文章编号:1007－6735(2022)04－0388－09　DOI: 10.13255/j.cnki.jusst.20211009003

基于隐马尔可夫模型的电力信息系统动态威胁定量分析

苏鹏涛[1]，　吴　贶[2]，　陈孟婕[3]，　张雪芹[3]
（1. 上海欣能信息科技发展有限公司，上海 200025；2. 上海挚达科技发展有限公司，上海 200433；
3. 华东理工大学 信息科学与工程学院，上海 200237）

摘要：针对典型电力信息系统的网络威胁定量评估问题，提出了基于网络入侵检测系统（network intrusion detection syetem，NIDS）报警信息和隐马尔可夫模型的网络威胁动态分析方法 HMM-NIDS。该方法充分利用 NIDS 报警信息，从优先级、严重度、资产值和可信度 4 个方面分析 NIDS 报警信息，给出了报警威胁定量描述和分类方法，优化了隐马尔可夫模型中的观测矩阵；基于贝叶斯网络分析攻击成功的可信度，避免 NIDS 误警信息干扰；基于改进的隐马尔可夫模型，融合得到系统的动态风险量化值。基于 Darpa2000 实验场景模拟 DDoS 攻击，通过对比实验，验证了所提方法的有效性和优越性。

关键词：电力信息系统；威胁定量分析；隐马尔可夫；入侵报警；贝叶斯网络
中图分类号：TN 918.9　　**文献标志码：**A

Dynamic threat quantitative analysis of power information system based on hidden Markov model

SU Pengtao[1], WU Kuang[2], CHEN Mengjie[3], ZHANG Xueqin[3]
(1. *Shanghai Shineenergy Information Technology Development Co., Ltd., Shanghai 200025, China*; 2. *Shanghai Zhida Technology Development Co., Ltd., Shanghai 200433, China*; 3. *School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China*)

Abstract: Aiming at the problem of network threat quantitative evaluation of typical power information system, a network threat dynamic analysis method hmm-ids based on network intrusion detection syetem (NIDS) alarm information and hidden Markov model was proposed in this paper. NIDS alarm information was fully used to analyzes alarm threats from four aspects: priority, severity, asset value and reliability. A quantitative description and classification method of alarm threats were given and the observation matrix in hidden Markov model was optimized. The reliability of successful attack based on Bayesian network was analyzed, which avoided the interference of NIDS false alarm information. Based on the improved hidden Markov model, the dynamic risk quantification value of the system was obtained by fusion. DDoS attacks were simulated based on DARPA2000 experimental scenario. Through comparative experiments, the effectiveness and superiority of the proposed method were verified.

收稿日期: 2021−10−09
第一作者: 苏鹏涛（1978−），男，高级工程师．研究方向：计算机与信息技术应用．E-mail：supt@shineenergy.com
通信作者: 张雪芹（1972−），女，教授．研究方向：信息安全、信息分析与处理．E-mail: zxq@ecust.edu.cn

Keywords: *power information system*; *quantitative threat analysis*; *hidden Markov*; *intrusion alarm*; *Bayesian network*

电力信息系统与其他计算机信息系统一样，面临着各种各样的安全威胁。传统的信息系统威胁评估通常采用 CORAS 方法、层次分析法、贝叶斯网络法、以及 ISSREM 方法等。但是，随着网络攻击的多元化、复杂化，这些方法不能有效地实现对网络威胁进行实时动态定量评估。对此，李欣等[1]提出了一种基于改进隐马尔可夫模型的网络安全态势评估方法，该方法通过结合人群搜索算法（seeker optimization algorithm，SOA）以解决 Baum-Welch 参数优化算法易陷入局部最优解的问题，实验表明，使用优化后的参数进行量化分析有效地提高了模型的准确率。梁智强等[2]设计了一种新型的信息安全风险评估模型，该模型通过在风险评估与定量分析中引入层次分析法，并在风险计算过程中融合模糊数学知识，有效地在降低评估成本的基础上提高了评估效率。马刚等[3]针对大规模分布式复杂信息系统的风险评估，提出了一种基于威胁传播树 (threat propagation trees，TPT) 的系统风险评估定量分析方法，该方法利用采样资产节点的转移状态和发出的威胁传播边来生成 TPT，通过计算 TPT 的概率和 TPT 中资产的期望损失定量评估系统的风险。Ciapessoni 等[4]提出了基于改进的概率风险动态安全评估方法，将概率风险与时间域模拟评估应急影响相结合，提供不稳定风险指标。该方法可应用于当前的电力系统以及预测电力系统状态，为操作和运营规划环境提供了有价值的支撑。Xiong 等[5]提出了基于动态攻防博弈的信息网络漏洞威胁评估模型，根据信息系统实际控制调度，兼顾可用资源的数量、资源的配置合理性、成本效益和攻击节点的数量等制约条件，建立了信息网络漏洞威胁评估模型，实现了对网络物理系统漏洞威胁的定量评估。周未等[6]提出了一种层次化的网络安全风险评估框架，通过计算单个脆弱点的置信度，并综合分析攻击危害指数以及危害指数，推算节点攻击安全风险值，继而依据各节点的权重值来量化全网的安全风险。张至元等[7]提出基于广义随机佩式网和网络安全拓扑确定的状态转移图 Cybernet，利用电力监控系统的历史流量数据对初始入侵概率分量进行建模，根据系统运行结果和马尔可夫链稳态概率量化电力系统信息安全和工程安全的相关关系，通过仿真验证了该模型的准确性和适用性。杨英杰等[8]提出一种基于属性攻击图的动态威胁风险分析模型，该方法结合通用漏洞评分标准和贝叶斯概率转移方法，设计了单步及综合威胁转移概率度量策略，并且使用动态威胁属性攻击图生成算法，实现了消解攻击图中威胁传递环路干扰的目的。王辉等[9]提出了一种基于新型贝叶斯模型的网络风险评估方法，该模型通过采用删除节点次序算法以确定消元顺序，并在计算节点的后验风险概率时，采用团树传播算法进行动态计算，有效降低了后验概率计算时间。张雪芹等[10]提出基于属性攻击图和贝叶斯网络的社会工程学威胁评估方法，通过定义社会工程学的可利用的脆弱性语义和攻击节点语义，提出相应的脆弱性可利用概率计算方法，根据属性攻击图构建社会工程学攻击图，采用贝叶斯网络模型对其进行量化评估。

