第 42 卷 第 11 期 计 算 机 工 程 2016 年 11 月
Vol. 42 No. 11 Computer Engineering November 2016

·安全技术· 文章编号:1000-3428(2016)11-0158-07 文献标志码:A 中图分类号:TP309

基于贝叶斯推理的 PASG 计算模型

王 辉[1],亢凯航[1],刘淑芬[1,2]

(1. 河南理工大学 计算机科学与技术学院,河南 焦作 454000;
2. 吉林大学 计算机科学与技术学院,长春 130012)

摘 要: 互联网中存在的内部威胁具有隐蔽性高、难以管理等特点。为此,建立一种针对内部威胁的贝叶斯网络攻击图(BNAG)模型。将攻击者在攻击过程中的行为作为研究对象,以行为在其动作期间的资源指向为基础,通过<行为,资源>二元组对模型内部进行分析,同时基于 BANG 模型,量化分析其中的资源与行为、行为与资源间的关联度,进一步给出贝叶斯网络节点概率计算模型(PASG)。将似然加权法作为评估抽样方法,对内部威胁进行预测分析。实例分析结果表明,PASG 模型对内部威胁能够起到有效的预测及防范作用。

关键词: 内部威胁;贝叶斯网络;网络攻击图;似然加权;信息融合

中文引用格式: 王 辉,亢凯航,刘淑芬. 基于贝叶斯推理的 PASG 计算模型[J]. 计算机工程,2016,42(11):158-164.

英文引用格式: Wang Hui, Kang Kaihang, Liu Shufen. PASG Computational Model Based on Bayesian Inference[J]. Computer Engineering, 2016, 42(11):158-164.

PASG Computational Model Based on Bayesian Inference

WANG Hui[1], KANG Kaihang[1], LIU Shufen[1,2]

(1. College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, China;
2. College of Computer Science and Technology, Jilin University, Changchun 130012, China)

【Abstract】 For the characteristics of insider threats in Internet, such as high concealment and difficult management, this paper proposes a Bayesian Network Attack Graph(BNAG) model which aims at insider threats. It takes the behavior of attacker in the attack process as the research object, takes the resource point of the behavior pointing as the basis, and analyses internal threats by the two-tuples <Action, Resource>. Relying on the BNAG model, it quantitatively analyses the relationship between resources and behavior, behavior and resources in the model, and proposes Probability of Attack Structure Graph(PASG) computational model based on Bayesian reseasoning. It forecasts and analyses the internal threats by the modified likelihood weighting method. The example analysis result shows that the proposed model can effectively forcast and prevent the insider threats.

【Key words】 insider threat; Bayesian network; network attack graph; likehood weighting; information fusion

DOI: 10.3969/j.issn.1000-3428.2016.11.026

0 概述

从首台大型机的诞生发展至今,人们的生活与办公越来越依赖于计算机,而有关于信息安全的问题也日益受到人们的重视。关于信息安全的问题,从最开始的网络安全,至今已逐渐发展到内部威胁问题。正因为内部威胁具有的高隐蔽性、难以管理、威胁性大等特点[1],在安全领域,如何防御来自于内部的威胁已成为一个十分严峻的问题。

《中国互联网网络安全态势报告》[2]指出,2011 年,我国相关部门接收到网络钓鱼类事件举报 9 463 起,被暗中植入后门的网站 52 324 个,其中包括政府网站 3 016 个。相比 2010 年,增长百分率分别达到 73.3%,213.7% 和 93.1%,这些攻击造成了大量用户信息泄露,直接造成的经济损失超过 194 亿元,同时相当一部分直接针对我国某些重要政

基金项目: 国家自然科学基金(61300216);高等学校博士学科点专项科研基金(20124116120004);河南省教育厅科学技术研究重点基金(13A510325)。

作者简介: 王 辉(1975—),男,副教授、博士,主研方向为网络安全、无线传感器网络;亢凯航,硕士研究生;刘淑芬,教授。

收稿日期: 2015-11-13 **修回日期:** 2015-12-13 **E-mail:** wanghui_jsj@hpu.edu.cn

府部门网站攻击,也严重威胁着我国的国家安全;中国互联网数据中心(IDC)的安全统计数据中也同样指出,在已知的安全威胁中,超过 70% 的攻击都是来自于企业内部终端的安全威胁;同时来自美国 CSI/FBI 的《计算机犯罪与安全调查报告》显示,在计算机安全事故所造成的损失中,有一半以上来自于内网安全漏洞;中国国家计算机网络应急技术处理协调中心(CNCERT/CC)在报告[3]中也明确说明,终端隐患已成为最大的安全威胁之一。由此可见,内网威胁已经成为企业乃至政府网络最为薄弱的环节,如何部署内网安全产品也成为现今企业安全的重中之重。

