(*r*)为响应措施 *r* 的代价,是指系统实施响应动作时所付出的代价,这些代价类似于攻击获利,也都是估算值,以通用的货币为单位.

因为本文只考虑攻击者的攻击、系统响应及攻击者下一步攻击的过程.因此,针对系统的每一个响应动作,攻击者在下一步攻击的选择上,均会选择其收益最大的攻击动作,即:

对系统响应动作 $r_i(i=1,\ldots,n)$,

攻击者的最佳下一步攻击动作为

$$a^{\Delta}(r_i) \in \arg\max_{a_j} u_a(r_i, a_j) \tag{10}$$

因此系统的最佳响应为

$$r^*(a^{\Delta}(r_i)) \in \arg\max_{a^{\Delta}(r_i)} u_s(r_i, a^{\Delta}(r_i)) \tag{11}$$

当得出系统最佳响应动作后,攻击者在整个博弈过程中的最佳响应下一步攻击动作为

$$a^*(r^*(a^{\Delta}(r_i))) \in \arg\max_{a_j} u_a(r^*(a^{\Delta}(r_i)), a_j) \tag{12}$$

根据纳什均衡的存在条件[14]:任意有限策略型博弈至少存在一个混合策略纳什均衡.因为 IRAG 模型中各参与方在每个信息集上的可选行动数目是有限的,所以,它的扩展型博弈是有限的,对应的策略型博弈也是有限的,因此 IRAG 模型至少存在一个混合策略纳什均衡.而根据文献[15]中的定理:如果有限扩展型博弈是完美信息的,则它还存在纯策略纳什均衡.综上所述,IRAG 模型至少存在一个纯策略纳什均衡.

对进行计算中存在多于一个均衡的情况(实验中这种情况很少),本文提出了一个简单而有效的方法:预定义参照响应模式对报警响应的博弈搜索进行指导,即系统预先定义一种响应方式,对出现的多个纳什均衡,使用最接近于预定义响应方式的措施(本文根据响应措施的阻止率判断)进行响应,从而确定唯一的博弈结果,便于系统自动响应的快速实施.

根据上述计算公式可以看出,在本文的 IRAG 模型下,系统的响应动作是针对攻击者所有可能的下一步动作做出的最优反应.根据均衡的定义[14]:均衡时每个参与方选择的策略都是其他参与方所选策略的最佳反应,任何一方单方面改变策略就会使其利益受损,因此攻击者和系统无论哪一方都不会偏离均衡结果选择行动.由此可见,IRAG 模型推导的系统的响应决策,是充分考虑攻击者和系统双方利益下的最优反应,其结果是稳定的.而单方面考虑系统收益的模型,由于攻击者的策略未被考虑进去,当攻击者改变原有的攻击策略时,响应模型计算出的最优解将会失效.另外,本文提出的 IRAG 模型使用了攻击图来描述攻击者的攻击策略,一方面,可将攻击之间的因果关系考虑进来,另一方面,根据各节点之间的连通性对攻击关系加以限制,因而能够较为准确地对攻击者的策略进行判断,同时也加强了系统推理最佳响应动作的准确性.

3 应用实例

如在图 2 的拓扑环境中,系统中有两台主机,位于 DMZ(DeMilitarized zone)区的是 Solaris 系统 L_1(节点价值为 6),运行 Sadmind 和 Apache,其中 Apache 用作对外门户网站的 Web 服务器.另一台是 Windows 系统 L_2(节点价值为 10),运行 Sql Server 数据库.本例中,Windows 节点信任 Solaris 节点.经扫描,Solaris 节点中的漏洞有 Sadmind 缓冲区溢出,Windows 节点有 Sql Server 缓冲区溢出.根据文献[10,11]的方法,建立本例的攻击图如图 3 所示.

