Baltic J. Modern Computing, Vol. 8 (2020), No. 2, 275-292
https://doi.org/10.22364/bjmc.2020.8.2.05

# Habits Attribution and Digital Evidence Object Models Based Tool for Cybercrime Investigation

Šarūnas GRIGALIŪNAS, Jevgenijus TOLDINAS

Faculty of Informatics, Kaunas University of Technology, Studentų g. 50, Kaunas, Lithuania

{sarunas.grigaliunas, eugenijus.toldinas}@ktu.lt

**Abstract.** The amount of data stored on computers is growing rapidly every year, which makes time-consuming investigation of digital evidence in cybercrime, because of the need to investigate a large amount of data and extract criminal evidence from it. Expert investigation begins with the collection, copying and authentication of each content on the digital medium. The following steps deal with the findings and extract evidence of crime using a variety of methods and tools. Our research deals with the frameworks, methods, models and tools of the search for digital evidence of cybercrime. However, there is as yet no specialized method and tool available to assist an expert in reducing the size of investigated data and to solve the problem of searching for and identifying digital evidence of cybercrime due to the lack of specialized tools and techniques to automate expert investigation. In this paper we propose cybercrime forensic investigation tool based on the digital evidence object model

**Keywords:** cybercrime, forensic investigation methods, models and tools, digital evidence

## 1. Introduction

The Term Bank of the Republic of Lithuania provides an approved definition of the term "Forensic investigation": “In accordance with the procedure established by the laws of the Republic of Lithuania and the Republic of Lithuania An investigation by a forensic expert or professional requiring special knowledge (forensics, object investigation and legal advice)". Forensic Law of the Republic of Lithuania No. IX-1161 determines „expert expertise“ as "the detailed knowledge necessary for the conduct an expertise, acquired in education, special training or professional activity in the field of science, technology, art or any other human activity". Computer crime is carried out using computers, computer networks, and modern information technologies. Search for digital evidence of these crimes requires specific expert knowledge and technical measures, as the computer (data contained therein) becomes the tool of illegal activity or computer technology is used for gathering information, planning and executing criminal activity and illicit data exchange.

Investigating cybercrime poses many challenges for law enforcement and those responsible for ensuring information security. The main ones are: understanding of the specifics of the objects under consideration and the ability to analyse them properly;

knowledge of the laws governing general criminal investigation processes, new legal documents regulating cyber space; ability to assess various risks. The challenges outlined above have influenced the evolution of tools, models and methods for investigating digital evidence in cybercrime, which has led to increasing demands on experts. But criminals have also become more cautious and realize that their actions can be tracked and that abandoned digital footprints may later become evidences in court. Recent trends indicate that criminals are taking steps to complicate the work of experts by using data encryption methods, using automated tools to hide digital evidences, and avoiding using their computers directly to commit crimes. Specialized methods and tools can reduce the amount of data analysed for digital evidences, help an expert to extract digital evidences, and shorten the time needed to perform an expert analysis.

## 2. Related Work

Cybercrime is defined as a crime in which a computer is an object of crime or used as a tool of crime. Cyber criminals can use computer technology to access personal information, business secrets, or use the Internet for malicious purposes. Digital forensics encompasses the recovery and investigation of objects found in digital devices. As was defined at Digital Forensic Research Conference (Palmer, 2001) they need to use scientifically validated methods for the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital data from digital sources, to facilitate or promote events that have been identified as criminals or to assist in unlawful actions which are found to interfere with the planned operations.

Cybercrime is growing in modern society, and the number of computers is growing as well, they change size, shape, speed and function. As computers become smaller, faster and cheaper, they are increasingly embedded in another larger systems and allow creation, storage, and processing of information transmitted in unprecedented ways. Therefore, digital evidence may occur unexpectedly places and forms that requires digital evidence to be aggregated to an objects, which later could be investigated.

The amount of digital information created and replicated in the world grew exponentially and today is calculated in zettabytes. Likewise security threats and different types of attack against communication networks, Internet-of-Things (IoT) infrastructure (Abdul-Ghani and Konstantas, 2019), cyber-physical systems (Sommer et al., 2019), Industry 4.0 (Ervural and Ervural, 2018), Wireless Sensor Networks (WSNs) (Karabiyik and Akkaya, 2019), cloud and fog end devices (Nagar et al., 2017), (Venčkauskas et al., 2018), smartphones (Odusami et al., 2018), social networks (Umair et al., 2017), (Salahdine and Kaabouch, 2019), etc., is growing unhinged, making the communication systems and private data of users vulnerable.

Accordingly grows the volume, variety, velocity, and veracity of digital data available for forensic investigation process that involves collection, preservation, analysis and presentation of evidence of attacks from various heterogeneous digital sources, such as mobile devices, networks, big data in the cloud, etc. (Quick and Choo, 2018). As a result, worldwide spending on Internet of Things (IoT) endpoint security solutions are predicted to reach 631 M$ in 2021 (Gartner, 2018). Unfortunately, there is very little evidence-based research to provide technical solutions and to reduce and analyse the increasing volume of data exists, raising the need for data reduction methods and more efficient data subset collection processes such as the one proposed in

(Quick and Choo, 2016). Another issue faced by modern digital forensics is the need to design effective methodologies and develop efficient tools to detect digital forensic attacks in real-time, which is especially urgent considering the dependability of our society on critical infrastructure such as smart power grids and the threats raised by hybrid warfare (Kurt et al., 2019). Due to the facts above the forensic investigation process is very time consuming, because it requires the examination of all available digital data capacities collected from the digital device used for cybercrime. The forensic investigation process commences with the collection, duplication, and authentication of every piece of digital media prior to examination. Moreover, every action taken has to adhere to the legitimacy rules so that the obtained digital evidence could be presented in the court. The essence of this approach is to prioritize the evidence recovery schedule so that the high probative value, low resource consuming evidential traces are recovered first, while low probative value, high resource intensive evidential traces are deferred until it is clear whether they are actually required for the probable success of the case.

Previous approaches to the modelling of the digital forensics domain included finite state machines (Gladyshev, 2004), theory of information (Cruz et al., 2015) and hypothesis testing (Brian, 2006). The digital forensic evidence model (Cohen, 2010) has defined the process in terms of Laws, Violations, Claims, Events, Traces, Internal Consistency, Demonstration Consistency, Forensic Procedures, Resources, Schedule Sequence using the elements of formal set theory. The digital investigation process model (Carrier and Spafford, 2003) has five categories: Readiness Phases, Deployment Phases, Physical Crime Scene Investigation Phases, Digital Crime Scene Investigation Phases and Presentation Phase. However, these early models of digital forensics did not scale well with the data deluge facing the digital forensics investigators.

