计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2015.20140184
52(7):1642-1659,2015

基于威胁传播采样的复杂信息系统风险评估

马 刚[1,2] 杜宇鸽[3] 安 波[1] 张 博[1,2] 王 伟[4] 史忠植[1]
[1](中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)
[2](中国科学院大学 北京 100049)
[3](中国信息安全测评中心 北京 100085)
[4](北京联索科技有限公司 北京 100080)
(mag@ics.ict.ac.cn)

Risk Evaluation of Complex Information System Based on Threat Propagation Sampling

Ma Gang[1,2], Du Yuge[3], An Bo[1], Zhang Bo[1,2], Wang Wei[4], and Shi Zhongzhi[1]
[1](*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)
[2](*University of Chinese Academy of Sciences, Beijing* 100049)
[3](*China Information Technology Security Evaluation Center, Beijing* 100085)
[4](*Beijing Lexo Technologies Co., Ltd., Beijing* 100080)

Abstract Information security has been one of the focuses of social concern in the age of Internet. There is no doubt that accurately assessing the security of information medium is becoming the focus of the present information security work. For evaluating the risk of the large-scale distributed complex information system, we propose a risk evaluation method of the complex information system based on threat propagation sampling. Firstly, when the threats propagate in the complex information system, the number of threat propagation trees (TPT) is reduced by sampling the transition states of the asset nodes and the threat propagation edges emitted by the asset nodes, then computing the expected value loss of each node in the threat propagation tree and the probability of each threat propagation tree to evaluate the risk of the complex information system. The experimental analysis not only shows the risk evaluation proposed in this paper has higher time-efficient compared with the traditional combined strategy when producing the threat propagation tree, but also can make an objective and accurate risk evaluation. Furthermore, it can give more comprehensive and rational security-guide advices for security-risk managers when developing some security protection strategies on the asset nodes of the complex information system.

Key words complex information system; threat propagation sampling; risk evaluation; asset node; threat propagation edge; threat propagation trees (TPT)

摘 要 互联网时代的信息安全已成为全社会关注的问题之一.信息系统是信息的载体,为有效评估大规模分布式复杂信息系统的风险,构建了一种基于威胁传播采样的复杂信息系统风险评估方法.该方法考虑到威胁在复杂信息系统中传播时,对资产结点的转移状态以及资产结点发出的威胁传播边进行采

收稿日期:2014-03-06;**修回日期**:2014-12-05
基金项目:国家"九七三"重点基础研究发展计划基金项目(2013CB329502);国家自然科学基金项目(61035003);国家"八六三"高技术研究发展计划基金项目(2012AA011003);国家科技支撑计划基金项目(2012BA107B02)

样来生成威胁传播树(threat propagation trees, TPT),然后通过计算威胁传播树中各资产结点的期望损失以及威胁传播树的概率来对整个复杂信息系统进行风险评估.实验分析表明,基于威胁传播采样的复杂信息系统风险评估方法,在生成威胁传播树时具有高效的时间效率,能够对复杂信息系统进行客观准确的风险评估,且在对复杂信息系统资产结点制定安全防护策略时,能够为安全风险管理者提供较为合理的安全指导建议.

关键词　复杂信息系统;威胁传播采样;风险评估;资产结点;威胁传播边;威胁传播树

中图法分类号　TP309

科学技术的发展以及信息时代的到来,让世界充满着各种各样的复杂信息系统,从 Internet 到 World Wide Web、从大型电力网络到全球交通网络、从科研合作网络到各种经济、政治、社会关系网络等等[1-2].一个典型的复杂信息系统可以抽象为由多个结点以及连接 2 个结点之间的边所组成,结点用来代表真实系统中的不同个体,而边则用来表示个体间的关系.通常 2 个结点之间具有某种特定的关系则连一条边,反之则不连边;有边相连的 2 个结点被看作是相邻的[3].例如,计算机网络可以看作是自主工作的计算机通过通信介质(如光缆、双绞线、同轴电缆)相互连接形成的网络,类似的还有电力网络,社会关系网等等[4].

但是,就在信息技术快速发展与信息网络系统广泛应用的同时,以复杂数据通信网络为技术支撑基础的信息系统建设也相应地得到了快速推进,在此基础上,各项信息业务开展的多元化综合利用和信息共享工作为社会管理涉及的各项工作提供强有力的信息技术支持,然而如何保证复杂数据通信网络的安全稳定运行是信息化工作在日常运行维护中遇到的最大问题之一[5].复杂信息系统安全的风险评估则是保障信息系统安全运行的关键,旨在分析说明系统中客观存在的潜在威胁对系统脆弱性造成的破坏和损失,最终将风险控制在可接受范围内[6].

信息安全风险评估工作是保障网络和信息系统安全的重要基础,现有的复杂网络信息安全技术依赖于防火墙、入侵检测和反病毒软件等,都属于静态片面的被动安全防御策略,强调的是以攻击为中心,检测攻击后才有所响应.事实上,为了保证网络系统的安全性和健壮性,尤其是复杂网络系统,必须提出一系列主动实时防护模型与技术战略,通过态势感知、风险评估、安全检测等手段对当前网络安全态势进行判断,并依据判断结果实施网络主动防御的安防体系[7].

因此,基于安全风险评估的主动安全防御技术逐渐成为主流,与传统的被动防御技术相比,基于安全风险评估的主动防御技术能够帮助用户预先识别网络系统脆弱性以及所面临的潜在的安全威胁,根据安全需求来选取符合最优成本效应的主动安全防御策略,从而能够达到提前避免危险事件发生[8].

基于安全风险评估的主动安全防御技术,最关键的一个环节就是安全风险评估,现有的复杂信息系统安全风险评估方法比较多,但总的概括起来主要有如下 3 类[9]:1)基于安全监测工具的评估方法,主要用于安全期望不高的评估对象;2)基于风险评估技术的评估方法,主要有 CORAS,RSDS,CRAMM 和 COBIT;3)基于系统综合的评估方法,使用先进的评估理论方法建立风险评估模型进行评估,得出科学合理的评估结果.经过多年的探索与实践,风险评估理论和方法在其他领域已逐渐发展成一套较为完整、成熟的手段和工具,但在信息安全领域,由于其领域的特殊性,仍旧存在着许多困扰管理者和研究人员的问题,如风险评估过程相关性处理、风险评估的动态性、人因可靠性以及不确定性[10].

从国内外的研究现状可以看出:一方面,信息系统安全评估标准与方法强调风险评估的必要性,通过评估系统中资产的安全属性来判断信息系统的安全等级,通常采用一些问卷的方式,这些往往只能给出面向过程的评估框架模型,不能给出定量的分析,缺乏可操作的工程数学方法,导致评估结果存在较大的主观性[11].另一方面,有些风险评估方法只给出了定量计算方法,但实施过于繁琐,且理论研究与实践结合不够紧密,没有很好地与我国信息化实践相结合[12].

针对以上问题,本文提出一种基于威胁传播进行采样的复杂信息系统风险评估定量分析方法,以图论为基础并结合概率论的相关知识,通过对复杂信息系统中受到威胁作用的资产结点的转移状态以及受到威胁作用的资产结点发出的威胁传播边进行采样的方式构建威胁传播树,然后再通过计算每棵

威胁传播树的期望风险来评估复杂信息系统的整体期望风险,最后为系统中的资产结点给出一个安全保护策略.

本文主要贡献,提出了一种新的威胁传播树生成方法,该方法能够模拟威胁在复杂信息系统中的传播过程,且在运行时间效率上明显优于文献[13]中所采用的组合策略方法.本文所提方法获得的威胁传播树用于复杂信息系统的风险评估时,能够保证与组合策略方法有着相同的风险评估准确度,而且在制定资产结点安全保护策略时,能够提供比组合策略方法更全面的安全建议.

1 复杂信息系统

复杂信息系统是一类子系统数量大、种类多,相互之间有非常复杂的非线性关联关系的信息系统,它具有开放性、复杂性、非线性、层次性、进化涌现性等特点,且子系统之间存在着强耦合现象,不能够将系统分成几个单独部分进行研究[14].复杂信息系统构成要素一般包括:资产、威胁、脆弱性、系统拓扑结构(如图 1 所示).

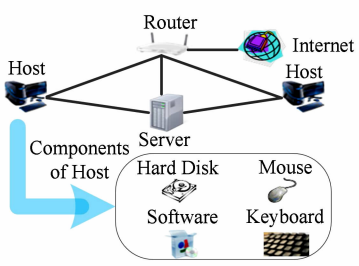


Fig. 1 Complex information system instance.

图 1 复杂信息系统实例

复杂信息系统一般由一系列硬件和软件组成,各个构成要素的定义如下.

定义 1. 资产.对系统具有价值的事物,用 A 表示资产的集合.

以图 1 中的信息系统为例,主机(host)、服务器(server)、路由器(router)以及由主机细化后的硬盘、鼠标等也可以当作系统中的资产.

系统在运行过程中,一个资产可能处于不同的状态,例如正常状态、被感染病毒、管理员权限被非法获取、系统崩溃等.设 S 为系统中资产所有状态的集合,定义映射 $W:A\times S\mapsto R$,$W(a,s)$ 表示资产 a 在状态 s 时的价值.本文假设对于任意 $a\in A$,$s\in S$,$W(a,s)$ 是已知的,这些值可以由系统管理者分析得出.

定义 2. 威胁.可能对资产造成损害的因素,用 T 表示所有威胁的集合.

如图 1 中的复杂系统所面临的威胁可能有缓冲区溢出攻击、恶意代码、人员操作失误等.

处于某一状态 s 的任一资产 a 面临的威胁可能是由系统中的其他资产传播而来,也可能直接来自系统外部或者资产自身产生(如硬件故障、长时间运行导致内存崩溃、操作人员误运行了病毒代码等)[15].我们将通过其他资产传播而来的威胁 t 定义为资产的传播性威胁,将直接来自系统外部或者资产自身产生的威胁 t 称为资产的原发性威胁.在本文中令 $\eta(t|a,s)$ 来表示某一段时间内处于状态 s 的资产 a 爆发原发性威胁 $t\in T$ 的概率.

定义 3. 脆弱性.能被威胁利用的、存在于资产上的弱点,用 V 表示所有脆弱性的集合.

如图 1 所示复杂信息系统的脆弱性可能包括操作系统漏洞、应用程序漏洞、硬件弱点、start_art.html 漏洞等[16].

由于资产处于某一状态时有可能暴露出多个不同的脆弱性,因此威胁 t 作用于资产 a 时,a 所处状态不同,则 t 可利用的脆弱性也不同.假设威胁 t 作用在处于状态 s 的资产 a 时,t 可以利用的脆弱性有多种,但是 t 每次作用时仅能利用其中一个脆弱性.因此,记 $\beta(v|a,s,t)$ 表示威胁 t 作用在处于状态 s 的资产 a 上时,利用资产 a 的脆弱性 v 的概率,且满足:

$$\sum_{v\in V}\beta(v\mid a,s,t)\leqslant 1. \tag{1}$$

式(1)表示威胁作用在处于状态 s 的资产 a 上时,利用该状态下所有脆弱性 v 的概率之和不超过 1.式(1)等于 1 的情况为状态 s 的所有脆弱性都有可能作用成功;小于 1 的情况为处于状态 s 时存在不能作用成功的脆弱性.

定义 4. 系统拓扑结构.系统间的资产因访问关系而形成的网络结构.

