基于贝叶斯网络的内部威胁预测研究*

王 辉†, 杨光灿, 韩冬梅
(河南理工大学 计算机科学与技术学院, 河南 焦作 454000)

摘 要: 在内部网络带给企业办公便利的同时,内部网络所带来的威胁也日渐突出,由于企业中内部威胁具有危害性大、难以检测等特点,内部威胁亟需解决。因此,提出了基于贝叶斯网络攻击图的内部威胁预测模型。以内部用户实际操作过程中的行为为研究对象,以内部用户攻击过程中所占有的资源状态和所进行的操作序列攻击证据为节点,构建贝叶斯网络攻击图;以网络攻击图来描述攻击者在攻击过程中的不同攻击路径和攻击状态,并且利用贝叶斯网络推理算法计算内部威胁的危险概率。在贝叶斯网络攻击图中定义了元操作、原子攻击、攻击证据等概念,量化了节点变量、节点变量取值和条件概率分布。以改进的似然加权算法为基础,使贝叶斯网络的参数计算更加简便,内部威胁的预测更加精确。最后,通过仿真实验证明了该方法建模速度快、计算过程简单、计算结果精确,在预测内部威胁时的有效性和适用性。

关键词: 内部威胁; 贝叶斯网络; 网络攻击图; 似然加权法
中图分类号: TP391 **文献标志码**: A **文章编号**: 1001-3695(2013)09-2767-05
doi:10.3969/j.issn.1001-3695.2013.09.052

Research of predicting insider threat based on Bayesian network

WANG Hui†, YANG Guang-can, HAN Dong-mei
(*College of Computer Science & Technology, Henan Polytechnic University, Jiaozuo Henan* 454000, *China*)

Abstract: Internal network brings convenience for corporate office, but increasing threats are also brought into enterprises. Insider threat causes great harm to enterprises, and is difficult to detect, so it is urgently to be solved. This paper put forward a predictive model of insider threat based on Bayesian network attack graphs. It considered the behaviors in attacking process as research objects, and considered the resources and operation sequence as nodes, established Bayesian network attack graphs. It described the different attack paths and attack state in the process of attacking by Bayesian network attack graphs, and used Bayesian network inference algorithm to calculate the risk probability of insider threat. In Bayesian network attack graphs, the concepts of meta-operation, atomic attack and intrusion evidence were defined, and node variable, its value and conditional probability distribution were quantified. Based on the improved likelihood weighted algorithm, the calculation of Bayesian network parameters is easier, and the prediction of insider threat is more accurate. Ultimately, by simulation experiment, it is proved that the modeling speed is fast, the process of calculation is simple, the result is exact, and it is valid and applicative in predicting insider threat.

Key words: insider threat; Bayesian network; network attack graphs; likelihood weighted algorithm

随着科学技术的发展和社会网络使用的普遍,有关网络方面的安全事件越来越多,这些网络安全事件不仅发生在企事业单位的外部网络,也发生在单位的内部网络。事实表明,内部威胁的危害性远远大于外部威胁。虽然目前存在很多防御工具和检测工具,并且这些工具的防御技术水平越来越高,但是它们主要是用来解决外部威胁的,有关保护内网安全的技术有待发展和重视。

贝叶斯网络是一种图形模式,主要用来描述随机变量间的依赖关系,应用于不确定性问题的求解,在分析问题,预测和防御方面已经得到广泛的应用。结合内部威胁的不确定性、脆弱性、复杂性,本文提出基于贝叶斯网络攻击图的内部威胁预测模型,利用贝叶斯网络的图形模式来形象地描述内部网络中的各种状态,形成网络攻击图,并且利用贝叶斯网络推理算法计算内部威胁的危险概率。

1 相关研究

Liu 等人[1]首次在信息风险评估方面利用贝叶斯网络建模,构建模型时采用的是基于主机间的安全状态变化概率,并没有利用风险的本质脆弱性方面度量信息系统的安全水准,构建的模型可扩展性差。

Frigault 等人[2]利用 TVA 构建贝叶斯网络模型进行风险评估,这种建模方法仅仅是提高了建模的速度,并没有指出模型中条件概率分布的确定问题。

王桢珍等人[3]提出利用贝叶斯网络模型计算信息安全风险概率。他们将风险概率因素导入规划渗透图,利用规划渗透图作为贝叶斯网络模型的网络结构,利用专家知识和最大熵验前分布获取模型网络参数的计算值。该模型构建速度快,并且利用贝叶斯网络学习机制建模生成的参数更为准确。

收稿日期: 2012-12-09; **修回日期**: 2013-02-28 **基金项目**: 国家自然科学基金资助项目(51174263);河南省教育厅自然科学基金资助项目(2011B520015);河南理工大学博士基金资助项目(B2010-61);河南省社科联基金资助项目(SKL-2012-849)

作者简介: 王辉(1975-),男(通信作者),河南焦作人,副教授,博士,主要研究方向为网络安全(wanghui_jsj@hpu.edu.cn);杨光灿,男,河南平顶山人,硕士,主要研究方向为网络安全;韩东梅,女,河南周口人,硕士,主要研究方向为网络安全.