Digital forensic investigation encompasses the whole process of collecting, analysing and reporting digital material from a crime scene according to certain standards and methods (Prayud et al., 2015). Cybercrime digital forensics consists of 4 main steps: preparation, collection, analysis and reporting (Geddes and Zadeh, 2016) which is based on Digital Forensics Process Model (Palmer, 2001), (Karabiyik and Akkaya, 2019). Forensic investigation process is raising the need for creating methods and tools based on intelligent technologies, such as artificial intelligence, computational modelling, and/or social network analysis, in order to keep pace with the development of new technologies (Irons and Lallie, 2014), (Jusas et al., 2017). Examples of such valuable contributions are:

- Relevancy-ranking algorithms for digital forensic string search based on 18 features as quantitative indicators of search hit relevancy (Beebe and Liu, 2014);
- The adoption of the Allen algebra, a kind of integral algebra for temporal reasoning, for a semantically rich representation of events related to the cyber incident and advanced digital forensics timeline analysis (Chabot et al., 2014);
- Answer Set Programming, which is a kind of declarative programming to address complex (mostly NP-hard) search problems, formulation of tangible investigative hypotheses and automated reasoning (Costantini et al., 2019);
- AFF4 (the Advanced Forensic Format 4) object model based on the Resource Description Framework (RDF) data model for unique identification and forensic analysis of digital evidence in real time (Cruz et al., 2015);
- Object-Oriented Diplomatics, a conceptual methodology for building digital records capable of supporting their authenticity over time (Jansen, 2015);

- Curated (digital) Forensic object (CuFA), an ontology based (semi-) formal model of digital objects in the cyber forensics domain (Harichandran et al., 2016);
- CybOX, the open-source schema for storing and sharing digital forensic information, associated with Digital Forensic Analysis eXpression (DFAX) ontology for representing common objects and their relationships in digital forensic investigations (Casey et al., 2015);
- Multiple layered orthogonal ontologies for digital forensics that capture relationships from low-level artefact to high-level connections between individuals and allow rule inferring and reasoning using SPARQL query language to automatically derive events from forensic artefacts (Turnbull and Randhawa, 2015).

The role of metadata in digital forensic science defines their importance, which is useful for finding a suspicious system to commit a crime or malicious activity. As the author observes, (Shindel et al, 2015) we can save time and storage in the digital trial by examining metadata. Another advantage of metadata is that it can be explored on any platform. The authors (Pladna, 2008) focused on the development of standard digital evidence by observing the various digital forensics tools, while keeping in mind the legal integrity of the digital evidence elements. In addition, an online questionnaire was used to gain knowledge of experienced stakeholders in digital forensics. Based on the findings, the authors propose a standard for digital evidence that includes case data, evidence source, evidence element, and chain of custody. The results of the study allowed the authors to create a defined XML schema for digital evidence.

The forensic model proposed by the authors (Siahaan and Rahim, 2017) is applied in many areas, and the model includes three components that are assembled, enabled, and managed in such a way that they are the ultimate goal for attaining high quality success. It consists of three parts: Human, Equipment, and Protocol. In other words, when we investigate cybercrime digital evidence, it is important to know: *Where* is the Crime Information; *When* was the cybercrime committed; *Who* did this.

As experts go through digital evidence investigation process, they need time savings abilities. During the case, the process steps can be repeated several times. Every case has to determine when to stop. When digital evidence is sufficiently prosecuted, the value of additional identification and analysis is reduced. Forensic expertise is clearly a very important sector in the world. The advancement of new technology has prompted forensic science to accelerate the growth cycle, since forensic services are now used by different sectors where they are. Unfortunately, there is very little evidence-based research to provide technical solutions and to reduce and analyse the increasing volume of data exists, raising the need for data reduction methods and more efficient data subset collection processes. Another issue faced by modern digital forensics is the need to design effective methodologies and develop efficient tools.

# 3. Modelling and implementing cybercrime forensic investigation method

## 3.1. Modelling digital evidence investigation using habits attribution profiling method

The paper (Grigaliūnas and Toldinas, 2016) presents an original solution of using habits attribution model for the digital evidence investigation. The proposed model focuses on digital evidence investigation that uses attributed habits decreasing number of the artefact search sequences from the set of digital user places. The authors presented a systematic approach to dealing with the problem of attribution, profiling and habits using feature diagram. The habits identification domain (HiD), they mean the profiling technique that is based on the attributed habits. Figure 1 below shows the model of the HiD represented using a feature diagram.

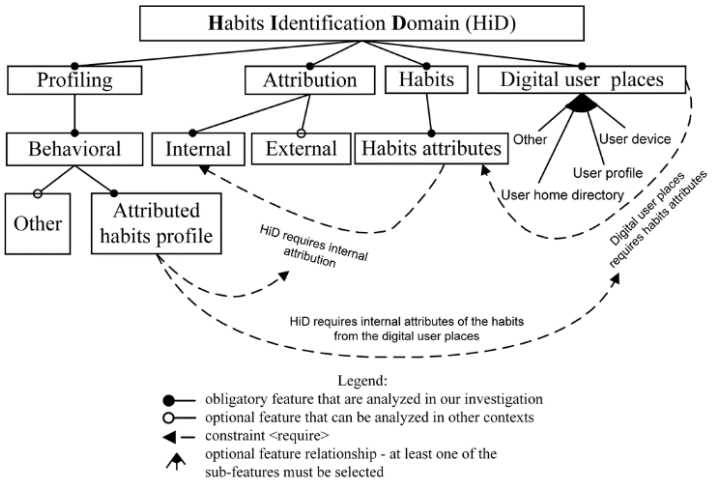


**Figure 1.** Feature diagram of the habits identification domain (HiD)

The profile creation of attributed habits starts from the research and analysis of all information that can be gathered from digital remnants, left on a digital device by its user. The computer user is a human being, tending to customize all the environments with which they interact. Thus, they cannot avoid leaving (even unconsciously) digital evidence artefacts based on detected, recognized and compared habits. The described in this chapter model is suitable to the digital devices, such as: personal computers, tablets, smartphones, etc. Digital evidence artefacts investigation, using proposed model that is based on the habits attribution method, can also be applied to the websites or social networks. The method based on HiD model decreases the number of the evidence investigation search sequences from the set of digital user places. It analyses data and metadata memorized into a digital device by applying specific methods taken from intelligence and traditional profiling in order to obtain information that helps to create digital profile with suspect user habits attributes and then consider it during evidence investigation.

