ŠARŪNAS GRIGALIŪNAS

DEVELOPMENT OF CYBERCRIME FORENSIC INVESTIGATION METHOD

DOCTORAL DISSERTATION

Kaunas
2020

KAUNAS UNIVERSITY OF TECHNOLOGY

ŠARŪNAS GRIGALIŪNAS

DEVELOPMENT OF CYBERCRIME FORENSIC INVESTIGATION METHOD

Doctoral dissertation
Technological sciences, Informatics engineering (T 007)

2020, Kaunas

This doctoral dissertation was prepared at Kaunas University of Technology, Faculty of Informatics, Department of Computer Sciences, during the period of 2015–2019.

Scientific Supervisor:

Prof. dr. Jevgenijus TOLDINAS (Kaunas University of Technology, Informatics Engineering T 007).

Doctoral dissertation has been published in:
http://ktu.edu

Editor: Armandas Rumšas (Publishing Office “Technologija”)

KAUNO TECHNOLOGIJOS UNIVERSITETAS

ŠARŪNAS GRIGALIŪNAS

NUSIKALTIMŲ ELEKTRONINĖJE ERDVĖJE EKSPERTINIO TYRIMO METODAS

Daktaro disertacija
Technologijos mokslai, Informatikos inžinerija (T 007)

2020, Kaunas

Disertacija rengta 2015-2019 metais Kauno technologijos universiteto Informatikos fakultete kompiuterių katedroje.

Mokslinis vadovas:
Prof. Dr. Jevgenijus TOLDINAS (Kauno technologijos universitetas, Technologijos mokslai, Informatikos inžinerija, T 007).

Interneto svetainės, kurioje skelbiama disertacija, adresas:
http://ktu.edu

Redagavo: Armandas Rumšas (Leidykla “Technologija”)

Abstract

Computer crime is carried out with the help of computers, computer networks, and modern information technologies. Searching for digital evidence of these crimes requires specific expert knowledge and technical means, as the tool or computer technology for gathering information, planning and executing criminal activity, and illicit data exchange becomes a tool of illegal activity. The amount of data stored on computers has been growing rapidly every year, which makes investigation of digital evidence in cybercrime both time-consuming and difficult because of the need to investigate a large amount of data and provide well-grounded detailed proof.

This dissertation examines the problem of the emergence of a specialized, appropriate method and tool that helps an expert reduce the sample of the investigated data and deliver digital crime investigation choices. So far, specialized tools and techniques to automate expert research are insufficient.

This dissertation consists of an introduction, five main sections, and general conclusions.

The first section reviews the problem of searching for digital evidence of crime, and looks at the already existing search tools, models, and methods.

The second section proposes a multilayered ontology-based system for selecting the right tool for searching digital evidence. A wireframe, a taxonomy of digital evidence, is proposed to allow the categorization of evidence and the selection of the right tool for digital forensic.

The third section proposes a model to reduce the sampling of the digital evidence found by applying user digital profiling to their digital ‘habits’.

The fourth chapter proposes the Digital Evidence Object model, a method of expert investigation of cybercrime aimed to search for digital evidence of these crimes.

The fifth section describes the proposed DEO model application experiment which compares the results with similar tools available on the market. There was a positive evaluation of the numerical error finding obtained; the developed model was compared to the COCOMO II model, which allows determining the applicability of the model.

The work is summarized by general conclusions confirming the need for the newly proposed object model of digital evidence and its suitability for use in experimental investigation of the digital evidence of cybercrime.

Reziumė

Nusikaltimai elektroninėje erdvėje (Computer crime) atliekami pasitelkus į pagalbą kompiuterius, kompiuterinius tinklus, šiuolaikines informacines technologijas. Šių nusikaltimų skaitmeninių įkalčių (Digital evidence) paieškai reikalingos specifinės eksperto žinios bei techninės priemonės, nes neteisėto veiksmo įrankiu tampa kompiuteris (jame esantys duomenys) arba kompiuterinės technologijos naudojamos informacijos rinkimui, nusikalstamos veiklos planavimui ir vykdymui, bei neteisėtiems duomenų mainams. Kompiuteriuose saugomų duomenų kiekis kasmet sparčiai auga, dėl to nusikaltimų elektroninėje erdvėje skaitmeninių įkalčių ekspertinis tyrimas reikalauja daug laiko, nes reikia ištirti didelį duomenų kiekį ir iš jų išskirti nusikaltimų įkalčius.

Šioje disertacijoje analizuojama problema kylanti specializuoto tinkamo metodo bei įrankio, padedančio ekspertui sumažinti tiriamų duomenų imtį bei spręsti nusikaltimų elektroninėje erdvėje skaitmeninių įkalčių paieškos pasirinkimo. Kol kas nėra pakankamai specializuotų įrankių bei metodų, skirtų automatizuoti ekspertinį tyrimą.

Disertaciją sudaro įvadas, penki pagrindiniai skyriai ir bendrosios išvados.

Pirmajame skyriuje apžvelgiama nusikaltimų skaitmeninių įkalčių paieškos atlikimo problema, apžvelgiami egzistuojantys paieškos įrankiai, modeliai bei metodai.

Antrame skyriuje siūloma daugiasluoksne ontologijomis grįsta sistema skirta skaitmeninių įkalčių paieškai tinkamo įrankio pasirinkimui. Pasiūlytas karkasas, skaitmeninių įkalčių taksonomiją, kuri sudarytų sąlygas įkalčių kategorizavimui ir tinkamo įrankio pasirinkimą skaitmeninių įkalčių surinkimui.

Trečiame skyriuje siūlomas modelis, kaip sumažinti skaitmeninių įkalčių paieškos įrankio surastų įkalčių imtį taikant naudotojo profiliavimą, nustatant jo skaitmeninius „įpročius“.

Ketvirtame skyriuje siūlomas skaitmeninių įkalčių objekto (Digital Evidence Object) modelis, nusikaltimų elektroninėje erdvėje ekspertinio tyrimo metodas, skirtas šių nusikaltimų skaitmeninių įkalčių paieškai.

Penktajame skyriuje aprašomas pasiūlytas DEO modelio taikymo eksperimentas, kurio rezultatai lyginami su analogiškais įrankiais esančiais rinkoje. Atliktas skaitmeninių įkalčių paieškos klaidos teigiamas rodiklio vertinimas, palygintas su COCOMO II modeliu, kurie leidžia nustatyti modelio taikymo tinkamumą.

Darbą apibendrina bendros išvados, patvirtinančios naujai pasiūlyto skaitmeninių įkalčių objekto modelio poreikį ir jo tinkamumą naudoti nusikaltimų skaitmeninių įkalčių eksperimentiniam tyrimui atlikti.

Notations

Symbols

(D) – digital user places – set of the user devices, files (folders), home directory, etc.
(E) – evidence – set of digital evidence
(P) – profile – set of hardware profile attributes
(S) – search – set of search rules
ΔT_{ev} – *T*ime duration between consecutive time values of investigated event period
ΔT_{inv} – *T*ime duration between consecutive time values of investigation period
BT_{ev} – *B*egin *T*ime indicate start of investigated event at that time
BT_{inv} – *B*egin *T*ime indicate start of investigation period at that time
ET_{ev} – *E*nd *T*ime indicate end of investigated event at that time
ET_{inv} – *E*nd *T*ime indicate end of investigation period at that time
$f(sr_1)$ – a set of excluded files and folders with the attribute identified that they belong to the computer operating system
$f(sr_2)$ – a nickname using habit with the attribute 'FirstLast' nickname that is used in every digital place
$f(sr_3)$ – a file name setting habit with the attribute 'V' in the file name (evaluated because the user has a habit of inserting character 'V' in the file name for versioning)
5W – Why, When, Where, What, Who
Cl – Digital Evidence on the Cloud (Cl)
Co – Computer Examination (Co)
D – General Digital Investigation (D)
HiD – The habits identification domain
N – Digital Evidence on the network (N)
O – Others (O) – Internet of Things, Mobile devices, artificial intelligence AI, Social Networks
P – *P*lace for investigation
S – *S*ource for investigation
TM - NIST CFTC and an ontology-based transformation model
E – *Entity* (process, file, directory, registry or system entry, etc.) that takes place in criminal activity
U – *P*erson who takes place in criminal activity
$f(sr_4)$ – user login habit with the login name attributes – 'First Name', 'Last Name', 'e-mail address'
$f(SR)$ – construct search rules-based function

Abbreviations

AutoPSY - digital forensics platform and graphical interface to The Sleuth Kit
CA – Child Abuse
CD – Criminal Damage
CEIS - cogenerated energy information system
CFO - Cyber forensics ontology
CFTC - NIST computer forensic tool catalog
CFTCO - NIST computer forensic tool catalog ontology
COCOMO - The Constructive Cost Model
CPV – Corporate Policy Violation
CSV - comma-separated values
CuFA - Curated (digital) Forensic object
CybOX - Cyber Observable eXpression
DEIC - Digital Evidence Investigation of Cybercrime
DEO - Digital Evidence Object
DEOF - Cybercrime digital evidence object fingerprint
DF - Digital forensic
DFF - Digital Forensics Framework
DFXML - Digital Forensics XML
FI – Financial Investigations
FORZA - FORensics ZAchman framework
FPR - False Positive Rate
FTK - Forensic Toolkit
HDD - Disk drive
IE – Industrial Espionage
IS - Information system
LPKTC - Expert Investigation at the Lithuanian Police Forensic Science Research Center
LTEC - Lithuanian Forensic Science Center
NIST - The National Institute of Standards and Technology
OBTM - An ontology-based transformation model
OBTS - Ontology-Based Transformation System
OPML - Outline Processor Markup Language
OWL - Web Ontology Language
TSK - The Sleuth Kit
UNIX - family of multitasking, multiuser computer operating systems that derive from the original AT&T
UTC - Coordinated Universal Time
XDT - XML document transformations
XML - Extensible Markup Language

Contents

INTRODUCTION

Problem formulation

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, an investigation by a forensic expert or a professional requiring special knowledge (forensics, object investigation and legal advice).” The Forensic Law of the Republic of Lithuania No. IX-1161 determines ‘expert expertise’ as “the detailed knowledge necessary to 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 by 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 (with the 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.

The amount of data stored on computers has been growing rapidly every year, which makes the investigation of digital evidence in cybercrime time-consuming because of the need to investigate a large amount of data and to 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 by using a variety of methods and tools. The research deals with the frameworks, methods and models of the search for digital evidence of cybercrime. However, there is as of yet no specialized method and no tool available to assist an expert in reducing the size of the 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.

Relevance of the thesis

Investigation of 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 analyze the data properly; knowledge of the laws governing general criminal investigation processes, new legal documents regulating the cyber space; ability to assess various risks. The challenges outlined above have influenced the evolution of tools, models and methods for the investigation of digital evidence in cybercrime, which has led to increasing demands imposed on experts. Yet, criminals have also become more cautious and realized that their actions can be tracked and that any digital footprints that they leave may later become evidence 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 evidence, and avoiding the use of their own computers directly to commit crimes.

Specialized methods and tools can reduce the amount of data analyzed as digital evidence, help an expert to extract digital evidence, and shorten the time needed to perform expert analysis

The object of research

The object of this dissertation is the methods of cybercrime investigation designed to improve the process of expert investigation by reducing the searching time for digital evidence.

The aim of the thesis

The main aim of the dissertation is to enrich the area of cybercrime forensic investigation process by proposing an expert investigation method that would help the expert find digital evidence, reduce the sample of data for expert investigation, and shorten the search time for digital evidence.

The objectives of the thesis

In order to achieve the objective of the dissertation and solve the scientific problem, the following tasks have been set:

1. To analyze the already existing methods, models and tools for expert investigation of cybercrime.
2. To develop a new approach which incorporates user-habits profiling and digital evidence object models.
3. To experimentally test and evaluate the tool developed by searching for digital evidence of cybercrime.

Research methodology

The following research methods were used to achieve the objective of the work:

1. Comparative analysis of scientific literature on the already existing techniques, models, and tools for cybercrime.
2. An experimental study of the proposed method and ontology-based transformations in the field of digital crime investigation, user-habits profiling, digital evidence, object models performed by using the developed tool in order to substantiate and evaluate the author's proposals.
3. A method of structured evaluation and generalization of the research and analysis results.

Scientific novelty of the thesis

The scientific novelty has been proven by the following results:

1. A method for profiling the attributes of user-habits for searching of cybercrime digital evidence has been proposed.
2. A two-stage ontology-based transformation model and an ontology-based transformation system have been proposed to enable the selection of an appropriate tool for searching for cybercrime digital evidence.
3. A method based on the digital evidence object model for cybercrime has been proposed.

Practical value of research findings

A tool has been developed and experimentally investigated for the search of cybercrime digital evidence which:

1. Allows reducing the sample size of the expert study data.
2. Enables the expert to more quickly evaluate the results provided by the tool and to extract digital evidence from the data by using patterns of user-habit attribute profiling and digital evidence object models.
3. Can be used when urgent evaluation of cybercrime digital evidence is needed.

Defended statements

1. User-habit attribute profiling allows reducing the amount of data for forensic investigations performed during search for cybercrime digital evidence.
2. The digital evidence object model allows the selection of a reduced representative sample of data focusing on the potential offender and thus helping the expert to make a decision and to reduce the time of expert investigation.
3. The method based on the profiling of habit attributes and digital evidence object models can be applied both for the case study of an expertise, as well as for using the implemented automatization tool.

Approval of research findings

The materials of the dissertation have been approved at three international scientific conferences. Two articles with the citation index in peer-reviewed journals of the *Clarivate Analytics Web of Science* database have been published:

1. Grigaliūnas, Šarūnas; Toldinas, Jevgenijus; Venčkauskas, Algimantas. An ontology-based transformation model for the digital forensics domain // Elektronika ir elektrotechnika. Kaunas: KTU. ISSN 1392-1215. eISSN 2029-5731. 2017, vol. 23, iss. 3, p. 78-82. DOI: 10.5755/j01.eie.23.3.18337.
2. Šarūnas Grigaliūnas, Jevgenijus Toldinas. "Habits Attribution and Digital Evidence Object Models Based Tool for Cybercrime Investigation". Baltic J. Modern Computing, Vol. 8 (2020), No. 2, 275-292. DOI: 10.22364/bjmc.2020.8.2.05.

Dissertation structure

The scientific work consists of an introduction of the dissertation, five chapters, general conclusions, references, a list of the author's publications, and 2 appendices. The total scope of the dissertation is 98 pages without annexes. There are 31 pictures and 23 tables in the text. 208 references were used in the dissertation text.

1 METHODS AND TOOLS FOR CYBERCRIME FORENSIC INVESTIGATION

Digital forensic investigation is a process of collecting, examining and analyzing digital data from various places, such as digital devices, networks, and big data in the cloud. The aim of forensic investigation is to provide situation awareness in terms of the identification and preservation of digital evidence, extraction of information, and analysis of the extracted information in order to facilitate time-critical decision making. To automate the tasks of forensic expert research and development of methods, models and tools must be considered which answer some of the main questions: has a computer crime been committed? Who was (is) the criminal? What digital evidence has been left?

In this chapter, systematic analysis of the main principles of methods, models and tools developed for forensic investigation are presented. The result of the analysis demonstrates the researcher's interests in the development of methods, models and tools for the forensic investigation domain that helps experts in finding digital evidence and making decisions to answer the question about whoever committed the computer crime.

The analysis presented in this Chapter was published in (Venčkauskas, Toldinas, Grigaliūnas, Damaševičius, & Jusas, 2015)[1].

1.1 Overview of cybercrime and digital evidence

A wide range of cybercrimes (including criminal damage, industrial espionage, financial investigations, child abuse, etc.) uses digital devices for storing and transmitting information. In such a way, cybercrime can be defined as criminal activity committed by using a computer, especially to illegally access, transmit, or manipulate data. Digital evidence that is stored in the digital form can play a major role in the forensic investigation process. The disciplines that comprise information assurance of digital forensics best define the legal requirements, and their evolution is informed and guided by case law, regulatory changes (Ryan & Shpantzer, 2002). The ability of cybercrime lawyers and digital forensics examiners to take the solutions of forensic tools and processes to court depends on the regulating acts.

Cybercrime is defined as a crime in which a computer is an object of crime (hacking, cheating, spam) or is used as a tool of crime (child pornography, hate crime). Cyber criminals can use computer technology to access personal information, business secrets, or use the Internet for malicious purposes. Criminals can also use computers to communicate and store documents or data. Criminals engaged in illegal activities are often called hackers. To investigate cybercrime and to collect all the possible and relevant digital evidence for all crimes, law enforcement is merging the collection and analysis of cybercrime digital evidence, also known as computer forensics, into their infrastructure. Even though digital forensic profession has so far become complicated, there are more interesting problems looming at the horizon.

Digital forensics is a process of disclosure and interpretation of electronic data. The purpose of the process is to preserve any original evidence in a structured study by

[1] The references are given in the list of publications by the author on the topic of the dissertation.

collecting, identifying and validating digital information to restore all the history of events. In response to a question what digital forensics is, we also need to identify the digital forensic science which encompasses the recovery and investigation of the material found in digital devices. As it was defined at the 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 criminal or to assist in unlawful actions which are found to interfere with the planned operations.

Cybercrime is growing in the modern society, and the number of computers is growing as well, they are changing in size, shape, speed and function. As computers are becoming smaller, faster and cheaper, they are increasingly embedded in other larger systems and allow creation, storage, and processing of information transmitted in unprecedented ways. Therefore, digital evidence may occur in unexpected places and forms. Space measurement for everyone from environmental monitoring to interactive pulse or high above sea level control means that digital evidence will be even more difficult to collect and analyze; this will make it harder to present in the traditional ways. It is especially important to choose from many clues or get advice on what to look for in the first place, to detect the clues which could help to decide the case by answering the question: who has committed the cybercrime?

The more sophisticated is the computer crime, the more the forensic computer science expands scientific studies from various aspects. It is therefore necessary to break down scientific concentration in the field of computer forensics; it is planned to investigate the crime and even to restore the system after the damage has been done because it was divided into several segments, such as:

1. Solid media (hard disk, usb flash, etc.) forensics
2. Operating systems (files, registry, catalogs, etc.) forensics
3. Data from network (netflows, tcpdump, etc,) forensics
4. Internet (email, browser cash, etc.) forensics

Today, the attention is being drawn to the fact that about ten years ago there was a hypothesis about how cyber digital evidence could be aggregated to objects, which could be investigated later. However, there is still no process of unifying the objects of digital evidence with the reduction of labor cost.

1.2 Cybercrime forensic investigation process

The amount of digital information created and replicated in the world has been growing exponentially, and today it is calculated in zettabytes. Likewise, security threats and various types of attack against communication networks, Internet-of-Things (IoT) infrastructure (Abdul-Ghani & Konstantas, 2019), cyber-physical systems (Sommer, Dürrwang, & Kriesten, 2019), Industry 4.0 (Ervural & Ervural, 2018), Wireless Sensor Networks (WSNs) (Karabiyik & Akkaya, 2019) (Wei & Wo, n.d.), cloud and fog end devices (Nagar, Nanda, He, & Tan, 2017), (Venčkauskas, Morkevicius, Bagdonas, Damaševičius, & Maskeliūnas, 2018), smartphones (Odusami *et al*., 2018), social networks (Umair, Nanda, & He, 2017), (Salahdine & Kaabouch, 2019), etc., has been growing unhinged, making the communication systems and the

private data of users vulnerable. Accordingly, the volume, variety, velocity, and veracity of the digital data available for forensic investigation process has been growing, and 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 & Choo, 2018). As a result, worldwide spending on the Internet of Things (IoT) endpoint security solutions is predicted to reach 631 M$ in 2021 (‘Gartner Says Worldwide IoT Security Spending Will Reach $1.5 Billion in 2018’, 2018).

Unfortunately, there is very little evidence-based research to provide technical solutions and to reduce and analyze the increasing volume of data in terms of raising the need for data reduction methods and more efficient data subset collection processes, such as the one proposed in (Quick & Choo, 2014). 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 the critical infrastructure such as smart power grids and the threats raised by hybrid warfare (Kurt, Yilmaz, & Wang, 2019).

Due to the facts outlined above, the forensic investigation process is very time consuming because it requires examination of all available digital data capacities collected from the digital device(s) 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 taken action 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, and 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.

There are several economics-related metrics that can be employed to prioritize the recovery of the evidential traces, most notably return-on-investment, and cost benefit ratio. Some authors (Overill, Silomon, & Roscoe, 2013) show examples of case-specific devices based on the role of digital evidence in their study and those responsible for these devices at the time of the events. They suggest dealing separately with individual evidence and additional evidence, as their roles are quite different. Alternatively, one can assign costs and weights to each evidential trace, and then schedule them in order of increasing the cost within the decreasing probative value (Overill, Kwan, Chow, Lai, & Law, 2009). The monetized cost of recovering a specific expected digital evidence trace is evaluated as the estimated (typically average) number of expert hours required multiplied by the estimated (typically average) hourly cost (including overheads) of an examiner plus the hourly cost of using any specialist equipment. The digital forensic investigation process is depicted in Figure. 1.

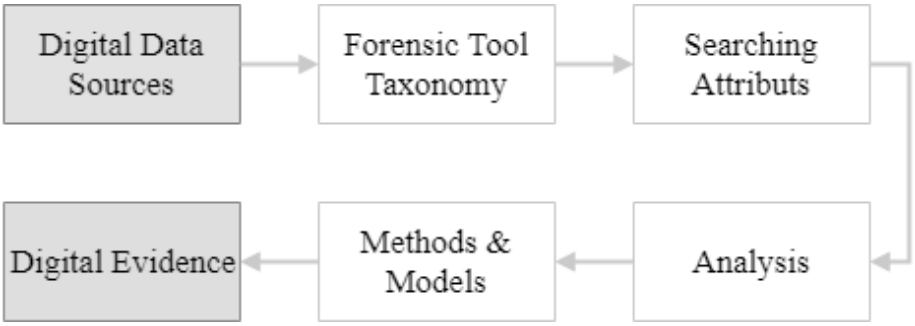

Fig. 1. Digital forensic investigation

The weights or probative values of the expected evidential traces are agreed and assigned by experienced expert examiners and normalized to sum to unity and then use for the recovery of the expected evidential trace. The challenge to the computer forensic science is to develop methods and models which provide valid and reliable results while protecting the real digital evidence from destruction.

The goal of any given cybercrime digital forensic examination is to find facts, and to recreate the truth of an event via these facts (Kävrestad, 2018). The examiner reveals the truth of an event by discovering and exposing the remnants (‘fingerprints’ or evidence) of the event that have been left on the digital device or source (see Fig. 2).

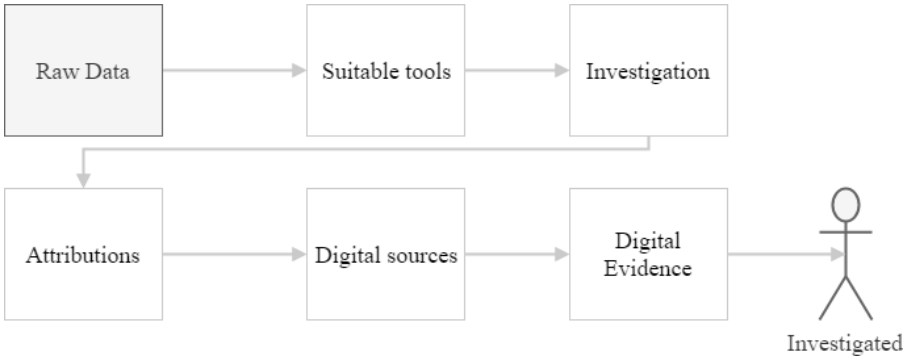

Fig. 2. Process of extracting digital evidence from raw data

The process always starts with gathering any type of information from raw data. That is the data that has not been processed for use and collected from a source. The digital device typically consists of the main board unit, the central processing unit, data storage devices, and peripheral units. Investigators are facing with the problem how to gather any information from a digital device after an incident is becoming an important issue. There are a lot of free or commercial forensic tools on the market, and, before starting an investigation, the examiner must make a decision which tool, where and when should be used. If so, the first question (before starting an investigation) is how to choose the appropriate tool for the investigation that will be the right one in that particular case. Choosing the right tool as early as in this phase can significantly reduce the time of the investigation.

The investigation may take some time because the information will be passed on to the experts and a quick decision by itself will not succeed. Attribution is a suspect action ('event' or 'touch point') that contributes in some way to the desired result and then assigns value to each of these events. It is very important to know when the event happened. The attributions of digital evidence make it possible to understand the combination of events that affect individuals wishing to engage in the desired behavior. The investigator repeats the process of identification for each digital object (Hui, 2012) on the extracted data list with the attribution. An object can also point to a completely new potential source of data. For example, investigators might find a new remote access account which the target was using. After this discovery, police may want to subpoena the contents of the new access. The investigators might also find evidence indicating the stored target, or they may overwrite files on a removable flash drive. For any new data object search leads, the investigators consider going back to the extraction step to process the data. Similarly, for any new source of data that might lead to new evidence, investigators consider going all the way back to the process of obtaining and imaging that new digital forensic data.

The next logical step is to examine the software related to the cybercrime digital evidence. When we talk about preserving digital evidence, digital researchers face many difficulties in trying to source information from a bootable disk by using only software. The software must always communicate with the operating system when the system is running. The investigators have been offered many hardware and software solutions. To be fully effective, they must be installed before the incident occurs.

It is quite clear that there is a need to have the skills to work with software or to give digital evidence to the expert. It takes time to get results. In Lithuania, this process lasts up to 6 months (LTEC, 2016), and, usually, it takes a year until a cybercrime is investigated. Digital forensic research is a process that develops and tests various theories through hypothesis, analyzing digital devices, media that provide relevant evidence in court proceedings by using scientific methods and technologies. The main purpose of the investigation is to establish the truth (*Who*) about the illegal activity (*What*) and the way (*Where*) the crime was committed (Casey, 2010). Digital evidence in this case is a digital object with reliable information that supports or disproves the hypothesis.

1.3 Comparison of tools for cybercrime forensic investigation

The goal of any given forensic examination is to find facts, and to recreate the truth of an event via these facts. The examiner reveals the truth of an event by discovering and exposing the remnants ('fingerprints' or evidence) of the event that have been left on the system (Altheide & Carvey, 2011). Twenty-five functionalities focused on the objects for evidence collection from various sources and a total of 156 tools have been classified (Venčkauskas *et al*., 2015). Open source digital forensic referenced tools are organized into the following categories.

Bootable environments – used to boot a suspect system into a trusted state (3 tools). **Data acquisition** – used to collect data from a dead or live suspect system (26 tools). **Volume system** – used to examine the data structures that organize media, such as partition tables and disk labels (15 tools). **File system** – used to examine a file

system or disk image and show the file content and other metadata (20 tools). **Application** – used to analyze the contents of a file (i.e., at the application layer) (54 tools). **Network** – used to analyze network packets and traffic. This does not include logs from network devices (11 tools). **Memory** – used to analyze memory dumps from computers (3 tools). **Frameworks** – used to build custom tools (8 tools).

Eight functionalities are focused on the objects for evidence collection from various sources, and a total 140 tools are referenced (the same tool may be referenced in more than one category). Table 1 Digital Forensic tools taxonomy illustrates the proposed categories and subcategories of computer forensic tools taxonomy and the number of referenced tools.

Table 1 Digital Forensic tools taxonomy

Taxonomy author	Categories	Subcategories	Number of referenced tools
Francia and Clinton (Francia, G., Clinton, K. 2005)	Imaging	4	n/a
	Analysing	7	
	Visualization	2	
NIST	25	n/a	154
Open Source	8	n/a	140

Internet Evidence Finder (IEF) (Magnet Forensics, 2019) is a digital forensics software solution used to find, analyze and present digital evidence found on computers, smartphones and tablets.

Forensic Toolkit (FTK) (AccessData, 2019) is a court-cited platform of digital investigations built for speed, stability and the ease of use.

Covert Forensic Imaging Device (CFID) (SCG Canada Inc., 2019) was designed for forward deployed military, intelligence, and law enforcement personnel who need a simple, small, portable and inconspicuous solution for imaging, cloning and wiping data from portable media such as USB and SD Cards.

BlackLight (BlackBag Technologies, 2019) is a multi-platform forensic analysis tool that allows examiners to quickly and intuitively analyze digital forensic media.

OSForensics (PassMark™ Software, 2019) allows the examiner to identify suspicious files and activity with hash matching, drive signature comparisons, e-mails, memory and binary data.

Steganography Analysis and Research Center (SARC, 2019) has developed state-of-the-art steganography detection and extraction capabilities that address the needs of the computer forensic examiners and information technology security personnel.

Infotainment and Vehicle System Forensics (iVe) (Berla, 2019) is a vehicle system forensic tool that acquires user data from vehicles and allows forensic examiners and investigators to analyze it.

XRY (MSAB, 2019) is a software application designed to run on the *Windows* operating system which allows the examiner to perform a secure forensic extraction of the data from a wide variety of mobile devices.

Automated Image and Restore (AIR) (Linuxlinks, 2019) is a graphical user interface designed to make the task of creating forensic images of the digital media easier for investigators and incident response personnel.

The ***Sleuth Kit®*** (TSK) (Basis Technology, 2019a) is a library and collection of

command line tools which allow a user to investigate disk images, analyze volume and file system data.

Digital Evidence & Forensic Toolkit (DEFT) (DEFT Association, 2019) is the Linux distribution dedicated to digital forensics and intelligence activities.

Digital Forensics Framework (DFF) (ArxSys, 2019) is a free and Open Source computer forensics software that can be used by both professional and non-expert people in order to quickly and easily collect, preserve and reveal digital evidence without compromising systems and data.

Defined features are used to compare the analyzed tools. For this reason, a decision was made to use four types of electronic devices (Holder, Robinson, & Rose, 2009) and the potential evidence for those types published in the FBI guide for first responders (Table 2).

Table 2 Types of electronic devices and potential evidence

Type of electronic device	Definition of electronic device	Potential evidence
Computer system	Typically consists of the main base unit, sometimes called a central processing unit (CPU), data storage devices, a monitor, a keyboard, and a mouse.	Most commonly found in files that are stored on hard drives and storage devices and media.
Handheld devices (portable, mobile devices)	Is a small device that can include computing, telephone/fax, paging, networking, and other features. A handheld device approaches the full functionality of a desktop computer system.	Includes address book, appointment calendar information, documents, email, handwriting, passwords, phone book, text messages, and voice messages.
Peripheral devices	Modems, routers, printers, scanners, docking stations, etc.	Potential evidence related to the devices themselves. In addition, for routers in the configuration files.
Computer networks	Two or more computer systems linked by data cables or by wireless connections to enable them to share resources and data.	The networked computers and connected devices themselves may be evidence that is useful to an investigation or prosecution. The data they contain may also be valuable evidence and may include software, documents, photos, etc.

In evaluating all the overviewed tools used for cybercrime forensic investigation according to the chosen types of electronic devices (whether it applies (+), does not apply (-)), and the results are presented in Table 3.