张少俊等人[4]提出使用网络攻击图描述网络安全中攻击者到达攻击目标所选择的所有路径,利用网络攻击图进行安全分析。但是有关网络攻击图的计算存在很多不足,有关攻击图节点的置信度计算复杂,推理缺乏数学理论支撑。根据以上研究,笔者基于概率统计理论,提出利用贝叶斯推理计算网络攻击图的节点置信度分布。实验证明,利用贝叶斯推理计算置信度能够显著提高计算的精确性,并且贝叶斯推理在计算复杂大规模网络攻击图的方面有特长。

内部威胁在网络安全领域中是一个严峻的问题,具有隐蔽性、难抵御、危害大、难管理的特点[5,6],要想研究分析内部威胁,首先需要一种安全分析工具来描述内部威胁的各种情况。而网络攻击图[7]是近些年安全分析的重要工具之一,而且有关网络攻击图的自动生成技术研究是国内外专家学者们的研究热点,目前已经取得不错的研究成果。贝叶斯网络是目前解决不确定性风险评估和预测的比较成熟的理论模型[8],具有完善的建模方法和推理算法,在建模方面语言严谨、计算准确。因此,本文提出基于贝叶斯网络的内部威胁预测模型,利用网络攻击图作为贝叶斯网络的模型结构,利用改进的似然加权算法计算内部威胁预测模型的参数;并且在建模的过程中提出了元操作、攻击证据和原子攻击的概念,这样的建模方法速度快、效率高,在各种环境里都适用。

2 基于贝叶斯网络的内部威胁预测模型

从贝叶斯网络的定义可以推知基于贝叶斯网络的模型构建需要从两个层面来理解。在定性方面,它是用一个有向无环图来刻画不同变量节点之间的依赖和独立关系,即模型的网络结构;在定量方面,它是用条件概率分布来描述子节点对其父节点的依赖关系,即模型的参数。

2.1 元操作的定义

网络攻击是由不同的命令或者操作集合组成的,其中,对不同种类的命令或者操作,按照某种方法分类形成的集合称为元操作[9]。举例说明如下:打开 Word 可以是通过双击“Word 2003”程序或者点中“Word 2003”,按“Enter”键的键盘操作,这两种方式都是由不同的操作组成但是功能相似,因此打开 Word 的操作集合是一个元操作。

2.2 原子攻击的定义

文献[10,11]定义了 SPRINT(signature powered revised instruction table)计划,指出根据用户登录系统的使用目的预测分析内部威胁,即在用户使用系统之前,按照〈主题,动作,对象,期间〉的列表形式提交用户使用意图,形成 SPRINT 计划。其中,内部网络中合法的元操作集合 Mos(material operate set)是用户提交的 SPRINT 计划中的〈动作,对象〉属性在元操作上的映射。

攻击过程中,两个资源状态之间的一系列操作(包括各种命令和操作)与元操作集合 Mos 中的一个子集 sub-Mos 相对应,这样的一系列操作称为原子攻击。原子攻击是攻击者从一个资源状态到达下一个资源状态所需要的最小元操作集合。原子攻击与元操作的关系如图 1 所示。

2.3 攻击证据的定义

为了表示攻击资源状态之间的一系列操作,引入攻击证据的概念,攻击证据就是攻击者从一个资源状态到达另一个资源状态日志所记录的一系列操作的集合。攻击证据是由多个元操作按照一定的顺序组成的集合。元操作的不同、顺序的不同,导致的攻击证据也会不同。

用攻击证据的置信度来度量攻击证据的大小。攻击证据的置信度是指攻击证据的元操作集合覆盖原子攻击中元操作集合的概率。在这里,覆盖的意思是按照原子攻击集合中元操作的先后顺序,系统监测到的攻击证据元操作集合中相同于原子攻击集合中的元操作的个数。假设用 m_i 来表示元操作集合,原子攻击集合为 $\{m_1,m_2,m_3,m_4\}$,顺序是 $m_1\rightarrow m_2\rightarrow m_3\rightarrow m_4$,则攻击证据 $\{m_1,m_5,m_6,m_2,m_7\}$ 的覆盖为 2,置信度为 50%;攻击证据 $\{m_5,m_6,m_7,m_8\}$ 的覆盖为 0,置信度为 0%;攻击证据 $\{m_2,m_1,m_5,m_6\}$ 的覆盖为 1,置信度为 25%。