### 3.2. Modelling digital evidence investigation using digital evidence object model method

The paper (Grigaliūnas et al., 2020) presents a new digital evidence object model (DEO) for forensic investigation. The proposed (DEO) model is based on the analysis of information extracted by due forensic process using the elements of category theory with respect to the 5Ws (*Why, When, Where, What, and Who*) (Lopez et al., 2016) while focusing on forensic investigation cases proposed in (Quick and Choo, 2014). For proposed DEO model category theory is used because it is well established in computer science and it has found proponents in several other fields as well (Delvenne, 2019). Specifically, it is well suited to model open, autonomous and networked dynamical systems, therefore they formalism can be applied to describe the digital objects as well.

The goal of the proposed DEO model is to formalize examination phase of the digital forensic investigation process, reduce amount of data from computer system or digital device for examination and accelerate digital evidence acquisition. The model follows the guide of The U.S. Department of Justice (Web, c) in which four main phases of the forensic process were defined: collection, examination, analysis, and reporting. The examination phase is divided in two parts: documentation (document the content and state of the evidence in its totality) and data reduction. Data reduction part of the examination phase is critical due to massive volume of data and information that is stored in computer systems.

A category is a class of objects and arrows linking objects (Delvenne, 2019) and a category consists of objects $X, Y, Z$ , arrows that go between them and given arrows (functions) $f : X \to Y, g : Y \to Z, h : X \to Z$, forming a diagram given in Figure 2.

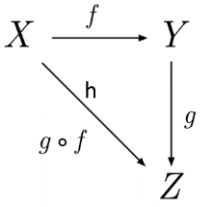


**Figure 2.** Schematic representation of a category with objects X, Y, Z.

Functions $f : X \to Y$ and $g : Y \to Z$ compose to a function $g \circ f : X \to Z$) commutes if and only if $g(f(x)) = h(x)$ for all $x$ in $X$. Arrows (functions) $f$ and $g$ are composable, and the composition of $f$ and $g$ is denoted by $g \circ f = h \; or \; g \circ f : X \to Z$ as shown in Figure 2.

The DEO model is formally defined by a tuple with five variables (1) and is summarized graphically in Figure 3:

$$DEO = (W_{hy}, W_{hen}, W_{here}, W_{hat}, W_{ho}). \quad (1)$$

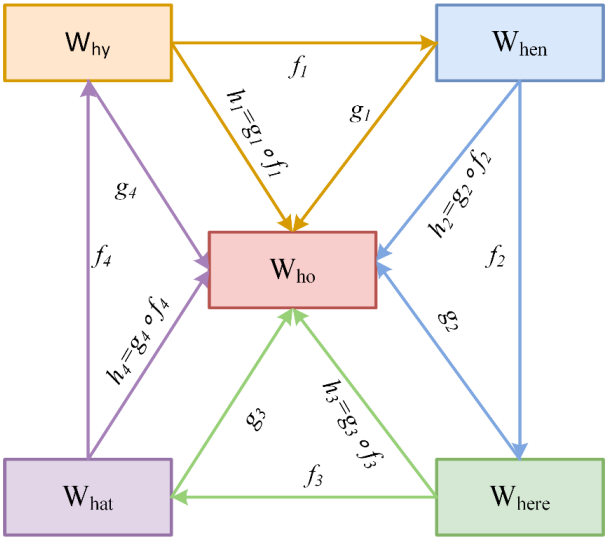


**Figure 3.** Digital Evidence Object (DEO) model

### 3.3. Implementing cybercrime forensic investigation tool

Cybercrime forensic investigation process contains four main phases: acquisition, analysis, presentation and management. According to this process, high-level architecture of the DEIC tool and data-flow diagram depicted in Figure 4.

At the first phase a first responder acquire digital evidence in the form of hard disk drive (HDD) from the suspicious computer and using available disk imaging software creates an exact copy of the suspicious HDD. Next, an examiner using forensic toolkit (FTK), prepare case of all data attributes from suspicious HDD image and exports it as comma separated file (csv). At the preparation phase an expert using DEIC tool imports FTK exported csv file, applies proposed HiD and DEO models and gets DEIC produced reports. If reports does not have appropriate evidence an expert at the management stage can reconfigure models parameters, than repeat presentation and management phases till appropriate evidence will be found. DEIC tool produced reports can be exported to the csv file.

The main purpose of the digital forensic investigation in our implementation is to provide a valid and reliable evaluation of proposed method for digital evidence investigation using presented habits attribution profiling model and digital evidence object (DEO) model. To achieve evaluation result Digital Evidence Investigation of Cybercrime (DEIC) tool (see Figure 5) was developed to perform this task.