如图 1 所示复杂信息系统中的 4 个资产相互连接而构成的网络结构.复杂信息系统里的资产因访问关系形成的网络结构(即拓扑结构)可用无向图 $G=G(A,E)$ 来描述,其中 A 为无向图 G 的结点集合,A 中的一个元素表示复杂信息系统中的一个资产,$E=A\times A$ 为边的集合,E 中的任一元素均为一个关于 2 个资产的 2 元组 $\langle a_i,a_j\rangle$,其中 $a_i,a_j\in A$ 且 $a_i\neq a_j$,边 $\langle a_i,a_j\rangle$ 用于表示 2 个资产 $a_i,a_j(a_i\neq a_j)$ 间存在的访问关系,威胁就是利用这些关系在系统的各资产间进行传播[17].

复杂信息系统各个构成要素之间的关系如图 2 所示. 资产拥有脆弱性,脆弱性暴露资产缺点,脆弱性是资产本身的属性,威胁会利用资产的脆弱性对资产造成损害,单纯的脆弱性本身不会对资产造成损害,资产间相互连接构成系统拓扑结构[18].

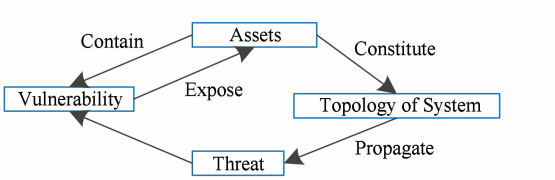


Fig. 2　Interrelationship of components in complex information system.

图 2　复杂信息系统构成要素及其相互关系

2　威胁在复杂信息系统中的传播

2.1　资产状态转移

复杂信息系统在运行过程中,威胁 t 作用在处于状态 s 的资产 a 后,使得资产 a 以一定概率从原状态向其他状态 s' 转变的过程称为资产状态转移.

一个资产处于正常状态时,可能由于溢出攻击利用操作系统漏洞而使资产状态转移到管理员权限被非法获取的状态,也可能由于溢出攻击利用某个应用程序的漏洞使系统到达不能提供服务的状态. 处于系统崩溃状态的资产面对溢出攻击时其状态不会发生转移[19]. 同样,资产处于正常状态时,如果威胁是病毒则可能使资产转移到系统崩溃的状态,所以威胁 t 作用于资产 a 后,a 向哪些状态转移依赖于 3 个因素:

1) 资产 a 的当前状态;

2) 作用于资产的威胁 t;

3) 威胁 t 作用于资产 a 时所利用的脆弱性 v.

设威胁每次作用于资产时只能利用资产的一个脆弱性,如果威胁对资产作用成功,那么资产的状态转移是一个逐步恶化的过程,而且转移到的下一状态一定是一个没有到达过的状态,即资产的状态转移过程不会出现环路. 例如,溢出攻击利用操作系统漏洞而使资产状态转移到管理员权限被非法获取的状态,在这个状态转移过程中,管理员权限被非法获取的状态一定是资产以前没有到达过的状态,而且管理员权限被非法获取的状态比正常状态更恶化[20]. 记 $\delta(s'|a,s,t,v)$ 表示资产 a 处于状态 s 时,t 利用 a 的脆弱性 v 作用于 a 后,a 的状态由 s 转移到 s' 的概率,满足 $\sum_{s'\in S}\delta(s'|a,s,t,v)=1$.

当一个威胁 t 作用于资产 a 后,首先判断 a 当前所处状态 s,然后确定当前状态下存在的脆弱性 v,最后威胁 t 以一定的威胁作用概率 $\beta(v|a,s,t)$ 利用资产的不同脆弱性对资产进行作用. 若威胁 t 对资产的脆弱性 v 作用成功,则资产 a 将以一定的状态转移概率 $\delta(s'|a,s,t,v)$ 从状态 s 转移到状态 s'. 因此,一个威胁 t 作用在处于状态 s 的资产 a,导致其状态转移到 s' 的概率 $\rho(s'|a,s,t)$ 为:资产 a 处于状态 s 时,威胁 t 利用资产在该状态下的每一种脆弱性 v 的概率 $\beta(v|a,s,t)$ 与威胁 t 利用资产在该状态下的每一种脆弱性 v 作用于 a 导致其状态转移到 s' 的概率 $\delta(s'|a,s,t,v)$ 的乘积之和,也即:

$$\rho(s'\mid a,s,t)=\sum_{v\in V}\beta(v\mid a,s,t)\delta(s'\mid a,s,t,v). \tag{2}$$

2.2　威胁传播

系统中的资产转移到某个状态后引发新的威胁作用于相邻资产的过程称为威胁传播,如图 3 所示即为 2 个资产之间的一种威胁传播形式.

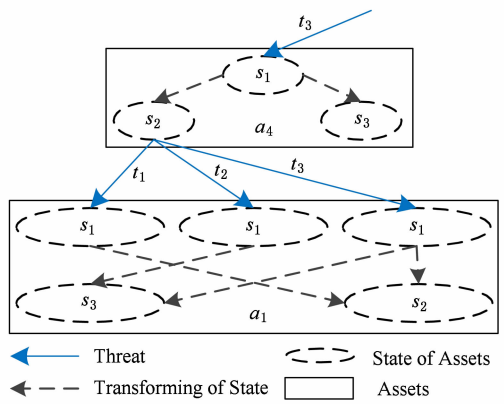


Fig. 3　Detailed process propagating threat from asset a_4 to a_1.

图 3　威胁从资产 a_4 传播到 a_1 的详细过程

在图 3 中,处于状态 s_1 的资产 a_4 受到威胁 t_3 作用后,a_4 的状态将有可能转移到状态 s_2 或 s_3,如果 a_4 的状态转移到 s_2 后,则处于状态 s_2 的资产 a_4 将有可能向处于状态 s_1 的相邻资产 a_1 发出 3 种威胁 t_1,t_2 或 t_3,a_1 在受到不同的威胁后将会发生不同的状态转移. 从图 3 中可以看出,处于状态 s_1 的资产受到威胁 t_1 作用后,其状态将转移到 s_2;受到威胁 t_2 作用后,其状态将转移到 s_3;受到威胁 t_3 作用后,其状态将转移到 s_2 或 s_3.

威胁的传播依赖于复杂系统的拓扑结构,且威胁在传播的过程中,一个威胁导致资产 a 状态发生

转移后,可能引发哪些威胁作用于相邻资产 a' 与 2 个因素有关:

1) 资产 a 受到威胁作用后的转移状态 s';

2) 资产 a 与其相邻资产 a' 之间的访问关系.

记 $\xi(t|a,s)$ 表示处于状态 s 的资产 a 向相邻资产发出威胁 t 的概率,且 $\sum_{t\in T}\xi(t|a,s)\leqslant 1$. 在实际中,一个资产能够向它的相邻资产发出威胁时,表明该资产已经发生了状态转移,而该资产能够发生状态转移也表明它已被其他相邻资产发来的威胁作用成功,也即整个过程是资产因为受到威胁作用而发生状态转移,紧接着该资产在状态转移后又发出新的威胁.

因此,当资产 a 受到威胁 t 作用,从初始状态 s 以概率 $\rho(s'|a,s,t)$ 转移到状态 s' 时,向相邻资产发出威胁 t' 的概率记为 $\varphi(t'|a,s,t)$,它由该资产 a 受到威胁 t 作用时的状态转移概率 $\rho(s'|a,s,t)$ 与资产转移到状态 s' 后向相邻资产发出威胁 t' 的概率 $\xi(t'|a,s')$ 的乘积组成,即:

$$\varphi(t'|a,s,t)=\sum_{s'\in S}\rho(s'|a,s,t)\xi(t'|a,s'). \quad (3)$$

威胁在传播的过程中,对于任一复杂信息系统,都可以将系统里面的资产结点 $A=\{a_0,a_1,\cdots,a_n\}$ 划分为 3 类(为便于表示以下均将"资产结点"称作"结点"):

1) 活动结点(active node, AN):当前时刻受到了威胁作用,并且将产生的新威胁向外传播的结点.

2) 死结点(inactive node, IAN):已经被威胁作用过的结点,且不再被威胁再次作用.

3) 可激活结点(activatable node, ATN):网络中还未受到威胁作用过的结点.

此外,本文假设威胁在传播的过程中,复杂信息系统中的结点还应具有如下 6 个特征:

1) 每个结点所拥有的状态 $S=\{s_0,s_1,\cdots,s_m\}$ 种类有限,脆弱性 $V=\{v_0,v_1,\cdots,v_N\}$ 种类有限,且每种状态可以传播多种威胁 $T=\{t_0,t_1,\cdots,t_M\}$,但每种威胁 1 次只能作用于资产的一个脆弱性.

2) 一个结点受到原发性威胁 t_ε 作用后,由该原发性威胁直接或间接引发的一系列其他威胁在网络中传播过程时,不会有其他原发性威胁发生.

3) 每时刻,活动结点只向相邻的可激活结点传播威胁,当威胁传播成功后,该活动结点立即变为死结点,将不再被威胁作用,而受到威胁作用的可激活结点成为了下一时刻的活动结点.

4) 处于状态 s 的活动结点 a 受到威胁 t 作用后,将以相同概率 $\varphi(t'|a,s,t)$ 向所有相邻可激活结点传播威胁 t'.

5) 同一时刻不存在多个威胁同时作用于同一个结点,但允许一个资产发出的威胁同时作用于多个相邻可激活结点.

6) 当系统中不存在活动结点时,威胁传播结束.

在复杂信息系统的工作过程中,原发性威胁 t_ε 有可能以概率 $\eta(t_\varepsilon|a,s)$ 作用于任一结点 a(初始状态为 s),若结点 a 受到原发性威胁 t_ε 作用成功,那么该结点的状态 s 将会以概率 $\rho(s'|a,s,t_\varepsilon)$ 转移到状态 s',然后以概率 $\varphi(t'|a,s,t_\varepsilon)$ 向相邻可激活结点发出威胁 t'. 在此过程中,a 发出的威胁 t' 作用于相邻可激活结点有 3 种情况:

1) 威胁 t' 发出失败,威胁传播到此终止. 该情况表明此次爆发的原发威胁 t_ε 只导致了复杂信息系统中的一个资产受到威胁作用.

2) 威胁 t' 发出成功,但对相邻可激活结点作用失败,威胁传播到此终止. 该情况表明此次爆发的原发威胁 t_ε 仍旧只导致了复杂信息系统中的一个资产受到威胁作用.

3) 威胁 t' 发出成功,且对相邻可激活结点作用成功. 该情况表明受到原发性威胁 t_ε 作用的结点 a,其状态转移后发出了新的威胁 t' 并成功作用于相邻可激活结点,若受到威胁 t' 作用成功的结点继续发出威胁作用于其他相邻可激活结点,不断重复这一过程,直到威胁传播在某一结点终止,便形成了一条多结点的威胁传播路径,本文中将采用树的结构描述该威胁传播路径.

不失一般性,任一威胁作用在结点上,都会有上述 3 种情况之一发生,也即在复杂信息系统中,就是通过上述规则威胁从一个结点发出并作用到另外一个结点,或者终止在某一个结点.

2.2.1 威胁传播树的定义

对于任一复杂信息系统拓扑结构图 $G=G(A,E)$(其中 $A=\{a_1,a_2,\cdots,a_n\}$)而言,在有限时间序列 $\Theta=(1,2,\cdots,n-1)$ 上,从初始时刻 1 开始,原发性威胁 t_ε 作用于图中的某一结点 a_1,结点 a_1 受到威胁作用成功后将发出新的威胁 t_1 并通过有向边 $\langle a_1,a_1'\rangle$ 由 a_1 向其相邻可激活结点 a_1' 传播;在下一时刻(即时刻 2),威胁 t_1 作用于结点 a_2,若作用成功,a_2 又将发出新的威胁 t_2 并沿着有向边 $\langle a_2,a_2'\rangle$ 由 a_2 向其相邻可激活结点 a_2' 传播,重复此过程,直到在某一时刻 $i\in\Theta$ 时,所有活动结点 $a_i\in AN$ 不再发出威胁或所发出的威胁对其相邻可激活结点均没

有作用成功为止,威胁传播结束.这里 $a_2=a_1'$,因为在时刻 i 的活动结点 a_i 与时刻 $i-1$ 威胁作用成功的可激活结点 a_{i-1}' 是同一个结点.由于在时刻 $i-1$ 活动结点 a_{i-1} 发出的威胁有可能同时作用于多个相邻可激活结点 a_{i-1}',因此在时刻 i 有可能存在多个活动结点 $a_i\in AN$.为方便描述,以下所涉及的 a_i,t_i 均表示在时刻 i 时的活动结点与该时刻的活动结点发出的威胁,并不指代具体的结点名称和威胁名称,仅表示时间顺序关系.

