The Necessity of Developing a Digital Evidence Ontology

Jasmin Ćosić
IT Section of Police administration
Ministry of Interior
77000 Bihać, B & H
jascosic@bih.net.ba

Zoran Ćosić
"STATHEROS", d.o.o.
Županjska 35
21216 Kaštel Stari, Split, Croatia
zoran.cosic@statheros.hr

Abstract. *The aim of this paper is to highlight the problems encountered by investigators in the pursuit of forensic investigations of digital devices, primarily because of misunderstanding or false understanding of certain important concepts. An ontology of digital evidence was proposed as one of possible method suitable as a solution of this problem.*

Keywords. digital evidence, chain of custody, ontology, DEMF, DCoCDE-on

1 Introduction

There are so many definitions of digital forensic and digital evidence. One of many definitions is „digital forensic can be defined as the application of science and engineering to the legal problem of digital evidence".[1] On the question "What is Digital Forensics?" Pollitt highlighted in [2] that digital forensics is not an elephant, it is a process and not just one process, but a group of tasks and processes in investigation. Digital evidence is defined as any data stored or transmitted using a computer that support of refute a theory of how an offense occurred or that address critical elements of the offense such as intent or alibi.[3] The definition proposed by the Standard Working Group on Digital Evidence (SWGDE) is any information of probative value that is either stored or transmitted in a digital form. [4] Another definition proposed by the International Organization of Computer Evidence - IOCE is „...information stored or transmitted in binary form that may be relied upon in court". [5], [6]

In all phases of forensic investigation, digital evidence is susceptible to external influences and coming into contact with many factors. Legal admissibility of digital evidence is the ability of that evidence to be accepted as evidence in a court of law. The evidential weight of digital evidence can only be safeguarded if it can be proven that the records are accurate i.e. by whom they were created and when and that no alteration has occurred. In order for the evidence to be accepted by the court as valid, chain of custody for digital evidence must be kept, or it must be *known who exactly, when and where* came into contact with evidence in each stage of the investigation. [7]

For the paper purposes "chain of custody" and "chain of evidence" would be considered like synonyms. The phrase "chain of custody" or "chain of evidence" refers to the accurate auditing control of original evidence material that could potentially be used for legal purposes. [8] Some authors use a term „chain of evidence" instead of chain of custody. The purpose of testimony concerning chain of custody is to prove that evidence has not been altered or changed through all phases, and must include documentation on how evidence is gathered, how was transported, analyzed and presented. Knowing the current location of original evidence, is not enough for court, there must be accurate logs tracking evidence material at all time. Access to the evidence must be controlled and audited. To prove the chain of custody, we must know all the details on how the evidence was handled every step of the way. The old formula used by police, journalists and researchers - Who, What, When, Where, Why, and How - "Five Ws" (and one H) can be applied to help in digital forensic investigation. [9] [10][11]

The authors in previous studies [11] attempted to make ontological approach to help better understand and clearly define the concept in a chain of digital evidence field. The aim was to set up a taxonomy diagram of a chain of digital evidence in all phases of forensic investigation. The reasons for this are many, methods of crimes are changing from the year to year, daily appears new data carriers which may contain digital evidence, all of them is harder and harder to find. Preserving the chain of evidence has become almost impossible without explicit knowledge of the problem domain. The authors have attempted to allow "reuse" of knowledge from the domain of digital forensics and digital chain of evidence, but it made the first step towards creating an open framework for the secure management with digital evidence.

2 Basic concept of ontology

According to Gruber [12] ontology is explicit specification of a conceptualization process. The term is borrowed from philosophy, where ontology is a systematic accounting of existence. In recent years the development of ontology's-explicit formal specifications of the terms in the domain and relations among them (Gruber 1993) has been moving from the realm of Artificial-Intelligence laboratories to the desktops of domain experts.[13]
Ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them.
The Artificial-Intelligence literature contains many definitions of ontology. For the purposes of this guide ontology is used like a formal explicit description of concepts in a domain of discourse - classes (sometimes called concepts), properties of each concept describing various features and attributes of the concept (slots, roles or properties), and restrictions on slots (facets, role restrictions). An ontology together with a set of individual instances of classes build a knowledge base.[13]
On the question "Why would someone want to develop an ontology? " [13] gave some of the reasons:

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

Sharing common understanding of the structure of information among people or software agents is one of the more common goals in developing ontology's.[12][14]

3 Ontology in digital forensic

There is a lack of scientific paper about using domain ontology in digital forensic field. Reasons for this is a multidisciplinary field of digital forensics, because knowledge of the technical aspects are not enough, it is necessary to know the law - legal aspects and implications of the process of presenting digital evidence in court. Some authors in scientific papers tried to present the groundwork for the "ontology of cyber forensics, digital forensics" and "ontology of small-scale devices". The aim was to define the basic concepts and create a new approach to the study of the scientific field.