为了便于说明,假设原子攻击中的元操作个数为 n,当攻击证据的覆盖个数小于 n 时称为取证不足;当攻击证据的覆盖个数等于 n 时称为取证完全。

以图 2 为例说明元操作、原子攻击和攻击证据之间的关系。v_1 和 v_2 是两个资源节点,o_1 和 o_2 是两个攻击证据,关系如图 2 所示。攻击者在占有资源节点 v_1 时可通过两种攻击证据 o_1 和 o_2 中的一种占有攻击资源 v_2。假设元操作集合 $M=\{m_i\mid i=1,2,3,\cdots,n\}$,若攻击者从占有资源节点 v_1 到占有资源节点 v_2 必须经过四个元操作,它们是 $m_2\rightarrow m_4\rightarrow m_5\rightarrow m_7$,箭头表示四个元操作集合之间的先后顺序,则称 $\{m_2,m_4,m_5,m_7\}$ 为原子攻击。当元操作集合个数少于这四个元操作或者出现这四个元操作的先后顺序不同时都不能称之为原子攻击。攻击证据则是不同元操作的组合,当攻击证据 $o_1=\{m_1,m_2,m_3,m_4,m_6\}$ 时,攻击证据的覆盖个数为 2,称之为攻击者取证不足,无法实现攻击效果;当攻击证据 $o_1=\{m_1,m_2,m_3,m_4,m_5,m_6,m_7\}$ 时,攻击证据⊇原子攻击,覆盖个数为 4,称之为攻击者取证完全,成功到达资源节点 v_2。

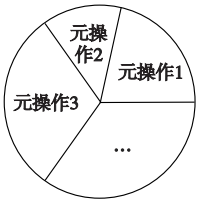

图1 原子攻击

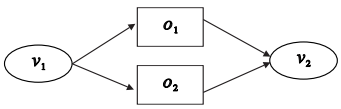

图2 例子

2.4 网络攻击图

为了描述攻击者实施攻击过程中的不同攻击路径和攻击特点,本文采用网络攻击图(network attack graph,NAG)表示攻击路径中不同攻击位置的相互关系。网络攻击图的定义为 $NAG=(V,V_0,G,O,E,P)$。其中:

a) V 是系统资源状态节点集合,$V=\{v_i\mid i=1,2,\cdots,N\}$,$v_i$ 是单个资源节点,用来描述攻击者在攻击过程中所占有的资源,取值为 true 或 false。当 v_i 取值为 true 时,表示攻击者成功占有该资源;当 v_i 取值为 false 时,表示攻击者未能占有该资源。

b) V_0 是指初始状态下,攻击者已占有的攻击资源集合,是 V 的一个子集。在网络攻击图中是起始节点集合。

c) G 是目标节点集合,即攻击者最终要达到的攻击目标资源集合。

d) O 是攻击证据节点集合,表示攻击者在已占有资源的情况下进行的一系列操作的集合。$O=\{o_i\mid i=1,2,\cdots,n\}$,$o_i$ 是单个攻击证据节点,取值为攻击证据的置信度,取值范围为 [0,1]。当攻击证据取值置信度小于 100% 时,表示此时攻击

者取证不足,未能成功占有下一个资源节点;当攻击证据置信度取值等于100%时,表示攻击者取证完全,成功占有下一个资源节点。

e)E 是关联各类节点的有向边集合。$E=(E_1\cup E_2)$,其中 $E_1\subseteq V\times O$,$E_2\subseteq O\times V$。E_1 表示攻击者只有在占有某些资源状态的情形下才能触发攻击证据的产生;E_2 表示攻击证据的实施情况可以使攻击者占有某些资源。

f)P 是网络攻击图中各节点的条件概率分布$(P_1\cup P_2)$,P_1 为攻击证据节点的条件概率分布,P_2 为资源状态节点的条件概率分布。一般地,定义 Pre(X)为节点 X 的父节点集合,Con(X)为节点 X 的孩子节点集合。为了网络攻击图的完备性,图中考虑了不同节点间的 AND 和 OR 关系,即不同的 Pre(o_i)节点间存在 AND 和 OR 关系,不同的 Pre(v_i)节点之间也存在 AND 和 OR 关系。具体介绍如下:

(a)不同 Pre(o_i)节点间的 AND 关系,表示攻击者需要同时占有不同的资源,才能更进一步地实施攻击,形成攻击证据,进而占有更多的资源;

(b)不同 Pre(o_i)节点间的 OR 关系,表示攻击者只要占有任何一种资源,都可以实施攻击证据,进而占有更多的攻击资源;

(c)不同 Pre(v_i)节点间的 AND 关系,表示要想使攻击者到达下个资源状态节点,只有资源节点的各父亲节点中的攻击证据都取证完全时才能达到目的;