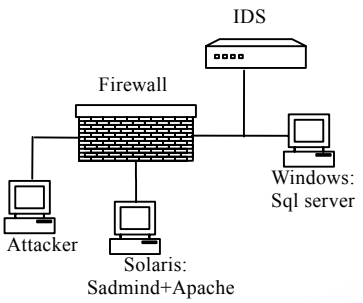

Fig.2 The network environment of the experiment

图 2 实例的网络环境

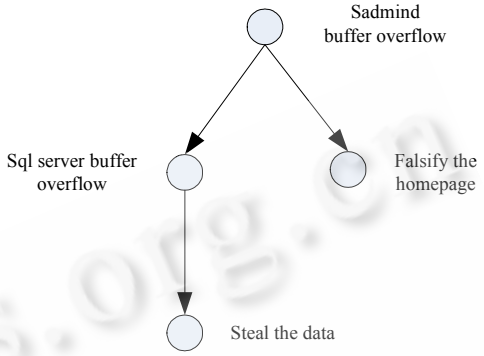

Fig.3 The attack graph of the experiment

图 3 实例的攻击图

实验中,系统入侵检测系统发出了 Sadmind 缓冲区溢出的报警,并评估报警的可信度为 90%,针对这个报警系统进行响应,即 p=Sadmind 缓冲区溢出,根据后继攻击集的定义,p 的后继攻击集为 $\{a_1,a_2,a_3,a_4,a_5\}$,其中 $a_1,a_2,\ldots,a_5$ 分别是指 Null、修改主页、Sql Server 缓冲区溢出、DOS 攻击和 Sadmind 缓冲区溢出,攻击代价分别为 0,2,5,10,5.系统的响应方法有记录、关闭 Sadmind 20 分钟、关闭 Solaris 主机 20 分钟、关闭 Windows 主机 Sql Server 数据库 20 分钟,分别用 m_1,m_2,m_3,m_4 表示,响应代价分别为 0,5,20,8.因为管理员使用 m_4 的时候通常要使用 m_2,因此系统针对报警 p 的响应集为 $\{m_1,m_2,m_2+m_4\}$,博弈场景如图 4 所示.

实验中,系统和攻击者的类型空间分别设为 $\Theta_s=\{\theta_{s1},\theta_{s2},\ldots,\theta_{s7}\}$,其中 $\theta_{s1}=(0.1,0.1,0.8)$,$\theta_{s2}=(0.1,0.8,0.1)$,$\theta_{s3}=(0.8,0.1,0.1)$,$\theta_{s4}=(0.45,0.45,0.1)$,$\theta_{s5}=(0.45,0.1,0.45)$,$\theta_{s6}=$,$(0.1,0.45,0.45)$ $\theta_{s7}=(1/3,1/3,1/3)$;$\Theta_a=\{\theta_{a1},\theta_{a2},\theta_{a3}\}$,其中 $\theta_{a1}=(1,0,0)$,$\theta_{a2}=(0,1,0)$,$\theta_{a3}=(0,0,1)$.系统受到攻击的损失、原子响应措施对攻击动作的阻止概率见表 1 和表 2.

首先,根据推理(推理过程略),得出系统对攻击者类型的先验信念为 $p_s(\theta_{a1})=0.5$,$p_s(\theta_{a2})=0.4$,$p_s(\theta_{a3})=0.1$;攻击者对系统类型的先验信念为 $p_a(\theta_{a1})=0.3$,$p_a(\theta_{s7})=0.7$.当系统发出 Sadmind 缓冲区溢出报警时,$p(r|\theta_s)$的值见表 3,经过贝叶斯公式计算,攻击者对系统类型的后验概率分别为 $\tilde{p}_a(\theta_{s1}|m_1)=0.533$,$\tilde{p}_a(\theta_{s7}|m_1)=0.467$,$\tilde{p}_a(\theta_{s1}|m_2)=0.125$,$\tilde{p}_a(\theta_{s7}|m_2)=0.875$,$\tilde{p}_a(\theta_{s1}|m_3)=0.125$,$\tilde{p}_a(\theta_{s7}|m_3)=0.875$,$\tilde{p}_a(\theta_{s1}|m_2+m_4)=0$,$\tilde{p}_a(\theta_{s7}|m_2+m_4)=1$.

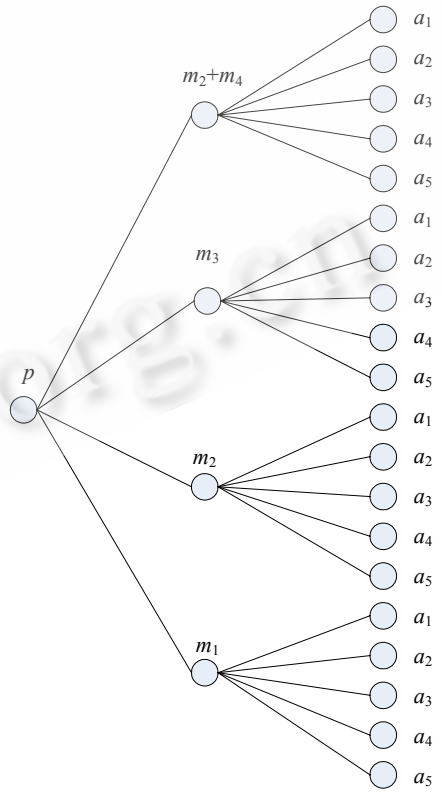

Fig.4 The game scenario

图 4 博弈场景

Table 1 The prevent probability of response methods

表 1 响应阻止率

Response m_i attack	m_1	m_2	m_3	m_4
a_1	0	0	0	0
a_2	0	90%	95%	0
a_3	0	0	0	95%
a_4	0	0	95%	0
a_5	0	90%	95%	0

