基于可信概率的电子数据取证有效性模型

孙国梓[1),2),3)] 耿伟明[1)] 陈丹伟[1),2),3)] 申 涛[4)]

[1)](南京邮电大学计算机学院 南京 210003)

[2)](江苏省无线传感网高技术研究重点实验室 南京 210003)

[3)](宽带无线通信与传感网技术教育部重点实验室 南京 210003)

[4)](中国移动通信集团河北有限公司石家庄分公司 石家庄 050021)

摘 要 针对当前证据有效性不足的缺点,结合概率论,提出了基于可信概率的电子数据取证有效性模型.以 Petri 网为基础,将取证后经形式化处理的数据抽象为 Petri 网中的库所,操作行为和取证方法抽象为变迁,后一节点为运用该操作方法对前一节点进行某种变换所形成.给出了取证过程中的基本定义和形式化处理方法,研究了概率计算的相关算法,描述了详细的推理过程.利用"可信度+数据源+取证规则"作为对所得证据的有效性说明,为可信取证的动态取证行为可信提供理论基础.通过概率计算的方法,最终得到具体的概率数据,在保证数据源信息可信的基础上(即静态属性可信的假设前提),通过可信概率(概率值接近 0 或者 1)的方法保证处理过程所使用的取证规则可信(即使用可信的动态取证方法或行为),最终实现电子数据作为证据的高的可信度.最后,设计了有效性证明系统,利用实际案例,分析并验证了可信概率在电子数据取证有效性模型中的具体应用.

关键词 可信取证;有效性;电子数据取证;Petri 网;概率

中图法分类号 TP309 **DOI 号**: 10.3724/SP.J.1016.2011.01262

One Validity Model of Digital Data Forensics Based on Trusted Probability

SUN Guo-Zi[1),2),3)] GENG Wei-Ming[1)] CHEN Dan-Wei[1),2),3)] SHEN Tao[4)]

[1)](*College of Computer, Nanjing University of Posts and Telecommunications, Nanjing* 210003)

[2)](*Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing* 210003)

[3)](*Key Laboratory of Broadband Wireless Communication and Sensor Network Technology of Ministry of Education, Nanjing* 210003)

[4)](*Shijiazhuang Branch, Hebei Co. Ltd, China Mobile Communication Group, Shijiazhuang* 050021)

Abstract According to the shortage of the current evidence's validity, one validity model of digital data forensics based on trusted probability is put forward. Based on Petri net, after collecting the evidence, the digital data processed through formalization is abstracted as the place of Petri net, at the same time, the operating behaviors and forensics methods are abstracted as the transitions. Then the backward nodes are formed by making some transformation on the forward nodes using the methods described above. The model puts forward the basic definitions and the methods of formalization processing. Moreover, it makes some researches on the related algorithms of probability calculation and describes the reasoning process in detail. The validity of evidence is proved by the combination architecture "Credit+Data source+Forensics rules", which provides theoretical basis for credibility of dynamic behavior in trusted forensics. Using the method of probability calculation, the concrete probability value can be finally gained. If the data source is supposed to be trusted, which means the data has trusted static attribute, then we can use the method of trusted probability, whose value is closing to 0 or 1, to ensure the forensics rules to be

收稿日期:2009-11-04;最终修改稿收到日期:2011-03-25.本课题得到国家科技部"十一五"科技支撑计划(2007BAK34B06)、国家自然科学基金(61073114)、江苏省高校自然科学基金(09KJD520007)、南京邮电大学攀登计划(NY208009)及江苏高校优势学科建设工程项目资助.**孙国梓**,男,1972 年生,博士,副教授,主要研究方向为电子数据取证、网络与通信安全.E-mail: sun@njupt.edu.cn.**耿伟明**,男,1985 年生,硕士研究生,主要研究方向为电子数据取证.**陈丹伟**,男,1970 年生,博士,副教授,主要研究方向为电子数据取证、嵌入式系统.**申 涛**,男,1983 年生,硕士,主要研究方向为电子数据取证.

trusted within the processing, and the methods or the behaviors of dynamic forensics are trusted too. These models and methods give high confidence to the digital data as the evidence. In the end, a system of validity proof is designed to analyze and verify the trusted probability through its concrete application in the validity model of digital data forensics.

Keywords　trusted forensics; validity; digital data forensics; Petri nets; probability

1　引　言

计算机证据出现在法庭上已有 30 多年历史，最初的电子证据是从计算机中获得的正式输出，法庭不认为它与普通的传统证据有什么不同. 但随着计算机技术的发展，与计算机相关的法庭案例的复杂性逐渐增加，电子证据与传统证据之间的类似性亦逐渐减弱. 后来，各国逐渐开始对电子证据进行立法，并在已有计算机取证工作的基础上，建立了许多专业的计算机取证部门、实验室及咨询服务公司，如 1984 年美国 FBI 实验室和其它法律执行部门建立的检查计算机证据的实验室. 现在美国至少有 70% 的法律部门拥有自己的计算机取证实验室[1]. 一些公司逐渐开发出了许多非常实用的取证产品，比较好的产品有 Encase、NetWitness、取证大师等.

随着对计算机取证相关问题的研究不断深入，人们的视线不再局限于取证工具的开发，逐渐转向计算机取证学相关理论的研究与完善. 1991 年，在美国召开的国际计算机专家会议上首次提出计算机取证这一术语[2]. 此后，几乎每年都召开以计算机取证为主题的学术会议，当前的计算机取证学正在逐步走向标准化、规范化，并将最终成为一门成熟、完善的学科.

2　现有计算机取证模型

在计算机取证发展的初期，人们关注的只是如何去获取数据，并不断开发出相关取证工具. 进入 20 世纪 90 年代中后期，随着取证技术的发展，人们逐渐发现了由计算机取证基本理论和基本方法的研究滞后所带来的种种弊端，并开始意识到此前过分注重应用产品的开发而忽略了基本理论和基本方法的研究，导致了在犯罪案件侦破中的取证过程没有一致性也没有可依据的统一标准，使得取证的可操作性较差，而且极易遭到法庭上法官的质疑. 由此，业内许多专家开始对取证及取证标准等科学问题进行研究.

自 20 世纪 90 年代以来，计算机取证的研究者们相继提出了很多计算机取证的模型. 有些研究人员提出通过分析网络取证的详细需求，建立包含犯罪行为案例、入侵行为案例和电子证据特征的取证知识库[1,3]；还有的学者提出采用 XML 数据模型、数据融合技术、取证知识库、专家推理机制和挖掘引擎的取证模型，并开始着手研究这些模型的评价机制[1,4-5]. 计算机取证的早期模型包括基本过程模型(basic process model)、事件响应过程模型(incident response process model)、法律执行过程模型(law enforcement process model)、过程抽象模型(an abstract process model)[1,6-9]. 后来，在总结前面这些模型的基础上，形成了综合模型(the integrated digital investigation model)[1,10]等.

另外，作为鉴定机构取证和鉴定的程序，本身也具有相当的重要性，其作用是保证科学的手段和方法得以有效实施，从而得出真实、客观、公正的取证结果和鉴定结论，因此，这种程序应该以技术标准的方式建立，并普遍适用，同时应符合以下要求①：(1) 所有的操作过程都应该记录并加以保存，以便于审查和监督；(2) 对原始材料不能修改或删除，可以复制和分析；(3) 所有的操作都由两人以上进行，并且是要重复的；(4) 使用的计算机软硬件必须是经过鉴定或充分测试的，确保不会对工作造成不应有的影响；(5) 应保证排除病毒或其它非正常因素干扰；(6) 应对证据的真实性进行分析和鉴定；(7) 应对证据的关联性进行分析和鉴定；(8) 应对证据的合法性进行描述.

3　可信取证理念的提出

对于计算机取证的研究，一直以来的工作重点都是电子数据的获取和对所取得数据的分析. 规范化的操作流程一般分为 4 步②：(1) 识别证据. 识别

① http://www.darkst.com/bbs/viewthread.php?tid=33498
② http://vpn.szelink.com/archiver/showtopic-961.aspx

可获取的信息的类型以及获取的方法.(2)保存证据.确保跟原始数据一致,不对原始数据造成改动和破坏.(3)分析证据.以可见的方式显示,结果要具有确定性,分析得出的结论一定要去验证,只有确定的才能作为证据提交.(4)提交证据.向管理者、律师或者法院提交证据.

