doi:10.3969/j.issn.1671-637X.2015.01.009

基于模糊—隐马尔可夫模型的复合式攻击预测方法

张艳雪[1a], 赵冬梅[1b], 刘金星[2]
(1.河北师范大学,a.数学与信息科学学院;b.信息技术学院,石家庄 050000;
2.空军第一航空学院航空军械系,河南 信阳 464000)

摘 要:通过对复合式攻击预测方法的研究,将关联规则、模糊评价法和隐马尔可夫模型相结合,提出了基于模糊—隐马尔可夫模型的复合式攻击预测方法。该方法首先将原始报警信息融合为超级报警信息,进而基于攻击行为的初始概率分布确定初始状态矩阵,根据关联规则确定状态转移矩阵,应用模糊判别法确定观察矩阵,最后应用隐马尔可夫模型中的Forward算法对报警信息隶属的攻击场景进行了识别,Viterbi算法对攻击意图序列进行了预测。仿真实验验证了该方法的有效性。

关键词:网络安全;模糊评价;隐马尔可夫模型;复合攻击;报警信息
中图分类号:V271.4;TP309 **文献标志码**:A **文章编号**:1671-637X(2015)01-0039-06

Approach to Forecasting Multi-Stage Attack Based on Fuzzy Hidden Markov Model

ZHANG Yan-xue[1a], ZHAO Dong-mei[1b], LIU Jin-xing[2]
(1. Hebei Normal University, a. College of Mathematics and Information Science; b. College of Information Technology, Shijiazhuang 050000, China; 2. The First Aeronautics College of PLAAF, Xinyang 464000, China)

Abstract: Through study on methods for forecasting multi-stage attack, we proposed a forecasting approach based on fuzzy, Hidden Markov Model(HMM) by integrating the association rule, fuzzy evaluation method and hidden Markov model together. Firstly, the original alarm information was fused into hyper alarm information. Secondly, the initial state matrix was obtained by the initial probability of the attack behaviors, the state transition matrix was determined according to the association rule, and the observation matrix was obtained by fuzzy evaluation. Finally, the attack scenarios leading to the alarm information were recognized by the Forward algorithm of HMM, and the next possible attack sequence was forecasted by the Viterbi algorithm of HMM. The results of simulation experiments verify the validity of this approach.

Key words: network security; fuzzy evaluation; hidden Markov model; multi-stage attack; alarm information

0 引言

目前,网络安全形势日趋复杂,复合式攻击逐渐成为网络攻击的主流。由中国国家互联网应急中心(CNCERT/CC)发布的2012年中国互联网网络安全报告可见,“蠕虫”、DDoS等复合式攻击占总体网络攻击行为中的60%。复合式攻击即攻击者利用攻击目标自身存在的安全漏洞,蓄意采用含有多个攻击步骤的攻击行为对攻击目标进行攻击,最终实现对攻击目标毁灭性打击[1]。在复合式攻击中,攻击步骤具有3个特点:1)攻击步骤之间存在着因果关系[2];2)攻击步骤具有时间上的顺序性[3];3)攻击步骤具有不确定性[3]。

早在1999年,HUANG M Y等人首次提出将攻击意图作为一个单独的因素,利用扩展的目标树对攻击意图进行建模,对攻击者后续攻击进行预测。自2005年以来,有一大批国内外学者对复合式攻击预测方法进行了大量的研究。

当前对复合式攻击方法的研究主要包括4类:1)

收稿日期:2014-01-17 **修回日期**:2014-10-15
基金项目:国家自然科学基金(60573036);河北省自然科学基金(F2013205193);河北省科技支撑计划(12213514D);河北师范大学研究生科研基金(201305)
作者简介:张艳雪(1989—),女,河北衡水人,硕士生,研究方向为网络信息安全。