Table 2 The lost caused by attack

表 2 攻击损失

Attacks Attack lost	a_1 (0,0,0)	a_2 (0,40,0)	a_3 (40,30,5)	a_4 (0,0,40)	a_5 (40,30,5)

Table 3 $p(r|\theta_s)$ value

表 3 $p(r|\theta_s)$值

$p(r\|\theta_s)$	θ_{s1}	θ_{s2}	θ_{s1}	θ_{s1}	θ_{s1}	θ_{s1}	θ_{s1}
m_1	0.8	0.1	0.1	0	0.1	0.5	0.3
m_2	0.1	0.6	0.3	0.6	0.6	0.4	0.3
m_3	0.1	0.2	0.3	0.2	0.2	0.2	0.3
m_2+m_4	0	0.1	0.3	0.2	0.1	0	0.1

将上述元素用本文算法进行计算,得出本次实验的唯一纯策略均衡结果为($r^*=m_2,a^*=a_3$.表示系统的最佳响应方式为 m_2,即关闭 Solaris 系统中的 Sadmind 20 分钟,此时系统和攻击者的收益为–66.5 和 37.3.在这个实验中,若系统管理员改变他的响应策略,如使用 m_3 或 m_1 进行响应,计算出系统的收益分别是–136.7 和–226.35,显然低于博弈均衡时的收益,即系统不可能通过单方面改变他的策略来增加收益.同样,对攻击者而言,也是如此,如在本实验中若系统采用 m_2 响应方式,攻击者不用均衡时的 a_3 攻击,而使用 a_2 或 a_4 攻击,则此时攻击者的收益仅为 19.7 和 27.1,均小于均衡时的 37.3,即攻击者不可能通过单方面改变他的策略来增加收益.另外,通过这个实验也可以看出,响应效果较好的响应方式,并不一定能够增加系统的收益,相反地,由于响应代价过高,收益反而降低.如实验中的看上去响应效果较好的 m_3 或 m_2+m_4,而计算出系统的收益分别是–136.7 和–124.3,小于采用本文的方法选择 m_2 时的收益–66.5.

另外,本文使用安全尺度的偏好及节点的价值对系统和攻击者的收益进行计算,较不使用安全尺度偏好及节点价值更加合理和精确.如在本实验中,由于攻击者的类型为(0,0,1)的概率仅为 10%,故其对 DOS 攻击的偏好不高,若不计算安全尺度的偏好,将会显著提高 DOS 攻击给攻击者带来的收益,因此,最终的计算结果必然会发生较大变化.本次实验中,在不计算安全尺度偏好的情况下进行计算,系统的最佳响应方式为 m_3,这也是其他基于成本分析的入侵响应研究[6]的最优结果,而这时,系统的实际收益为–136.7,低于采用本文方法选择 m_2 进行响应时的–66.5.

因此,从本次实验可以看出,本文的基于 IRAG 模型的响应方法,比较全面和准确地考虑到系统和攻击者在攻防时的收益,较之其他响应方式更加准确和有效.同时,基于 IRAG 模型推理里的最终结果确保了系统和攻击者双方都接受且不会擅自改变的,即对系统而言其响应措施为稳定的最优响应措施.

4 结束语

在当前报警响应工作中广泛存在着不能充分考虑攻击者的策略变化以及不能充分考虑系统的收益因素等问题,对此本文提出了一种基于攻击图的报警响应模型.该模型运用攻击图对攻击者的攻击策略进行建模,并使用博弈理论中计算攻防双方均衡的方式进行最优响应措施的推断.经过分析和实验验证,一方面,IRAG 模型通过使用攻击图将攻击者可能的策略变化纳入模型当中,另一方面,对攻击者和系统双方收益考虑得比较全面.因此,系统在报警响应后的收益得到了保证,使得比单从系统方面考虑的最优响应决策更加稳定、可靠.因此,使用 IRAG 模型进行报警响应,能够有效提高响应效果.当然,目前的 IRAG 模型还不够完善,比如使用攻击图对攻击者的攻击策略进行建模,需要消耗大量的时间和空间,暂时还难以实用等,因此对攻击图在建图方法上进行优

化,以提高建图的效率将是我们下一步工作的重点.