本文结合现有的网络内部威胁模型,以攻击者的行为为基础,设计基于 < 行为,资源 > 二元组模型的贝叶斯网络攻击图(Bayesian Network Attack Graph,BNAG),并进一步将其优化为基于贝叶斯推理的节点概率计算模型 PASG,最后通过改进的似然加权算法验证 PASG 模型的可行性。

1　相关研究

基于内部威胁的特点进行威胁研究,首先需要一种安全分析工具对内部威胁的各种情况进行分析,并加以解释。从现有研究中不难得出,攻击图能够通过对系统网络配置和脆弱性信息的分析,得到网络威胁的传播路径[4],因此,攻击图已成为分析内部威胁最为有效的工具。

文献[5]根据用户登陆的目的,对内部威胁加以分析,提出了一种检测攻击的方法,即 SPRINT 计划。而文献[6]指出了该方法的缺陷,即攻击者仍然可以通过对系统弱点进行符合 SPRINT 计划所规定的操作发起有效攻击。文献[7]在信息风险评估方面基于主机的安全状态变化概率对风险进行贝叶斯网络建模,但是该方法并没有反映出风险的本质,因此,模型的可扩展性较差。文献[8]提出一种基于 TVA 构建的贝叶斯网络模型对风险进行评估,虽然在一定程度上加快了建模的进度,但并没有指出如何计算模型中各个节点条件概率的分布情况。

在具体的网络威胁传播路径场景中,节点间的相关性,以及目标节点的复杂性和重要性的不同,都会导致节点概率的错误计算,影响传播路径的预测,而文献[9]则未考虑这点。文献[10]指出:导致节点概率计算出现错误的原因之一是节点之间存在相关性,并基于贝叶斯网络的 d-分隔定理,给出解决这一问题的方法,但未考虑条件概率对节点概率的影响,且增加节点的方法使得对攻击图的理解更加复杂。文献[11-12]利用改进的贝叶斯推理——似然加权法来计算攻击图节点置信度。通过引入攻击证据间的时间偏序关系,简化了节点置信度的计算,在一定程度上提升了准确性,但只给出了条件概率的假设值,并未给出计算方法。文献[12]对于大规模网络的分析与计算复杂度较高。文献[13]提出使用隐马尔科夫模型(Hidden Markov Model,HMM)的概率方法消除节点间关联性的影响,讨论了成本和收益因素对不同类型节点概率的关联性影响,并通过消除一些关联性较低的节点,提高了节点计算的准确性。但是该方法并未考虑到攻击成本对攻击者的选择策略的影响。文献[14]提出了基于代价的攻击者选择性攻击策略来对网络攻击进行分析,其消除冗余节点是通过攻击者的主观选择,未能分析攻击成本和节点概率的关联性,因而无法通过攻击成本对攻击路径进行预测。

目前,由于贝叶斯网络具有严谨的建模方面语言,在计算更为准确,建模方法与推理算法也已较成熟,因此贝叶斯网络在解决不确定性风险评估和预测问题方面是一种比较完善的理论模型[15]。本文在 SPRINT 计划的基础上给出基于内部威胁的贝叶斯网络攻击图 BNAG,并通过资源节点和行为节点间的关系,将其进一步优化为贝叶斯网络计算模型 PASG,对其进行定性构建和定量计算。其中的关键在于利用攻击者对资源进行操控的过程,量化分析资源与行为、行为与资源间的关系,形成贝叶斯网络攻击图,针对内部威胁的三大特点,使用贝叶斯理论对模型内部节点的威胁度进行定性分析和定量计算,进而对内部威胁进行预测,达到防御的效果。由于模型内的参数并不依靠操作的合法性,而是依赖于行为与资源之间的关联度,因此能较好地弥补 SPRINT 的缺陷。本文从系统全局出发,由大至小,再至大,对系统进行威胁预测,较之以往更加符合实际规律。

2　PASG 计算模型

2.1　PASG 模型定义

定义 1　资源节点 o_i 是攻击者已占有或者想要占有的系统对象;行为节点 a_i 是攻击者对系统或资源进行操作的指令;行为节点的操作对象为资源节点;资源节点指向的是下一个行为节点;资源节点与资源节点、行为节点与行为节点之间存在 AND-OR 的逻辑关系。

定义 2　根据用户登录系统的使用目的对内部威胁进行预测分析的攻击模型,即 SPRINT 计划,得出一个内部威胁网络攻击图 BNAG,其网络结构由行为

节点、资源节点以及节点间存在的因果关系确定,其中,o_0 为初始节点,是攻击者最初持有的资源。o_i,a_i 为其叶节点。图 1 表现了一个攻击者在相应的环境中,通过对系统资源的截获,使自己拥有更多的系统资源,以此对目的资源节点进行攻击的网络攻击图。