以上4步是计算机取证的一般流程,而对于要建立一个可信的电子数据取证模型,本文进一步细化,把取证流程分为归类、保存、收集、检查、分析和陈述[11].首先对原始数据进行分门别类的归类;根据类别采取相应的保存方式;接着根据需求进行收集数据以便于采用科学的分析方法进行分析;检查是为了确保这些数据没有被篡改;接着采用数据挖掘、神经网络等分析方法进行分析;最后生成陈述报告.为了确保电子数据的篡改、伪造等破坏证据行为的出现,整个取证过程需要建立监督记录.具体记录内容包括案件编号、发现以及收集的时间地点、证据流转的记录、证据的固定方法、取证工作人员的签名.

由于电子证据的易被破坏性,所以对电子证锯的取证主体有特殊要求,不同主体所取的电子证据的证明力不同,如公证人员所取电子证据的证明力一般大于当事人所取电子证据的证明力.取证程序合法性的规定.应对取证的各个环节加以规范,规定对电子证据的保全及保管的要求,禁止违法取证行为.在分析证据阶段,对于电子数据有效性和可信性的分析还缺乏相应的方法,或者说没有一个标准方式来衡量电子数据的有效性和可信性,这样取得的数据容易受到质疑,也阻碍了对证据有效性的认定.而在计算机取证过程中,公安等执法机关还缺乏有效的可信的工具,目前只是利用国外一些常用的取证工具或凭借取证工作人员自身技术经验,在程序上还缺乏一套保证电子数据可信性的评估体系,提出的证据很容易遭到质疑.

事实上,如果考虑将当前的取证模式扩充为"人+工具+证明",即在着眼于利用工具取得证据的同时,尽可能分享经验丰富的取证人员的知识和经验,并建立一套以逻辑、自动机等数学理论为基础的形式化取证模型,同时从理论上证明模型在取证中的有效性(即取证是可信的),最终形成一套可信取证的推导验证体系,以使电子数据的取证系统更加完备,所取得的数据更具说服力.目前的推理主要是人工操作即经验性的工作,难以保障数据的篡改、伪造等破坏证据行为,而这些活动又不易被发现.为此,本模型从保障实施、进行核查两个方面来避免篡改、伪造等破坏证据的行为.

借鉴可信计算的思想,我们提出"可信取证"(trusted forensics)理念[12],结合电子数据的固有特征和取证行为,将电子数据可信取证宏观分为两方面:①可信的静态属性(电子数据本身的静态特征可信);②可信的动态行为(由电子数据转换为证据所需的过程或行为可信).以此为基础,实现指定电子数据可信获取(发现、固定、提取)和可信展现(分析、表达).

4 理论基础

本文提出一个有效性推理的新思路,该推理以概率论和Petri网为基础,将取证后经形式化处理的数据抽象为Petri网中的库所,操作行为和取证方法抽象为变迁,后一节点为运用该操作方法对前一节点进行某种变换所形成;同时每一节点根据其不同类型具有不同的属性,后一节点概率为前一节点概率与边的概率的乘积.该模型为可信取证的动态取证行为可信提供理论基础.

4.1 证据与概率

在证据学的研究过程中,通常面临一些认识论的问题:(1)已经过去的事实是否能够在诉讼中再现;(2)何种程度的再现才能完全排除第2种解释的可能性;(3)使用何种方法或途径再现才具有不可怀疑的可靠性;(4)如何准确地定义事实的真相等.下面分别予以对应的简单说明.

对于上述的第1点,从绝对意义上来说,过去的事实是一种时间和空间的终结,我们只能相对模拟而无法绝对再现,任何意义上的模拟充其量只是最大程度地接近事实;上述的第2、3、4点的本质是相同的,除了这种再现能够为法律和社会伦理准则所认可外,从技术角度上讲,对过去模拟的程度有赖于概率论的应用.也就是说,我们应当回答的问题是:准确模拟出过去的事实的概率究竟是多大,这中间可能发生的偏差会是多少.

以取证过程中的指纹鉴定为例[10]:统计分析证明,人类十指指纹完全相同的概率大约为60亿分之一,其概率基数基本上等同于全球现有人口,因此利用指纹比对做出的同一认定,通常具有非常可靠的排他性.也就是说,在犯罪现场如果出现嫌疑人的指纹,那么认定嫌疑人出现在现场(再现出"嫌疑人出现在现场"这一事实)的概率基本上接近1,也就是

可以非常可靠地再现出过去的事实(除非有人能够伪造他人的指纹).指纹做为证据的有效性完全依赖于经验上的统计概率,也就是从经验来讲,两个人指纹相同的概率接近 0,但这实际上并不能够完全排除出现特例的情形,如果在一起凶杀案中发现一个指纹,而恰好某个人的指纹与作案人员的指纹相同而被错误判刑的可能性仍然存在,只是这种可能性很低,因此能够被社会伦理和法律所接受,实际上这也就是"冤枉概率" 问题,也就是说一旦你的证据再现过去事实的准确程度使得"冤枉嫌疑人的概率"非常的低,以至于能够被社会伦理和法律所接受,那么你的证据就是有效且可信的,本文称概率值接近于 0 或 1 的概率为可信概率.

4.2　概率论及 Petri 网基础

有关概率论的概念和公式,可参见文献[13],我们需要关注相应的条件概率和独立性的一些基本性质,特别是全概率公式和贝叶斯公式.

设$(\Omega,\mathbb{F},P)$为一概率空间,$A_1,A_2,\cdots,A_n$是Ω的一个有穷剖分,即

$$\sum_{i=1}^{n} A_i = \Omega \tag{1}$$

且 $A_i\in\mathbb{F}$,$P(A_i)>0$, $i=1,2,\cdots,n$. 则

(1) 对任一事件 $B\in\mathbb{F}$,有全概率公式

$$P(B) = \sum_{i=1}^{n} P(B/A_i)P(A_i) \tag{2}$$

(2) 对任一事件 $B\in\mathbb{F}$,且 $P(B)\geq 0$,有贝叶斯公式

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{j=1}^{n} P(B/A_j)P(A_j)}, \quad i = 1,2,\cdots,n \tag{3}$$

全概率公式应用的环境可解释为:导致某一事件发生的原因可能有多种,每一种原因对该结果的发生做出一定的"贡献",已知各原因的概率,求该结果可能发生的概率值.

贝叶斯公式实际上是全概率公式的"逆问题",可叙述为:若已知各种"原因"概率,设在进行的随机试验中该事件已发生,问在这样的条件下,各原因发生的条件概率是多少.

Petri 网是对离散并行系统的数学表示. Petri 网是 20 世纪 60 年代由卡尔·A·佩特里发明的[14],它既有严格的数学表述方式,也有直观的图形表达方式,既有丰富的系统描述手段和系统行为分析技术,又为计算机科学提供坚实的概念基础.

经典的 Petri 网是简单的过程模型,由两种节点(库所和变迁)、有向弧、令牌等元素组成的,有关 Petri 网的形式化描述请参见文献[14].

5　基于 Petri 网的有效性推理过程

5.1　基本定义

基于 Petri 网,通过对文献[14-19]的研究,结合实际推理过程,本文给出以下定义:

(1) 原始数据集合:$D'=\{d'_1,d'_2,\cdots,d'_m\}$.

它是指在被犯罪分子破坏之前的数据,即系统中原有数据,该数据根据可由数据恢复软件和分析过程进行推理两方面获得并验证.

(2) 现有数据集:$D=\{d_1,d_2,\cdots,d_n\}$.

它是指被犯罪分子破坏之后的数据,取证人员可以直接在系统中发现或恢复的数据.

(3) 现有结果集(破坏结果):$R=\{r_1,r_2,\cdots,r_k\}$.

