网络取证隐马尔可夫模型证据融合方法

杨 珺[1]，马秦生[1]，王 敏[2]，曹 阳[1]

(1. 武汉大学电子信息学院 武汉 430079; 2. 通信指挥学院二系 武汉 430010)

【摘要】针对网络取证因果关联证据融合方法存在的算法复杂、重现场景不够精确等问题，提出了基于隐马尔科夫模型的网络取证证据融合方法，阐述了应用隐马尔科夫模型进行证据融合的可行性。该方法以元证据序列作为随机观察序列，以网络入侵步骤作为随机状态序列，通过对元证据序列进行解码操作，找寻最可能的网络入侵步骤并据此回溯证据链。实验结果表明，与基于贝叶斯网络的多源证据融合方法相比，该方法的算法复杂度和抵御干扰项的能力均得到了明显的改善，该方法能够以较小的代价较精确地重现网络入侵的犯罪现场。

关 键 词 计算机取证； 数据融合； 隐马尔可夫模型； 网络安全

中图分类号 TN915.08 **文献标志码** A doi:10.3969/j.issn.1001-0548.2013.03.006

Evidence Fusion of the Network Forensics on the Hidden Markov Models

YANG Jun[1], MA Qin-sheng[1], WANG Min[2], and CAO Yang[1]

(1. School of Electronic Information, Wuhan University Wuhan 430079;

2. Second Department, Commanding Communications Academy Wuhan 430010)

Abstract To improve the algorithm complexity and the accuracy of reproduced scene, a new method for the evidence fusion of the network forensics on the hidden Markov models (HMM) is proposed. The feasibility of this method is expounded. By taking the sequence of the meta-evidence as the random observation sequence, and the network intrusion step as the random state sequence, the most likely network intrusion step is inferred by the decoding operation aimed at the sequence of the meta-evidence and the chain of the evidence is backtracked accordingly. When they are applied in the same problem, the algorithm complexity and the anti-interference ability of the proposed method are dramatically modified compared with the method of Bayesian network. Therefore, the proposed method has a good ability in the cost to reproduce the scene of the crime.

Key words computer forensics; data fusion; hidden Markov models; network security

在计算机网络取证领域，取证分析过程中的证据分析阶段主要用于确定疑似证据可能分布的范围，而证据融合阶段主要用于提取多个疑似证据间的相关性，以形成证据链、重现犯罪现场[1]，因此，在网络取证分析的证据融合阶段，对疑似证据进行关联性分析是十分重要、也是十分必要的工作。

目前，主流的网络取证证据融合方法有：1) 图示关联法，该方法认为一次入侵过程可以用图示方法再现。优点是入侵场景可视化，缺点则是证据间的关联性较弱，如文献[2-3]提出的将时序聚类结果表示为证据图的多源证据融合方法；2) 时间序列关联法，该方法认为组成一次入侵过程的各个单步入侵行为可能以较高的概率发生在同一个时间窗口内。优点是可以发现新的入侵过程，缺点同样是证据间的关联性较弱，如文献[4-5]提出的基于序列模式挖掘的多源证据融合方法；3) 因果关联法，该方法认为一次入侵过程由多个单步入侵行为组成，这些入侵行为构成多个入侵步骤，而入侵步骤间具有前因和后果的关系，优点是证据间有较强的关联性，缺点则是证据融合算法复杂、重现场景不够精确，如文献[6-7]提出的基于贝叶斯网络的多源证据融合方法，该方法也是目前国内外最具代表性的。由于证据间具有较强的关联性，因此在向法庭提交取证结论时，因果关联法获取的结果更加易于被理解和采纳。

隐马尔可夫模型(HMM)是一个双重的随机过

收稿日期： 2011 – 08 – 28；修回日期：2012 – 06 – 21
基金项目：高等学校博士学科点专项科研基金(20040486049)
作者简介：杨珺(1973 –)，女，博士，主要从事信息安全和SoC高层次设计方面的研究.