基于攻击前提和攻击结果的复合式攻击预测方法[4]，该方法通过事件间的前驱后继关系，预测今后一段时间内，攻击者要实施的攻击行为，但是由于复合式攻击行为的多样性和复杂性，通过事件间的前驱后继关系进行预测难以实现；2）基于隐着色 Petri 网（HCPN）的复合式攻击预测方法[2,5]，该方法是对传统 Petri 网的改进，通过 HCPN 对复合式攻击场景建模，对报警信息进行关联，但是该方法的重点在于对攻击行为的检测[6]，而非预测；3）基于 Bayes 博弈的复合式攻击预测方法[7-8]，该方法能够理性预测出在下一个博弈阶段，攻击者选择攻击的概率和防御者选择防御的概率，但是该方法具有一定程度的主观性，目前的研究只建立了二人博弈模型，具有一定的局限性；4）基于攻击意图的复合式预测方法[6,9-10]，该方法应用扩展的有向图来描述攻击行为之间的逻辑关系，根据逻辑关系进行下一步的预测，其不足在于难以确定复合式攻击的匹配度。

为实现对复合式攻击行为的有效预测，本文提出了一种新的复合式攻击预测方法——基于模糊—隐马尔可夫模型的复合式攻击预测方法。该方法首先将原始报警信息融合为超级报警信息[11]，进而基于攻击行为概率分布确定初始状态矩阵，根据关联规则确定状态转移矩阵，应用模糊判别法确定观察矩阵，最后应用隐马尔可夫模型中的 Forward 算法对报警信息隶属的攻击场景进行识别，应用隐马尔可夫模型中的 Viterbi 算法对攻击意图序列进行预测，并通过仿真实验验证了该方法的有效性。

1 基于模糊—隐马尔可夫模型的复合式攻击预测模型

隐马尔可夫模型[12-13]（Hidden Markov Model，HMM）常用来处理与时间序列有关的问题，现已被广泛应用于语音识别、信号处理、生物信息学等领域，近年来也被应用于入侵检测领域。

一个完整的 HMM 模型包括以下 5 项元素（S，V，$\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{p}$）：S 为一组状态的集合；V 为一组输出符号组成的集合；$\boldsymbol{A}$ 为状态转移矩阵，是一个方阵；$\boldsymbol{B}$ 为观察矩阵，即输出符号的概率分布；$\boldsymbol{p}$ 为初始状态矩阵，即初始状态概率分布。

由于攻击者的攻击意图隐藏于系统中，因此无法由外界直接观察到，只能通过系统的表面特征推算系统的真实状态，也就是通过报警信息推算攻击意图。基于该特点，构建了一个模糊—隐马尔可夫模型，以实现对复合式攻击行为的识别和预测：首先将原始报警信息融合为超级报警信息；然后分别通过攻击行为概率分布、关联规则和模糊评价法来确定初始状态矩阵、状态转移矩阵和观察矩阵；最后通过 HMM 模型中的 Forward 算法对报警信息隶属的攻击场景进行识别；通过 HMM 模型中的 Viterbi 算法对攻击意图序列进行预测。该方法的流程如图 1 所示。