可见，采用隐马尔可夫模型、攻击图、贝叶斯网络等方法可以有效实现对信息系统威胁的动态量化评估，但是这些方法都没有充分利用信息系统中部署的网络入侵检测系统（network intrusion detection syetem，NIDS）的报警信息，而 NIDS 的报警信息具有威胁检测实时性强的特点，能够很好地反映当前系统所面临的威胁变化。对此，本文提出一种基于 NIDS 报警信息和改进的隐马尔可夫模型的网络威胁动态分析方法 HMM-NIDS。该方法提出综合优先级、严重度、资产值和可信度 4 个方面对 NIDS 报警信息进行分析，基于贝叶斯网络分析 NIDS 报警信息的可信度，基于改进的隐马尔可夫模型动态计算系统风险值，实现电力信息系统威胁的动态定量评估。

1　相关理论

1.1　HMM 隐马尔可夫模型

1.1.1　马尔可夫链

马尔可夫链通常用于描述一组互相转化关系之间具有无后效性关系的状态 [x]。设在任一时刻 t，随机序列O_t可以处于的状态为$s_1, s_2, \cdots, s_N$，假

设它在$m+k$时刻处于某一个状态q_{k+m}的概率只与它在m时刻的状态q_m有关，而与m时刻以前的状态无关，即$P(O_{m+k}=q_{m+k}|O_m=q_m, O_{m-1}=q_{m-1},\cdots,O_1=q_1)=P(O_{m+k}=q_{m+k}|O_m=q_m)$, $q_1, q_2,\cdots,q_m,q_{m+k}(s_1, s_2,\cdots,s_N)$，则称$O_t$为Markov链，并且称$P_{ij}(m, m+k)=P\left(q_{m+k}=s_j\middle|q_m=s_j\right),1\leqslant i,j\leqslant N$，为$k$步转移概率。

1.1.2 隐马尔可夫模型

隐马尔可夫模型(hidden Markov model，HMM)，是由Baum和Welch提出的一种统计分析模型，被广泛应用在语音识别、图像处理、故障诊断等领域[8-9]。隐马尔可夫模型通常用于描述一个存在隐含参数的马尔可夫过程，是一个双重随机过程。一个HMM模型包括如下元素：

a. s，HMM中隐藏Markov链。设有N个状态$s_1,s_2,\cdots,s_N$，在t时刻所处的状态为q_t，则$q_t\in s_1, s_2,\cdots,s_N$。

b. $\boldsymbol{V}$，观测状态集合。设有M个观测状态，$\boldsymbol{V}=\{v_1,v_2,\cdots,v_M\}$。观测序列$\boldsymbol{O}=\{o_1,o_2,\cdots,o_T\}$，其中$o_t\in\boldsymbol{V}$，$T$表示观测序列的长度。

c. $\boldsymbol{\pi}$，初始状态概率分布。$\pi_i=P(q_1=v_i)$，其中$1\leqslant i\leqslant N$。向量表示$\boldsymbol{\pi}=(r_1,\cdots,r_N)$。

d. $\boldsymbol{T}$，状态转移矩阵。

$$T_{ij}=P(q_{t+1}=s_j|q_t=s_i),1\leqslant i,j\leqslant N \qquad (1)$$

上式表示在t时刻状态为s_i，则$t+1$时刻状态为s_j的概率。

e. $\boldsymbol{O}$，观测矩阵。

$$O_{nm}=P(o_t=v_m|q_t=s_n) \qquad (2)$$

式中：$1\leqslant n\leqslant N$，$1\leqslant m\leqslant M$。O_{nm}表示在时刻t、主机处于s_n状态下观察到v_m的概率。

隐马尔可夫模型通常由两个部分构成，如图1所示：第一部分采用π，P状态转移概率描述马尔可夫链；第二部分采用Q描述随机过程。第一部分的输出结果为状态向量，第二部分的输出结果为观察值向量。

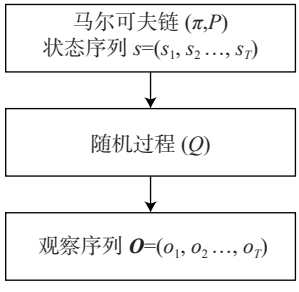

图1 隐马尔可夫模型过程

Fig. 1 Process of HMM

1.2 贝叶斯网络

贝叶斯网络是一种概率图模型，其网络拓扑结构是一个有向无环图(directed acyclic graph，DAG)，即图中所有边都是有方向的，并且没有循环，用来刻画随机变量之间的依赖关系。贝叶斯网络为了能够表示随机变量之间的因果关系或非条件独立，把某个研究系统中所关联到的所有随机变量，依据变量之间是否满足条件独立将其绘制在一个有向图中。图2展示了一个贝叶斯网络，边集是$E=\{(A,C),(B,C)\}$，由于A, B相互独立，因此$P(A|C,B)=P(A|C)$。这意味着，对于这个贝叶斯网络，A的概率只取决于C，而B值与这个局部概率无关。对图2，可推出$P(A,B,C)=P(A|C)*P(C)*P(B|C)$成立。