Table 3 Summary of cybercrime forensic investigation tools comparison results

Tool Name	Traditional forensics and NIST taxonomy category	Suitability for IoT&S domain				
		Computer system	***Handheld Devices***	***Peripheral devices***	***Computer networks***	***Other IoT&S devices***
Autopsy	Deleted File Recovery	+	-	-	-	-
	Hash Analysis					
	String Search					
Automated Image and Restore	Disk Imaging	+	-	-	-	-
BlackLight	Deleted File Recovery	+	+	-	+	-
	Email Parsing					
	File Carving					
	Forensic Tool Suite (Mac					

Tool Name	Traditional forensics and NIST taxonomy category	Suitability for IoT&S domain				
		Computer system	***Handheld Devices***	***Peripheral devices***	***Computer networks***	***Other IoT&S devices***
	Investigations)					
	Forensic Tool Suite (Windows Investigations)					
	Hash Analysis					
	Image Analysis (Graphics Files)					
	Instant Messenger					
	Mobile Device Acquisition and Analysis					
	Social Media					
	String Search					
	Web Browser Forensics					
	Windows Registry Analysis					
Covert Forensic Imaging Device (CFID)	Disk Imaging	+	-	-	-	-
Digital evidence & forensic toolkit (DEFT)	Deleted File Recovery	+	+	-	+	-
	Email Parsing					
	File Carving					
	Forensic Tool Suite (Windows Investigations)					
	Hash Analysis					
	Memory Capture and Analysis					
	Software Write Block					
	String Search					
	Windows Registry Analysis					
Digital Forensics Framework (DFF)	Deleted File Recovery	+	-	-	+	-
	Email Parsing					
	File Carving					
	Forensic Tool Suite (Windows Investigations)					
	Hash Analysis					
	Memory Capture and Analysis					
	Software Write Block					
	String Search					
	Windows Registry Analysis					
Forensic Toolkit (FTK)	Deleted File Recovery	+	-	-	+	-
	Email Parsing					
	File Carving					
	Forensic Tool Suite (Windows Investigations)					
	Hash Analysis					
	Memory Capture and Analysis					
	String Search					
Internet Evidence Finder (IEF)	Cloud Services	+	+	-	+	-
	Email Parsing					
	Forensic Tool Suite (Windows Investigations)					
	Instant Messenger					
	Memory Capture and Analysis					
	Mobile Device Acquisition and Analysis					
	P2P Analysis					
	Social Media					
	Web Browser Forensics					

Tool Name	Traditional forensics and NIST taxonomy category	Suitability for IoT&S domain				
		Computer system	***Handheld Devices***	***Peripheral devices***	***Computer networks***	***Other IoT&S devices***
Infotainment and Vehicle System Forensics (iVe)	Infotainment & Vehicle Forensics	-	+	-	-	+
OSForensics	Deleted File Recovery	+	-	-	+	-
	Email Parsing					
	File Carving					
	Forensic Tool Suite (Windows Investigations)					
	Hash Analysis					
	Memory Capture and Analysis					
	String Search					
	Web Browser Forensics					
	Windows Registry Analysis					
Steganography Analyzer	Steganalysis Evidence Scanner	+	-	-	+	-
	Steganalysis Field Scanner					
	Steganalysis Real-Time Scanner					
	Steganalysis Signature Scanner					
The Sleuth Kit	Deleted File Recovery	+	-	-	-	-
	Hash Analysis					
XRY Complete	Deleted File Recovery	-	+	+	-	+
	Hash Algorithms					
	Mobile Device Acquisition and Analysis					
	GPS Devices Physical Examinations					
	Memory Card Logical and Physical Examinations					
	File Signature Analysis					
	3G Modems and Portable Music Players Analysis					

There is a huge number of available computer forensic tools from standalone packages to complex integrated tools developed for a wide range of crime investigations. The results showed that most of the analyzed tools are suitable for computer system forensic investigation and are oriented to extract from a computer system a wide range of raw data. This proves the idea that there is lack of tools that would be considered as assistance for experts in the context of the cyber situation awareness and help in the digital evidence investigation process.

1.4 Digital forensic science

In the methodological part, many digital forensic researches focus on the viability of the methodologies in use, rather than on the tools used to conduct the research (Wilsdon & Slay, 2006). This has led to a limited number of cybercrime digital forensics tools (DFT) assessment and validation studies that have limited resources for evaluating their tools. Carrier's Model of Abstraction Layers (Carrier, 2003) was one of the first attempts to establish a DFT assessment methodology, and it developed a model focused on the identification and analysis phases of the study. So, it removed the collection and approval part, but the abstraction layers are already available in the digital survey. For example, when data is acquired, it is usually untreated and very

complex for humans. Thus, tools are designed to render the raw data abstract.

To make a research more effective, it is required to establish communication with the digital forensic science (DFC) and define the investigative process for identifying the essential research by monitoring the steps used in field operations. DFC is a multidisciplinary and interdisciplinary field covering a wide range of disciplines, such as criminology, law, ethics, computer engineering, information and communication technologies (ICT), computer science and forensic science. This is the process of disclosing and interpreting electronic data in order to preserve all the original evidence. Although the area of digital forensics is still young, DFC is further developing with each passing day. Colleges and universities around the world have started providing DFC courses in the information security program.

Digital forensics examines a wide range of electronic devices and information formats in terms of further ownership of various software developers and equipment manufacturers. In fact, creating such a large and diverse stakeholder group is a complex task. It is also difficult as the participants do not want to accept certain standards and rules, which often causes potential conflicts of interest with each other.

The academic community and practitioners have always been complaining about the shortage of standard operating procedures in digital forensics and expressed a strong opinion for systematic and reliable methods of forensic research. However, very little domain is denoted by partially productive standards and procedures. The term ‘forensic science’ should be used carefully, as the digital forensics process does not always comply with the rules of the court procedure. In addition, the legitimacy of the process depends on the jurisdictional system of the particular country.

The evolution of modern digital devices goes beyond the scale and efficiency of digital forensic techniques. Digital triage is one of the ways to solve this problem, as it can quickly extract intelligence and provide valuable information to a forensic expert. This triad is known as a live triad. Similarly, such a methodology can be used in the laboratory to identify digital media analysis and to facilitate the absence of examinations. High level training from field examiners is required. It would be economically advantageous to hire a less technically competent specialist for this job.

One way to do this is to change the software to make it friendlier to less qualified technical examiners, or to create specific tools that can incorporate expert knowledge. The main concept of this approach is that expert systems can find and interpret low-level computational evidence and provide advanced concepts. Large-scale data can be obtained for forensic examination. As each desktop computer has multiple processors, the available resources can be used to speed up the calculations. Parallel processing methods are already in place, but, so far, only in small quantities.

In 2010, an article on the history of digital forensics research was published (Cohen, 2010) where the author discussed the history which may not be chronologically short, but still is very complex. In less than thirty years, digital forensics has been witnessed to grow immensely, and has been expanded by investigators to its current cyber digital evidence searching and analysis methods. The author unambiguously revealed that predicting the future will require digital forensics tools.

To curb the explosive growth of expenses on security, digital forensics should be grounded on a sound methodology with strong scientific roots and address the four

main aspects of observation and measurement, automation, complexity, and artificial intelligence (Olivier, 2016). The latter is defined by Olivier as a "science that may be trusted to produce scientifically-justifiable evidence." Compared to another well-settled research domain, such as Computer Science, or Information Security, digital forensics is still considered as a relatively new research direction that continues to evolve rapidly. Consequently, digital forensics still needs thoroughly defined and validated process models, datasets, procedures, techniques, as well as formal research methodologies (Montasari, Carpenter, & Hill, 2019).

Despite the rapid development of the field, the researchers admit that the gap between the theory and the practice in digital forensics remains (Sremack, 2007), which has led to a gap between digital forensics practitioners and researchers. The latter group ensures that the pursuit of new models for the digital forensic science is still relevant and is continuing. Specifically, there is a need to present the results of digital investigation clearly and understandably in order to give decision-makers increased confidence in the digital forensic investigation results (Casey, 2018).

1.5 Methods and frameworks for cybercrime forensic investigation

The fundamental principles of digital analysis were highlighted by Ricci (Ieong, 2006) who defined the investigation tasks and outlined different roles and their responsibilities in a digital forensic investigation. The authors combined the roles, responsibilities and procedures together, and developed the technology-independent digital forensics investigation framework FORZA (Ieong, 2006) (see Fig**. 3**).

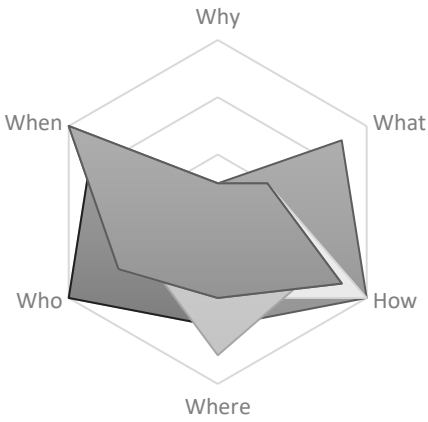

Fig. 3. Layers of FORZA framework

This is very important because they tried to use the 5W (see Fig. 3) methodology for the first time. The FORZA framework was trying to collect resources without focusing on the case of digital evidence and cutting it out. Thus, it can be concluded that, through this FORZA framework, if the motivation is clear, better decisions can be taken at the time of the data loss while assuring that it is done at the right time in the right place while providing information about whoever is suitable for performing the task.

The cybercrime in the Internet network can generally be stored in any type of files that use a specific data format. For that particular case, in 2012, some authors (Riadi, Istiyanto, & Ashari, 2012) presented a clustering technique as one of the methods that can be used to facilitate the identification process. If a disk or a data set is encrypted and readable, as long as the computer is powered on and logged on with the owner's username and password, the right solution for this case is to use the live acquisition cybercrime digital evidence collection method (Akbal, Dogan, & Dogan, 2018).

Live forensics and traditional forensic methods are, essentially, methods of identification, storage, analysis and presentation. However, 'live' forensics was a traditional forensic method that could not get information from the data and information that could be obtained if the system were to work, such as the memory, the network process, the swap file, the system process and system files. The principle is to preserve digital evidence in the form of a process and any computer activity when it is turned on and connected to a computer network because the digital proof on the computer containing the file will be lost when the computer has been turned off. The biggest obstacle to this method is that information can be lost if the procedures are not maintained.

When a cybercrime occurs, the first responders arrive on the scene to identify and protect digital devices for the reliability of the forensic science (while preserving the integrity of evidence). By providing evidence facilities, digital forensic researchers shall collect digital evidence for further investigation and analysis. Throughout the phases of collection, investigation and analysis, researchers shall use digital forensic tools (hardware and software). These measures help the police find and restore digital evidence that can serve as evidence of guilt or become evidence of innocence for the defendants. During the reporting phase, the investigators draw up a report which they include in their testimonies. When the investigator is asked to testify and provide evidence in court, the admissibility of the evidence will be examined on the basis of the investigator's order. The most important factor of admissibility is to check whether the evidence device has not been changed during the investigation. In the case of the IoT environment, this can be quite difficult as there is no universal standard for collecting, preserving and analysing IoT data. Therefore, we need to understand where and what the cloud forensics includes (see Fig. 4).

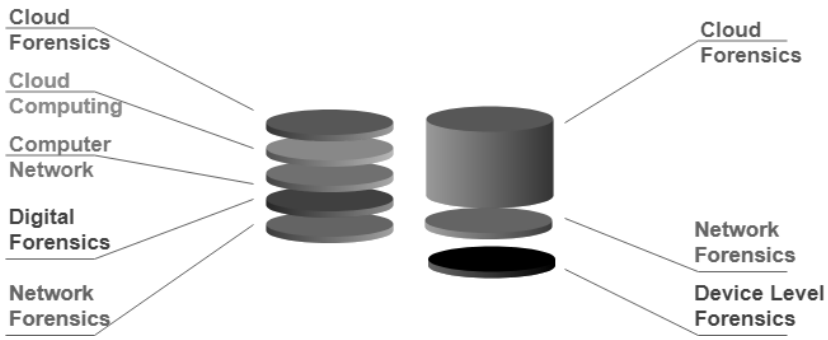

Fig. 4. Part of Cloud and IoT (Zawoad & Hasan, 2015) cybercrime forensics

The main problem is the different interfaces and storage devices. The IoT devices

used in everyday life have different interfaces which allow users to access services or manage devices. Examples of interfaces may include the appropriate software, mobile software, hardware, or embedded software which provides an invisible interface. Because of the different interfaces, digital forensic research is a frustrating process as digital forensic tools do not automatically detect all types of interfaces, file systems, and even data.

As we can see from the NIST definition, cloud computing is an evolving paradigm with complex aspects. Its key features have dramatically reduced IT costs, which are driving the global trend for the strategic adoption of cloud computing. In order to ensure the availability and economy of services, major cloud providers, such as *Amazon*, *Salesforce.com* and *Google*, have data centers around the world in variousa jurisdictions providing cloud services. The data stored in one data center is replicated in several locations to ensure abundance and to reduce the risk of a single failure. For this reason, multiple jurisdiction and multiple leasing have become the default forensic detection, thus creating many more legal challenges. Most cloud forensics files require complex collaboration between the service providers and the client, as well as several tenants who are using the same resources, or cooperation between international law enforcement agencies.

In order to analyze the problem further and emphasize that cloud forensics is a multifaceted problem and not just a technical issue, we would like to broaden the definition of cloud forensics in three key aspects: organizational, legal and technical. The Secure Provisioning Module is a module which stores the chain of evidence stored in the database. This can be done by using PASS as a storage system that performs automated collection, management, storage and retrieval of objects (Muniswamy-Reddy, Holland, Braun, & Seltzer, 2006). The Secure Provenance Module provides records of PASS stored in the original database. Finally, the police can access the evidence and its original records by using the proposed APIs which ensure that the evidence is confidential when using encryption algorithms. To make this possible, researchers need a web server to access the requested data through the API.

Another work (Karabiyik & Akkaya, 2019) proposed the Internet of Things (IoT) Digital Forensics Framework (IDFF). That system provides step-by-step operation in order to show how the operation is performed. According to the authors, the use of automated, intelligent and independent detection and research also identifies security issues for users in smart homes.

IoT and Clouds provide a significant source of potential evidence, but the heterogeneous nature and ways in which data is disseminated, compiled and processed raises the challenges that digital forensic research needs to overcome. At the end, new methods are needed not only to become barriers, but also to influence the architecture and processes in order to gain access to this rich source of potential evidence in IoT and thus in the Cloud environment.

According to some authors (Soni & Bharti, 2015), one of the key aspects of the digital forensic science is the ability to provide evidence of crimes committed by using the digital social media to be brought to court. Network Logs are important evidence. They proposed common framework for forensic analysis of live networks. (Datta & Pan, 2016) suggested an ontology method for collecting digital evidence in a cloud

environment. The cloud belongs to distributed architecture, so the traditional digital forensic approach cannot be directly applied. In cloud computing, most forensic data comes from data logs and their comparison before and after the attack. However, manual comparison of these journals is a very tedious and impossible task, depending on the number of logs one creates. An automated system is therefore proposed.

The rise of cloud computing is driving digital forensics to a new horizon. Many of the current challenges, such as the Jurisdiction and international weaknesses, are worsening, and the new environment also offers unique opportunities for core standards and policies. Cloud computing is a new field of cybercrime as well as a new basis for new research methods.

In this work (Roman, Lopez, & Mambo, 2018), all the research presented by the authors is classified according to the original paradigm for which they were designed. One obvious conclusion is that there are very few studies specifically designed for fog counting and mobile edge computing compared to the quantification of research focused on mobile cloud computing. These paradigms have been recently developed, and their infrastructure is not fully defined, whereas the mobile cloud computing paradigm has been studied for longer. Collecting information from Clouds and IoT equipment is the most complex part as several components are involved: legal regulation, geo location, technologies in use, causes, and the time of the cybercrime. Another aspect is how to identify cyber-attacks. Some authors (Patil, Dharaskar, & Thakare, 2017) revealed that it is still possible to make a connection between cyber-attacks and cloud services. The system proposed by these authors considered three parameters for assessing the violations of cloud service providers, and only two types of attacks are considered. This suggests that more network attacks can be traced and correlated with the information from the IoT device.

(Venčkauskas *et al*., 2018) provided separate, easy-to-read digital evidence of the secure flow protocol for the layer of the fog node end device. This protocol is lightweight, without communication, supports broadcasting and multicast operations, and can provide the authenticity of the data source, data integrity and confidentiality. This area makes one wonder because it will probably take the longest time to find a criminal. Therefore, the most promising strategy is to create a police investigator-friendly method by collecting electronic evidence for future cybercrime investigations.

Computer digital devices store data on internal and external storage devices. The stored data must be collected in certain ways. The Shadow, which only copies the criminal part of the protected data or all of it from the device, is called Image Acquisition (IA) (Carrier, 2005). In order to achieve the standards of a police image, the data must be taken from recorded devices (write block protection), and the image must be encrypted by using such hash methods as MD5, SHA256. When an image is not taken from the original hard disk, the integrity and correctness of the evidence can be destroyed. The expression value calculated at the end of the IA process indicates that the image contains the data from the original device (Akalın & Uluyol, 2016). Some authors (Akbal *et al*., 2018) compare the time taken to capture the device image by tools: *FTK Imager*, *Forensic Imager*, *Encase* and *Xway Imager*. In terms of hardware and software IA over bridges, *FTK Imager* completes the process faster than *Forensic Imager*. It is necessary to mention the fact that all the taken information is not

classified, and the police investigator must decide for him/herself what information s/he requires and how to look for a computer disk image. Related techniques for searching information text extraction in physical devices are sought to obtain information from the sizeable unstructured text data set, such as classifying and dividing text documents (Feldman & Sanger, 2007). Tulowiecki (Tulowiecki, 2018) in his study presents an information retrieval method intended to find meaningful and geographically resolved historical descriptions in large digital historical dossiers. By introducing a biogeography application, this method creates a 'search engine', and uses enhanced boosted regression trees to help find compound descriptions based on the text properties in the history set.

Another method of forensic investigators is called the Case-Based Reasoning Forensic Triager (CBR-FT) (Horsman, Laing, & Vickers, 2014); this is a method to collect and reuse previous digital forensic information in order to highlight the probable area of evidence in the suspected operating system, thereby helping the investigator decide where to look for evidence. The CBR-FT has been shown to be a more effective three-way approach compared to the use of basic commercial applications. The CBR-FT framework provides a mechanism for collecting the results of previous digital court trials and uses this knowledge to predict evidence of a particular type of system.

(Lohiya & Shah, 2015) proposed a new method of segmenting objects when seeking to identify a region of interest that exceeds the provided data set. In addition, the face detection algorithm includes a cascade object detector which allows the algorithm to update the expected position in another frame. This continuity of video frames was not exploited by CAMShift and KLT algorithms. Thus, unlike the CAMShift algorithm and the Kanade-Lucas-Tomasi tracking tool, the proposed face tracking tool retains information about the upcoming positive and provides better results. If we are talking about Apple or Linux, their authors (Quick & Choo, 2016) developed (Quick & Choo, 2014) the idea of data reduction. The authors provide a methodology to reduce the amount of data by using selective imaging. The methodology only offers the choice of basic files and data. The forensic investigator decides to include certain file types. The solution is based on the file data in these file types.

Another option to consider when reducing the amount of data is the thumbnail of video and movie files. (Peersman, Schulze, Rashid, Brennan, & Fischer, 2016) provided a method of learning techniques (assistive vectors) for artificial intelligence and machines to automatically mark a new medium of a child's sexual abuse (CSA). This method uses two steps to mark unknown CSA files. In the first stage, text categorization methods are used to determine whether the file contains CSA content based on its file name. The classification of the text is based on the following characteristics: pre-defined keywords, clear language usage patterns, child-related expressions and family relationships in English, French, German, Italian, Dutch and Japanese. In addition, all two, three, and four consecutive lines are extracted from file names. The second stage receives the files from the first level and examines the visual content of the images and audio files. The second stage is based on the decision on several modal functions.

Other authors (Grier & Richard, 2015) introduced a new approach called

screening collectors intended to depict the regions of selected disk drives. Sifting collectors create a sector by sector, a bit dedicated to a bit region with a forensic value. The forensic image is created by Advanced Forensics Format v3 (Garfinkel, Malan, Dubec, Stevens, & Pham, 2006) and is fully compatible with the already existing forensic tools. These authors use this standard in their experiment because it has become a disk space and squeezes images of the clips. The choice of regions with the forensic value is based on profiles. These authors do not expect examiners to create profiles by themselves, thus the profiles must be created and stored in the library. Sift collectors first collect metadata according to a set profile. Later, the authors (Dalins, Wilson, & Carman, 2015) install a scanning and search method that can be used for digital broadcasting. The proposed method adopts the Monte Carlo tree search strategy used for file system search called the Monte Carlo File System Search (MCFS). Initial random sampling is done by using a non-linear number of points to conduct a guided search. There are three methods for evaluating files, each built on the previous one: similarity-based score, simple scorer, and type of interest scorer. However, the integration of the domain knowledge and the skin tone detection scores showed lower results than expected. An additional study is required to improve these applications. In general, the proposed improved approach is promising because of the complexity of the system due to the complex file system (Roussev, Quates, & Martell, 2013). I conclude that many of the challenges await researchers of digital fingerprints using the triage techniques and tools to maintain the pace with the development of new technologies. This again demonstrates the need for highly skilled professionals to develop new methods.

Some authors (Martinez-Romo & Araujo, 2010) offered various techniques to automatically find candidate web pages to substitute the broken links. Such ways can be used for good and bad purposes in the digital cybercrime. They compared different information extraction methods to construct the queries submitted to the search engine and the ranking of the suspicious web pages that it provides, in order to help the investigators to find changes of the content. The authors found that the most suitable results to expand the query are those with a high probability in the document which is the source of the term and use vector space model intended to identify semantic relationships. This method is automated, but it is noticeable that the results do not make it possible for the decision-maker or cybercrime investigator find who made the content changes.

1.6 Evaluation metrics of the results of forensic investigation methods

It is extremely important to note that open source tools come down to commercial functionality even though they are free. Open source testing for the forensic method does not work without common data sets (Arshad, Jantan, & Abiodun, 2018). However, if automated tools are implemented to support the best practices and latest technologies and techniques, the user community, such as Researchers and Lawyers, will be able to thoroughly check these datasets for possible errors. In addition, the test process would provide an easy-to-access method to test both tools and key methods. Users and scientists will be able to make significant software package comparisons and discover their shortcomings and offer additional or revised features and requirements to support

automated tools.

Because of the diversity and development of the medium, it is difficult to create universal standards of digital forensic science. At the same time, it is also difficult to adopt the case-law on traditional research, such as Testing Standard Data Enclosures. As a result, scientists can contribute to improving the accuracy of the proposed methods and the reliability of the proposed methods so that the courts meet the legal criteria.

In the digital forensics tools and error rates (Lyle, 2010), key explanations indicate that the findings of expert witnesses must be based on scientific methods and defined as scientific knowledge. The error rate reflects the level of uncertainty in the scientific method. There are two types of error. The first one corresponds to the last decision and it is called the false positive; it is also called error Type I. The second option is illustrated by the second solution and is referred to as a false negative, or Type II error (Buchanan, 2011). Lyle (Lyle, 2010) explained that, at first sight, it is possible to determine the level of cybercrime digital evidence tools errors. The simplest solution is:

$$\frac{n}{k}. \tag{1}$$

where n is the total number of the received bits, and k is the number of incorrectly acquired bits.

Typical metrics used in information retrieval and classification assessment domains (Tharwat, 2018) known as the False Positive Rate (FPR) may be used to evaluate results of forensic investigation methods. FPR is the ratio of the irrelevant objects in a set of retrieved objects:

$$FPR = \frac{FP}{FP + TN}. \tag{2}$$

where FP is the number of the irrelevant objects that were retrieved, and TN is the number of irrelevant objects that were not retrieved.

Other authors (Pan & Batten, 2009) carried out additional work which was intended to define the metrics. They proposed a solution to measure the accuracy rate and the precision rate. The first evidence is that it is correctly extracted from the list of evidence, and the second aspect is the number of the extracted files from the file list. The authors also proposed a methodology for evaluating the tool performance. These indicators cover many aspects of cybercrime digital forensic tools, but some of them are not sufficiently detailed. For example, Ayers (Ayers, 2009) does not provide detailed definitions of 'correct results' that can be interpreted by different investigators. In addition, the extraction process may not receive accurate data for the problems that are not due to the tool.

Software managers in the 1980s found that they needed a way to estimate the cost of software development in software engineering. One of the developments was the open-internal Constructive Cost Model (Boehm, Valerdi, Lane, & Brown, 2005). COCOMO, besides other metrics, allowed software managers to reason about the cost, performance, and 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 and (EM) is the multiplicative, 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 the software development effort.

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

1.7 Models for cybercrime forensic investigation

Previous approaches to the modelling of the domain of digital forensics included finite state machines (Gladyshev, 2004), the theory of information (Cruz, Moser, & Cohen, 2015) and hypothesis testing (Brian, 2006). The digital forensic evidence model (Cohen, 2010) defined the process in terms of Laws, Violations, Claims, Events, Traces, Internal Consistency, Demonstration Consistency, Forensic Procedures, Resources, Schedule Sequence by using the elements of the formal set theory. The digital investigation process model (Carrier & Spafford, 2003) involves 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 investigators of digital forensics.

There have been several schemas proposed in the past for representing digital forensic information, but these have not been widely adopted (Turner, 2005). One schema that is in use is *Digital Forensics XML* (Garfinkel, 2012). This schema was primarily developed to represent the output from tools used to analyze storage media, including file system parsers, file carvers, and hash set generators.

Using XML is the best choice for storing digital information in a structured format. User-defined tags may be XML to represent digital information. Metadata provides information about the properties of files and directories that are useful for finding a suspicious system. It also adds value to behaviour and attributes. A cybercrime expert searches the data needed to investigate digital crime. Another way to deal with digital crime is to get digital forensic Metadata instead of data. Thus, digital

forensic experts only need digital forensic information that can prove a criminal activity. There are several tags created by using XML that can store digital information and prove to be a crime or a suspicious computer.

DFXML also provides the ability to store metadata in an XML structure format (Garfinkel, 2012). DFXML allows sharing structured information between independent tools and organizations. Thanks to these transformations, the development of the intelligent ontologies transformation system was provided (Grigaliunas, Toldinas, & Venckauskas, 2017).

The role of metadata in the digital forensic science defines its importance, which is useful for finding a suspicious system likely to commit a crime or to get involved in malicious activity. As (Patel & Sharma, 2015) observed, 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. (Brett Pladna, 2008) focused on the development of standard digital evidence by observing 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 proposed a standard for digital evidence which includes case data, evidence source, evidence element, and the chain of custody. The results of the study allowed the authors to create a defined XML schema for digital evidence.

A common ontological definition regarding a particular domain in the field of science has been created. According to it, common information structure and reusable knowledge can be formed, moreover, assumptions in a domain can be created, and the most important aspect is that in each section at a specific stage, the selected items can be analyzed (Luthfi, 2014). In the field of computer forensics, the concept of Ontology is used to describe and classify specific stages in the process of investigation. While a number of ontologies related to security and intrusion detection were identified (Carrier, 2003), (Razzaq, Anwar, Ahmad, Latif, & Munir, 2014), in developing a framework, the authors for the sake of simplicity and tractability developed their own ontology. Currently, the authors are limiting the use of the features of OWL to the class taxonomy definition, property definition and specification of the event instance declaration. These authors do not use any constraints over the classes or object properties, or property hierarchies.

(Turnbull & Randhawa, 2015) described an ontology approach to examinations. The purpose of this method is to enable a less technically specific user to run the triad tool. This is done by collecting low-level evidence and conclusions from the collected facts. This method focuses on the automatic extraction of events from the basis of forensic evidence. The resource description system is used as the basis for ontology. A representative feature of this method is that different layered ontologies are created through the same data set. The description of the ontologies in use is unclear.

There are many triage definitions that slightly differ depending on the attributes (Roussev *et al*., 2013), (Montasari, 2016), (Cantrell & Dampier, 2012), (Koopmans & James, 2013). The variety of triad definitions reflects the diversity of attitudes and indicates the immaturity of the field. As Cantrell states: “Digital triage is not a judicial process by definition” (Cantrell & Dampier, 2012). However, there are other

definitions, and this statement is not suitable for all cases (Roussev *et al*., 2013), (Montasari, 2016), (Koopmans & James, 2013), (Hong, Yu, Lee, & Lee, 2013), especially when it comes to cybercrime digital evidence. Koopmans (Koopmans & James, 2013) and (Roussev *et al*., 2013) use the term 'digital forensic science'. If the digital triage is not a forensic process, the term 'forensic' cannot be used in conjunction with the term 'digital triage' because it is misleading.

Hong (Hong *et al*., 2013) introduced the triage model, and adapted it to the legal requirements of the Korean system. Thus, the proposed triage model complies with the rules of court proceedings. In addition, Hong (Hong *et al*., 2013) proposed that the triage model of the legal system of a particular country should be identified separately. This article discusses many results. Following the results of the questionnaire, a new triad model was proposed. The Triage process is divided into four phases: planning, execution, categorization and solution.

(Marturana & Tacconi, 2013) summarized the research presented at conferences (Marturana, Me, Berte, & Tacconi, 2011), (Marturana, Tacconi, Berte, & Me, 2012) and presented a model for the living and post mortem triage using machine learning methods. The presented model consists of the following four stages: forensic examination, acquisition and normalization of properties, definition of context and priorities, and classification of data.

A critical review of the triage in live forensics is concluded in (Bashir & Khan, 2013). Considerable amount of work is required to develop pertinent triage for live digital forensic analysis. Currently, the use of triage in the digital forensic investigation is hindered by many obstacles, such as the forensic investigation process and its rules as forensics does not always comply with the rules, thus raising the need for creating methods and tools of digital triage 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 & Lallie, 2014), (Jusas, Birvinskas, & Gahramanov, 2017). Examples of such valuable contributions are:

1. Relevancy-ranking algorithms for digital forensic string search based on 18 features as quantitative indicators of search hit relevancy (Beebe & Liu, 2014);
2. 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, Bertaux, Nicolle, & Kechadi, 2014);
3. 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, De Gasperis, & Olivieri, 2019);
4. 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);
5. Object-Oriented Diplomatics, a conceptual methodology for building digital records capable of supporting their authenticity over time (Jansen, 2015);
6. Curated (digital) Forensic object (CuFA), an ontology based (semi-) formal model of digital objects in the cyber forensics domain (Harichandran,

Walnycky, Baggili, & Breitinger, 2016);

7. 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, Back, & Barnum, 2015);
8. 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 & Randhawa, 2015).

These are the most commonly used models (Brain Adams, 2012):

1. An extended model of cybercrime investigation;
2. Enhanced digital investigation process model;
3. Computer forensic field triage process model;
4. Proactive and reactive digital forensics investigation models;
5. Model for critical phases in network forensics;
6. Proactive network forensic evidence analysis model.

The forensic model proposed by authors (Siahaan & 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* the cybercrime was committed; *Who* did this.

Digital forensic investigation encompasses the whole process of collecting, analyzing and reporting digital material from the crime scene according to certain standards and methods (Prayudi, Ashari, Priyambodo, & Priyambodo, 2015). Cybercrime digital forensics consist of 4 main steps: preparation, collection, analysis, and reporting (Geddes & Zadeh, 2016), which is based on the Digital Forensics Process Model (Palmer, 2001), (Karabiyik & Akkaya, 2019).

The Secure Evidence Retention Module (Zawoad & Hasan, 2015) ensures continuous tracking of the registered devices for forensic evidence by using log files or sensor-collected data. If the evidence is recognizable, it is stored in the evidence store. The data warehouse database is designed on the hadoop distributed file system in order to modify and reliably process large data.

Since existing methods cannot unanimously express, exchange, and reuse digital evidence verification information, a solution is provided for the model of ontology-based models of digital evidence review elements (Wang, 2017). Firstly, in combination with the knowledge, classification and characterization of multimedia analysis of digital evidence is performed. Secondly, based on the principles of constructing ontologies, a model of the knowledge base of elements of a review of digital evidence is developed covering subject ontology, applied ontology and atomic ontology. Finally, the model can effectively gain knowledge about the digital data survey by analyzing the review scenario.