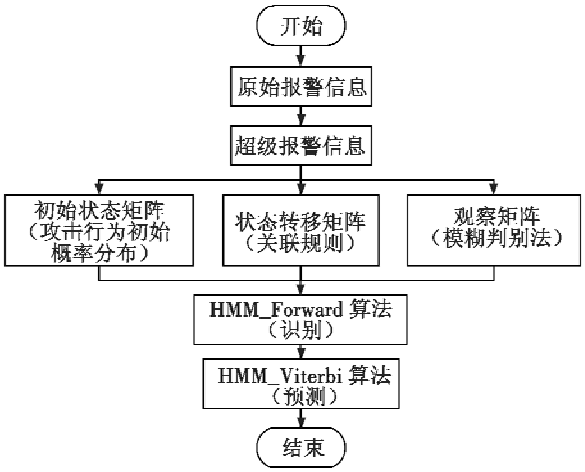

图 1 基于模糊—隐马尔可夫模型的复合式攻击预测方法流程图

Fig. 1 Flow chart for forecasting multi-stage attack based on fuzzy HMM

基于模糊—隐马尔可夫模型的复合式攻击行为识别、预测模型如图 2 所示。

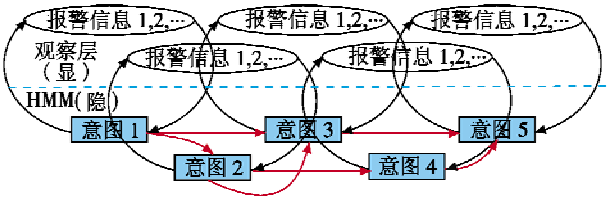

图 2 基于模糊—隐马尔可夫模型的复合式攻击识别、预测模型

Fig. 2 The model for recognizing and forecasting multi-stage attack based on fuzzy HMM

该模型的研究对象是一个数据序列（由报警信息属性值组成），该序列中的每个值被称为观察值。该模型认为：数据序列背后隐藏着另外一个序列，此序列便是一系列的状态值（由意图序列组成）。每个观测值都是在某个状态下发生的，该状态（隐）是不能直接观测到的，只能通过外部观测值（显）推断其特性。

2 相关技术问题的解决方案

2.1 原始报警信息的处理

由于原始报警信息数据具有很多属性，本文通过对原始报警信息特征事件关联的语义分析，将原始报警信息的格式统一定义为 RawAlert（RawAlert_ID，RawAlert_Type，Source_IP，Destination_IP，Start/End_Time），

将超级报警信息的格式统一定义为 SuperAlert(SuperAlert_ID, SuperAlert_Type, Source_IP, Destination_IP, Start/End_Time, Alert_Count)。

将原始报警信息融合为超级报警信息的规则定义如下所述。

1) 在原始报警信息中,除 RawAlert_ID 和 Start/End_Time 这 2 个属性值不同外,其他的 3 个属性值均相同,则认为是一个报警事件的多个重复报警信息,这类报警信息只留一条,其余的直接丢弃。

2) 将具有"Source_IP 属性值不同,Destination_IP 属性值相同"特点[5]的原始报警信息融合为一条超级报警信息。如表 1、表 2 所示,其中,m 为原始报警信息个数。

表 1　原始报警信息

Table 1　Raw alarm information

Alert_ID	Alert_Name	Source_IP	Destination_IP	Start/End_Time
1	ICMP Echo Reply	172.16.112.1	172.16.113.168	05:18/05:18
2	ICMP Echo Reply	172.16.112.2	172.16.113.168	05:18/05:18
⋮	⋮	⋮	⋮	⋮
m	ICMP Echo Reply	172.16.112.m	172.16.113.168	05:31/05:31

表 2　超级报警信息

Table 2　Hyper alarm information

SuperAlert_ID	SuperAlert_Name	Source_IP	Destination_IP	Start/End_Time	Alert_Count
001	ICMP Echo Reply	172.16.112.*	172.16.113.168	05:18/05:31	m

2.2　状态转移矩阵的确定——关联规则

通过对关联规则[9,14]现阶段的研究发现,使用关联规则挖掘技术有助于计算复合式攻击行为中攻击意图的状态转移矩阵,能有效反映出复合式攻击行为中攻击意图间的依赖关系。

设:1)复合式攻击X中的两个攻击意图分别为意图 i、意图 j;2)两个攻击意图之间的状态转移概率为 a_{ij}, $a_{ij}=p$(意图 $i\rightarrow$意图 j)。如果攻击意图项集中存在关联规则:意图 $i\rightarrow$意图 j,则称意图 i 和意图 j 之间存在着状态转移关系。a_{ij}的算式为

$$a_{ij}=\text{count}(\text{意图 } i\rightarrow\text{意图 } j)/\text{count}(\text{意图 } i) \quad (1)$$

式中:count(意图 $i\rightarrow$意图 j)表示意图 i 和意图 j 在攻击意图项集中同时出现并且意图 i 紧跟意图 j 的攻击意图的个数;count(意图 i)表示在整个攻击意图集中意图 i 出现的个数。举例说明如表 3 所示。根据式(1),计算结果,见矩阵 1(初始化)。

