第 35 卷 第 8 期 计 算 机 工 程 2009 年 4 月
Vol.35 No.8 Computer Engineering April 2009
·博士论文· 文章编号：1000—3428(2009)08—0020—03 文献标识码：A 中图分类号：TP309

一种计算机数据取证有效性的证明方法

陈丹伟[1,2]，孙国梓[1,2]，唐 娟[1,2]，王海平[1,2]
(1. 南京邮电大学计算机学院，南京 210003；2. 南京邮电大学计算机技术研究所，南京 210003)

摘 要：简述计算机数据取证的基本要求，给出一种计算机数据取证有效性的证明系统，对电子数据取证有效性理念及其体系进行研究。通过对取证方法有效性和所取数据有效性进行一系列的定义和推导，研究计算机数据取证有效性的一种形式化证明方法。利用上述证明方法对一个计算机取证实例进行取证有效性的形式化证明。

关键词：计算机数据；取证；有效性；计算机犯罪

Proving Method for Validity of Computer Data Forensics

CHEN Dan-wei[1,2], SUN Guo-zi[1,2], TANG Juan[1,2], WANG Hai-ping[1,2]
(1. College of Computer, Nanjing University of Posts & Telecommunications, Nanjing 210003;
2. Institute of Computer Technology, Nanjing University of Posts & Telecommunications, Nanjing 210003)

【Abstract】After describing the basic demands of computer forensics, this paper puts forward a system of proving the reliability of computer forensics. Thought of digital forensics and its system is studied. With the definitions and illations of the forensic methods and gained data, it investigates a formalized method of proving the validity of computer forensics. The method is put forward to proving the reliability of a computer forensics as an example.

【Key word】computer data; forensics; validity; computer crime

1 概述

随着互联网和计算机的发展，计算机犯罪逐渐引起人们的广泛关注。在计算机犯罪中，计算机被用作犯罪的工具或成为犯罪侵害的目标[1]。为了打击犯罪分子，遏制计算机犯罪的趋势，计算机数据取证应运而生。

对于计算机数据取证，一直都没有统一的定义。可以把它简单地表述为[2]：使用计算机软件和工具进行分析、研究，从特定活动中寻找和提取法律证据的过程。计算机取证必须小心谨慎地遵循一定的取证准则，才可以在最大程度上保证取证的真实性、客观性和全面性。一般有以下 5 点要求[3]：(1)保护所要取证的目标计算机系统；(2)尽可能发现目标系统中的所有文件，包括恢复被删除文件、查找隐藏文件、进入缓存或交换区文件、打开受保护或已加密的文件等；(3)尽可能分析所有相关资料；(4)提交全面分析的结果；(5)给出相关专家意见或证明。

2 数据取证有效性证明系统的体系结构

2.1 问题的提出

对计算机数据取证的研究重点一直都放在计算机数据的获取和对所取得数据的分析上。对于数据取证的有效性还缺乏相应的论证，这样取得的数据容易受到置疑，其采信度也会大打折扣。因此，本文将重点放在证明体系结构的提出和计算机数据取证有效性的证明上。

2.2 体系结构的建立

根据以上对计算机数据取证的的相关分析，本文构建了计算机取证有效性证明的体系结构，描述了体系中各功能模块的作用及模块之间的关系。

图 1 给出了一个计算机数据取证有效性证明系统的体系结构。

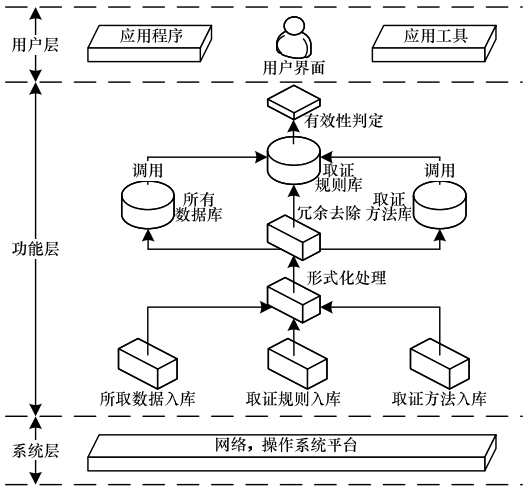

图 1 计算机数据取证有效性证明系统

(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。系统层包括的网络和操作系统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)功能层