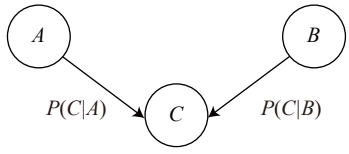

图2 贝叶斯网络结构

Fig.2 Bayesian network structure

一般来说，在给定DAG中的节点$X=\{X_1,X_2,\cdots,X_n\}$的条件下，任何贝叶斯网络的联合概率函数都是$P(X)=\prod_{i=1}^{n}P(X_i|X_{i-1})$，其中$X_{i-1}$为$X_i$的父结点。

2 基于NIDS报警信息和HMM的电力信息系统威胁定量评估

电力信息系统是一类特定的信息系统。一个典型的电力信息系统通常包括生产控制区和信息管理区。控制区通常包括分散控制系统DCS、电能计量系统、调度系统和保护管理系等。信息管理区通常有办公自动化系统、生产管理信息系统等，信息管理区的关键硬件资产包括交换机、网关、操作员站、工程师站等。

为了实现对电力信息系统遭受威胁的动态定量评估，通常可以先计算出系统中每个资产(主机、路由器、系统等)的风险值，继而得到整个系统的风险值。但是，由于威胁的动态变化，要定量分析系统中每个资产的安全动态通常是一个难点问题。通过文献分析可以看到，基于隐马尔可夫模型能够对信息系统的安全态势进行有效量化评估。同时，电力信息系统中通常在防火墙后部

署有网络入侵检测系统 NIDS，用于实时检测和发现网络攻击。为此本文提出基于 HMM 和 NIDS 报警信息分析和资产相关的网络安全事件，通过综合评估感知资产的威胁状态，最终实现对系统面临威胁的定量评估。

2.1　基于 HMM 的电力信息系统威胁定量评估

设在电力信息系统中，每个资产的安全状态变化构成一个马尔可夫链，信息系统中 NIDS 的报警序列是一个随机过程，每发生一个报警，就会对资产安全状态的转移带来影响（不同类型和不同威胁度的报警带来的影响不同）。因此，NIDS 的报警序列和资产的安全状态之间就形成了一个隐马尔可夫模型，可根据观察一段时间的报警序列，计算出该资产的安全状态以及风险值。

2.1.1　主机的安全状态

定义系统中的资产有 4 个安全状态，$S=\{\mathrm{G},\mathrm{P},\mathrm{A},\mathrm{C}\}$，含义如表 1 所示。4 个状态的转换关系如图 3 所示。

表 1　网络威胁状态描述

Tab.1　Description of network threat state

状态序列	状态	状态描述
S_1	G(Good)	良好状态，该主机没有遭受攻击
S_2	P(Probed)	该主机处于被攻击者探测到的状态
S_3	A(Attacked)	该主机被攻击过，增加了被入侵的概率
S_4	C(Compromised)	该主机已经被入侵，处于最危险的状态

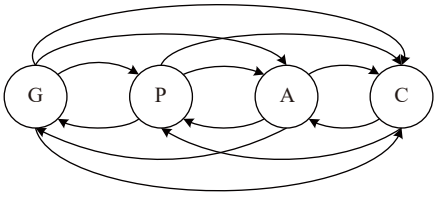

图 3　网络威胁状态转换图

Fig.3　Network threat state transition diagram

2.1.2　主机安全状态的分布概率

HMM包含一个三元组$\lambda=(\boldsymbol{T},\boldsymbol{O},\boldsymbol{\pi})$。其中，$\boldsymbol{T}$表示资产状态转换概率的状态转换矩阵；$\boldsymbol{O}$表示当资产处于某一特定状态时，观察到某种攻击概率的观察矩阵；初始状态$\boldsymbol{\pi}$则代表计算开始时资产处于各个状态的概率。

资产状态的初始值可依据主机的漏洞信息取得，其后资产处于各个状态的概率可通过$\boldsymbol{O}$和O计算。状态分布在t时刻表示为$r_t=\{r_t(i)\}$，$1\leqslant i\leqslant N$，状态的分布概率公式表示为

$$r_1(i)=\pi_i\times O_i(o_1) \tag{3}$$

$$r_{t+1}(i)=\left[\sum_{j=1}^{N}r_t(j)\times T_{ji}\right]\times O_i(o_{t+1}) \tag{4}$$

$O_i(o_{t+1})$是在 t+1 时刻观察到观测序列o_{t+1}的概率。

2.1.3　主机和系统风险值的定量计算

为了能够定量计算资产的风险值，引入代价向量$\boldsymbol{C}=\{c_1,c_2,\ldots,c_N\}$，该向量表示某资产在各个状态下的量化风险值，通过该量化处理可以将资产状态转化为定量风险值 R

$$R=\sum_{i=1}^{N}r_ic_i \tag{5}$$

式中：R 为 1～4 表示攻击有一定可能被探测到；R 为 4～7 表示资产已经遭受到攻击；R 为 7～10 则表示攻击比较严重。

假设一个系统中共有 L 个资产，那么整个系统的风险值表示为

$$R_{\mathrm{net}}=\sum_{i=1}^{L}R_i \tag{6}$$

2.2　基于 NIDS 报警信息的威胁分析

2.2.1　NIDS 报警类型分类

在电力信息系统中，借助入侵检测系统的报警信息可以了解某资产处于某一特定状态时，遭受到某种攻击的情况。但是由于 NIDS 报警类型繁多，如 Snort 的基本报警类型就有几千个[11]，由于直接将 NIDS 报警和 HMM 中的观察矩阵 $\boldsymbol{O}$ 关联会使得观察矩阵的规模相当庞大，导致算法运行效率严重降低。因此需寻找一种合适的报警归类方法，缩小观察矩阵的规模，加快计算速度。