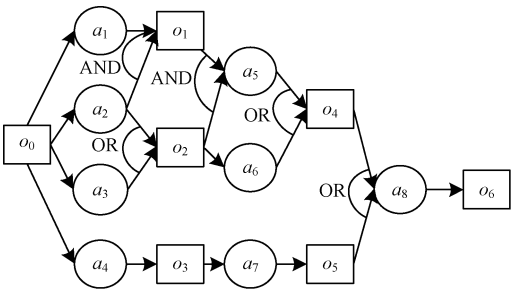


图 1 基于 SPRINT 计划的内部威胁网络攻击图

定义 3 攻击节点 $Pas = < o_{i-1}, a_i, o_i >$。

如图 2 所示,在 BNAG 中,a_i 的前提条件是 o_{i-1},即 $o_{i-1} = pre(a_i)$,o_{i-1} 是在攻击节点 Pas 中 a_i 的父节点的集合;o_i 是 a_i 的后置条件,记为 $o_i = next(a_i)$,是在攻击节点 Pas 中 a_i 的子节点集合。满足以上的前提条件及后置条件的 Pas 节点即为 BNAG 中的 Pas 节点,即:

$$\forall o_{i-1}, a_i, o_i \in Pas$$

$$o_{i-1} = pre(a_i), o_i = next(a_i)$$

$$pas_{i+1} = next(pas_i), pas_{i-1} = pre(pas_i)$$

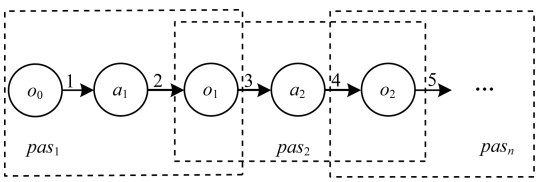


图 2 PAS 攻击节点模型

定义 4 基于内部威胁的贝叶斯网络攻击模型 BNAG = < PASG, P >。

PASG 为基于贝叶斯推理的内部威胁网络攻击计算模型。其中,PASG = < A, O, T, Pas >,A 表示的是攻击者的行为节点;O 为该行为的资源子节点,即 $O = next(A)$;T 为攻击者执行攻击的时间域;Pas 则是 $PASG$ 中各节点的条件概率分布。$PASG$ 反映的是在攻击过程中,各攻击节点之间存在着的因果关系,它决定了 BNAG 的网络拓扑结构;$P = \{P(Att \mid Pas)\}$ 是攻击发生的条件概率的集合,即威胁度,其中发生攻击记为事件 $Att = 1$,Pas 为攻击节点,且 Pas 决定了 BNAG 的网络参数。

因此,图 1 的网络结构可转换为图 3 的结构,即基于贝叶斯推理的 PASG 计算模型。

从图 3 中可以看出,对于一个特定的 PASG,其网络结构由攻击节点模型(PAS)确定,说明了攻击者获取目标节点的整个过程,与其对应的概率分布情况。参数 $P(Pas = Pas_i)$ 表明了攻击节点被攻击者执行的可能性大小。P(PAS)是本次操作执行的概率均值。PAS 决定了 PASG 的网络参数。

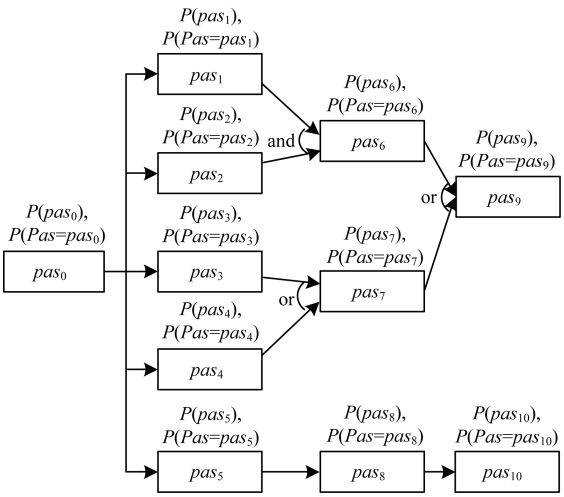


图 3 PASG 计算模型

定义 5 $Pas = pas_0$ 为初始攻击节点,在内部威胁的贝叶斯网中,初始攻击节点能以一定的概率转变成为 PASG 中任意的攻击节点。

定义 6 PAS 节点模型 PAST = $\{AO, W_{PO}, OP_O, W_O\}$。攻击对资源节点的安全威胁与对资源节点的攻击频度 f、攻击者的行为与资源节点指向集的关联度 AO、行为节点指向集的危险程度 W_{PO}、上一个行为节点占有的资源 O 的危险程度 W_O 和上一个行为节点占有的资源 O 及其行为节点指向集的关联度 O_{PO} 有关。而且,不同时间内相同攻击行为对资源的攻击次数也不同[16]。给定时间分析窗口 Δta,定义 t 时刻对 o_i 进行操作时对目标的攻击进度 pas_{oi} 为:

$$\begin{aligned} pas_{o_i}(t) &= f(f, AO[1,k], W_{P_O}[k,1](t), \\ &\quad OP_O[1,k], W_O[k,1](t)) \\ &= f \cdot (AO[1,k] \cdot W_{P_O}[k,1](t) \\ &\quad + OP_O[1,k] \cdot W_O[k,1](t)) \end{aligned}$$