矩阵 1(初始化)

$$\begin{array}{c} \\ \{1\} \\ \{2\} \\ \{3\} \\ \{4\} \\ \{5\} \end{array}\begin{array}{c} \begin{array}{ccccc} \{1\} & \{2\} & \{3\} & \{4\} & \{5\} \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}。$$

表 3　结果

Table 3　Results

意图号	攻击意图项集
1	1 3 4
2	2 3 5
3	1 2 3 5
4	2 5

扫描 →

意图集	个数
{1}	2
{2}	3
{3}	3
{4}	1
{5}	3

2.3　观察矩阵的确定——模糊判别法

模糊判别法是一种基于模糊数学的综合评标方法,该方法根据模糊数学的隶属度理论,把定性评价转化为定量评价,即用模糊数学对受到多个因素制约的事物做出一个综合决策。模糊评价法能够有效地处理在评价过程中本身带有的主观性,以及客观所遇到的模糊性现象[15]。

设:1) $\boldsymbol{B}$ 为各报警信息在某准则下的相对权重;2) $\boldsymbol{A}$ 为权重分配集;3) $\boldsymbol{R}$ 为隶属度矩阵,即

$$\boldsymbol{R}=\begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.490 & 0.490 & 0.020 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.200 & 0.200 & 0.200 & 0.200 & 0.200 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.660 & 0.170 & 0.170 \end{bmatrix}。$$

其中,$\boldsymbol{B}=\boldsymbol{A}\circ\boldsymbol{R}=(a_1,a_2,\cdots,a_m)\circ\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{pmatrix}$,

$\circ$ 为模糊算子。本文采用 $M(\bullet,\oplus)$模糊算子计算,举例如下。

1) 因素集 $U=\{Alert_1,Alert_2,Alert_3,Alert_4,Alert_5,Alert_6,Alert_7,Alert_8,Alert_9,Alert_{10},Alert_{11},Alert_{12},Alert_{13}\}$。

2) 因素集 U 的评判集 $V=\{v_1,v_2,v_3,v_4,v_5\}$,其含义如表 4 所示。

由专家参照表 4 对因素集 U 进行概率评价,每位专家对各个因素确定其影响概率为 v_1,v_2,v_3,v_4,v_5 中的一种。结合各位专家的评定意见,计算每个报警信

息对攻击意图完成的影响程度,得到的隶属度矩阵 $\boldsymbol{R}$ 作为观察矩阵。对 v_1,v_2,v_3,v_4,v_5 的权重依次为 1/25,3/25,5/25,7/25,9/25。按照公式 $\boldsymbol{B}=\boldsymbol{A}\circ\boldsymbol{R}$ 得到的相对权重为[0.040,0.059,0.059,0.024,0.040,0.040,0.040,0.040,0.040,0.280,0.238,0.010,0.010],归一化处理后的向量为[0.043,0.065,0.065,0.027,0.043,0.043,0.043,0.043,0.043,0.304,0.259,0.011,0.011]。由上述结果可以看出:$Alert_{10}$,$Alert_{11}$对攻击意图的完成影响程度最大,今后在这类复合式攻击中,当检测到该类报警信息时,要加大防范力度。

表4 影响等级定义[16]

Table 4 Definition of impact levels

影响等级	描述	影响等级	描述
v_1	可忽略的	v_4	严重
v_2	微小	v_5	关键
v_3	一般		

2.4 基于 HMM 的复合网络攻击识别算法——Forward 算法

Forward 算法的步骤如下所述。

1) 初始化。

$$\alpha_1(i)=p_ib_i(o_1)\quad 1\leqslant i\leqslant N \tag{2}$$

式中,p_i 为在 $t=1$ 时的初始概率,且 $p_i=p(q_i=s_i)$。

2) 迭代计算。

$$\alpha_{t+1}(j)=\left[\sum_{i=1}^{N}\alpha_t(i)a_{ij}\right]b_j(o_{t+1})\quad 1\leqslant t\leqslant T-1,\quad 1\leqslant j\leqslant N \tag{3}$$

