计算机证据元数据表示方法*

杨 珺[1],李 晶[1],王 敏[2],陈 晨[1],廖伟辉[1]
(1.武汉大学 电子信息学院,湖北 武汉 430079;
2.通信指挥学院 二系,湖北 武汉 430010)

摘 要:针对计算机证据格式繁杂不利于形成证据链的问题,提出了计算机证据元数据表示方法。该方法将计算机证据的描述层次划分为数据表示层、证据表示层、事件表示层和案件表示层,并在这些层次上分别采用证据元数据来描述计算机证据。通过对计算机异常事件证据元数据的设计,实现对计算机证据的内在属性和关系进行统一的表达,能够方便地组织、分析、融合和提交计算机证据。

关键词:计算机安全;电子犯罪对策;计算机取证;证据;元数据

中图分类号:TP309 **文献标识码**:A

Notation for computer evidential metadata

YANG Jun[1], LI Jing[1], WANG Min[2], CHEN Chen[1], LIAO Wei Hui[1]
(1.School of Electronic Information, Wuhan University, Wuhan 430079, China;
2.Second Department, Commanding Communications Academy, Wuhan 430010, China)

Abstract: In view of the difficulty of constructing evidential chain due to the miscellaneous format of the computer evidence, a notation method for the computer evidential metadata was proposed. The objects were described by different form of computer evidential metadata in the data-level, the evidence-level, the event-level and the case-level. The computer evidential metadata of computer abnormal event was designed. The results indicate that the attributes and relations of computer evidence can be expressed uniformly, and the computer evidence can be organized, analyzed, melted and submitted conveniently.

Key words: computer security; electronic crime countermeasures; computer forensics; evidence; metadata

目前,计算机犯罪活动日趋猖獗,因此有效地组织计算机犯罪证据信息,形成计算机证据链是一项非常重要的工作。计算机证据来源众多,格式繁杂[1-2]。一次入侵事件的证据可能留存于网络、网络设备以及计算机系统中,这些记录了入侵者的入侵手段、入侵时间和入侵结果的证据相互间存在着紧密的联系,为了能够对这些格式不尽相同却具有相关性的证据进行有效地组织、分析、融合和提交,迫切需要对它们进行统一的表示。元数据(Metadata)是描述数据的数据[3],是对信息对象编码式的结构化描述,它主要用来描述数据的内容、质量、所有者、提供方式、覆盖范围以及管理方式等,是数据与数据用户之间的桥梁。

本文拟用元数据对计算机证据进行描述,以便计算机证据元数据能够统一地表达计算机证据的内在属性及其相互间的关系,为形成计算机证据链奠定良好的基础。

1 计算机证据元数据

1.1 计算机证据

计算机证据也称电子证据,是指在计算机系统和网络设备运行过程中产生的能够表征某种事件事实的数据集合。

计算机证据源分为计算机系统证据源和计算机网络证据源。计算机系统证据源包括:系统日志文件、数据文件、内存映像文件、临时文件、空闲磁盘空间等;计算机网络证据源包括:网络数据包、杀毒软件日志文件、防火墙日志文件、入侵检测系统日志文件、各种服务器日

* 基金项目:高等学校博士学科点专项科研基金(20040486049)

志文件等。

1.2 计算机证据元数据

目前,元数据还没有统一的定义,一些学者将元数据简单定义为:关于数据的数据[4],而该定义无法清晰地反映元数据的内涵。于是,不同领域的学者从不同角度对该定义进行了扩展和深化[3]。当前,在计算机取证领域中,对计算机证据元数据的定义还未见报道。

通过分析和总结元数据在其他领域中的定义,本文将计算机证据元数据定义为:一种构建在计算机证据基础上具有统一格式的结构化数据,它是计算机证据的结构化描述。其目的是描述计算机证据的基本特征、基本属性和相互关系,以便那些格式不同的相关证据能够进行有效地组织、分析、融合和提交。

1.3 计算机证据元数据特点

作为对计算机证据结构化描述的一种方式,计算机证据元数据的基本特点为:

(1)描述性:计算机证据元数据按照规则描述对象,并以此组织和管理证据资源。

(2)动态性:计算机证据元数据会随着描述对象的变化而变化。

(3)多样性:从不同角度对描述对象的特征进行划分并产生多种计算机证据元数据表示形式。

(4)复杂性:计算机证据元数据可以嵌套,既是元数据的集合,也是可选择性的元数据。

(5)多层性:计算机证据元数据的描述对象可以是多层次的。

(6)支撑性:计算机证据元数据能够有效地维护描述对象的原始性和完整性,能够与描述对象共存。

2 计算机证据元数据表示

2.1 计算机证据元数据表示原则

本文依据以下原则对计算机证据元数据进行描述。

(1)模块化:根据内容将计算机证据划分为不同模块,各个模块分别采用不同类型的计算机证据元数据表示,以满足不同的应用需求。

(2)一致性:计算机证据元数据的描述应尽量与元数据的国际标准、国家标准或行业标准保持一致。

(3)可扩展:通过复用、嵌接、延伸、细化、修改等方式,构建新的计算机证据元数据;预留计算机证据元数据的元素空间以适应未来的需求。

(4)稳定性:将基本的、共同的和必需的内容归纳为1个核心元素集,核心元素集具有相对的稳定性,能够满足基本的应用需求。

(5)互操作:计算机证据元数据应支持异构系统间的互操作,它可在不同系统间实现传输、交换或转换。

(6)递归性:计算机证据元数据能被逐层地描述、定义、确认和验证。在每个层次上,计算机证据元数据均具有独立元素。

(7)开放性:一旦出现新的计算机证据源,即可设计出相应的计算机证据元数据。

2.2 计算机证据元数据的表示层次

根据计算机取证分析递进构造证据链的特点以及计算机证据元数据的特点和表示原则,将描述层次划分为数据表示层、证据表示层、事件表示层和案件表示层,并在这些层次上分别采用计算机证据元数据对描述对象进行描述,它们的相互关系如图1所示。

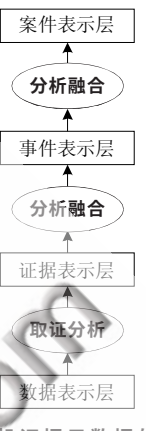

图1 计算机证据元数据的表示层次

数据表示层的任务是描述原始的取证数据。如,对于系统日志文件、防火墙日志文件、数据文件等具有严谨记录格式的原始取证数据,可借鉴其本身的描述格式来构建相应的计算机证据元数据;对于网络数据包,可参照TCP/IP协议树和网络数据特征属性来构建相应的计算机证据元数据。

证据表示层的任务是描述取证分析后的数据。经取证分析方法得到的证据通常与取证数据具有相同的格式,因此在证据表示层可采用与数据表示层相同或相似的计算机证据元数据描述。

事件表示层的任务是描述计算机异常事件的所有证据,即在事件表示层,对计算机异常事件的所有证据进行统一的规范化描述和表示,是对证据的进一步分析、关联和融合的过程,是形成计算机证据链的核心和关键步骤。

案件表示层的任务是描述计算机案件(一系列计算机异常事件)的所有证据,是提交证据、形成调查报告的基础。

2.3 事件表示层的描述

通常,一个事件由事件主体、事件目标、事件时间、事件地点以及事件操作来描述。异常事件元数据根据计算机证据元数据表示原则,并参考通用入侵检测对象GIDO(Generalized Intrusion Detection Objects)中的事件描述类[5]、事件对象描述、交换格式IODEF(Incident Object Description and Exchange Format)中的若干事件类[6]来设计,该描述如表1所示。

表中省略了部分同类的、繁冗的数据项子项和子元素,将描述事件地点的数据项分布到各个相应的数据项子项或子元素中,并用位置节点标示,增加了复合事件描述,复合事件是单一事件的集合运算(并、与、异或等运算),它复用了单个事件的一般性描述,保持了元数据