The network has the opportunity to utilize the tools of machine training. With their help, authors (Trifonov, Manolov, Yoshinov, Tsochev, & Pavlova, 2017)

embraced security models based on Cyber threat mobility/dynamics. This should also include the use of Cyber Threat Intelligence in order to assess the effectiveness and efficiency of the implemented security controls.

1.8 Research on methods for computer crime forensic investigation in Lithuania

Various approaches have been developed over the past decade, but most of them have not been able to process large amounts of data, study evidence and competently improve the comprehensibility of the charts to help the investigator. Opportunities are being sought in Lithuania to develop trends of intelligent information systems and technologies as well as future possibilities for their implementation in the area of legal application will be examined. (Dzemydiene, 2010) poses a question of how different data can be combined into a single structure for intelligence purposes as it has become an important issue. The increased mobility and new communication channels allow cyber criminals to better plan and organize their activities over large geographical areas. Therefore, the patterns that reflect fraudulent activities are becoming increasingly difficult to identify due to the large amount of data that is distributed across files in independent police and legal structures. (Dzemydiene, 2010) made a recommendation regarding information systems and methods for their implementation.

(Jusas *et al*., 2017) suggested digital sorting as the first phase of a forensic investigation. They proposed two types of digital sorting: live and posthumous. The main goal of live sorting is to quickly extract information from potential sources. Live sorting raises legitimate concerns. Posthumous sorting is carried out in the laboratory. The main purpose is to assess the seized devices for the availability of relevant evidence. Digital sorting allows items that may contain evidence to be quickly identified. Therefore, this is a solution to the open case problem. The currently existing digital sorting methods and tools have limitations, especially in the forensic context. However, they do not offer a better solution yet, and the authors concluded that the developers are facing many tasks in creating methods and tools for digital sorting to keep pace with the development of new technologies.

The P2P (torrent) is a popular utility for sharing large files over the Internet. Sometimes, this powerful utility is used to commit cybercrime, such as the exchange of illegal material or the illegal exchange of legal material. To help forensic investigators manage this cybercrime, (Venčkauskas, Jusas, Paulikas, & Toldinas, 2016) examined the artefacts left behind by a BitTorrent client and proposed a method for detecting the artefacts which indicate the completed activity of the BitTorrent client.

The problems with recognizing security information in legal and other administrative data are being discussed in Lithuanian articles. Authors analyze the prospects of using data mining when seeking to solve two main problems: the frequency and indirectness of this data. Security research uses two types of data. The first is scientific data that has been specially developed and collected to test certain security theories. This data is criminological, sociological, psychological surveys, experimental data, etc. The second type of data is the data that is not intended for research in the field of security. Most data is by-products of the national legal system, particularly its criminal justice system, national and local institutions responsible for maintaining the public order, and other state institutions. All the data contains a wealth

of information for monitoring and managing the activities of these institutions. There are two types of problems when using this information in a safety study. First, this information should be 'decrypted' from the data describing the activities of the respective institutions. Second, security data is lost in large amounts of other publicly available data. Therefore, it has to be 'dismantled'. (Justickis, 2010) describes the perspectives of the modern method of information processing – data mining in solving both problems, and proposes a general algorithm for the extraction of such data.

In the field of digital forensics, it is not an easy task to get a clear idea of the events and artefacts which occur over time. (Bhandari & Jusas, 2020) focused on recovering the timeline of events and artefacts, which would allow digital investigators to understand the timeline of a digital crime and interpret the findings as digital evidence. One of the most important and complex tasks in the field of digital forensics is the analysis of huge events due to the explosive growth of the Internet, interconnected devices, and innovative technologies today. They are introducing a methodology based on the concept of abstraction and forensic tools that can help the investigators recover, understand the schedule of events and artefacts, and interpret the evidence by tracking user actions on a typical computer system. A single structure is defined for all the sources at a certain level of abstraction over a number of fields that should be considered as an abstract time axis. The main goal of this approach is to optimally support chart analysis. The experiment of these authors shows that the proposed approach improves the analysis of the time axis.

In Lithuania, a methodological book was also published. The aim of this textbook is to acquaint the students with the main types of cybercrime, the legal aspects of their definition and investigation, basic research processes and tools, to provide practical knowledge in case studies, and to outline possibilities. (Goranin & Mažeika, 2011) hope that this textbook will also be useful to students of other similar or general information security courses, lawyers dealing with cybercrime, law enforcement officers, and anyone interested in information security. This textbook is one of the first attempts to describe the field of this issue in Lithuanian; therefore, the authors apologize for possible mistakes and inaccuracies and will appreciate constructive criticism and suggestions.

When summarizing, it can be started that, among Lithuanian researchers, there is significant interest being paid to the Computer crime forensic investigation topic, and a few new approaches or solutions for forensics can be highlighted.

1.9 Comparison of methods, models and frameworks for cybercrime forensic investigation

As shown in the previous sections, there exist a large number of methods, models and frameworks for cybercrime forensic investigation proposed by contemporary scientists. All of these methods, models and frameworks concentrate on specific evidence or propose their own evaluation points of view. Therefore, in this section, we shall provide an overview of some relevant methods, models and frameworks for cybercrime forensic investigation in order to form a general view on the currently existing solutions.

Beebe and Liu (Beebe *et al*., 2014) proposed an initial set of relevancy ranking

features and obtained very promising empirical results thus delivering a valuable analytical technique of text string searching allowing to reduce the analytical burden, which is important for digital forensic practitioners. As a result, they proposed eighteen features as quantitative indicators of search hit relevancy.

Chabot, Bertaux, Nicolle and M-Tachar Kechadi (Chabot *et al*., 2014) represented the SADFC (Semantic Analysis of Digital Forensic Cases) approach as a synergy of three elements: (i) knowledge model for advanced digital forensics timeline analysis, (ii) investigation process model, (iii) ontology-centered architecture. It allows reducing the tedious character of the analysis for the investigators, and to focus on the tasks for which their expertise and experience are most needed, such as the interpretation of results, validation of hypotheses, etc.

Costantini, Gasperis and Olivieri (Costantini *et al*., 2019) presented the formalization of realistic investigative cases via simple answer set programming programs and showed how such a methodology can lead to the formulation of tangible investigative hypotheses. Moreover, a design for a feasible decision support system especially meant for investigators and based on artificial intelligence tools was proposed. The evidence analysis phase was addressed, and a first step was made which aimed to create an infrastructure for the application of artificial intelligence and automated reasoning in the field of digital forensics.

Datta and Pan (Datta & Pan, 2016) sought to mitigate the problem of investigating cloud crime incident manually rather than focusing on automation; proposed a two-component framework: the ontology-enabled forensic blackboard, and the ontology-enabled forensic controller and processor. According to the authors, the use of the concept of ontology makes this automation faster and more accurate. Another point of the authors' view is the fact that the existence of knowledge bases owing to both of the model components makes the investigation an easier one. The knowledge base of the proposed ontology-enabled forensic controller and processor helps the investigator to propagate the correct request for evidence to the correct malicious actor identifier which allows ultimately collecting the relevant evidence and information at the reported crime scene by the investigator. Datta and Pan declare that populating this knowledge base each and every time makes these knowledge bases more and more enriched.

Hikmatyar, Prayudi and Riadi (Hikmatyar *et al*., 2017) presented a network forensics framework development using the interactive planning approach. The authors concluded that Integrated Digital Forensics Investigation Framework version 3 is more comprehensive and useful for network investigation than the other existing models and that it had the strategy approach phase for adapting case handling.

Horsman, Laing and Vickers (Horsman *et al*., 2013) presented the Case-Based Reasoning Forensic Triager (CBR-FT) method that collects and reuses past digital forensic investigation information and then highlights the likely evidential areas on a suspect operating system, thereby helping the investigator decide where to search for evidence. The authors introduced evidence relevance rating (ERR) that is given by the investigating practitioner and is the evidence related to the investigation. Next, CBR-FT uses the ERRs as a prior probability distribution in a Bayesian model to determine the priority of particular locations for searching during the device triage. In the paper, it

was shown that, in 17 cases, *CBR-FT* recovered more evidence and less non-relevant data than *EnCase*.

Ieong (Ieong, 2006) defined the sets of six key questions among all the investigation processes that each practitioner would always ask: (1) What (the data attributes), (2) Why (the motivation), (3) How (the procedures), (4) Who (the people), (5) Where (the location), and (6) When (the time) questions. The author linked together eight roles and their responsibilities and introduced the technology-independent digital forensics investigation framework through the Zachman framework derivatives – FORensics ZAchman framework (FORZA). By using this framework, questions and answers in a digital forensics' investigation could be systematically thought through.

Luthfi (Luthfi, 2014) discussed the Ontology Framework approach. The proposed model uses a structured hierarchy of layers that create connectivity between the variant and the searching investigation of an activity so that computer forensic analysis activity can be carried out automatically. The simplicity of the mechanism and the rules that are used becomes an important factor for the development of automated systems while aiming to render this framework into a system that can automatically integrate the phases of digital forensic investigations while using the Ontology framework.

Marturana and Tacconi (Marturana and Tacconi, 2013) performed a benchmark study by using Bayes Networks, Decision Trees, Locally Weighted Learning and Support Vector Machines. They presented the triage model that consisted of four steps: (i) forensic acquisition, (i) feature extraction and normalization, (iii) context and priority definition, (iv) data classification. The proposed methodology is aimed at extracting and analyzing crime-related features concerning the user's habits, skills and interests from digital devices, and categorizing them accordingly.

Overill, Kwan, Chow, Lai and Law (Overill *et al*., 2009) proposed a two-phase digital forensic investigation model which achieved the twin goals of reliability and cost-effectiveness by incorporating a pre-processing phase which runs in parallel with the data collection phase. In the first phase (pre-processing), the methodology of detecting traces is used. In the second phase, a Bayesian Network is used to analyze the traces.

Peersman, Schulze, Rashid, Brennan and Fischer (Peersman *et al*., 2016) presented a new intelligent forensics approach which incorporates the advantages of the artificial intelligence and machine learning theory to automatically flag new/previously unseen child sexual abuse (CSA) media to investigators. The presented iCOP toolkit consists of two key components: the filename categorization module, and the media classification module. The filename and image classification approaches are synthesized in the iCOP toolkit in order to identify and prioritize new/previously unknown CSA media.

Roussev and Quates (Roussev and Quates, 2012) introduced content triage with similarity digests and demonstrated that by applying similarity digests in a systematic manner, the scope of examination can be narrowed down within a matter of minutes to hours. The authors used a sizeable case study to demonstrate the utility of similarity digests as a triage tool and showed that a single and simple data correlation tool (*sdhash*) can provide a systematic and efficient path to triage with minimal assumptions and knowledge of the case. The experiment of these authors showed that in

approximately 20 min, they were able to reach the following preliminary conclusions: a) yes, the initial suspicion of smuggling is correct; b) only Jo's computer is involved; c) the images were likely introduced on Nov 18; d) the USB drive is the apparent mode of transmission; and e) there is every reason to believe that somebody (most likely Jo) deliberately placed kitty material on his computer.

We evaluated all the overviewed methods, models and frameworks targeting cybercrime forensic investigation according to the selected characteristics, and the results are presented in Table 4.

Table 4 Summary of methods, models and frameworks for cybercrime forensic investigation

Proposed	Starting point	Model type	Evaluation equation	Results	Experiment	Reference
Ranking algorithms for digital forensic string search hits	Eighteen features as quantitative indicators of search hit relevancy	Support vector machine (SVM) models TF-IDF: used normalized, logarithmic, corpus level term frequency	Recall Precision Average precision	Significant improvements in information retrieval effectiveness with rank-ordered list output.	Synthetic case 'M57 Patents'	(Beebe & Liu, 2014)
Model for advanced digital forensics timeline analysis SADFC	Not applicable	Knowledge Model Investigation Process Model, Ontology-centred architecture	Subject, object, event and footprint. Allen algebra used.	SADFC approach allows to reduce the tedious character of the analysis.	Case study and designed a fictitious investigation	**(Chabot *et al.*, 2014)**
Answer Set Programming (ASP) approach	Realistic investigative cases	Artificial Intelligence (AI) and Automated Reasoning	Computational Logic	Explain ability and accountability are in fact of particular importance in this field.	Case study	(Costantini *et al.*, 2019)
An Intelligent Forensic Framework towards Cloud	VM snapshots provided by VMM (Virtual Machine Manager)	Ontology-Enabled Forensic Blackboard (OFB) and Ontology-Enabled Forensic Controller and Processor (OFCP).	Web Ontology Language (OWL)	The proposed framework is an approach towards mitigating the problem of investigating cloud crime incident manually rather giving it an ability of automation.	Not defined	(Datta & Pan, 2016)
Integrated Digital Forensics Investigation Framework (IDFIF)	Testing IDFIF version 3 performed by applying on a Distributed Denial of Service (DDoS)	A strategy approach phase for adapting case handling.	Disadvantages Advantages	IDFIFv3 comprehensive and useful for network investigation than the other existing model.	IDFIFv2, EMCI, HOBF, EDIMP, CFFTPM, GFNI	**(**Hikmatyar, Prayudi, & Riadi, 2017)
Case-Based Reasoning Forensic Triager (CBR-FT) method	Is given an evidence relevance rating (ERR) by the investigating practitioner.	CBR-FT uses the ERRs as a prior probability distribution in a Bayesian model to determine the priority of particular locations for searching during DT (device triage).	Based on probabilities $P(L\|E) = \frac{P(L\|E)P(L)}{P(L\|E)P(L) + P(E\|-L)}$	Higher values of Recall and Precision	EnCase commercial tool recall (in 15 of the 20 cases) was as good as or better than the commercial tool. In 17 cases precision was higher than that of EnCase, that	(Horsman *et al.*, 2014)

Proposed	Starting point	Model type	Evaluation equation	Results	Experiment	Reference
					is, CBR-FT recovered more evidence and less non-relevant data than nCase.	
FORensics ZAchman (FORZA) framework	Outlined eight different roles and their responsibilities in a digital forensics investigation.	What Why How Who Where When	Reconnaissance, Reliability and Relevancy	Using this framework, questions and answers in a digital forensics investigation could be systematically thought through	A web hacking case	(Ieong, 2006)
Ontology Framework for Automation Digital Forensics Investigation	Ontology Framework.	This model uses a structured hierarchy of layers that create connectivity between the variant and searching investigation of activity	Two layers hierarchical structure where the layer has atomic construction in Technology focused on Hardware and Software	The simplicity of the mechanism and rules that are used	Not defined	(Luthfi, 2014)
A Machine Learning-based Triage methodology	The preliminary step of the categorization process is to train a Machine Learning classifier	Bayes Networks, Decision Trees, Locally Weighted Learning and Support Vector Machines	precision= TP/(TP/FP) recall= TP/(TP+FN) f-measure= 2*recall*precision	Workflow consisting of four phases: forensic acquisition, feature extraction and normalization, context and priority definition, data classification, is aimed at extracting and analyzing crime-related features concerning user's habits, skills and interests from digital devices, and categorizing them accordingly	Two case studies conducted on the crimes of copyright infringement and child pornography exchange	**(**Marturana & Tacconi, 2013)
Model for conducting swift, practical and cost-effective digital forensic investigations	Pre-processing – Detecting Traces	Bayesian Network – Analyzing Traces	cumulative evidentiary weight $W = \sum_{i=1}^{m} \omega_i$ relative fractional evidentiary weight $\boldsymbol{\omega_i}$ of each trace Ti is	Twin goals of reliability and cost-effectiveness. Obtained probability value of 0.94 ± 0.06.	BitTorrent Case Study	(Overill *et al.*, 2009)

Proposed	Starting point	Model type	Evaluation equation	Results	Experiment	Reference
			either assigned by an expert or, by default, is set to one.			
A new intelligent forensics approach	Support Vector Machines (SVM) to Naive Bayes (NB) and Logistic Regression (LR).	Incorporates the advantages of artificial intelligence and machine learning theory	Recall, Precision, Fscore $F_{core} = 2 \frac{Precision \cdot Reca}{Precision + Rec}$	iCOP toolkit that is designed to highlight sharers of new/previously unknown child sexual abuse media in P2P networks.	Adopted a highly skewed data distribution	(Peersman et al., 2016)
Content triage with similarity digests	Disk images, network captures, RAM snapshots, and USB flash media that consists of 1.5 TB of raw data	Similarity digests	Generation performance in Data Set Size (GB) Time (min) Rate (MB/s)	Single and simple data correlation tool (sdhash) can provide a systematic and efficient path to triage with minimal assumptions and knowledge of the case.	A sizeable case study of the 2009-M57-Patents scenario	(Roussev & Quates, 2012)

The provided overview and comparison of twelve relevant methods, models and frameworks for cybercrime forensic investigation (see Table 4) forms a general view on the currently existing solutions. During the comparison, we noted that, mostly, the used evaluations are recall, precision, f-measure, which are based on TP, TN, FP, FN and could apply to evaluate the proposed models in the dissertation. Also, the case study is mostly the usable strategy for the experiments.

The results showed that most of the analyzed solutions propose various methods, models and frameworks for the main purpose to reduce the tedious character of the analysis thus mitigating the problem of investigating cloud crime incident manually; instead, the researchers are seeking possibilities of employing automation and getting better cost-effectiveness as well as reliable results. This proves the idea there is lack of methods, models and frameworks for cybercrime forensic investigation which would consider reducing the amount of analyzed data and supporting experts in terms of evidence evaluation when making decisions.

1.10 Critical analysis and discussion

Modern digital forensics is expected to to run effective digital investigations in computer systems, handheld (mobile) devices, peripheral (fog-edge) devices, computer networks, cloud, and IoT devices. Analysis in the above mentioned systems is often time-consuming, highly complex, and costly. In addition, modern digital forensics requires from the experts a high level of skills, knowledge, and the use of specialized software when facing the need to analyze some amount of digital data. Roy (Roy, Dixit, Naskar, & Chakraborty, 2020) argues that R&D must make vigorous effort in order to produce efficient forensic techniques because of the cyber criminals' ever-increasing activities in the digital world and that the process may render the present digital forensic state-of-the-art techniques obsolete.

In order to compare different forensics investigation models, a set of investigation criteria was defined (Takwa, Belgacem, & Adel, 2016): (i) digital evidence time gaining, (ii) digital evidence transparency and privacy, (iii) digital evidence reliability and consistency, (iv) digital evidence reusability. The digital evidence time gaining criteria were firstly defined as one of the most valuable criteria besides others. The authors expressed hope that the new proposed comparison of models would help prospective research participants understand many incidents.

Extracting digital evidence and then using the established traces of crime is a challenge. In the previous chapters, the analysis of assignment, profiling and forensics domain was complied, as well as the research of the areas of computer systems where the user left some digital images. These areas can include computer users, their home directories, files and folders, nicknames, and more. I have reviewed tools, methods, frameworks and models which can be applied in relation to digital evidence investigation from the domain of digital forensics.

As a result, I propose five groups of digital evidence investigation for methods, frameworks and models to use:

1. General Digital Investigation (GDI);
2. Computer Examination (CE);

3. Digital Evidence on the Network (DEN);
4. Digital Evidence on the Cloud (DEC);
5. Others (O) – Internet of Things, Mobile devices, AI, Social Networks.

The results of our analysis are summarized in Figures 5–6.

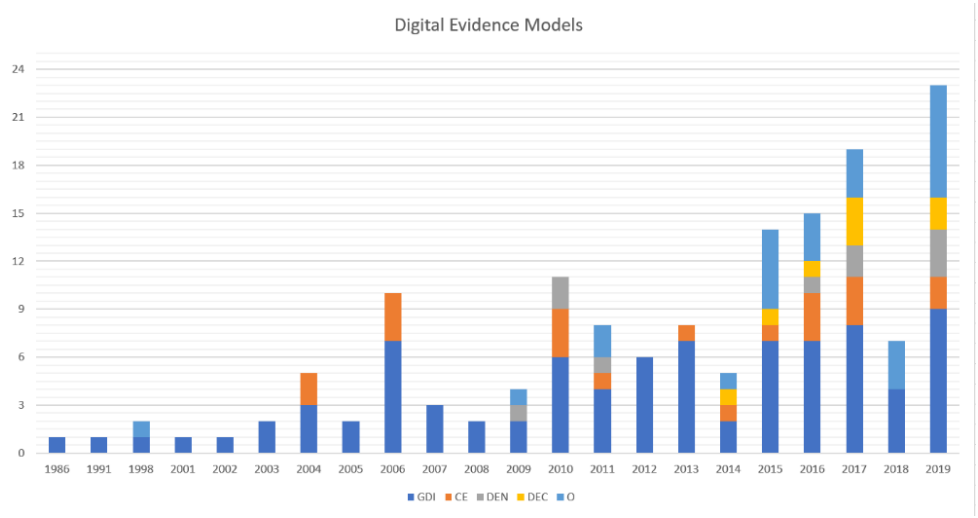

Fig. 5. Evaluation of models used for cybercrime forensic investigation based on five groups of digital evidence investigation

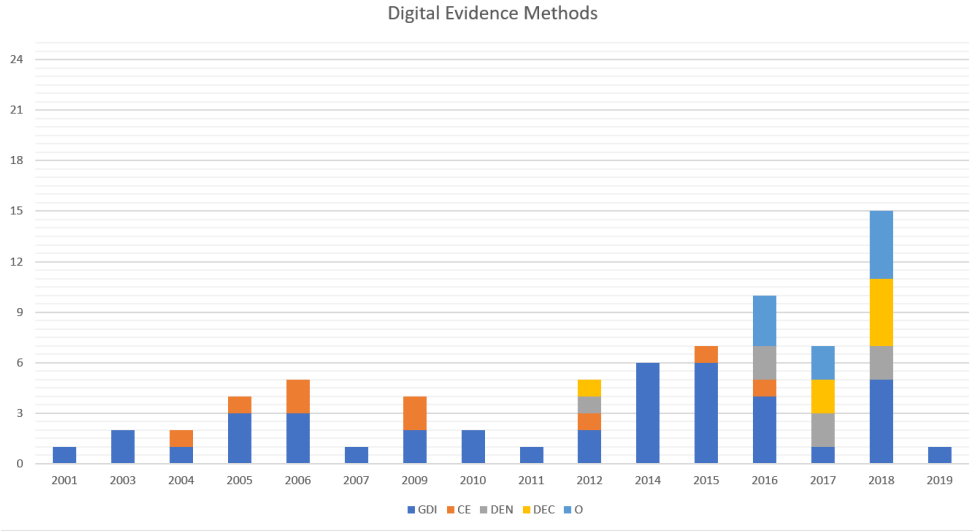

Fig. 6. Evaluation of methods used for cybercrime forensic investigation based on five groups of digital evidence investigation

A summary of results shows that the number of new digital investigation methods for computer examination (CE) has declined significantly over the last five years. There is also a growing trend towards models which focuses on the digital

evidence investigation on the cloud (DEC). A striking trend observed when looking at the models is the research on models for general digital investigation (GDI).

Quite opposite is the case with methods. In 2019, research on new methods declined significantly. I make an assumption that it is due to the amount of data that has been growing exponentially, and there is need to do research and have new methods developed which will be able to quickly process such quantities of data.

In addition, the methods simplify the task but do not respond to the quarrel: who made the cybercrime. Digital evidence should indicate the suspect as accurately as possible. In such a case, there is lack of attention committed to ensuring new models that can help experts in digital evidence investigation and decision making when answering the question: who committed the cybercrime?

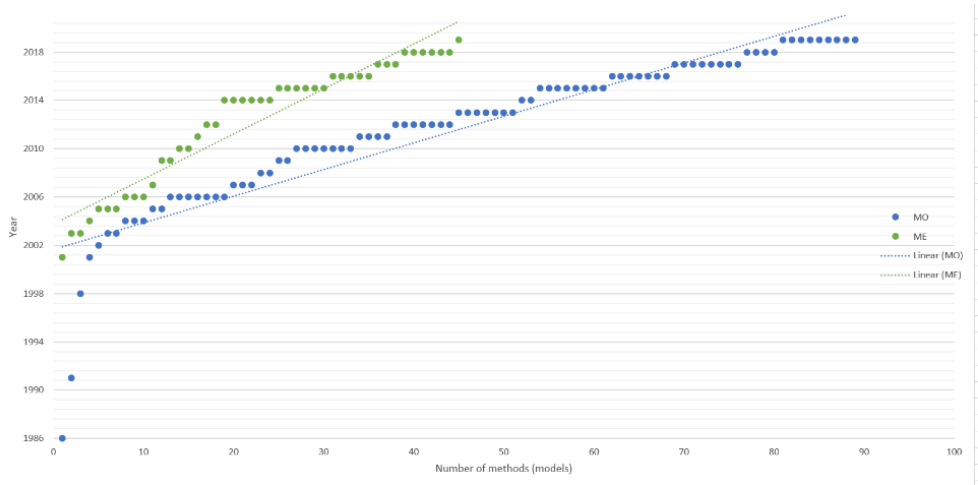

Fig. 7. Annual trend in models and methods for digital evidence investigation research

With regard to the annual trend (see Fig. 7) in the methods and models, I are clearly seeing a proliferation of methods in the future that will facilitate investigations in cybercrime and reduce time costs.

1.11 Conclusions of the First Chapter and Formulation of the Objectives of the Thesis

The first chapter of this thesis provides an overview of digital evidence, forensic investigation process and the already existing research in this area. More significant attention is dedicated to the analysis of existing tools, methods, frameworks and models used for digital evidence investigation. The following conclusions have been drawn:

1. Systemic analysis of scientific papers in the field of cybercrime forensics investigation domain revealed that the majority of the papers in the period from 2002 to 2019 were written on the topic of methods and models for general digital investigation. The importance of this topic in the scientific society showed that this problem is highly important. However, there is still lack of methods unifying the objects of digital evidence prioritizing the evidence recovery with the reduction of time and cost of digital

investigation.

2. There is a huge number of available computer forensic tools from standalone packages to complex integrated tools developed for wide range crime investigations. If so, the first question (before starting the investigation) is how to choose the appropriate tool for investigation that will be the right one in that particular case. Choosing the right tool as early as in this phase can significantly reduce the time of the investigation.
3. The currently existing methods, models and frameworks for cybercrime forensic investigation were compared. The comparison proved that there is no single superior method, model or framework which would be able to cover the exponential growth of the amount of digital information in the field of the main cybercrime forensics related areas. Therefore, a new, more holistic cybercrime forensic investigation method is needed to concentrate on reducing the expertise time and cost. In the course of our comparison, I noted that the most-used evaluations are recall, precision, f-measure, which are based on TP, TN, FP. Also, the case study is the most usable strategy for the experiments.
4. Another issue faced by modern digital forensics is the need to design effective methodologies and develop efficient tools. The results showed that most of the analyzed tools are suitable for computer system forensic investigation and oriented to extracts from the computer system wide range raw data. This proves the idea that there is lack of tools that would be considered as assistance to experts in the context of cyber situation awareness and help in the digital evidence investigation process.

Based on the conclusions, the following tasks are formulated to achieve the aim:

1. To propose a new transformation system for the digital forensics domain that would assist and facilitate computer forensic experts as the right tool for digital evidence investigation selection.
2. To propose a new model for digital evidence investigation while using the habits attribution profiling method through analyses of attribution, profiling and habits domains with the aim to decrease the number of the objects in the search sequences from the set of the digital user places.
3. To propose a new digital evidence object model which should combine the crime investigation process with the object-oriented programming model informatively in order to reduce the amount of data extracted for the experts for analysis and to help experts in decision making when requiring to answer the question: who committed the crime?

2 PROPOSED MULTI-LAYERED TRANSFORMATION SYSTEM FOR THE DOMAIN OF DIGITAL FORENSICS

In general, ontology creates a common vocabulary to analyze the domain information within a certain area. Therefore, by creating the ontology, it is possible

to form the common information structures, to reuse the knowledge, to make assumptions within the domain, and to analyze every piece of knowledge. The aim of this chapter is to create an ontology transformation model and a system for the digital forensics domain that enables separate formulation and incorporation of domain-specific concepts as ontologies, and rules that refer to those ontologies to be developed for computer forensic experts in their respective domains. I strive to assist and facilitate the work of computer forensic experts by providing the right tool for digital evidence investigation selection; I propose the transformation model and a multi-layered architecture of the ontologies transformation system. I consider the use of the XDT transformations attributes in XML transform files to reflect the cyber forensic ontology to the National Institute of Standards and Technology proposed Computer Forensic Tool Catalog taxonomy.

The analysis and research presented in this chapter was published in (Grigaliunas *et al*., 2017).

2.1 Cyber forensics ontology and a two-stage transformation model

Collection, examination, analysis and reporting are the main activities in the digital forensic evidence investigation process. Preparation will assist in selecting the right tool, fulfilling the necessary legal requirements, deciding the level of management, and arranging the necessary support (Saleem, Popov, & Bagilli, 2014). The National Institute of Standards and Technology (NIST) realized the need for searching the forensics tools by technical parameters based on specific forensics tool functionality and proposed the Computer Forensic Tool Catalog (NIST, 2019). The primary goal of the Tool Catalog is to provide an easily searchable catalog of forensic tools. In addition, NIST proposed a forensic tool taxonomy based on forensic tool functionalities.

In (Brinson, Robinson, & Rogers, 2006) the cyber forensics ontology was presented that consists of a five-layer hierarchical structure with the resulting final layer being the specified areas for certifying and specializing. Those layers belong to the technology and profession domains and are described as follows: hardware, software (the technology domain), law, academia, military, private sector (the profession domain).

Two domains were analyzed: cyber forensics ontology (CFO) (Brinson *et al*., 2006) and NIST computer forensic tool catalog (CFTC) (NIST, 2019). The relationship between the analyzed domains is depicted in is as shown below:

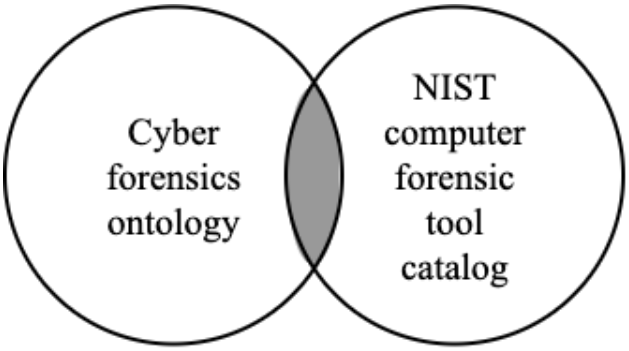

Fig. 8. Relationship between the cyber forensic ontologies and NIST computer forensic tool catalog.

As mentioned above, at the preparation stage, a computer forensic expert makes the significant decision of selecting the right tool for digital evidence investigation. On the one hand, the NIST CFTC proposes many suitable tools classified by their functionality, and, on the other hand, the computer forensic expert uses cyber forensic ontology. As shown in Fig. 8. *Relationship between the cyber forensic ontologies and NIST computer forensic tool catalog.* only a small number of domains intersects.