该集合为案发时被犯罪分子破坏的现场的情况,包括传统意义上的结果和计算机或移动设备中的数据或结果.

(4) 取证方法集:$M=\{m_1,m_2,\cdots,m_j\}$.

它是指取证人员在取证过程中可以采用的取证方法,如镜像备份、MD5 校验、数据恢复及其采用的工具等.

(5) 入侵方法集:$I=\{i_1,i_2,\cdots,i_q\}$.

它包括所有可能的入侵手段或方法、采用的工具等.

(6) 规则集:$N=\{n_1,n_2,\cdots,n_p\}$.

它是一些规则的集合,包括取证规则和操作规则两部分.这些规则存储在有效性推理过程中有效性规则库中.

取证规则说明了取证人员在取证时采用何种方式得到的数据有效.以下为部分取证有效性规则.

规则 1.　$\exists d_i\in D$, $d_j=\mathrm{Copy}(d_i)$,则 d_j.content$=d_i$,取证有效.

规则 2.　$\exists d_i,d_j\in D$, $\mathrm{MD5}(d_i)=\mathrm{MD5}(d_j)$,则 $d_i\equiv d_j$,取证有效.

规则 3.　$\exists d'_i\in D'$, $d_i=\mathrm{Mirror}(d'_i)$,则 $d_i\equiv d'_i$,取证有效.

规则 4.　$\exists d'_i\in D'$, $d_i=\mathrm{Recover}(d'_i)$,则 $d_i\equiv d'_i$,取证有效.

规则 5.　$\exists e_i$, $e_j\in E$, if $e_i\rightarrow e_j$,即证据 e_i可以验证 e_j,如果 e_i有效,则 e_j有效.

规则 6. $\exists d_i' \in D', \exists d_i \in D, D', D \neq NIL$，若 $E(d_i')/R$，则若保存缓存数据 $\in E$，则 $D' \equiv D$，且 d_i 有效.

规则 7. $\exists d_i' \in D', \exists d_i \in D, D', D \neq NIL$，若 $S \xrightarrow{\text{自备无病毒启动盘}} R, E(d_i')/R$，则 $D' \equiv D$，且 d_i 有效.

操作规则说明了日常操作可能对文件属性所做的修改等. 以下为部分操作规则.

规则 8. $\exists d_i \in D$, Visit $\in OP$, if (Visit(d_i)), then Change(d_i.lastVisitTime).

规则 9. $\exists d_i \in D$, Edit $\in OP$, if (Edit(d_i)), then Change(d_i.content, d_i.visitTime, d_i.modifyTime).

规则 10. $\exists d_i \in D$, Create $\in OP$, if (Create(d_i)), then New(d_i).

(7) 入侵结果集：$IR=\{ir_1, ir_2, \cdots, ir_l\}$.

它包括针对所有入侵手段或方法、采用的工具等所产生的可能的各种入侵结果.

(8) 操作动作集：$OP=\{op_1, op_2, \cdots, op_x\}$.

它包括日常对计算机的操作动作，定义该集合是因为犯罪分子在入侵过程中不免要对计算机进行一些不属于入侵方法中的操作.

(9) 属性集：$A=\{at_1, at_2, \cdots, at_t\}$.

它包括文件的属性名称，如创建者、创建时间、修改者、修改时间、访问时间等属性，以及某些特定文件格式所具有的属性.

(10) 经过计算机取证，所得证据集合：$E=\{e_1, e_2, \cdots, e_n\}$.

(11) 二元关系 $IIR=I\times IR=\{(i_q, IR_q)\}$ 为由入侵方法 i_q 所产生的结果集合 IR_q，其中 $IR_q=\{ir_{q1}, ir_{q2}, \cdots\}$ 为集合 IR 的子集.

(12) 二元关系 $MN=M\times N=\{(m_j, N_j)\}$ 为取证方法 m_j 所满足的取证规则集合 N_j，其中 $N_j=\{n_{j1}, n_{j2}, \cdots\}$ 为集合 N 的子集.

源数据属性中具有获取方法属性，根据对应关系，某种方法在二元集合中对应某些取证规则，即满足这些规则，说明有效性，这样，该数据间接满足这些规则，方便查找.

(13) 二元关系 $OPA=OP\times A=\{(op_i, A_i)\}$ 为某一操作动作 op_i 所修改的文件属性集合 A_i，其中 $A_i=\{at_{i1}, at_{i2}, \cdots, at_{in}\}$ 为集合 A 的子集.

根据所调查的文件的属性修改情况，并结合某一操作动作可以对文件所做的修改，判断其是否发生. 若文件属性修改情况与集合中某一规则不符，则该操作不发生.

(14) 取证开始前，可疑计算机的状态定义为以下 4 种：S(关闭)、R(运行)、C(接入网络)、$\sim C$(未接入网络).

(15) 通过取证方法集 M 中的技术方法获取取证对象中的数据，即生成现有结果集 R 的过程必须进行监督记录. 记取证过程中的监督记录为 $W=\{w_1, w_2, \cdots, w_m\}$，则在证据审核过程中必须满足：

$\forall r_k \in R, \exists w_j \in W, w_j.\text{content}=\text{Record}(m_i, r_k)$.

(16) 事件 A 在情况 B 下发生，表示为 A/B.

(17) 在 Petri 网中，$\Sigma=(S, T; F, K, W, M_0)$ 为网系统. 根据网系统定义，结合取证过程实际，定义为

系统 $\Sigma=(S, T; F, K, W, P, A, G, U)$，

其中：S 为数据或某一事件或具有某一属性的实物的集合，为库所集，又称 S_元，原始数据集 $D'\subseteq S$，结果集 $R\subseteq S$.

T 为变迁，又称 T_元，文中是操作方法的集合.

K 为最终数据源函数，即若 $x\in S$，则 $K(x)$ 表示可以得出该 S_元 x 的属于现有数据集 D 的所有元素的集合.

W 为权函数，表示推理过程中某一方法的概率，或前一节点在推导出后一节点所占的权重.

P 为库所的概率.

A 为属性集，其中存储有元素的属性值. 如 $A(d)$.content 表示文件 d 的内容.

G 为操作方法，$G(d)$ 表示对 $^{\cdot}d$ 中元素采取的操作方法，即变换的集合.

U 为该 S_元 x 所描述的事件. 若 x 本身为某一文件则 $U(x)$ 为文件本身.

假设各个原始数据源 D' 中元素概率为 1，即 $\forall d_i \in D', P(d_i)=1$.

根据以上定义，可得出如下性质：

(1) $\forall x\in S, 0\leq P(x)\leq 1$.

在日常生活中，概率值不大于 1 同时又不小于 0 是毫无疑问的，因此必须保证模型中所有概率值符合人们的习惯.

在证明过程中需要用到节点之间的概率计算，具体证明过程参见 5.3 节的证明.

(2) 若 $\exists z\in x^{\cdot}, {}^{\cdot}z=\{x\}$，则 $|{}^{\cdot}z|=|f(x)|=1$.

上式是说，若 z 为 x 的输出集，且 z 的输入集只有元素 x，则 z 只由 x 经某一操作 f 得出.

(3) $K(x)$中元素个数的关系：

$$|K(x)|=\sum_{y\in{}^{\cdot}x}|K(y)|-\sum_{y_i,y_j\in{}^{\cdot}x}|K(y_i)\cap K(y_j)|+\cdots+(-1)^{k-1}\cdot\sum_{y_{i_1},y_{i_2},\cdots,y_{i_k}\in{}^{\cdot}x}|K(y_{i_1})\cap K(y_{i_2})\cap\cdots\cap K(y_{i_k})|+\cdots+(-1)^{|{}^{\cdot}x|-1}\cdot\left|\bigcap_{y_i\in{}^{\cdot}x}K(y_i)\right| \tag{4}$$

其中$|K(x)|$表示集合 $K(x)$中元素的个数.