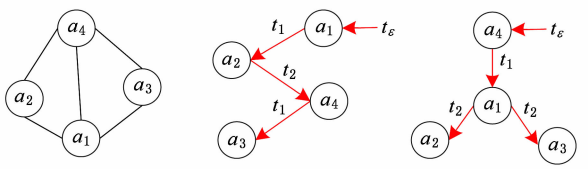
在以上过程中所形成的各相邻死结点间的有向边 $E=\{\langle a_1,a_1'\rangle,\cdots,\langle a_{n-1},a_{n-1}'\rangle\}$ 以及每条有向边上的威胁序列 $\Gamma=(t_1,\cdots,t_{n-1})$ 所组成的有向无环图称作威胁传播树(threat propagation trees, TPT).其中,受到原发性威胁 t_ε 作用的结点称作树根;受到威胁作用后不再发出新的威胁或所发出的新威胁对其相邻可激活结点没有作用成功的结点称作树叶.因此,一棵高度为 $h(0<h<|A|)$ 的威胁传播树可定义为:

$$\gamma(E,\Gamma)=\{t_\varepsilon,\bigcup_{i=1}^{h}(\bigcup_{a_i\in AN}\bigcup_{a_i'\in B_i}(\langle a_i,a_i'\rangle,t_i))\}.\quad(4)$$

由前面对威胁传播的分析可知,威胁在每个时刻最多只会向前传播 1 步,而这一步恰好对应树的高度增加 1,也即威胁传播的时间序列长度与威胁传播树的高度相对应.因此在式(4)中,AN 表示时刻 i 的活动结点集;B_i 表示 a_i 发出的威胁作用成功的相邻可激活结点集,因为在时刻 i,结点 a_i 发出的威胁 t_i 有可能同时作用于 a_i 的多个相邻可激活结点,也即在此时会有多条威胁传播边 $(\langle a_i,a_i'\rangle,t_i)$ 产生.如图 4(c)所示,在时刻 2 的活动结点 a_1 发出的威胁 t_2 对其相邻可激活结点 $B_1=\{a_2,a_3\}$ 同时作用成功,因此同一时刻产生了 2 条威胁传播边 $(\langle a_1,a_2\rangle,t_2)$ 与 $(\langle a_1,a_3\rangle,t_2)$.$E$ 表示构成威胁传播树的边集(利用含有 2 个结点 a_i,a_i' 的二元组 $\langle a_i,a_i'\rangle\in E$ 来表示一条边);Γ 表示边集 E 中每条边上传播的威胁序列,如 t_i 对应于由 a_i 发出并通过边 $\langle a_i,a_i'\rangle$ 传播向结点 a_i' 的威胁.

当树高 $h=0$ 时,威胁传播树 $\gamma(E,\Gamma)$ 只含有单个结点,即 $\gamma(E,\Gamma)=\{(\langle a_1,a_1\rangle,t_\varepsilon)\}$.所以,威胁 Γ 在拓扑结构图 $G=G(A,E)$ 中进行传播时,其高度为 $h=0,1,\cdots,|A|-1$ 的所有类型的威胁传播树可表示为:$\psi=\{(\langle a_1,a_1\rangle,t_\varepsilon)\}\bigcup_{h=1}^{|A|-1}\gamma(E,\Gamma)$.

针对一个实际的复杂信息系统拓扑结构(如图 4(a)),图 4(b)与图 4(c)为威胁 $T=\{t_\varepsilon,t_1,t_2\}$ 在 4 结点网络中传播时所形成的 2 棵不同高度的威胁传播树.在图 4(b)中,威胁传播树的边集为 $E=\{\langle a_1,a_2\rangle,\langle a_2,a_4\rangle,\langle a_4,a_3\rangle\}$,每条边所传播的威胁序列为 $\Gamma=(t_1,t_2,t_1)$,因此该棵树描述为 $\gamma(E,\Gamma)=\{t_\varepsilon,(\langle a_1,a_2\rangle,t_1),(\langle a_2,a_4\rangle,t_2),(\langle a_4,a_3\rangle,t_1)\}$,其中 t_ε 为原发性威胁.同理,在图 4(c)中,威胁传播树的边集为 $E=\{\langle a_4,a_1\rangle,\langle a_1,a_2\rangle,\langle a_1,a_3\rangle\}$,每条边所传播的威胁序列为 $\Gamma=(t_1,t_2,t_2)$,则该棵树描述为 $\gamma(E,\Gamma)=\{t_\varepsilon,(\langle a_4,a_1\rangle,t_1),(\langle a_1,a_2\rangle,t_2),(\langle a_1,a_3\rangle,t_2)\}$.



(a) Topology of system　(b) Propagation tree Ⅰ　(c) Propagation tree Ⅱ

Fig. 4　A case of two threats propagating on complex information system with four nodes.

图 4　2 种威胁在 4 结点复杂信息系统中的 2 种传播情况

由于原发性威胁 t_ε 是以一定概率 $\eta(t_\varepsilon|a,s)$ 作用于复杂信息系统的任一结点 a(初始状态为 s),因此复杂信息系统中的每个结点在时刻 1 都有可能成为树根,而对于每个结点作为 1 次树根时又将可能衍生出多种威胁传播树.因此,如何获得确实有效的威胁传播树是复杂信息系统风险评估的首要步骤.

2.2.2　威胁传播树的生成

在第 1 节中,已将复杂信息系统拓扑结构定义为 $G=G(A,E)$(结点集 A 中的元素个数为 n),根据 2.2.1 节中威胁传播树的定义,可以将原发性威胁作用于复杂信息系统的某一结点时所产生的所有可能的威胁传播树按其高度 h 做如下划分:高度为 0 的威胁传播树(单个结点)、高度为 1 的威胁传播树、…、高度为 $n-1$ 的威胁传播树.

事实上,根据威胁传播树的定义可知,每一棵威胁传播树都是 G 的一个连通子图中的一棵生成树与威胁集 T 的一种组合,也即在一棵生成树的所有边上通过传播威胁集 T 中的不同威胁而得到的组合情况.因此,首先考虑在文献[13]中所描述的威胁传播树生成方法——组合策略(如算法 1 所描述)来生成威胁传播树,将连通图的连通子图生成算法与图的生成树算法相结合来获得复杂信息系统拓扑结构的所有类型的生成树集 Ω(高度为 0 的生成树,高度为 1 的生成树,…,高度为 $n-1$ 的生成树),然后再通过威胁集 T 中的威胁与生成树的边进行组合

而得到所有类型的威胁传播树,最后再计算每棵威胁传播树的形成概率.

算法 1. 组合策略生成威胁传播树算法 *CombinedStrategy*(G).

输入:系统拓扑结构 $G=G(A,E)$;

输出:威胁传播树集合 ψ.

① $\Omega\leftarrow\Omega_0$;

② for $i=1$ to $|A|-2$ do

③ for each $e\in E$ do

④ if $e\in\omega_i$ then continue;

⑤ $\omega_i=\omega_i\bigcup e$;

⑥ if ω_{i+1}不是一棵树 then continue;

⑦ if $\omega_{i+1}\in\Omega_{i+1}$ then continue;

⑧ $\Omega_{i+1}=\Omega_i\bigcup\omega_{i+1}$;

⑨ end for

⑩ $\Omega\leftarrow\Omega\bigcup\Omega_{i+1}$;

⑪ end for

⑫ $\psi\leftarrow\Omega$ are propagated by T;

⑬ return ψ.

在算法 1 的描述中,ω_i 表示高度为 i 的生成树,Ω_i表示只含有高度为 i 生成树集合.但是,该方法在获取生成树的过程中将会面临一个问题——生成树的规模十分庞大.通过程序计算可知,对于具有 n 个结点($a_1,a_2,\cdots,a_n$)的完全图 G,在它的所有连通子图中,以结点 a_1 为根的全部生成树(高度为 0 的生成树、高度为 1 的生成树、…、高度为 $n-1$ 的生成树)的规模如表 1 所示,完全图以任一结点为树根,计算其所有连通子图的全部生成树的数目是相同的,所以不妨以结点 a_1 作为树根.具有 n 个节结点的完全图,其所有连通子图中以某一结点为根的全部生成树规模可由函数 $f(n)=a\times n^b+c$ 近似计算得到,通过表 1 中数据可以得到函数 $f(n)$的参数估计 $a=(1.458\pm1.0360)^{-13}$,$b=20.43\pm0.34$,$c=0\pm469.5$.

在表 1 中,当完全图 G 结点数达到 8 个时,所有连通子图中以结点 a_0 为根的生成树规模就已经达到了 4×10^5 数量级,这仅仅是生成树的规模就已经相当庞大了,若再对生成树与威胁集进行组合得到威胁传播树,其威胁传播树的规模将远超过生成树的规模.因此,通过计算生成树,再根据威胁集与生成树边进行组合的策略,在实际的系统风险评估中,无论是在时间还是空间效率上都有着很大的缺陷.为了克服这一缺陷,需要根据威胁在威胁传播树中的具体传播过程,考虑另外的策略来缩小威胁传播树的规模.

Table 1 Number of Spanning Trees Changes Along with Number of Complete Graph's Nodes

表 1 生成树规模随完全图结点个数的变化

Number of Nodes in Complete Graph	Number of Spanning Trees Rooted at a_1
1	1
2	2
3	6
4	29
5	212
6	2 117
7	26 830
8	412 015

由 2.2 节所描述的复杂信息系统威胁传播所具有的特征 5)可知,在任一时刻 $i\in\Theta$,活动结点 a_i 发出的威胁可以同时作用于多个相邻可激活结点,结合前面生成威胁传播边的过程可知,如果 a_i 的相邻可激活结点集 A_i 的元素个数为 $|A_i|$,且 a_i 可发出的威胁种类数为 N_t,则威胁传播边($\langle a_i,a_i'\rangle,t_i$)的形式有 $N_t\times(2^{|A_i|}-1)$种,显然在每个时刻威胁传播树边的数量均以 $2^{|A_i|}$ 的指数级规模增长,从而也导致了威胁传播树的数量规模也呈现该指数级增长.对于任何一种威胁在复杂信息系统中进行传播时,其所对应的时间序列 Θ 的长度并不唯一(可能的取值为$|\Theta|=1,\cdots,|A|-1$,$|A|$为复杂信息系统结点个数),也即每一个时刻 $i(i\in\Theta)$都将会产生一类高度为 $h=i$ 的威胁传播树(这里不包括单结点威胁传播树).因此,为了降低在时间序列 Θ 上所产生的威胁传播树规模,考虑一种基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法(以下简称"采样策略"),即在威胁的传播过程中对威胁传播树的生成进行如下 2 步采样操作:

1) 在威胁传播过程中,对每时刻活动结点的转移状态进行采样.

根据威胁在复杂信息系统中的传播过程可知,在时刻 $i\in\Theta$,如果受到威胁 $t_{i-1}(t_{i-1}\in\Gamma)$作用的活动结点 a_i(初始状态 s)要发出新的威胁 t_i,则其结点自身必需要经历一个状态转移过程 $s\rightarrow s'$;反之,如果该结点状态转移不成功,则不能发出相应的新威胁,也即威胁传播到此终止.因此,可以基于这样一个过程,考虑对每个活动结点的转移状态进行采样,

而不是考虑所有的可能转移状态,具体过程描述如下.