To assist and facilitate the computer forensic expert regarding the adequate tool for digital evidence investigation selection, ontology was created (CFTCO) by using the NIST CFTC, and the proposed model is depicted in Fig. 9. *Ontology-based transformation model.*

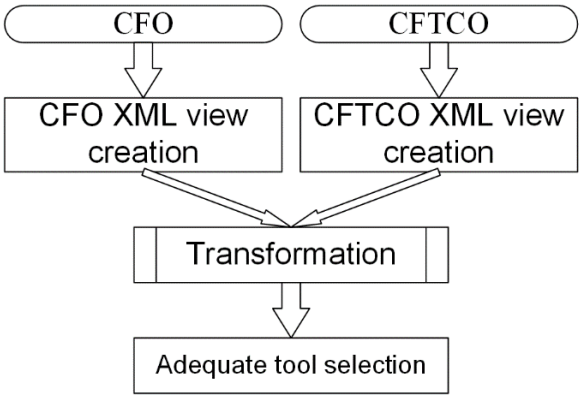

Fig. 9. Ontology-based transformation model.

An ontology-based transformation model (OBTM) consists of two stages (see Fig. 9. *Ontology-based transformation model.*). The first stage relies on XML view creation for the selected ontologies, whereas the second stage focuses on the development of transformation rules. Transformation rules will reflect XML views regarding the adequate tool selection for further use in the digital evidence investigation process.

Formally, an ontology is a pair $O = (D, R)$, where D is the domain and R is a

set of relations defined in D and $R \subseteq D^n$.

To define the model, I formulate the following set A of ontology axioms:

Axiom A1. If O_{cf} is computer forensic domain ontology, and E_{xcf} is the corresponding custom XML elements domain of CFO, then there exists a function $f_{cf}: O_{cf} \rightarrow E_{xcf}$.

Axiom A2. If O_{cftc} is computer forensic tool catalog domain ontology, and E_{xcftc} is corresponding custom XML elements domain of CFTCO, then there exists a function $f_{cftc}: O_{cftc} \rightarrow E_{xcftc}$.

Axiom A3. If E_{xcf} is custom XML elements domain of CFO, and O_{cftc} is computer forensic tool catalog domain ontology, then there exists function$g_{cf} \circ f_{cf}: E_{xcf} \rightarrow O_{cftc}$.

Axiom A4. If E_{xcftc} is custom XML elements domain of CFTCO, and O_{cf} is computer forensic domain ontology, then there exists function $g_{cftc} \circ f_{cftc}: E_{xcftc} \rightarrow O_{cf}$.

I define an OBTM as follows:

$$TM(O) = \left(O_{cf}, f_{cf}, E_{xcf}, O_{cftc}, f_{cftc}, E_{xcftc}, g_{cf}, g_{cftc}\right) \quad (4)$$

2.2 Framework to develop an ontology-based transformation system

Extensible Markup Language (XML) is a mark-up language for documents encoding in a format that is human-readable and machine-readable. It is widely used in computer systems and applications.

The OWL 2 Web Ontology Language, informally OWL 2, is an ontology language for the Semantic Web with a formally defined meaning. OWL 2 ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents (W3C, 2019).

Digital Forensics XML (DFXML) is an XML language designed to represent a wide range of forensic information and forensic processing results; because of its abstractions to the needs of forensics tools and analysts, DFXML allows the sharing of structured information between independent tools and organizations (Garfinkel. 2012).

Cyber Observable eXpression (CybOX™) is a standardized language for encoding and communicating high-fidelity information about cyber observables (MITRE, 2019). It uses XML schemas for creating objects and other resources.

In ASP.NET, most applications have settings in the Web.config (Microsoft, 2019) file that must be different when the application is deployed. To automate the process of changing the Web.config file when publishing (deploy) a Visual Studio web project to different destination environments, XML document transformations (XDT) are used (Microsoft, 2019) Web.config Transformation Syntax for Web Project Deployment Using Visual Studio.

A transform file is an XML file that specifies how the Web.config file should be changed when it is deployed. Transformation actions are specified by using XML attributes that are defined in the XML-Document-Transform namespace, which is

mapped to the XDT prefix. The XML-Document-Transform namespace defines two attributes: the Locator and the Transform. The Locator attribute specifies the Web.config element or a set of elements that the user wants to change in some way. The Transform attribute specifies what the user wants to do to the elements that the Locator attribute is going to find (Muniswamy-Reddy *et al*., 2006). The Locator and Transform attributes and their meanings are shown in Table 5 XDT Locator and Transform attributes.

Table 5 XDT Locator and Transform attributes.

XDT attribute	Attribute parameters	Explanation
Locator	Condition	Specifies an XPath expression that is appended to the current element's XPath expression
	Match	Selects the element or elements that have a matching value for the specified attribute or attributes
	XPath	Specifies an absolute XPath expression that is applied to the development Web.config file
Transform	Replace	Replaces the selected element with the element that is specified in the transform file
	Insert	Adds the element that is defined in the transform file as a sibling to the selected element or elements
	InsertBefore	Inserts the element that is defined in the transform XML directly before the element that is selected by the specified XPath expression
	InsertAfter	Inserts the element that is defined in the transform XML directly after the element that is selected by the specified XPath expression
	Remove	Removes the selected element. If multiple elements are selected, removes the first element
	RemoveAll	Removes the selected element or elements
	RemoveAttributes	Removes specified attributes from the selected elements
	SetAttributes	Sets attributes for selected elements to the specified values

To realize the proposed model for two-way ontology transformations in the development of the Ontology-Based Transformation System (OBTS), I use XDT transformations attributes in XML transform files.

2.3 Architecture of an ontology-based transformation system

In this section, I describe the multi-layered architecture of the OBTS and its component for assisting the computer forensic expert in the adequate tool selection for digital evidence investigation. The architecture of the OBTS is depicted in *Fig. 10. Architecture of the Ontologies Transformation System.*

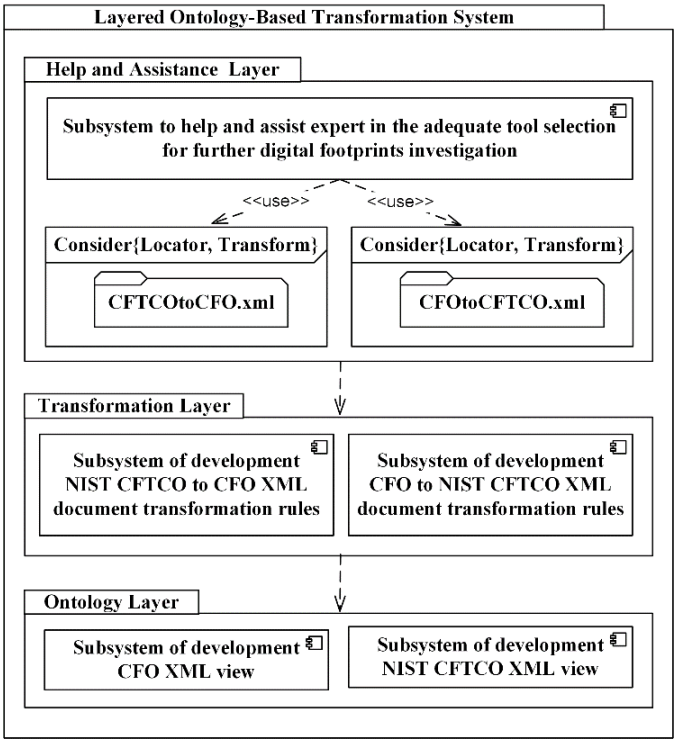

Fig. 10. Architecture of the Ontologies Transformation System.

The OBTS has the following components for supporting two-way ontology transformations:

1. Subsystem of development cyber forensics ontology view intended for cyber forensics ontology view and XML encoding creation.
2. Subsystem of development NIST CFTCO view intended for NIST CFTCO view and XML encoding creation.
3. Subsystem of XML transformation rules development from cyber forensics ontologies to NIST CFTCO, intended for XDT rules generation using Locator and Transform attributes and saving them to the CFOtoCFTC.xml file.
4. Subsystem of XML transformation rules development from NIST CFTCO to cyber forensics ontologies intended for XDT rules generation by using Locator and Transform attributes and saving them to the CFTCtoCFO.xml file.
5. Subsystem that uses CFTCOtoCFO.xml, CFOtoCFTCO.xml files and is intended to help and assist the computer forensic expert with the right tool for digital evidence investigation selection.

As a case study of the proposed ontologies transformation system, I present XML documents transformation example from CFO to CFTCO.

At the ontology layer, XML views are developed. I use a fragment of the proposed in (Brinson *et al*., 2006) ontology as an example of CFO (see Fig. 11). The developed XML view is depicted in Figure 12.

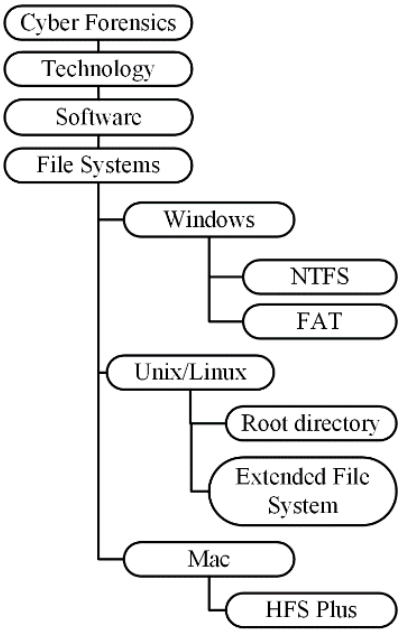

Fig. 11. CFO example

```
<?xml version="1.0" encoding="UTF-8"?>
- <technology>
  - <software>
    - <filesystem>
      - <operatingsystem name="Windows">
          <filesystemtype>NTFS</filesystemtype>
          <filesystemtype>FAT</filesystemtype>
        </operatingsystem>
      - <operatingsystem name="Unix_Linux">
          <filesystemtype>Root_directory</filesystemtype>
          <filesystemtype>Extended_File_System</filesystemtype>
        </operatingsystem>
      - <operatingsystem name="Mac">
          <filesystemtype>HFS_Plus</filesystemtype>
        </operatingsystem>
      </filesystem>
    </software>
</technology>
```

Fig. 12. CFO representation in XML view

I create the CFTCO from NIST proposed forensic tool taxonomy, and, for our case study, I select the file carving domain (see Fig. 13. *CFTCO file carving domain*). The developed XML view is depicted in Fig. 14. *CFTCO representation in XML view*.

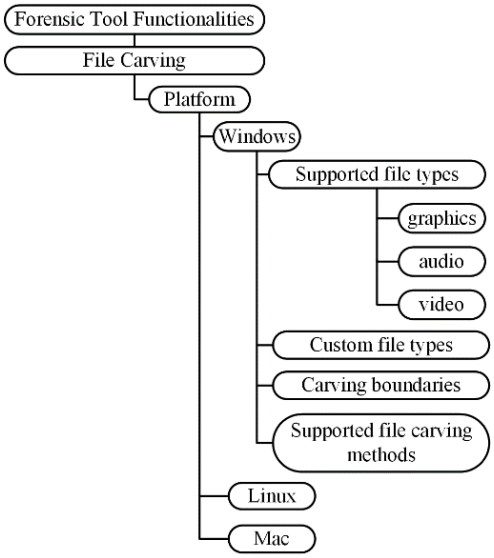

Fig. 13. CFTCO file carving domain

```
<?xml version="1.0" encoding="UTF-8"?>
- <filecarving>
  - <platform name="Windows">
    - <supportedfiletypes name="graphics">
        <fileformat>jpg</fileformat>
        <fileformat>png</fileformat>
        <fileformat>bmp</fileformat>
        <fileformat>gif</fileformat>
      </supportedfiletypes>
    + <supportedfiletypes name="audio">
    + <supportedfiletypes name="video">
    + <customfiletypes>
    + <carvingboundaries>
    + <supportedfilecarvingmethods>
    </platform>
  + <platform name="Linux">
  + <platform name="Mac">
</filecarving>
```

Fig. 14. CFTCO representation in XML view

At the transformation layer, I are using XDT locator and transform attributes (see Table 5 XDT Locator and Transform attributes.), and XML document transformation rules are developed. In our case study, I developed a transformation rule example shown in Fig. 15. *XML document transformation rule from CFOtoCFTCO.xml*

Fig. 15. XML document transformation rule from CFOtoCFTCO.xml

At the Help and Assistance Layer XML document transformation, rules developed at the transformation layer are applied. The result is depicted in Fig. 16. *XML document of the CFO fragment.*

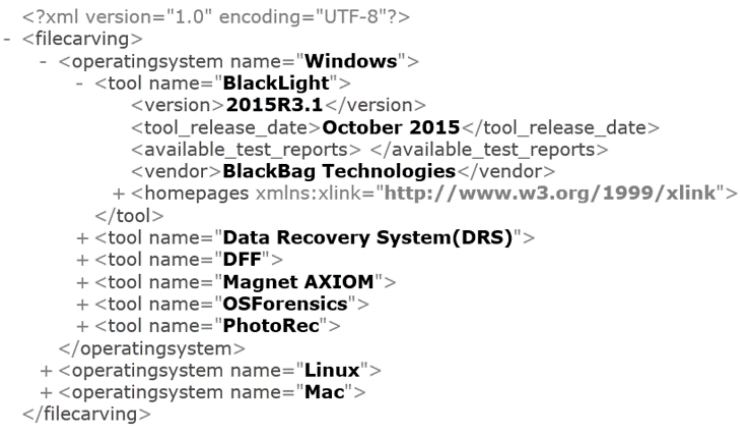

Fig. 16. XML document of the CFO fragment

After XML document transformation rules have been applied and transformed, the XML document is ready to use the subsystem, which will help and assist the computer forensic expert as the right tool for digital evidence investigation selection by showing the suitable tool names. OPML is an XML-based format which allows exchange of outline-structured information between applications running on different operating systems and environments. Portable digital format (pdf) readers

can open an OPML file. In the proposed OBTS, I use OPML to encode the transformed XML files which are exported and then opened with the pdf reader depicted in Fig. 17**.** *OPML file exported and opened with pdf reader*.

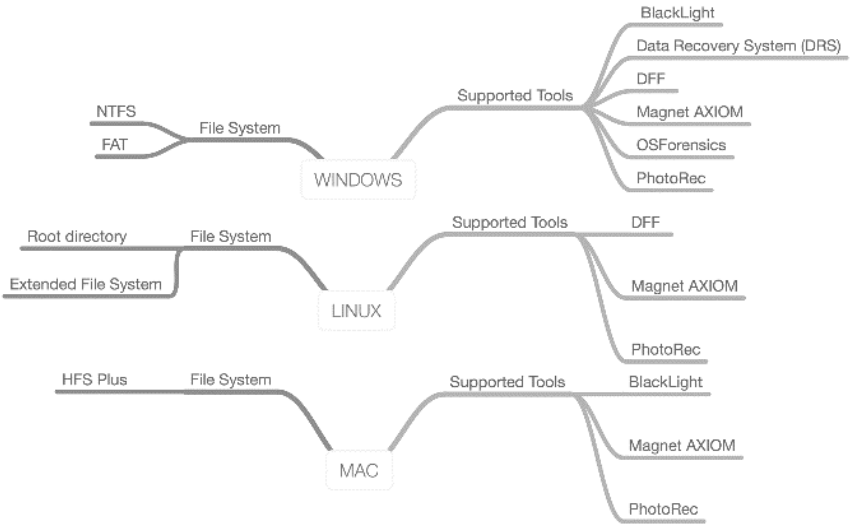

Fig. 17. OPML file exported and opened with pdf reader.

The subsystem also holds homepage URL for every proposed tool from the list. The CFO and CFTCO have created common definitions in the digital forensics domain. While both tools belong to the same digital forensics domain, they are very different, and only a small amount of artefacts, when expressed through ontologies, intersect. Typically, computer forensics experts operate in terms of the CFO, but the NIST taxonomy of forensic tools for digital evidence investigation is given in CFTC terms. In this chapter, I consider three challenging tasks:

1.to propose a two-stage model for the transformations of ontologies from CFO to CFTCO and vice versa;

2.to suggest XML document transformations (XDT) to map CFO and CFTCO representations from one to the other;

3.to develop a multi-layered architecture and ontology-based transformation system (OBTS) in which the proposed model and XDT are realized.

In the case study, I create a set of transformation rules and show that OBTS transforms the CFO to the NIST tool list and is able to assist computer forensics experts in selecting an appropriate tool for further digital evidence investigation.

2.4 Conclusions of the Second Chapter

1. The analysis of tools for cybercrime forensic investigation revealed that there is a huge number of available computer forensic tools and no solutions how to select the appropriate tool. CFO and CFTCO have created common definitions in the digital forensics domain. While both tools belong to the

same digital forensics domain, they are very different, and only a small amount of artefacts, when expressed through ontologies, intersect. Typically, computer forensics experts operate in terms of CFO, but the NIST taxonomy of forensic tools for digital evidence investigation is given in CFTC terms.

2. The proposed two-stage transformation model provides a synergistic approach conducting the computer forensics expert's domain (CFO) with the NIST taxonomy of forensic tools domain (CFCTO) for digital evidence investigation transforming ontologies from CFO to CFTCO and vice versa. In the model, an XML view creation is proposed for CFO and CFCTO ontologies, which ensures a highly flexible semi-structured document with a tag-driven structure that is adaptable to universalize the data structure. The proposed model is adapted to cover XML document transformation rules (XDT) to map CFO and CFTCO XML views from one to the other.
3. The proposed multi-layered architecture and ontology-based transformation system (OBTS) in which the proposed model and XDT are realized can serve these experts who operate in terms of the forensics domain with regard to reducing the time needed for the appropriate tool for digital evidence investigation selection from the NIST tool catalog. In the case study, I created a set of transformation rules and showed that OBTS transforms the CFO to the NIST tool list and is able to assist computer forensics experts in selecting the appropriate tool for further investigation of digital evidence.
4. A multi-layered transformation system and a two-stage transformation model for the digital forensics domain are distinguished from the currently existing solutions by their integration of the important CFO and CFCTO domains ontologies and the defined XDT transformations between them.

3 PROPOSED MODEL FOR DIGITAL EVIDENCE INVESTIGATION USING THE HABITS ATTRIBUTION PROFILING METHOD

The acquisition of digital remnants and their use in order to find cybercrime evidence in the digital user places (device, profile, home directory, etc.) is the challenge of this Chapter. Through analyses of attribution, profiling and habits domains, in this chapter, I propose an approach for modelling certain issues of digital evidence investigation. The proposed model focuses on digital evidence investigation that uses the habit attribution profiling method in order to decrease the number of the objects in search sequences from the set of digital user places.

The analysis and research presented in this chapter was published in (Grigaliunas & Toldinas, 2016).

3.1 General framework for the analysis and digital evidence investigation of cybercrime

A general framework for the analysis and digital evidence investigation of cybercrime is depicted in Fig. 18. *General framework for the analysis and*

acquisition of digital evidence. At the core of the framework there is the Y-diagram which links together three basic domains: attribution, profiling and habits. The framework also outlines a context of the selected domains: for the attribution domain – the metadata and other logs can be used to attribute actions to personality identification, for profiling domain – profiles help reconstruct the crime when there are too many unknowns, and for the habits domain – an approach to examining and classifying user habits is suggested.

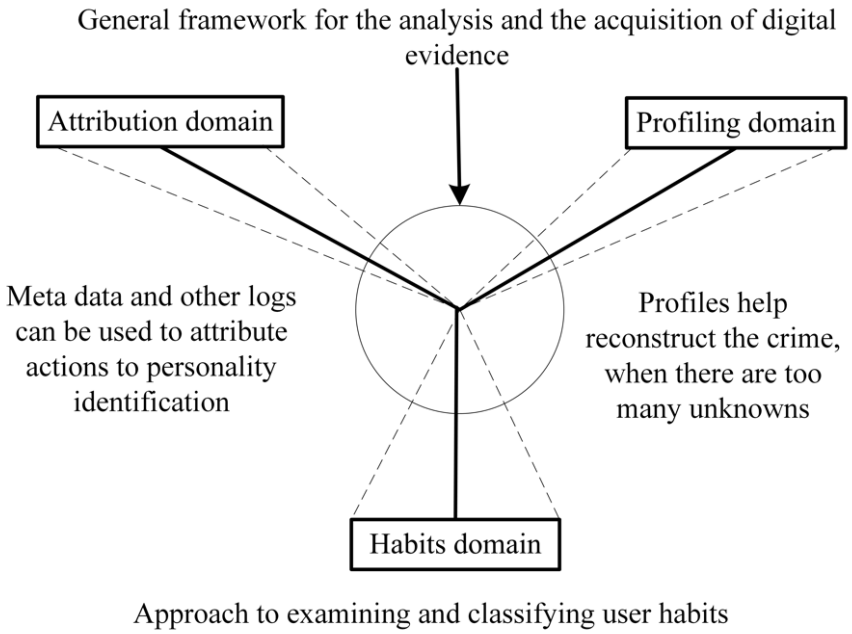

Fig. 18. General framework for the analysis and acquisition of digital evidence

The intersection between the domains (presented by a circle in Fig. 18. *General framework for the analysis and acquisition of digital evidence*) identifies the space for the proposed attributed habits profiling method. The space of the attribution domain is to be considered for each user's habit from the digital habits domain. The attribution domain (adopted from (Fiske & Taylor, 1991)) is depicted in Fig. 19**.** *Attribution domain*.

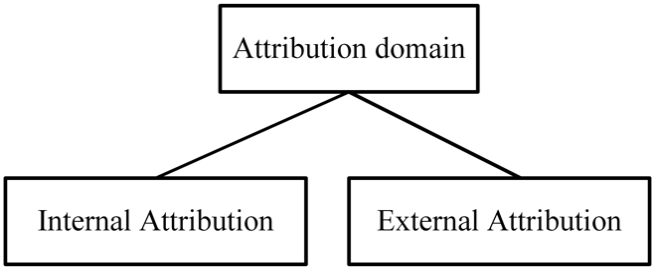

Fig. 19. Attribution domain

If we look at the attribution domain, it includes two subdomains. The Internal Attribution subdomain is the process of assigning the cause of behaviour to some internal characteristic, rather than to outside forces. When we explain the behaviour of others, we look for enduring features, such as personality traits. For example, we attribute the behavior of a person to their personality, motifs or beliefs (Fiske & Taylor, 1991). The External Attribution subdomain is the process of assigning the cause of behaviour to some situation or event outside a person's control rather than to some internal characteristic. When we try to explain our own behaviour, we tend to make external attributions, such as situational or environment features (Fiske & Taylor, 1991).

The space of the profiling domain has two main approaches: basic approaches to profiling and criminal profiling models. Profiling domain (adopted from (Douglas, Ressler, Burgess, & Hartman, 1986), (Nykodym, Taylor, & Vilela, 2005), (Rogers, 2003), (Štuikys, Burbaitė, & Bespalova, 2015) is depicted in Fig. 20**.** *Profiling domain.*

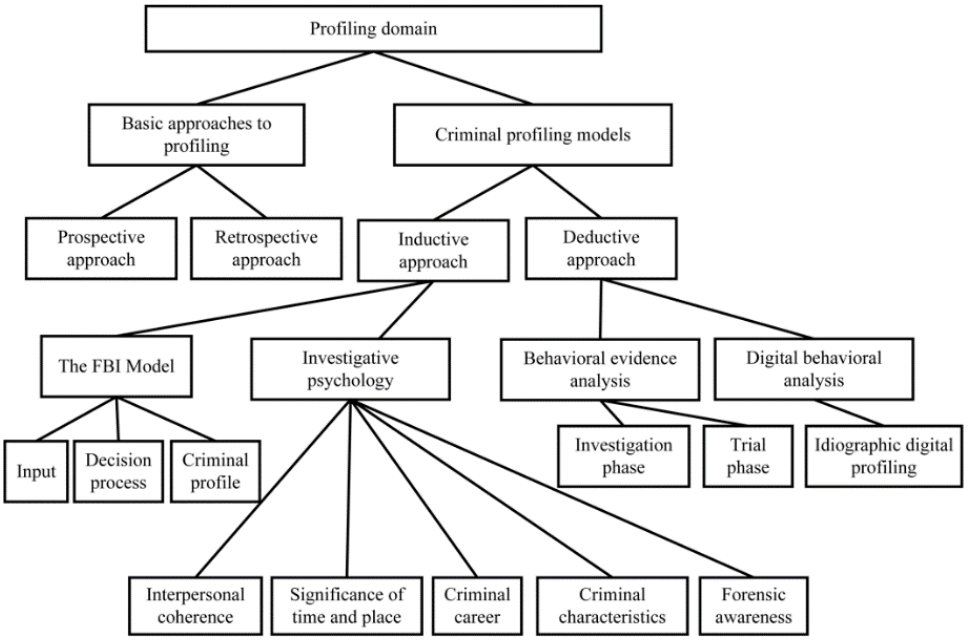

Fig. 20. Profiling domain

Basic approaches are usually common in the profiling domain. A template of an offender for a type of crime is used to narrow down the pool of suspects in prospective approach, and in retrospective approach, the details of a crime are used to produce a description of the offender (Nykodym *et al*., 2005).

Criminal profiling models manifest two main approaches – inductive and deductive. Statistical information about previously committed crimes is used to generalize the behaviour exhibited in these crimes in the inductive approach, whereas, in the deductive approach, the offender's characteristics are derived from the specific case (Rogers, 2003). The main weakness of inductive approaches are the

immeasurable error rates caused by the variable levels of honesty and perception bias of the questionnaire respondents, or by case studies taken from a wide range of time and increased possibility of unpredictable external influences on the case (Štuikys *et al*., 2015). A deductive method takes a more personal approach as it examines every case contextually (Turvey, 2012).

Methodologies following the inductive approach are as follows: the FBI model (Douglas *et al*., 1986), and investigative psychology (Canter, 2004). The FBI model consists of three stages: input, decision process, and criminal profile. The crime scene is assessed, and evidence is collected in the input stage. In the decision process, the input is organized and then analyzed in order to establish the patterns, and to conduct the crime scene assessment that deals with the crime. In the criminal profile, the offender is described, and the criminal profile is used to aid the investigation (Douglas *et al*., 1986).

Investigative psychology uses statistics from offender databases, and, when building the profile, it relies on the following factors: interpersonal coherence, significance of time and place, criminal characteristics, criminal career and forensic awareness (Rogers, 2003). This model is mostly suited for offline crimes, however, in digital investigation, not all of these factors have equal significance (e.g., estimating the offender's forensic awareness (Canter, 2004) can help estimate the offender's level of technical skill).

The behavioural evidence analysis methodology (Turvey, 2012) is the deductive profiling approach which consists of two phases: the investigation phase occurs when there is a criminal event, but the offender is unknown, and the trial phase occurs when the offender becomes known.

Steel (Steel, 2014) first proposed an idiographic approach to digital profiling by examining particular subjects, Internet activities and electronic media for the purposes of using digital footprints left behind for immediate use in an ongoing investigation. The guidance presented in Steel's paper (Steel, 2014) is provided to investigators to assist in creating an idiographic digital behavioural profile in active criminal cases. The profile can be developed iteratively and refined during the course of an investigation. When multiple potential users are involved, as it may be the case with judicially authorized data intercepts of the Internet traffic (e.g., from a wireless access point), profiling can assist in subject disambiguation. Ultimately, a successful profile will provide immediate value to the investigators in case planning, subject identification, lead generation, obtaining and executing warrants, and prosecuting offenders (Steel, 2014).

The space of habits domain (adopted from (Wood & Rünger, 2016) is depicted in Fig. 21**.** *Habits domain* and has a goal system with two main processes: exposure and activation, which include three ways in which habits interface with goals to guide the behaviour (Wood & Rünger, 2016).

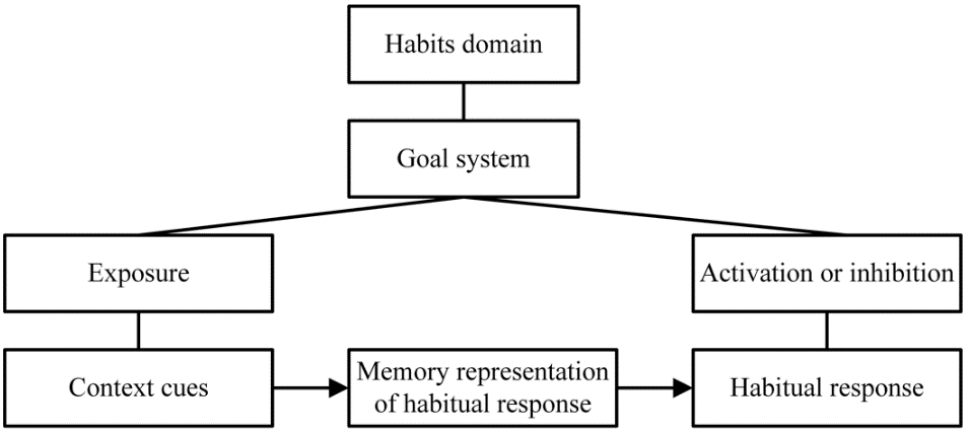

Fig. 21. Habits domain

Starting from exposure context cues automatically activate the habit representation in memory and then forming habitual response based on habit activation by tailoring people's behaviour to the current circumstances (Wood & Rünger, 2016).

3.1.1 Representation of habits identification domain using feature diagram

I are presenting a systematic approach to dealing with the problem of attribution, profiling and habits while using a feature diagram. By the habits identification domain (HiD), I mean the profiling technique that is based on the attributed habits. Fig. 22**.** *Feature diagram of the habits identification domain (HiD)* outlines a model of the HiD which is represented by using a feature diagram.

In general, a feature diagram is a tree-like notation or a directed acyclic graph which consists of a set of nodes, a set of directed edges, a set of edge decorations, relationships, and constraints among its features. A feature is understood as an externally visible characteristic of an item (i.e., a concept, entity, algorithm, system or domain). The root represents the top-level feature. The intermediate nodes represent compound features, and leaves represent the atomic features that are non-decomposable to smaller ones in a given context. The edges are used to progressively decompose a compound feature into more detailed features. Edges of the graph also denote relationships or dependencies between features. One can learn more about the notation from (Schobbens, Heymans, & Trigaux, 2006).

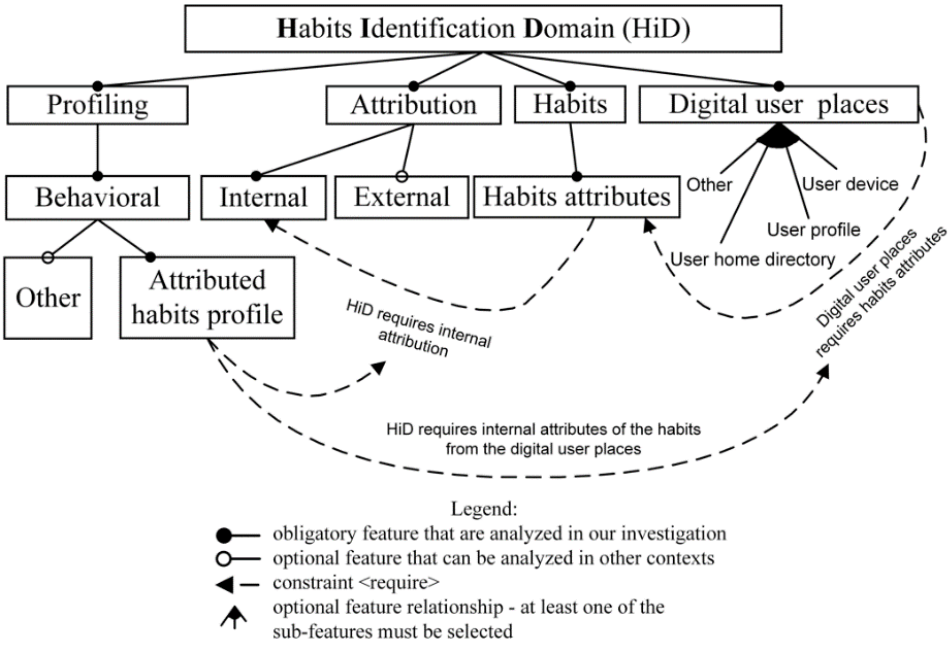

Fig. 22. Feature diagram of the habits identification domain (HiD)

In Fig. 22**.** *Feature diagram of the habits identification domain (HiD)* the mandatory features express the common aspects of the concept, whereas the optional and alternative features express variability. All the basic features may appear either as a solitary feature or in groups. If all the mandatory features in the group are derived from the same parent in the parent-child relationship, we can speak about a relationship among those features. An optional feature is the one which may be included or not if its parent is included in the feature model (Štuikys & Damaševičius, 2013).