(4) 若 $x\in S$,根据集合论知识,有关于 $K(x)$的计算公式：

$$K(x)=\bigcup_{y\in{}^{\cdot}x}K(y)-\bigcup_{y_i,y_j\in{}^{\cdot}x}(K(y_i)\cap K(y_j))+\cdots+(-1)^{k-1}\cdot\bigcup_{y_{i_1},y_{i_2},\cdots,y_{i_k}\in{}^{\cdot}x}(K(y_{i_1})\cap K(y_{i_2})\cap\cdots\cap K(y_{i_k}))+\cdots+(-1)^{|{}^{\cdot}x|-1}\cdot\bigcap_{y_i\in{}^{\cdot}x}K(y_i) \tag{5}$$

其中，

$$\sum_{y_i,y_j\in{}^{\cdot}x}K(y_i)\cap K(y_j)$$

表示任意两个 x 前集中的元素所公共的前导元素的个数. 其余类同.

上式清晰地表示出了 S_元 x 与其输入集 ${}^{\cdot}x$ 之间关于数据源集合 D 的关系,方便了计算.

(5) $x\in S$,若$|K(x)|=1$,则 $K(x)=K({}^{\cdot}x)$.

显然,若 x 的前集中只有一个元素,则 x 的最终数据源与其前集中元素的最终数据源相同.

(6) $x\in S$,若$|K(x)|=1$,则 $W({}^{\cdot}x,\ x)=1$, $P(x)=P({}^{\cdot}x)$.

若 x 的前集中只有一个元素,则该方法必然对元 x 起决定作用,所以 $W({}^{\cdot}x,\ x)=1$.

5.2　形式化处理

在获取原始数据后,数据的格式各式各样. 对于主机数据而言,文件有着各种文件类型及时间属性等,而网络数据则包含各种协议的数据包. 若不对它们进行统一的格式处理,则在以后的分析中会产生很大的麻烦. 因此,在分析数据之前,有必要进行形式化处理.

所谓形式化处理,就是在获取原始数据后,对所有数据进行处理,处理后的数据有着统一的、系统易识别的格式.

下面定义数据属性.

数据类型

{1:存储设备数据;2:手持设备数据;3:外围设备数据;4:其它潜在的数字证据数据;5:与网络相关的数据}.

本文所列出的数据类型 5 是指能反映上网信息的日志、文档等静态存储在本地计算机或者网络服务器中的文件. 之所以将与网络相关的数据单独列为一种类型,是因为这种类型的数据反映了一定的网络行为,相对于一般电子证据地位特殊,对案件的侦破工作具有重要意义. 当然,本文所分析的电子证据还是传统意义上的以文件为核心的数据证据. 对于网络数据的电子化取证过程更为复杂,限于篇幅,本文未作详细描述,本文相关推理方法的形式化过程也是没有针对网络数据源的.

数据源

{1:硬盘驱动器;2:外部硬盘驱动器;3:可移动媒体;4:拇指驱动器;5:记忆卡;}.

文件类型

{1:Word 文档;2:TXT 文件;3:电子邮件;4:图片文件;5:可执行文件;6:日志文件;7:IE 缓存文件;8:聊天记录;9:音频/视频文件}.

在存储属性时,可直接存储其代码,如利用两字节表示上述属性,字节从高至低依次定义为:第 1～2 位表示该节点为数据(00)或事件(01)或某一属性特征(10),第 3～5 位为数据类型,第 6～8 位为数据源,第 9～12 位为文件类型,其它暂未定义.

每个节点根据其不同的数据类型定义不同的属性,如：

(1) IE 缓存文件:{数据源,链接地址,访问时间,内容关键字,取证方法};

(2) Word 文档或 TXT 文件:{数据源,作者,创建时间,修改时间,访问时间,内容关键字,取证方法};

(3) 电子邮件文件:{数据源,收件人,发件人,主题,时间,转发信息,取证方法};

(4) 病毒或木马文件:{数据源,病毒类型,特性,IP 地址,取证方法};

(5) 恢复的数据:{修改日期,恢复工具,原文件存放位置,取证方法},

其中,数据源指明源数据的存放位置,取证方法为该数据所获得的方法.

5.3　概率的计算

各节点有其概率值,同时变迁(即操作方法)有其自己的概率,表示该方法对结果的重要性. 可视为

概率论中的条件概率,其大小值规定为

(1)若 $z\in x^{\cdot}$,且$|{}^{\cdot}z|=1$,则 $W(x,z)=1$. 例如图 1 中的 d_{11},d_{22},d_{23} 等.

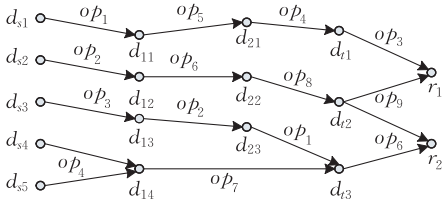

图 1 推理过程示意图

(2)若结果由多个方法得出,则多个方法概率之和为 1,每个方法具体值根据重要性确定;即

$$\exists y_i\in{}^{\cdot}x,\ \sum_{y\in{}^{\cdot}x}W(y,x)=1 \tag{6}$$

参见图 1 中节点 d_{14},d_{t3}. 对于 d_{t3},由两个方法 op_1 和 op_7 得出,则 $op_1+op_7=1$,假如节点 d_{23} 对 d_{t3} 的影响有 70%,则可设 $op_1=0.7$,$op_7=0.3$.

各节点概率计算方法如下:

(1)$z\in x^{\cdot}$,若 $W(x,z)=1$,所以 $P(z)=P(x)$.

(2)若$|K(x)|>1$,对于 $\forall i\in K(x)$,$W(y_i,x)$已知,且有式(6),则应用全概率公式计算 $P(x)$得

$$P(x)=\sum_{y\in{}^{\cdot}x}P(y)\cdot W(y,x) \tag{7}$$

(3)根据具体情况计算:多发生在由多个节点推出的事件中,由于节点间的相关性问题而不能或很难直接利用前一节点概率值得出. 结果应保证$0\leq P\leq 1$.

例如图 1 中节点 d_{t1},由于其只由数据 d_{21} 得出,因此操作方法 op_4 的权重值为 1,其可信度与 d_{21} 相同;而对于数据 d_{t3},由于该节点由 d_{14} 和 d_{23} 两个数据得出,假设方法 op_1 概率为 p,则 op_7 概率为 $1-p$,所以 $P(d_{t3})=p\cdot P(d_{14})+(1-p)\cdot P(d_{23})$. 若有 n 个数据,则各个方法概率之和为 $p_1+p_2+\cdots+p_n=1$.

事实上,第 1 种情况(即 $W(x,z)=1$ 时)是第 2 种情况的一个特例. 只是由于概率特殊,所以单独说明.

下面利用节点之间的概率计算公式对 5.1 节中性质(1)进行证明.

证明. 对于紧接 d_i 的节点 $d_{1,j}$,首先证明 $d_{1,j}\leq 1$.

(1)若$|{}^{\cdot}d_{1,j}|=1$,不妨设其前集为 d_{i0},则 $P(d_{1,j})=P(d_{i0})=1$.

(2)若$|{}^{\cdot}d_{1,j}|>1$,设其前集为${}^{\cdot}d_{1,j}$,则

$$P(d_{1,j})=\sum_{d\in{}^{\cdot}d_{1,j}}P(d)\cdot W(d,d_{1,j})\leq \sum_{d\in{}^{\cdot}d_{1,j}}1\cdot W(d,d_{1,j})=\sum_{d\in{}^{\cdot}d_{1,j}}W(d,d_{1,j})=1 \tag{8}$$

(3)由其它方法计算出来的,在计算过程中已保证 $0\leq P\leq 1$. 所以 $P(d_{1,j})\leq 1$.

假设证明的节点为 $d_{i,j0}$,其前集为${}^{\cdot}d_{i,j0}$. 设集合$\{P(d)\mid \forall d\in{}^{\cdot}d_{i,j0}\}$的最大值为 $P_{\max}$,则有 $0\leq P_{\max}\leq 1$.

$$P(d_{i,j_0})=\sum_{d\in{}^{\cdot}d_{i,j_0}}P(d)\cdot W(d,d_{i,j_0})\leq \sum_{d\in{}^{\cdot}d_{i,j_0}}P_{\max}\cdot W(d,d_{i,j_0})=P_{\max}\cdot\sum_{d\in{}^{\cdot}d_{i,j_0}}W(d,d_{i,j_0})=P_{\max} \tag{9}$$

显然又有 $P_{i,j0}\geq 0$.