程，它包含了一个随机观察序列和一个隐藏的具有因果关系的状态转移马尔可夫链。在已知模型参数和随机观察序列的情况下，最可能的状态转移链可采用简洁的解码算法精确地求解 [8]。在网络取证证据融合中，为了提高信息的处理效率，疑似证据将被聚合为元证据 [9]，若能将元证据序列作为 HMM 的随机观察序列，将具有因果关系的网络入侵步骤作为 HMM 的状态转移链，那么在已知模型参数和元证据序列的情况下，便可由解码算法精确地求出入侵步骤，并能够据此回溯出证据链。这样，HMM便能够以较小的代价很好地重现网络入侵的犯罪现场。

本文将基于因果关联法探用HMM对网络取证证据融合方法进行研究。

1　HMM证据融合原理

HMM善于处理序列型数据，它是一种描述随机观察序列和随机状态序列间多对多对应关系的概率模型。在HMM中，状态是隐藏的，但它可以通过观察序列推断出来，即对于给定的HMM和观察序列，HMM的解码算法可以求出最有可能产生该观察序列的状态序列[8]。

网络入侵过程往往分为多个入侵步骤，每个入侵步骤的实施仅仅依赖于上一个入侵步骤实施的结果。每个入侵步骤可以通过多种入侵行为实现，而某些入侵行为亦可能在多个入侵步骤中出现，即入侵行为和入侵步骤间是多对多的对应关系。某个时刻使用哪种入侵行为、处于哪个入侵步骤是无法确定的，即入侵行为和入侵步骤均是随机的。因此，HMM能够很好地对网络入侵过程进行建模。在网络取证分析中，入侵行为即为疑似证据，疑似证据可以通过证据分析得到，而统一格式并经聚合后的疑似证据序列构成了元证据序列[9]。因此，元证据序列可以作为网络入侵过程HMM的随机观察序列，而入侵步骤则可以作为网络入侵过程HMM的随机状态序列。当模型参数和元证据序列给定时，HMM的解码算法便可以推断出最可能的产生该元证据序列的入侵步骤。由此，HMM能够很好地描述元证据序列和隐藏的具有因果关系的网络入侵步骤间的关系，而入侵步骤回溯的结果便是证据链，如图1所示。

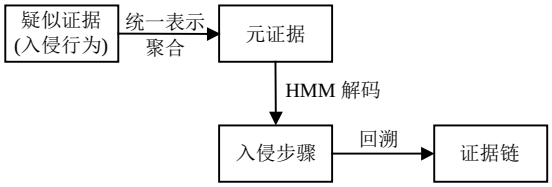

图1　HMM证据融合原理

设作为观察值的元证据的有限集为 $V=\{V_1,V_2,\cdots,V_M\}$，其中M为有限集的规模；元证据序列为$O=\{o_1,o_2,\cdots,o_t,\cdots,o_T\}$，其中$o_t$为$t$时刻的元证据；作为状态的入侵步骤的有限集为 $S=\{S_1,S_2,\cdots,S_N\}$，其中N为有限集的规模；入侵步骤为$Q=\{q_1,q_2,\cdots,q_t,\cdots,q_T\}$，其中$q_t$为$t$时刻的入侵步骤。则在网络入侵过程HMM中，初始状态分布为 $\pi=\{\pi_i\}$，$\pi_i=P(q_1=S_i)$；状态转移概率为 $A=\{a_{ij}\}$，$a_{ij}=P(q_{t+1}=S_j\mid q_t=S_i)$；观察值生成概率为 $B=\{b_{jk}\}$，$b_{jk}=P(o_t=V_k\mid q_t=S_j)$ $(i,j=1,2,\cdots,N;\ k=1,2,\cdots,M)$。

若已知 $O=\{o_1,o_2,\cdots,o_t,\cdots,o_T\}$ 和模型参数 $\lambda=(\pi,A,B)$，则入侵步骤即为当

$$p(O\mid S)P(S)=p(o_1\mid S_1)P(S_1)\prod_{2}^{T}p(o_t\mid S_t)P(S_t\mid S_{t-1})$$