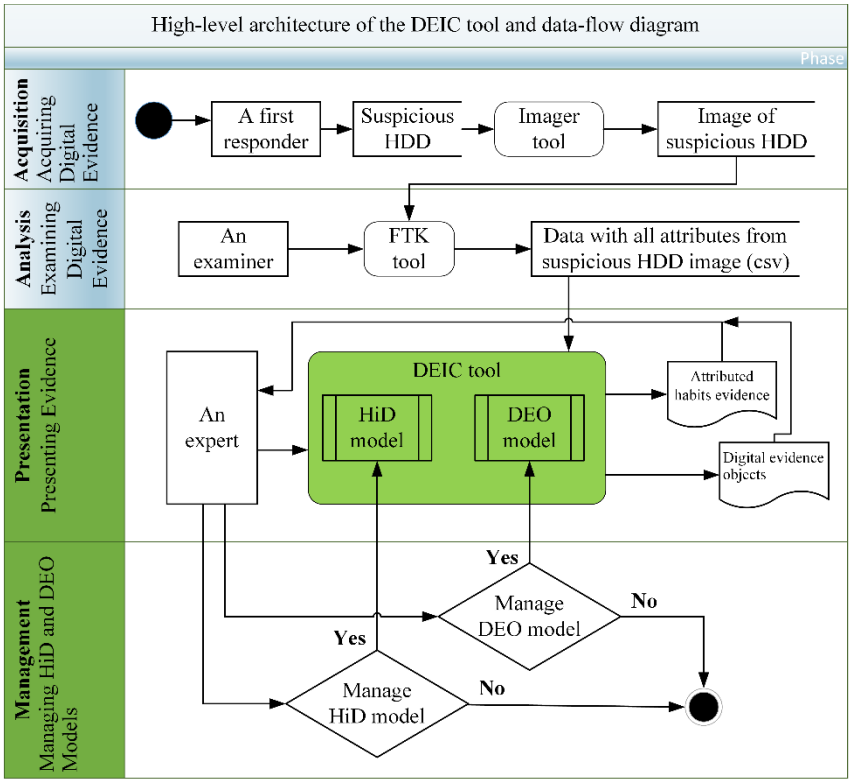


**Figure 4.** High-level architecture of the DEIC tool and data-flow diagram

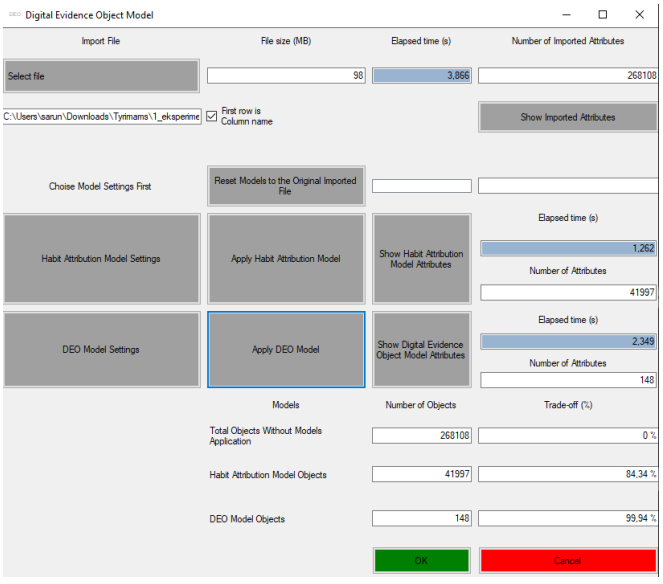


**Figure 5.** Main user interface of implemented cybercrime forensic investigation tool

In Figure 6 are shown settings user interface of proposed method for digital evidence investigation using habits attribution model.

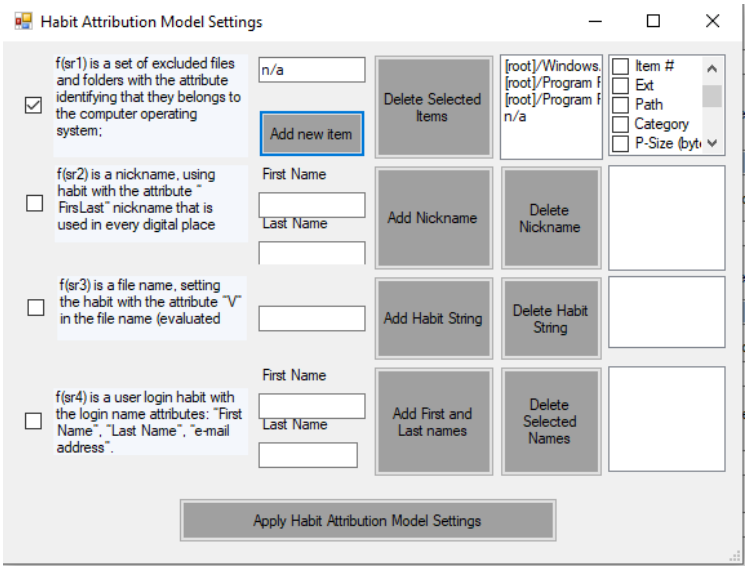


**Figure 6.** Settings user interface of proposed method for digital evidence investigation using habits attribution model

By applying a f(sr1) function which is a set of excluded files and folders with the attribute identified that they belong to the computer operating system, we can suggest that experts do not adjust the attributes of the operating system. In Figure 6 you can see that we suggest using: Path – path to all of evidences, CAM – Create, Access, Modified files, Users – system users, user – specific user and n/a - not applicable, for sample reduction. All parameters are: Name, Label, Item #, Ext – file extension, Path, Category, P-Size (bytes) - physical size, L-Size (bytes) – logical size, MD5, SHA1, SHA256, Created, Accessed, Modified. All files have physical and logical sizes, often the physical size is larger than the logical size, and sometimes it is equal. But the logical size should never be larger than the physical size, otherwise the file system is corrupted or something unusual happens.

By applying a f(sr2) function which is a nickname using habit with the attribute "FirsLast" nickname that is used in every digital evidence line. By applying a f(sr3) function which is a file name setting habit with the attribute "V" in the file names. By applying a f(sr4) function which is a user login habit with the login name attributes – "First Name", "Last Name".

There we can set the *Why* parameter. Expand it by selecting additional parameters. All parameters are: Name, Label, Item #, Ext – file extension, Path, Category, P-Size (bytes) - physical size, L-Size (bytes) – logical size, MD5, SHA1, SHA256, Created, Accessed, Modified. By applying *When* and *What* function which is the time interval indicate investigation and event periods. This was especially helpful in the second part of the experiment - working with a real case. By applying *Where* is a possibility to refine digital evidence search by specific person or process.

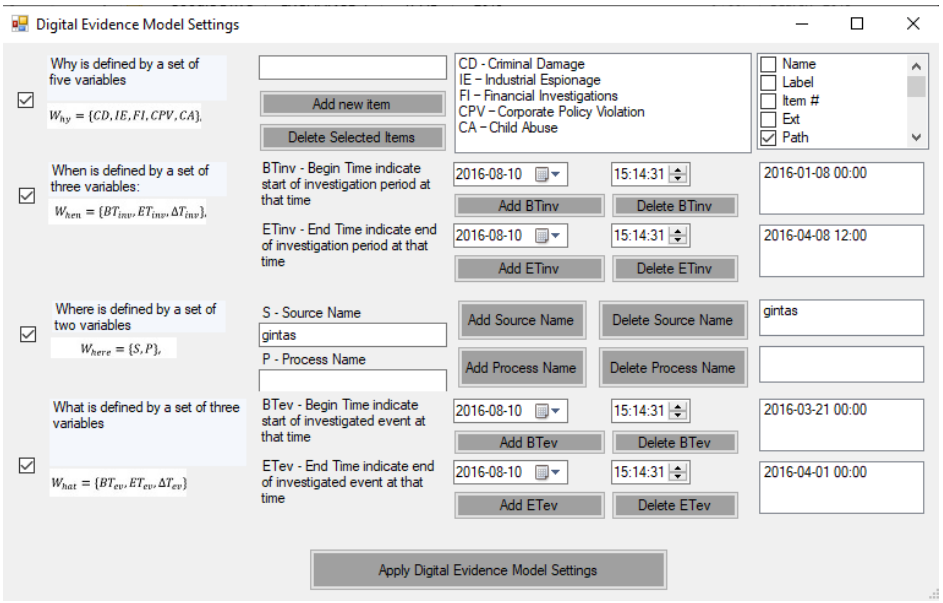


**Figure 7.** Settings user interface of proposed method for digital evidence investigation using DEO model

# 4. Case study

The main purpose of the digital forensic investigation in our case study is to provide a valid and reliable collection of DEOs that can help forensic expert to uncover evidence.

**Context of the case study.** An information system (IS), which controls the power cogeneration plant system, has malfunctioned due to suspected hacking activity. As a result, the power plant caught fire on March 22, 2016, leading to significant material losses to the plant owners. The insuring company of the power plant started investigation to determine the causes of the incident. The logs of the IS were suspected to be modified between March 21, 2016 and April 1, 2016.