所以 $0\leq P_{i,j0}\leq 1$. 证毕.

由以上证明可知,模型中概率符合日常人们的思维习惯.

在电子数据取证过程中,为使其达到法庭信任的程度,可根据需要设定两个代表可信度的概率值 δ_1 和 δ_2,使推理过程中各个节点可信度达到 δ_1,而得出的最终结果可信度达到 δ_2.

5.4 概率的验证

在计算出概率后,可以利用贝叶斯公式对推理过程和概率值进行验证.

根据贝叶斯公式(式(3)),在整个过程的概率计算完成后,$P(A_i)$,$P(B/A_i)$都为已知. 其中,$P(B/A_i)$即为由 $A_i\mid(i=1,2,\cdots,n)$计算出的概率值. 也就是说事件 B 的前集为:${}^{\cdot}B=\{A_i\mid(i=1,2,\cdots,n)\}$.

由贝叶斯公式,可以计算出各个前集元素 $A_i\mid(i=1,2,\cdots,n)$的条件概率 $P(A_i/B)$,$i=1,2,\cdots,n$.

在实际中,根据以往案例的经验或其它方面的参考,取证人员对于每个事件 $A_i\mid(i=1,2,\cdots,n)$及其方法 f_i在其结果中所占比例可以有一个估计值. 因此,取证人员可以据此对概率值做出判断.

5.5 推理过程

根据以上定义,推理过程可描述为:已知原始数据集合 D'和结果集合 R 分别为起始节点集和终点集(两集合不一定是原始状态的数据,而是经过形式化处理后在源文件中提取出的数据),根据取证方法集合 M 和入侵方法集合 I,如何将两个集合连接,并使连接最大化(在推理过程中可能会产生新的节

点，即新数据），并且保证图中每个节点的概率值大于某一预设值 δ，以保证推理过程的可信度.

设形式化处理后的数据 $DS=Formalize(D')=\{d_{s1},d_{s2},\cdots,d_{sr}\}$，则 DS 也就是推理过程的起始集. 推理过程的中间节点集用 d_{ij} 表示（其中，i 表示其在推理过程中的第几步，j 表示该步骤中的第几个节点. 但 i 并不是严格的，因推理中节点可能在多次用到），终止节点以 $DT=\{d_{t1},d_{t2},\cdots,d_{ts}\}$ 表示. 推理过程示意图如图 1 所示.

图 1 中，省略表示变迁节点的方形点，并在连接线上方表示该变迁. 仅以少数节点和步骤表示，而实际中的过程可能要比图示复杂得多.

综上所述，整个推理过程如图 2 所示.

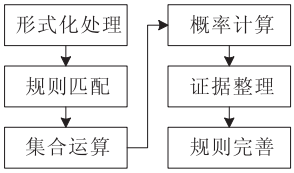

图 2　推理过程

（1）首先将原始数据进行形式化处理，按照预定格式将数据进行格式统一，标明原始数据源、时间、创建者、访问者、最后修改者等属性，以及该数据的获取方式（采用何种取证工具获得）. 在形式化处理的过程中要按照原始数据的可信级别进行划分. 具体的可信级别划分可以参考文献[20]，从 C0～C6 共 7 个级别. 对数据源可信度的划分也是实现可信取证的重要方面.

（2）以形式化的数据为初始节点，现有结果集合为终结点，根据收集到的文件系统及其属性的变化，结合模型中操作方法及其对应的操作结果的规则限制，采取纵向或横向对比等方法，从前后两端开始进行推理，进行规则匹配. 例如，对于现有初始节点，通过哪种操作，可能得出怎样的结论；要达到现有的结果，需要怎样的条件和数据，采取什么方法；根据操作规则，某文件的某一属性或几个属性的修改需要对其进行怎样的操作等. 在推理过程中标明操作以及该操作后前一节点所做的改动等.

（3）根据节点数据源的性质，利用集合的交、并等运算，得出各个节点的原始数据源.

（4）根据节点概率计算方法，计算各个节点的概率值.

（5）对整个分析过程进行整理，得出证据，必要时可以根据分析过程，附上对最终证据的说明（如证据的由来、其可能性大小等），分析过程的几个关键点、对证据具有最大影响力的原始证据等.

（6）对于特定案件，在实施过程中应对入侵规则、取证规则等进行修正和完善，使之更加准确.

6　系统及实例

6.1　系统架构

根据前述的取证模型和有效性分析过程，通过对文献[1-3,10,21-27]等的研究，提取其可取之处，本节描述取证有效性证明系统的结构（如图 3 所示）. 该系统将静态取证和动态取证结合，在强调静态取证、严格按照静态取证步骤执行的同时，加入了动态取证，即入侵检测模块，必要时可以进行实时监控. 在进行动态监控的同时，又不影响静态分析. 各个模块既相互关联，相互协作，又具有一定的独立性.

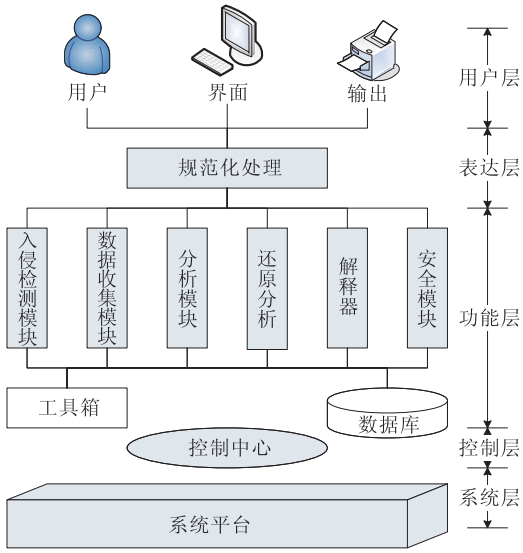

图 3　系统架构图

（1）系统层

系统层支撑有效性证明系统的运行，主要包括软件和硬件的平台，可以提供证明所需要的数据存取和应用程序运行的各种环境.

（2）控制层

控制层位于系统层之上，是整个取证分析和有效性证明系统的控制核心. 负责对各个模块（如入侵检测模块、数据获取模块、数据分析模块等）的启动、调用等发送控制信息，由各相关模块完成工作以及系统设置等其它功能. 电子证据很容易被修改、销毁或以令人信服的方式被制造. 不正确的信息所具有潜在的破坏性比没有信息更大. 在取证工作中，要避免电子证据被偶然或恶意的修改是控制层需要完成的工作.

一方面，要保障标准操作流程的严格执行. 通过加强监督取证流程，进一步规范取证操作等来保障取证执行. 另一方面，要进行相应的核查工作. 对于推理分析过程中用到的证据，要对其可信度进行归类，例如对于单一来源的证据，要试图找到能够作为佐证的其它证据来完成对该证据的核查.

(3) 功能层

该层分为以下几个部分：

① 入侵检测模块. 入侵检测系统负责进行实时监控. 当需要进行监控时，可以启动入侵检测模块，并将获取的数据存储至数据库中，以备后用.

② 数据收集模块. 该模块主要负责进行数据的获取，通过调用相关工具，完成原始数据的收集，并对所收集数据进行校验以及存储等.

③ 分析模块. 根据 5.5 节描述的分析过程，对数据进行分析，图 4 给出该部分包含的几个功能.