为了合理根据报警信息评估主机状态，除了考虑攻击报警本身之外，还应综合考虑被攻击主机的信息，以及 NIDS 虚警的可能性。为此，本文通过对报警严重程度、目标资产关键程度、攻击目标的优先等级和攻击成功执行的可能性进行分析，提出采用严重度（severity）、资产值（asset）、优先级（priority）、可信度 (reliability) 4 个因素相结合的方式来综合分析和分类报警信息，如表 2 所示。

表 2　报警数据处理分类

Tab.2　Alert data processing classification

报警威胁	严重度	报警本身的严重程度
	资产值	攻击所针对的目标资产关键程度
	优先级	攻击类型或攻击目标的优先等级
	可信度	网络环境下攻击成功执行的可能性

a. 严重度

NIDS，如 Snort，对入侵威胁类型有较为明确的分类和严重等级定义，如表 3 所示，通过查询该表单可以得到相应等级的严重度 S。

表 3 基于 Snort 的严重度分级

Tab.3 Severity classification of snort alert

等级	攻击类型	攻击描述
9	远程管理员权限	获得超级用户权限(administrator 和root 等),可以完全控制系统,能够对目标主机造成最严重的影响。
8	远程普通用户权限	获得一定的权限,可以连接到系统,并且执行一些操作来影响目标主机的机密性、完整性和可用性。在此基础上可能进一步提升至超级用户权限。攻击者的下一步入侵目标就是获得系统的全部控制权。
7	拒绝服务	系统的可用性遭到破坏,不能正常运行一些服务和执行一些操作。系统的机密性和完整性将要遭到进一步的破坏。
6	远程文件写	完整性遭到破坏,数据被损坏。
5	远程文件读	机密性破坏,机密信息遭到窃取。
4	攻击命令中继	攻击被执行,即将造成较严重的攻击。
3	恶意代码执行	恶意代码执行,具有一定的破坏性,但对机密性、完整性和可用性不构成较大破坏。
2	安全信息泄露	违背安全策略行为,造成系统信息泄露。可能作为攻击者执行下一步攻击步骤的前提,如被嗅探和扫描。
1	无效攻击	攻击手段无效,无法突破被试网。
0	无威胁	正常网络用户行为(IDS 误报警)。

b. 资产值

通常在进行威胁分析时，可以由相关管理人员给出电力信息系统资产清单，并根据信息技术安全评估标准（information technology security evaluation criteria，ITSEC）给出的资产 CIA 三性(机密性、完整性与可用性) 计算资产值 A 并划分资产等级。设资产值为 v，则

$$v = f(x,y,z) = \sqrt{\sqrt{x \times y} \times z} \tag{7}$$

式中：x 为资产的机密性；y 为资产的完整性；z 为资产的可用性值，通常由相关管理人员结合企业业务来确定。以某电力信息系统为例，表 4 给出了关键资产清单及其资产等级示例。表中资产等级 1，2，3，4，5 分别代表资产等级为很低、低、中、高、很高。

表 4 系统资产清单示例

Tab.4 Example of system asset list

资产编号	资产名称	资产等级
ASSET_1	单向隔离器-01	4
ASSET_2	网关机-01	4
ASSET_3	Cisco路由器-01	4
ASSET_4	PC_01	2
ASSET_5	MIS服务器	5
ASSET_6	电子数据	4
ASSET_7	人员文档	3
ASSET_8	安全管理制度	3

c. 优先级

优先级用于表示攻击类型的优先级。Snort 中含有攻击优先级信息，级别从 1 到 3。1 代表级别最高，3 代表级别最低。表 5 给出优先级 P 的量化值。

表 5 NIDS 报警优先级分类

Tab.5 Classification of priority

优先级	1	2	3
量化优先级	0.9	0.6	0.3

2.2.2 基于贝叶斯网络的可信度分析

可信度用于衡量攻击发生的真实性以及成功执行的可能性。由于 NIDS 报警信息中不可避免地存在着虚警或误报情况，为了更合理地评估风险状态，需要对报警的可信度做进一步分析。为此，本文提出结合目标资产的配置信息、运行状态以及成功执行攻击所依赖条件，基于贝叶斯网络对可信度 R 进行计算。

a. 构建贝叶斯网络的拓扑结构

贝叶斯网络方法使用概率值来表示变量之间的依赖关系。影响攻击成功执行概率的相关因素可以通过贝叶斯网络进行推测。考虑Microsoft公司安全小组对攻击向量的定义，同时考虑到网络环境中的威胁和攻击同样会对漏洞的脆弱性产生一定的影响，建立动态威胁分析贝叶斯网络的拓扑结构如图 4 所示。

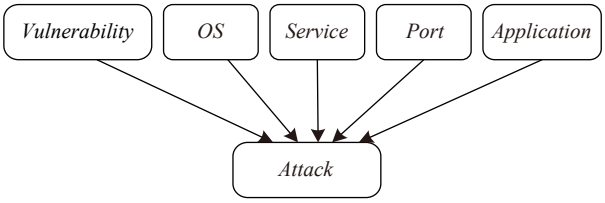

图 4　动态威胁分析贝叶斯拓扑结构

Fig.4　Bayesian network topology of network threat

在上述贝叶斯网络拓扑中的 5 个节点分别代表漏洞、操作系统、服务、端口和应用。设每个节点有“Yes”和“No”两种状态，各个节点处在“Yes”状态的含义如表 6 所示。

b. 确定条件概率分布

条件概率表（conditional probability table，CPT）通常可以通过专家知识和对历史数据的学习获得[12]，示例如表 7 所示。

表 6　贝叶斯网络各节点信息

Tab.6　Node information of Bayesian network

节点信息	“Yes”状态含义
Attack	攻击成功(Succeed)
Vulnerability	目标主机上检测到攻击针对的漏洞
OS	目标主机的操作系统受该攻击的影响
Service	目标主机上攻击针对的服务正在运行
Port	目标主机上攻击针对的端口可连接
Application	目标主机上应用程序版本受该攻击的影响