功能层是整个有效性证明系统的核心，其中最重要的部分是有效性判定模块。在功能层中，首先实现的是取证规则，

基金项目：国家自然科学基金资助项目(60703086)；国家“863”计划基金资助项目(2004BA811B04)
作者简介：陈丹伟(1970－)，副教授、博士，主研方向：信息安全；孙国梓，副教授、博士；唐 娟，硕士；王海平，讲师、硕士
收稿日期：2008-09-22 **E-mail：**chendw@njupt.edu.cn

某次是取证所用到的取证方法和所取得数据的入库操作；数据入库后有一个形式化处理的过程，考虑到该系统的适用范围，采用产生式规则表示方法来描述形式语言的语法，模拟人的认知和判断过程；在形式化处理后，需要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；最后根据取证规则对取证方法和所取得数据的有效性进行判定，综合 2 种判定结果，从而得到最终结果。

(3)用户层

用户层包括用户接口层和各类应用工具层。各种不同的用户界面是用户与系统之间联系的纽带，包括图形化浏览器、菜单、对话框等。

2.3 取证有效性判定模块

有效性判定模块是取证有效性形式化证明中最关键的模块。对取证有效性判定采用对取证方法和数据先分别判定、再综合判定的方法。有效性判定模块对于取证方法和所取数据具有通用性。判定的过程借用了人工智能中专家系统的研究方法，在传统的专家系统基础上，结合实际做了相应的演变。传统的专家系统模型[4]如图 2 所示。演变后的有效性判定模块如图 3 所示。

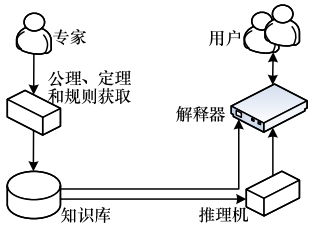

图 2　专家系统结构

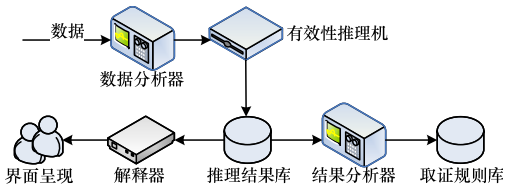

图 3　有效性判定模块结构

3　取证有效性证明的理论准备

所取的计算机数据要能够被法庭采信，作为办案的有效依据，必须确定计算机数据取证的有效性。而国内外关于有效性判定方面的研究还比较少。

本文借鉴刑事诉讼法关于传统证据在何种情况下可以作为定案证据的规定，认为对计算机数据取证有效性的证明过程就是一个对取证过程和取证方法的“查证”过程。“查证”体现到计算机数据上，则表现为证明计算机数据的完整性和与原始数据的一致性。完整性可理解为所取的数据与原始数据相比，没有增加或删减内容；一致性可以理解为对于同一数据，原始数据与所取数据在内容上完全相同。