Heum Park et al. in his paper Cyber forensic ontology for cyber criminal investigation [15] develop Cyber Forensic Ontology for the cyber investigation in cyber space. Cyber crime is classified into two classes - cyber terror and general cyber crime. Those two classes are connected with each other. Investigation of cyber terror requires high technology, system environment and experts. General cyber crime is connected with general crime by evidence (digital evidence). Authors defined the concepts and relations among crime types, evidence collection, criminals and crime case and law. The limitation of this ontological model is that it is less based on digital evidence and other phases that are important in the process of digital investigation and it is related to dealing with digital evidence. The only stage in the process of dealing with digital evidence, which authors mention is "collection", while they ignored all other phases (identification, searching, transporting, storing, examination, analysis and presentation).

David Christopher Harrill and Richard P. Mislan [16] presented small scale digital device forensics ontology in 2007, in order to develop an ontological to provide law enforcement with the appropriate knowledge regarding the devices found in the SSDD (Small Scale Digital Devices) domain. The paper categorized SSDDs according to certain criteria and gave detailed description of each of them. The purpose of this paper was to provide a guiding framework in which to place small scale digital devices. According to authors this ontology can be used as a method to further develop a set of standard and procedures at which to approach SSDD.

Ashley Brinson et al. [17] in 2007 developed the cyber forensic ontology for the purpose of finding the correct layer for specialization, certification and education within the cyber forensic domain. Topic of cyber forensic consisted of two subtopics: technology and profession. Technology subtopic is broken down into hardware and software. Profession side is broken down into law, academia, military and private sector. Hardware section of his model is broken up five different parts: large scale digital devices, small scale digital devices, computers, storage devices and obscure devices. The software section of his model contains three categories: analysis tools, operating system and file system. The law section focuses on law enforcement and courts and legal aspects of cyber forensic. Profession academia is broken down in research and education, while a military categories focuses on what cyber forensic duties military personnel perform. Military section can be defensive and offensive. Private sector was broken down into consulting and industry. This ontological model can be utilized for the purpose of curriculum development.

DIALOG: A framework for modeling, analysis and reuse of digital forensic knowledge by Kahvedzic and Kechadi [18] provides a general, application independent vocabulary that can be used to describe an investigation at different level of detail. His framework is defined to encapsulate all concepts of

the digital forensic field and the relationship between them. Presented model encapsulates the knowledge associated with digital investigation cases. Paper and presented ontology are based on modeling the Windows registry and registry structure and authors limit the scope of this paper to the encoding of forensics knowledge associated with the Windows Registry.

Carver et all. [19] [20] in his early work discussed the need for the application of ontology's to support digital forensics, but no specific ontology was recommended. It is stressed the lack of open ontology's in digital forensics and the needed to create a knowledge base of formal and uniform representation.

Morton Swimmer in his work Towards An Ontology of Malware Classes [21] present a formal ontology of Malware that intends to facilitate precise communication of Malware type. The ontology consist of two parts. The ontology is expressed in OWL and published so that it can be used.

According to the [22] it is not possible to build an ontology that would be sufficiently "large" to include all concepts that occur and which are of interest to people who conduct forensic investigations.

4 Proposed methodology

With conceptualization process we must determine the objects and sets of objects and relations that rule between them.
The main attributes that determine whether the digital evidence to be accepted by the court are:

- Time (time stamp)
- Place (gps location)
- Summary of digital evidence (hash value)
- Biometric characteristics of investigators
- Procedures (rules to be complied with)
- Reason (for digital investigation)

These attributes also represents the hypothetical variables - qualitative and quantitative characteristics of digital evidence. Hypothetical constructs are concluded or suspected concepts.
As a digital evidence to be accepted by the court it must be:

- Acceptability
- Relevancy
- Authenticity
- Integrity (non-repudiation)
- Confidentiality
- Availability
- Accountability [23] [24]

Besides the listed attributes that determine the acceptability of digital evidence essential are also:

- Repeatability
- Reconnaissance
- Availability

There are a lot of question about this! One of the question that arises here is: "What is the appropriate metrics and how to measure these attributes?" In the previous studies and research and in the practice, the integrity of digital evidence was proved with „hash value" and a summary of its calculation and comparison. Confidentiality and Availability of evidence was proved with the chain of evidence (chain of custody) which was usually in paper form or some type of electronic forms. Other attributes are more or less will be given to the judge to accept them or not (Acceptability, Relevancy, etc.)!
The main idea is to develop ontology of digital evidence and an automated system which can decide whether the evidence will be acceptable or not.
The system would be, on the basis of ontology's and set of rules (OCL) automatically decide on the admissibility of digital evidence.