(d)不同 Pre(v_i)节点间的 OR 关系,表示任何一个该资源节点的父亲节点的攻击证据取证完全,都可以使攻击者成功占有该资源。

根据以上定义,生成网络攻击图如图 3 所示。

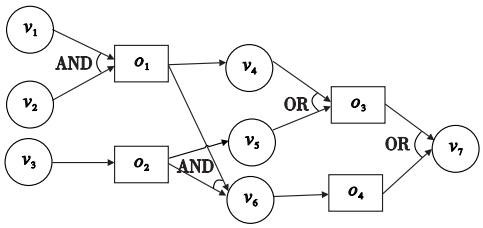

图3　网络攻击图

用有向边表示攻击者拥有的资源状态与攻击证据之间的关联关系,并且考虑各种情况,生成目标网络攻击图的主干结构。其中:v_1、v_2 和 v_3 是初始节点,v_7 是目标节点,要想成功占有资源 v_7,共有三条途径(用符号 ∧ 表示节点之间的 AND 关系):

a)$(v_1\wedge v_2)\rightarrow o_1\rightarrow v_4\rightarrow o_3\rightarrow v_7$;

b)$v_3\rightarrow o_2\rightarrow v_5\rightarrow o_3\rightarrow v_7$;

c)$((v_1\wedge v_2\rightarrow o_1)\wedge(v_3\rightarrow o_2))\rightarrow v_6\rightarrow o_4\rightarrow v_7$。

为了使网络攻击图中的节点之间的关系更加明确,下面详细介绍一下网络攻击图节点之间的线序关系的形成,即攻击路径的形成过程,用 STEP 表示线序关系。首先求出网络攻击图的偏序集合。这里,引入离散数学中入度和出度的概念。在有向图中,射入一个节点的边数称为该节点的入度,射出一个节点的边数称为该节点的出度。例如图 4 中攻击节点 o_1 的入度为 2,出度为 2;v_1 节点的入度为 0,出度为 1。在判断节点之间的偏序关系的过程中,考虑节点的入度和出度,根据入度和出度作出路径的选择:首先找出入度为 0 的节点 m,然后切断节点 m 关联的有向边,如果这个有向边关联的另外一个节点为 n,则此时记下节点 m 和节点 n 的偏序关系〈m,n〉;依此类推,最终得到网络攻击图中所有节点之间的偏序关系。偏序集合(partially ordered set,POS)中,虚线表示要切断的有向边,具体过程如图 4 所示。

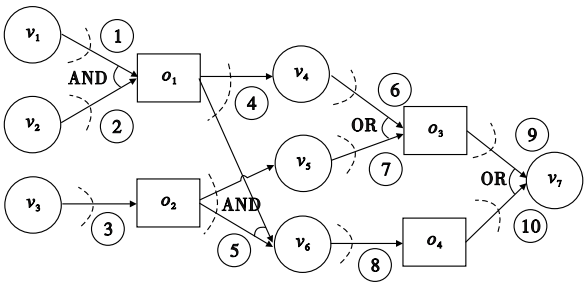

图4　偏序关系生成

本文作如下定义:R 表示节点的入度为 0 的集合;$R_{s_1,\cdots,s_n}$ 表示节点 $s_1,\cdots,s_n$ 的入度为 0;C 表示边的切断;C_{ab} 表示切断 $a\rightarrow b$ 的边;A 表示节点间的关系为 AND;A_{mn} 表示节点 m、n 之间为 AND 关系;POS_I 表示最初的偏序关系集合;POS_O 表示旧的偏序关系集合;POS_N 表示新的偏序关系集合;$A=B$ 表示把 B 的值赋值给 A。

其过程各步骤阐明如下:

a)入度为 0 的节点为 v_1、v_2、v_3。首先切断 $v_1\rightarrow o_1$ 边,且记下 v_1 和 v_2 之间是 AND 关系:$v_1\wedge v_2$,记 $\mathrm{POS}_I=\{\langle v_1,o_1\rangle\},(v_1\wedge v_2)$;即 $R_{v1}\wedge C_{V1O1}\wedge A_{V1O1}\rightarrow\mathrm{POS}_I$,$\mathrm{POS}_O=\mathrm{POS}_I$。

b)判断入度为 0 的节点 v_2。切断 $v_2\rightarrow o_1$ 边,记 $\mathrm{POS}_N=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle\},(v_1\wedge v_2)$;即 $R_{v2}\wedge C_{V2O1}\wedge\mathrm{POS}_O\rightarrow\mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

c)判断入度为 0 的节点 v_3,切断 $v_3\rightarrow o_2$ 有向边,记 $\mathrm{POS}_N=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle\},(v_1\wedge v_2)$;即 $R_{v3}\wedge C_{V3O2}\wedge\mathrm{POS}_O\rightarrow\mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

d)入度为 0 的节点为 o_1,o_2。切断 $o_1\rightarrow v_4$ 和 $o_1\rightarrow v_6$ 两条边,记 $\mathrm{POS}_N=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle,\langle o_1,v_4\rangle,\langle o_1,v_6\rangle\},(v_1\wedge v_2),(o_1\wedge o_2)$;即 $R_{o1}\wedge C_{o1v4}\wedge C_{o1v6}\wedge A_{o1o2}\wedge\mathrm{POS}_O\rightarrow\mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