本文所要讨论的是，对于通过一定方法取得的数据，如何通过形式化的方法证明其有效性。

有效性证明过程中的定义、公理、定理及规则如下。

定义 1 原始数据。指计算机或计算机网络中包括的所有数据。

定义 2 所取数据。指运用一定的取证手段从计算机中取得的数据。

定义 3 计算机数据取证。指通过一定的合乎取证规范的技术手段，从计算机和计算机网络中获取各种数据、为计算机犯罪调查提供证据的证据获得过程。

定义 4 计算机数据取证规范。指进行计算机数据取证过程中必须遵守的一系列行为准则。

定义 5 计算机数据取证有效。指所取得的计算机数据具有完整性与原数据的一致性。

定义 6 计算机数据取证有效性证明系统可以看作一个元组：

$M=\{D', D, S, A, M, C, N, I\}$

其中，D'是所有可疑计算机或网络中包含的所有计算机数据，即原始数据集合，有 $D'=\{d_0',d_1',\cdots,d_m'\}$；$D$ 是通过一定的取证方法取得的数据，即所取得数据的集合，有 $D=\{d_0,d_1,\cdots,d_n\}$；S 是已知的计算机犯罪类型，有 $S=\{s_0,s_1,\cdots,s_r\}$；$A\subseteq D\times D$ 是状态转换函数，即分析方法集合，有 $A=\{a_0,a_1,\cdots,a_p\}$；C 是所确定数据的特征集合，有 $C=\{c_0,c_1,\cdots,c_k\}$；M 是取证方法集合，有 $M=\{m_0,m_1,\cdots,m_l\}$；N 是取证规范集合，有 $N=\{n_0,n_1,\cdots,n_p\}$；I 为某个犯罪对于计算机的影响，$I=\{i_0,i_1,\cdots,i_n\}$。

定义 7 设 D, C 均为非空集合，$P(C)$表示集合 C 的幂集，$F\subseteq D\times P(C)$，则定义 D 和 $P(C)$的合成如下：

$F=F\{f_0,f_1,\cdots,f_i\}=\{(d_i,C_i)\mid \exists\, i,0\leqslant i\leqslant n\}$

其中，F 为所取得数据 d_i 所具有的特征集合。

定义 8 设 D, M 均为非空集合，$P(M)$表示集合 M 的幂集，$T\subseteq D\times P(M)$，则定义 D 和 $P(M)$的合成如下：

$T=T\{t_0,t_1,\cdots,t_i\}=\{(d_i,M_i)\mid \exists\, i,0\leqslant i\leqslant n\}$

其中，T 为取得数据 d_i 所用的取证方法集合。

定义 9 设 M, N 均为非空集合，$P(M)$表示集合 M 的幂集，$P(N)$表示集合 N 的幂集，$P\subseteq P(M)\times P(N)$，则定义 $P(M)$和 $P(N)$的合成如下：

$P=P\{p_0,p_1,\cdots,p_i\}=\{(M_i,N_i)\mid \exists\, i,\ 0\leqslant i\leqslant 2^l-1\}$

其中，P 为取证方法集 M_i 中元素所用的取证规范集合。

定义 10 设 S, M 均为非空集合，$P(S)$表示集合 S 的幂集，$P(M)$表示集合 M 的幂集，$Q=P(S)\times P(M)$，则定义 $P(S)$和 $P(M)$的合成如下：

$Q=Q\{q_0,q_1,\cdots,q_i\}=\{(S_i,M_i)\mid \exists\, i,\ 0\leqslant i\leqslant 2^r-1\}$

其中，Q 为取得已知计算机犯罪集 S_i 证据的取证方法集合。

定义 11 设 S, I 均为非空集合，$P(S)$表示集合 S 的幂集，$P(I)$表示集合 I 的幂集，$Y=P(S)\times P(I)$，则定义 $P(S)$和 $P(I)$的合成如下：

$Y=Y\{y_0,y_1,\ldots,y_i\}=\{(S_i,I_i)\mid \exists\, i,\ 0\leqslant i\leqslant 2^r-1\}$

其中，Y 为已知计算机犯罪集 S_i 对于计算机的影响集合。

定义 12 设 I, M 均为非空集合，$P(I)$表示集合 I 的幂集，$P(M)$表示集合 M 的幂集，$Z=P(I)\times P(M)$，则定义 $P(I)$和 $P(M)$的合成如下：

$Z=Z\{z_0,z_1,\cdots,z_i\}=\{(I_i,M_i)\mid \exists\, i,\ 0\leqslant i\leqslant 2^n-1\}$

其中，Z 为取得已知犯罪对计算机的影响集 I_i 的取证方法集合。

定义 13 取证开始前，可疑计算机的状态有以下 4 种：S(关闭)，R(运行)，C(接入网络)，$\sqcup C$(未接入网络)。

定义 14 计算机取证的动作为 E。

定义 15 事件 A 在情况 B 下发生，表示为 $A\mid B$。

公理 1 $\exists\, d_i\in D$, $d_j=copy(d_i)$，则 $d_i.lmt=d_j.lmt$, $d_i.lvt\neq d_j.lvt$, $d_i.et\neq d_j.et$;(lmt:最后一次更改的日期时间；lvt：最后一次访问时间；et：创建的日期时间)。