在时刻 $i\in\Theta$,威胁 $t_{i-1}(t_{i-1}\in\Gamma)$作用于结点 a_i(初始状态为 s),若结点 a_i 受到威胁 t_{i-1}作用成功,那么该结点的状态 s 将会以概率 $\rho(s'|a_i,s,t_{i-1})$转移到状态 s'. 在 2.1 节的资产状态转移描述中,资产的状态转移是一个逐步恶化的过程,这里用 $d(s_{ij})$表示资产 a_i 当前状态 s_{ij} 的恶化程度,j 表示当前状态的可选转移状态编号. 状态的恶化程度越高,其 $d(\cdot)$值越大. 设 a_i 的状态集 S 中的元素已根据$d(\cdot)$值升序排列,显然,a_i(当前状态 s_{ij})的转移状态 s_{ik} 的取值范围为 $S=\{s_{ij},s_{i(j+1)},s_{i(j+2)},\cdots,s_{i(j+m)}\}$(其中 $s_{ik}=s_{ij}$ 表示状态没有发生转移,视为一种特殊的转移状态),且它服从概率分布 $s_{ik}\sim\mathcal{G}(s')$,其中:

$$\mathcal{G}(s'=s_{ik})=\rho(s_{ik}\mid a_i,s_{ij},t_{i-1}). \tag{5}$$

在式(5)中,$s_{ij},s_{ik}\in S$,这里 s_{ij} 表示结点 a_i 受到威胁作用前的状态,s_{ik} 表示结点 a_i 受到威胁作用后的转移状态,且 s_{ik} 可以取值 s_{ij} 表示状态没有发生转移.

因此,根据式(5)计算得到结点 a_i(当前状态 s_{ij})的所有可能转移状态 s_{ik} 的概率分布值 $\mathcal{G}(\cdot)=\{\mathcal{G}(s_{ij}),\mathcal{G}(s_{i(j+1)}),\cdots,\mathcal{G}(s_{i(j+m)})\}$,利用 $\mathcal{G}(\cdot)$中的元素将连续区间(0,1)映射为 $m+1$ 个子区间,使得满足:第一个子区间长度为 $\mathcal{G}(s_{ij})$,第 2 个子区间长度为 $\mathcal{G}(s_{i(j+1)})$,…,第 $m+1$ 个子区间长度为 $\mathcal{G}(s_{i(j+m)})$,如图 5 所示:

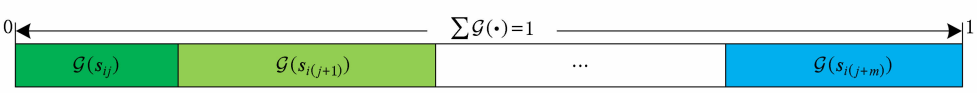


Fig. 5　The mapping from probability distribution values to continuous interval (0,1).

图 5　概率分布值对连续区间(0,1)的映射

然后,从均匀分布 $u\sim U(0,1)$中获取一个样本 u,观察样本 u 落入图 5 的哪一个子区间,也即样本 u 落入区间所对应的随机变量 s_{ik} 就是结点 a_i 的转移状态. 至此,对结点转移状态的采样过程结束,采样算法如算法 2 所示.

算法 2. 结点转移状态采样算法 *GetTransitionState*(a_i).

输入:当前活动结点 a_i;

输出:结点 a_i 的转移状态 s_{ik}.

① 获取结点 a_i 的当前状态及可能的转移状态集$\{s_{ij},s_{i(j+1)},\cdots,s_{i(j+m)}\}$;

② 对$\{s_{ij},s_{i(j+1)},\cdots,s_{i(j+m)}\}$中的每一个状态利用式(5)计算 $\mathcal{G}(\cdot)$;

③ 根据 $\mathcal{G}(\cdot)$将连续区间[0,1]划分为 $m+1$ 个不同的子区间;

④ 从均匀分布 $u\sim U(0,1)$中获取一个随机值 u;

⑤ for each $\mathcal{G}(s_{ik})\in\mathcal{G}(\cdot)$ do

⑥　if u 落在区间 $\mathcal{G}(s_{ik})$ then

⑦　　a_i 的转移状态取值 s_{ik};

⑧　　break;

⑨　end if

⑩ end for

⑪ return a_i 的转移状态 s_{ik}.

在威胁传播树中,活动结点处于不同的状态将发出不同的威胁,通过对结点转移状态的采样,保留一部分转移状态,控制了威胁传播树中的威胁种类(不同的状态有可能发出不同的威胁),进而能够有效地减少威胁传播树中威胁传播边的种类(在威胁传播树中,同一条边可以传播多种威胁. 因此,减少了威胁种类,也就减少了威胁传播边的种类,进而减少威胁传播树的规模).

2) 在威胁传播过程中,对每时刻产生的威胁传播边进行采样.

根据 2.2.1 节中威胁传播树的定义,在时刻 $i\in\Theta$,当威胁 $t_{i-1}(t_{i-1}\in\Gamma)$作用于结点 a_i(初始状态为 s)时,若结点 a_i 受到威胁 t_{i-1}作用成功,那么该结点的状态 s 将会以概率 $\rho(s'|a_i,s,t_{i-1})$转移到状态 s',然后该结点再以概率 $\varphi(t_i|a_i,s,t_{i-1})$向 a_i 的相邻可激活结点集 A_i 发出新的威胁 t_i,如果威胁 t_i 对相邻可激活结点 $a'_i\in B_i$ 作用成功(即威胁 t_i 作用在处于状态 s 的相邻可激活结点 a'_i 上时,成功利用资产 a'_i 的脆弱性 v_k),则在结点 a_i 与 $a'_i\in B_i$ 之间形成威胁传播边($\langle a_i,a'_i\rangle,t_i$),这里 $A_i=B_i\cup C_i$,其中 B_i 表示从结点 a_i 发出威胁并作用成功的相邻可激活结点集,C_i 表示从结点 a_i 发出威胁但作用不成功的相邻可激活结点集. 由于在时刻 i,威胁 t_i 可以同时对 a_i 的多个相邻可激活结点 a'_i 作用成功,所以在时刻 i 产生的威胁传播边有可能是多条,也即 a'_i 可能是多个结点的一个组合情况,如图 4(c),威胁 t_2 对 a_1 的相邻可激活结点 a_2,a_3 同时作用成功,产生 2 条

威胁传播边,应将这 2 条威胁传播边联合起来视为一种类型威胁传播边.因此,在时刻 i 产生威胁传播边$(\langle a_i,a_i'\rangle,t_i)$的概率为

$$p((\langle a_i,a_i'\rangle,t_i))=\varphi(t_i\mid a_i,s,t_{i-1})\times\prod_{a_i'\in B_i}\sum_{v_k\in V}\beta(v_k\mid a_i',s,t_i)\times\prod_{a_i''\in C_i}(1-\sum_{v_k\in V}\beta(v_k\mid a_i'',s,t_i)).\quad(6)$$

在式(6)中,$A_i=B_i\cup C_i$,其中 B_i 表示从结点 a_i 发出威胁并作用成功的相邻可激活结点集,C_i 表示从结点 a_i 发出威胁但作用不成功的相邻可激活结点集,V 表示结点的脆弱性集.

通过步骤 1 在时刻 $i\in\Theta$ 对 a_i 的转移状态进行采样后,a_i 将会转移到一个确定状态,在该状态下活动结点 a_i 向其相邻可激活结点 $a_i'\in A_i$ 发出威胁 $t_i\in T_i$,T_i 表示 a_i 在某一状态下发出的威胁集,若 t_i 对 a_i' 作用成功则形成威胁传播边$(\langle a_i,a_i'\rangle,t_i)$.因此,在时刻 i,由活动结点 a_i 发出的所有可能的威胁传播边为 $E_i=\{\bigcup_{t_i\in T_i}\bigcup_{a_i'\in A_i}(\langle a_i,a_i'\rangle,t_i)\}$,且每种威胁传播边的概率可通过式(6)计算得到.显然,E 中的元素服从概率分布$(\langle a_i,a_i'\rangle,t_i)\sim\mathcal{H}(e)$,其中:

$$\mathcal{H}(e=(\langle a_i,a_i'\rangle,t_i))=p(\langle a_i,a_i'\rangle,t_i)).\quad(7)$$

在式(7)中,$p(\langle a_i,a_i'\rangle,t_i))$表示结点 a_i 向 a_i' 发出威胁并作用于 a_i' 时所产生的威胁传播边$(\langle a_i,a_i'\rangle,t_i)$的概率.

令 $e_j=(\langle a_i,a_i'\rangle,t_i)(e_j\in E_i,j=0,1,\cdots,|E_i|)$,根据式(7)可计算得到,在时刻 $i\in\Theta$,活动结点 a_i 与其相邻可激活结点 a'_i 之间的威胁传播边 e_j 的概率分布值 $\mathcal{H}(\cdot)=\{\bigcup_{j=0}^{|E_i|}\mathcal{H}(e_j)\}$,其中 $j=0$ 表示 a_i 与 A_i 中的结点之间没有形成任何威胁传播边,即 a_i 发出的威胁对 A_i 中的任何一个结点都没有作用成功,其概率为 $\mathcal{H}(e_0)=1-\{\bigcup_{j=1}^{|E_i|}\mathcal{H}(e_j)\}$.利用 $\mathcal{H}(\cdot)$中的元素将连续区间(0,1)映射为$|E_i|+1$ 个子区间,使其满足:第一个子区间长度为 $\mathcal{H}(e_0)$,第 2 个子区间长度为 $\mathcal{H}(e_1)$,…,第$|E_i|+1$ 个子区间长度为 $\mathcal{H}(e_{|E_i|})$,如图 6 所示:

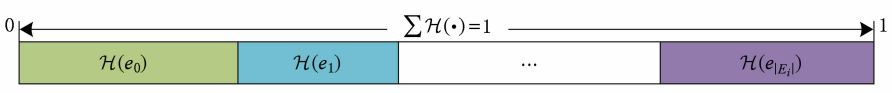


Fig. 6 The mapping from probability distribution values $\mathcal{H}(\cdot)$ to continuous interval (0,1).

图 6 概率分布值 $\mathcal{H}(\cdot)$对连续区间(0,1)的映射

然后,从均匀分布 $u\sim U(0,1)$中获取一个样本 u,观察样本 u 落入图 6 的哪一个子区间,也即样本 u 落入区间所对应的随机变量 e_j 就是当前时刻活动结点 a_i 与其相邻可激活结点 a'_i 之间的威胁传播边$(\langle a_i,a_i'\rangle,t_i)$.至此,对威胁传播边进行采样的过程结束,采样算法如算法 3 所示.

算法 3. 威胁传播边采样算法 *GetPropagationEdge*(a_i,s).

输入:当前活动结点 a_i 和发出威胁 t_i 的状态 s;

输出:威胁传播边$(\langle a_i,a_i'\rangle,t_i)$.