式中:a_{ij}为状态转移概率,且 $a_{ij}=p(q_{t+1}=s_j\mid q_t=s_i)$;$b_j(o_t)$为在观测值 o_t 状态下产生的概率值,$b_j(o_t)=p(o_t\mid q_t=s_i)$。

3) 终止条件。

$$p(O\mid\lambda)=\sum_{i=1}^{N}\alpha_T(i) \tag{4}$$

式中:λ 为给定的 HMM 模型;O 为观察序列,$O=\{O_1,O_2,\cdots,O_K\}$。

2.5 基于 HMM 的复合网络攻击预测算法——Viterbi 算法

Viterbi 算法的步骤如下所述。

1) 初始化,即

$$\delta_1(i)=\pi_ib_i(o_1)\quad 1\leqslant i\leqslant N,\psi_1(i)=0\ 。 \tag{5}$$

2) 迭代计算,即

$$\begin{cases}\delta_t(j)=\max(\delta_{t-1}(i)a_{ij})b_j(o_t)\quad 1\leqslant i\leqslant N\\ \psi_t(j)=\arg\max(\delta_{t-1}(i)a_{ij})\end{cases}。 \tag{6}$$

3) 终止条件为

$$\begin{cases}p^*=\max(\delta_T(i))\\ q_T^*=\arg\max(\delta_T(i))\end{cases}。 \tag{7}$$

4) 求解最佳路径,即

$$q_t^*=\psi_{t+1}(q_{t+1}^*)\quad t=T-1,T-2,\cdots,1\ 。 \tag{8}$$

3 仿真实验与分析

本文用于实验的数据集为 DAPRA(Defense Advanced Research Projects Agency)于2000年提供的攻击场景测试数据集 LLDOS1.0(inside),从中提取出两种复合式攻击行为,分别为 DDoS 和 FTP Bounce。根据1节,本文建立了2个模糊—隐马尔可夫模型,分别为 DDoS_HMM 和 FTP Bounce_HMM,模糊—隐马尔可夫模型参数如表5、表6所示。

1)基于 DDoS 的隐马尔可夫模型的攻击意图、报警信息如表5所示。

表5 DDoS_HMM

Table 5 DDoS_HMM

序号	攻击意图	报警信息编号	报警信息
1	IPSweep	$Alert_1$	ICMP Echo Reply
2	Sadmind Ping	$Alert_2$	PRC portmap sadmind request UDP
		$Alert_3$	RPC portmap Solaris sadmind port query udp request
		$Alert_4$	RPC sadmind UDP Ping
3	Sadmind Exploit	$Alert_5$	RPC portmap Solaris sadmind port query udp request
		$Alert_6$	RPC port sadmind request UDP
		$Alert_7$	RPC sadmind UDP
			RPC sadmind UDP
		$Alert_8$	NETMGT_PROC_SERVICE CLIENT_DOMAIN Overflow attempt
		$Alert_9$	RPC portmap Solaris sadmind port query udp portmapper sadmind port query attempt
4	InstallDDoS Tools	$Alert_{10}$	RSERVICES rsh root
5	LaunchDDoS Attack	$Alert_{11}$	SNMP AgentX/tcp request
		$Alert_{12}$	SNMP trap tcp
		$Alert_{13}$	SNMP request tcp

表6 FTP Bounce_HMM

Table 6 FTP Bounce_HMM

序号	攻击意图	报警信息编号	报警信息
1	IPSweep	$Alert_1$	ICMP Echo Reply
		$Alert_2$	ICMP Ping NMAP
2	PortScan	$Alert_3$	Scan NMAP TCP
		$Alert_4$	Scan synscan port scan
3	FTPExploit	$Alert_5$	FTP anonymous login attempt
		$Alert_6$	FTP anonymous ftp login attempt
		$Alert_7$	FTP forward
4	RhostModify	$Alert_8$	FTP rhosts
5	LaunchFTP BounceAttack	$Alert_9$	RSERVICES rsh root
		$Alert_{10}$	RSERVICES rlogin root