公理 2 $\exists d_i \in D, d_j = open(d_i)$,则 $d_i.lmt = d_j.lmt$, $d_i.lvt \neq dj.lvt$, $d_i.et = d_j.et$。

公理 3 $\exists d_i \in D, \exists d_i' \in D'$，若 $E(d_i) = d_i'$，则 E 有效。

定理 1 $\exists d_i \in D, \exists t_i \in T, \forall m_i \in M_i$，有 $m_i \uparrow N$, 则 d_i 有效；$\exists m_i \in M_i$，有 $m_i \downarrow N$，则 d_i 有效性不确定（↑表示符合，↓表示不符合)。

证明：因为 $m_i \uparrow N$，根据 N 的定义，符合 N 的 m_i 均可以保证取证的完整性、一致性，所以 d_i 有效。

反之，若 $m_i \downarrow N$，则 d_i 有效性不确定。

定理 2 $\exists d_i \in D, \exists t_i \in T, \exists m_j, m_k, \cdots \in M_i$, $j \neq k \neq \cdots$, $m_j \uparrow N$, $m_k \uparrow N$, …，则 d_i 有效。

证明：假设 d_i 有效性不确定，不妨设 $i=1$，用方法 m_0 取得的数据为 d_0，用方法 m_1 取得的数据为 $d_1, d_0 \neq d_1$。

根据 T 的定义及定理 1，有 $m_0 \uparrow N, m_1 \uparrow N$，则 d_0 和 d_1 均有效，即 $d_0 = d_1$。

由此假设不成立，即 d_i 的有效性确定。

同理可以证明 $i=2,3,\cdots,n$(n 为自然数)的情况。

定理得证。

规则 1 $\exists d_i, d_j \in D, \exists f_i, f_j \in F, i \neq j$, $f_i.C_i \sqsubset f_j.C_j$，则 d_i, d_j 有效。

规则 2 $\exists d_i \in D, \exists (d_i, C_i) \in F, C_i = \{c_{i0}, c_{i1}, \cdots\}$，若 $\exists c_{im} \sqsubset c_{in}, m \neq n$，则 d_i 有效。

规则 3 $\exists d_i \in D$，若 $E(d_i) \mid C$，则 d_i 有效性不确定；若 $E(d_i) \mid \sqcup C$，则 d_i 有效。

规则 4 $\exists d_i' \in D'$, $\exists d_i \in D, D', D \neq NIL$，若 $E(d_i) \mid R$，保存缓存数据 $\in E$，则 $D' \equiv D$，且 d_i 有效。

规则 5 $\exists d_i' \in D'$, $\exists d_i \in D, D', D \neq NIL$，若 $E(d_i) \mid R$，$S \xrightarrow{\text{自备无病毒启动盘}} R$，则 $D' \equiv D$，且 d_i 有效。

规则 6 $\exists s_i \in S$，且已知 s_i 对应的 Q_i，若取证过程中，Q_i 每个元素被穷举，则取证有效。

规则 7 $\exists s_i \in S$，$\exists z_i, z_j \in Z, \exists d_i' \in D'$, $\exists d_i, d_j \in D$, $d_i' \xrightarrow{z_i.M_i} d_i, d_i' \xrightarrow{z_j.M_j} d_j$，若 $d_i = d_j$，则取证有效。

规则 8 对于可能包含 D'的所有地方，都要尽可能搜查、收集。

规则 9 $\exists d_i \in D, d_j = copy(d_i)$，则 $d_j.content = d_i$；取证有效。

规则 10 $\exists d_i, d_j \in D, MD5(d_i) = MD5(d_j)$，则 $d_i \equiv d_j$；取证有效。

规则 11 $\exists d_i' \in D', d_i = mirror(d_i')$，则 $d_i \equiv d_i'$；取证有效。

规则 12 $\exists d_i' \in D', \exists d_i \in D, d_i' \xrightarrow{M_i} d_i$, $F\{(d_i, C_i)\} = F\{(d_i', C_i')\}$，则该次计算机取证有效。