e)入度为 0 的节点为 o_2。切断 $o_2\rightarrow v_5$ 和 $o_2\rightarrow v_6$ 两条边,记 $\mathrm{POS}_N=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle,\langle o_1,v_4\rangle,\langle o_1,v_6\rangle,\langle o_2,v_5\rangle,\langle o_2,v_6\rangle\},(v_1\wedge v_2),(o_1\wedge o_2)$;即 $R_{o2}\wedge C_{o2v5}\wedge C_{o2v6}\wedge\mathrm{POS}_O\rightarrow\mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

f)入度为 0 的节点为 v_4、v_5、v_6。切断 $v_4\rightarrow o_3$ 有向边,v_4 和 v_5 节点之间是 OR 关系,此时可以直接把每一个有向边加入偏序集合,无须备注它们之间的关系。记 $\mathrm{POS}_N=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle,\langle o_1,v_4\rangle,\langle o_1,v_6\rangle,\langle o_2,v_5\rangle,\langle o_2,v_6\rangle,\langle v_4,o_3\rangle\},(v_1\wedge v_2),(o_1\wedge o_2)$;即 $R_{v4}\wedge C_{v4o3}\wedge\mathrm{POS}_O\rightarrow\mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

g)入度为 0 的节点为 v_5、v_6。切断 $v_5\rightarrow o_3$,v_4 和 v_5 节点之间是 OR 关系,此时可以直接把每一个有向边加入偏序集合,无须备注它们之间的关系。记 $\mathrm{POS}_N=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle,\langle o_1,v_4\rangle,\langle o_1,v_6\rangle,\langle o_2,v_5\rangle,\langle o_2,v_6\rangle,\langle v_4,o_3\rangle,\langle v_5,o_3\rangle\},(v_1\wedge v_2),(o_1\wedge o_2)$;即 $R_{v5}\wedge C_{v5o3}\wedge\mathrm{POS}_O\rightarrow\mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

h)入度为 0 的节点为 v_6 和 o_3。切断 $v_6\rightarrow o_4$,记 $\mathrm{POS}=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle,\langle o_1,v_4\rangle,\langle o_1,v_6\rangle,\langle o_2,v_5\rangle,\langle o_2,v_6\rangle,\langle v_4,o_3\rangle,\langle v_5,o_3\rangle,\langle v_6,o_4\rangle\},(v_1\wedge v_2),(o_1\wedge o_2)$;即 $R_{v6}\wedge C_{v6o4}\wedge\mathrm{POS}_O\rightarrow\mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

i)入度为 0 的节点为 o_3 和 o_4。切断 $o_3\rightarrow v_7$ 有向边,记 $\mathrm{POS}=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle,\langle o_1,v_4\rangle,\langle o_1,v_6\rangle,\langle o_2,v_5\rangle,$

$\langle o_2,v_6\rangle,\langle v_4,o_3\rangle,\langle v_5,o_3\rangle,\langle v_6,o_4\rangle,\langle o_3,v_7\rangle\}$,$(v_1\wedge v_2)$,$(o_1\wedge o_2)$;即 $R_{o3}\wedge C_{o3v7}\wedge \mathrm{POS}_O\rightarrow \mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

j)入度为 0 的节点是 o_4。切断 $o_4\rightarrow v_7$,记 $\mathrm{POS}=\{\langle v_1,o_1\rangle,\langle v_2,o_1\rangle,\langle v_3,o_2\rangle,\langle o_1,v_4\rangle,\langle o_1,v_6\rangle,\langle o_2,v_5\rangle,\langle o_2,v_6\rangle,\langle v_4,o_3\rangle,\langle v_5,o_3\rangle,\langle v_6,o_4\rangle,\langle o_3,v_7\rangle,\langle o_4,v_7\rangle\}$,$(v_1\wedge v_2)$,$(o_1\wedge o_2)$;即 $R_{o4}\wedge C_{o4v7}\wedge \mathrm{POS}_O\rightarrow \mathrm{POS}_N$,$\mathrm{POS}_O=\mathrm{POS}_N$。

根据偏序集合 POS 可以得出节点集合之间不考虑节点间的 AND 关系时的偏序集合:

$\mathrm{POS}_1=\{\langle v_1,o_1\rangle,\langle o_1,v_4\rangle,\langle v_4,o_3\rangle,\langle o_3,v_7\rangle\}$;

$\mathrm{POS}_2=\{\langle v_2,o_1\rangle,\langle o_1,v_4\rangle,\langle v_4,o_3\rangle,\langle o_3,v_7\rangle\}$;