DDoS_HMM 初始状态矩阵 $\boldsymbol{p}$,状态转移矩阵 $\boldsymbol{A}$,观察矩阵见矩阵2~4。

矩阵2(初始状态矩阵)

$$\begin{matrix} State_1 & State_2 & State_3 & State_4 & State_5 \end{matrix}$$
$$[0.250 \quad 0.750 \quad 0.000 \quad 0.000 \quad 0.000]\ ;$$

矩阵 3(状态转移矩阵)

$$\begin{matrix} & State_1 & State_2 & State_3 & State_4 & State_5 \\ State_1 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ State_2 & 0.000 & 0.177 & 0.823 & 0.000 & 0.000 \\ State_3 & 0.000 & 0.228 & 0.688 & 0.028 & 0.056 \\ State_4 & 0.000 & 0.000 & 0.000 & 0.750 & 0.250 \\ State_5 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{matrix}\ ;$$

矩阵 4(观察矩阵)

$$\begin{matrix} & Alert_1 & Alert_2 & Alert_3 & Alert_4 & Alert_5 & Alert_6 & Alert_7 & Alert_8 & Alert_9 & Alert_{10} & Alert_{11} & Alert_{12} & Alert_{13} \\ State_1 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ State_2 & 0.000 & 0.490 & 0.490 & 0.020 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ State_3 & 0.000 & 0.000 & 0.000 & 0.000 & 0.200 & 0.200 & 0.200 & 0.200 & 0.200 & 0.000 & 0.000 & 0.000 & 0.000 \\ State_4 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ State_5 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.660 & 0.170 & 0.170 \end{matrix}\ 。$$

2) 基于 FTP Bounce 的隐马尔可夫模型的攻击意图、报警信息如表 6 所示。

FTP Bounce_HMM 初始状态矩阵 $\boldsymbol{p}$,状态转移概率矩阵 $\boldsymbol{A}$,报警信息产生概率 $\boldsymbol{B}$ 见矩阵 5 ~ 7。

矩阵 5(初始状态矩阵)

$$\begin{matrix} State_1 & State_2 & State_3 & State_4 & State_5 \end{matrix}$$
$$[0.667 \quad 0.333 \quad 0.000 \quad 0.000 \quad 0.000]\ ;$$

矩阵 6(状态转移矩阵)

$$\begin{matrix} & State_1 & State_2 & State_3 & State_4 & State_5 \\ State_1 & 0.600 & 0.400 & 0.000 & 0.000 & 0.000 \\ State_2 & 0.000 & 0.823 & 0.177 & 0.000 & 0.000 \\ State_3 & 0.000 & 0.000 & 0.625 & 0.375 & 0.000 \\ State_4 & 0.000 & 0.000 & 0.000 & 0.750 & 0.250 \\ State_5 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{matrix}\ ;$$

矩阵 7(观察矩阵)

$$\begin{matrix} & Alert_1 & Alert_2 & Alert_3 & Alert_4 & Alert_5 & Alert_6 & Alert_7 & Alert_8 & Alert_9 & Alert_{10} \\ State_1 & 0.250 & 0.750 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ State_2 & 0.000 & 0.000 & 0.118 & 0.882 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ State_3 & 0.000 & 0.000 & 0.000 & 0.000 & 0.625 & 0.375 & 0.000 & 0.000 & 0.000 & 0.000 \\ State_4 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 \\ State_5 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.833 & 0.167 \end{matrix}\ 。$$

