symmetry

MDPI

*Review*

# Methods and Tools of Digital Triage in Forensic Context: Survey and Future Directions

**Vacius Jusas [1,*], Darius Birvinskas [2] and Elvar Gahramanov [1]**

[1] Software Engineering Department, Kaunas University of Technology, Studentu St. 50, LT-51368 Kaunas, Lithuania; elvar.gahramanov@ktu.edu
[2] Computer Department, Kaunas University of Technology, Studentu St. 50, LT-51368 Kaunas, Lithuania; darius.birvinskas@ktu.lt
* Correspondence: vacius.jusas@ktu.lt; Tel.: +370-65-676-159

Academic Editor: Shu-Ching Chen
Received: 7 February 2017; Accepted: 22 March 2017; Published: 28 March 2017

**Abstract:** Digital triage is the first investigative step of the forensic examination. The digital triage comes in two forms, live triage and post-mortem triage. The primary goal of the live triage is a rapid extraction of an intelligence from the potential sources. The live triage raises legitimate concerns. The post-mortem triage is conducted in the laboratory and its main goal is ranking of the seized devices for the possible existence of the relevant evidence. The digital triage has the potential to quickly identify items that are likely to contain the evidential data. Therefore, it is a solution to the problem of case backlogs. However, existing methods and tools of the digital triage have limitations, especially, in the forensic context. Nevertheless, we have no better solution for the time being. In this paper, we critically review published research works and the proposed solutions for digital triage. The review is divided into four sections as follows: live triage, post-mortem triage, mobile device triage, and triage tools. We conclude that many challenges are awaiting for the developers in creating methods and tools of digital triage in order to keep pace with the development of new technologies.

**Keywords:** digital forensic; digital triage; live triage; post-mortem triage; triage tools

## 1. Introduction

The volume of data for forensic investigation keeps constantly growing. This is a result of the continuing technology development when scale and bounds of the Internet rapidly change and social networks come to everyday use. The storage capacity expands to new areas when smart phones become part of the Internet devices and cloud storage services are offered. The digital forensic process is very time consuming, because it requires the examination of all available data volumes collected from the cybercrime scene. The digital forensic process commences with the collection, duplication, and authentication of every piece of digital media prior to examination. Moreover, every action taken has to adhere to the legitimacy rules so that the obtained digital evidence could be presented in the court. However, life is very dynamic, and the situations, in which some information about a possible cybercrime has to be obtained as promptly as possible without adhering to the rules of long legal scrutiny, arise. Of course, the information obtained in a such way cannot be directly used in the court; however, a quick access to such knowledge can speed up the future process of digital forensics and, is some situations, can even save somebody's life. Therefore, such actions are justifiable.

A process that takes place prior to the standard forensic methodology is called digital triage. It can provide valuable intelligence without subjecting digital evidence to a full examination. This quick intelligence can be used in the field to guide the search and seizure, and in the laboratory to determine if a media is worth to be examined.

The term "triage" comes from the field of medicine, in which it refers to the situations when because of having limited resources, the injured people are ranked according to the necessity to receive treatment. Such ranking ensures the achievement of the least damage to patients when resources are limited [1].

Rogers et al. [2], the authors of the first field triage model in computer forensics, define triage as a process of ranking objects in terms of importance or priority. Casey et al. [3] define triage in digital forensics as part of forensic examination process. The forensic examination is described as three-tier strategy consisting of three levels: (i) survey/triage forensic inspection, (ii) preliminary forensic examination, and (iii) in-depth forensic examination. The first stage, in which many potential sources of digital evidence for specific information are reviewed, is alternatively referred to as survey or triage. The same idea that triage is part of forensic examination, is supported in later works [4–8]. Casey [4] underlines that triage is effective for prioritizing, but it is not a substitute for a more thorough review. Casey [5] argues that triage is a technical process, which can be performed outside a laboratory by professionals with basic training and limited oversight. Categorizing digital triage as a technical process makes it more clear that the information has not undergone rigorous quality assessment and its legitimacy has not been evaluated.

There are many other definitions of triage, which slightly differ depending on the attributed qualities [7–12]. The diversity of triage definitions reflects the variety of the views and indicates the immaturity of the field. However, it is not the main problem. The focus should be devoted to the decision whether digital triage is a forensic process. As Cantrell et al. [13] state, "Digital triage is not a forensic process by definition". It is not clear to which definition Cantrell et al. [13] refer. It is possible to suppose that it is the definition by Rogers et al. [2]. However, other definitions exist, and this statement is not true for all the cases [7,11,14]. Koopman and James [11], and Roussev et al. [7] use the term "digital forensic triage". If digital triage is not the forensic process, then the term "forensic" cannot be used together with the term "digital triage", because it misleads. Hong et al. [14] introduce a triage model that is adapted to the requirements of the legal Korean system. Consequently, the proposed triage model adheres to the rules of the forensic process. Moreover, Hong et al. [14] suggest establishing a triage model individually for the legal system of a specific country.