$\mathrm{POS}_3=\{\langle v_1,o_1\rangle,\langle o_1,v_6\rangle,\langle v_6,o_4\rangle,\langle o_4,v_7\rangle\}$;

$\mathrm{POS}_4=\{\langle v_2,o_1\rangle,\langle o_1,v_6\rangle,\langle v_6,o_4\rangle,\langle o_4,v_7\rangle\}$;

$\mathrm{POS}_5=\{\langle v_3,o_2\rangle,\langle o_2,v_5\rangle,\langle v_5,o_3\rangle,\langle o_3,v_7\rangle\}$;

$\mathrm{POS}_6=\{\langle v_3,o_2\rangle,\langle o_2,v_6\rangle,\langle v_6,o_4\rangle,\langle o_4,v_7\rangle\}$。

图中只有 v_1 和 v_2、o_1 和 o_2 两对节点之间存在 AND 关系,在这里,AND 关系表示只有这两个节点同时都满足时才能与下一个节点之间形成真正意义上的偏序关系。因此改进后的偏序集合形成了线序集合,如下所示就是本文的网络攻击图抽出的线序集合:

$\mathrm{STEP}_1=\{\langle\{v_1,v_2\},o_1\rangle,\langle o_1,v_4\rangle,\langle v_4,o_3\rangle,\langle o_3,v_7\rangle\}$;

$\mathrm{STEP}_2=\{\langle\{\langle\{v_1,v_2\},o_1\rangle,\langle v_3,o_2\rangle\},v_6\rangle,\langle v_6,o_4\rangle,\langle o_4,v_7\rangle\}$;

$\mathrm{STEP}_3=\{\langle v_3,o_2\rangle,\langle o_2,v_5\rangle,\langle v_5,o_3\rangle,\langle o_3,v_7\rangle\}$。

将 STEP_1、STEP_2 和 STEP_3 与前面根据网络攻击图得出的三条路径相比,可以看出由以上方法求出的线序关系与实际的攻击路径是一致的,同时证明了本方法的正确性和有效性。这个思想将作为下文中算法 2 的核心表达出来。

2.5 贝叶斯网络模型的计算

后验概率、最大可能解释和最大后验假设问题是进行贝叶斯网络推理的三种类型,本文主要是计算后验概率分布。在本文中,后验概率问题是指已知网络攻击图中起始节点集合的概率,计算目标节点集合后验概率分布的问题。为了便于说明问题,定义证据变量和查询变量的含义。证据变量就是指现有的数据分析已经得出概率的起始变量,用大写字母 E 来表示,取值用小写字母 e 来表示。查询变量则是指需要求出后验概率分布的目标变量,用大写字母 Q 表示,其值用小写字母 q 来表示。

求解后验概率问题的推理方法有很多种,通常分为精确推理和近似推理。但是由于网络结构的复杂性,精确推理很难适应结构庞大的贝叶斯网络结构,因此本文采用近似推理。近似推理目前已经有很多比较成熟的推理算法,本文采用随机推理算法中的一种——似然加权法[12]。似然加权法的核心思想就是根据适合的概率分布 $P(X)$ 进行随机抽样,抽取的样本组成样本空间,最后根据样本空间近似计算贝叶斯网络的后验概率分布。具体操作是:按照贝叶斯网络的网络攻击图中节点的偏序关系进行顺序抽样,在抽样的过程中,如果抽取的变量是已知的证据变量,则按照分布 $P(X)$ 进行抽样,并且把抽样的结果设为证据变量的观测值;如果抽取的变量不是证据变量,按照概率分布 $P(X|\pi(X))$ 进行抽样,并且抽样的结果就是样本的取值结果。似然加权法的用意旨在不造成抽样样本的浪费,使抽取的每一个证据变量的结果都与要求的后验概率分布中的证据变量的取值一致,从而使每一个样本都能够有效地发挥作用。

假定抽样结果中抽取到 $E=e$ 的概率为 $e_i\,|\,i=1,2,\cdots,n$,在满足 $Q=q$ 的前提下抽取到 $E=e$ 的概率为 $q_i\,|\,i=1,2,\cdots,n$,那么计算后验概率的公式为 $P(Q=q|E=e)\approx\dfrac{\sum_{i=n}e_i}{\sum_{i=n}q_i};i=1,2,\cdots,n$。由于本文提出的贝叶斯网络攻击图中节点之间存在 AND 和 OR 关系,在计算似然加权时,节点间的拓扑顺序受到影响,因此本文提出似然加权法的改进算法,即算法 1 和算法 2。

算法 1 似然加权法算法 likelihood weighted (NAG,m,E,e,Q,q)