表 7　条件概率表示例

Tab.7　Parts of CPT reliability assessment

V_ID	*OS*	*Ser*	*Port*	*App*	*Attack*	*P*(*Attack*|*V_ID*,*OS*,*Ser*,*Port*,*App*)	*P*(*V_ID*)	*P*(*OS*)	*P*(*Ser*)	*P*(*Port*)	*P*(*App*)	*P*(*Jiont*)
Yes	Yes	Yes	Yes	Yes	Succeed	0.99	0.4	0.8	0.7	0.6	0.75	0.099 8
Yes	Yes	No	Yes	Yes	Succeed	0.8	0.4	0.8	0.3	0.6	0.75	0.034 5
Yes	Yes	Yes	No	Yes	Succeed	0.3	0.4	0.8	0.7	0.4	0.75	0.020 2
Yes	Yes	No	No	Yes	Succeed	0.25	0.4	0.8	0.3	0.4	0.75	0.007 2
Yes	Yes	No	Yes	Yes	Succeed	0.2	0.4	0.8	0.3	0.6	0.75	0.008 6
Yes	No	No	Yes	Yes	Succeed	0.1	0.4	0.2	0.3	0.6	0.75	0.001 1
Yes	No	No	No	Yes	Succeed	0.05	0.4	0.2	0.3	0.4	0.75	0.000 4
No	Yes	Yes	Yes	Yes	Succeed	0.8	0.6	0.8	0.7	0.6	0.75	0.121 0
No	Yes	No	Yes	Yes	Succeed	0.4	0.6	0.8	0.3	0.6	0.75	0.025 9
No	Yes	No	No	Yes	Succeed	0.2	0.6	0.8	0.3	0.4	0.75	0.008 1
.....												

表 7 中，*V_ID* 表示漏洞信息；*Ser* 表示正在运行的服务；*App* 表示应用程序。以第一行数据为例，$P(V_ID)=0.4$表示 V_ID 节点处在“Yes”状态的先验概率为 0.4，其后 4 个概率分别表示 *OS*，*Ser*，*Port*，*App* 处在“Yes”状态的先验概率；$P(Attack|V_ID, OS, Ser, Port, App)=0.99$表示 5 个节点均处在“Yes”状态时攻击成功的概率为 0.99；$P(Jiont)=P(Attack=\text{succeed}, App=\text{yes}, V\_ID=\text{yes}, OS=\text{yes}, Port=\text{yes}, Ser=\text{yes})=0.099\,8$，表示全部节点均处在“Yes”状态的联合概率为 0.099 8。

c. 可信度计算

首先可使用 Nmap 等工具获取实际环境中的主机节点信息，包括操作系统、端口、应用、服务等，接下来可以通过漏洞扫描工具获取在目标资产中存在的漏洞，漏洞信息可以在美国国家通用漏洞数据库（national vulnerability database，NVD）[13]中查询得到。

在获取 5 个节点的状态后，使用贝叶斯网络计算攻击成功的概率。设节点状态数据集为 *D*，需要进行推理的网络结构为 *B*，*P*(*D*) 和 *P*(*B*) 为先验概率，则贝叶斯网络的推理目标计算后验概率 *P*(*B*|*D*) 的公式为

$$P(B|D)=\frac{P(D,B)}{P(D)} \tag{8}$$

在现实网络环境中，若一个攻击所针对的漏洞信息、操作系统和应用程序都与目标主机运行状态相匹配，而其他 3 个节点的状态均是未知的，则状态 *D* 表示为

$$D=(V\_ID=\text{yes}, OS=\text{yes}, App=\text{yes})$$

根据表 7 的前 5 行，可信度 *R* 可由下式计算：

$$R = P(Attack = \text{succeed}|D) = \frac{P(Attack = \text{succeed}, V\_ID = \text{yes}, OS = \text{yes}, App = \text{yes})}{P(V\_ID = \text{yes}, OS = \text{yes}, App = \text{yes})} = 0.701\,9$$

d. 报警信息分类

根据上述分析可以得到攻击严重度、资产价值、优先级和可信度的评分指标，为了实现对报警信息的分类，首先将各因素做归一化处理，之后采用加权融合，得到威胁的综合严重程度分析步骤如下。

第一步：归一化处理。定义攻击严重度为 V_S，资产值为 V_A，优先级为 V_P，以及可信度为 V_R。归一化处理的公式为

$$\left.\begin{aligned} T_S = \frac{V_S}{10}\times 100\%,\ T_A = \frac{V_A}{5}\times 100\% \\ T_P = V_P \times 100\%,\ T_R = V_R \times 100\% \end{aligned}\right\} \quad (9)$$