To summarize the diversity of views on digital triage, we stress the following features:

1. Digital triage is a technical process to provide information for the forensic examination, but does not involve the evaluation of digital evidence
2. The goal of digital triage is to rapidly review many potential sources of digital evidence for specific information and prioritize the digital media to make the subsequent analysis easier
3. The term "forensic" cannot be used together with the term "digital triage" if the process of digital triage does not adhere to the rules of the forensic process specific to the country

Digital triage comes in two forms: live and post-mortem. The post-mortem form of triage, which is conducted on the digital image, is not always recognized as triage. We suppose that both forms of digital triage are equally important. Live triage raises many concerns, because it is conducted on the live system, and the destruction of the likely evidence is possible. However, live digital triage has several advantages:

1. It enables a rapid extraction of intelligence that can be used for suspect interrogation
2. Some data can be lost if the computer is shut down

The primary concern inherent to both forms of digital triage is that the evidential data can remain unnoticed [15]. Pollitt [16] argues that the process of digital triage in the context of forensics is an admission of failure. However, he recognizes that for now a better approach does not exist.

Moreover, the term "triage" becomes the common word to indicate the initial and rapid step in the different areas of the forensic investigation. For example, it is used in the retail industry [17], in the internet of things [18], in the fraud of identity and travel documents [19].

We review the research works related to digital triage. We divide the review into four sections as follows: live triage, post-mortem triage, mobile device triage, and triage tools. The largest section is on the triage tools. Such abundance of research works highlights the practical need for triage tools.

In the next section, we review the models and methods of live triage.

## 2. Models and Methods of Live Triage

Rogers et al. [2] introduce the model for the field triage process in computer forensics and name it the Cyber Forensic Field Triage Process Model (CFFTPM). The CFFTPM has six phases: planning, triage, usage/user profiles, chronology/timeline, internet activity, and case specific evidence. Each phase has several sub-tasks and considerations that vary according to the specifics of the case and operating system under investigation. The CFFTPM originates from child pornography cases. Nevertheless, it is general enough to be applicable to other possible cases; however, the model cannot be considered as the ultimate solution for every case. It is also important to note that the proposed model does not preclude transporting the system to a laboratory environment for a more thorough investigation.

Cantrell et al. [13] discuss a proposed model for digital triage. The proposed model is a linear framework, except the preservation phase that is an investigative principle preserved throughout all the phases. The first phase is planning and readiness that occurs before the investigation onsite. The next phase is live forensic that is included as an optional step, depending on the need and expertise, and it must occur prior to the following phases because the volatile memory can be lost very quickly. The middle three phases: computer profile phase, crime potential phase, and presentation phase are intended to be an automated process, coded as a computer program or script using the existing tools. The last phase, triage examination phase, is optional depending on the need. The triage examination should be an automated process that is guided by the examiner using predefined templates specific to each case.

Hong et al. [14] propose a theoretical framework for implementing a triage model. The requirement for the triage model is to consider the limiting factors of the onsite search and seizure. The framework consists of three phases: assessment, triage model, and reassessment. The proposed framework is based on the assumption that reassessments are performed periodically according to the changes in search and the conditions of the onsite seizure. To establish a triage model, a questionnaire that consists of 48 questions, which are provided in the paper, was prepared; it was answered by 58 respondents in total. The paper presents a large discussion of the results. After assessing the results of the questionnaire, a new triage model is proposed. The triage process is divided into four steps: planning, execution, categorization, and decision. The properly collected information mostly depends on the execution step. The execution step prioritizes the file types for the search according to three types of crime: personal general crime, personal high-tech crime, and corporate general crime. Next, the file search is conducted in the following order: timeline of interest; filename- or contents-based keywords search; and file/directory path-based search. Another important procedure in the execution step is the detection of suspicious files. The proposed triage model can be applied only to personal computers and it is tailored to the Korean legal system requirements for the privacy protection.

Overill et al. [20] propose an attractive idea to introduce triage template pipelines into the investigative process for the most popular types of digital crimes, enabling digital evidence to be examined according to a number of prioritised criteria. Each specific digital crime has its own template of prioritised devices and the data based on the cost-effectiveness criteria of front-loading probative value and back-loading resource utilisation. The authors declare that about 80% of all digital crimes in Hong Kong are accounted for just five types of crime. However, they do not enumerate these types of crime. The authors state that "the work this far has addressed the set of five digital crime templates", however, the examples of templates for only two digital crimes are provided. To be more precise, they are the Distributed Denial of Service (DDoS) template diagram and the Peer-to-Peer (P2P) template diagram. Moreover, the construction of these example templates is not discussed in detail. An advantage of the triage template pipeline approach over the triage tools is that

the evidential recovery process can be terminated as soon as it becomes apparent that the probative value criterion has been fulfilled. Therefore, the triage time can be shorter in some cases. The essence of the proposed triage template pipelines is formalized common sense.