图 4 展示了一个简单的 PAST 节点模型的内部结构及其内部元素之间的关联关系。对于节点 pas_j,假设其内部模型 PAST = $\{AO_j, W_{POj}, OP_{Oj}, W_{Oj}\}$,其中,$AO_j$ 表示了攻击者行为 A_k 与其行为指向集 $[O_1, O_2, \cdots, O_n]$ 之间的量化关系,W_{POj} 则为 t 时刻 A_k 占有 $[O_1, O_2, \cdots, O_n]$ 后对目标资源的威胁的量化值,OP_{Oj} 为攻击者行为 A_k 的前置资源节点 $[O_m]$ 与其指向的资源节点 $[O_1, O_2, \cdots, O_n]$ 之间的量化关系,W_{Oj} 为 t 时刻攻击者行为 A_k 所占有的前置资源节点 $[O_m]$ 对目标资源的威胁的量化值。

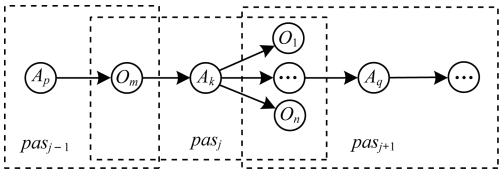


图 4　PAST 节点模型

对于图 4,可设 $AO_j=[1n,2n,\cdots,nn]$ 表示 A_k 与 $[O_1,O_2,\cdots,O_n]$ 之间的关联度向量为 $[1n,2n,\cdots,nn]$,$W_{POj}=[w_1,\ w_2,\cdots,w_n]$ 表示 t 时刻攻击节点 PAS_j 中的行为 $[A_k]$ 占有其指向集资源对目标资源的危险度向量为 $[w_1,\ w_2,\cdots,w_n]$,$OP_{Oj}=[1n,2n,\cdots,nn]$ 表示了 A_k 的前置资源节点与其指向的资源节点之间的关联度向量为 $[1n,2n,\cdots,nn]$,$W_{Oj}=[wo_1,wo_2,\cdots,wo_n]$ 表示在 t 时刻攻击节点 pas_j 的前置攻击节点 pas_{j-1} 的行为 $[A_p]$ 所占有的指向集资源对目标资源的危险度向量为 $[wo_1,\ wo_2,\cdots,wo_n]$。

由此可得,对图 4 中的简单模型,其攻击进度 pas_{oi} 为:

$$\begin{aligned}pas_{o_i}(t)&=f(f,AO[1,k],W_{P_O}[k,1](t),\\&\quad OP_O[1,k],W_O[k,1](t))\\&=f\cdot([1n,2n,\cdots,nn][w_1,w_2,\cdots,w_n]\\&\quad+[1n,2n,\cdots,mn][wo_1,wo_2,\cdots,wo_n])\end{aligned}\tag{1}$$

因此,PASG 模型有以下特点:

1)PASG 为有向无环图,反映的是攻击过程中,各攻击节点之间的因果关系,决定了 BNAG 的网络拓扑结构。在 BNAG 模型中,它以为 pas_0 初始节点,pas_i 为叶节点。任意 PAS 攻击节点即为满足定义 3 的 PAS 节点。

2)对 $P(Pas=pas_i|Att)$ 是攻击发生的条件概率的集合,其中攻击成功,即攻击者获取到该资源节点,记为事件 $Att=1$,pas_i 为攻击节点。它表明了该攻击节点与其父节点之间的关系。其中,其先验概率为初始攻击节点的条件概率。

3)$P(Pas=pas_0|Att)$ 为系统的初始状态,在基于内部威胁的贝叶斯网络中,所有的攻击节点都是由初始状态开始,并以一定概率转变的。

2.2　PASG 模型构建

PASG 模型的构建包括两部分:1)定性构建 BNAG 模型的结构,并转换为 PASG;2)定量构建 BNAG 模型的参数。具体步骤如下:

步骤 1　确定攻击模型 BNAG 与计算模型 PASG 的网络结构。

贝叶斯网络计算模型 PASG 作为攻击模型 BNAG 的结构基础,表现了对系统特定资源的攻击过程中各个攻击节点之间的偏序关系,其转化算法如下:

算法 1　模型转化

描述根据内部威胁攻击模型 BNAG,将其转化为贝叶斯网络计算模型 PASG

输入　内部威胁攻击模型 BNAG

输出　贝叶斯网络计算模型 PASG

1)寻找初始资源节点,并在 BNAG 中建立相应的节点 o_0。

2)对步骤 1)中所建立初始资源节点的 o_0,遍历其所有的行为子节点,并在 BNAG 中建立相应的节点 a_i,如果 o_0 含有多个直接后续的行为子节点,则在 BNAG 中也相应地建立这些多个子节点,确立这些多个行为子节点的 AND-OR 关系,并在 BNAG 中表示出来。

3)对步骤 2)中建立的行为节点 a_i,遍历其所有的资源子节点,并在 BNAG 中建立相应的节点 o_i,如果 a_i 含有多个直接后续的资源子节点,则在 BNAG 中也相应地建立这些多个子节点,确立这些多个资源子节点的 AND-OR 关系,并在 BNAG 中表示出来。

4)对步骤 3)中建立的行为节点 o_i,遍历其所有的行为子节点,并在 BNAG 中建立相应的节点 a_j,如果 o_i 含有多个直接后续的资源子节点,则在 BNAG 中也相应地建立这些多个子节点,确立这些多个资源子节点的 AND-OR 关系,并在 BNAG 中表示出来。

5)重复执行步骤 3)和步骤 4),直至所有的行为节点和资源节点遍历完全。此时,BNAG 中有且仅有一个资源终节点。

6)在 PASG 中建立初始 PAS 节点 pas_0。其中 pas_0 仅含有 BNAG 中的初始资源节点 o_0。

7)对 BNAG 中 o_0 的子节点 a_i,在 PASG 中建立相应的 PAS 节点 pas_i,其中包含了 a_i 的父节点 o_0 与子节点 o_i。如果存在多个 a_i 且这些个 a_i 之间存在 AND-OR 关系,则在与之对应的节点 pas_i 中也建立相应的 AND-OR 关系。

8)对步骤 7)中建立的 PAS 节点 pas_i,遍历其对应的 BNAG 模型中的行为节点 a_i,寻找其直接的后续行为节点 a_j,在 PASG 中建立节点 pas_j,其包含了 a_j 的父节点 o_i 与子节点 o_j。如果存在多个 a_j 且这些个 a_j 之间存在 AND-OR 关系,则在与之对应的 PAS 节点 pas_j 中也建立相应的 AND-OR 关系。

9)重复执行步骤 8),直至 BNAG 中所有的节点全部转化为 PASG 中的 PAS 节点。

步骤 2　确定贝叶斯网络计算模型 PASG 的网络参数。

由于实验无法获取足够的数据支持,因此很难确定出模型的网络参数,基于此,本文采用多源前验信息融合前验概率分布计算 PASG 模型的攻击节点的条件概率来确定模型的网络参数,而不是经典的贝叶斯试验确定网络节点的条件概率。

PASG 网络攻击节点之间的逻辑关系有以下 3 种:

1)一对多,即一个 pas_j 能导致多个 pas_k($k=1,2,\cdots,n$)的发生,其中多个 pas_k 之间的逻辑关系为 OR。

2)多对一,即多个 pas_j($j=1,2,\cdots,n$)能导致一个 pas_k 的发生,其中多个 pas_k 之间的逻辑关系可能是 OR 或者 AND。

3)一对一,即一个 pas_j 能导致一个 pas_k 的发生。

图 5 给出了 PASG 模型中各个攻击节点之间可能存在的 3 种逻辑关系。

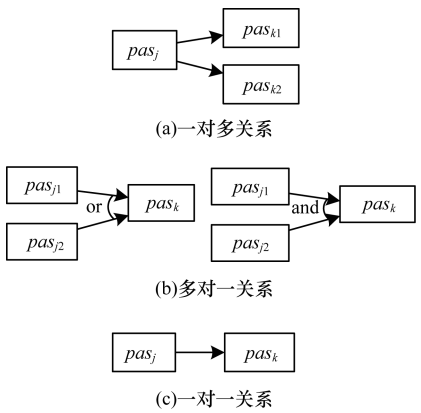


图 5 PAS 攻击节点间的关系

PASG 中各攻击节点的攻击状态具有二态性,即节点事件状态只可能是成功(1)或失败(0),因此事件“成功执行攻击节点”发生的条件概率 θ 的前验概率分布服从 Beta 分布。本文使用多源前验信息融合方法计算 PASG 模型中节点的条件概率 θ 的前验概率分布。在获取专家意见后,设定参数 θ 的验前均值为 μ,通过计算 Beta(θ,a,b)的极大熵,即可得出 Beta(θ,a,b)的分布参数。