**Object of the case study.** Image of the 40 GB Samsung hard disk drive (HDD) that was seized from the suspicious computer. The image mounted in the expert computer and prepared for the examination and analysis.

## 4.1. Evaluation of implemented tool and experimental result

The tools that were used for the experiment:

- Forensic Toolkit 5 (Web, a);
- The proposed DEIC tool.

Order of the digital evidence amount is very important in sample reduction. At the very beginning, we import the attributes (sequentially) obtained from each experiment and then applying proposed method.

Ten different digital cybercrime device images were used. The main purpose is to evaluate the functionality of the proposed method implemented in the DEIC tool using

HiD and DEO models, compare tool results with FTK obtained. The hypothesis raised by the expert - suspicious action was performed maliciously affecting cogenerated energy information system (CEIS) by which journal information was modified and may be operating system traces were modified too. The expert has selected the period for his investigation from 2016-01-08 to 2016-04-08. The possible time of potential suspicious action is defined from 2016-03-21 to 2016-04-01.

The experimental results (see Table 1) demonstrate how different proposed method for cybercrime investigation using habits attribution profiling model functions reduce the number of digital evidences.

**Table 1** Number of attributes using habits attribution profiling model (HiD)

| Cybercrime device image | File size (MB) | DEIC tool elapsed time (s) | Number of attributes | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | FTK | HiD model functions | | | | |
| | | | | f(sr1) Path | f(sr1) Path CAMn/a | f(sr2) Path Users | f(sr3) Path Users | f(sr4) Path user |
| 1 | 98 | 5,164 | 268108 | 41997 | 41997 | 16496 | 16633 | 10181 |
| 2 | 89 | 4,175 | 245581 | 38902 | 3208 | 13779 | 13779 | 13779 |
| 3 | 93 | 8,665 | 257531 | 40633 | 27511 | 12691 | 15693 | 9726 |
| 4 | 196 | 16,066 | 536216 | 83994 | 53696 | 7388 | 33222 | 10164 |
| 5 | 160 | 10,673 | 804324 | 125991 | 85671 | 16209 | 49833 | 30477 |
| 6 | 297 | 14,835 | 804324 | 125991 | 85671 | 16209 | 49833 | 30477 |
| 7 | 93 | 3,427 | 255989 | 39091 | 27511 | 4814 | 15693 | 9726 |
| 8 | 490 | 19,498 | 1340544 | 209989 | 142734 | 27070 | 83165 | 50905 |
| 9 | 98 | 4,25 | 267954 | 41861 | 28422 | 5395 | 16602 | 10157 |
| 10 | 96 | 5,586 | 263783 | 40513 | 27635 | 5322 | 15973 | 15973 |

Analysing results we have got from DEIC tool after proposed model for cybercrime forensic investigation using habits attribution profiling method functions where applied for the first image we draw the following conclusions: the first image after applying FTK tool has 268108 attributes, with f(sr1) function and parameter Path we immediately reduces the amount of attributes (see Table 1); not all cases to reduce attributes by apply CAM - Create, Access, Modified in f(sr1) function, so the results of Path and Path plus

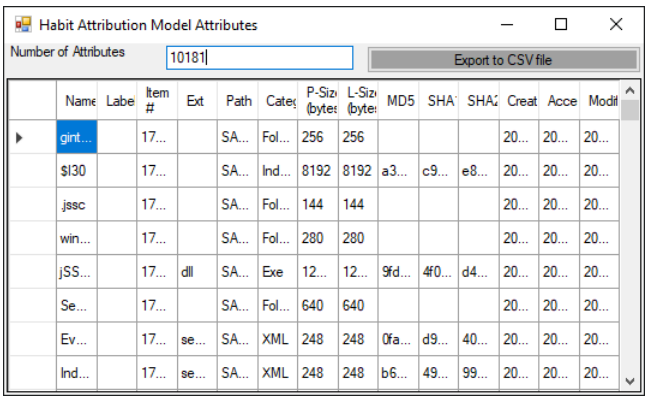


**Figure 8.** DEIC tool results export window

CAM are the same (if you look at the table, this is certainly not the case at all); a f(sr2) function and a f(sr3) function with parameter Users reduce the amount of attributes to 16496 (of course we stay better for the last feature); a f(sr4) function minimizes the maximum number of attributes and reaches the 10181 quantity. The software provides ability for exporting to an expert the summary of founded attributes in csv file (see Figure 8).

By applying proposed $DEO = (W_{hy}, W_{hen}, W_{here}, W_{hat}, W_{ho})$ model which is formally defined by a tuple with five variables obtained the ability to reduce the quantity of attributes even further (Table 2).

**Table 2** Number of attributes using DEO model

| Cybercrime device image | File size (MB) | DEIC tool elapsed time (s) | Number of attributes | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | FTK | Digital evidence object model | | | | |
| | | | | Path (Why) | Path Users (5W) | Path user (5W) | Ext csv (5W) | Path CAM (5W) |
| 1 | 98 | 5,164 | 268108 | 38658 | 13401 | 10182 | 19924 | 153 |
| 2 | 89 | 4,175 | 245581 | 18495 | 10765 | 7749 | 3903 | 166 |
| 3 | 93 | 8,665 | 257531 | 37513 | 12691 | 9727 | 19926 | 127 |
| 4 | 196 | 16,066 | 536216 | 68679 | 18121 | 1522 | 39848 | 332 |
| 5 | 160 | 10,673 | 804324 | 115974 | 40137 | 30477 | 59778 | 39 |
| 6 | 297 | 14,835 | 804324 | 115974 | 40137 | 30477 | 59772 | 498 |
| 7 | 93 | 3,427 | 255989 | 36212 | 12591 | 9726 | 19924 | 90 |
| 8 | 490 | 19,498 | 1340544 | 193294 | 67005 | 50905 | 99630 | 150 |
| 9 | 98 | 4,25 | 267954 | 30649 | 13371 | 10157 | 19924 | 121 |
| 10 | 96 | 5,586 | 263783 | 37199 | 5322 | 4032 | 19163 | 70 |

Using DEIC tool and applying DEO Model for the first image (see Table 2) we draw the following conclusions: using FTK first image has 268108 attributes, with DEO (Why) and parameter Path (Figure 7) we immediately reduces the amount of attributes. Since in the DEO model (see Table 2) all Why, When, Where, What, and Who (5W) are interrelated, so to reduce the amount of attributes in the search to find the Users, user (Who), Ext of the file and CAM. So, at the very end, with DEO Path CAM (5W), we only have 153 attributes that will need to be examined by the case investigator to make the final decision.