取最大值时的$Q=\{q_1,q_2,\cdots,q_t,\cdots,q_T\}$。

2　证据链构造算法

证据链构造算法由HMM解码算法和证据回溯算法组成，HMM解码算法用于寻找产生已知元证据序列的最可能的入侵步骤，而证据回溯算法则用于还原疑似证据序列。

典型的HMM解码算法是Viterbi算法，该算法采用递归法寻求已知观察值序列最可能的隐状态序列，它的算法复杂度为$O(N^2T)$[8]，其核心步骤为：

$$\delta_t(j)=\max_{1\leqslant i\leqslant N}[\delta_{t-1}(i)a_{ij}b_j(o_t)]\quad 2\leqslant t\leqslant T,\ 1\leqslant j\leqslant N \tag{1}$$

$$\varphi_t(j)=\arg\max_{1\leqslant i\leqslant N}[\delta_{t-1}(i)a_{ij}]\quad 2\leqslant t\leqslant T,\ 1\leqslant j\leqslant N \tag{2}$$

式中，变量$\delta_t(j)$表示截止到t时刻，状态序列$Q=\{q_1,q_2,\cdots,q_t=S_j\}$产生出观察值序列$O=\{o_1,o_2,\cdots,o_t\}$的最高概率；数组$\varphi_t(j)$表示在$t$时刻，使$\delta_t(j)$取得最大值的$t-1$时刻的状态号$i$。由式(1)和式(2)可以看出：当模型参数$\lambda=(\pi,A,B)$精确、元证据序列$O=\{o_1,o_2,\cdots,o_t,\cdots,o_T\}$纯净时，Viterbi算法可以精确地递推出入侵步骤$Q=\{q_1,q_2,\cdots,q_t,\cdots,q_T\}$。

然而，在疑似证据中往往包含一些干扰项，这些干扰项可能是证据分析过程的误报，也可能是其他的入侵行为，甚至可能是入侵者为掩盖入侵目的故意添加的行为。这些含有干扰项的疑似证据被聚合为元证据序列后会导致元证据序列中亦含有干扰

项。而Viterbi算法主要适用于纯净的观察值序列，它严格地按照前一个观察值的状态来解析后一个观察值的状态。尽管干扰项出现的概率很低且与前后项的关联不甚紧密，可是干扰项的存在依然可能使Viterbi算法导出的状态序列出现链偏离、甚至链断裂的问题。因此，有必要使Viterbi算法具备区分并处理观察值序列中干扰项的能力。

定义关联阈值Th、关联状态 $r=\{r_i\}$ $(i=1,2,\cdots,t,\cdots,T)$，当 $a_{ij}\cdot b_j(o_t)<\text{Th}$ 时，o_t 为干扰项，其对应的状态不作为隐状态序列的有效部分，相应的 $r_i=0$；否则，为非干扰项，相应的 $r_i=1$。

在HMM解码算法(改进后的Viterbi算法)删除干扰项、递推出入侵步骤后，证据回溯算法用同时刻的元证据分别对应替换入侵步骤，再用疑似证据分别对应替换元证据序列，即可回溯出证据链。