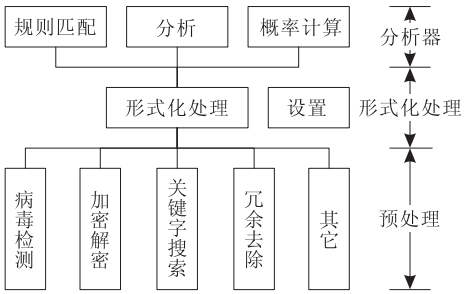

图 4 分析模块

i. 木马病毒检测. 结合安全模块对数据、文件进行分析，防止木马或病毒的破坏. 对于有病毒木马的系统，期望通过分析可以获取有价值的信息，例如获得入侵者 IP 地址，获得与身份相关的用户名等信息.

ii. 加密解密处理. 首先分析是否有密码保护，或经过加密处理的文件，并对经过加密处理的文件，利用工具箱中密码破解工具尝试进行破解.

iii. 其它预处理. 执行除上述两种处理之外的处理工作，如进行关键字搜索、排序，去除所得数据中重复、错误或无用的冗余数据，以便简化后面的有效性判定，提高执行效率等操作.

iv. 形式化处理. 将经过预处理的数据根据预先设定的格式进行格式化处理，以便分析器进行分析.

v. 分析器. 数据经过形式化处理后，由分析器利用规则，根据前述分析过程对数据进行分析，并计算相应概率值. 分析完成后，将中间数据、结论等交由解释器处理. 该部分可分为 3 个小模块，分析部分负责根据经过形式化处理过的文件信息，分析入侵者可能对其进行的操作，规则匹配模块负责根据操作规则和取证规则，分析经过某操作后产生的文件的信息，概率计算模块负责根据 5.3 节描述的概率计算方法计算各个数据的概率.

④ 安全模块. 该模块主要负责保障系统的安全，并可以根据需要进行设置. 例如：与分析模块结合对取证后的数据进行木马病毒的分析及病毒库的更新，当需要利用网络进行数据传输时确定是否对数据进行加密以及采用的加密方式. 还包括数据的本地存储安全性设置及其它安全相关管理等功能.

⑤ 还原分析. 根据已有数据，包括取证过程中所获取的原始数据和分析过程中所得的数据，对整个入侵过程进行还原分析.

⑥ 解释器. 解释器根据原始数据、中间数据以及犯罪过程还原情况等信息，结合知识库，形成“可信度＋数据源＋取证规则”的证据格式，并负责对证据进行必要性的说明，以便在法庭上呈现时更具说明力.

⑦ 数据库. 用于存储各种数据，存储内容可分为两类：

规则库：包括入侵检测系统规则、取证规则、有效性规则、知识库等.

数据：包括各种数据、文件，包括取证的原始数据、分析过程中产生的中间数据、结果数据以及以往案例数据等.

⑧ 工具箱. 工具箱包含了取证过程中的常用工具，例如：数据恢复工具、磁盘镜像工具、密码破解工具、案例工具、完整性校验工具等等. 在取证过程中由相关模块调用.

(4) 表达层

功能层完成对数据的分析，形成证据后，交由表达层对证据格式进行处理. 根据法律法规，对由功能层形成的“可信度＋数据源＋取证规则”的证据的合法性进行检查. 并按照法律法规规定的格式对所得的证据进行处理.

(5) 用户层

用户层是整个系统的重要部分，是用户与系统进行人机交互的纽带. 负责界面呈现和人机交互，主要包括图形化菜单、对话框、输入框的显示、最终证据的显示、输出打印等功能.

6.2 实例分析

举一个简单的例子具体说明概率在推理过程的应用，说明前面的推理过程.

例 1. 跟踪一个传播淫秽信息的网站. 根据该

网站维护的日志找到维护该淫秽网站的嫌疑人张，据张交代，张本人并不是网站的建设人员，他只是受李委托建设该网站，但是李本人对此矢口否认. 根据张交代，他和李是多年好友，李向其支付费用时均是直接向其提供现金，因此并无直接的证据证明李向其提供了资金. 经对李的主机分析，李经常访问该淫秽网站，但并未发现维护该网站的记录. 但李声称自己是从张那里得知这一网站，因此偶尔访问. 但是根据张交代，他每次向网站上上传完新的内容之后，都会通过电话告知李，李对此矢口否认，表示自己对网站的管理一无所知. 经调查发现表 1 所示基本情况.

表 1　案件相关基本情况

序号	时间	行为
1	8 月 10 日 10 时 20 分	张上传电影 1. mpg 2. mpg 3. mpg
2	8 月 10 日 10 时 25 分	李的主机访问电影 1. mpg
3	8 月 30 日 9 时 41 分	张上传电影 4. mpg 5. mpg 6. mpg
4	8 月 30 日 9 时 50 分	李的主机访问电影 5. mpg
5	9 月 2 日 22 时 51 分	张上传电影 7. mpg 8. mpg 9. mpg
6	9 月 2 日 22 时 56 分	李的主机访问电影 7. mpg

在推理分析之前，有两点需要论述一下：

首先是数据源的可信性. 假设李未使用别的手段(如新闻推送、能自动搜索网站更新的搜索工具等). 根据本案例的情况，证明张上传电影行为的电子证据的可信级别属于 C4 级. 如果有多来源的证据可用，那么数据的可信性是有支撑有支持的，因为有诸多细节可以佐证. 这样就降低了数据在被取证之前或者在取证工作过程中被篡改、伪造等破坏证据行为的可能性.

其次是推理分析的过程. 在推理分析过程中，通常会在一开始调查员提出一个初步的假设，基于这个假设，一系列支持假设的电子数据将被选择用来推理，最后很有可能会得出一个结论. 当然，得出的结论在没有其它证据相互佐证、相互印证之前只能是作为一个线索.

分析：通过研究表 1 发现，李每次在张上传电影 10min 内立即进行访问确认. 假设李所说的是事实，则他可以在任意时间访问，又经调查发现李通常在 8 点至 22 点上网，则他恰在电影上传的时间内访问的概率为 1/84，又由于每次上传和访问是互不相关的，即相互独立的，故可得李每次在张上传 10min 内访问的概率为 $1/84^3$，即 1/592704. 假如有 5 次事件，则可得出概率为 40 亿分之一. 几次重复事件发生的概率如此之小，因此可以确认“李在张上传 10min 内访问而张未告知李”的事件不可能发生，故证明李所说为假. 图 5 显示了整个推理过程：

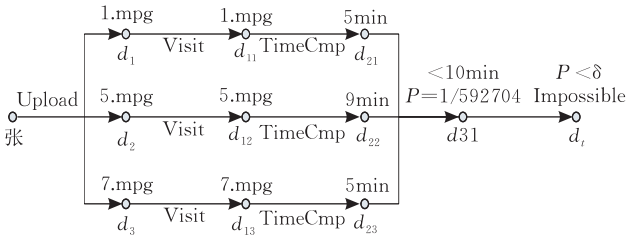

图 5　推理过程

图 5 中，节点上方表示该节点所代表的意义，下方表示该节点的概率(概率为 1 的点未标明)，箭头下方标明对前一节点所做的操作.

根据前面的推理过程描述，分析过程中节点 d_1 的属性值如图 6 所示.

图 6　节点 d_1 属性窗口

Visit 之前，1. mpg 的节点属性为(记该节点为 d_1)

$P(d_1)=1$.

$G(d_1)=\{\text{Upload}\}$.

$K(d_1)=\{1.\text{mpg}\}$.

$U(d_1)=\{1.\text{mpg}\}$.

根据操作规则，Visit 方法只修改属性中的 visitTime 属性，其余属性不变.

$A(d_1)=$ {数据源：磁盘文件；数据类型：视频文件；创建者：张；创建时间：8. 10. 10：20；最后访问者：张；访问时间：8. 10. 10：20}.

Visit 之后，1. mpg 的节点属性为(记该节点为 d_{11})

$G(d_{11})=\{\text{Visit}\}$.

$U(d_{11})=\{1.\text{mpg}\}$.

$\cdot d_{11}=\{d_1\}$.

由于 $|G(d_{11})|=1$，所以 $P(d_{11})=P(d_1)=1$，$K(d_{11})=K(d_1)=\{1.\text{mpg}\}$.

$A(d_{11})=$ {数据源：磁盘文件；数据类型：视频文件；创建者：张；创建时间：8. 10. 10：20；最后访问者：李；访问时间：8. 10. 10：25}.

另外，对于 d_1 和 d_{11}，$W(d_1,d_{11})=1$.