第二步：给定各个因素的权重因子 δ，一个攻击的综合威胁程度T'的计算式为

$$T' = \delta_1 P + \delta_2 S + \delta_3 R + \delta_4 A$$

这里，δ 可以由专家确定或者根据历史数据采用层次分析法等方法确定[14]。

第三步：参考 CVSS 对于漏洞评级的分类，将报警分为四类[15]，威胁程度和报警分类的对应关系如表 8 所示。

表 8　报警分类标准

Tab.8　Snort alert classification standard

报警分类	低	中	高	临界
报警分值	0.0～0.5	0.5～0.7	0.7～0.9	0.9～1.0

2.3　HMM-NIDS 算法

将本文所提的基于 NIDS 报警信息和改进 HMM 的威胁定量分析方法命名为 HMM-NIDS 算法，该算法描述如下：

步骤 1：根据电力信息系统拓扑结构，定义资产当前安全状态，并确定各安全状态概率密度。

步骤 2：优化 HMM 中的 $\boldsymbol{T}$ 和 $\boldsymbol{O}$ 的矩阵参数。

a. 获取 Snort 报警数据信息。

b. 根据报警信息对优先级、资产值和严重度进行量化分析。

c. 对动态威胁建立贝叶斯拓扑结构，并建立条件概率表，确定条件概率分布；利用工具获取节点的状态信息；通过贝叶斯网络分析计算得到攻击成功的概率，量化可信度评价指标。

d. 将上述四类因素进行融合，确定威胁值和报警类别。

e. 根据报警类别对 HMM 中原始 Trans 和 Obs 矩阵的参数进行优化，并设置初始概率和代价向量。

步骤 3：利用 HMM 对系统动态威胁量化分析，计算主机当前的风险值。

步骤 4：计算当前系统风险值。

3　实验及结果

为了验证本文所提方法的有效性，实验模拟某电力信息系统遭受到 DDoS 网络攻击，采用本文所提方法对其进行动态威胁定量风险分析。

3.1　实验场景描述

实验采用 Darpa2000 LLDOS1.0 攻击数据集[16]模拟 DDoS 攻击。Darpa2000 LLDOS1.0 的实验环境包含 4 个 C 类子网。整个数据集包含了由两台 Snort 采集得到的 190 min 的 NIDS 报警数据。一条 Snort 报警包含的信息包括：时间戳、报警类型编号、报警名称、攻击类型、优先级、协议、源地址和端口、目标地址和端口等。通常情况下，一个攻击可触发多条 NIDS 报警，一条报警表示攻击在该时间节点上的攻击特征。

图 5 给出了该次 DDoS 攻击的步骤，描述如下：

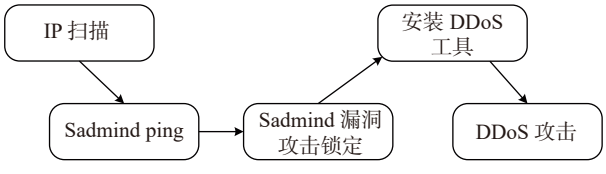

图 5　DARPA 实验场景攻击步骤

Fig.5　Scenario steps of DARPA experiment

第一阶段：IP 扫描。攻击入侵者发送 ICMP 请求监听 ICMP 应答，进行 IP 扫描是为了寻找网络中的活跃主机。

第二阶段：Sadmind Ping。通过探查发现的活跃主机，确定正在执行远端管理工具的主机，以此锁定了 Pascal，Mill 和 Locke3 台主机。

第三阶段：攻击 Pascal，Mill 和 Locke3 台主机。3 台主机被攻击者使用 Sadmind 漏洞攻击锁定。由于是一个远程缓冲区溢出攻击，攻击者需不断尝试入侵，直至成功。

第四阶段：3 台主机 Pascal，Mill 和 Locke 遭受了入侵攻击，成为傀儡机。攻击者在 Pascal，Mill 和 Locke 上安装 DDoS 工具，经由 RSH 服务登录，并装配攻击代理。

第五阶段：开始 DDoS 攻击。傀儡机上所有的

服务均被攻击者使用攻击代理控制，攻击者通过伪造 IP 地址对远程服务器发起 DDoS 攻击。

3.2　实验及结果

3.2.1 报警分类

对实验中每条 Snort 报警信息分析后可得到严重度、资产值、优先级和可信度量化值，根据经验设优先级、严重度、可信度和资产值的权重分别为 δ_1=0.30，δ_2=0.33，δ_3=0.21，δ_4=0.16，融合得到该条 Snort 报警信息的威胁值以及类别。受篇幅限制，表 9 仅列出部分 Snort 报警信息的威胁类别。

表 9　Snort DARPA 实验报警分类

Tab.9　Snort DARPA alert classification

Snort sid	步骤	报警名称	严重度	资产值	优先级	可信度	报警威胁	分类
408	1	ICMP Echo Reply	0.3	0.85	0.3	0.4	0.429	低
1957	2	RPC sadmind UDP PING	0.3	0.85	0.3	0.61	0.493	低
718	3	INFO TELNET login incorrect	0.3	0.85	0.6	0.52	0.539	中
1911	4	RPC sadmind UDP NETMGT_PROC_SERVICECLIENT_DOMAIN overflow attempt	0.3	0.85	0.3	0.66	0.508	中
247	5	DDoS mstream client to handler	0.7	0.85	0.9	0.69	0.787	高
.....								