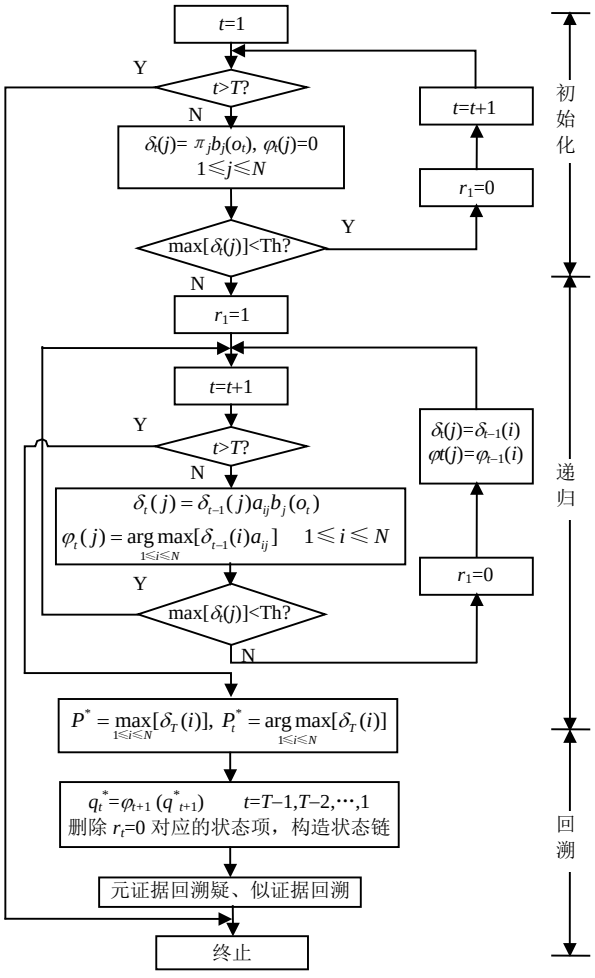

图2 证据链构造算法流程图

证据链构造的算法流程如图2所示。若 $\max[\delta_1(j)]$ 不满足关联阈值条件，则 o_1 的观察值对应的状态不作为状态链的初始状态，依此类推。若无 $\max[\delta_t(j)]$ 满足关联阈值条件，则相应的观察值序列无法构造状态链。若 $\max[\delta_t(j)]$ 不满足关联阈值条件，则将 $t-1$ 时刻观察值序列的最高概率作为 t 时刻观察值序列的最高概率。状态链由 $r_i=1$ 对应的状态项构成。疑似证据回溯依据聚合时记录在其属性项中的元证据的标记进行。在Viterbi算法中引入的关联状态判断功能并未增加算法的循环量，因此证据链构造算法的算法复杂度依然为 $O(N^2T)$。

3 实验结果及分析

实验数据选自DARPA2000下的LLDOS1.0内网数据集[10]，该数据集是由林肯实验室(Lincoln Laboratory)构造的入侵场景关联评测数据集，其攻击场景分为IPSweep、SadmindPing、SadmindExploit、InstallDDoSTool、DDoS共5个步骤。

采用snort入侵检测系统对LLDOS1.0数据集进行检测，得到疑似证据。采用改进的Leader-Follower算法[8]对疑似证据进行聚合，得到元证据。采用HMM解码算法对元证据序列进行关联分析，算法中的主要参数设置为：关联阈值Th=0.01；初始概率分布 π= [0.4 0.6 0 0 0]；状态转移概率 A=

$$\begin{bmatrix} 0.1 & 0.9 & 0 & 0 & 0 \\ 0 & 0.1 & 0.85 & 0.05 & 0 \\ 0 & 0.1 & 0.2 & 0.7 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

；观察值生成概率

$$B=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0.65 & 0.1 & 0.05 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0.35 & 0.55 \end{bmatrix}$$

。

关联分析后**可以**得到HMM状态转移链，如图3和表1所示。

图3 LLDOS1.0 HMM状态转移链

表1 节点与元证据的对应关系

节点	元证据
*N*1	ip sweep
*N*2	sadmind port scan
*N*3	sadmind exploit
*N*4	rsh
*N*5	ddos

采用证据回溯算法对HMM状态转移链进行回溯，得到LLDOS1.0的证据链，如图4所示。

分析表及图的信息可以得到：

1) HMM解码算法能够准确地描述入侵步骤。

图3描述的HMM状态转移链反映了在LLDOS1.0内网数据集中具有因果关系的DDoS攻击步骤，依次为地址扫描(ip sweep)、端口扫描(sadmind port scan)、权限提升(sadmind exploit)、DDoS安装(rsh)和DDoS启动(ddos) 5步，它们很好地吻合了MIT林肯实验室公布的LLDOS1.0内网数据集的攻击步骤。