规则 13 $\exists d_i \in D, \exists a_j, a_k \in A$，若 $a_j(d_i) \leftarrow a_k(d_i)$，则 a_j, a_k 有效。

以上一系列的形式化推导是以定义为基础，把在取证操作中可能出现的行为、情况分别形式化地规定为公理和规则，并在此基础上进行形式化推导，从而得出更丰富的定理，充实有效性的判定规则系统。

4 取证有效性形式化证明实例

下面通过华盛顿大学信息安全服务小组专家 Dave Dittrich 的一个计算机取证实例[5]，运用前面所给出的公理、定理和规则进行计算机取证有效性的形式化证明。

Dave Dittrich 在有证据显示计算机被入侵时首先马上切断系统电源，其次对原始硬盘进行物理拷贝，同时用 MD5 对原始硬盘上的数据做摘要，保存原始证据和摘要信息。在取证过程中，他对取证中的每一个步骤和发现都进行详细的记录，具体步骤如下：

(1)将物理复制的磁盘以只读方式安装在取证系统上。首先使用标准的 Unix 工具进行分析，发现了可疑账号和与这些账号相关的可疑文件，找到了安装黑客文件的证据及黑客文件。

(2)使用 TCT(The Coroner's Toolkit)工具对系统中所有文件按 MAC 时间进行排序，并恢复出所有被删除的文件，寻找线索。

(3)汇总所有证据，形成取证报告。后续的调查也证明，这次取证为确定犯罪嫌疑人提供了进一步的证据。

首先证明取证过程的有效性。在取证过程中，计算机的初始状态为 R，取证人员首先切断了电源，若采取正常方式关闭 Unix 系统，系统的缓存文件将被清除。采用断电的非正常关闭系统方式则不会丢失这些文件。这样可以保证接下来要取得的数据与原始数据一致。满足公理 3 及规则 4，方法有效。取证人员对原始数据进行物理拷贝。在此次取证过程中，重点是要发现入侵者是谁，根据公理 1 及规则 10，拷贝中会改变的文件属性对最后结果无影响，方法有效。在拷贝的同时，用 MD5 算法做了摘要，满足规则 10，方法有效。

在对证据的分析过程中，取证人员以只读方式分析所取证据，满足公理 3，分析过程有效。同时取证人员先后应用了标准 Unix 工具和 TCT 工具对所取数据进行分析，2 种工具的分析方法互补，满足定理 2 和规则 13，分析过程有效。

综合对取证过程及分析过程的形式化分析，可知取证方法和所取数据均有效，因此，可得此次计算机取证有效。事实也证明此次取证达到了要求，确定了犯罪嫌疑人。

5 结束语

计算机取证的形式和方法虽然根据不同的取证需要变化很多，但万变不离其宗。计算机取证就是为了给犯罪调查提供有效的证据，而取证的有效性也可以用抽象化的形式描述和证明。本文仅仅是一个开始，如何深化计算机数据取证有效性的形式化证明，还有待进一步研究。

参考文献

[1] Kruse W G, Heiser J G. Computer Forensics: Incident Response Essentials[M]. 北京: 人民邮电出版社, 2003.

[2] 刘宝旭, 马建民, 池亚平. 计算机网络安全应急响应技术的分析与研究[J]. 计算机工程, 2007, 33(10): 128-130.

[3] Robbins J. An Explanation of Computer Forensics[EB/OL]. (2008-01-02). http://www.computerforensics.net/forensics.htm.

[4] 尹朝庆, 尹 皓. 人工智能与专家系统[M]. 北京: 中国水利水电出版社, 2002.

[5] Dittrich D. Basic Steps in Forensic Analysis of Unix Systems[EB/OL]. (2007-09-10). http://staff.washington.edu/dittrich/forensics/.

编辑 张正兴