3.2.2 观测矩阵优化

根据上述实验场景，在设置 HMM 参数时要综合考虑到以下几点因素：a. 转移矩阵 $\boldsymbol{T}$ 中，下一时刻，主机和网络维持原状态的可能性最大；b. 观测矩阵 $\boldsymbol{O}$ 中，当网络和主机处在危险状态时，更有可能观测到威胁度高的报警；c. 一旦主机进入“攻破（C）”状态，处于该状态的时长将比其他状态的更久；d. 为了便于实验比较，假设所有主机的参数相同。

设置初始概率为：$\boldsymbol{\pi}=(\pi_G,\pi_P,\pi_A,\pi_C)=(0.8,\ 0.1,\ 0.05,\ 0.05)$，代价向量设置为：$\boldsymbol{C}=(C_G, C_P, C_A, C_C)=(0.1, 0.4, 0.7, 1)$，假设最初有 100 种 NIDS 报警分类，则原始的转移矩阵为 4×100 的矩阵，观测矩阵为 4×4 的矩阵，受篇幅限制不在此列出。

采用本文改进的报警分类，由于优化了报警类型数量和一定程度上克服虚警干扰，计算出的风险值以及对应的安全状态概率也随之改变，转移矩阵 $\boldsymbol{T}$ 和观测矩阵 $\boldsymbol{O}$ 为

$$\boldsymbol{T}=\begin{bmatrix} pGG & pGP & pGA & pGC \\ pPG & pPP & pPA & pPC \\ pAG & pAP & pAA & pAC \\ pCG & pCP & pCA & pCC \end{bmatrix}= \begin{bmatrix} 0.6 & 0.3 & 0.09 & 0.01 \\ 0.3 & 0.4 & 0.25 & 0.05 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0.01 & 0.09 & 0.1 & 0.8 \end{bmatrix}$$

$$\boldsymbol{O}=\begin{bmatrix} qG(1) & qG(2) & qG(3) & qG(4) \\ qP(1) & qP(2) & qP(3) & qP(4) \\ qA(1) & qA(2) & qA(3) & qA(4) \\ qC(1) & qC(2) & qC(3) & qC(4) \end{bmatrix}= \begin{bmatrix} 0.01 & 0.14 & 0.15 & 0.7 \\ 0.05 & 0.25 & 0.25 & 0.45 \\ 0.05 & 0.2 & 0.35 & 0.4 \\ 0.05 & 0.2 & 0.25 & 0.4 \end{bmatrix}$$

式中：pGP 表示由 Good 状态转换为 Probed 状态的概率；$qG(1)$，$qG(2)$，$qG(3)$，$qG(4)$ 分别表示在 Good 状态下观察到报警类别为优先级、严重度、可信度和资产值的概率，以此类推。这里，各概率值根据式（1）和式（2）计算得到。

可见，若 NIDS 报警类型很多，则原始矩阵 $\boldsymbol{O}$ 规模很大，运算量大，本方法通过对 NIDS 报警信息归类，有效地缩小了矩阵 $\boldsymbol{O}$ 的规模，减少了算法的复杂度。

3.2.3 实时风险分析

为了进一步验证所提方法能够有效描述系统所受的威胁状态，对这段攻击场景，根据式（4）和式（5）计算系统风险值，对传统报警分类和改进分类方法分别计算其实时风险值，使用 Sigmaplot 进行作图，计算结果如图 6 所示，图中以 min 为单位并选取 130 min 的实验数据作为展示。

从图 6(a) 和图 6(b) 可以看到，采用 HMM 和 HMM-NIDS 方法可以还原出攻击场景。

第一阶（0~20 min）：IP 扫描。这段时间的威胁值并不高，图 6(a) 和图 6(b) 均显示存在 4 分以

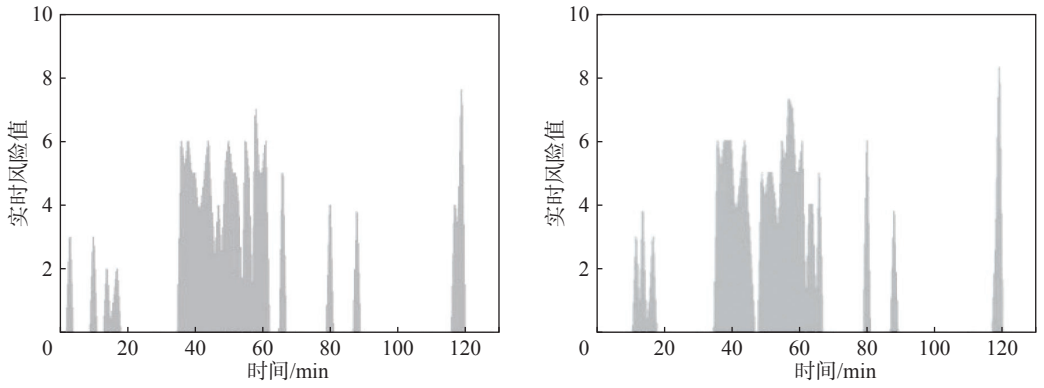

图 6 威胁实时分析图

Fig.6 Real-time threat analysis diagram

下威胁。

第二阶段(约 35～48 min)：尝试入侵。可以发现在这一阶段，两种方法表征的风险值都有了明显的提升。

第三阶段(约 48～68 min)：漏洞攻击。图 6(a)和图 6(b)都在约 58 min 左右出现较高风险值，HMM-NIDS 计算得到的最高风险值达到了 7.865 8。

第四阶段(约 68～118 min)：安装 DDoS 程序。该段时间 Snort 检测的网络风险值大多趋于零，这是因为 Snort 只能检测到网络攻击行为。

第五阶段(118～120 min)： DDoS 攻击。此时攻击者发起 DDoS 攻击，图 6(a)显示威胁值达到 7.962， 图 6(b)显示威胁值达到 8.216。

可见，采用 HMM 和 HMM-NIDS 分析得到的威胁值与网络安全态势变化状况均与实际情况相符合，说明采用 HMM 模型能够实现电力信息系统动态威胁定量风险分析。从图中可以看到，HMM-NIDS 对 1，2，3 阶段的刻画更显著和清晰。在第一阶段，大约 3 min 左右，图 6(a)中出现虚警；在第二阶段，48 min 左右，图 6(b)更清晰地刻画出了查探结束和发起攻击前的间歇；在第三阶段，在图 6(b)对攻击趋势(弱–强–弱)刻画与实际更加一致，更好地反映了连续性的攻击状态，而图 6(a)在 62 min 左右出现了风险为 0 的情况。