References:

[1] Kabiri P, Ghorbani A. Research on intrusion detection and response: A survey. Int'l Journal of Network Security, 2005, 1(2):84−102.

[2] Musman S, Flesher P. System or security managers adaptive response tool. In: Proc. of the Information Survivability Conf. and Exposition 2000. 2000. http://doi.ieeecomputersociety.org/10.1109/DISCEX.2000.821509

[3] Schnackenberg D, Djahandari K, Sterne D. Infrastructure for intrusion detection and response. In: Proc. of the DARPA Information Survivability Conf. and Exposition. 2000. 3−11. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=17794&arnumber=821505

[4] Carver CA, Hill JM, Surdu JR, Pooch UW. A methodology for using intelligent agents to provide automated intrusion response. In: Proc. of the IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop. West Point, 2000. 110−116. http://www.bucksurdu.com/Professional/Documents/IntrusionResponsePaper.pdf

[5] Ragsdale DJ, Carver CA, Humphries JW, Pooch UW. Adaptation techniques for intrusion detection and intrusion response systems. In: Proc. of the IEEE Int'l Conf. on Systems, Man, and Cybernetics at Nashville. Tennessee, 2000. 2344−2349. http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/7099/19129/00884341.pdf?isNumber=19129&arNumber=00884341&isnumber=19129&arnumber=00884341

[6] Lee W, Fan W, Miller M, Stolfo SJ, Zadok E. Toward cost-sensitive modeling for intrusion detection and response. Journal of Computer Security, 2002,10(1-2):5−22.

[7] Foo B, Wu YS, Mao YC, Bagchi S, Spafford E. ADEPTS: Adaptive intrusion response using attack graphs in an e-commerce environment. In: Proc. of the 2005 Int'l Conf. on Dependable Systems and Networks (DSN 2005). 2005. 508−517. http://doi.ieeecomputersociety.org/10.1109/DSN.2005.17

[8] Cuppens F, Miege A. Alert correlation in a cooperative intrusion detection framework. In: Proc. of the IEEE Symp. on Research in Security and Privacy. Oakland, 2002. 202−215. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1004372

[9] Ning P, Xu D. Learning attack strategies from intrusion alerts. In: Proc. of the 10th ACM Conf. on Computer and Communication Security. Washington, 2003. 200−209. http://doi.acm.org/10.1145/948109.948137

[10] Sheyner O, Joshua H, Jha S, Lippmann R, Wing JM. Automated generation and analysis of attack graphs. In: Proc. of the IEEE Symp. on Security and Privacy. Oakland, 2002. 273−284.

[11] Sheyner O, Wing J. Tools for generating and analyzing attack graphs. LNCS 3188, 2004. 344−371. http://www.cs.cmu.edu/~scenariograph/sheynerwing04.pdf

[12] Debar H, Wespi A. Aggregation and correlation of intrusion-detection alerts. In: Proc. of the Recent Advances in Intrusion Detection: The 4th Int'l Symposium (RAID 2001). Davis: Springer-Verlag, 2001. 85−103. http://portal.acm.org/citation.cfm?id=645839.670735&coll=GUIDE&dl=GUIDE

[13] Phillip AP, Martin WF, Alfonso V. A mission-impact-based approach to INFOSEC alarm correlation. In: Proc. of the 5th Int'l Symp. on Recent Advances in Intrusion Detection (RAID) 2002. Zurich: Springer-Verlag, 2002. 95−115. http://www.springerlink.com/content/2487wb0an7qq8art/

[14] Osborne MJ, Rubinstein A. A Course in Game Theory. Cambridge, London: MIT Press, 1994.

[15] Kuhn H. Extensive games and the problem of information. Annals of Mathematics Studies, 1953,2(28):193−216.

石进(1976－),男,安徽和县人,博士,CCF学生会员,主要研究领域为网络安全,系统安全.

陆音(1978－),男,博士,主要研究领域为网络安全,下一代互联网.

郭山清(1976－),男,博士,讲师,CCF 会员,主要研究领域为网络安全.

谢立(1942－),男,教授,博士生导师,CCF高级会员,主要研究领域为分布式系统,系统安全,网络安全.