For evaluation of results, we use typical metrics used in information retrieval and classification assessment domains (Tharwat, 2018). An error rate in the false detection of reported objects is known as type I error or false positive rate (FPR). An error rate related to objects that falsely not detected is known as type II error or false negative rate (FNR). For proposed DEIC tool evaluation FNR is not highly important, because of the evidence presentation process peculiarity that requires evidence extraction. In such case if evidence objects falsely not detected the evidence presentation process will be cicely repeated while evidence will be presented. At the end of forensic investigation process an expert finally selects finite set of appropriate objects and attributes, from what evidence could be extracted and presented to the court.

FPR is the ratio of irrelevant objects in a set of retrieved objects and is suitable for DEIC tool evaluation because it shows how much objects and attributes using FTK tool where extracted at all and their ratio with HiD and DEO evidence objects and attributes that could be presented to the court. FPR is calculated by equation (2).

$$FPR = \frac{C}{C+D}. \quad (2)$$

where $C$ is the number of irrelevant objects that were retrieved, and $D$ is the number of irrelevant objects that were not retrieved.

**Table 3** Evaluation of results using False Positive Rate

| Cybercrime device image | Numbers of attributes imported from FTK | HiD f(sr4) Path user | False Positive Rate (FPR) | DEO Path CAM (5W) | False Positive Rate (FPR) |
|---|---|---|---|---|---|
| 1 | 268108 | 10181 | 0,03658 | 153 | 0,00057 |
| 2 | 245581 | 13779 | 0,05313 | 166 | 0,00068 |
| 3 | 257531 | 9726 | 0,03639 | 127 | 0,00049 |
| 4 | 536216 | 10164 | 0,01860 | 332 | 0,00062 |
| 5 | 804324 | 30477 | 0,03651 | 39 | 0,00005 |
| 6 | 804324 | 30477 | 0,03651 | 498 | 0,00062 |
| 7 | 255989 | 9726 | 0,03660 | 90 | 0,00035 |
| 8 | 1340544 | 50905 | 0,03658 | 150 | 0,00011 |
| 9 | 267954 | 10157 | 0,03652 | 121 | 0,00045 |
| 10 | 263783 | 15973 | 0,05710 | 70 | 0,00027 |

Software managers in 1980s found they needed a way to estimate the cost of software development in software engineering one of this was open-internal Constructive Cost Model (Boehm et al., 2005). COCOMO besides others metrics allowed software managers to reason about cost, performance, functionality trade-offs.

The COCOMO form is a hypothesis that is tested by the data. The general COCOMO form is:

$$PM = A \times \left(\sum Size\right)^{\sum B} \times \Pi(EM) \quad (3)$$

$\sum$ is the additive, $\sum B$ is the exponential, $(EM)$ is the multiplicative$\left(\mathrm{W_{hy}}, \mathrm{W_{hen}}, \mathrm{W_{here}}, \mathrm{W_{hat}}, \mathrm{W_{ho}}\right)$ where,

PM = person months

A = calibration factor

Size = measure(s) of functional size of software module that has an additive effect on software development effort

B = scale factor(s) that have an exponential or nonlinear effect on software development effort

EM = effort multipliers that influence software development effort.

Currently, COCOMO II is designed to estimate the software effort associated with the analysis of software requirements and the design, implementation, and test of software. Cybercrime forensic expert responsibility is to examine electronics devices that may have been used in cybercrime with main task to find digital evidence of crime activity. Cybercrime forensic expert make a lot of effort on searching and analysing tremendously number of unstructured data from computer hard drives, networks, data storage devices like e-mails, photos, documents and etc. In such a manner forensic investigation process may be evaluated using COCOMO models as cybercrime forensic expert investigation process closely comparable with software engineering process.

It can also be an important component of digital evidence forensics or evidence image acquisition (it is an investigation project based on extracted lines of evidence)

models depend upon the two main equations: Development or in digital forensics analogous part it's analysis effort and time: $E = a * (KLOC)^b$. Which is based on MM - man-month / person month / staff-month is one month of effort by one person. $E$ - treats the number of person-hours per month, PH/PM, as an adjustable factor with a nominal value of 152 hours/PM (it's in COCOMO'81 model), but we will apply this model in Lithuania, where the average monthly management is 160/hours/PM. Embedded Effort is chosen because digital footprint search is a very time consuming job and can take different types of evidences (computer, mobile device, network, cloud, others). $KLOC$ = Kilo (1000) line of code. In our case it will be the result of extracted lines from digital evidence images. The constant, $a$, approximates a productivity constant in PM/KSLOC for the case where $E = 1.0$. The above formula is used for the cost estimation of for the basic COCOMO II model, and also is used in the subsequent models. The constant values a and b for the Basic Model for the different categories of system: $a = 3.6$ and $b = 1.2$ (Web, b). According to the Lithuanian Forensic Science Center (FSCL) typology of presentation of the findings (Table 4):

1. Categorical Positive Conclusion: Formulated when there is an enough set of attributes.
2. Probable: Missing signs formulated for categorical inference.
3. Unable to detect: All or all parts of the test object required for testing are missing, test objects are damaged, inoperative, FSCL does not have technical means.

**Table 4** Ques of Expert Investigation at the Lithuanian police forensic science centre (LPFSC) and the FSCL (source FSCL, 2016)

| Type of expert study | Queue LPFSC (months) | Queue FSCL (months)) |
|---|---|---|
| | | |
| Information technology research | 9 | 12 |

This means that it will take about a year (12 months) for FSCL to come to a conclusion in order to find digital evidences (not the fact that it will be found). In the case of the DEIC tool, the set of digital evidences is reduced by just a few clicks (takes a couple of minutes) and requires less special knowledge.

We also calculated the theoretical time for forensic investigation applying COCOMO II model for our experimental cybercrime digital images. It would take a while to look at all the number of digital evidence attributes imported from FTK tool (see Table 5).

**Table 5** Theoretical time for forensic investigation applying COCOMO II model FTK case

| Cybercrime device image | File size (MB) | Numbers of attributes imported from FTK | PM | Hours |
|---|---|---|---|---|
| 1 | 98 | 268108 | 2953,07 | 42897,28 |
| 2 | 98 | 268108 | 2953,07 | 42897,28 |
| 3 | 93 | 257531 | 2813,83 | 41204,96 |
| 4 | 196 | 536216 | 6784,38 | 85794,56 |
| 5 | 160 | 804324 | 11036,20 | 128691,84 |
| 6 | 297 | 804324 | 11036,20 | 128691,84 |
| 7 | 93 | 255989 | 2793,62 | 40958,24 |
| 8 | 490 | 1340544 | 20372,29 | 214487,04 |
| 9 | 98 | 268080 | 2952,70 | 42892,80 |
| 10 | 97 | 264627 | 2907,12 | 42340,32 |

Using the COCOMO II model, we calculated how would change the time of cybercrime forensic investigation after DEO and HiD models applied (see Table 6).