4 结 论

针对典型电力信息系统的网络威胁定量评估问题，提出了基于 NIDS 报警信息和 HMM 的网络威胁动态分析方法 HMM-NIDS。该方法充分利用系统中部署的网络入侵检测系统 NIDS 的报警信息，提出从优先级、严重度、资产值和可信度 4 个方面对当前威胁进行分类，简化隐马尔可夫模型中的观测矩阵，并给出了优先级、严重度、资产值和可信度的定量描述方法，特别是提出基于贝叶斯网络分析 NIDS 报警信息的可信度，防止了误报和虚警信息的干扰。最后，基于优化的 HMM 模型，融合计算得到系统的动态威胁量化值。基于 Darpa2000 数据集，模拟系统发生 DDoS 攻击时的状态，实验验证了本文方法的有效性和优越性。未来，可进一步结合脆弱性等其他因素，对电力信息系统做更全面的风险评估。

参考文献：

[1] 李欣, 段詠程. 基于改进隐马尔可夫模型的网络安全态势评估方法 [J]. 计算机科学, 2020, 47(7): 287–291.

[2] 梁智强, 林丹生. 基于电力系统的信息安全风险评估机制研究 [J]. 信息网络安全, 2017(4): 86–90.

[3] 马刚, 杜宇鸽, 安波, 等. 基于威胁传播采样的复杂信息系统风险评估 [J]. 计算机研究与发展, 2015, 52(7): 1642–1659.

[4] CIAPESSONI E, CIRIO D, MASSUCCO S, et al. Risk-Based dynamic security assessment for power system operation and operational planning[J]. Energies, 2017, 10(4): 475.

[5] XIONG J X, WU J Z. Construction of information network vulnerability threat assessment model for CPS risk assessment[J]. Computer Communications, 2020, 155: 197–204.

[6] 周未, 张宏, 李博涵. 基于攻防状态图模型的网络风险评估方法 [J]. 东南大学学报 (自然科学版), 2016, 46(4): 688–694.

[下转第 416 页]

[13] YADAV K S, SINGHA J. Facial expression recognition using modified Viola-John's algorithm and KNN classifier[J]. Multimedia Tools and Applications, 2020, 79(19): 13089–13107.
[14] XU Y T, ZHANG Y Q, ZHAO J, et al. KNN-based maximum margin and minimum volume hyper-sphere machine for imbalanced data classification[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(2): 357–368.
[15] 殷小舟. 一种改进的结合 K 近邻法的 SVM 分类算法 [J]. 中国图象图形学报, 2009, 14(11): 2299–2303.
[16] 李欢, 焦建民. 简化的粒子群优化快速 KNN 分类算法 [J]. 计算机工程与应用, 2008, 44(32): 57–59.
[17] 江涛, 陈小莉, 张玉芳, 等. 基于聚类算法的 KNN 文本分类算法研究 [J]. 计算机工程与应用, 2009, 45(7): 153–155,158.
[18] 朱军, 胡文波. 贝叶斯机器学习前沿进展综述 [J]. 计算机研究与发展, 2015, 52(1): 16–26.
[19] 姚勇,赵辉,刘志镜. 一种非线性支持向量机决策树多值分类器 [J]. 西安电子科技大学学报, 2007(6): 873–876.
[20] 刘秋阁, 何清, 史忠植. 一种新的非线性支持向量机分类算法 [C]//中国人工智能学会第 12 届全国学术年会. 哈尔滨: 中国人工智能学会, 2007.

（编辑：石 瑛）

[上接第 396 页]

[7] 杨至元, 张仕鹏, 孙浩, 等. 基于 Cyber-net 与学习算法的变电站网络威胁风险评估 [J]. 电力系统自动化, 2020, 44(24): 19–27.
[8] 杨英杰, 冷强, 常德显, 等. 基于属性攻击图的网络动态威胁分析技术研究 [J]. 电子与信息学报, 2019, 41(8): 1838–1846.
[9] 王辉, 张娟, 赵雅, 等. 一种新型贝叶斯模型的网络风险评估方法 [J]. 小型微型计算机系统, 2020, 41(9): 1898–1904.
[10] 张雪芹, 张立, 顾春华. 社交网络中社会工程学威胁定量评估 [J]. 浙江大学学报 (工学版), 2019, 53(5): 837–842.
[11] AL-KARAKI J N, GAWANMEH A, ALMALKAWI I T, et al. Probabilistic analysis of security attacks in cloud environment using hidden Markov models[J]. Transactions on Emerging Telecommunications Technologies, 2020: e3915.
[12] MEROUANE M. An approach for detecting and preventing DDoS attacks in campus[J]. Automatic Control and Computer Sciences, 2017, 51(1): 13–23.
[13] National Vulnerability Database[EB/OL]. (2000-07-01)[2022-05-18]. https://nvd.nist.gov/
[14] STEPHENSON T A. An introduction to Bayesian network theory and usage[R]. Switzerland: IDIAP, 2000: 00-03.
[15] ZHANG X Q, CHEN M J. CVDE based industrial system dynamic vulnerability assessment A. T. Balkema & G. Westers[C]//Proceedings of the 2014 International Conference on Network Security and Communication Engineering (NSCE 2014). Hong Kong, China: CRC Press, 2014: 66–73.
[16] Lincoln Laboratory[EB/OL]. (2022-05-12)[2022-05-19]. https://www.ll.mit.edu/r-d/datasets/2000-darpa-intrusion-detection-scenario-specific-datasets

（编辑：董 伟）