The presented model focuses on the specialization of the habits, their attributes and profiles. I permit that for each user habit there is a set of specialized habit attributes. A set of specialized habit attributes is digitally identified from the digital user places (see Fig. 22. *Feature diagram of the habits identification domain (HiD)*). Every habit from the set of habits is attributed with internal attributes and then saved in the profile that is further used in the evidence investigation process.

3.1.2 Model for digital evidence investigation using habits attribution method

HiD is based on a model which includes mining, comparison and recognition of digital profiles of a user's digital places. Identification is done through the comparison of a digital first profile taken from a Computer certainty attached to a known subject, and the profiles are extracted from other digital devices, with which those crimes were committed but cannot be attributed with certainty to the subject. It should be noted that the principle upon which the method is based is two-way, that is, it can also start from the user's digital profile 'anonymous' of the device, for comparison with profiles of other devices (also, not involved in the offense)

attributed with certainty to particular subjects (Schultz & Shumway, 2002). The creation of HiD starts from the study of information characterizing the detected areas, such as computer users, their home directories, files and folders, nicknames, etc. on a computer, or digital devices.

In general, the investigator uses a set of search rules during the examination of a user's digital place. As an example, when digital remnants are pictures investigated as evidence, the investigator uses search rules as follows: *.jpg or *.gif or *.png, etc.

Further, I formalize the investigation task in the way adopted from (Štuikys *et al*., 2015). Let us have four sets: evidence (*E*) – a set of digital evidence, search (*S*) – a set of the search rules, profile (*P*) – a set of the hardware profile attributes, digital user places (*D*) – a set of user devices, files (folders), home directory, etc.

Each device has its own specific classification of the areas containing features, according to its specific characteristics and the installed applications. Here, I find a generic classification of the basic areas related to a Personal Computer. The number of the research areas of the feature is flexible because it depends on the target of the research and the applications available on the device. The starting point is the log files (D) providing all the information (P) about the user's machine configuration. The second step is the analysis of the files stored in folders created for any user by the operating system. In fact, they contain the most 'personalized' files made by the user. The creation of the user's folder profile is not sufficient to delineate the entire profile of the user's machine, since other features can be detected from the files stored in areas not included in the generic user folders. The Profile (on an *apple* device) includes those files contained, for example, into directories on other partitions, additional hard disks, including also deallocated files, etc. The profiles extrapolated so far (see Fig. 23**.** *HiD based process of search in a multi-user digital place flow chart*) consist of all the elements necessary for creating the user habits profile called HiD. It is the digital behavioural model that describes the user's interaction with the digital device under analysis. It is therefore composed of:

• all the characterizing information that is recognized on the entire machine during the analysis.

• all the files that contain it.

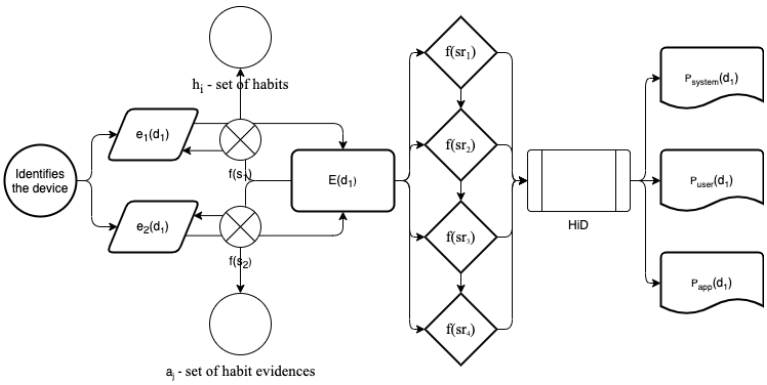

Fig. 23. HiD based process of search in a multi-user digital place flow chart

Every set has its own variants. Regarding the introduced sets, I propose the sequence of digital evidence investigation in general as defined by Equation (5).

$$\begin{Bmatrix} e_1 \\ \dots \\ e_i \\ \dots \\ e_n \end{Bmatrix} \rightarrow \begin{Bmatrix} s_1 \\ \dots \\ s_i \\ \dots \\ s_m \end{Bmatrix} \rightarrow \begin{Bmatrix} p_1 \\ \dots \\ p_i \\ \dots \\ p_k \end{Bmatrix} \rightarrow \begin{Bmatrix} d_1 \\ \dots \\ d_i \\ \dots \\ d_l \end{Bmatrix}. \quad (5)$$

When evaluating Equation (5), I will have the total sequence number $|E|\times|S|\times|P|\times|D|$. Sequences have constraints, such as investigator's evidence preference, or the search rule which might be influential in selecting the variants from *D*. An example of the investigator's evidence preference may be search for some types of files or folders.

For a HiD-based method, I introduce a set of habits $H = \{h_i\},\ i = [1, m]$ of the same HiD, a set of habit evidence $A = \{a_j\},\ j = [1, q]$, and a search rule construction function (6).

$$f(SR) = H \cup A\ . \quad (6)$$

The function $f(SR)$ construct search rules are based on the following assumptions:

The computer user is a human being tending to customize all the environments with which she or he interacts based on her or his habits.

Evidence may be applied to the habit for habit identification.

Attributed habits profile may be used to construct search rules.

Next, I replace the set of the search rules $\{s_1, \dots, s_i, \dots, s_m\}$ in Equation (5) with the set of Function (6) calculated values. Equation (5) now may be rewritten as (7).

$$\begin{Bmatrix} e_1 \\ \dots \\ e_i \\ \dots \\ e_n \end{Bmatrix} \rightarrow \begin{Bmatrix} f(sr_1) \\ \dots \\ f(sr_i) \\ \dots \\ f(sr_m) \end{Bmatrix} \rightarrow \begin{Bmatrix} p_1 \\ \dots \\ p_i \\ \dots \\ p_k \end{Bmatrix} \rightarrow \begin{Bmatrix} d_1 \\ \dots \\ d_i \\ \dots \\ d_l \end{Bmatrix}. \quad (7)$$

When evaluating Equation (7), I will have sequences $|E|\times|F(SR)|\times|P|\times|D|$. Sequences have constraints, such as investigator evidence preference, or a crime habit, which might be influential in selecting variants from *D*. As an example of a suspect's habit, we may use the naming manner of files and folders. For further formalization, I shall use the proposed equation (7).

Due to the above mentioned constraints, the number of sequences is less than evaluated in Equation (7). I assume that those constraints can be expressed by using

the constraint operators: r*equires*, *requires_any_of*, *excludes* and *subset*.

Let us have variants of evidence $E\{e_1, e_2\}$ for user hardware, variants of search rule construction function $F(SR)\{sr_1, sr_2, sr_3, sr_4\}$, variants of profiles $P\{p_1, p_2\}$, and variants of digital user places $D\{d_1, \dots, d_i, \dots d_l\}$. I assume that the investigator shall search for two pieces of evidence, I have four search rule values calculated by (6), two profiles will be investigated (two devices seized for investigation), and a huge number of files and folders may be found on the seized devices for evidence.

In our case, I have a number of variants *Nv* evaluated in (8) as sixteen sequences, each of which shall contain subsets of *D*:

$$N_v = |E| \times |F(SR)| \times |P|. \quad (8)$$

For further research, I assume that the investigator searches for evidence $\{e_1, e_2\}$ (9–12) when evidence $\{e_1\}$ is based on $f(sr_2)$ and $f(sr_3)$ search rule construction function values (10). Evidence $\{e_2\}$ is based on $f(sr_1), f(sr_2), f(sr_3)$ search rule construction function values (12). The selected search rule construction function $f(sr_1)$ requires $\{p_2\}$ profile (15) and search rule construction functions $f(sr_2), f(sr_3)$ requires any of $\{p_1, p_2\}$ profiles (16). According to the selected profiles, files and folders shall be investigated as evidence (13, 14).

$$\{e_1\}\ excludes\ \{f(sr_1), f(sr_4)\}; \quad (9)$$
$$\{e_1\}\ requires\_any\_of\ \{f(sr_2), f(sr_3)\}; \quad (10)$$
$$\{e_2\}\ excludes\ \{f(sr_4)\}; \quad (11)$$
$$\{e_2\}\ requires\_any\_of\ \{f(sr_1), f(sr_2), f(sr_3)\}; \quad (12)$$
$$\{p_1\}\ requires\ \{d_1, \dots, d_i, \dots d_l\}; \quad (13)$$
$$\{p_2\}\ requires\ \{d_1, \dots, d_i, \dots d_l\}; \quad (14)$$
$$\{f(sr_1)\}\ requires\ \{p_2\}; \quad (15)$$
$$\{f(sr_2), f(sr_3)\}\ requires\_any\_of\ \{p_1, p_2\}. \quad (16)$$

Let $SD = subset\{d_1, \dots, d_i, \dots d_l\}$ and we shall write evidence investigation sequences as follows:

$$e_1 \rightarrow f(sr_2) \rightarrow p_1 \rightarrow SD;\ e_1 \rightarrow f(sr_2) \rightarrow p_2 \rightarrow SD; \quad (17)$$
$$e_1 \rightarrow f(sr_3) \rightarrow p_1 \rightarrow SD;\ e_1 \rightarrow f(sr_3) \rightarrow p_2 \rightarrow SD; \quad (18)$$
$$e_2 \rightarrow f(sr_1) \rightarrow p_2 \rightarrow SD; \quad (19)$$
$$e_2 \rightarrow f(sr_2) \rightarrow p_1 \rightarrow SD;\ e_2 \rightarrow f(sr_2) \rightarrow p_2 \rightarrow SD; \quad (20)$$
$$e_2 \rightarrow f(sr_3) \rightarrow p_1 \rightarrow SD;\ e_2 \rightarrow f(sr_3) \rightarrow p_2 \rightarrow SD. \quad (21)$$

During the research, technical analysis of objects, a computer Macbook Air and a mobile phone iPhone, software and information stored in it, logical link analysis of files, and information grouping were performed. The investigation itself is conducted more than a year later after the incident (in late 2018). A search for the relevant information is performed which can describe the actions of the equipment

owner. This is a technical dismantling of the iPhone (technically damaged), with the aim of gaining access to the information storage device (flash).

The iPhone (black, operating system version 10.3.1) has been inspected and connected to a power source until it is fully charged. The phone was not protected by encryption or pin means. The phone uses cloud sync (iCloud) with access to <name> @ gmail.com. Software has been activated on the phone (23 applications in total): WhatsApp (there is a saved reminder to contact Person1 12/11/17 23:53 to 13/11/17 00:53 – one hour, Gmail (email <name> @ gmail.com, contacts sync, calendar – no bookmarks), dropbox, icloud drive, apple wallet (never used or synced). It is important to note that this device did not sync email and photos to iCloud or Apple. The positioning service was also disconnected on the device, which is why not all the photos on the device can provide location information. One number is blocked by the phone owner from being able to contact him. The last email was aimed to detect Facebook activity (assuming that it could have been a police investigator). The total saved on the backup media: Video 327, Photo 1796 files. Macbook Air mobile device copy: Library, Application Support, Backup: <ID>. The Macbook Air laptop, operating system version 10.11.6, was inspected and connected to a power source until it was fully charged. The device was submitted without user (Admin) access. The change of the password allowed to start the analysis of the user's behaviour, the actions after the date of 07.11.2017 were analysed separately (catalog After2017-11-07). It was noted that the device was last active on 2017 December 5. I assume this according to the logical sequence of actions (by using the digital evidence object (DEO) method) that it was possibly a technician, a police investigator. This is also shown by the collected IP addresses. The Facebook profile export was detected in the trash. I believe that reading and further interpreting this information may be helpful in the case of compiling and interpreting the USER's personality profile. A lot of personal information, lifestyle, leisure, sexual priorities are revealed (correspondence with friends). A relevant, compiled file, 'Nickname', was found, which clearly shows the problems of 'sobriety' (possibly drug use, by the way, this tendency is also revealed by Facebook (profile found in the trash) with other friends) and access to the room.

I note that this device contains a copy of the mobile device (provided), only the content is from 2017 (the beginning). Email accounts (specified by the owner), (facebook profile), (Iphone icloud) are detected. I believe that, wherever possible, the contents of mailboxes should continue to be analyzed in order to reveal contacts and exchanges of information with individuals, and what may be relevant in the case: the login history. In the study information tag document, I provided information about the login history to Facebook by making the IP addresses and connections of different devices visible. According to the records, it was connected from Sweden, later Lithuania to upgrade the Macbook Air and the mobile device iPhone, but, according to the login header, it can be seen that it is not the device that dominates. Attention is also drawn to the fact that the user's behaviour indicates his desire to be as anonymous as possible. Many VPN (virtual private network) profiles are used. The logbook records show that this is not a remote workplace, it is a global vpn network. The user does not have a Macbook air connection to the cloud

(icloud), and therefore does not use the data synchronization option. A dropbox is installed on the computer, which contains photos. The Photos are not relevant as the dates are 2017 May, June. The user very frequently used the services of wetransfer software to transfer larger files. In most cases, these are personal photo sessions. In order to form a user profile about his priorities, the history of web browsing was examined. The history did not in itself reveal or show information that could be examined in the future. The social profile (social- <user> .zip) in the Trash directory will provide opportunities and more data to see contacts, correspondence until the date of 05/12/2017. It can be unzipped to be read in the browser by selecting and opening the file.

With regard to the information obtained in the case presented here, it shows no qualitative assessment (however, this is the only investigator's responsibility, in this case), as the specific research described by the example given here was aimed solely to collecting coincident (i.e., in possession of only two values: match/no match), which could bring with certainty the identity of the same subject in question.

By using our proposed method when the search rule construction function evaluates sets of habits and their evidence, it is possible to reduce the number of subset *D* sequences (as an example of our case, from sixteen to nine).

3.1.3 Case study of the proposed model for digital evidence investigation based on the habits attribution method

We shall use this method to prove two user profiles: first, profile p_1 coincides with all files on the user's hard disk drive, and next, profile p_2 coincides with all the files on the user's hard disk snapshot. The attributed habits evaluated by the proposed function (6) have four values as follows:

$f(sr_1)$ – is a set of excluded files and folders with the attribute identified that they belong to the computer operating system;

$f(sr_2)$ – is a nickname using habit with the attribute 'FirstLast' nickname that is used in every digital place (the nickname is composed from the four initial letters of the user's first name and from the four first letters of the user's last name);

$f(sr_3)$ – is a file name setting habit with the attribute 'V' in the file name (evaluated because the user has a habit of inserting the character 'V' in the file name for versioning);

$f(sr_4)$ – is the user's login habit with the login name attributes – 'First Name', 'Last Name', 'email address'.

The case study sample was investigated for two pieces of evidences: first, evidence e_1 collected from the files on the user's hard disk drive, and next, evidence e_2 collected from the files on the user's hard disk snapshot. It is carried out in the statistical way by calculating the percentage of coincident search results trade-offs founded by comparing the general investigation case with the results when search rules construction functions $f(sr_1)$, $f(sr_2)$ and $f(sr_3)$ are used.

The case study sample is tested in two tests:

Test 1. Evidence e_1 is collected from the files on the user's hard disk drive. Search with coincident values $f(sr_2)$ and $f(sr_3)$ – 60 files are selected which have evidence objects;

Test 2. Evidence e_2 is collected from the files on the user's hard disk snapshot. Search with coincident values $f(sr_1)$, $f(sr_2)$ and $f(sr_3)$ – 30 files are selected which have evidence objects.

Quantitative assessment of the test results is presented in Table 6. Test results of digital evidence investigation using habit evidence.

Table 6. Test results of digital evidence investigation using habit evidence

	Evidence e_1 collected from the files on the user's hard disk drive		**Evidence e_2 collected from the files on the user's hard disk snapshot**	
	Number of files	**Trade-offs**	**Number of files**	**Trade-offs**
Before test	15429	0%	15097	2.2%
Excluded from investigation after $\boldsymbol{f(sr_1)}$ applied	Not applied		332	
Further investigation after $\boldsymbol{f(sr_1)}$ applied	15429		14765	
Files that has evidence objects after general investigation	125	48%	250	12%
Files that has evidence objects after $\boldsymbol{f(sr_2)}$, $\boldsymbol{f(sr_3)}$ being applied	60		30	

After the values $f(sr_2)$ and $f(sr_3)$ were applied and used as searching rules for evidence e_1 collected from the files on the user's hard disk drive, test results indicate 48% trade-offs, and after the values $f(sr_1)$, $f(sr_2)$ and $f(sr_3)$ were applied and used as searching rules evidence e_2 collected from the files on the user's hard disk snapshot, the test results indicates 2.2% and 12% trade-offs (a trade-off is obtained by comparing the ratio of the objects detected in the model to the actual representative sample), accordingly.

In this chapter, I have analyzed attribution, profiling and habits domains. I have presented a systematic approach to dealing with the problem of analyzed domains by using the feature diagram model. The proposed habits identification domain (HiD) model deals with the profiling method that is based on the attributed habits and focuses on the specialization of the habits, their attributes, and profiles.

The method based on our proposed HiD model decreases the number of the evidence investigation search sequences from the set of digital user places. It analyzes data and metadata memorized in a digital device by applying specific methods taken from intelligence and traditional profiling in order to obtain information that helps to create a digital profile with suspect user habits attributes and then considers it during evidence investigation.

The profile creation of the attributed habits starts from the research and analysis of all the 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 uses the proposed model that is based on the habit's

attribution method; it can also be applied to websites or social networks.

3.2 Conclusions of the Third Chapter

1. The analysis of attribution, profiling and habits domains disclosed that forensic investigation in order to find crime digital evidence in the digital user places (device, profile, home directory, etc.) is much broader in scope than in other areas of forensic analysis. Metadata and other logs can be used in order to attribute actions for personality identification. The proposed systematic approach deals with the problem of the analyzed domains by using the feature diagram model. The proposed habits identification domain (HiD) model deals with the profiling method that is based on the attributed habits and focuses on the specialization of the habits, their attributes, and profiles.
2. The method based on the proposed HiD model is distinguished from the currently existing solutions by its integration of the specific methods adopted from intelligence and traditional profiling in order to obtain information that helps to create a digital profile with the suspect user's habits attributes and then consider it during evidence investigation.
3. The profile creation of the attributed habits starts from the research and analysis of all the 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.
4. The model described in this chapter is suitable to such digital devices as: personal computers, tablets, smartphones, etc. The case study of the proposed model for digital evidence investigation based on habits attribution method test results indicates 48% trade-offs when comparing files on the user's hard disk drive that contain evidence objects after general investigation with the files that contain evidence objects after the proposed method's functions have been applied, and 12% trade-offs when evidence has been collected from the files on the user's hard disk snapshot.

4 PROPOSED DIGITAL EVIDENCE OBJECT MODEL

In cybercrime investigation, the theoretical methodology and practical tools have become two essential technologies. Theoretical methodologies define the models to investigate the cybercrime, and practical tools crime. The known models integrate the knowledge of experts from the fields of digital forensics and software development, use event reconstruction, automatic knowledge extraction, and preservation of data integrity. The aim of our research is to propose a digital evidence object model that combines the crime investigation process with the object-oriented programming model informatively. I propose a novel Digital Evidence Object (DEO) Model that is defined as $DEO = (W_{hy}, W_{hen}, W_{here}, W_{hat}, W_{ho})$. The

proposed model provides a methodology for digital investigation by minimizing investigation cost and time. The DEO model can be directly mapped into a tool applicable to investigate various types of cybercrime.

The described research and its results described in this Chapter were published as (Grigaliūnas & Toldinas, 2017) and (Grigaliūnas & Toldinas, 2020).

4.1 Theoretical background of the Digital Evidence Object model

Here I propose the Digital Evidence Object (DEO) model that is based on the analysis of information extracted by due forensic process using the elements of the category theory (Delvenne, 2019) with respect to the 5Ws (Why, When, Where, What, and Who) (Miranda Lopez, Moon, & Park, 2016) while focusing on forensic investigation cases proposed in (Quick & Choo, 2014). I use the category theory for our proposed DEO model because it is well-established in the 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 formally can be applied to describe digital objects as well.

The goal of the proposed DEO model is to formalize the examination phase of the digital forensic investigation process, to reduce the amount of data from the computer system or the digital device for examination, and to accelerate digital evidence acquisition. The model follows the guide of the U.S. Department of Justice (National Institute of Justice Technical Working Group for Electronic Crime Scene Investigation. Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition. USA, 2008), 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 (we document the content and the state of the evidence in its totality), and data reduction. The data reduction part of the examination phase is critical due to the 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). 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 Fig. 24. *Schematic representation of a category with objects X, Y, Z.* The diagram (Fig. 25**.** *Functions* $\boldsymbol{f : X \to Y}$ *and* $\boldsymbol{g : Y \to Z}$ *compose to a function* $\boldsymbol{g \circ f : X \to Z}$.) 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 Fig. 24. *Schematic representation of a category with objects X, Y, Z.* and in Fig. 25. *Functions* $\boldsymbol{f : X \to Y}$ *and* $\boldsymbol{g : Y \to Z}$ *compose to a function* $\boldsymbol{g \circ f : X \to Z}$.

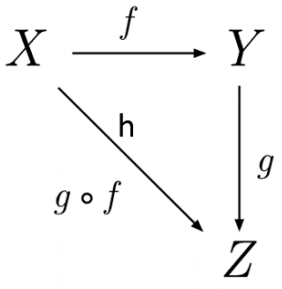

Fig. 24. Schematic representation of a category with objects X, Y, Z.

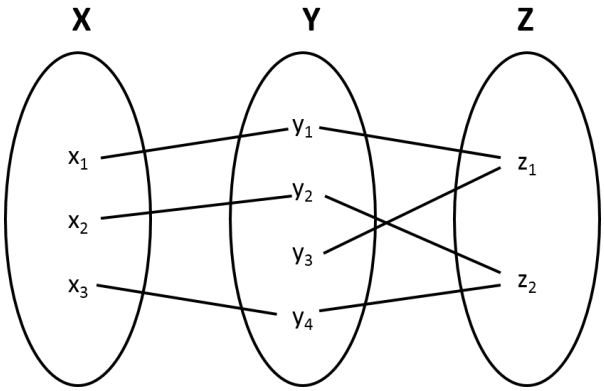

Fig. 25. Functions $\boldsymbol{f : X \rightarrow Y}$ and $\boldsymbol{g : Y \rightarrow Z}$ compose to a function $\boldsymbol{g \circ f : X \rightarrow Z}$.

When a physical crime is being considered, the investigators analyze objects at the scene to determine evidence objects. The goal of an investigation is to learn as much as possible about the history of the objects at a crime scene. The digital evidence history includes the states and events that have occurred in computer systems. The history includes all users made activities including applications, the operating system and other processes. In other words, the investigator must find answers to the 5W's questions: Why, When, Where, What, and, finally, Who has committed the criminal activity (Damaševičius *et al*., 2019). The investigator must make assumptions about the previous relations between digital objects and events based on final, and, possibly, intermediate, states of the computer system.

The DEO (22) model is formally defined by a tuple with five variables:

$$DEO = \left(W_{hy}, W_{hen}, W_{here}, W_{hat}, W_{ho}\right). \quad (22)$$

I define W_{hy} by a set of five variables:

$$W_{hy} = \{CD, IE, FI, CPV, CA\}; \quad (23)$$

where *CD* – *C*riminal *D*amage; *IE* – *I*ndustrial *E*spionage; *FI* – *F*inancial *I*nvestigations; *CPV* – *C*orporate *P*olicy *V*iolation; *CA* – *C*hild *A*buse.

W_{hen} is defined by a set of three variables:

$$W_{hen} = \{BT_{inv}, ET_{inv}, \Delta T_{inv}\}; \quad (24)$$

where BT_{inv} – *B*egin *T*ime indicates the start of the investigation period at that time; ET_{inv} – *E*nd *T*ime indicates the end of investigation period at that time; ΔT_{inv} – *T*ime duration between the consecutive time values of the investigation period.

W_{here} is defined by a set of two variables:

$$W_{here} = \{S, P\}; \quad (25)$$

where *S* – *S*ource for investigation; *P* – *P*lace for investigation.

W_{hat} is defined by a set of three variables:

$$W_{hat} = \{BT_{ev}, ET_{ev}, \Delta T_{ev}\}; \quad (26)$$

where: BT_{ev} – *B*egin *T*ime indicates the start of the investigated event at that time; ET_{ev} – *E*nd *T*ime indicates the end of the investigated event at that time; ΔT_{ev} – *T*ime duration between the consecutive time values of the investigated event period.

W_{ho} is defined by a set of two variables:

$$W_{ho} = \{\mathrm{U}, \mathrm{E}\}; \quad (27)$$

where: *U* – *P*erson who takes part in criminal activity; *E* – *Entity* (process, file, directory, registry or system entry, etc.) that take place in criminal activity.

For the DEO model, the following assumptions are formulated:

If $f_1 : W_{hy} \rightarrow W_{hen}$ and if $g_1 : W_{hen} \rightarrow W_{ho}$ then there exists composition $h_1 = g_1 \circ f_1 : W_{hy} \rightarrow W_{ho}$

If $f_2 : W_{hen} \rightarrow W_{here}$ and if $g_2 : W_{hhere} \rightarrow W_{ho}$ then there exists composition $h_2 = g_2 \circ f_2 : W_{hen} \rightarrow W_{ho}$

If $f_3 : W_{here} \rightarrow W_{hat}$ and if $g_3 : W_{hat} \rightarrow W_{ho}$ then there exists composition $h_3 = g_3 \circ f_3 : W_{here} \rightarrow W_{ho}$

If $f_4 : W_{hat} \rightarrow W_{hy}$ and if $g_4 : W_{hy} \rightarrow W_{ho}$ then there exists composition $h_4 = g_4 \circ f_4 : W_{hat} \rightarrow W_{ho}$

Definition. To formally define the possible DEO set for the investigation, the n-ary Cartesian product is calculated as follows:

$$H_1, \cdots, H_n = \{(h_1, \cdots, h_n) \mid h_i \in H_i \; for \; every \; i \in \{i, \cdots, n\}\}. \quad (28)$$

The DEO model is summarized graphically in Fig. 26. *Digital Evidence Object (DEO) model.*

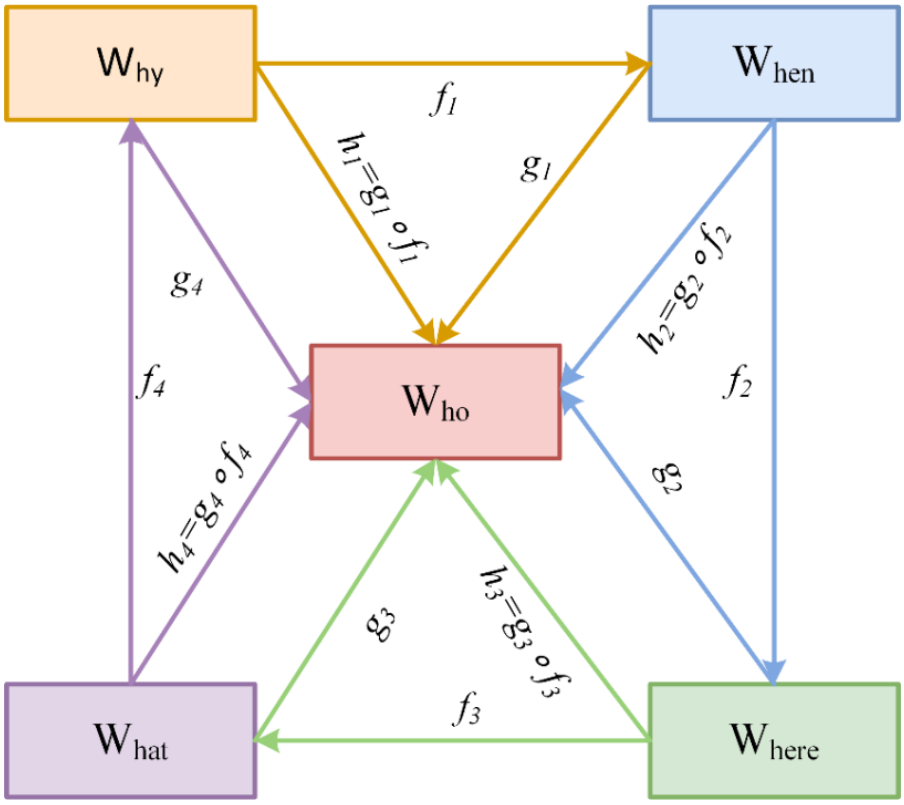

Fig. 26. Digital Evidence Object (DEO) model.

4.2 Case study and evaluation of the proposed digital evidence object model for digital evidence investigation

In our experiment, Accuracy, Precision, Recall, Miss Rate and F1-score are used as metrics to measure the performance of the proposed approach. Accuracy is used as the main performance indicator. Besides, the confusion matrix widely used in the classification model is also used in our experiment. In the confusion matrix, the true positive (TP) is the number of object records which are correctly classified as digital evidence objects, the true negative (TN) is the number of n-ary Cartesian product which is the number of irrelevant objects that were not retrieved, the false positive (FP) is the number of normal records which are incorrectly classified as digital evidence objects from operating the system layer, the false negative (FN) is the number of digital evidence objects from the application layer records which are incorrectly classified as digital evidence. For the evaluation of results, I use typical metrics used in the information retrieval and classification assessment domains (Tharwat, 2018):

Accuracy (ACC): the percentage of all the records correctly classified in the total records.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (29)$$

Precision: the percentage of the correctly identified digital evidence records in all the identified digital evidence records.

$$Precision = \frac{TP}{TP + FP} \quad (30)$$

Recall: the percentage of the correctly identified digital evidence records in all the digital evidence records. It is also called the true positive rate (TPR).

$$Recall = \frac{TP}{TP + FN} \quad (31)$$

Miss rate: it is the ratio of irrelevant objects in a set of retrieved objects. It is also called the false negative rate (FNR).

$$Miss\ rate = \frac{FN}{FN + TP} \quad (32)$$

F1-measure: it is the harmonic mean of Precision and Recall.

$$F1 = 2 \times \frac{Recall \times Precesion}{Recall + Precesion} \quad (33)$$

The NIST and ISO/lEC forensic guidelines explain the forensic investigation in five stages: (1) Identification, (2) Collection and/or Acquisition, (3) Preservation, (4) Examination and Analysis, and (5) Reporting. The Examination and Analysis process begins with a copy of the seized device and uncovers digital evidence by using the approved guidelines and the right forensic tools. Here, we demonstrate the the 4th stage of the forensic investigation – Examination and Analysis.

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 the forensic expert to uncover evidence. To achieve it, the n-ary Cartesian product as the set of all possible sets of DEOs is calculated.

Context of the case study. An information system (IS) which controls a 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 in order 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 a suspicious computer. The image mounted in the expert computer and prepared for the 4th stage of the forensic investigation –

Examination and Analysis.

Assumption for Examination and Analysis. The forensic expert knew about the cogenerated energy information system (CEIS) that is installed on the seized HDD. The main task of the installed CEIS is to control the amount of the produced cogenerated energy. For some reason (possibly fraud with the purpose of hiding the amount of the produced cogenerated energy), the system did not provide real data or did not work properly.