After defining a set of variables and construct the steps that need to be made are:

- Determining and defining the scope and domain ontology's
- Methodological elements analysis
- Enumeration of essential terms, the definition of concepts from the domain of digital forensics
- The semantic description of the identified variables necessary to create ontology's
- Define class, setting a hierarchy among classes, defining classes and properties set limits on these properties
- The definition and development of ontology's in some of the available tools (Protégé)
- The development and concretization of the system based on open-ontology ("To-Be")

Figure 1 present a first version taxonomy diagram of chain of evidence concept.

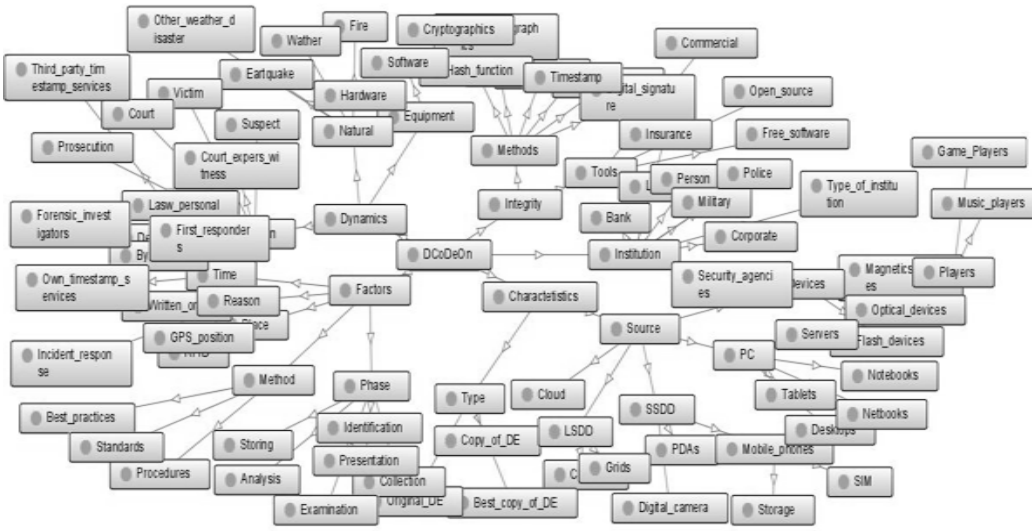

Figure 1 Proposed taxonomy of chain of evidence - early work [11]

Proposed Chain of Digital evidence (CoDe) can be presented like a *function of secure management* that consists of few factors:

- Fingerprint of digital evidence
- Biometrics characteristics
- Time stamp
- GPS locations of person who handles evidence
- Write order or incident response - reason (Why) and
- Standards, set of procedures and best practices.

```
CoDe = f {  fngrprnt _of _file,      //what
            bio_charact,             //who
            time_stamp,              //when
            gps_location,            //where
            reason,                  //why
            set_of_procedures};      //how
```

Those factors are essential for acceptance of digital evidence by the court. Today many of these factors are ignored and the impact they make are unknowns for persons who perform a digital investigation. Therefore, many questions cannot be answered, and digital investigations fall into the water. The cases fall on court or even do not prosecute.
Such a developed system must provide an answer to key question:

- What is the digital evidence
- Where are the digital evidence
- Who manage (make contact) with digital evidence
- Why (reason) to do it
- When digital evidence is handled
- How is handled with digital evidence

5 Conclusion

In this research authors deals with digital evidence and necessity for creating an ontology of digital evidence. It is important because today chain of custody of digital evidence is essential and most vulnerable part of digital investigation process. Proposed methodology is developing a ontology of digital evidence. With this ontology we can share common understanding of the structure of this domain (digital forensic) among forensic investigators and other personal that has to do with digital evidence, among software agents and between forensic investigator and software. It can also enable reuse of knowledge in digital investigation process. This ontology will be a basics for creating a automated open-system for managing with digital evidence.