当收到报警信息“ICMP Echo Reply”和“RPC portmap Solaris sadmind port query udp request”时,根据隐马尔可夫模型的 Forward 算法分别求出基于两个复合式攻击的隐马尔可夫模型产生该报警信息的概率 $\boldsymbol{p}$(Alerts | DDoS_HMM) = 0.122 5和 $\boldsymbol{p}$(Alerts | FTP Bounce_HMM) = 0.007 9。由这两个概率值可以看出 $\boldsymbol{p}$(Alerts | DDoS_HMM) > $\boldsymbol{p}$(Alerts | FTP Bounce_HMM),说明现在正在进行的攻击行为可能是 DDoS 复合式攻击。

由隐马尔可夫模型的 Forward 算法,将正在进行的攻击行为识别为:DDoS,当控制台收到的报警信息序列为{$Alert_1$, $Alert_3$, $Alert_4$}时,可根据隐马尔可夫模型中的 Viterbi 算法求出现在已经完成的攻击意图序列为[1,2],即现在完成的是前 2 个攻击意图,IPSweep 和 SadmindPing,下面要进行的攻击意图为 SadmindExploit。

当前 4 类主流的复合式攻击预测方法,即基于攻击前提和攻击结果的复合式攻击预测方法(A),基于 HCPN 的复合式攻击预测方法(B),基于 Bayes 博弈的复合式攻击预测方法(C)和基于攻击意图的复合式攻击预测方法(D),与本文提出的基于模糊—隐马尔可夫模型的复合式攻击预测方法(E)的性能对比如表 7 所示。

表 7　不同方法性能对比

Table 7　Performance contrast of different methods

预测方法	先验知识	在线关联	多元报警信息关联	预测新场景	客观性
A	√	√	×	√	√
B	√	√	√	×	√
C	√	√	√	√	×
D	√	√	×	√	√
E	×	√	√	√	√

可以明显看出:在当前复合式攻击预测方法中,本文提出的基于模糊—隐马尔可夫模型的复合式攻击预测方法在报警信息预处理、报警信息关联、对复合式攻击的识别和预测等方面的综合性能最优。

4　结束语

针对当前复合式攻击预测方法存在的对攻击行为难预测和对复合式攻击匹配度难确定的问题,本文通过对现有的复合式攻击预测方法的研究,结合了关联规则、模糊评价法和隐马尔可夫模型,提出了基于模糊—隐马尔可夫模型的复合式攻击预测方法。通过仿真实验,实现了对复合式攻击行为的识别和预测,证明了该方法的有效性。

参 考 文 献

[1] 王莉. 网络多步攻击识别方法研究[D]. 武汉:华中科技大学,2007. (WANG L. Study on method of network multi-stage attack plan recognition[D]. Wuhan:Huazhong University of Science and Technology, 2007.)

[2] 严芬,黄皓,殷新春. 基于 CTPN 的复合攻击检测方法研究[J]. 计算机学报,2006,29(8):1383-1391. (YAN F, HUANG H, YIN X C. A detection algorithm for mulit-step attack based on CTPN[J]. Chinese Journal of Computers, 2006, 29(8):1383-1391.)

[3] 袁晨. 基于 GCT 的多步攻击检测方法研究[D]. 长春:吉林大学,2010. (YUAN C. Research on multi-step attack detection method based on GCT[D]. Changchun:Jilin University, 2010.)

[4] 王祖俪,程小平. 入侵响应中基于事件相关性的攻击预测算法[J]. 计算机科学,2005,32(4):144-146. (WANG Z L, CHENG X P. An attack predictive algorithm based on the correlation of intrusion alerts in intrusion response[J]. Computer Science, 2005, 32(4):144-146.)

[5] WU R Y, LI W G, HUANG H. An attack modeling on hierarchical colored Petri nets[C]//International Conference on Computer and Electrical Engineering, 2008:918-921.

[6] 陈灿,阎保平. 针对复合攻击的网络攻击预测算法[J]. 计算机工程,2011,37(5):172-174. (CHEN C, YAN B P. Network attack forecast algorithm for multi-step attack[J]. Computer Engineering, 2011, 37(5):172-174.)

[7] 曹晖,王青青,马义忠,等. 基于动态贝叶斯博弈的攻击预测模型[J]. 计算机应用,2007,27(6):1545-1547. (CAO H, WANG Q Q, MA Y Z, et al. Attack predition model based on dynamic Bayesian games[J]. Journal of Computer Applications, 2007, 27(6):1545-1547.)