首先,设定参数的前验均值概率,则:

$$\mu=\frac{a}{a+b} \tag{2}$$

然后,计算 Beta(θ,a,b)的极大熵,即:

$$\begin{aligned}\max H(\delta) &= -\int\delta(\theta)\mathrm{lb}\left(\frac{\delta(\theta)}{\delta_0(\theta)}\right)\mathrm{d}\theta\\ &= -\int\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\\ &\times\mathrm{lb}\left(\int\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\right)\mathrm{d}\theta\end{aligned} \tag{3}$$

其中,θ 的前验概率 $\delta_0(\theta)=1$。结合式(2)和式(3),计算得到使 $\max H(\delta)$ 取得最大值的最大熵前验概率分布参数 a 和 b。

步骤 3 更新模型内部参数。

基于贝叶斯方法的网络推理的基本类型有3 种:1)基于网络模型的计算后验概率;2)最大可能解释;3)最大后验假设问题。本文则是以通过对网络攻击图中起始资源节点集合的前验概率分布进行分析,进而计算得出目标资源节点集合的后验概率分布。具体过程如下:

1)执行步骤 1)和步骤 2),得到计算模型 PASG。获取 PASG 中第 i 个攻击节点 pas_i,而在 pas_i 父节点的第 j 个组合状态下,可得该节点的前验概率分布为:

$$\delta(\theta_{ij})=\mathrm{Beta}(\theta_{ij}\mid a_{ij},b_{ij}) \tag{4}$$

2)分析样本数据可得,在第 i 个攻击节点的父节点的第 j 个组合状态发生 n 次的条件下,攻击节点 pas_i 共发生了 m 次,即 pas_i 符合 n 重贝努利实验,则其似然函数为:

$$P(D\mid\theta_{ij})=C_n^m\theta_{ij}^m(1-\theta_{ij})^{n-m} \tag{5}$$

在实验过程中,系统通过对攻击者对攻击节点的执行状态进行收集,得到模型中相应节点与其父节点的部分状态组合。在获取大量样本之后,后验概率分布越来越依赖于样本数据,先验知识的影响逐渐减小,此时通过融合先验概率和样本信息来得到后验信息。

3)假定获取了完整的数据,此时的前验概率分布为 $\delta(\theta)$,样本信息为 $p(D\mid\theta)\delta(\theta)$,在给定样本 $D=(D_1,D_2,\cdots,D_n)$ 的条件下 θ 的后验概率分布为:

$$\delta(\theta\mid D)=\frac{p(D\mid\theta)\delta(\theta)}{\int_\theta p(D\mid\theta)\delta(\theta)\mathrm{d}\theta} \tag{6}$$

4)将式(4)和式(5)同时代入式(6)中,得到参数的后验概率分布:

$$\delta(\theta_{ij}\mid D)=\mathrm{Beta}(\theta_{ij}\mid a_{ij}+m,b_{ij}+n-m) \tag{7}$$

由此可以得出,后验概率分布的均值 θ_{ij} 即为在通过一定的试验数据训练后,在 pas_i 的父节点的第 j 个组合状态下的节点 pas_i 的条件概率。

2.3 基于 PASG 的攻击概率计算

在 PASG 模型中,利用似然加权法可以避免逻辑抽样因舍弃成本而造成的浪费,它按攻击网络的拓扑序对其每个变量 X 进行抽样分析:1)若 X 不为证据变量,此时抽样方式使用逻辑抽样;2)若 X 为证据变量,则此时抽样结果则为 X 的观测值。这样可以保证任意一个样本都和证据 $E=e$ 相同,不需要舍弃从而造成浪费,是一种较逻辑抽样法更为有效的贝叶斯网计算方法。

在 PASG 模型中,由于需要考虑到各个攻击节点之间的 AND-OR 关系,因此在使用似然加权法时需要对节点间的拓扑序进行处理。

算法 2　线序关系 *STEP*(PASG,pas_i)

描述　根据贝叶斯网络的网络攻击图 PASG,找出 PASG 中所有节点的线序关系集合

输入　网络攻击图 PASG,任意 2 个节点 *X* 和 *Y*,偏序关系集合 *POS*,AND 关系集合 *AC*

输出　贝叶斯网络中节点的线序关系集合 *SETP*

```
POS←Ø;POS_i←Ø;AC←Ø;
for(Pre(X)=0 的每一个 X);
  在 PASG 中找出节点 X 的所有子节点变量 Y;
  if(Pre(Y)>1 且 Y 指向其子节点的 2 条有向边之间存在 AND 关系;
    AC←AC∪{X1∧X2};
  endif;
  POS←POS∪{(X,Y)}
  Next(Y)←0;
endfor;
POS_i←forearch(所有不考虑 AND 关系的偏序集 in POS);
SETP_i←forearch(满足 AND 关系的线序集合 in POS_i);
return SETP
```