同样，Visit 之前，5. mpg 的节点属性为(记该节点为 d_2)

$P(d_2)=1$.

$G(d_2)=\{\text{Upload}\}$.

$K(d_2)=\{5.\text{mpg}\}$.

$U(d_2)=\{5.\text{mpg}\}$.

$A(d_2)=\{$数据源:磁盘文件;数据类型:视频文件;创建者:张;创建时间:8.30.09:41;最后访问者:张;访问时间:8.30.09:41$\}$.

Visit 之后,5.mpg 的节点属性为(记该节点为 d_{12})

$G(d_{12})=\{\text{Visit}\}$.

$U(d_{12})=\{5.\text{mpg}\}$.

$\cdot d_{12}=\{d_2\}$.

由于$|G(d_{12})|=1$,所以 $P(d_{12})=P(d_2)=1$,$K(d_{12})=K(d_1)=\{5.\text{mpg}\}$.

$A(d_{12})=\{$数据源:磁盘文件;数据类型:视频文件;创建者:张;创建时间:8.30.09:41;最后访问者:李;访问时间:08.30.09:50$\}$.

对于 d_2 和 d_{12},$W(d_2, d_{12})=1$.

Visit 之前,7.mpg 的节点属性为(记该节点为 d_3)

$P(d_3)=1$.

$G(d_3)=\{\text{Upload}\}$.

$K(d_3)=\{7.\text{mpg}\}$.

$U(d_3)=\{7.\text{mpg}\}$.

$A(d_3)=\{$数据源:磁盘文件;数据类型:视频文件;创建者:张;创建时间:09.02.22:51;最后访问者:张;访问时间:09.02.22:51$\}$.

Visit 之后,7.mpg 的节点属性为(记该节点为 d_{13})

$G(d_{13})=\{\text{Visit}\}$.

$U(d_{13})=\{7.\text{mpg}\}$.

$\cdot d_{13}=\{d_3\}$.

由于$|G(d_{13})|=1$,所以 $P(d_{13})=P(d_3)=1$,$K(d_{13})=K(d_3)=\{7.\text{mpg}\}$.

$A(d_{13})=\{$数据源:磁盘文件;数据类型:视频文件;创建者:张;创建时间:09.02.22:51;最后访问者:李;访问时间:09.02.22:56$\}$.

对于 d_3 和 d_{13},$W(d_3,d_{13})=1$.

随后,进行访问时间和创建时间的纵向对比 TimeCmp,得出时间差分别为 5 min,9 min,5 min.该步骤虽是对比操作,涉及两个文件,但由于是纵向对比而非横向对比,不涉及其它文件,因此不需要计算对比的概率,概率值同前一节点,得出各节点属性为

$d_{21}:G(d_{21})=\{\text{TimeCmp}\}$.

$U(d_{21})=\{\text{TimeMargin}(\text{visitTime}-\text{createTime})=5\,\text{min}\}$.

$\cdot d_{21}=\{d_{11}\}$.

同样,由于$|G(d_{21})|=1$,所以 $P(d_{21})=P(d_{11})=1$,$K(d_{21})=K(d_{11})=\{1.\text{mpg}\}$.

$d_{22}:G(d_{22})=\{\text{TimeCmp}\}$.

$U(d_{22})=\{\text{TimeMargin}(\text{visitTime}-\text{createTime})=9\,\text{min}\}$.

$\cdot d_{22}=\{d_{12}\}$.

$P(d_{22})=P(d_{12})=1$.

$K(d_{22})=K(d_{12})=\{5.\text{mpg}\}$.

$d_{23}:G(d_{23})=\{\text{TimeCmp}\}$.

$U(d_{23})=\{\text{TimeMargin}(\text{visitTime}-\text{createTime})=5\,\text{min}\}$.

$\cdot d_{23}=\{d_{13}\}$.

$P(d_{23})=P(d_{13})=1$.

$K(d_{23})=K(d_{13})=\{7.\text{mpg}\}$.

在分别对比完成后,对这几个文件类型相同的文件横向对比,可以发现,时间差值都在 10 min 之内,也就是说,相同类型的文件,时间间隔都很小.这种现象比较可疑,于是对于计算该事件的概率:

$d_{31}:G(d_{31})=\{\text{Compare}\}$.

$U(d_{31})=\{\text{TimeMargin}(\text{visitTime}-\text{createTime})<10\,\text{min}\}$.

$\cdot d_{31}=\{d_{21},d_{22},d_{23}\}$.

由于是多个文件的相互比较,因此:

$K(d_{31})=\{1.\text{mpg},5.\text{mpg},7.\text{mpg}\}$.

其概率为:$P(d_{31})=1/592704$.

图 7 给出了分析过程中节点 d_{31} 的属性值.

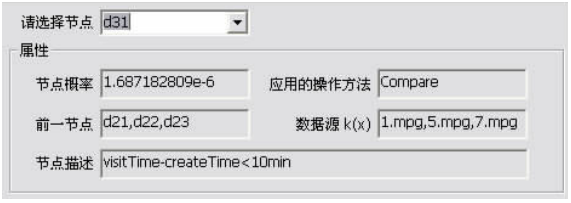

图 7 节点 d_{31} 属性窗口

通过计算概率可知概率值非常小,已小于系统所定义的概率值 δ(设 $\delta=0.00001$,正如前面分析所说,若多于两个事件,概率更小).小到如果不是张告知李,李连续多次随后访问文件几乎是不可能的事件.因此,可以断定李所说为假.

$d_t:G(d_t)=\{\text{Decision}\}$.

$U(d_t)=\{$李在张上传 10 min 内访问而张未告知李$\}$.

$\cdot d_t=\{d_{31}\}$.

$K(d_t)=\{1.\text{mpg},5.\text{mpg},7.\text{mpg}\}$.

$P(d_t)=1/592704<\delta$.

则 d_t 即为最终证据.

普通情况下，调查者只能发现“李多次在张上传后 10min 内访问该文件”这个可疑事件，需要由犯罪嫌疑人进行必要的解释，若解释不清或解释不合理，法庭不予信任，则该证据可以信任，证据有效. 然而，就证据本身而言，虽然由于证据为在磁盘上发现的文件，有效性不容质疑，但该事件发生的可能性为多少是未知的. 因此，该证据在法庭上缺乏说服力.

相比来说，本文通过将 Petri 网推导分析方法引入到可信取证模型中，达到了通过概率来衡量推导分析所获得的结论的可信程度，为案件的侦破提供了支持. 特别需要关注的是，在电子数据取证过程中，应避免一些不科学或不符合规则的取证行为或取证方法. 节点属性的变化即为某一方法操作的结果（例如，Visit 操作使得 1. mpg 的访问属性改变），而该方法在数据库中又对应着某些适用规则. 这样连接起来，每个节点在推理过程所做的变化及其所适用的有效性规则一目了然，证据的来源、有效性可以清晰地说明，法官不难判断推理过程的合理性. 再加上推理过程的可信度值，或接近于 1 的极高值表示该事件为真，或接近于 0 的值表示该事件为假（本例中结论概率接近于 0，事件为假），直观性强，证据的说服力也显而易见.

7　结束语

当前，计算机犯罪事件随着网络的普及而迅猛的发展，如何获取计算机犯罪的相关证据，实施计算机取证，积极打击计算机犯罪，成为近来计算机安全领域新的研究方向.

文章以概率论和 Petri 网为基础，结合相关知识，描述了一种基于动态取证行为可信的可信取证推理过程，也即有效性说明过程. 该过程采用了“可信度＋数据源＋取证规则”作为对所得证据的有效性进行了评估，可信度为接近 0 或 1 的概率值，用以说明该证据所描述事件的可能性，数据源则说明了得出该证据的数据源，说明该证据的存在性，而取证规则说明了由数据源至证据整个阶段所使用的规则及规则的有效性等情况. 依照本文提出的基于可信概率的有效性取证模型，证据的可信性得到了有效的评估.