**Table 6** Theoretical time for forensic investigation applying COCOMO II model DEIC tool case

| DEO Attributes | PM | Hours | HiD Attributes | PM | Hours |
|---|---|---|---|---|---|
| 153 | 0,38 | 60,54 | 10181 | 58,30 | 9327,62 |
| 166 | 0,42 | 66,77 | 13779 | 83,82 | 13411,70 |
| 127 | 0,30 | 48,42 | 9726 | 55,19 | 8829,65 |
| 332 | 0,96 | 153,39 | 10164 | 58,18 | 9308,94 |
| 39 | 0,07 | 11,74 | 30477 | 217,30 | 34768,76 |
| 498 | 1,56 | 249,52 | 30477 | 217,30 | 34768,76 |
| 90 | 0,20 | 32,03 | 9726 | 55,19 | 8829,65 |
| 150 | 0,37 | 59,12 | 50905 | 402,17 | 64347,99 |
| 121 | 0,29 | 45,68 | 10157 | 58,13 | 9301,24 |
| 70 | 0,15 | 23,69 | 15973 | 100,08 | 16013,50 |

If the goal of the Case is to get a profile of a potential offender as quickly as possible, we can achieve an average 27 times reduction of objects, using HiD (see Table 7).

**Table 7** Digital evidence objects set reduction (in times)

| Cybercrime device image | HiD f(sr4) Path user | DEO Path Users (5W) |
|---|---|---|
| 1 | 26 | 20 |
| 2 | 18 | 23 |
| 3 | 26 | 20 |
| 4 | 53 | 30 |
| 5 | 26 | 20 |
| 6 | 26 | 20 |
| 7 | 26 | 20 |
| 8 | 26 | 20 |
| 9 | 26 | 20 |
| 10 | 17 | 50 |
| **Average** | **27** | **24** |

By analogy with the DEO model and knowing nothing about the user, up to 24 times (see Table 7) reduction of objects can be achieved.

## 5. Conclusions

There is a huge number of available computer forensic tools from standalone packages to complex integrated tools, developed for wide range crime investigations. NIST (National Institute of Standards and Technology) registered number 154, open source 140 of tools. A very striking trend when looking at models is the search for digital investigation of cybercrime. Due to this abundance of digital evidence for cybercrime investigation, it became clear why in Lithuania it could take up to a year.

We propose a novel attributed habits profile model based on habits that can be detected, recognized and compared. It is suitable to the digital devices such as personal computers, tablets, smartphones etc. An experiment is conducted to reduce the set of cybercrime digital evidence and the programmed tool was succeeded in demonstrating

the reduction of attributes about 27 times. We propose a novel digital evidence object model that is based on the analysis of information extracted by due forensic process using the elements of category theory with respect to the 5Ws (Why, When, Where, What, and Who) while focusing on forensic investigation cases proposed in. Specifically, it is well suited to model open, autonomous and networked dynamical systems, therefore they formalism can be applied to describe the digital objects as well. An experiment is conducted to reduce the set of cybercrime digital evidence and the programmed tool was succeeded in demonstrating the reduction of attributes about 24 times.

The implemented DEIC tool can help forensic investigator, first, to reduce amount of data for examination, next, to analyse and extract digital evidence from reduced amount of information and smaller data set. By examining the smaller amount up to 99 percent of information and data from a suspected digital image, the digital forensics examiner can increase his/her performance and reduce the error rate.

## References

Abdul-Ghani, H. A., Konstantas, D. (2019). A Comprehensive Study of Security and Privacy Guidelines, Threats, and Countermeasures: An IoT Perspective. Journal of Sensor and Actuator Networks, 8(2), 22.

Beebe, N. L., Liu, L. (2014). Ranking algorithms for digital forensic string search hits. Digital Investigation, 11, S124–S132.

Boehm, B., Valerdi, R., Lane, J. A., Brown, A. W. (2005). COCOMO suite methodology and evolution. CrossTalk, (4), 20–25.

Brian D., C. (2006). A hypothesis-based approach to digital forensic investigations. Dissertation. Spafford, Purdue University. Available at: https://docs.lib.purdue.edu/dissertations/AAI3232156/

Carrier, B., Spafford, E. H. (2003). Getting Physical with the Digital Investigation Process. International Journal of Digital Evidence. Fall 2003, Volume 2, Issue 2. Available at: https://pdfs.semanticscholar.org/915b/524318e2f0689b586ba7ae89ea39e9b22ce3.pdf

Casey, E., Back, G., Barnum, S. (2015). Leveraging CybOXTM to standardize representation and exchange of digital forensic information. Digital Investigation, 12, S102–S110.

Chabot, Y., Bertaux, A., Nicolle, C., Kechadi, M.-T. (2014). A complete formalized knowledge representation model for advanced digital forensics timeline analysis. Digital Investigation, 11, S95–S105.

Cohen, F. (2010). Toward a Science of Digital Forensic Evidence Examination. In K.-P. Chow, S. Shenoi (Eds.), Advances in Digital Forensics VI (Vol. 337, pp. 17–35). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved August 22, 2019.

Costantini, S., De Gasperis, G., Olivieri, R. (2019). Digital forensics and investigations meet artificial intelligence. Annals of Mathematics and Artificial Intelligence, 86(1–3), 193–229.

Cruz, F., Moser, A., Cohen, M. (2015). A scalable file based data store for forensic analysis. Digital Investigation, 12, S90–S101.

Cruz, F., Moser, A., Cohen, M. (2015). A scalable file based data store for forensic analysis. Digital Investigation, 12, S90–S101.

Delvenne, J.-C. (2019). Category Theory for Autonomous and Networked Dynamical Systems. Entropy, 21(3), 302.

Ervural, B. C., Ervural, B. (2018). Overview of Cyber Security in the Industry 4.0 Era. Industry 4.0: Managing The Digital Transformation (pp. 267–284). Cham: Springer International Publishing. Retrieved August 22, 2019.

FSCL. (2016, June 29). Forensic Science Centre of Lithuania (FSCL). Available at: http://www.ltec.lt/index.php?id=883

Gartner. (2018). Gartner Says Worldwide IoT Security Spending Will Reach $1.5 Billion in 2018. Available at: https://www.gartner.com/en/newsroom/press-releases/2018-03-21-gartner-says-worldwide-iot-security-spending-will-reach-1-point-5-billion-in-2018

Geddes, M., Zadeh, P. B. (2016). Forensic analysis of private browsing. 2016 International Conference On Cyber Security And Protection Of Digital Services (Cyber Security) (pp. 1–2). Presented at the 2016 International Conference On Cyber Security And Protection Of Digital Services (Cyber Security), London, United Kingdom: IEEE. Retrieved August 22, 2019.