2) HMM可以精确地对复杂入侵过程建模。

文献[6-7]采用贝叶斯网络对LLDOS1.0内网数据集建立的模型如图5和表2所示。尽管贝叶斯网络模型重构了LLDOS1.0内网数据集的攻击过程，但是图中的*N*6(Email_Almail_Overflow)并非是一个攻击步骤，而是一个噪音节点。HMM构造的LLDOS1.0内网数据集的攻击过程能够区分观察值序列中的干扰因素，精确地重现入侵过程。

证据链构造算法的复杂度是多项式级的复杂度，而贝叶斯网络推理却是NP-难问题[8]。因此，贝叶斯网络通常采用近似推理的方法，这也是文献[6-7]在状态链中产生噪音节点的主要因素。

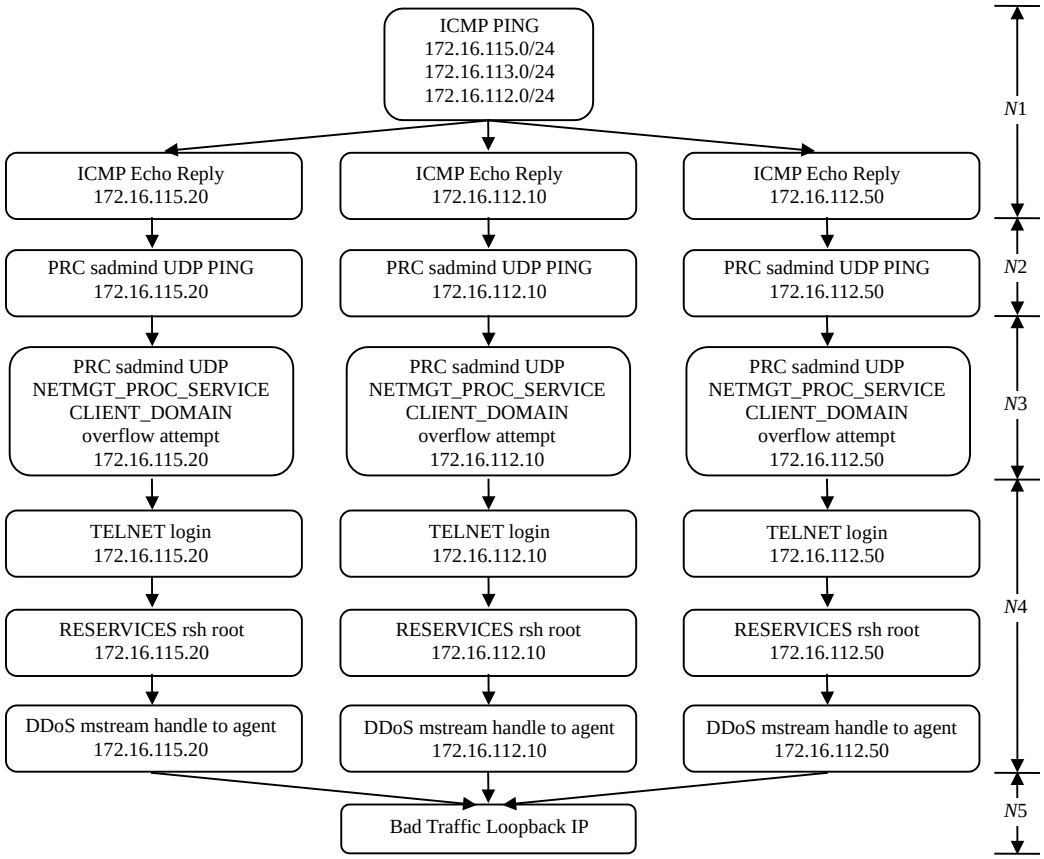

图4　LLDOS1.0证据链

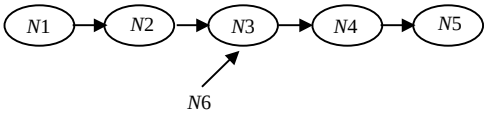

图5　LLDOS1.0元报警贝叶斯网络

表2　节点与元报警的对应关系

节点	元报警
*N*1	Sadmind_Ping
*N*2	Sadmind_Amslverify_Overflow
*N*3	Rsh
*N*4	Mstream_Zombie
*N*5	Strem_Dos
*N*6	Email_Almail_Overflow