描述:本算法根据贝叶斯网络的网络攻击图和已知的攻击证据变量,求出查询变量的后验概率分布。

输入:贝叶斯网 NAG,样本量 m,证据变量 E,证据变量的取值 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。

```
1:A←STEP(NAG,D);
2:ω_e←0;ω_{q,e}←0;
3:for(i=1 to m)
4:D_i←0;
5:for(A 中的每一个变量 X)
6:if(X∈E)
7:x←X 的观测值;
8:else
9:x←从 P(X|π(X)) 抽样的结果;
10:end if
11:end for
12:D_i←D_i∪{X=x};
13:ω_i←∏_{X∈E} P(X|π(X))|D_i;
14:ω_e←ω_e+ω_i;
15:if(D 与 D=q 一致)
16:ω_{q,e}←ω_{q,e}+ω_i;
17:end if
18: end for
19: return ω_{q,e}/ω_e.
```

算法 2 线序关系 STEP(NAG,SETP_i)

描述:根据贝叶斯网络的网络攻击图 NAG,找出具有顺序的攻击路径节点集合 D。

输入:网络攻击图 NAG,入度 λ_1,出度 λ_2,偏序关系 POS,线序关系 STEP,X 和 Y 任意两个节点,AND 关系集 M;

输出:贝叶斯网络的线序关系节点集合。

```
1:POS←∅;POS_i←∅;SETP_i←∅;M←∅;
2:for(每一个 λ_1=0 的节点变量 X);
3:在 NAG 中找出与节点 X 之间存在有向边的所有的节点变量 Y;
4:if(Y 的入度>1 且射入 Y 的两条有向边之间存在 AND 关系)
5:M←M∪{X_1∧X_2};
6:end if;
7:POS←POS∪{⟨X,Y⟩}
8:X 的出度 λ_2←0;
9:end for;
10:在 POS 集合中,遍历查找所有不考虑 AND 关系的偏序集合 POS_i;
11:在 POS_i 集合中, 查找满足 AND 关系的线序集合 SETP_i;
12:return SETP_i。
```

3 实验数据及分析

实验的环境为 Windows 系统的小型局域网组成的内部网络。对于图 3 的网络攻击图,利用随机数产生器对于抽样过程中涉及到的变量进行随机抽样。举例说明:对于服从伯努利分布的随机变量 X,取值为 1 或 0。假设 $P(X=0)=p$,$P(X=1)=1-p$,对 $P(X)$ 抽样的过程如下:利用随机数产生器产生一个实数 $x\in[0,1]$。如果 x 的值在 $[0,p]$ 区间里,则抽样的样本结果为 $X=0$;否则,$X=1$。

似然加权法的核心思想是按照网络结构图的拓扑序对每一

个节点变量进行多次抽样,最后根据抽样组合数据统计分析后验概率。按照采样时间的不同进行抽样统计分析,每组抽样个数为 30 000 个。用 N_i 表示采样的时间点。以 $P(v_7|v_1,v_2,v_3)$ 的概率值作为抽样结果的统计项,$p = P(v_7 = \text{true}|v_1 = \text{true}, v_2 = \text{true}, v_3 = \text{true})$ 为所求概率值。本次实验总共进行了四次采样过程,用 P_i 表示每次采样的概率抽样组合,采样结果如表 1 所示。

表 1　采样结果

	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}
P_1	0.278 9	0.333 1	0.496 5	0.542 5	0.606 1	0.677 8	0.747 5	0.798 2	0.853 6	0.952 7
P_2	0.192 7	0.265 0	0.302 4	0.452 1	0.534 7	0.602 4	0.718 9	0.803 4	0.815 6	0.830 0
P_3	0.124 9	0.302 4	0.365 4	0.453 6	0.536 7	0.602 4	0.678 9	0.724 6	0.790 0	0.820 0
P_4	0.250 0	0.354 7	0.482 1	0.527 8	0.619 6	0.705 6	0.759 8	0.823 6	0.895 6	0.920 0

由实验所得采样结果可以看出,随着采样时间的不同,攻击者的攻击概率逐渐增大,这与实际攻击途径中的效果是一致的。由于随着攻击进度的推进,攻击者拥有和熟悉的攻击资源越来越多,靠近攻击目标越来越近,对系统造成的威胁可能性也越来越大。

图 5 和 6 显示了本文提出的内部威胁预测模型计算得出的用户攻击进度。假设管理人员设定警戒值为 0.85,当监测到用户攻击概率达到 0.85 时,阻止用户的攻击行为。从走势图可以看出,该模型能够实时监测内部用户的攻击行为,通过后验概率值量化内部用户的攻击概率,从而给内部网络的管理决策人员提供参考,以便作出正确的决策。