Gladyshev, P. (2004). Formalizing Event Reconstruction in Digital Investigations. University College Dublin. Available at: http://formalforensics.org/publications/thesis/

Grigaliunas, S., Toldinas J., Venckauskas A., Morkevicius N., Damasevicius R. Digital Evidence Object Model for Situation Awareness and Decision Making in Digital Forensics Investigation. Mar/Apr 2020 - Situation Awareness in Human Computer Interaction. IEEE Intelligent Systems. (in review Round 2).

Grigaliūnas, Š., Toldinas, J. Digital evidence investigation using habits attribution // RCITD - Proceedings in research conference in technical disciplines. Zilina : EDIS - Publishing Institution of the University of Zilina. ISSN 2453-6571. 2016, vol. 4, iss. 1, p. 30-35. DOI: 10.18638/rcitd.2016.4.1.86.

Harichandran, V. S., Walnycky, D., Baggili, I., Breitinger, F. (2016). CuFA: A more formal definition for digital forensic artifacts. Digital Investigation, 18, S125–S137.

Irons, A., Lallie, H. (2014). Digital Forensics to Intelligent Forensics. Future Internet, 6(3), 584–596.

Jansen, A. (2015). Object-oriented diplomatics: Using archival diplomatics in software application development to support authenticity of digital records. (D. Luciana Duranti, Ed.)Records Management Journal, 25(1), 45–55.

Jusas, V., Birvinskas, D., Gahramanov, E. (2017). Methods and Tools of Digital Triage in Forensic Context: Survey and Future Directions. Symmetry, 9(4), 49.

Karabiyik, U., Akkaya, K. (2019). Digital Forensics for IoT and WSNs. In H. M. Ammari (Ed.), Mission-Oriented Sensor Networks and Systems: Art and Science (Vol. 164, pp. 171–207). Cham: Springer International Publishing. Retrieved October 2, 2019.

Karabiyik, U., Akkaya, K. (2019). Digital Forensics for IoT and WSNs. In H. M. Ammari (Ed.), Mission-Oriented Sensor Networks and Systems: Art and Science (Vol. 164, pp. 171–207). Cham: Springer International Publishing. Retrieved October 2, 2019.

Kurt, M. N., Yilmaz, Y., Wang, X. (2019). Real-Time Detection of Hybrid and Stealthy Cyber-Attacks in Smart Grid. IEEE Transactions on Information Forensics and Security, 14(2), 498–513.

Lopez, E. M., Moon S. Y., Park J. H. (2016). Scenario-Based Digital Forensics Challenges in Cloud Computing. Symmetry. 8. 107. 10.3390/sym8100107.

Nagar, U., Nanda, P., He, X., Tan, Z. (2017). A framework for data security in cloud using collaborative intrusion detection scheme. Proceedings of the 10th International Conference on Security of Information and Networks—SIN '17 (pp. 188–193). Presented at the the 10th International Conference, Jaipur, India: ACM Press. Retrieved August 22, 2019.

Odusami, M., Abayomi-Alli, O., Misra, S., Shobayo, O., Damasevicius, R., Maskeliunas, R. (2018). Android Malware Detection: A Survey. In H. Florez, C. Diaz, J. Chavarriaga (Eds.), Applied Informatics (Vol. 942, pp. 255–266). Cham: Springer International Publishing. Retrieved August 22, 2019.

Palmer, G. (2001). "A Road Map for Digital Forensic Research," Technical Report (DTR-T001-01) for Digital Forensic Research Workshop (DFRWS), New York, 2001.

Palmer, G. (2001, November 6). A Road Map for Digital Forensic Research. Report From the First Digital Forensic Research Workshop (DFRWS). Available at: https://www.dfrws.org/sites/default/files/session-files/a_road_map_for_digital_forensic_research.pdf

Pladna, B. (2008). Computer Forensics Procedures, Tools, and Digital Evidence Bags: What They Are and Who Should Use Them. Citeseer

Prayudi, Y., Ashari, A., K Priyambodo, T. (2015). A Proposed Digital Forensics Business Model to Support Cybercrime Investigation in Indonesia. International Journal of Computer Network and Information Security, 7(11), 1–8.

Quick, D., Choo, K.-K. R. (2016). Big forensic data reduction: Digital forensic images and electronic evidence. Cluster Computing, 19(2), 723–740.

Quick, D., Choo, K.-K. R. (2018). Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINT + OSINT): A timely and cohesive mix. Future Generation Computer Systems, 78, 558–567.

Quick, D., Choo, K.-K.R. (2014). Data reduction and data mining framework for digital forensic evidence: Storage, intelligence, review and archive. Trends & Issues in Crime and Criminal Justice. AIC. 1-11.

Salahdine, F., Kaabouch, N. (2019). Social Engineering Attacks: A Survey. Future Internet, 11(4), 89.

Shindel V., Prajapati P., Patel V. Large-Scale Cluster File Systems: Metadata. IJSRD - International Journal for Scientific Research & Development. Vol. 3, Issue 09, 2015. ISSN (online): 2321-0613. Available at: http://www.ijsrd.com/articles/IJSRDV3I90038.pdf

Siahaan, A. P. U., Rahim, R. (2017). Post-Genesis Digital Forensics Investigation (preprint). INA-Rxiv. Retrieved October 2, 2019.

Sommer, F., Darwin, J., Kriesten, R. (2019). Survey and Classification of Automotive Security Attacks. Information, 10(4), 148.

Tharwat, A. Classification assessment methods. Applied Computing and Informatics 2018. doi:10.1016/j.aci.2018.08.003

Turnbull, B., Randhawa, S. (2015). Automated event and social network extraction from digital evidence sources with ontological mapping. Digital Investigation, 13, 94–106.

Umair, A., Nanda, P., He, X. (2017). Online Social Network Information Forensics: A Survey on Use of Various Tools and Determining How Cautious Facebook Users are? 2017 IEEE Trustcom/BigDataSE/ICESS (pp. 1139–1144). Presented at the 2017 IEEE Trustcom/BigDataSE/ICESS, Sydney, Australia: IEEE. Retrieved August 22, 2019.

Venčkauskas, A., Morkevicius, N., Bagdonas, K., Damaševičius, R., Maskeliūnas, R. (2018). A Lightweight Protocol for Secure Video Streaming. Sensors, 18(5), 1554.

Web (a). Forensic Toolkit (FTK). AccessData. Available at: https://accessdata.com/products-services/forensic-toolkit-ftk

Web (b). Geeks for geeks. (2019). COCOMO Model. Available at: https://www.geeksforgeeks.org/software-engineering-cocomo-model/

Web (c). National Institute of Justice. Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition. Available at: https://www.ncjrs.gov/pdffiles1/nij/219941.pdf.

Received April 8, 2020, revised May 3, 2020, accepted May 4, 2020