① 获取结点 a_i 的相邻可激活结点集 A_i;

② 根据活动结点 a_i 的状态 s 获取威胁集 T_i;

③ for each $a_i'\in A_i$ do

④ for each $t_i\in T_i$ do

⑤ 利用式(6)计算 $p(\langle a_i,a_i'\rangle,t_i)$;

⑥ $P\leftarrow P\cup p(\langle a_i,a_i'\rangle,t_i))$;

⑦ end for

⑧ end for

⑨ 利用式(7)计算 P 中每一个元素的 $\mathcal{H}(\cdot)$;

⑩ 利用 $\mathcal{H}(\cdot)$将连续区间[0,1]划分为$|\mathcal{H}(\cdot)|$个不同的子区间;

⑪ 从均匀分布 $u\sim U(0,1)$中获取一个随机值 u;

⑫ for each $\mathcal{H}(\langle a_i,a_i'\rangle,t_i)\in\mathcal{H}(\cdot)$ do

⑬ if u 落在区间 $\mathcal{H}(\langle a_i,a_i'\rangle,t_i)$ then

⑭ $e\leftarrow(\langle a_i,a_i'\rangle,t_i)$;

⑮ break;

⑯ end if

⑰ end for

⑱ return e.

在威胁的传播过程中,某一时刻有可能存在多个活动结点,而每个活动结点如果存在多个相邻可激活结点,则有可能产生多种威胁传播边.因此,通过算法 3 对威胁传播边进行采样,能够为威胁传播树的当前活动结点 a_i 保留一种威胁传播边,进一步降低了威胁传播树的规模.

在威胁的传播过程中,通过算法 2 与算法 3 对当前时刻活动结点的转移状态及其发出的威胁所形成的威胁传播边分别进行采样后,迅速地降低了威胁传播树的规模,提高了威胁传播树的生成速度.因此,对任一复杂信息系统拓扑结构图 $G=G(A,E)$

(其中 $A=\{a_1,a_2,\cdots,a_n\}$)而言,从某一时刻原发性威胁 t_ε 作用于 G 中的某一个结点开始,在有限时间序列 $\Theta=(1,2,\cdots,n-1)$ 上生成威胁传播树的算法如算法 4 所述.

算法 4. 采样策略生成威胁传播树算法 *CreateThreatPropagationTree*(G,a_r).

输入:系统拓扑结构 $G=\mathcal{G}(A,E)$和树根 $a_r\in A$;

输出:威胁传播树集合 γ.

① $a_1\leftarrow a_r,\Theta\leftarrow(1,2,\cdots,|A|-1)$;
② $s\leftarrow GetTransitionState(a_1)$;
③ if $s=$Null then
④　return Null;
⑤ end if
⑥ $e\leftarrow GetPropagationEdge(a_1,s)$;
⑦ if $e=$Null then
⑧　$e\leftarrow(\langle a_1,a_1\rangle,t_\varepsilon)$;
⑨　$\gamma\leftarrow\{t_\varepsilon,(\langle a_1,a_1\rangle,t_\varepsilon)\}$;
⑩　return γ;
⑪ end if
⑫ $\gamma\leftarrow\{t_\varepsilon,e\}$;
⑬ for $i=2$ to $|\Theta|$ do
⑭　获取当前活动结点集合 AN;
⑮　$E_i\leftarrow$Null;
⑯　for each $a_i\in AN$ do
⑰　　$s\leftarrow GetTransitionState(a_i)$;
⑱　　if $s=$Null then
⑲　　continue;
⑳　　end if
㉑　　$e\leftarrow GetPropagationEdge(a_i,s)$;
㉒　　if $e\neq$Null then
㉓　　$\gamma\leftarrow\gamma\cup e$;
㉔　　end if
㉕　　$E_i\leftarrow E_i\cup e$;
㉖　end for
㉗　if each $e\in E_i$ is Null then
㉘　　威胁传播结束;
㉙　　break;
㉚　end if
㉛ end for
㉜ return γ.

通过算法 4 可知,执行该算法 1 次至多会产生一棵威胁传播树,因此在时间序列 Θ 上,若以相同结点作为威胁传播树的根结点,进行多次模拟该过程(多次执行该算法),便可以得到规模可控的威胁传播树集合 $\psi=\{\gamma_1,\gamma_2,\gamma_3,\cdots\}$,并且所获得的威胁传播树集合 ψ 中的元素(威胁传播树)将服从一定的概率分布,进而可以利用集合 ψ 中的威胁传播树对复杂信息系统的风险进行估计.

3　复杂信息系统风险评估

一定时间段内,复杂信息系统中所有受到威胁作用的资产结点的期望损失之和称为系统风险.损失是威胁通过利用资产的脆弱性导致资产价值的改变而产生,威胁在复杂信息系统中传播形成了威胁传播树,一定时间段内产生的威胁传播树所带来的系统资产损失期望即为系统风险.因此,系统风险的计算应包括 2 个方面:1)威胁传播树的产生概率;2)威胁传播树中各资产结点的价值损失.

3.1　威胁传播树的概率

由 2.2.1 节中威胁传播树的定义可知,本文是将威胁在复杂信息系统中的传播路径定义为一棵威胁传播树.事实上,通过这样一种定义,就将在一个时间段内对复杂信息系统的风险评估转换为计算对该时间段内可能产生的的威胁传播树所带来的风险.由于威胁在复杂信息系统中的传播具有一定的概率性,显然,计算一定时间段内所产生的威胁传播树以及该威胁传播树的产生概率是一必需的过程.

因此,计算一定时间段内复杂信息系统风险发生的概率,也即是计算在该时间段内威胁传播所形成的每棵威胁传播树 γ 所出现的概率 $P(\gamma)$.在 2.2.1 节中,威胁传播树被定义为 2 元组 $\gamma(E,\Gamma)$,其中 $E=\{\langle a_1,a_1'\rangle,\cdots,\langle a_{n-1},a_{n-1}'\rangle\}$为连结各死结点间的有向边集合,$\Gamma=(t_1,\cdots,t_{n-1})$为 E 中各条边上所传播的威胁序列.由于威胁在复杂信息系统中作用于结点是具有时间先后顺序的,故可将组成 $\gamma(E,\Gamma)$中所有边的结点描述为结点序列 $\Lambda=\Lambda_i\cup\Lambda_n$,其中 $i<n$ 表示威胁作用的结点序列编号,$\Lambda_i=(a_1,a_2,\cdots,a_i)$为非叶结点序列,$\Lambda_n$ 为叶结点序列,a_1 为根结点;同理,在 $\gamma(E,\Gamma)$上传播的威胁描述为一个威胁序列 $\Gamma=\Gamma_i\cup\Gamma_n$,其中 $i<n$ 表示威胁序列编号,$\Gamma_i=(t_\varepsilon,t_1,t_2,\cdots,t_i)$为原发性威胁 t_ε 与非叶结点发出的威胁所组成的序列,Γ_n 为叶结点发出的威胁序列,t_ε 直接作用于根结点 a_1.结点序列 Λ 与威胁序列 Γ 有着一一对应关系:威胁 $t_i(\neq t_\varepsilon)$是由结点 a_i 所发出的.威胁在传播过程中,设 B_i 表示从结点 a_i 发出威胁并作用成功的相邻可激活结点集,

C_i 表示从结点 a_i 发出的威胁没有作用成功的相邻可激活结点集，$A_i=B_i\cup C_i$ 表示结点 a_i 的所有相邻可激活结点集. 因此，结合威胁在复杂信息系统中传播过程的描述及其所形成的威胁传播树的定义，可将威胁在传播时所形成的威胁传播树 γ_j 的概率 $P(\gamma_j)$ 描述如下：

$$P(\gamma_j)=\eta(t_\varepsilon\mid a_1,s)\prod_{a_i\in\Lambda_i,t_i\in\Gamma_i-t_\varepsilon}[\varphi(t_i\mid a_i,s,t_{i-1})\times\prod_{a_i'\in B_i}\sum_{v_k\in V}\beta(v_k\mid a_i',s,t_i)\prod_{a_i'\in C_i}(1-\sum_{v_k\in V}\beta(v_k\mid a_i',s,t_i))]\times\prod_{a_n\in\Lambda_n}[\sum_{t_n\in\Gamma_n}(\varphi(t_n\mid a_n,s,t_{n-1})\times\prod_{a_n'\in\Lambda_n}(1-\sum_{v_k\in V}\beta(v_k\mid a_n',s,t_n)))+(1-\sum_{t_n\in\Gamma_n}\varphi(t_n\mid a_n,s,t_{n-1}))],\ 2\leqslant n\leqslant|A|. \tag{8}$$

在式(8)中，$\eta(t_\varepsilon\mid a_i,s)$ 为威胁传播树 $\gamma_j(\Lambda,\Gamma)$ 的根结点 a_i 爆发原发性威胁的概率；$\prod[\varphi\prod(\sum\beta)\prod(1-\sum\beta)]$ 为威胁传播树中所有非叶结点的威胁传播概率，即任一结点的发出威胁概率 φ 乘以所有作用成功的相邻结点概率 $\prod(\sum\beta)$ 与作用不成功的相邻结点概率 $\prod(1-\sum\beta)$；$\prod[\sum(\varphi\prod(1-\sum\beta))+(1-\sum\varphi)]$ 为威胁传播树中所有叶结点的威胁传播概率，由于叶结点为威胁传播的终止结点，威胁在每一个叶结点上终止传播有 2 种可能情况：1) 威胁已向相邻可激活结点发出，但是没有作用成功，其概率为 $\sum(\varphi\prod(1-\sum\beta))$；2) 威胁没有向相邻可激活结点发出，其概率为 $(1-\sum\varphi)$，因此每个叶结点的威胁传播概率 $\sum(\varphi\prod(1-\sum\beta))+(1-\sum\varphi)$ 的乘积即为威胁传播树中所有叶结点的威胁传播概率，这里 $\sum$ 与 $\prod$ 分别表示连加与连乘. 在式(8)中，当 $n=1$，也即威胁传播树 γ_j 只有一个结点时，公式演化为

$$P(\gamma_j)=\eta(t_\varepsilon\mid a_i,s)[\sum_{t_i\in\Gamma_i-t_\varepsilon}(\varphi(t_i\mid a_i,s,t_\varepsilon)\times\prod_{a_i'\in\Lambda_i}(1-\sum_{v_k\in V}\beta(v_k\mid a_i',s,t_i)))+(1-\sum_{t_i\in\Gamma_i-t_\varepsilon}\varphi(t_i\mid a_i,s,t_\varepsilon))]. \tag{9}$$

在式(9)中，由于当前威胁传播树只有一个结点，即根结点 a_i 也是叶结点，在公式中不存在非叶结点的传播概率，因此威胁传播树的形成概率只需计算根结点的原发性威胁的发生概率与叶结点的传播概率之积.

3.2 资产结点的价值损失期望

威胁在复杂信息系统中传播时，如果原发性威胁作用的结点不同，或每个活动结点发出的传播性威胁不同，即便是具有相同结构的威胁传播树，其结点所产生的价值损失也将不同，而且结点的损失还与结点被威胁作用后的转移状态有关. 显然，在计算威胁传播树中的资产结点的价值损失时，必需同时考虑威胁传播树 γ_j 中单个资产结点 a 受到威胁 t 作用时的状态转移($s\to s'$)概率 $\rho(s'\mid a,s,t)$，以及状态转移所造成的结点价值损失 $\Delta W(\Delta W=W(a,s)-W(a,s'))$. 因此，一棵威胁传播树所带来的系统期望价值损失表述为威胁传播树中所有受到威胁作用的结点的状态转移概率与其状态转移所造成的价值损失之积的累加和，即对于任意一棵威胁传播树 γ_j，期望价值损失 $L(\gamma_j)$ 计算如下：

$$L(\gamma_j)=\sum_{a\in\Lambda}\rho(s'\mid a,s,t)(W(a,s)-W(a,s')). \tag{10}$$

复杂信息系统中资产数量庞大，原发性威胁在不同资产上的发生概率不尽相同，而且即便是同一个资产上的相同原发性威胁也有可能产生不同的威胁传播树. 因此，对于任一复杂信息系统拓扑结构图 $G=G(A,E)(A=\{a_1,a_2,\cdots,a_n\})$ 而言，在有限时间序列 $\Theta=(1,2,\cdots,n-1)$ 上，威胁 $T=\{t_0,t_1,\cdots,t_M\}$ 作用在该拓扑结构所对应的复杂信息系统的结点上时，将产生出一棵威胁传播树 γ. 若复杂信息系统经历由多个时间序列 Θ 所组成的时间段，则将可能产生出多棵威胁传播树而形成威胁传播树集合 $\psi=\{\gamma_1,\gamma_2,\gamma_3,\cdots\}$. 算法 4 只能生成复杂信息系统在时间序列 Θ 上可能出现的一棵威胁传播树，因此在由多个时间序列 Θ 所组成的时间段上，可通过多次运行该算法来生成复杂信息系统在一段时间内可能产生多棵不同的威胁传播树. 由于不同种类的威胁传播树除了其本身结构不同，而且其产生的概率也不同. 因此，不同的威胁传播树将会给复杂信息系统带来不同的风险.