算法 3　似然加权法

描述　根据贝叶斯网络的网络攻击图 PASG 和已执行的攻击节点,求出查询变量的后验概率分布

输入　贝叶斯网 PASG,样本量 *m*,已执行的攻击节点 *E*,攻击节点 *E* 的取值 *e*,查询变量 *Q*,查询变量的取值 *q*

输出　$P(Q=q|E=e)$的近似值

```
S←STEP(PASG,D);
ω_e←0;ω_{q,e}←0;
for(i=1 to m)
  Di←0;
  for(S 中的每一个变量 X)
    if(X∈E)
      x←X 的观测值;
    else
      x←从 P(X|π(X))抽样的结果;
    endif
  end for
  D_i←D_i∪{X=x};
  ω_i← ∏X∈EP(X|π(X))|D_i;
  ω_e←ω_e+ω_i;
  if(D 与 D=q 一致)
    ω_{q,e}←ω_{q,e}+ω_i;
  endif
Endfor
return ω_{q,e}/ω_e
```

3　实例分析

本文采用 Windows 系统下的 Matlab 作为主要编程语言,实现贝叶斯推理似然加权算法以验证算法的功能。实验环境为 Matlab 组建的虚拟贝叶斯攻击网。通过计算机模拟生成一个 BNAG 攻击图,从而对组建的虚拟网络进行检测和防御。

似然加权法的核心思想是根据攻击图的拓扑序对每一个节点变量进行若干次抽样,最后通过抽样得到的结果,对后验概率进行统计分析。按照采样时间以及采样的交互次数进行抽样统计分析,每组抽样数为 1 000 个。

由于攻击者的行为是不确定的,而其所需要获取的目标资源则为确定的,因此本文利用随机数产生器生成符合条件的随机数对抽样过程中所涉及到的变量进行随机抽样。

利用式(1)获取攻击者初始的攻击进度 *Pas*,将其代入式(2)和式(3)得到攻击者获取目标资源的初始威胁度。结合步骤 3 计算出相应的后验概率 *p*。假定随机变量 *X*,其服从贝努利分布,且取值为 0 或 1。假设 $P(X=0)=p$,$P(X=1)=1-p$,对 $P(X)$抽样的过程如下:随机生成服从伯努利分布的一个数 $x\in[0,1]$。如果 $x\in[0,p]$,则抽样的样本结果为 $X=1$,否则 $X=0$。其中,每次抽样后,将得出的后验概率 *p* 作为后置步骤的先验概率对后验概率重新抽样计算。采样计算结果如表 1 所示,其中 $P_1\sim P_4$ 表示后验概率。

表 1　实验采样计算结果

采样时间	P_1	P_2	P_3	P_4
1	0.228 0	0.150 0	0.128 0	0.123 0
2	0.232 5	0.140 0	0.109 4	0.105 7
3	0.256 1	0.172 7	0.092 6	0.113 2
4	0.227 9	0.153 9	0.109 4	0.126 3
5	0.243 8	0.137 1	0.113 6	0.065 0
6	0.287 9	0.187 1	0.172 9	0.073 8
7	0.365 5	0.342 7	0.269 7	0.104 6
8	0.609 6	0.690 6	0.453 3	0.278 2
9	0.882 5	0.823 6	0.803 5	0.460 0
10	1.000 0	0.942 5	1.000 0	0.915 9

由采样计算结果可以看到,用户的攻击概率随着采样时间的增加也在逐渐增大,符合实际攻击效果。图 6 显示,在攻击进度进一步推进下,攻击者获取和拥有的资源逐渐增加,离目标资源越来越近,因此,对系统的威胁也越来越大,从而攻击成功的可能性也在逐渐增大。

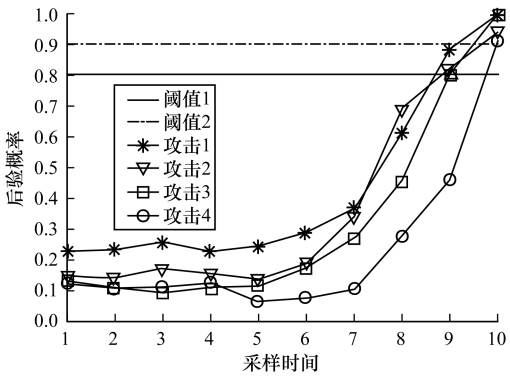


图 6 内部威胁模型中用户攻击监测图

为便于对系统进行实时检测和防御,内部威胁攻击模型需要一个量化值对攻击者的行为进行判断,为此本文设置了警戒阈值和危险阈值。假设警戒阈值为 0.8,危险阈值为 0.9,当检测到攻击概率大于等于 0.8 时,应对该用户实施重点监控,并向该用户发出警告;当检测到攻击概率大于等 0.9 时,终止该用户的行为操作,以达到对系统防御的目的。