3) 证据回溯算法能够还原入侵场景。

图4描述了针对地址172.16.115.20、172.16.112.10和172.16.112.50共3个受害主机的证据链，再现了202.077.162.213攻击主机的犯罪场景，该场景与MIT林肯实验室描述的入侵场景[10]基本相同。

4) 证据链构造算法具有较高的运行效率。

基于HMM的证据链构造算法的时间复杂度为$O(N^2T)$，而基于贝叶斯网络的多源证据融合算法的时间复杂度为$O(N^4T)$[6-7]，因此前者的执行效率更高。

4　结　论

证据融合技术是从疑似证据中找寻计算机犯罪证据链的技术，是计算机取证分析技术中最重要的

环节之一。本文基于因果关联法，论证了HMM应用于网络取证证据融合的可行性，提出了网络取证HMM证据融合方法。该方法充分地利用HMM善于描述具有多对多对应关系随机序列的特点及善于构建具有因果关系状态转移链的特点，实现了对网络入侵证据的关联分析。仿真实验表明，该方法对干扰项具有较强的处理能力，能够以较小的代价较精确地重现网络入侵的犯罪现场。

参 考 文 献

[1] 杨珺, 曹阳, 马秦生, 等. 人工免疫行为轮廓取证分析方法[J]. 电子科技大学学报, 2010, 39(6): 911-914.
YANG Jun, CAO Yang, MA Qin-sheng, et al. Forensic analysis method of behavior profiling on artificial immunity[J]. Journal of University of Electronic Science and Technology of China, 2010, 39(6): 911-914.

[2] WANG Wei, DANIELS T E. A graph based approach toward network forensics analysis[J]. ACM Transactions on Information and System Security, 2008, 12(1): 4:1-4:33.

[3] WANG Wei, DANIELS T E. Network forensics analysis with evidence graphs[C]//2005 Digital Forensic Research Workshop. New Orleans: DFRWS, 2005: 1-6.

[4] ABRAHAM T. Event sequence mining to develop profiles for computer forensic investigation purposes[C]// Proceedings of the 2006 Australasian workshops on Grid computing and e-research. Darlinghurst, Australia: Australian Computer Society, 2006: 145-153.

[5] LEE W, QIN X. Statistical causality of INFOSEC alert data[J]. Computer Science, 2005, 5(2): 101-127.

[6] 张有东, 曾庆凯, 王建东. 网络协同取证计算研究[J]. 计算机学报, 2010, 33(3): 504-513.
ZHANG You-dong, ZENG Qing-kai, WANG Jian-dong. Studies of network coordinative forensics computing[J]. Chinese Journal of Computers, 2010, 33(3): 504-513.

[7] ZHANG You-dong. Cooperation forensic computing research[C]//Proceedings of the 1st International Workshop on Knowledge Discovery and Data Mining. Adelaide, Australia: Australian Computer Society, 2008: 25-30.

[8] DUDA R O, HART P E, STORK D G. Pattern classification [M]. 2nd ed. New York: Wiley-Interscience, 2001.

[9] 杨珺, 李晶, 王敏, 等. 计算机证据元数据表示方法[J]. 微型机与应用, 2009, 28(19): 63-65.
YANG Jun, LI Jing, WANG Min, et al. Notation for computer evidential metadata[J]. Microcomputer & Its Applications, 2009, 28(19): 63-65.

[10] MIT Lincoln Laboratory. 2000 DARPA intrusion detection scenario specific data sets[EB/OL]. (2000-2-24) [2009-5-8]. http://www.ll.mit.edu/IST/ideval/data/2000/2000_data _ index.html.

编 辑 张 俊