3.3 系统风险评估

根据系统风险的定义可知，任意一棵威胁传播树 γ_j 的期望风险 $R(\gamma_j)$ 可表示为该棵威胁传播树的期望损失 $L(\gamma_j)$ 与形成该棵威胁传播树的概率 $P(\gamma_j)$ 之积，也即：

$$R(\gamma_j)=P(\gamma_j)L(\gamma_j). \tag{11}$$

在式(11)中，每棵威胁传播树所带来的期望价值损失 $L(\gamma_j)$ 可通过式(10)计算得到，而形成任意一棵威胁传播树 $\gamma_j(\Lambda,\Gamma)$ 的概率 $P(\gamma_j)$ 则可根据

3.1 节中的式(8)或式(9)(其中式(8)为多结点威胁传播树产生概率计算公式,式(9)为单结点威胁传播树产生概率计算公式)得到.一般情况下,构成一个复杂信息系统的资产有多种,也即组成复杂信息系统拓扑结构 $G=\mathcal{G}(A,E)$ 的结点有多个,结合算法 4 可知,威胁 $T=\{t_0,t_1,\cdots,t_M\}$ 在复杂信息系统中传播时,将会产生多种威胁传播树而构成集合 $\psi=\{\gamma_1,\gamma_2,\gamma_3,\cdots\}$(在实验中通过多次运行算法 4 即可得到该威胁传播树集合).因此,复杂信息系统风险 $R(A,T)$ 即为威胁传播树集合 ψ 中所有威胁传播树的期望风险之和,即:

$$R(A,T)=\sum_{\gamma_j\in\psi}R(\gamma_j). \tag{12}$$

通过算法 4 并结合式(12)便能计算出复杂信息系统在一段时间内所面临的系统期望风险.因此,在时间序列 $\Theta=(1,2,\cdots,n-1)$ 上,对于任一复杂信息系统拓扑结构图 $G=G(A,E)$(其中 $A=\{a_1,a_2,\cdots,a_n\}$)而言,威胁 $T=\{t_0,t_1,\cdots,t_M\}$ 作用在该拓扑结构所对应的复杂信息系统上时,所产生的系统风险的计算过程可描述为算法 5 所示.

算法 5. 复杂信息系统风险评估算法 *ComputeExpectedRisk*(G,a_r).

输入:系统拓扑结构 $G=G(A,E)$ 和树根 $a_r\in A$;

输出:$G=G(A,E)$ 的期望风险 $R(A,T)$.

① $\psi\leftarrow$Null;

② for $i=1$ to *samplingNum* do

③　$\gamma_i\leftarrow$*CreateThreatPropagationTree*(G,a_r);

④　计算式(10)所描述的威胁传播树 γ_i 的期望损失 $L(\gamma_i)$;

⑤　计算式(11)所描述的威胁传播树 γ_i 的期望损失 $R(\gamma_i)$;

⑥　$\psi\leftarrow\psi\cup\gamma_i$;

⑦ end for

⑧ for each $\gamma\in\psi$ do

⑨　计算式(12)描述的系统期望风险 $R(A,T)$;

⑩ end for

⑪ return $R(A,T)$.

4　实验分析

为验证本文提出的基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法的有效性,本节将选取文献[13]中的组合策略生成威胁传播树的方法与本文的基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法进行对照实验.在实验中,主要集中在 2 方面展开对比:1) 算法 1 与算法 4 针对不同结点规模的复杂信息系统生成威胁传播树时,在时间效率方面进行对比;2) 算法 1 与算法 4 针对同一个复杂信息系统进行风险评估时,对其评估结果的准确度及安全性建议进行对比.实验均是在相同的系统平台(处理器 Intel® Core™ i7-3770 CPU 3.4 GHz,内存 2.00 GB,操作系统 Windows 7 Professional)和编程环境(Java 1.7)下完成.

4.1　威胁传播树生成算法时空效率分析

在 2.2.2 节中已经叙述到,组合策略生成威胁传播树方法(算法 1),是将连通图的连通子图生成算法与图的生成树算法相结合来获得复杂信息系统拓扑结构的所有类型的生成树,然后再通过威胁集 T 中的威胁与生成树的边进行组合而得到所有类型的威胁传播树.然而,基于结点状态及威胁传播边进行采样的威胁传播树生成算法,是在威胁传播的过程中对每一时刻活动结点的转移状态以及状态转移后向相邻可激活结点发出的威胁传播边进行采样.

为证明文章所提出的基于转移状态和威胁传播边进行采样的威胁传播树生成算法(简称"采样策略")在时空效率上的优越性,本节针对复杂信息系统拓扑结构图是一个完全图的情况下,对于不同规模结点集,分别运行算法 1 所示组合策略方法与算法 4 所示采样策略算法各 10 000 次,然后得到其各自的平均时空耗费如表 2 所示.表 2 描述了 2 种不同算法对于拥有不同结点规模的复杂信息系统拓扑结构在生成威胁传播树时的平均时空效率.

从表 2 中的数据可以看出,作为拓扑结构图为完全图的复杂信息系统,当其结点规模超过 6 个时,组合策略方法生成威胁传播树所耗时间与内存明显多于基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法;当结点规模超过 9 个时,在 2 GB 系统内存的情况下,运行组合策略方法已经内存耗尽,无法继续运行,而基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法所耗时间还不到 1 s,且内存消耗也没有达到 100 MB.显然,与组合策略方法生成威胁传播树相比较,基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法对于拥有结点规模较大的复杂信息系统进行风险评估时,在运行时空效率方面有着明显的优势.

Table 2 Time-Memory Efficiency of Constructing Threat Propagation Trees for Different Number of Nodes

表 2 不同结点规模下生成威胁传播树时空效率比较

Number of Nodes	Combined Strategy		Sampling Strategy	
	Time/s	Memory/MB	Time/s	Memory/MB
1	0.000 1	2.001 0	0.015 8	4.028 0
2	0.000 6	2.040 3	0.017 8	4.347 1
3	0.001 3	2.160 0	0.018 6	4.829 9
4	0.008 9	3.586 0	0.035 4	6.747 6
5	0.025 0	23.580 0	0.046 8	13.948 0
6	0.094 6	103.929 7	0.068 4	24.829 4
7	1.198 0	687.804 6	0.114 5	31.401 1
8	55.447	1 031.852 0	0.154 7	47.188 7
9	Over Time	Over Memory	0.277 4	97.007 9
10	Over Time	Over Memory	0.368 0	193.132 9

在表 3 中,列出了针对更多结点数目的完全图利用采样策略生成威胁传播树的时间效率. 可以看出,当结点规模达到 30 个时所耗时间才相当于组合策略对应于 8 个结点所耗时间,很明显,采样策略在生成威胁传播树的时间效率上明显优于组合策略.

Table 3 Time Efficiency of Sampling Strategy for Different Number of Nodes

表 3 不同结点规模下采样策略生成威胁传播树时间效率

Number of Nodes	Time Efficiency/s	Number of Nodes	Time Efficiency/s
11	0.53	23	71.63
12	0.99	24	116.29
13	1.12	25	168.56
14	1.91	26	243.25
15	2.71	27	379.24
16	3.87	28	545.84
17	7.07	29	831.19
18	8.35	30	1 245.80
19	15.31	31	1 876.88
20	19.85)	32	2 812.23
21	31.68	33	4 201.69
22	47.03	34	6 309.42

事实上,无论是组合策略还是采样策略,其算法的时空复杂度均是关于复杂系统拓扑结构图边的数量的. 组合策略需要遍历拓扑结构图中所有边的组合情况,而采样策略只需在每次进行采样时,仅遍历当前活动结点与相邻可激活结点之间的边的组合情况,这样就大大降低了生成威胁传播树时所需空间与时间的消耗. 在实际生活中,复杂网络拓扑结构图是完全图的可能性较低,也即在实际生活中所遇到的复杂网络拓扑结构图大部分都是非完全图的. 因此采样策略在实际使用中的效率将会更高.

4.2 安全风险评估结果分析

为进一步验证本文提出的基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法(采样策略)在复杂信息系统风险评估中的优势性,本节将从系统期望风险及安全建议 2 个方面与采用组合策略生成威胁传播树方法进行横向比较和分析,以阐述基于采样策略的风险评估模型在复杂信息系统领域中的应用及优势.

在实验中,以图 7 所示的 8 结点网络拓扑结构 $G=G(A,E)$, $A=\{0,1,\cdots,7\}$ 为例,计算若其中某一个结点爆发原发性威胁时,对系统可能造成的期望风险.

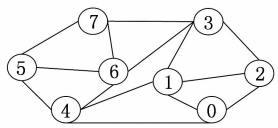


Fig. 7 The complex information system topology graph with 8 nodes.

图 7 拥有 8 个结点的复杂信息系统拓扑结构

在实际应用中,是通过复杂信息系统的历史运行记录来确定系统中每个结点所拥有的状态集 S,脆弱性集 V,威胁集 T. 这里不妨设图 7 所示拓扑结构图中每个结点的状态集 $S=\{s_0,s_1\}$(s_0 表示结点初始状态,s_1 表示结点受到威胁作用后的转移状态),脆弱性集 $V=\{v_0,v_1\}$,威胁集 $T=\{t_\varepsilon,t_0,t_1\}$(其中 t_ε 为原发性威胁). 根据本文第 1 节和第 2 节所述的相应约束条件进行随机初始化每个结点的状态转移概率 δ,原发威胁概率 η,发出威胁概率 ξ 以及威胁作用概率 β;每个结点的状态转移价值损失 ΔW 在满足 $\sum_{i\in A}\Delta W_i=100$ 的条件下进行随机初始化,如表 4 所示:

Table 4 Value Loss ΔW from Node's State Transition

表 4 结点状态转移价值损失 ΔW

Node ID	Original State	Transitional State	Loss Value ΔW
0	s_0	s_1	20.389
1	s_0	s_1	6.892
2	s_0	s_1	15.073
3	s_0	s_1	21.465
4	s_0	s_1	3.074
5	s_0	s_1	7.402
6	s_0	s_1	13.839
7	s_0	s_1	11.866

4.2.1　系统期望风险

对于图 7 所示的 8 结点网络拓扑结构,不妨设原发性威胁 t_ε 作用于结点 a_6,则通过算法 1 所述的组合策略方法生成的威胁传播树共有 69 750 棵,面对规模如此庞大的威胁传播树,如果计算每一棵威胁传播树所带来的系统期望风险则显得不切实际.在实际应用中,人们更多关注的是发生概率较高的威胁传播树,因此,在该实验中只将本文所提出的威胁传播树生成算法(见表 5)与通过组合策略方法生成的发生概率最高的 20 棵威胁传播树(见表 6)进行对照分析.