目前,传统的内部威胁预测模型主要依赖于攻击图模型以及模型内部系统风险因素的概率分布,虽然这种二元组预测模型能够快速地对模型进行构建,能准确地找出系统风险的关键点,但在实际的攻击过程中,风险因素间的相关性,以及目标节点的复杂性及重要性的不同,都可能导致节点概率的错误计算,影响预测的准确性。本文则是利用系统风险因素之间的关系,对系统的可靠性进行分析评估,提高了模型对内部威胁预测的准确性。

4 结束语

本文从贝叶斯网络模型角度出发,对内部威胁的网络攻击更深层次的结构进行描述,构建一个可以从定性及定量 2 个层面上体现内部威胁内部各个风险因素之间依赖关系的攻击概率计算模型,并使用似然加权作为主要的抽样方法分析内部威胁的网络结构,最后计算得出内部威胁网络攻击中的后验概率。

本文模型具有以下优势:1)以攻击者的行为、攻击资源作为攻击图的基础,保证了准确性和有效性;2)能实时反映系统当前的安全状况,且更客观实际;3)通过对模型参数量化和对传统模型优化,使攻击图更直观地反映出攻击者的意图行为,简化了计算过程。

尽管本文模型在一定程度上解决了内部威胁的检测和感知方面的问题,但依然存在不足。下一步研究将逐步优化该模型,降低算法的时间复杂度,提高执行效率,加快检测速度。

参考文献

[1] Tsay J K, Mjølsnes S. Computational Security Analysis of the UMTS and LTE Authentication and Key Agreement Protocols[J]. Computer Science, 2013, 360(12): 12-17.

[2] 国家计算机网络应急技术处理协调中心. 2011 年中国互联网网络安全态势报告[EB/OL]. (2012-03-20). http://www.cert.org.cn/publish/main/46/index.html.

[3] 中国互联网协会.《中国互联网站发展状况及其安全报告(2015)》发布[EB/OL]. (2015-03-20). http://www.cac.gov.cn/2015-03/20/c_1114714197.htm.

[4] 陈 锋,张 怡,苏金树,等. 攻击图的两种形式化分析[J]. 软件学报,2010,21(4):838-848.

[5] Chinchani R, Upadhyaya S, Kwiat K. Towards the Scalable Implementation of a User Level Anomaly Detection System[C]//Proceedings of MILCOM'02. Washington D.C., USA: IEEE Press, 2002: 1503-1508.

[6] 王 辉,刘淑芬. 一种可扩展的内部威胁预测模型[J]. 计算机学报,2006,29(8):1346-1355.

[7] Yu Liu, Hong Man. Network Vulnerability Assessment Using Bayesian Networks[J]. Proceedings of SPIE, 2005, 5812: 61-71.

[8] Frigault M, Wang L, Singhal A, et al. Measuring Network Security Using Dynamic Bayesian Network[C]//Proceedings of ACM Workshop on Quality of Protection. New York, USA: ACM Press, 2008: 23-30.

[9] 石 进,郭山清. 一种基于攻击图的入侵响应方法[J]. 软件学报,2008,19(10):2746-2753.

[10] Homer J, Ou X, Schmidt D. A Sound and Practical Approach to Quantifying Security Risk in Enterprise Networks[EB/OL]. [2015-10-11]. http://people.cs.ksu.edu/~xou/publications/tr_homer_0809.pdf.

[11] 张少俊,李建华,宋珊珊,等. 贝叶斯推理在攻击图节点置信度计算中的应用[J]. 软件学报,2010,21(9):2376-2386.

[12] Publishing S. A Novel Attack Graph Posterior Inference Model Based on Bayesian Network[J]. Journal of Information Security, 2011, 2(1): 8-27.

[13] Wang Shuzhen, Zhang Zonghua, Kadobayashi Y. Exploring Attack Graph for Cost-benefit Security Hardening: A Probabilistic Approach[J]. Computers & Security, 2013, 32(1): 158-169.

[14] 吴泓润,覃 俊,郑波尽. 基于代价的复杂网络抗攻击性研究[J]. 计算机科学,2012,39(8):224-227.

[15] 杨 健,高文逸,刘 军. 一种基于贝叶斯网络的威胁估计方法[J]. 解放军理工大学学报(自然科学版),2010,11(1):43-48.

[16] 陈秀真,郑庆华,管晓宏,等. 层次化网络安全威胁态势量化评估方法[J]. 软件学报,2006,17(4):885-897.

编辑 金胡考