[8] 曹晖,王青青,马义忠,等. 基于静态贝叶斯博弈的攻击预测模型[J]. 计算机应用研究,2007,24(10):122-124. (CAO H, WANG Q Q, MA Y Z, et al. Attack predition model based on static Bayesian game[J]. Application Research of Computers, 2007, 24(10):122-124.)

[9] 张松红. 基于隐马尔可夫模型的网络安全预警技术研究[D]. 郑州:解放军信息工程大学,2007. (ZHANG S H. Research on network security early warning technology based on hidden Markov model[D]. Zhengzhou: PLA Information Engineering University, 2007.)

[10] 鲍旭华,戴英侠,冯萍慧,等. 基于入侵意图的复合攻击检测和预测算法[J]. 软件学报,2005,16(12):2132-2138. (BAO X H, DAI Y X, FENG P H, et al. A detection and forecast algorithm for multi-step attack based on intrusion intention[J]. Journal of Software, 2005, 16(12):2132-2138.)

[11] 翟光群,周双银. 多步攻击告警关联模型构建与实现[J]. 计算机应用,2011,31(5):1276-1279. (ZHAI G Q, ZHOU S Y. Construction and implementation of multi-step attacks alert correlation model[J]. Journal of Computer Applications, 2011, 31(5):1276-1279.)

[12] ALSERHANI F, AKHLAQ M, AWAN I U, et al. MARS: Multi-stage attack recognition system[C]//The 24th IEEE International Conference on Advanced Information Networking and Applications, 2010:753-759.

[13] LEE D H, KIM D Y, JUNG J I. Multi-stage intrusion detection using hidden Markov algorithm[C]//International Conference on Information Science and Security, 2008:72-77.

[14] 张松红,王亚弟,韩继红. 基于隐马尔可夫模型的复合攻击预测方法[J]. 计算机工程,2008,34(6):131-133. (ZHANG S H, WANG Y D, HAN J H. Approach to forecasting multi-step attack based on HMM[J]. Computer Engineering, 2008, 34(6):131-133.)

[15] 赵冬梅,刘金星,马建峰. 基于模糊小波神经网络的信息安全风险评估[J]. 华中科技大学学报:自然科学版,2009,37(11):43-45. (ZHAO D M, LIU J X, MA J F. Risk assessment of information security using fuzzy wavelet neural network[J]. J. Huazhong Univ. of Sci. & Tech: Natural Science Edition, 2009, 37(11):43-45.)

[16] 赵冬梅. 信息安全风险评估量化方法研究[D]. 西安:西安电子科技大学,2007. (ZHAO D M. Study on the risk assessment quantitative method of information security[D]. Xi'an: Xidian University, 2007.)

(上接第33页)

trend of the airborne radar warning receiver[J]. Electronic Information Warfare Technology, 2008, 23(3):51-54.)

[9] 王强,张安,陈超. 机载有源相控阵雷达发现概率改进算法[J]. 火力与指挥控制,2011,36(11):57-59. (WANG Q, ZHANG A, CHEN C. One improved algorithm of active airborne phased array radar detection probability[J]. Fire Control & Command Control, 2011, 36(11):57-59.)

[10] 蓝伟华,陈晓风. 单机多目标攻击的目标威胁排序[J]. 电光与控制,2006,13(5):16-17,30. (LAN W H, CHEN X F. Target threat sequencing in single-aircraft multi-target combat[J]. Electronics Optics & Control, 2006, 13(5):16-17, 30.)

[11] 张涛,于雷,周中良,等. 基于人工势场启发粒子群算法的空战机动决策[J]. 电光与控制,2013,20(1):77-82. (ZHANG T, YU L, ZHOU Z L, et al. Decision-making of air combat maneuvering based on APF and PSO[J]. Electronics Optics & Control, 2013, 20(1):77-82.)