在表 6 中,不妨以威胁传播树"$6:47t_1;7:5t_0$"为例来叙述其表示意义(表 5 中的威胁传播树表示的意义与表 6 中的相同),结点 6 受到原发性威胁 t_ε 作用成功并发生状态转移,在转移到新状态后又向其所有相邻可激活结点发出新的威胁 t_1,然而此时威胁只对相邻可激活结点 4 和 7 作用成功并发生状态转移,结点 4 和 7 发生状态转移后又向其各自的相邻可激活结点发出新的威胁,但是结点 4 发出的新威胁对其相邻可激活结点均没有作用成功,而结点 7 发出的新威胁 t_0 只对其相邻可激活结点 5 作用成功,结点 5 发生状态转移后也会向其相邻可激活结点发出新威胁,但是新威胁对其相邻可激活结点并没有作用成功.至此,威胁在拓扑结构中的传播终止,形成一棵威胁传播树.由图 7 可知,在结点 6 受到原发性威胁 t_ε 作用时,结点 6 的相邻可激活结点集为{3,4,5,7},由于不同结点对威胁的安全防御能力有可能不同,因此不一定所有相邻可激活结点都被威胁作用成功.因此出现了"$6:47t_1;7:5t_0$"类型的威胁传播树.

Table 5　Threat Propagation Trees Constructed by Sampling Strategy with 2 000 Sampling

表 5　采样策略 2 000 次采样获得的威胁传播树

TPT	Expected Loss of Value for Nodes								Probability of TPT	Expected Risk of TPT	Expected Risk of System
	0	1	2	3	4	5	6	7			
6	0.00	0.00	0.00	0.00	0.00	0.00	2.57	0.00	4.89E−02	1.26E−01	
$6:357t_0$	0.00	0.00	0.00	4.40	0.00	0.34	2.56	1.68	7.33E−04	6.59E−03	
$6:37t_0$	0.00	0.00	0.00	4.40	0.00	0.00	2.56	1.68	6.38E−04	5.52E−03	
$6:57t_0$	0.00	0.00	0.00	0.00	0.00	0.34	2.56	1.68	6.28E−04	2.88E−03	
$6:7t_0$	0.00	0.00	0.00	0.00	0.00	0.00	2.56	1.68	5.47E−04	2.32E−03	
$6:7t_1$	0.00	0.00	0.00	0.00	0.00	0.00	2.56	4.71	5.16E−04	3.76E−03	
$6:47t_1$	0.00	0.00	0.00	0.00	0.58	0.00	2.56	4.71	4.83E−04	3.80E−03	
$6:57t_1$	0.00	0.00	0.00	0.00	0.00	1.44	2.56	4.71	4.79E−04	4.17E−03	
$6:457t_1$	0.00	0.00	0.00	0.00	0.58	1.44	2.56	4.71	4.48E−04	4.17E−03	
$6:37t_1$	0.00	0.00	0.00	1.41	0.00	0.00	2.56	4.71	4.48E−04	3.89E−03	1.83E−01
$6:347t_1$	0.00	0.00	0.00	1.41	0.58	0.00	2.56	4.71	4.19E−04	3.88E−03	
$6:357t_1$	0.00	0.00	0.00	1.41	0.00	1.44	2.56	4.71	4.15E−04	4.21E−03	
$6:3457t_1$	0.00	0.00	0.00	1.41	0.58	1.44	2.56	4.71	3.89E−04	4.16E−03	
$6:3457t_0$	0.00	0.00	0.00	4.40	0.21	0.34	2.56	1.68	3.46E−04	3.18E−03	
$6:347t_0$	0.00	0.00	0.00	4.40	0.21	0.00	2.56	1.68	3.01E−04	2.67E−03	
$6:457t_0$	0.00	0.00	0.00	0.00	0.21	0.34	2.56	1.68	2.97E−04	1.42E−03	
$6:47t_0$	0.00	0.00	0.00	0.00	0.21	0.00	2.56	1.68	2.58E−04	1.15E−03	
$6:347t_1;4:01t_1$	1.97	1.05	0.00	1.41	0.58	0.00	2.56	4.71	6.63E−06	8.15E−05	
$6:457t_1;4:0t_0$	4.00	0.00	0.00	0.00	0.58	1.44	2.56	4.71	1.95E−06	2.60E−05	
$6:345t_0;3:7t_0$	0.00	0.00	0.00	4.40	0.21	0.34	2.56	1.68	2.62E−08	2.41E−07	

假设表 6 中的数据是通过模拟一段时间内复杂信息系统的工作情况而得到的,则通过该表中的每棵威胁传播树期望风险再结合式(12),即可计算得到未来一段时间内该系统的期望风险为 1.87E−01.但是组合策略方法并不能为复杂信息系统给出一个人为可控的工作时间,且每运行组合策略方法 1 次,

Table 6 Threat Propagation Trees with Top 20 Probability Constructed by Combined Strategy

表 6 组合策略方法生成的发生概率最高的 20 棵威胁传播树

TPT	Expected Loss of Value for Nodes								Probability of TPT	Expected Risk of TPT	Expected Risk of System
	0	1	2	3	4	5	6	7			
6	0.00	0.00	0.00	0.00	0.00	0.00	2.56	0.00	4.90E−02	1.26E−01	
$6:357t_0$	0.00	0.00	0.00	4.40	0.00	0.34	2.56	1.68	7.33E−04	6.59E−03	
$6:37t_0$	0.00	0.00	0.00	4.40	0.00	0.00	2.56	1.68	6.38E−04	5.52E−03	
$6:57t_0$	0.00	0.00	0.00	0.00	0.00	0.34	2.56	1.68	6.28E−04	2.88E−03	
$6:7t_0$	0.00	0.00	0.00	0.00	0.00	0.00	2.56	1.68	5.47E−04	2.32E−03	
$6:7t_1$	0.00	0.00	0.00	0.00	0.00	0.00	2.56	4.71	5.16E−04	3.76E−03	
$6:47t_1$	0.00	0.00	0.00	0.00	0.58	0.00	2.56	4.71	4.83E−04	3.80E−03	
$6:57t_1$	0.00	0.00	0.00	0.00	0.00	1.44	2.56	4.71	4.79E−04	4.17E−03	
$6:457t_1$	0.00	0.00	0.00	0.00	0.58	1.44	2.56	4.71	4.48E−04	4.17E−03	
$6:37t_1$	0.00	0.00	0.00	1.41	0.00	0.00	2.56	4.71	4.48E−04	3.89E−03	1.87E−01
$6:347t_1$	0.00	0.00	0.00	1.41	0.58	0.00	2.56	4.71	4.19E−04	3.88E−03	
$6:357t_1$	0.00	0.00	0.00	1.41	0.00	1.44	2.56	4.71	4.15E−04	4.21E−03	
$6:3457t_1$	0.00	0.00	0.00	1.41	0.58	1.44	2.56	4.71	3.89E−04	4.16E−03	
$6:3457t_0$	0.00	0.00	0.00	4.40	0.21	0.34	2.56	1.68	3.46E−04	3.18E−03	
$6:347t_0$	0.00	0.00	0.00	4.40	0.21	0.00	2.56	1.68	3.01E−04	2.67E−03	
$6:457t_0$	0.00	0.00	0.00	0.00	0.21	0.34	2.56	1.68	2.97E−04	1.42E−03	
$6:47t_0$	0.00	0.00	0.00	0.00	0.21	0.00	2.56	1.68	2.58E−04	1.15E−03	
$6:47t_1;7:3t_1$	0.00	0.00	0.00	1.41	0.58	0.00	2.56	4.71	1.24E−04	1.15E−03	
$6:47t_1;7:5t_0$	0.00	0.00	0.00	0.00	0.58	0.34	2.56	4.71	1.24E−04	1.01E−03	
$6:47t_1;7:5t_1$	0.00	0.00	0.00	0.00	0.58	1.44	2.56	4.71	1.19E−04	1.11E−03	

都需要计算出所有可能的威胁传播树,当面临结点规模较大的复杂信息系统时,在有限的计算资源上,该方法将会失效.然而,基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法(采样策略)除了能够计算得到与组合策略方法有着极小误差的系统期望风险(见表 5 所示),而且算法还能够控制复杂信息系统的模拟工作时间(采样策略算法 4 的运行次数),面对规模庞大的复杂信息系统时还能保持较高的运行效率.

采样策略通过对图 7 所示拓扑结构进行 2 000 次威胁传播采样获得了表 5 所示的 20 棵威胁传播树的信息.在算法工作的过程中,1 次采样过程可以视为复杂信息系统的一个工作时间段,因此可以通过采样次数的设置来达到模拟控制复杂信息系统的工作时间.通过式(12)结合表 5 中的威胁传播树期望风险,便可计算得到在 2 000 个工作时间段内,复杂信息系统的期望风险为 1.83E−01.

对比表 6 与表 5 中的系统期望风险值可以得出,采样策略获得的系统期望风险值略小于组合策略方法所得到的系统期望风险值,即在系统期望风险的计算精度上,采样策略较组合策略降低了约 2.14%,这显然是非常符合生成威胁传播树的 2 种不同算法的特性.组合策略是从全局整体解空间去搜索最优解,得到的威胁传播树信息较为全面,而采样策略是从一个局部空间出发以一定的概率分布去近似逼近全局最优解,显然会带来一定的威胁传播树信息的损失,但是这种对实际最优解的近似方法所带来的信息损失是可以接受的.虽然在系统风险值的计算中采样策略获得了较组合策略稍低的精度,但是随着复杂网络结点数目的增加,采样策略在运行时间效率上却得到了显著的提升,从表 2 中可以看出,当结点数目达到 8 个时,采样策略算法的时间效率提升超过了 300 多倍,且内存的消耗量也大大低于组合策略.所以作为系统整体风险值的精度损失所带来的评估效率的提升是完全有价值的.

再次通过表 5 与表 6 中的数据对比可以得到,2 种算法获得的威胁传播树有 85%的相同率,且结点{3,4,5,6,7}都包括在这些威胁传播树中;此外,采

样策略比起组合策略方法生成的威胁传播树还多包含另外 2 个结点{0,1}.因此,可以看出采样策略在复杂信息系统风险评估过程中,与组合策略方法相比,在能够保证风险评估准确性的情况下,还具有较高的时间效率和更全面的结点覆盖性.

4.2.2　安全建议

信息的安全需要信息系统的安全来保证,而信息系统的安全则需要通过系统的安全检测机制、安全防护机制、和安全恢复机制这一过程来完成.事实上,系统安全检测机制就是对系统的风险评估,通过分析系统中的哪些资产结点最有可能出现风险来制定相应的安全防护机制.因此,在复杂信息系统的安全风险评估过程中,确定各资产结点受到威胁攻击的可能性显得极为重要.

在 4.2.1 节中已经知道每棵威胁传播树的详细信息,因此可以通过统计每个资产结点在威胁传播树中出现的频率作为其结点重要度的一个评估值(频率越高则重要度越高).如 4.2.1 节中的表 5 与表 6 所示,通过组合策略与采样策略各自获得的威胁传播树集合中,同一类结点在各自的集合中出现的频率并不相同,各结点出现频率的详细统计结果如图 8 所示:

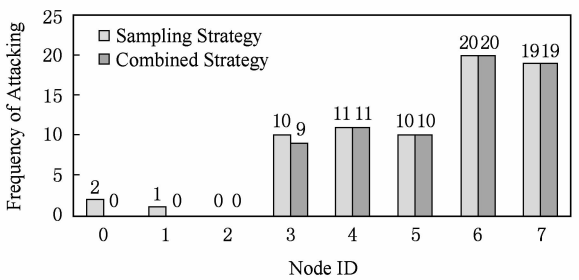


Fig. 8　The frequency of different nodes emerging in different threat propagation trees.