The hypothesis raised by the expert. Suspicious action was performed maliciously affecting CEIS, by which, journal information was modified and, maybe, the operating system's traces were modified as well. The possible time of the potential suspicious action is defined from March 21, 2016 to April 1, 2016. The expert has selected the period for his investigation from January 08, 2016 to April 08, 2016.

The tools that were used by the expert. AutoPSY 4.9 (Basis Technology, 2019b), Forensic Toolkit 5 (FTK) (AccessData, 2019), and the proposed DEO model.

In order to properly perform digital evidence analysis in the first series, we need to catalogue all of them (see Table 7. Table of the type of DEO attributes in the case study)

Table 7. Table of the type of DEO attributes in the case study

No.	Number of DEO	Type of DEO
1	1	7-Zip
2	5831	ASCII
3	9	Adobe
4	33	Alternate
5	686	Bitmap
6	2	Bookmarks
7	25	Cache
8	1	Category
9	5	Cookies
10	2	Corel
11	16	DER
12	1	Disk
13	17	Document
14	50	EFS
15	637	ESE
16	17	Empty
17	16	Excel
18	26719	Exe
19	2	File
20	3	Flash
21	20899	Folder
22	426	GIF
23	134	GZip
24	687	HTML
25	2	History
26	7198	IE
27	1	IIS
28	3720	Index
29	55	Internet
30	981	JPEG
31	728	JSON
32	35482	Java
33	70	LibreOffice
34	2	Login
35	10	MPEG
36	92	MS
37	65	MSIE
38	2	MacPaint
39	2	Macintosh
40	2	Metadata
41	82	Microsoft
42	6	Midi
43	3356	OLE
44	1	OpenOffice
45	4	PGP
46	1	PICT2
47	2997	PKCS7
48	28967	PNG
49	1	Partition
50	4357	Placeholder
51	11	RIFF
52	664	Riff
53	24	SQLITE
54	332	Scalable
55	58343	Slack
56	33	StarOffice
57	11	StarView
58	155	Summary
59	21475	Text
60	15	Thumbcache
61	10	Thumbs.db
62	3	Tiff
63	2	Top
64	601	TrueType
65	818	UTF-8
66	1542	Unallocated
67	1590	Unicode
68	10577	Unknown
69	1	Unpartitioned
70	36	WMP
71	1586	Windows
72	1	WordPerfect
73	23862	XML
74	1572	Zero
75	440	Zip
76	2	dBase

By using hypothesis that is raised by the expert, I define W_{hy1} in the sense of Eq. (21) and composition function h_1:

$$\mathrm{W_{hy1}} = \{\mathrm{CD}\} \subset \mathrm{W_{hy}}, \tag{34}$$

$$\mathrm{h_1} = \mathrm{g_1} \circ \mathrm{f_1} : \mathrm{W_{hy1}} \to \mathrm{W_{ho}}, \tag{35}$$

By using hypothesis that the possible time of the potential suspicious action is defined from March 21, 2016 to April 1, 2016, I define W_{hen1} in the sense of Eq. (36) and composition function h_2:

$$W_{hen1} = \{1452211200,1460116800,7905600\} \subset W_{hen}, \tag{36}$$

$$h_2 = g_2 \circ f_2 : W_{hen1} \to W_{ho}, \tag{37}$$

In Equation (36), Unix timestamps are calculated as follows: BT_{inv} = 2016-01-08 @ 12:00am (UTC) =1452211200; ET_{inv} = 2016-04-08 @ 12:00am (UTC) =1460116800; ΔT_{inv} = 91 days, 12 hours, 0 minutes and 0 seconds =7905600.

By using the mounted image of the hard disk from the suspicious computer, I define W_{here1} in the sense of Eq. (38) and composition function h_3:

$$W_{here1} = \{S, P\} \subset W_{here}, \tag{38}$$

$$h_3 = g_3 \circ f_3 : W_{here1} \to W_{ho}, \tag{39}$$

Since place P of the hard disk for investigation may be dependent on the application or operating system responsibility, I define P as a set of two variables:

$$P = P_{App} \cup P_{Os}, \tag{40}$$

Place P_{App} of the application domain relational calculus has the following form:

$$\{< a_1, a_2, a_3, a_4 | P_{App}(a_1 \dots a_4) >\}, \tag{41}$$

$$P_{App} = \{P_{App_1}, P_{App_2}, \dots P_{App_n}\}, \tag{42}$$

where a_1 – the name of the application object, a_2 – the path to the application object, a_3 – the timestamp in the UNIX format of the application object, a_4 – the type of the application object.

Place P_{Os} of the operating system domain relational calculus has the following form:

$$\{< o_1, o_2, o_3, o_4 | P_{Os}(o_1 \dots o_4) >\}, \tag{43}$$

$$P_{Os} = \{P_{Os_1}, P_{Os_2}, \dots P_{Os_n}\}, \tag{44}$$

where o_1 – the name of the operating system object, o_2 – the path to the operating system object, o_3 – the timestamp in the UNIX format of the operating system object, o_4 – the type of the operating system object.

The places of application responsibility objects are shown in Table 8. Places of Application responsibility.

Table 8. Places of Application responsibility

Places of Application responsibility				
P_{App}	Name	Path	Time (unix)	Type
No.	a_1	a_2	a_3	a_4
P_{App_1}	gintas	E250E233-13AA-42D9-BF81-4C5037BAE598	22 03 2016 18:49:13 (1458672553)	Remote Control
...	...	...	...	...
$P_{App_{59}}$	gintas	SAMSUNG-HDD-SP4002H.001/Partition 1/NONAME [NTFS]/[root]/statistic/2016/March/22/19.csv	22 03 2016 19:03:48 (1458673428)	Modified
...	...	...	...	...
$P_{App_{1547}}$	gintas	USBSTOR\DiskJetFlashTranscend_8GB___USBSTOR\DiskJetFlash	22 03 2016 12:18:39 (1458649119)	Physical Access
...	...	...	...	...
$P_{App_{12365}}$	gintas	SAMSUNG-HDD-SP4002H.001/Partition 1/NONAME [NTFS]/[root]/Program Files (x86)/TeamViewer/TeamViewer_Desktop.exe	22 03 2016 18:49:13 (1458672553)	TeamViewer

The places of the operating system responsibility objects are shown in Table 9. Places of Operating System responsibility.

Table 9. Places of Operating System responsibility.

Places of Operating System responsibility				
P_{Os}	Name	Path	Time (unix)	Type
No.	o_1	o_2	o_3	o_4
P_{Os_1}	gintas	E250E233-13AA-42D9-BF81-4C5037BAE598	22 03 2016 18:49:13 (1458672553)	System Service
...	...	...	...	...
$P_{Os_{14865}}$	gintas	SAMSUNG-HDD-SP4002H.001/Partition 1/NONAME [NTFS]/[root]/Program Files (x86)/LibreOffice 4/program/resource/dbaen-ZA.res	22 03 2016 19:03:48 (1458673428)	Software
...	...	...	...	...
$P_{Os_{85478}}$	admin	USBSTOR\DiskGeneric_ USBSTOR\Generic_Flash_Disk_8 Generic_Flash_Disk_8	22 03 2016 12:18:39 (1458649119)	System Service
...	...	...	...	...
$P_{Os_{215987}}$	gintas	SAMSUNG-HDD-SP4002H.001/Partition 1/NONAME [NTFS]/[root]/Program Files(x86)/TeamViewer/TeamViewer_Desktop.exe	22 03 2016 18:49:13 (1458672553)	Software

Based on the hypothesis that was raised by the expert, I define W_{hat1} in the sense of Equation (45) and composition function h_4:

$$\mathrm{W_{hat1}} = \{1458518400,1459468800,950400\} \subset \mathrm{W_{hat}}, \quad (45)$$

$$\mathrm{h_4} = \mathrm{g_4} \circ \mathrm{f_4} : \mathrm{W_{hat1}} \to \mathrm{W_{ho}}, \quad (46)$$

In Equation (45), the Unix timestamps are calculated as follows: BT_{ev} = 2016-03-21 @ 12:00am (UTC) = 1458518400; ET_{ev} = 2016-04-01 @ 12:00am (UTC) = 1459468800; ΔT_{ev} = 11 days, 0 hours, 0 minutes and 0 seconds 950400.

The possible DEO set for the investigation in the application domain is calculated as the n-ary Cartesian product in the sense of Eq. (47).

$$H_1 = \{1452211200,1460116800,7905600, S, P_{App_1}, 1458518400,1459468800,950400\}, \quad (47)$$
$$\ldots$$
$$H_i = \{1452211200,1460116800,7905600, S, P_{App_i}, 1458518400,1459468800,950400\}, \quad (48)$$
$$\ldots$$
$$H_n = \{1452211200,1460116800,7905600, S, P_{App_n}, 1458518400,1459468800,950400\}, \quad (49)$$

The possible DEO set for the investigation in the operating system domain is calculated as the n-ary Cartesian product in the sense of Eq. (50).

$$H_1 = \{1452211200,1460116800,7905600, S, P_{Os_1}, 1458518400,1459468800,950400\}, \quad (50)$$
$$\ldots$$
$$H_i = \{1452211200,1460116800,7905600, S, P_{Os_i}, 1458518400,1459468800,950400\}, \quad (51)$$
$$\ldots$$
$$H_n = \{1452211200,1460116800,7905600, S, P_{Os_n}, 1458518400,1459468800,950400\}, \quad (52)$$

By applying the DEO model, I define W_{ho1} in the sense of Equation (53):

$$W_{ho1} = \{U_1, E_1\} \subset W_{ho}, \quad (53)$$

where U_1 – *P*erson who takes part in cybercriminal activity; E_1 – *Entity* (process, file, directory, registry or system entry, etc.) that takes place in cybercriminal activity as expressed by using Eqs. (54) – (57).

$$\{< u_1, u_2, \ldots, u_n | U(u_1 \ldots u_n) >\}, \quad (54)$$
$$\{< e_1, e_2, \ldots, e_n | E(e_1 \ldots e_n) >\}, \quad (55)$$
$$\{< u_i > | \exists a_1 (a_1 \in P_{App_i} \wedge (a_2 \in P_{App_i} \wedge (\exists a_3 (< a_3 > \in P_{App_i} \wedge a_3 = W_{hen1})))) \}, \quad (56)$$
$$\{< e_i > | \exists o_1 (o_1 \in P_{Os_i} \wedge (o_2 \in P_{Os_i} \wedge (\exists o_3 (< o_3 > \in P_{Os_i} \wedge o_3 = W_{hat1})))) \}. \quad (57)$$

In the sense of Equations (54)–(57), 277 W_{ho1} DEOs were selected to analyze, which include 4 Persons (U_1) and 273 Entitles (E_1) as follows:

$$U_1 = \{< Administrator, admin, gintas, guest >\}, \quad (58)$$

$$E_1 = \{< 1 >, \dots, < 19.csv >, \dots, < \text{DiskJet Flash Transcend 8GB } (App_{1547}) >, \dots, < \text{TeamViewer\_Desktop.exe} >, \dots, < 2 >, \dots, < \text{dbaen} - \text{ZA.res} >, \dots, < \text{Generic\_Flash\_Disk\_8} >, .. < \text{TeamViewer\_Desktop.exe} >\} . \quad (59)$$

The results of the case study are summarized in Table 10. Summary of the case study. When using FTK, *268108* objects were acquired from the disk image. With AutoPSY examination, *194103* objects were selected for expert investigation. If the expert were not using the DEO model, he would have to analyse all of them. When the DEO model was applied, only 277 objects were selected for digital evidence acquisition, which resulted in significant reduction of the data required to be analyzed.

The research based on the DEO model refers to the criminal damage (CD) activities that were investigated during the period from March 21, 2016 to April 1, 2016. The expert, after applying the DEO model, selected 277 DEO's as U_1 and E_1 and, after analysis, he made the following decisions:

1. On March 22, 2016, at 12:18:39 USB DiskJet Flash Transcend 8GB (see Table 8, $P_{App_{1547}}$) drive was physically attached.
2. User gintas using TeamViewer on March 22, 2016 at 18:49:13 (see Table 8, $P_{App_{12365}}$) was remotely connected to the OPERATORS controller.
3. By using the LibbreOffice tool (see Table 9. Places of Operating System responsibility., $P_{Os_{14865}}$), user gintas made a modification of files in the CEIS statistic directory (C:\statistic\2016\March\). That directory contains catalogues named by days of the month, so, in the March directory we could find subdirectories with names: 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and 22.
4. By analysing 277 DEO's, the expert found out that the content of the subdirectories 21 and 22 was modified. The amount of the produced cogenerated energy to the appropriate csv file is written every hour with the name according to the hour of the file creation time. As an example, the file 18.csv was created on March 22, 2016 at 18:00, and the next hour csv file must be written in sequence, presuming that it is in the directory with the same name as the day but with file name 19.csv. Therefore, file 19.csv must have the creation date and time March 22, 2016 19:00. However, the expert analyzing the DEO's found out that file 19.csv (see Table 8. Places of Application responsibility) was created on March 22, 2016 at 19:03:48 and, according to the file creation time 19:03:48, the expert made an assumption that file 19.csv was improperly modified. Additionally, the expert did not find the file 20.csv which must have been created on March 22, 2016 at 20:00 and did not find files 21.csv, 22.csv, and 23.csv, either.
5. Based on the DEO's analysis, the expert created a report in which, as digital evidence, directory C:\statistic\2016\March\ was selected with all the files located in it.
6. The expert made an assumption that user **gintas** is responsible for fraud by hiding the amount of the produced cogenerated energy.

Table 10. Summary of the case study.

Entity	Acquisition	Examination	Examination		Analysis
	FTK	AutoPSY 4.9	n-ary Cartesian product		After application of DEO model
			P_{App}	P_{Os}	
Archives	11692	1885	0	0	0
Databases	11663	98	0	0	0
Documents	54420	24621	1854	10643	110
Email	3	0	0	0	0
Executable	82201	75901	687	142658	0
Graphics	42009	33045	0	0	0
Internet and Chat	7373	0	0	0	0
Multimedia	1860	1718	0	0	0
Encrypted	3067	0	0	0	0
Others	14858	0	271	1784	47
Windows Registry	26778	56835	8928	60618	24
Unknown	12184	0	625	284	83
Total:	**268108**	**194103**	**12365**	**215987**	**277**
			228352		

I evaluate the efficiency of all the compared methods (tools) by using Eq. 29. I note that the number of objects based on which a forensics expert made his conclusions (i.e., the number of true positives), is **4**. By using the number of objects discovered by FTK as the total number of objects available for analysis, the DEO model allows achieving an error rate of 0.0001 (see Table 11 Evaluation of the proposed digital evidence object model), which means that the number of false positives was significantly reduced.

Table 11 Evaluation of the proposed digital evidence object model

DEO Objects	False Positive Rate (FPR)	Accuracy (ACC)	Precision	Recall	Miss rate (FNR)	F1-measure
268108	**0,001**	**0,460**	**0,001**	**0,986**	**0,014**	**0,002**

The representative sample obtained in the experiment is **277** (see Table 11), therefore, the DEO model allows achieving an error rate of **0.014**.

I proposed a novel Digital Evidence Object (DEO) model for digital forensic investigation and described its application. The proposed DEO model is based on the principles of the category theory and is used for digital investigation analysis with respect to the 5Ws (Why, When, Where, What, and Who). The model supports situation-aware intelligent time-critical decision making and automated knowledge discovery in the domain of digital forensics. Below is a comparison of the results of DEO and other analyzed methods (see Table 12 DEO and comparison of the result of analogous methods). The closest performance to the experiment value was achieved by Case-based CBR-FT (Horsman *et al*., 2014). However, the result in terms of the accuracy of the extracted digital traces was not inferior to that of machine-learning algorithms.

Table 12 DEO and comparison of the result of analogous methods

Case-Based Reasoning Forensic Triager (CBR-FT} (Horsman et al., 2014)						
			EnCase			
	Files Retrieved	Evidenced	Recall (TPR)	Precision (FPR)	F1-measure	Percentage correct
Maximum files (195896)	95553	14	1	0.0001465	0.000293	n/a
Similar amount of evidence (262)	15597	262	1	0.016798	0.033	n/a
			CBR-FT			
	Files Retrieved	Evidenced	Recall (TPR)	Precision (FPR)	F1-measure	Percentage correct
Maximum files (195896)	195896	14	1	0.00007	0.00014	n/a
Similar amount of evidence (262)	5896	0	0	0	0	n/a
			Digital Evidence Object (DEO)			
	Files Retrieved	Evidenced	Recall (TPR)	Precision (FPR)	F1-measure	Percentage correct
Files (objects): 268108 evidence 277	268108	277	0.986	0.001	0.002	n/a
A Machine Learning-based Triage methodology (Marturana & Tacconi, 2013)						
	Files Retrieved	Evidenced	Recall (TPR)	Precision (FPR)	F1-measure	Percentage correct
BN	24.1 GB	n/a	0.99	0.99	0.99	99
DT	24.1 GB	n/a	0.9	0.88	0.89	89.5
LWL	24.1 GB	n/a	0.79	0.77	0.77	78.5
SVM	24.1 GB	n/a	0.94	0.93	0.93	93.5
iCOP: Live forensics to reveal previously unknown criminal media on P2P networks (Peersman *et al*., 2016)						
			CSA			
	Files Retrieved	Evidenced	Recall (TPR)	Precision (FPR)	F1-measure	Percentage correct
Naive Bayes	1,000,000 regular filenames	10,000 CSA filenames	5.7	62.3	10.4	n/a
Support vector machines			43.1	79.7	55.8	n/a
Logistic regression			37.6	83	51.7	n/a
			NON-CSA			
	Files Retrieved	Evidenced	Recall (TPR)	Precision (FPR)	F1-measure	Percentage correct
Naive Bayes	1,000,000 regular filenames	10,000 CSA filenames	99.9	99.2	99.6	n/a
Support vector machines			99.9	99.5	99.7	n/a
Logistic regression			99.3	99.4	99.3	n/a

In order to demonstrate the applicability of the model, I presented a real-world case study for assisting a computer forensics expert in the digital evidence

investigation process of the fraud made with the purpose of hiding the amount of produced cogenerated energy by the power plant. Our results show that the proposed DEO model can formalize the examination phase of the digital forensic investigation process, reduce the amount of data from a computer system or a digital device for examination, accelerate digital evidence acquisition, and improve the cyber situation awareness. The DEO model can help a forensic investigator, first, to reduce the amount of data for examination, next, to analyze and extract digital evidence from the reduced amount of information and a smaller data set. By examining the smaller amount of information and data from a computer system, the digital forensics expert can increase his/her time-critical performance and reduce the error rate.

4.3 Conclusions of the Fourth Chapter

1. The analysis of the currently existing methods, models and frameworks for cybercrime forensic investigation proved that there is no superior method, model or framework which could be able to cover the exponential growth amount of digital information with the main cybercrime forensics related areas. I proposed a novel Digital Evidence Object (DEO) model for digital forensic investigation which is based on the principles of the category theory and is used for digital investigation analysis with respect to the 5Ws (Why, When, Where, What, and Who).
2. The model distinguishes the existing solutions because it supports situation-aware intelligent time-critical decision making and automated knowledge discovery in the domain of digital forensics.
3. I was presented with a real-world case study for assisting a computer forensics expert in the digital evidence investigation process of a fraud made with the purpose of hiding the amount of the produced cogenerated energy by a power plant. The obtained results prove that the proposed DEO model can formalize the examination phase of the digital forensic investigation process, reduce the amount of data from the computer system or the digital device for examination, accelerate digital evidence acquisition, and improve cyber situation awareness.
4. The DEO model can help the forensic investigator, first, to reduce the amount of data for examination, next, to analyze and extract digital evidence from the reduced amount of information and a smaller data set. By examining the smaller amount of information and data from the computer system, the digital forensics expert can increase his/her time-critical performance and reduce the error rate. The obtained results shows values of Recall (TPR) 0.986, Precision (FPR) 0.001 and F1-measure 0.002.

5 EVALUATION OF THE PROPOSED MODELS AND EXPERIMENTAL RESULTS

The 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 is depicted in Figure 27.

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

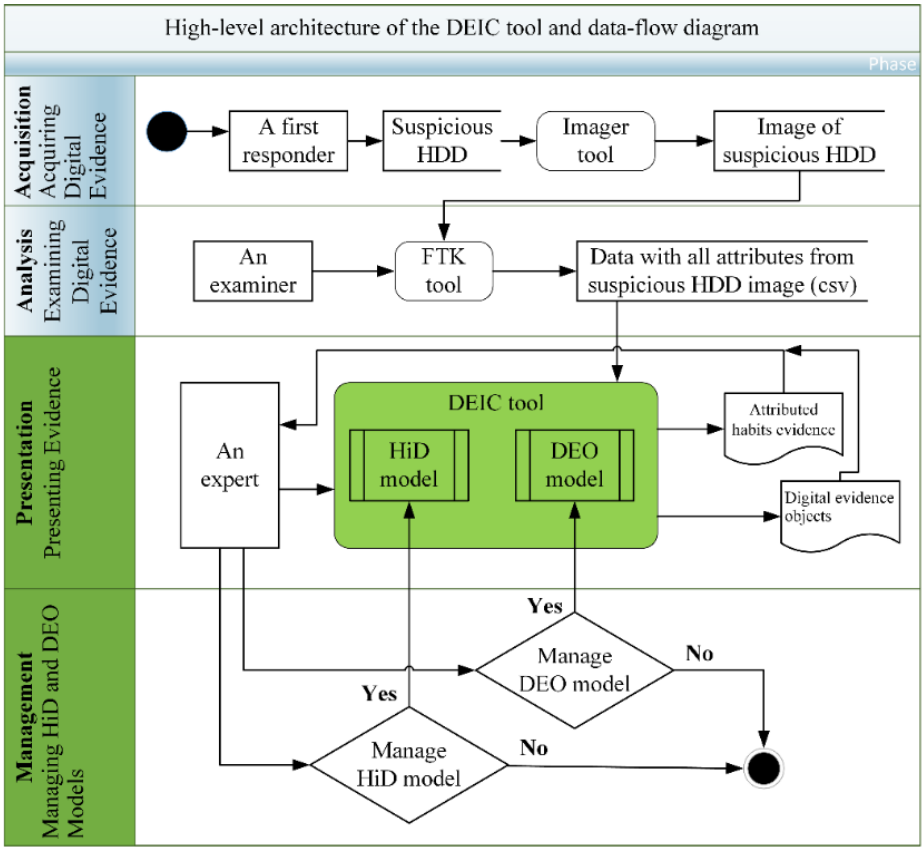

Figure 27 High-level architecture of the DEIC tool and data-flow diagram

The main purpose of the digital forensic investigation in our experiment is to provide a valid and reliable evaluation of the proposed model for digital evidence investigation by using the habits attribution profiling method (HiD) and the proposed digital evidence object model (DEO). In order to achieve the evaluation result, the Digital Evidence Investigation of Cybercrime (DEIC) tool (Grigaliūnas, 2019) (see Fig. 28) was developed to perform this task.

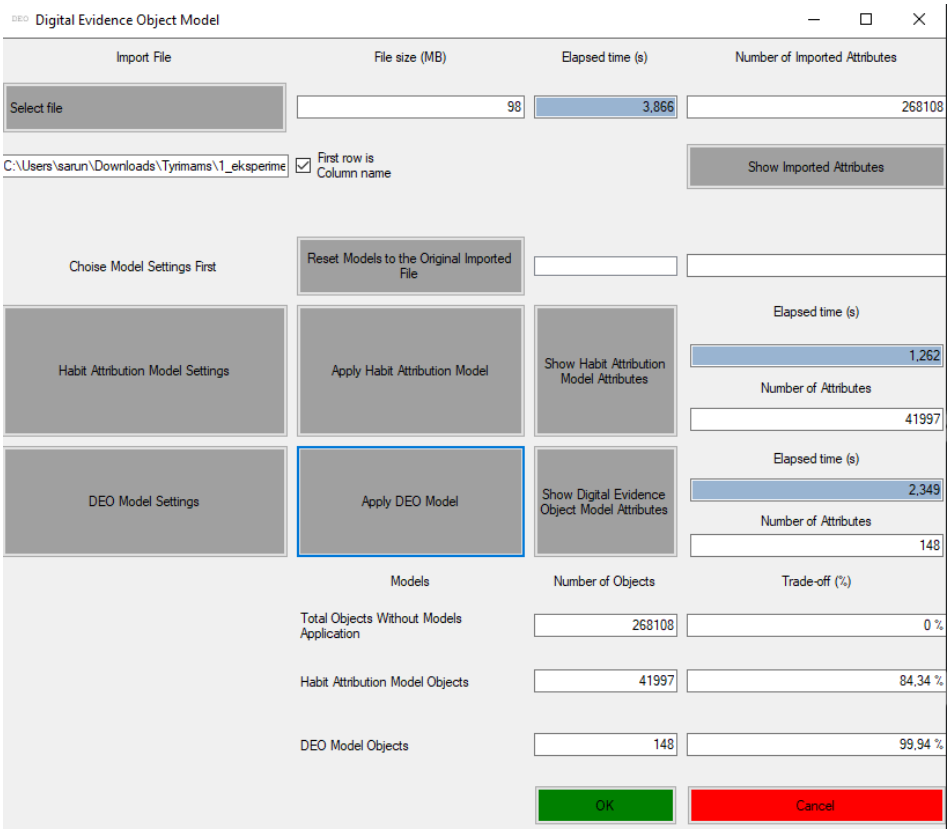

Fig. 28 DEIC tool to perform the experiment

I perform the experiment in two parts. *First part* – I use 10 (ten) different digital evidence device images in this section. The main purpose is to demonstrate the functionality of the HiD and DEO models. *Second part* – I investigate a real incident and apply our DEO model to detect digital evidence. The hypothesis is raised by the expert. Suspicious action was performed maliciously affecting CEIS, by which, journal information was modified, and, maybe, the operating system traces were modified as well. The possible time of the potential suspicious action is defined from March 21, 2016 to April 01, 2016. The expert selected the period for his investigation from January 08, 2016 to April 08, 2016.

5.1 Digital evidence investigation of cybercrime using DEIC Tool

The tools that were used by the expert are AutoPSY 4.9 (Basis Technology, 2019b), Forensic Toolkit 5 (FTK) (AccessData, 2019), and the proposed DEIC tool (Grigaliūnas, 2019) (Fig. 28 *DEIC tool to perform the experiment).*

The order of the digital evidence amount is very important in sample reduction. At the very beginning, I import the attributes (sequentially) obtained from each experiment, and then apply the proposed models. In Figure 29, shows the settings of the proposed model for digital evidence investigation by using the habits attribution profiling method (HiD).

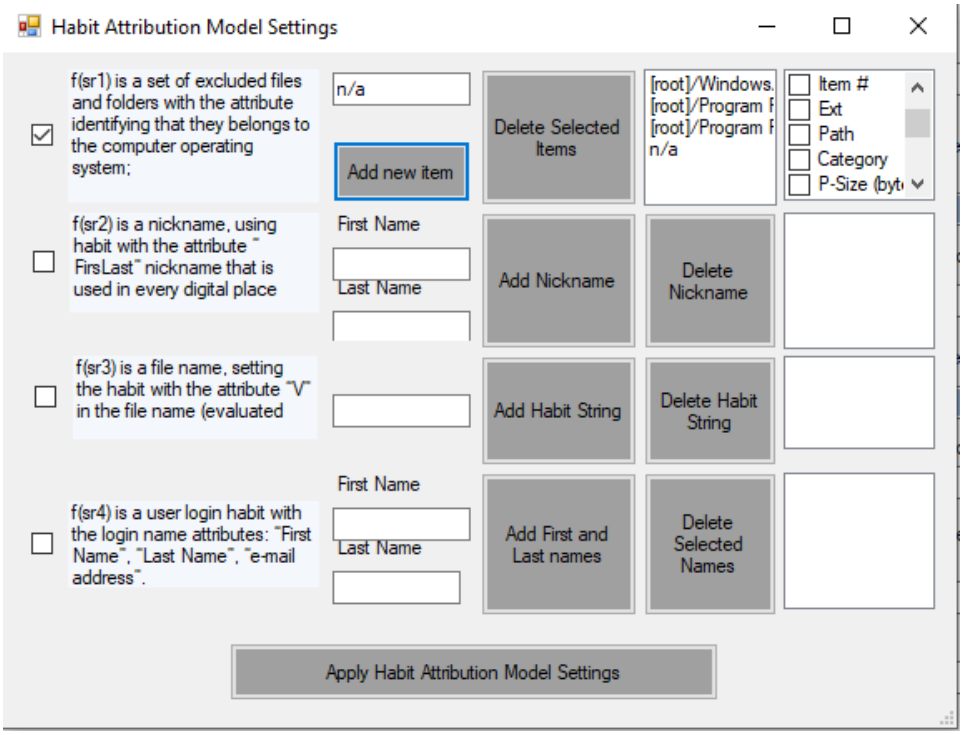

Fig. 29. Settings window of the proposed model for digital evidence investigation using the habits attribution profiling method

By applying the $f(sr_1)$ function which is a set of excluded files and folders with the attribute identified that they belong to the computer operating system, i can suggest that experts do not adjust the attributes of the operating system. In the table below we can see that I suggest using: Path – path to all evidence, CAM – Create, Access, Modified files, Users – system users, user – specific user and n/a – not applicable, for sample reduction. All the 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 the files have physical and logical sizes, often the physical size is larger than the logical size, sometimes it is equal. Yet, the logical size should never be larger than the physical

size, otherwise the file system is corrupted, or something unusual happens.

After applying the $f(sr_2)$ function which is the nickname using habit with the attribute 'FirstLast' nickname that is used in every digital evidence line. We apply the $f(sr_3)$ function which is the file name setting habit with the attribute 'V' in the file names. We apply the $f(sr_4)$ function which is a user's login habit with the login name attributes – 'First Name', 'Last Name'.

The experimental scenario as I do not know anything about digital evidence in advance (a black box) and got the following results (Table 13 Number of attributes by using the proposed model for digital evidence investigation using the habits attribution profiling method) after the experiment.

Table 13 Number of attributes by using the proposed model for digital evidence investigation using the habits attribution profiling method

Imported digital evidence image	**File size (MB)**	**Elapsed time (s)**	**Numbers of imported Attributes**	**HiD f(sr1) Path**	**HiD f(sr1) Path CAM n/a**	**HiD f(sr2) Path Users**	**HiD f(sr3) Path Users**	**HiD f(sr4) Path user**
1	98	5	268108	41997	41997	16496	16633	10181
2	89	4	245581	38902	3208	13779	13779	13779
3	93	9	257531	40633	27511	12691	15693	9726
4	196	16	536216	83994	53696	7388	33222	10164
5	160	11	804324	125991	85671	16209	49833	30477
6	297	15	804324	125991	85671	16209	49833	30477
7	93	4	255989	39091	27511	4814	15693	9726
8	490	20	1340544	209989	142734	27070	83165	50905
9	98	4	267954	41861	28422	5395	16602	10157
10	96	6	263783	40513	27635	5322	15973	15973

The obtained results (see Table 13) demonstrate how a different proposed model for digital evidence investigation using the habits attribution profiling method functions reduces the number of digital evidence. The quantity that will be reviewed later by an expert to make a decision (either deeper exploration with advanced tool settings, or it is sufficient to institute legal proceedings). When analyzing the results of DEIC after the proposed model for digital evidence investigation using the habits attribution profiling method functions where applied for the first image, I draw the following conclusions (this image is taken from a real case and used in the *Second part*): the first image has *268108* attributes; with the $f(sr_1)$ function and parameter *Path* we immediately reduce the amount of the attributes (see Table 18 Trade off after was applied HiD model was applied); not all the cases to reduce attributes apply *CAM* – Create, Access, Modified in the $f(sr_1)$ function, so the results of *Path* and *Path* plus *CAM* are the same (if we look at the table, this is certainly not the case at all); the $f(sr_2)$ function and the $f(sr_3)$ function with the parameter Users reduces the amount of attributes to *16496* (of course, we stay better for the last feature); the $f(sr_4)$ function minimizes the maximum number of attributes and reaches the quantity of *10181*. The software provides for exporting (see Fig. 30 *Habit Attribution Model export window*) the results to the csv format.