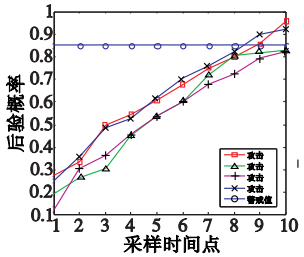

图5　内部威胁模型中用户攻击监测图1

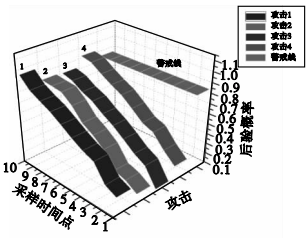

图6　内部威胁模型中用户攻击监测图2

4　结束语

本文提出用贝叶斯网络模型来描述内部威胁的网络结构。并且利用贝叶斯网络的推理算法—似然加权法,计算内部威胁网络攻击图中的后验概率问题。从而定性和定量分析了内部威胁的网络结构。最后通过实验证明,该模型能够准确有效地预测内部威胁。

参考文献:

[1] LIU Yu, MAN Hong. Network vulnerability assessment using Bayesian networks[C]//Proc of SPIE 5812. New York: ACM Press, 2005: 61-71.

[2] FRIGAULT M, WANG Ling-yu. Measuring network security using bayesian network-based attack graphs[C]//Proc of IEEE International Computer Software and Applications Conference. Washington DC: IEEE Computer Society, 2008: 698-703.

[3] 王桢珍,姜欣,武小悦,等. 信息安全风险概率计算的贝叶斯网络模型[J]. 电子学报,2010,38(2A):18-23.

[4] 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.

[5] GHINITA G. The optimization of situational awareness for insider threat detection[C]//Proc of the 1st ACM Conference on Data and Application Security and Privacy. New York: ACM Press, 2011: 231-235.

[6] NELLIKAR S, NICOL D M, CHOI J J. Role-based differentiation for insider detection algorithms[C]//Proc of ACM Conference on Computer and Communications Security. New York: ACM Press, 2010: 55-62.

[7] 王国玉,王会梅,陈志杰,等. 基于攻击图的计算机网络攻击建模方法[J]. 国防科技大学学报,2009,31(4):74-81.

[8] 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.

[9] 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):134-6-1355.

[10] CHINCHANI R, UPADHYAYA S, KWIAT K. Towards the scalable implementation of a user level anomaly detection system [C]//Proc of IEEE Military Communications Conference. 2002: 1503-1508.

[11] UPADHYAYA S, CHINCHANI R, KWIAT K. An analytical framework for reasoning about intrusions [C]//Proc of the 20th IEEE Symposium on Reliable Distributed Systerms. 2001: 99-105.

[12] FUNG R, CHANG K C. Weighting and integrating evidence for stochastic simulation in Bayesian networks[C]//Proc of the 1st Annual Conference on Uncertainty in Artificial Intelligence. 1989: 209-219.

(上接第 2758 页)

参考文献:

[1] 3GPP TS 33.401 v11.1.0, 3rd generation partnership project; technical specification group services and system aspects; 3GPP system architecture evolution (SAE) security architecture[S]. Valbonne: 3GPP, 2012.

[2] HAN C K. Security analysis and enhancements in LTE-advanced networks [D]. Seoul: Sungkyunkwan University, 2012.

[3] GINS E A, RAPHAEL C W, PARISH D J. Analysis and design of security for next generation 4G cellular networks[C]//Proc of the 13th Annual Post Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting. 2012.

[4] FORSBERG D, HUANG Le-ping, TSUYOSHI K, *et al*. Enhancing security and privacy in 3GPP E-UTRAN radio interface[C]//Proc of Personal, Indoor and Mobile Radio Communications. 2007: 1-5.

[5] MJØLSNES S, TSAY J K. Computational security analysis of the UMTS and LTE authentication and key agreement protocols [EB/OL]. (2012-03-17)[2013-01-08]. http://arxiv.org/abs/1203.3866.

[6] FUHR T, GILBERT H, REINHARD J R, *et al*. Analysis of the initial and modified versions of the candidate 3GPP integrity algorithm 128-EIA3[C]//Proc of the 18th International Workshop on Selected Areas in Cryptography. 2011: 230-242.

[7] 3GPP TS 33.102 v11.5.0, 3rd generation partnership project; technical specification group services and system aspects; 3G secu-rity; security architecture[S]. Valbonne: 3GPP, 2012.

[8] 3GPP TS 24.301 v11.4.1, 3rd generation partnership project; technical specification group core network and terminals; Non-access-stratum (NAS) protocol for evolved packet system (EPS) [S]. Valbonne: 3GPP, 2012.

[9] ORHANOU G, HAJJI S E, BENTALEB Y. SNOW 3G stream cipher operation and complexity study [J]. Contemporary Engineering Sciences, 2010, 3(3): 97-111.

[10] TUNSTALL M. Practical complexity differential Cryptanalysis and fault analysis of AES[J]. Journal of Cryptographic Engineering, 2011, 1(3): 219-230.

[11] ETSI TC SAGE v1.6, 3GPP confidentiality and integrity algorithms 128-EEA3 & 128-EIA3 [EB/OL]. (2011). http://www.etsi.org/index.php/services/security-algorithms/3gpp-algorithms.