图 8　不同结点在不同威胁传播树集合中出现的频率

按照传统的安全防护策略,当原发性威胁作用于结点 6 时,首先需要实施安全保护的是结点 6 的所有相邻结点{3,4,5,7},由于结点{3,4,5,7}都与结点 6 直接相连,则会对这些结点施以同等的防护力度.事实上,实际的安全防护资源是有限的,并不可能做到结点 6 的每个相邻结点都进行相同力度的防护,即便是做到了对结点 6 的所有相邻结点进行同等力度的防护,也会出现某些结点因为威胁作用频繁而缺乏保护力度被攻破,某些结点因为过度保护而造成保护资源浪费的问题.本文提出的基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法(采样策略)在评估系统风险的过程中却能很好地解决这一问题.

从图 8 所示的结点在不同威胁传播树集合中出现的频率可以得出,当原发性威胁作用于结点 6 时,的确应该先保护结点 6 的所有相邻结点{3,4,5,7},但是在实施保护策略时,根据图 8 所示的结点攻击频率,能够给出一个合理的资源保护分配方案.在结点 6 的所有相邻结点{3,4,5,7}中,结点 7 受到威胁攻击的频率最高(几乎达到结点{3,4,5}的 2 倍),因此应该实施最强的保护力度,且应该以相对于结点{3,4,5}的 2 倍保护力度进行保护.在结点{3,4,5}的保护策略中,结点 4 的保护力度应该略大于结点{3,5},对于结点{3,5}应施以相同的保护力度.

在图 8 中,采样策略和组合策略在不同结点攻击频率的对比中,组合策略方法并没有提供对结点{0,1}的保护建议,而采样策略却提供了对结点{0,1}的保护建议.因此,采样策略在复杂信息系统结点安全保护建议方面,比起组合策略方法能够提供一种更全面的资源保护策略建议.由此可见,本文提出的基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法(采样策略)在评估系统风险的过程,无论是在运行效率上还是在安全建议方面都比组合策略方法更优.

5　结　　论

为了评估复杂信息系统的安全风险,本文提出了一种在威胁的传播过程中基于结点转移状态及威胁传播边进行采样的威胁传播树生成算法,并通过实验阐明了本文所提算法能够对一个复杂信息系统在一定时间段内的系统风险进行客观正确的评估,且相对于组合策略方法在运行的时间效率和风险评估结果方面更具有优越性.由于影响复杂信息系统威胁传播行为的因素繁多而复杂,且复杂系统本身也具有多样性、复杂性等特点,现有的研究成果还远没有达到完全揭示其传播规律,不能精确地估算其风险损失、控制其传播的程度.因此,还有许多方面需进一步的深入研究与探索.

参　考　文　献

[1]　Jamin S, Raz D, Shavitt Y, et al. Guest editorial Internet and WWW measurement, mapping, and modeling [J]. IEEE Journal on Selected Areas in Communications, 2003, 21(6): 877-878

[2] Jeong H, Tonbor B, Albert R, et al. The large-scale organization of metabolic networks [J]. Nature, 2000, 407 (6804): 651-654

[3] Wang Zhanshan, Wang Junyi, Liang Hongjing. Research and progress of complex networks [J]. Communications of China Association of Automation, 2013, 34(170): 4-16 (in Chinese)
(王占山, 王军义, 梁洪晶. 复杂网络的相关研究及其进展[J]. 自动化学会通讯, 2013, 34(170): 4-16)

[4] Watts D J, Strogatz S H. Collective dynamics of "small-world" networks [J]. Nature, 1998, 393(6684): 440-442

[5] Ji Yong, Zhang Weihua, Zhang Zhengqi, et al. Risk assessment of complex data communication network [J]. Journal of Northeast Normal University: Natural Science Edition, 2013, 45(3): 57-61 (in Chinese)
(纪勇, 张伟华, 张证崎, 等. 复杂数据通信网络风险评估研究[J]. 东北师范大学报: 自然科学版, 2013, 45(3): 57-61)

[6] He Daren, Liu Zonghua, Wang Binghong. Complex Systems and Complex Networks [M]. Beijing: Higher Education Press, 2009: 187-234 (in Chinese)
(何大韧, 刘宗华, 汪秉宏. 复杂系统与复杂网络[M]. 北京: 高等教育出版社, 2009: 187-234)

[7] Jiang Wei, Fang Binxing, Tian Zhizhong, et al. Evaluating network security and optimal active defense based on attack-defense game model [J]. Chinese Journal of Computers, 2009, 32(4): 817-827 (in Chinese)
(姜伟, 方滨兴, 田志宏, 等. 基于攻防博弈模型的网络安全测评和最优主动防御[J]. 计算机学报, 2009, 32(4): 817-827)

[8] Nicol D M, Liljenstam M. Models and analysis of active worm defense [C] //Proc of the 3rd Int Workshop on Mathematical Methods, Models, and Architectures for Computer Network Security. Berlin: Springer, 2005: 38-53

[9] Wu Xiaoping, Fu Yu. Textbook for Information Security Risk Assessment [M]. Wuhan: Wuhan University Press, 2011: 1-9 (in Chinese)
(吴晓平, 付钰. 信息安全风险评估教程[M]. 武汉: 武汉大学出版社, 2011: 1-9)

[10] Zhang Li, Peng Jianfen, Du Yuge, et al. Information security risk assessment survey [J]. Journal of Tsinghua University: Science and Technology, 2012, 52(10): 1364-1368 (in Chinese)
(张利, 彭建芬, 杜宇鸽, 等. 信息安全风险评估的综合评估方法综述[J]. 清华大学学报: 自然科学版, 2012, 52(10): 1364-1368)

[11] Zhang Yongzheng, Fang Bingxing, Chi Yue, et al. Risk propagation model for assessing network information systems [J]. Journal of Software, 2007, 18(1): 137-145 (in Chinese)
(张永铮, 方滨兴, 迟悦, 等. 用于评估网络信息系统的风险传播模型[J]. 软件学报, 2007, 18(1):137-145)

[12] Li Xiaorong, Zhuang Yi, Xu Bin. Risk assessment model for information security based on danger theory [J]. Journal of Tsinghua University: Science and Technology, 2011, 51 (10): 1231-1235 (in Chinese)
(李晓蓉, 庄毅, 许斌. 基于危险理论的信息安全风险评估模型[J]. 清华大学学报: 自然科学版, 2011, 51(10): 1231-1235)

[13] Ma Gang, Du Yuge, Rong Jiang, et al. Risk assessment of the complex information system based on threat propagation [J]. Journal of Tsinghua University: Science and Technology, 2014, 54(1): 35-43 (in Chinese)
(马刚, 杜宇鸽, 荣江, 等. 基于威胁传播的复杂信息系统安全风险评估[J]. 清华大学学报: 自然科学版, 2014, 54(1): 35-43)

[14] Jin Hongzhang, Wei Qi, Guo Jian, et al. Vulnerability Theory and Application of Complex System [M]. Xi'an: Northwestern Polytechnical University Press, 2010: 1-28 (in Chinese)
(金鸿章, 韦琦, 郭建, 等. 复杂系统的脆性理论及应用[M]. 西安: 西北工业大学出版社, 2010: 1-28)

[15] Mu Chengpo, Huang Houkuan, Tian Shengfeng. Hierarchical online risk assessment for intrusion scenarios [J]. Journal of Computer Research and Development, 2010, 47(10): 1724-1732 (in Chinese)
(穆成坡, 黄厚宽, 田盛丰. 入侵进程的层次化在线风险评估[J]. 计算机研究与发展, 2010, 47(10): 1724-1732)

[16] Shi Yunfeng, Zhang Jinxiang, Feng Jianhua. Critical vulnerability analysis and exploitation based on exception capture [J]. Journal of Software, 2010, 21(11): 2944-2958 (in Chinese)
(时云峰, 张金祥, 冯建华. 基于异常捕获的强脆弱性分析与利用[J]. 软件学报, 2010, 21(11): 2944-2958)

[17] Zhao Gang, Kuang Xiaohui, Li Jin. A structural vulnerability analysis algorithm for large-scale distributed system [J]. Journal of Computer Research and Development, 2011, 48(5): 906-912 (in Chinese)
(赵刚, 况晓辉, 李津. 一种基于权值的大规模分布式系统结构脆弱性分析算法[J]. 计算机研究与发展, 2011, 48(5): 906-912)

[18] Zhou Liang, Li June, Lu Tianbo, et al. Research on quantitative assessment model on vulnerability risk for information system [J]. Journal on Communications, 2009, 30(2): 71-76 (in Chinese)
(周亮, 李俊峨, 陆天波, 等. 信息系统漏洞风险定量评估模型研究[J]. 通信学报, 2009, 30(2): 71-76)

[19] Lindquist J, Ma J, van den Driessche P, et al. Effective degree network disease models [J]. Journal of Mathematical Biology, 2011, 62(2): 143-164

[20] Zhao Gang, Liu Huan. Practical risk assessment based on multiple fuzzy comprehensive evaluations and entropy weighting [J]. Journal of Tsinghua University: Science and Technology, 2012, 52(10): 1382-1387 (in Chinese)
(赵刚, 刘焕. 基于多层次模糊综合评判及熵权理论的实用风险评估[J]. 清华大学学报: 自然科学版, 2012, 52(10): 1382-1387)

Ma Gang, born in 1986. PhD candidate. Student member of China Computer Federation. His main research interests include artificial intelligence, machine learning, neural computing and complex networks.

Zhang Bo, born in 1981. PhD candidate, lecturer. Student member of China Computer Federation. His main research interests include artificial intelligence, machine learning, data mining and cloud computing.

Du Yuge, born in 1980. Assistant professor, master. Her main research interests include risk assessment, classified protection, safety evaluation.

Wang Wei, born in 1968. PhD, president of Lexotech. His main research interests include intelligent search engine, artificial intelligence, machine learning.

An Bo, born in 1980. Associate professor, PhD supervisor. His main research interests include artificial intelligence, multi-agent systems, game theory, automated negotiation, resource allocation, and optimization.

Shi Zhongzhi, born in 1941. Professor, PhD supervisor. His main research interests include artificial intelligence, machine learning, neural computing and cognitive science.

2015 年起《计算机研究与发展》双月将固定领域专题

致广大读者和作者:

本刊从 2015 年起将双数期约 1/2 版面固定为某个领域,每年将策划该领域的一个热点主题进行集中报道.具体的征文通知将在专题发表前 6 个月发布,请关注期刊网站!

此外,本刊依然欢迎自由来稿.谢谢!

具体领域分布及执行领域编委如下:

刊期	领域	领域编委	投稿方式	截稿日期
2 期	软件技术(含数据库)	孟小峰 xfmeng@ruc.edu.cn 中国人民大学	期刊网站投稿, 备注填写“年+专题名称”	上一年 10 月 1 日左右 (以具体征文通知为准)
4 期	网络技术	林闯 chlin@tsinghua.edu.cn 清华大学	期刊网站投稿, 备注填写“年+专题名称”	上一年 12 月 1 日左右 (以具体征文通知为准)
6 期	体系结构	刘志勇 zyliu@ict.ac.cn 中国科学院计算技术研究所	期刊网站投稿, 备注填写“年+专题名称”	当年 2 月 1 日左右 (以具体征文通知为准)
8 期	人工智能	周志华 zhouzh@nju.edu.cn 南京大学	期刊网站投稿, 备注填写“年+专题名称”	当年 4 月 1 日左右 (以具体征文通知为准)
10 期	信息安全	曹珍富 zfcao@sjtu.edu.cn 上海交通大学	期刊网站投稿, 备注填写“年+专题名称”	当年 6 月 1 日左右 (以具体征文通知为准)
12 期	应用技术	郑庆华 qhzheng@mail.xjtu.edu.cn 西安交通大学	期刊网站投稿, 备注填写“年+专题名称”	当年 8 月 1 日左右 (以具体征文通知为准)