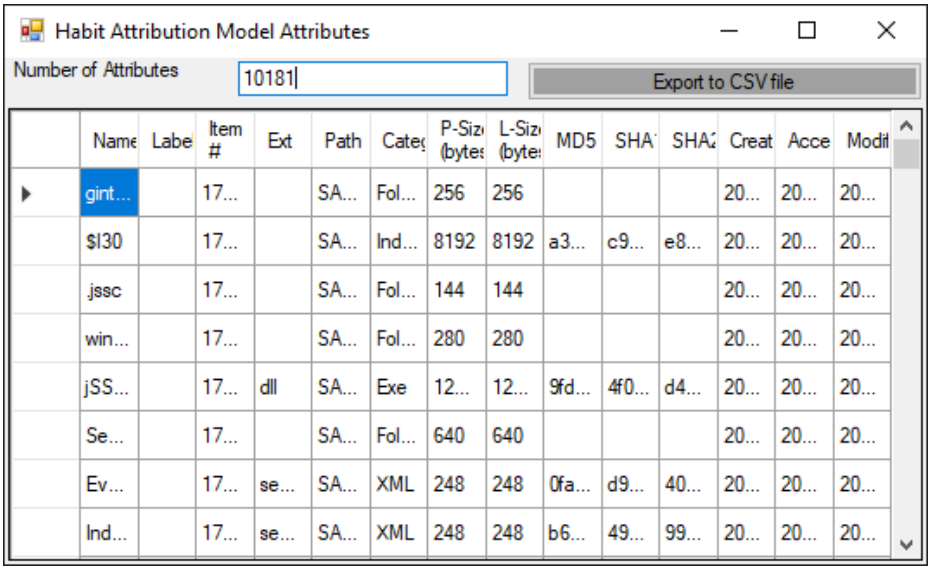

Fig. 30. Habit Attribution Model export window

This result shows how many innocent attributes (up to *96,20* percent) should be examined by experts. The large number of attributes makes the time required for experts take Years (Table 15 Lines of Expert Investigation at the Lithuanian Police Forensic Science Research Center (LPKTC) and the Lithuanian Forensic Science Center (LTEC) (LTEC. 2016.) to conduct Acquisition, Examination, Analysis of the suspect cybercrime digital evidence image. By using the COCOMO model, we can calculate (in Table 17 Examination time after applying DEO and HiD models) that the analysis of the cybercrime digital evidence shall take *9327.62* hours. Such a result is not acceptable, although it is significantly smaller compared to Table 16 Theoretical time for digital evidence acquisition. Therefore, I continue our expertise and continue to apply the DEO model.

By applying the $DEO = (W_{hy}, W_{hen}, W_{here}, W_{hat}, W_{ho})$ model which is formally defined by a tuple with five variables we look for possibility to reduce the quantity of the attributes even further. This requires proper information or information existing information to set up (see Fig. 31. *DEO Model Settings window*) DEIC tool (Grigaliūnas, 2019).

We choose the experimental scenario as we do not know anything about digital evidence in advance (the black box), and we get the following results (Table 14 Number of attributes by using DEO model) after the experiment.

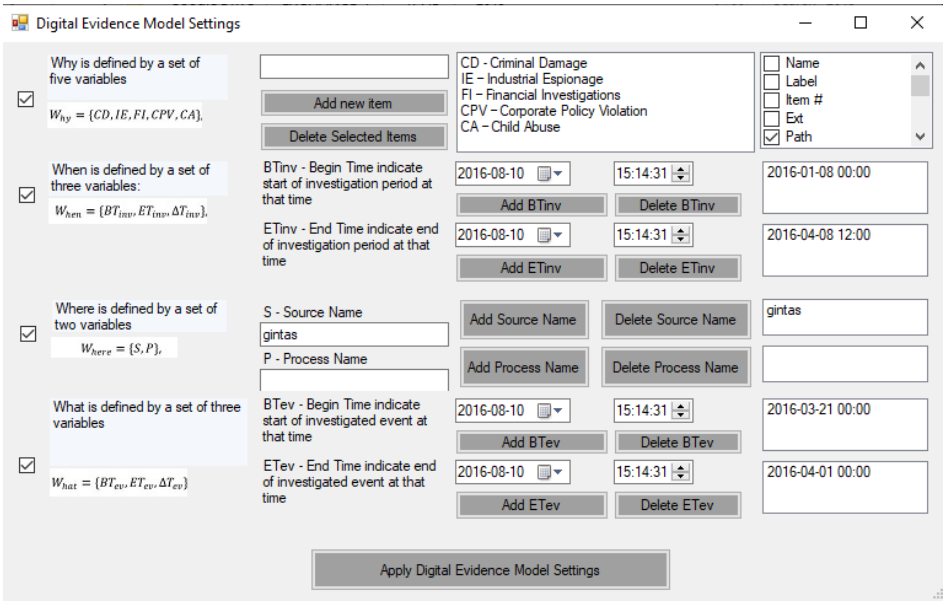

Fig. 31. DEO Model Settings window

DEO settings control is described in Formulas 23 – 27. There, we can set the W_{hy} parameter. We expand it by selecting additional parameters. All the 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 the W_{hen} and W_{hat} function which is the time interval, we indicate the investigation and event periods. This was especially helpful in the second part of the experiment – when working with a real case.

By applying W_{here} there is a possibility to refine our digital evidence search by a specific person or process.

Table 14 Number of attributes by using DEO model

Imported digital evidence image	File size (MB)	Elapsed time (s)	Numbers of imported Attributes	DEO Path (Why)	DEO Path Users (5W)	DEO Path user (5W)	DEO Ext csv (5W)	DEO Path CAM (5W)
1	98	5	268108	38658	13401	10182	19924	153
2	89	4	245581	18495	10765	7749	3903	166
3	93	9	257531	37513	12691	9727	19926	127
4	196	16	536216	68679	18121	1522	39848	332
5	160	11	804324	115974	40137	30477	59778	39
6	297	15	804324	115974	40137	30477	59772	498
7	93	4	255989	36212	12591	9726	19924	90
8	490	20	1340544	193294	67005	50905	99630	150
9	98	4	267954	30649	13371	10157	19924	121
10	96	6	263783	37199	5322	4032	19163	70

By using the DEIC tool DEO Model for the first image (Table 14 Number of attributes by using DEO model), I draw the following conclusions (this image is taken from the real case and used in the *Second part*): the first image has *268108* attributes, with DEO (W_{hy}) and parameter *Path* (Fig. 31. *DEO Model Settings window*), we immediately reduce the amount of attributes (Table 19 Trade off after applying DEO model). Since in the DEO model (Table 14 Number of attributes by using DEO model) all the Why, When, Where, What, and Who (5W) are interrelated, thus, we seek to reduce the amount of attributes in the search to find the *Users*, *user* (W_{ho}), *Ext* of the file and *CAM*. So, at the very end, with the *DEO Path CAM (5W)*, we only have *153* attributes that will need to be examined by the case investigator we make the final decision. It will take *60.54* hours (Table 17 Examination time after applying DEO and HiD models) by using the COCOMO effort calculator to make such a decision. That is how long it will take to examine the last digital evidence. It is obvious, that the digital evidence of each cybercrime contains unnecessary information. With the HiD and DEO models, this unnecessary amount can be reduced to *99.943* percent (Table 19 Trade off after applying DEO model).

5.2 Evaluation of the experiment by using COCOMO model

COCOMO II. Estimation of Effort (Sharma, 2011). COCOMO II is an objective model for planning and executing software projects. It can also be an important component of digital evidence forensics or evidence image acquisition (it is an investigation project based on the extracted lines of evidence) models depend upon the two main equations: Development, or, in digital forensics, analogous part, its 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 is an adjustable factor with the nominal value of 152 hours/PM (it is in COCOMO'81 model), but we shall apply this model in Lithuania where the average monthly management is 160/hours/PM. Embedded Effort is chosen because the digital footprint search is a very time-consuming job, and it can take different types of evidence (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 the 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 is also 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 $b = 1.2$ (Geeks for geeks, 2019).

According to the Lithuanian Forensic Science Center (LTEC) typology of presentation of the findings (see Table 15 Lines of Expert Investigation at the Lithuanian Police Forensic Science Research Center (LPKTC) and the Lithuanian Forensic Science Center (LTEC) (LTEC. 2016.)):

1. Categorical Positive Conclusion: Formulated when there is a sufficient set of attributes.
2. Probable: Missing signs formulated for categorical inference.
3. Unable to detect: All parts of the test object required for testing are missing,

and / or test objects are damaged, inoperative, LTEC does not have technical means.

Table 15 Lines of Expert Investigation at the Lithuanian Police Forensic Science Research Center (LPKTC) and the Lithuanian Forensic Science Center (LTEC) (LTEC. 2016.)

Type of expert study	Queue LPKTC (months)	Queue LTEC (months))
Information technology research	9	12

This means that it will take about a year (12 months) for LTEC to come to a conclusion in order to find digital evidence (it is not the fact that it will be found). In the case of the DEO Tool, the set of digital evidence is reduced by just a few clicks (it takes a couple of minutes) and requires no special knowledge.

I also calculated the maximum Table 16 Theoretical time for digital evidence acquisition. It would take a while to look at all the cybercrime digital evidence attributes.

Table 16 Theoretical time for digital evidence acquisition

Imported digital evidence image	File size (MB)	Numbers of imported Attributes	PM
1	98	268108	2953
2	98	268108	2953
3	93	257531	2813
4	196	536216	6784
5	160	804324	11036
6	297	804324	11036
7	93	255989	2793
8	490	1340544	20372
9	98	268080	2952
10	97	264627	2907

By using the COCOMO II model, I calculated how the time would change (Table 17 Examination time after applying DEO and HiD models) to examine the digital evidence of cybercrime .

Table 17 Examination time after applying DEO and HiD models

DEO Attributes	Hours	HiD Attributes	PM
153	60	10181	58
166	67	13779	84
127	48	9726	55
332	153	10164	58
39	12	30477	217
498	249	30477	217
90	32	9726	55
150	59	50905	402
121	46	10157	58
70	24	15973	100

With the HiD results, I can calculate efficiency (Table 18 Trade off after was applied HiD model).

Table 18 Trade off after was applied HiD model

Numbers of imported Attributes	HiD f(sr1) Path	Trade off (%)	HiD f(sr1) Path CAM n/a	Trade off (%)	HiD f(sr2) Path Users	Trade off (%)	HiD f(sr3) Path Users	Trade off (%)	HiD f(sr4) Path user	Trade off (%)
268108	41997	84.34	41997	84.34	16496	93.85	16633	93.80	10181	96.20
245581	38902	84.16	28546	88.38	13401	94.54	16633	93.23	10182	95.85
257531	40633	84.22	27511	89.32	12691	95.07	15693	93.91	9726	96.22
536216	83994	84.34	53696	89.99	7388	98.62	33222	93.80	10164	98.10
804324	125991	84.34	85671	89.35	16209	97.98	49833	93.80	30477	96.21
804324	125991	84.34	85671	89.35	16209	97.98	49833	93.80	30477	96.21
255989	39091	84.73	27511	89.25	4814	98.12	15693	93.87	9726	96.20
1340544	209989	84.34	142734	89.35	27070	97.98	83165	93.80	50905	96.20
267954	41987	84.33	28548	89.35	5395	97.99	16602	93.80	10157	96.21
263783	41351	84.32	28473	89.21	5334	97.98	15985	93.94	9659	96.34

With DEO results, I can calculate efficiency (Table 19 Trade off after applying DEO model)

Table 19 Trade off after applying DEO model

Numbers of imported Attributes	DEO Path (Why)	Trade off (%)	DEO Path Users (5W)	Trade off (%)	DEO Path user (5W)	Trade off (%)	DEO Ext csv (5W)	Trade off (%)	DEO Path CAM (5W)	Trade off (%)
268108	38658	85.58	13401	95.00	10182	96.20	19924	92.57	153	99.943
245581	38658	84.26	13401	94.54	2592	98.94	19924	91.89	153	99.938
257531	37513	85.43	12691	95.07	9727	96.22	19926	92.26	90	99.965
536216	68679	87.19	18121	96.62	1522	99.72	39848	92.57	180	99.966
804324	115974	85.58	40137	95.01	30477	96.21	59778	92.57	39	99.995
804324	115974	85.58	40137	95.01	30477	96.21	59772	92.57	270	99.966
255989	36212	85.85	12591	95.08	9726	96.20	19924	92.22	90	99.965
1340544	193294	85.58	67005	95.00	50905	96.20	99630	92.57	150	99.989
267954	30649	88.56	13371	95.01	10157	96.21	19924	92.56	90	99.966
263783	38037	85.58	12777	95.16	9659	96.34	19924	92.45	90	99.966

Finally, for evaluation of results, I use the typical metrics used in information retrieval and classification assessment domains. It is the ratio of the irrelevant objects in a set of retrieved objects (see: Table 20 HiD evaluation of results when using False Positive Rate) and (Table 21 DEO evaluation of results when using False Positive Rate).

Table 20 HiD evaluation of results when using False Positive Rate

Numbers of imported Attributes	HiD f(sr4) Path user	False Positive Rate (FPR)
268108	10181	0.0365
245581	13779	0.0532
257531	9726	0.0364
536216	10164	0.0186
804324	30477	0.0365
804324	30477	0.0366
255989	9726	0.0367
1340544	50905	0.0366
267954	10157	0.0365
263783	15973	0.0571

Table 21 DEO evaluation of results when using False Positive Rate

Numbers of imported Attributes	DEO Path CAM (5W)	False Positive Rate (FPR)
268108	153	0.0255
245581	166	0.0235
257531	127	0.0305
536216	332	0.0119
804324	39	0.0930
804324	498	0.0080
255989	90	0.0426
1340544	150	0.0260
267954	121	0.0320
263783	70	0.0541

Whether the digital evidence is added to the Case, it will be decided by the investigator (see Table 22 DEO evaluation of results when using False Negative Rate).

Table 22 DEO evaluation of results when using False Negative Rate

Numbers of imported Attributes	DEO Path CAM (5W)	False Negative Rate (FNR)
268108	153	0.0255
245581	166	0.0235
257531	127	0.0305
536216	332	0.0119
804324	39	0.0930
804324	498	0.0080
255989	90	0.0426
1340544	150	0.0260
267954	121	0.0320
263783	70	0.0541

However, it is clear that the HiD and DEO models can help to quickly understand whether such an inscription may be useful or whether we need to look into other digital evidence.

In order to achieve the evaluation result of the proposed models, the Digital Evidence Investigation of Cybercrime (DEIC) tool was developed (Grigaliūnas, 2019). I used the DEIC tool for experiments with digital evidence investigation of cybercrime from ten disk images with forensic artefacts.

I propose the use of the COCOMO II method to evaluate the time which is needed to perform analysis of artefacts obtained by the FTK and DEIC tool. I used ten different digital pieces of evidence, for which, FTK was selected by using the proposed ontology-based tool selection transformation. It is a great commercial tool that does excellent job of digital evidence acquisition and examination. Then, there are two (possibly, three) ways: to provide digital evidence to experts, to use a sophisticated commercial tool, or to use DEIC. From the experimental results presented in the tables (Table 18 Trade off after was applied HiD model and Table 19 Trade off after applying DEO model), we can clearly see the benefits of using the HiD and DEO models. Importantly, both of these models are complementary and are able to provide results (see: Table 23 Digital evidence set reduction (in times)) with digital evidence objects offered for review at the beginning of the cybercrime investigation.

Table 23 Digital evidence set reduction (in times)

HiD f(sr4) Path user	**DEO Path Users (5W)**
26	20
18	23
26	20
53	30
26	20
26	20
26	20
26	20
26	20
17	50
27	**24**

If the goal of the Case is to get the user profile of a potential offender as quickly as possible, we can achieve an average 27 times reduction of objects by using HiD (Table 23 Digital evidence set reduction (in times)). By analogy with the DEO model and when knowing nothing about the user, up to 24 times (Table 23 Digital evidence set reduction (in times)) reduction of objects can be achieved. The investigator no longer has to think about the digital evidence that is designed to keep the information system running. This type of object is included in the 5W model function and is removed from the set of sets with no value.

The DEO model has another unique model of object degradation. Depending on the information in the Case study, the number of objects can be minimized (Table 19 Trade off after applying DEO model, column: DEO Path CAM (5W)). This is what the DEO model accomplishes by degrading when it knows the preliminary date

of the cybercrime (the more accurate is the date, the better) W_{hen} and for W_{hat} period we look for. The W_{hy} feature is very useful, for example, if an investigation is linked to child pornography or copyright infringement. This is where the DEO task evolves, and, as we continue to search, we want to have all the objects that fall within our time of Case study (W_{here}). In all the cases, the model focuses on the 'who did it' (W_{ho}), and so we have a result that, as compared by the experts, includes real digital evidence objects (Table 8. Places of Application responsibility and Table 9. Places of Operating System responsibility.).

Our goal is to reduce the number of pieces of digital evidence in cybercrime investigations; it has been achieved. Experiments with models and the DEIC tool (Grigaliūnas, 2019) have shown a real opportunity to contribute to reducing cybercrime.

5.3 Conclusions of the Fifth Chapter

1. To achieve practical evaluation result of the proposed models, the Digital Evidence Investigation of Cybercrime (DEIC) tool was developed. When using the DEIC tool for experiments with digital evidence, investigation of cybercrime from ten disk images with forensic artefacts was conducted.
2. By using the proposed ontology-based transformation system for comparative studies of the proposed method, the FTK tool was selected. From the experimental results, we can clearly see the benefits of using the HiD and DEO models. Importantly, both of these models are complementary and are able to provide results with digital evidence objects offered for review at the beginning of the cybercrime investigation.
3. If an expert has some information of the suspicious user, such as the user name, nickname or other data, he/she could use the HiD for getting the user profile of the potential offender as quickly as possible, and achieve an average of 27 times reduced number of the representative objects for the first review. When an expert does not know anything about the suspicious user, the DEO model can be applied, and an average of 24 times reduced number of the representative objects for the first review is achieved.

6 GENERAL CONCLUSIONS

1. There is a huge number of available computer forensic tools from standalone packages to complex integrated tools developed for a wide range crime investigations. The analysis of methods, models and frameworks for cybercrime forensic investigation proved that there is no single superior method, model or framework, which would be able to cover the exponential growth of the amount of digital information with the main cybercrime forensics related areas. Therefore, a new, more holistic cybercrime forensic investigation method is needed to concentrate on reducing the expertise time and cost.
2. The newly proposed multi-layered architecture and ontology-based transformation system (OBTS), in which, the proposed model and XDT are realized, can serve experts who operate in terms of the forensics domain in reducing the time needed for the appropriate tool for digital evidence investigation selection from the NIST tool catalog.
3. The newly proposed method is based on:
    - Habits identification domain (HiD) model distinguished from the already existing solutions by its integration of the specific methods adopted from intelligence and traditional profiling in order to obtain information that helps to create a digital profile with the suspect user's habits attributes and then consider it during evidence investigation.
    - Digital evidence object (DEO) model is distinguished from the currently existing solutions because it supports situation-aware intelligent time-critical decision making and automated knowledge discovery in the digital forensics domain.
4. In situations when an expert has some information of the suspicious user, such as the user name, the nickname or other data, the habits identification domain (HiD) model may be used for getting the user profile of the potential offender as quickly as possible, and we achieve an average of 27 times reduced number of the representative evidence objects for the first review.
5. In situations when the expert does not know anything about the suspicious user, the digital evidence object (DEO) model may be used and achieve an average of 24 times reduced number of representative evidence objects for the first review.
6. The developed DEIC (Digital Evidence Investigation of Cybercrime) tool proves the practical applicability of the proposed method for assisting a computer forensics expert and formalizing the examination phase of the digital forensic investigation process, reducing the amount of data to investigate from a suspicious system or a digital device thus accelerating acquisition of digital evidence.

7 REFERENCES

1. 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.
2. AccessData. (2019). AccessData Group. Forensic Toolkit (FTK). Retrieved from http://accessdata.com/solutions/digital-forensics/forensic-toolkit-ftk
3. Ajijola, A., Zavarsky, P., & Ruhl, R. (2014). A review and comparative evaluation of forensics guidelines of NIST SP 800-101 Rev.1:2014 and ISO/IEC 27037:2012. *World Congress on Internet Security (WorldCIS-2014)* (pp. 66–73). Presented at the 2014 World Congress on Internet Security (WorldCIS), London, United Kingdom: IEEE. Retrieved August 22, 2019, from http://ieeexplore.ieee.org/document/7028169/
4. Akbal, E., Dogan, S., & Dogan, S. (2018). Forensics Image Acquisition Process of Digital Evidence. *International Journal of Computer Network and Information Security*, *10*(5), 1–8.
5. Alabdulsalam, S., Schaefer, K., Kechadi, T., & Le-Khac, N.-A. (2018). Internet of Things Forensics – Challenges and a Case Study. In G. Peterson & S. Shenoi (Eds.), *Advances in Digital Forensics XIV* (Vol. 532, pp. 35–48). Cham: Springer International Publishing. Retrieved October 1, 2019, from http://link.springer.com/10.1007/978-3-319-99277-8_3
6. Altheide, C., & Carvey, H. A. (2011). *Digital forensics with open source tools*. Burlington, MA: Syngress.
7. Arshad, H., Jantan, A. B., & Abiodun, O. I. (2018). Digital Forensics: Review of Issues in Scientific Validation of Digital Evidence. *Journal of Information Processing Systems*, *14*(2), 346–376.
8. ArxSys. (2019). Digital Forensics Framework (DFF). Retrieved from https://github.com/arxsys/dff
9. Ashton, K. (2009). That 'Internet of Things' Thing. *RFiD Journal*, *22(7)*, 97–114.
10. Ayers, D. (2009). A second generation computer forensic analysis system. *Digital Investigation*, *6*, S34–S42.
11. Baggili, I., BaAbdallah, A., Al-Safi, D., & Marrington, A. (2013). Research Trends in Digital Forensic Science: An Empirical Analysis of Published Research. In Marcus Rogers & K. C. Seigfried-Spellar (Eds.), *Digital Forensics and Cyber Crime* (Vol. 114, pp. 144–157). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved September 30, 2019, from http://link.springer.com/10.1007/978-3-642-39891-9_9
12. Barske, D., Stander, A., & Jordaan, J. (2010). A Digital Forensic Readiness framework for South African SME's. *2010 Information Security for South Africa* (pp. 1–6). Presented at the 2010 Information Security for South Africa (ISSA), Johannesburg, South Africa: IEEE. Retrieved August 22, 2019, from http://ieeexplore.ieee.org/document/5588281/
13. Bashir, M., & Khan, M. (2013). Triage in Live Digital Forensic Analysis. *The International Journal of Forensic Computer Science*, *8*(1), 35–44.
14. Basis Technology. (2019a). Autopsy. Retrieved from https://www.autopsy.com/
15. Basis Technology. (2019b). Sleuth Kit. Retrieved from : http://www.sleuthkit.org/sleuthkit/
16. Beebe, N. L., & Liu, L. (2014). Ranking algorithms for digital forensic string search

hits. *Digital Investigation*, *11*, S124–S132.

17. Berla. (2019). Infotainment and Vehicle System Forensics (iVe). Retrieved from https://berla.co/ecosystem/
18. Bhandari, S., & Jusas, V. (2020). An Abstraction Based Approach for Reconstruction of TimeLine in Digital Forensics. *Symmetry*, *12*(1), 104.
19. BlackBag Technologies. (2019). BlackLight. Retrieved from https://www.blackbagtech.com/
20. Boehm, B., Valerdi, R., Lane, J. A., & Brown, A. W. (2005). COCOMO suite methodology and evolution. *CrossTalk*, (4), 20–25.
21. Bottrill, M. C., Joseph, L. N., Carwardine, J., Bode, M., Cook, C., Game, E. T., Grantham, H., et al. (2008). Is conservation triage just smart decision making? *Trends in Ecology & Evolution*, *23*(12), 649–654.
22. Brain Adams, R. (2012). The Advanced Data Acquisition Model (ADAM): A Process Model for Digital Forensic Practice. Philosophy of Murdoch University.
23. Brett Pladna. (2008). Computer Forensics Procedures, Tools, and Digital Evidence Bags: What They Are and Who Should Use Them. Retrieved from http://www.infosecwriters.com/Papers/BPladna_Computer_Forensic_Procedures.pdf
24. Brian D., C. (2006, May 1). A HYPOTHESIS-BASED APPROACH TO DIGITAL FORENSIC INVESTIGATIONS.
25. Brinson, A., Robinson, A., & Rogers, M. (2006). A cyber forensics ontology: Creating a new approach to studying cyber forensics. *Digital Investigation*, *3*, 37–43.
26. Buchanan, B. (2011). Module Leader: Module number: Email: Telephone: Web page: Within ProfSIMS: MSN Messenger: Version:, 139.
27. Canter, D. (2004). Offender profiling and investigative psychology. *Journal of Investigative Psychology and Offender Profiling*, *1*(1), 1–15.
28. Cantrell, G., & Dampier, D. (2012). Implementing the Automated Phases of the Partially-Automated Digital Triage Process Model. *Journal of Digital Forensics, Security and Law*, 96–116.
29. Cantrell, G., Dampier, D., Dandass, Y. S., Niu, N., & Bogen, C. (2012). Research toward a Partially-Automated, and Crime Specific Digital Triage Process Model. *Computer and Information Science*, *5*(2), p29.
30. Carrier, B. (2003). Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers, *1*(4), 12.
31. Carrier, B. (2005). *File system forensic analysis*. Boston, Mass. ; London: Addison-Wesley.
32. Carrier, B. D., & Spafford, E. H. (n.d.). An Event-Based Digital Forensic Investigation Framework, 12.
33. Carrier, B., & Spafford, E. H. (2003). Getting Physical with the Digital Investigation Process, *2*(2), 21.
34. Casey, E. (2010). *Handbook of digital forensics and investigation*. Amsterdam ; Boston: Academic.
35. Casey, E. (2018). Clearly conveying digital forensic results. *Digital Investigation*, *24*, 1–3.
36. Casey, E., Back, G., & Barnum, S. (2015). Leveraging CybOX™ to standardize representation and exchange of digital forensic information. *Digital Investigation*, *12*,

S102–S110.

37. Chabot, Y., Bertaux, A., Nicolle, C., & Kechadi, M.-T. (2014a). A complete formalized knowledge representation model for advanced digital forensics timeline analysis. *Digital Investigation*, *11*, S95–S105.
38. Chabot, Y., Bertaux, A., Nicolle, C., & Kechadi, T. (2014b). Automatic Timeline Construction and Analysis for Computer Forensics Purposes. Unpublished. Retrieved May 17, 2020, from http://rgdoi.net/10.13140/2.1.3595.1040
39. Chen, H., Finin, T., & Joshi, A. (2003). An ontology for context-aware pervasive computing environments. *The Knowledge Engineering Review*, *18*(3), 197–207.
40. 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, from http://link.springer.com/10.1007/978-3-642-15506-2_2
41. Ćosić, J., & Bača, M. (2010). A Framework to (Im)Prove „Chain of Custody“ in Digital Investigation Process, 5.
42. Ćosić, J., & Ćosić, Z. (2012). The Necessity of Developing a Digital Evidence Ontology. Unpublished. Retrieved May 17, 2020, from http://rgdoi.net/10.13140/RG.2.1.4184.5843
43. Cosic, J., Cosic, Z., & Baca, M. (2011). An Ontological Approach to Study and Manage Digital Chain of Custody of Digital Evidence. *Journal of Information and Organizational Sciences*, *35*. Retrieved from https://pdfs.semanticscholar.org/f498/174e390ec63ac9d84948900b5a20028069e9.pdf
44. 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.
45. Cruz, F., Moser, A., & Cohen, M. (2015). A scalable file based data store for forensic analysis. *Digital Investigation*, *12*, S90–S101.
46. Dalins, J., Wilson, C., & Carman, M. (2015). Monte-Carlo Filesystem Search – A crawl strategy for digital forensics. *Digital Investigation*, *13*, 58–71.
47. Damaševičius, R., Toldinas, J., Venčkauskas, A., Grigaliūnas, Š., Morkevičius, N., & Jukavičius, V. (2019). Visual Analytics for Cyber Security Domain: State-of-the-Art and Challenges. In R. Damaševičius & G. Vasiljevienė (Eds.), *Information and Software Technologies* (Vol. 1078, pp. 256–270). Cham: Springer International Publishing. Retrieved March 27, 2020, from http://link.springer.com/10.1007/978-3-030-30275-7_20
48. Datta, S., & Pan, C. (2016). An Intelligent Forensic Framework towards Cloud: Its Ontological Aspects. *International Journal of Computer Applications*, *138*(9), 1–8.
49. DEFT Association. (2019). DEFT. Retrieved from http://www.deftlinux.net/
50. Delvenne, J.-C. (2019). Category Theory for Autonomous and Networked Dynamical Systems. *Entropy*, *21*(3), 302.
51. Douglas, J. E., Ressler, R. K., Burgess, A. W., & Hartman, C. R. (1986). Criminal profiling from crime scene analysis. *Behavioral Sciences & the Law*, *4*(4), 401–421.
52. Dzemydiene, D. (2010). Intelligent decision support systems for assistance in forensic investigation procese. *Handbook of Electronic Security and Digital Forensics*, 603–630.

53. 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, from http://link.springer.com/10.1007/978-3-319-57870-5_16
54. Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge ; New York: Cambridge University Press.
55. Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. MacGraw-Hill series in social psychology. New York, NY: MacGraw-Hill.
56. Fred Cohen. (2009). *Digital forensic evidence examination.* Place of publication not identified: Asp Press.
57. Garfinkel, S. (2012). Digital forensics XML and the DFXML toolset. *Digital Investigation*, *8*(3–4), 161–174.
58. Garfinkel, S. L. (2007). Carving contiguous and fragmented files with fast object validation. *Digital Investigation*, *4*, 2–12.
59. Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *Digital Investigation*, *7*, S64–S73.
60. Garfinkel, S., Malan, D., Dubec, K.-A., Stevens, C., & Pham, C. (2006). Advanced Forensic Format: An Open Extensible Format for Disk Imaging. In M. S. Olivier & S. Shenoi (Eds.), *Advances in Digital Forensics II* (Vol. 222, pp. 13–27). Boston, MA: Springer New York. Retrieved October 2, 2019, from http://link.springer.com/10.1007/0-387-36891-4_2
61. Gartner Says Worldwide IoT Security Spending Will Reach $1.5 Billion in 2018. (2018, March 21). . Retrieved from 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.
62. 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, from http://ieeexplore.ieee.org/document/7502341/
63. Geeks for geeks. (2019). COCOMO Model. Retrieved from https://www.geeksforgeeks.org/software-engineering-cocomo-model/
64. Giova, G. (2011). Improving Chain of Custody in Forensic Investigation of Electronic Digital Systems, *11*(1), 10.
65. Gladyshev, P. (2004). *Formalizing Event Reconstruction in Digital Investigations*. University College Dublin. Retrieved from http://formalforensics.org/publications/thesis/
66. Goranin, N., & Mažeika, D. (2011). *Nusikaltimai elektroninėje erdvėje ir jų tyrimo metodikos* (1st ed.). TEV. Retrieved March 29, 2020, from http://www.ebooks.ktu.lt/eb/239/nusikaltimai_elektronineje_erdveje_ir_ju_tyrimo_metodikos/
67. Grabosky, D. P. (2010). Computer Crime: A Criminological Overview, 21.
68. Grier, J., & Richard, G. G. (2015). Rapid forensic imaging of large disks with sifting collectors. *Digital Investigation*, *14*, S34–S44.
69. Grigaliūnas, Š. (2019). DEO Model Tool. Retrieved from https://digitalevidenceobject.com/

70. Grigaliūnas, Š., & Toldinas, J. (2017). *DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS. Digital evidence object model for cybercrime investigation*. Vilnius University. Retrieved March 28, 2020, from https://www.mii.lt/damss/index.php?page=doi_2017&lang=en

71. Grigaliūnas, Š., & Toldinas, J. (2020). Habits Attribution and Digital Evidence Object Models Based Tool for Cybercrime Investigation. *Baltic J. Modern Computing*, *8*, 275–292.