表 1 异常事件元数据表示

数据项	子项	子元素	描述
一般性描述	事件标识		ID，数据库表中唯一
	事件类别		网络、系统
	事件历史		相同事件统计次数
	事件评级		事件紧急程度
	事件来源		传感器标识
事件主体	位置节点	节点名称	网络节点(外部,内部)
		节点地址	IP 地址或域名
	身份鉴别	身份名称	攻击者、嗅探者等
		身份描述	上述身份描述
	……	……	……
事件目标	文件对象	文件标识、文件属性描述、文件位置节点	建立各类事件目标对象 ID，并进行属性描述(通常为对象定义)，其位置节点复用事件主体中的位置节点(Node)
	进程对象	进程标识、进程属性描述、进程位置节点	
	应用程序对象	程序标识、程序对象描述、程序位置节点	
	用户对象	系统用户标识、用户属性描述、用户帐户位置节点	
	网络对象	网络对象标识、网络对象属性描述、网络对象位置	
	信息资源对象	资源标识、资源属性描述、资源位置节点	
	……	……	……
事件时间	事件起始时间		采用国际通用标准时间，精度随应用而定
	事件结束时间		
	事件检测时间		
	事件报告时间		
	……	……	
事件操作	文件对象操作		复制、移动、删除等
	进程对象操作		执行、暂停、中断、终止等
	网络对象操作		扫描、连接、断开等
	应用程序操作		打开、关闭等
	用户权限操作		登录、退出、修改等
	……	……	……
事件复合关联	关联事件集合		相关事件 ID
	事件关联类别		时序关联、因果关联等
	事件关联描述		关联表达式
	……	……	……

描述结构上的完整性。

元数据是一种基本信息组织方法，是信息的标准化表示，它能够为信息系统各个层次的内容提供规范的定义、描述、交换和解析，能够为分布的、复杂的信息系统提供互操作和整合平台，可以为计算机智能识别、处理、集成各种信息提供工具。

采用计算机证据元数据对计算机证据进行统一描述有利于计算机证据的组织、分析、融合和提交，有利于计算机证据链的形成。

参考文献

[1] 殷联甫.计算机取证技术[M].北京：科学出版社，2008.

[2] 宁勇.电子证据的基本问题与取证初探[D].北京：清华大学，2004.

[3] 许永涛.基于 E-R-P 建模体系的政务信息资源元数据模型与应用研究[D].大连：大连理工大学，2008.

[4] HILLMANN D.Usingg Dublin Core [EB/OL].(2001-04-12)[2008-12-20].http://www.dublincore.org/documents.

[5] A common intrusion specification language [EB/OL]. (1999-6-11)[2008-8-12]. http://gost.isi.edu/cidf/drafts/ language.txt.

[6] RFC5070: The incident object description exchange format [EB/OL].(2007-12-1)[2009-3-11]. http://www.rfc-editor.org/rfc/ rfc5070.txt.

(收稿日期：2009-06-10)

(上接第 62 页)

报的各项数据的准确性、完整性及一致性，防止出现竣工资料与实际情况不符、管道线位不准、管道技术数据错误等问题，为西部管道的运营管理提供真实准确的数据资料。管理者可以通过网络查看不同比例管道及其沿线周边环境的直观信息，每根钢管都有完整的数据记录，可查看某一天、某一道工序环节的进度，甚至每道焊口的坐标值及埋深、焊工信息、无损检测影像等基本信息。这些数据全过程地记录在“数字化管道”系统中，一旦出现问题，即可进行可逆性数据跟踪，查出源头。

(4)在工程管理方面，实现对工程参建队伍、各标段基本情况、施工计划、施工进度等整体管理，发布工程建设的重要信息。公司内部及与参建单位之间实现电子公文收发和无纸化办公。自动汇总生成工程进度、质量等过程控制报表，并用地理信息系统等图形化的方式展示。通过设计数据与施工过程数据的对比查询功能，保证施工单位严格按照设计线路走向施工，数字化系统协助建设单位有效控制工程变更。

(5)在管道巡检方面，运用 GPS 巡线系统，结合管道坐标数据，设定科学合理的巡检路线和计划，实现对巡线人员的实时监督管理，准确及时地定位发现事故点位置，提高维抢修反应效率。

(6)在数据应用方面，分别为西部原油成品油管道 LMS 系统、西气东输二线设计、永登支线设计、新疆自治区地震局等提供管道中线成果、里程、高程等数据。为加快工作进度提供了有力支持。

“数字化管道”是应用信息技术提升管道工程建设和运营管理水平的大课题，本文仅是结合西部管道工程开展“数字化管道”建设的实践经验，进行了一些总结和思考。随着信息技术在管道建设和运营阶段应用的不断深入，对传统管道建设和管理方法将带来一次深层次的变革。

参考文献

[1] 杨祖佩，王维斌.油气管道完整性管理体系研究进展[J].油气储运，2006，25(8).

(收稿日期：2009-06-04)