References

[1] A. Sammes and B. Jankinson, *Forensic Computing: A Practitioner's Guide*

(Practitioner Series). Springer-Verlag, NewYork, 2000.

[2] M. Pollit, "Six blind men from Indostan.Digital forensic research workshop," in *DFRWS-Digital FOrensic Research Workshop*, 2004.

[3] W. J. Chisum, *Crime Reconstruction and Evidence Dynamics*, 2011st ed. SpringerLink.

[4] S. W. G. on D. E. (SWGDE)International O. on D. E. (IOCE), "Digital evidence standard and principles." 2000.

[5] "IOCE Princips and Definition," *IOCE COnference*, 1999. [Online]. Available: http://www.ioce.org/core.php?ID=5.

[6] IOCE, "Guidelines for best practice in the forensic examination of digital technology," 2002.

[7] J. Cosic and M. Bača, "Do We Have Full Control Over Integrity in Digital Evidence Life Cycle?," in *32nd International Conference on Information Technology Interfaces (ITI), 2010*, 2010, pp. 429-434.

[8] R. Yaeger, "Criminal computer forensics management," in *Proceedings of the 3rd annual conference on Information security curriculum development InfoSecCD 06*, 2006, p. 168.

[9] D. WYLD, "The Most Important Chain of Custody: Improving Evidence Tracking with RFID," 2009. [Online]. Available: http://www.rfidglobal.org/News/2009-10/200910271454481135.html. [Accessed: 05-Sep-2011].

[10] J. Cosic and M. Bača, "A Framework to (Im)Prove „Chain of Custody" in Digital Investigation Process," in *Proceedings of the 21st Central European Conference on Information and Intelligent Systems*, 2010, no. Im.

[11] J. Cosic, Z. Cosic, and M. Bača, "An Ontological Approach to Study and Manage Digital Chain of Custody of Digital Evidence," *Journal of Information and Organizational Sciences*, vol. 35, no. 1, pp. 1-13, 2011.

[12] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications," *Knowledge Creation Diffusion Utilization*, vol. 5, pp. 199-220, 1993.

[13] N. F. Noy and D. L. Mcguinness, "Ontology Development 101□: A Guide to Creating Your First Ontology," 2001.

[14] M. Mussen, "Dimensions of knowledge sharing and reuse," *Computers and Biomedical Research*, vol. 25, no. 5, pp. 435–467, 1992.

[15] H. Park, S. Cho, and H.-chul Kwon, "Cyber Forensics Ontology for Cyber Criminal Investigation," in *E-FORENSICS 2009 - 2nd International ICST Conference on Forensic Applications and Techniques in Telecommunications, Information and Multimedia*, 2009, pp. 160 - 165.

[16] D. C. Harrill and R. P. Mislan, "A Small Scale Digital Device Forensics ontology," *SMALL SCALE DIGITAL DEVICE FORENSICS JOURNAL*, vol. 1, no. 1, pp. 1-7, 2007.

[17] A. Brinson, A. Robinson, and M. Rogers, "A cyber forensics ontology: Creating a new approach to studying cyber forensics," *Digital Investigation*, vol. 3, pp. 37-43, Sep. 2006.

[18] D. Kahvedžić and T. Kechadi, "DIALOG: A framework for modeling, analysis and reuse of digital forensic knowledge," *Digital Investigation*, vol. 6, p. S23-S33, Sep. 2009.

[19] L. D. Carver and M. A. Hoss, "Weaving ontologies to support digital forensic analysis," in *ISI'09 Proceedings of the 2009 IEEE international conference on Intelligence and security informatics*, 2009, pp. 203-205.

[20] S. L. Garfinkel, "Digital forensics research: The next 10 years," *Digital Investigation*, vol. 7, p. S64-S73, Aug. 2010.

[21] M. Swimmer, "Towards An Ontology of Malware Classes," pp. 1-16, 2008.

[22] J. Huang, A. Yasinsac, and P. J. Hayes, "Knowledge Sharing and Reuse in Digital Forensics," *Digital Investigation*, pp. 1-6, 2011.

[23] J. Richter, N. Kuntze, and C. Rudolph, "Security Digital Evidence," in *2010 Fifth IEEE International Workshop on Systematic*

Approaches to Digital Forensic Engineering, 2010, pp. 119-130.

[24] E. Casey, *Digital evidence and computer crime-Third edition*. Academic Press, 2011, p. 807.