72. Grigaliunas, S., Toldinas, J., & Venckauskas, A. (2017). An Ontology-Based Transformation Model for the Digital Forensics Domain. *Elektronika ir Elektrotechnika*, *23*(3), 78–82.

73. Grosof, B. N., Horrocks, I., Volz, R., & Decker, S. (2003). Description logic programs: Combining logic programs with description logic. *Proceedings of the twelfth international conference on World Wide Web—WWW '03* (p. 48). Presented at the the twelfth international conference, Budapest, Hungary: ACM Press. Retrieved August 22, 2019, from http://portal.acm.org/citation.cfm?doid=775152.775160

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

75. Harris, R. (2006). Arriving at an anti-forensics consensus: Examining how to define and control the anti-forensics problem. *Digital Investigation*, *3*, 44–49.

76. Henseler, H., & Hyde, J. (2019). Technology assisted analysis of timeline and connections in digital forensic investigations. The 2019 edition of the International Conference on Artificial Intelligence and Law (ICAIL).

77. Hikmatyar, M., Prayudi, Y., & Riadi, I. (2017). Network Forensics Framework Development using Interactive Planning Approach. *International Journal of Computer Applications*, *161*(10), 41–48.

78. Holder, H. E., Robinson, O. L., & Rose, K. (2009). *Electronic Crime Scene Investigation: An On-the-Scene Reference for First Responders* (No. NCJ 227050). United States. Office of Justice Programs. Retrieved from https://www.hsdl.org/?view&did=30477

79. Hong, I., Yu, H., Lee, S., & Lee, K. (2013). A new triage model conforming to the needs of selective search and seizure of electronic evidence. *Digital Investigation*, *10*(2), 175–192.

80. Horsman, G., Laing, C., & Vickers, P. (2014). A case-based reasoning method for locating evidence during digital forensic device triage. *Decision Support Systems*, *61*, 69–78.

81. Hui, Y. (2012). What is a Digital Object?: What is a Digital Object? *Metaphilosophy*, *43*(4), 380–395.

82. Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2014). Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains, 14.

83. Ieong, R. S. C. (2006). FORZA – Digital forensics investigation framework that incorporate legal issues. *Digital Investigation*, *3*, 29–36.

84. Iii, G. A. F., & Clinton, K. (2005). COMPUTER FORENSICS LABORATORY AND TOOLS. *Journal of Computing Sciences in Colleges*, *Volume 20 Issue 6, June 2005*, 143–150.

85. Ingram, S. (1998). If the Profile Fits: Admitting Criminal Psychological Profiles into

Evidence in Criminal Trials, *54*, 29.

86. Internet of Things Architecture. (2013, September). IoT-A Internet of Things – Architecture. Retrieved from https://web.archive.org/web/20130918185701/http://www.iot-a.eu/public/terminology
87. Irons, A., & Lallie, H. (2014). Digital Forensics to Intelligent Forensics. *Future Internet*, *6*(3), 584–596.
88. Jahankhani, H., Watson, D. L., Me, G., & Leonhardt, F. (2010). *Handbook of Electronic Security and Digital Forensics*. WORLD SCIENTIFIC. Retrieved March 27, 2020, from http://www.worldscientific.com/worldscibooks/10.1142/7110
89. 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.
90. Jusas, V., Birvinskas, D., & Gahramanov, E. (2017). Methods and Tools of Digital Triage in Forensic Context: Survey and Future Directions. *Symmetry*, *9*(4), 49.
91. Justickis, V. (2010). *Criminal Data Mining*. WORLD SCIENTIFIC. Retrieved March 29, 2020, from http://www.worldscientific.com/worldscibooks/10.1142/7110
92. 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, from http://link.springer.com/10.1007/978-3-319-92384-0_6
93. Kävrestad, J. (2018). *Fundamentals of Digital Forensics: Theory, Methods, and Real-Life Applications*. Cham: Springer International Publishing. Retrieved August 22, 2019, from http://link.springer.com/10.1007/978-3-319-96319-8
94. Kohn, M. D., Eloff, M. M., & Eloff, J. H. P. (2013). Integrated digital forensic process model. *Computers & Security*, *38*, 103–115.
95. Koopmans, M. B., & James, J. I. (2013). Automated network triage. *Digital Investigation*, *10*(2), 129–137.
96. 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.
97. Lassila, O., & McGuinness, D. (2014). The Role of Frame-Based Representation on the Semantic Web, 11.
98. Lim, K.-S., & Lee, C. (2013). A framework for unified digital evidence management in security convergence. *Electronic Commerce Research*, *13*(3), 379–398.
99. Lim, K.-S., & Lee, S. (2008). A Methodology for Forensic Analysis of Embedded Systems. *2008 Second International Conference on Future Generation Communication and Networking* (pp. 283–286). Presented at the 2008 Second International Conference on Future Generation Communication and Networking (FGCN), Hainan, China: IEEE. Retrieved October 1, 2019, from http://ieeexplore.ieee.org/document/4734223/
100. Linuxlinks. (2019). Automated Image and Restore. Retrieved from http://www.linuxlinks.com/AutomatedImageandRestore/
101. Lohiya, R., & Shah, P. (2015). Video Based Face Detection and Tracking for Forensic Applications, *7*(1), 9.
102. LTEC. (2016, June 29). Ekspertinių tyrimų eilės Lietuvos policijos kriminalistinių tyrimų centre (LPKTC) ir Lietuvos teismo ekspertizės centre. Retrieved from

http://www.ltec.lt/index.php?id=883

103. Luthfi, A. (2014). The Use Of Ontology Framework For Automation Digital Forensics Investigation. Retrieved August 22, 2019, from https://zenodo.org/record/1091430
104. Lyle, J. R. (2010). If error rate is such a simple concept, why don't I have one for my forensic tool yet? *Digital Investigation*, *7*, S135–S139.
105. Magalingam, P., Manaf, A., Ahmad, R., & Yahya, Z. (2009). A New Digital Evidence Retrieval Model for Gambling Machine Forensic Investigation. *The International Journal of Forensic Computer Science*, 49–56.
106. Magnet Forensics. (2019). Magnet Forensics. Internet evidence finder (IEF). Retrieved from http://www.magnetforensics.com
107. Manuel Suárez-Albela, Tiago Fernández-Caramés, Paula Fraga-Lamas, & Luis Castedo. (2017). A Practical Evaluation of a High-Security Energy-Efficient Gateway for IoT Fog Computing Applications. *Sensors*, *17*(9), 1978.
108. Martinez-Romo, J., & Araujo, L. (2010). Analyzing Information Retrieval Methods to Recover Broken Web Links. In C. Gurrin, Y. He, G. Kazai, U. Kruschwitz, S. Little, T. Roelleke, S. Rüger, et al. (Eds.), *Advances in Information Retrieval* (Vol. 5993, pp. 26–37). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved August 22, 2019, from http://link.springer.com/10.1007/978-3-642-12275-0_6
109. Marturana, F., Me, G., Berte, R., & Tacconi, S. (2011). A Quantitative Approach to Triaging in Mobile Forensics. *2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications* (pp. 582–588). Presented at the 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Changsha, China: IEEE. Retrieved October 1, 2019, from http://ieeexplore.ieee.org/document/6120868/
110. Marturana, F., & Tacconi, S. (2013). A Machine Learning-based Triage methodology for automated categorization of digital media. *Digital Investigation*, *10*(2), 193–204.
111. Marturana, F., Tacconi, S., Berte, R., & Me, G. (2012). Triage-based automated analysis of evidence in court cases of copyright infringement. *2012 IEEE International Conference on Communications (ICC)* (pp. 6668–6672). Presented at the ICC 2012 - 2012 IEEE International Conference on Communications, Ottawa, ON, Canada: IEEE. Retrieved October 1, 2019, from http://ieeexplore.ieee.org/document/6364819/
112. Michael B, M., Jeffrey L, S., & David W, H. (n.d.). *Crime Scene Investigation. Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition*. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/219941.pdf
113. Microsoft. (2019). Web.config Transformation Syntax for Web Project Deployment Using Visual Studio. Retrieved from https://docs.microsoft.com/en-us/aspnet/web-forms/overview/deployment/visual-studio-web-deployment/web-config-transformations
114. Miranda Lopez, E., Moon, S., & Park, J. (2016). Scenario-Based Digital Forensics Challenges in Cloud Computing. *Symmetry*, *8*(10), 107.
115. MITRE. (2019). Cyber Observable eXpression (CybOX™). Retrieved from http://cyboxproject.github.io/about/
116. Mongay Batalla, J., & Krawiec, P. (2014). Conception of ID layer performance at the network level for Internet of Things. *Personal and Ubiquitous Computing*, *18*(2), 465–480.
117. Montasari, R. (2016). A Formal Two Stage Triage Process Model (FTSTPM) for

Digital Forensic Practice, 19.

118. Montasari, R., Carpenter, V., & Hill, R. (2019). A road map for digital forensics research: A novel approach for establishing the design science research process in digital forensics. *International Journal of Electronic Security and Digital Forensics*, *11*(2), 194.
119. MSAB. (2019). XRY. Retrieved from https://www.msab.com/xry/what-is-xry
120. Muniswamy-Reddy, K.-K., Holland, D. A., Braun, U., & Seltzer, M. I. (2006). Provenance-Aware Storage Systems. *USENIX Annual Technical Conference, General Track*, 43–56.
121. 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, from http://dl.acm.org/citation.cfm?doid=3136825.3136905
122. Nance, K., & Ryan, D. J. (2011). Legal Aspects of Digital Forensics: A Research Agenda. *2011 44th Hawaii International Conference on System Sciences* (pp. 1–6). Presented at the 2011 44th Hawaii International Conference on System Sciences (HICSS 2011), Kauai, HI: IEEE. Retrieved August 22, 2019, from http://ieeexplore.ieee.org/document/5719007/
123. Newsham, T., Palmer, C., Stamos, A., & Burns, J. (n.d.). Breaking Forensics Software: Weaknesses in Critical Evidence Collection, 30.
124. NIST. (2019). Computer Forensics Tool Catalog. Forensic Tool Taxonomy. Retrieved from http://toolcatalog.nist.gov/taxonomy/index.php?ff_id=5
125. Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology, 25.
126. Nykodym, N., Taylor, R., & Vilela, J. (2005). Criminal profiling and insider cyber crime. *Computer Law & Security Review*, *21*(5), 408–414.
127. OASIS. (2014). MQTT and the NIST Cybersecurity Framework Version 1.0. Retrieved from http://docs.oasis-open.org/mqtt/mqtt-nist-cybersecurity/v1.0/mqtt-nist-cybersecurity-v1.0.html
128. OASIS. (2019). Open Source Digital Forensic. Retrieved from https://web.archive.org/web/20150228083211/http://www2.opensourceforensics.org/tools
129. 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, from http://link.springer.com/10.1007/978-3-030-01535-0_19
130. Olivier, M. (2016). On a Scientific Theory of Digital Forensics. In G. Peterson & S. Shenoi (Eds.), *Advances in Digital Forensics XII* (Vol. 484, pp. 3–24). Cham: Springer International Publishing. Retrieved August 22, 2019, from http://link.springer.com/10.1007/978-3-319-46279-0_1
131. OpenText Corp. (2019). EnCase,. Retrieved from https://www.guidancesoftware.com/encase-forensic
132. Oracle. (2019). VirtualBox. Retrieved from https://www.virtualbox.org/
133. Oriwoh, E., Jazani, D., Epiphaniou, G., & Sant, P. (2013). Internet of Things Forensics:

Challenges and Approaches. *Proceedings of the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*. Presented at the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, Austin, United States: ICST. Retrieved March 28, 2020, from http://eudl.eu/doi/10.4108/icst.collaboratecom.2013.254159

134. Overill, R. E., Silomon, J. A. M., & Roscoe, K. A. (2013). Triage template pipelines in digital forensic investigations. *Digital Investigation*, *10*(2), 168–174.
135. Overill, R., Kwan, M., Chow, K.-P., Lai, P., & Law, F. (2009). A Cost-Effective Model for Digital Forensic Investigations. In G. Peterson & S. Shenoi (Eds.), *Advances in Digital Forensics V* (Vol. 306, pp. 231–240). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved October 1, 2019, from http://link.springer.com/10.1007/978-3-642-04155-6_17
136. Palmer, G. (2001, November 6). A Road Map for Digital Forensic Research. Report From the First Digital Forensic Research Workshop (DFRWS).
137. Pan, L., & Batten, L. M. (2009). Robust performance testing for digital forensic tools. *Digital Investigation*, *6*(1–2), 71–81.
138. PassMark™ Software. (2019). OSForensics. Retrieved from http://www.osforensics.com/
139. Patel, C. P., & Sharma, B. K. (2015). IJSRD - International Journal for Scientific Research & Development| Vol. 2, Issue 04, 2014 | ISSN (online): 2321-0613, *3*(08), 3.
140. Patil, S., Dharaskar, R., & Thakare, V. (2017). Cloud Forensics: A Framework for Digital Forensic in Cloud Based Environment by Identifying SLA Breaches by Cloud Actors, 6.
141. Peersman, C., Schulze, C., Rashid, A., Brennan, M., & Fischer, C. (2016). iCOP: Live forensics to reveal previously unknown criminal media on P2P networks. *Digital Investigation*, *18*, 50–64.
142. Peron, C. S. J., Legary, M., & Labs, S. (n.d.). Digital Anti-Forensics: Emerging trends in data transformation techniques, 11.
143. Petherick, W. (2005). The science of criminal profiling: All killers have their own modus operandi. New York, NY: Barnes & Noble Books.
144. Pollitt, M. (2010). A History of Digital Forensics. In K.-P. Chow & S. Shenoi (Eds.), *Advances in Digital Forensics VI* (Vol. 337, pp. 3–15). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved October 1, 2019, from http://link.springer.com/10.1007/978-3-642-15506-2_1
145. Portnoff, R. S., Afroz, S., Durrett, G., Kummerfeld, J. K., Berg-Kirkpatrick, T., McCoy, D., Levchenko, K., et al. (2017). Tools for Automated Analysis of Cybercriminal Markets. *Proceedings of the 26th International Conference on World Wide Web—WWW '17* (pp. 657–666). Presented at the the 26th International Conference, Perth, Australia: ACM Press. Retrieved August 22, 2019, from http://dl.acm.org/citation.cfm?doid=3038912.3052600
146. Prayudi, Y., Ashari, A., K Priyambodo, T., & 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.
147. Quick, D., & Choo, K.-K. R. (2016). Big forensic data reduction: Digital forensic images and electronic evidence. *Cluster Computing*, *19*(2), 723–740.
148. 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.

149. Quick, D., & Choo, K.-K. R. (n.d.). Data reduction and data mining framework for digital forensic evidence: Storage, intelligence, review and archive, 11.
150. Rajaboina, R., Reddy, P. C., & Kumar, R. A. (2015). Performance comparison of TCP, UDP and TFRC in static wireless environment. *2015 2nd International Conference on Electronics and Communication Systems (ICECS)* (pp. 206–212). Presented at the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India: IEEE. Retrieved October 1, 2019, from http://ieeexplore.ieee.org/document/7124893/
151. Razzaq, A., Anwar, Z., Ahmad, H. F., Latif, K., & Munir, F. (2014). Ontology for attack detection: An intelligent approach to web application security. *Computers & Security*, *45*, 124–146.
152. Reith, M., Carr, C., & Gunsch, G. (2002). The digital age can be characterized as the application of computer technology as a tool that enhances traditional methodologi. *International Journal of Digital Evidence*, *1*(3), 12.
153. Rekhis, S., & Boudriga, N. (2009). A Formal Rule-Based Scheme for Digital Investigation in Wireless Ad-hoc Networks. *2009 Fourth International IEEE Workshop on Systematic Approaches to Digital Forensic Engineering* (pp. 62–72). Presented at the 2009 Fourth International IEEE Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), Berkeley, California, USA: IEEE. Retrieved October 2, 2019, from http://ieeexplore.ieee.org/document/5341557/
154. Riadi, I., Istiyanto, J. E., & Ashari, A. (2012). Log Analysis Techniques using Clustering in Network Forensics, *10*, 9.
155. Rogers, M. D. (1999). Psychology of hackers: Steps toward a new taxonomy. Retrieved from http://www.dvara.net/HK/hacker_doc.pdf
156. Rogers, Marc. (2003). The role of criminal profiling in the computer forensics process. *Computers & Security*, *22*(4), 292–298.
157. Rogers, Marcus, Goldman, J., Mislan, R., Wedge, T., & Debrota, S. (2006). Computer Forensics Field Triage Process Model. *The Journal of Digital Forensics, Security and Law*, 27–40.
158. Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, *78*, 680–698.
159. Roussev, V., & Quates, C. (2012). Content triage with similarity digests: The M57 case study. *Digital Investigation*, *9*, S60–S68.
160. Roussev, V., Quates, C., & Martell, R. (2013). Real-time digital forensics and triage. *Digital Investigation*, *10*(2), 158–167.
161. Rowlingson, R. (2004). A Ten Step Process for Forensic Readiness, *2*(3), 28.
162. Roy, A., Dixit, R., Naskar, R., & Chakraborty, R. S. (2020). *Digital Image Forensics: Theory and Implementation*. Studies in Computational Intelligence (Vol. 755). Singapore: Springer Singapore. Retrieved March 29, 2020, from http://link.springer.com/10.1007/978-981-10-7644-2
163. Ryan, J. D., & Shpantzer, G. (2002). Legal Aspects of Digital Forensics. Retrieved from http://euro.ecom.cmu.edu/program/law/08-732/Evidence/RyanShpantzer.pdf
164. Salahdine, F., & Kaabouch, N. (2019). Social Engineering Attacks: A Survey. *Future*

Internet, *11*(4), 89.

165. Saleem, S., Popov, O., & Bagilli, I. (2014). Extended Abstract Digital Forensics Model with Preservation and Protection as Umbrella Principles. *Procedia Computer Science*, *35*, 812–821.
166. SARC. (2019). Steganography Analysis and Research Center (SARC). Retrieved from http://sarc-wv.blogspot.com/
167. SCG Canada Inc. (2019). Covert Forensic Imaging Device (CFID). Retrieved from http://www.scgcanada.com/
168. Schobbens, P.-Y., Heymans, P., & Trigaux, J.-C. (2006). Feature Diagrams: A Survey and a Formal Semantics. *14th IEEE International Requirements Engineering Conference (RE'06)* (pp. 139–148). Presented at the 14th IEEE International Requirements Engineering Conference, Minneapolis/St. Paul, MN: IEEE. Retrieved October 1, 2019, from http://ieeexplore.ieee.org/document/1704057/
169. Schultz, E. E., & Shumway, R. (2002). Incident response: A strategic guide to handling system and network security breaches (1st ed.). Indianapolis, Ind: New Riders Pub.
170. Scripting News. (2019). Outline Processor Markup Language. Retrieved from http://dev.opml.org
171. Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan, Rahman, S., & N. A. Khan, M. (2016). Digital Forensics through Application Behavior Analysis. *International Journal of Modern Education and Computer Science*, *8*(6), 50–56.
172. Sharma, H., Kanwal, N., & Batth, R. S. (2019). An Ontology of Digital Video Forensics: Classification, Research Gaps & Datasets. *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 485–491). Presented at the 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates: IEEE. Retrieved May 17, 2020, from https://ieeexplore.ieee.org/document/9004331/
173. Sharma, T. N. (2011). Analysis_of_Software_Cost_Estimation_using_COCOMO_II, *2*(6), 5.
174. Siahaan, A. P. U., & Rahim, R. (2017). *Post-Genesis Digital Forensics Investigation* (preprint). INA-Rxiv. Retrieved October 2, 2019, from https://osf.io/h5bds
175. Singer, D. A., & Kouda, R. (1999). A Comparison of the Weights-of-Evidence Method and Probabilistic Neural Networks. *Natural Resources Research*, *8*(4), 287–298.
176. Sommer, F., Dürrwang, J., & Kriesten, R. (2019). Survey and Classification of Automotive Security Attacks. *Information*, *10*(4), 148.
177. Soni, M., & Bharti, M. K. (2015). FraaS: A Framework for Digital Forensic Services in a Cloud-based Environment. *The International Journal of Forensic Computer Science*, *10*(1), 15–22.
178. Sremack, J. C. (2007). The Gap between Theory and Practice in Digital Forensics, 85.
179. Stamm, M. C., Lin, W. S., & Liu, K. J. R. (2012). Temporal Forensics and Anti-Forensics for Motion Compensated Video. *IEEE Transactions on Information Forensics and Security*, *7*(4), 1315–1329.
180. Steel, C. (2014). Idiographic Digital Profiling: Behavioral Analysis Based On Digital Forensics. *Journal of Digital Forensics, Security and Law*. Retrieved October 1, 2019, from http://commons.erau.edu/jdfsl/vol9/iss1/1/

181. Štuikys, V., Burbaitė, R., & Bespalova, K. (2015). The LO Sequencing Problem and Its Solution Using Meta-Programming-Based Approach. In G. Dregvaite & R. Damasevicius (Eds.), *Information and Software Technologies* (Vol. 538, pp. 151–164). Cham: Springer International Publishing. Retrieved October 2, 2019, from http://link.springer.com/10.1007/978-3-319-24770-0_14
182. Štuikys, V., & Damaševičius, R. (2013). *Meta-Programming and Model-Driven Meta-Program Development*. Advanced Information and Knowledge Processing (Vol. 5). London: Springer London. Retrieved October 2, 2019, from http://link.springer.com/10.1007/978-1-4471-4126-6
183. Takwa, O., Belgacem, C. R., & Adel, D. (2016). A New Digital Investigation Frameworks Comparison Method, *3*(4), 5.
184. Technical Analysis Group. (2004). *ISTS. Law enforcement tools and technologies for investigating cyber attacks: A national re-search and development agenda* (p. 35). Institute for security technology studies. Retrieved from http://index-of.es/Misc/pdf/ISTSLawEnforcementResearchandDevelopmentAgendaJune2004.pdf
185. Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, S2210832718301546.
186. Trifonov, R., Manolov, S., Yoshinov, R., Tsochev, G., & Pavlova, G. (2017). Artificial Intelligence Methods for Cyber Threats Intelligence, *2*, 7.
187. Tulowiecki, S. J. (2018). Information retrieval in physical geography: A method to recover geographical information from digitized historical documents. *Progress in Physical Geography: Earth and Environment*, *42*(3), 369–390.
188. Turnbull, B., & Randhawa, S. (2015). Automated event and social network extraction from digital evidence sources with ontological mapping. *Digital Investigation*, *13*, 94–106.
189. Turner, P. (2005a). Digital provenance – interpretation, verification and corroboration. *Digital Investigation*, *2*(1), 45–49.
190. Turner, P. (2005b). Unification of digital evidence from disparate sources (Digital Evidence Bags). *Digital Investigation*, *2*(3), 223–228.
191. Turvey, B. E. (2012). *Criminal profiling: An introduction to behavioral evidence analysis*. Oxford; Burlington, MA: Academic Press. Retrieved September 30, 2019, from http://site.ebrary.com/id/10480741
192. 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, from http://ieeexplore.ieee.org/document/8029567/
193. Van Buskirk, E., & Liu, V. T. (2006). Digital Evidence: Challenging the Presumption of Reliability. *Journal of Digital Forensic Practice*, *1*(1), 19–26.
194. Venčkauskas, A., Jusas, V., Paulikas, K., & Toldinas, J. (2016). A Methodology and Tool for Investigation of Artifacts Left by the BitTorrent Client. *Symmetry*, *8*(6), 40.
195. Venčkauskas, A., Morkevicius, N., Bagdonas, K., Damaševičius, R., & Maskeliūnas, R. (2018). A Lightweight Protocol for Secure Video Streaming. *Sensors*, *18*(5), 1554.
196. Venčkauskas, A., Morkevicius, N., Jukavičius, V., Damaševičius, R., Toldinas, J., & Grigaliūnas, Š. (2019). An Edge-Fog Secure Self-Authenticable Data Transfer Protocol. *Sensors*, *19*(16), 3612.

197. Venčkauskas, A., Toldinas, J., Grigaliūnas, Š., Damaševičius, R., & Jusas, V. (2015). Suitability of the digital forensic tools for investigation of cyber crime in the Internet of Things and Services (pp. 86–97). Presented at the The 3rd International Virtual Research Conference In Technical Disciplines. Retrieved October 1, 2019, from https://www.rcitd.com/archive/?vid=1&aid=2&kid=140301-67
198. Vukašinović, M. (2017). A Software System for automatic reaction to network anomalies and in Real Time Data Capturing necessary for investigation of digital Forensics, *11*, 8.
199. W3C. (2019). Web Ontology Language. Retrieved from http://www.w3.org/TR/2012/REC-owl2-overview-20121211/
200. Wahyudi, E., Riadi, I., & Prayudi, Y. (2018). Virtual Machine Forensic Analysis And Recovery Method For Recovery And Analysis Digital Evidence, *16*(2), 8.
201. Wang, N. (2017). A Knowledge Model of Digital Evidence Review Elements Based on Ontology: *International Journal of Digital Crime and Forensics*, *9*(3), 49–57.
202. Wei, W., & Wo, M. (n.d.). Algorithm Research of Known-plaintext Attack on Double Random Phase Mask Based on WSNs, 10.
203. Wilsdon, T., & Slay, J. (2006). Validation of Forensic Computing Software Utilizing Black Box Testing Techniques, 10.
204. Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314.
205. Yaqoob, I., Hashem, I. A. T., Ahmed, A., Kazmi, S. M. A., & Hong, C. S. (2019). Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges. *Future Generation Computer Systems*, *92*, 265–275.
206. Yunianto, E., Prayudi, Y., & Sugiantoro, B. (2019). B-DEC: Digital Evidence Cabinet based on Blockchain for Evidence Management. *International Journal of Computer Applications*, *181*(45), 22–29.
207. Yusoff, Y., Ismail, R., & Hassan, Z. (2011). Common Phases of Computer Forensics Investigation Models. *International Journal of Computer Science and Information Technology*, *3*(3), 17–31.
208. Zawoad, S., & Hasan, R. (2015). FAIoT: Towards Building a Forensics Aware Eco System for the Internet of Things. *2015 IEEE International Conference on Services Computing* (pp. 279–284). Presented at the 2015 IEEE International Conference on Services Computing (SCC), New York City, NY, USA: IEEE. Retrieved October 2, 2019, from http://ieeexplore.ieee.org/document/7207364/

8 THE LIST OF SCIENTIFIC PUBLICATIONS BY THE AUTHOR ON THE TOPIC OF THE DISSERTATION

Papers in the Reviewed Scientific Journal

Grigaliūnas, Šarūnas; Toldinas, Jevgenijus; Venčkauskas, Algimantas. An ontology-based transformation model for the digital forensics domain // Elektronika ir elektrotechnika. Kaunas: KTU. ISSN 1392-1215. eISSN 2029-5731. 2017, vol. 23, iss. 3, p. 78-82. DOI: 10.5755/j01.eie.23.3.18337.

Šarūnas Grigaliūnas, Jevgenijus Toldinas. „Habits Attribution and Digital Evidence Object Models Based Tool for Cybercrime Investigation“. Baltic J. Modern Computing, Vol. 8 (2020), No. 2, 275-292. DOI: 10.22364/bjmc.2020.8.2.05.

Papers in other Editions

Damaševičius, Robertas; Toldinas, Jevgenijus; Venčkauskas, Algimantas; Grigaliūnas, Šarūnas; Morkevičius, Nerijus; Jukavičius, Vaidas. Visual analytics for cyber security domain: state-of-the-art and challenges // Information and software technologies: 25th international conference, ICIST 2019, Vilnius, Lithuania, October 10–12, 2019: proceedings / Robertas Damaševičius, Giedrė Vasiljevienė (Eds.). Cham: Springer, 2019. ISBN 9783030302740. eISBN 9783030302757. p. 256-270. (Communications in computer and information science, ISSN 1865-0929, eISSN 1865-0937; vol. 1078). DOI: 10.1007/978-3-030-30275-7_20.

Grigaliūnas, Šarūnas; Toldinas, Jevgenijus. 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.

Toldinas, Jevgenijus; Venčkauskas, Algimantas; Grigaliūnas, Šarūnas; Damaševičius, Robertas; Jusas, Vacius. Suitability of the digital forensic tools for investigation of cyber crime in the internet of things and services // RCITD 2015: 3rd international virtual research conference in technical disciplines, October, 19-23, 2015. Zilina: EDIS - Publishing Institution of the University of Zilina, 2015. ISBN 9788055411255. ISSN 2453-6571. eISSN 1339-5076. 2015, vol. 3, iss. 1, p. 86-97. DOI: 10.18638/rcitd.2015.3.1.

Other conference abstracts and non-peer reviewed conference papers

Grigaliūnas, Šarūnas; Toldinas, Jevgenijus; Lozinskis, Borisas. Habits attribution and digital evidence object tool with fuzzy logic for cybercrime investigation // 11th international workshop on data analysis methods for software systems, Druskininkai, Lithuania, November 28-30, 2019 / Lithuanian Computer Society, Vilnius University Institute of Data Science and Digital Technologies, Lithuanian Academy of Sciences. Vilnius: Vilnius University, 2019. ISBN 9786090703243. eISBN 9786090703250. p. 29.

Grigaliūnas, Šarūnas; Toldinas, Jevgenijus. Digital evidence object model for cybercrime investigation // 9th International workshop on data analysis methods for

software systems, DAMSS: Druskininkai, Lithuania, November 30 - December 2, 2017 / Lithuanian Computer Society, Vilnius University, Institute of Data Science and Digital Technologies, Lithuanian Academy of Sciences. Vilnius: Vilnius University, 2017. ISBN 9789986680642. p. 18-19. DOI: 10.15388/DAMSS.2017.

9 ANNEXESS[2]

Annex A. Author's Declaration of Academic Integrity

Annex B. Copies of Scientific Publications by the Author on the Topic of Dissertation

[2] The annexes are provided in the enclosed usb disk.

SL344. 2020-07-23, 14 leidyb. apsk. l. Tiražas 12 egz.
Išleido Kauno technologijos universitetas, K. Donelaičio g. 73, 44249 Kaunas
Spausdino leidyklos „Technologija“ spaustuvė, Studentų g. 54, 51424 Kaunas