本文对取证过程中的动态行为的取证进行了有效评估，这是在假设静态属性可信的前提下进行的. 电子数据静态属性可信问题在本模型中通过建立监督记录得以解决，相关问题的研究是下一步工作的重点内容. 除此之外，在现有的取证案例和取证过程中，一些事实存在的取证方法，从根源上来说就未必可信，如何寻找更好的新的取证方法去解决取证过程的可信性问题值得深入研究.

致　谢　本文的研究工作曾得到南京邮电大学王绍棣教授、江苏省公安厅科技处刘捷警官和江苏省徐州市公安局耿涛警官的悉心指导，在此表示诚挚的谢意！

参　考　文　献

[1] Sun Bo. Research on key aspects of computer forensic methods[Ph. D. dissertation]. Chinese Academy of Sciences, Beijing, 2004(in Chinese)
(孙波. 计算机取证方法关键问题研究[博士学位论文]. 中国科学院，北京，2004)

[2] Daphyne S T, Karen A F. Legal methods of using computer forensics techniques for computer crime analysis and investigation. Issues in Information System, 2004, 4(2): 692-698

[3] Liao N D, Tian S F, Wang T H. Network forensics based on fuzzy logic and expert system. Computer Communications, 2009, 32(17): 1881-1892

[4] Alink W, Bhoedjang R A F, Boncz P A et al. XIRAF—XML-based indexing and querying for digital forensics. Digital Investigation, 2006, 3(S1): 50-58

[5] Richard A, Michael T, John B. Use of data mining techniques to model crime scene investigator performance. Knowledge-Based Systems, 2007, 20(2): 170-176

[6] Florian B, Eugene S. On the role of file system metadata in digital forensics. Digital Investigation, 2004, 1(4): 298-309

[7] Sarandis M, Dimitrios P, Christos D. On incident handling and response: A state-of-the-art approach. Computers & Security, 2006, 25(5): 351-370

[8] Marc R. The role of criminal profiling in the computer forensics process. Computers & Security, 2003, 22(4): 292-298

[9] Khatir M, Hejazi S M, Sneiders E. Two-dimensional evidence reliability amplification process model for digital forensics//Proceedings of the 3rd International Annual Workshop on Digital Forensics and Incident Analysis (WDFIA'08). Malaga, Spain, 2008: 21-29

[10] Shen Tao. The forensic reliability of electronic data based on Petri net [M. S. dissertation]. Nanjing University of Posts and Telecommunications, Nanjing, 2008(in Chinese)
(申涛. 基于 Petri 网的电子数据取证有效性模型[硕士学位论文]. 南京邮电大学，南京，2008)

[11] Peter Stephenson. The application of formal methods to root cause analysis of digital incidents. International Journal of Digital Evidence, 2004, 3(1): 1-15

[12] Sun Guo-Zi, Yu Chao, Chen Dan-Wei et al. The methods of trusted forensic based on waterfall model. Netinfo Security, 2009, (7): 4-6, 20(in Chinese)
(孙国梓，俞超，陈丹伟等. 基于瀑布模型的可信取证方法. 信息网络安全，2009，(7)：4-6，20)

[13] Liang Zhi-Shun, Deng Ji-Xian et al. Probability Theory and Mathematical Statistics. Beijing: Higher Education Press, 2002(in Chinese)

(梁之舜，邓集贤等. 概率论与数理统计. 北京：高等教育出版社，2002)

[14] Yuan Chong-Yi. The Principle and Application of Petri Net. Beijing: Publishing House of Electronics Industry, 2005(in Chinese)
(袁崇义著. Petri 网原理与应用. 北京：电子工业出版社，2005)

[15] Lin Chuang, Yang Hong-Kun, Shan Zhi-Guang. Application of Petri nets to bioinformatics. Chinese Journal of Computers, 2007, 30(11): 1889-1900(in Chinese)
(林闯，杨宏坤，单志广. Petri 网在生物信息学中的应用. 计算机学报，2007，30(11)：1889-1900)

[16] Carmen-Veronica B, Eugene J H K, Hendrik V L. Modeling of discrete event systems: A holistic and incremental approach using Petri nets. ACM Transactions on Modeling and Computer Simulation (TOMACS), 2004, 14(4): 389-423

[17] Thomas B H, Saïd H, Pascal Y. Mathematical programming approach to the Petri nets reachability problem. European Journal of Operational Research, 2007, 177(1): 176-197

[18] Liu D S, Wang J M, Stephen C et al. Modeling workflow processes with colored Petri nets. Computers in Industry, 2002, 49(3): 267-281

[19] Aalst W M P. The application of Petri nets to workflow management. The Journal of Circuits, Systems and Computers, 1998, 8(1): 21-66

[20] Eoghan Casey. Error, uncertainty, and loss in digital evidence. International Journal of Digital Evidence, 2002, 1(2): 1-45

[21] Dai Jiang-Shan, Xiao Jun-Mo, Zhang Zeng-Jun. Research and design of a distributed network real forensics system. Journal of University of Electronic Science and Technology of China, 2005, 34(3): 347-350(in Chinese)
(戴江山，肖军模，张增军. 分布式网络实时取证系统研究与设计. 电子科技大学学报，2005，34(3)：347-350)

[22] Wu Yao-Rui. Research of computer forensics method based on active acquisition and implementation techniques[Ph. D. dissertation]. Jilin University, Changchun, 2009(in Chinese)
(吴姚睿. 基于主动获取的计算机取证方法及实现技术研究[博士学位论文]. 吉林大学，长春，2009)

[23] Qi Zhao-Hui. The key technology research on computer intrusion forensics[Ph. D. dissertation]. Tianjin University, Tianjin, 2006(in Chinese)
(綦朝晖. 计算机入侵取证关键技术研究[博士学位论文]. 天津大学，天津，2006)

[24] Mohamed S, Ali R A, Assaad S et al. Forensic analysis of logs: Modeling and verification. Knowledge-Based Systems, 2007, 20(7): 671-682

[25] Florian B, Eugene H S. Run-time label propagation for forensic audit data. Computers & Security, 2007, 26(7-8): 496-513

[26] Himanshu K, Jim B, Mehedi B et al. Palantir: A framework for collaborative incident response and investigation//Proceedings of the 8th Symposium on Identity and Trust on the Internet (Dtrust '09). Gaithersburg, ACM, 2009: 38-51

[27] Cai Zi-Xing, John Durkin, Gong Tao. Advanced Expert System: Principles, Design and Application. Beijing: Science Press, 2005(in Chinese)
(蔡自兴，约翰·德尔金，龚涛. 高级专家系统：原理，设计及应用. 北京：科学出版社，2005)

SUN Guo-Zi, born in 1972, Ph. D., associate professor. His research interests mainly include digital data forensics, security of network and communication.

GENG Wei-Ming, born in 1985, M. S. candidate. His research interests focus on digital data forensics.

CHEN Dan-Wei, born in 1970, Ph. D., associate professor. He is currently working on digital data forensics, embedded system.

SHEN Tao, born in 1983, M. S.. His research interests focus on digital data forensics.

Background

During digital data forensics, the validity of the evidence has its shortage, so the validity model is needed. Based on trusted probability and Petri net, one validity model for digital data forensics is introduced in this paper.

This paper is supported by the 11th Five Years Plans of the National Key Technology R&D Program (No. 2007BAK34B06), the Chinese National Natural Science Foundation (No. 61073114), the Natural Science Foundation of Jiangsu University (No. 09KJD520007), the Climbing Program of Nanjing University of Posts and Telecommunications (No. NY208009), and the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

During the No. 2007BAK34B06 program, the authors' main work is to provide effective tools of digital data forensics for the department of computer forensics. This work provides the necessary theoretical and practical basis for the validity model of the digital data forensics. In the No. 61073114 program, the authors' main work is doing some formalized analysis, and the formalization methods have provided theoretical support for this work, such as the basic definition and formalization processing in the process of digital forensics. Otherwise, the authors study the key technology of trusted forensics and formalization analysis, decomposes trusted forensics into trusted static attribute and trusted dynamic behavior, and research the related theories and methods of formalization analysis. This article is also part of the work of No. 09KJD520007 and No. NY208009.