Roussev et al. [7] argue and analyze forensic triage as a real-time computation problem, which has allotted limited time and resources. One hour is considered to be an acceptable time limit for triage. The authors assume that an increase in the performance can be achieved if the acquisition and processing start and complete at almost the same time. It means that the processing should be as fast as data the cloning. The suitability of the most common open-source implementations and of most common forensic procedures to fit into the time constraints is investigated experimentally. The authors state that the triage investigation can be carried out in the field and in the laboratory. For the fieldwork, they consider 8-core workstation and for the laboratory, they consider 48-core server. The obtained results show that only a few basic methods, like file metadata extraction, crypto-hashing, and registry extraction, can fit into the time budget in the workstation triage. To increase the performance of the file acquisition, Roussev et al. [7] implement a Latency-Optimized Target Acquisition (LOTA) scheme. The main idea of this scheme is that the metadata of a filesystem is parsed to make an inverse map from blocks to files before cloning the target. This procedure allows sequential scanning of blocks and reconstructing the files. The LOTA scheme enables an improvement of a factor of two for files larger than 1 M and a factor of 100 for smaller files. It is recommended to use the scheme in the forensic environment routinely. The authors advocate employing parallel computations to obtain higher processing rates.

Lim and Lee [21] describe a unified evidence container XeBag for storing diverse digital evidence from different sources. The XeBag can be used for selective evidence collection and searching on the live system. The file structure of XeBag is based on well-known compression file formats, PKZip and WinRAR. To record forensic metadata, an Extensible Markup Language (XML) document is included additionally for each stored object. The XML format is a popular data exchange format, therefore, it enables easy access to the data. The authors provide a description of a video surveillance system to show how its digital evidence is stored and can be retrieved from the unified evidence container XeBag.

Grier and Richard III [22] introduce a new approach, called sifting collectors, for imaging of the selected regions of disk drives. The sifting collectors create a sector-by-sector, bit-for-bit exact image of disk regions that have forensic value. The forensics image is produced in an Advanced Forensics Format v3 [23], and it is fully compatible with the existing forensic tools. The selection of the regions that have forensics value is based on profiles. The authors do not expect that the examiners can prepare the profiles themselves, therefore, the profiles must be created and stored in a library. The sifting collectors firstly collect the metadata according to the defined profile. Then they interpret metadata, determine sectors of interest, and assemble them in the disk order. As a result, their methods are not suitable for unknown filesystems. If profiles are not possible to define, the alternative proposes to include a person in the scanning loop to decide what is relevant. The implemented prototype targets New Technology File System (NTFS) as a file system and uses the Master File Table as its primary source. The conducted experiment shows a speed up from 3 to 13 times in comparison to the forensic image acquisition tool Sleuthkit [24] for the test cases. The absolute values of runtimes are not provided. The accuracy of the region selection is between 54% and 95% for the considered test cases. Faster image acquisition time gives less accuracy. One important limitation of sifting collectors is their susceptibility to steganography and anti-forensics.

Penrose et al. [25] present an approach for fast contraband file detection on the device itself. The approach is based on clusters scanning, hash calculating, and comparison to the database. The cluster size is 4 KiB. A Bloom filter is used to store the cluster hashes of the contraband files. The Bloom filter reduces the size of the database of the block level Message-Digest Algorithm 5 (MD5) hashes by an order of magnitude; however, it costs a small false positive rate. The designed Bloom filter is 1 GiB in size and it uses eight hash functions. A larger Bloom filter enables faster access to the hashes of the contraband files. The performed experiment shows that the approach achieves 99.9% accuracy

scanning for contraband files in minutes. Some false positives are encountered; however, the results are positive for the existence of all contraband files. The experiment was conducted in legitimate computing environment. The authors draw a conclusion that this type of case can be further investigated in a forensically sound environment.

Turnbull and Randhawa [26] describe an ontology-based approach to assist examiner-led triage. The purpose of the approach is to enable a less technically intrinsic user to run a triage tool. This is implemented by collecting low-level artifacts and inferencing hypotheses from the collected facts. The approach is oriented to automatically deriving events from the base of the forensics artefacts. A Resource Descriptive Framework (RDF) is used as the basis of the ontology. The representative feature of the approach is that the layered multiple ontologies are designed over the same dataset. The description of the ontologies used is vague. The authors find some advantages of the RDF; however, they recognize that a Web Ontology Language (OWL) could provide more possibilities. The authors suggest that the approach is applicable for the extraction of information from social networks, though, no evidence of such application can be found in the paper. The implemented system to provide a proof-of-concept consists of a knowledge base, data ingestors, reasoners, and a visualiser. The visualiser is hardcoded into the used ontology. Neither test, nor real cases are provided. To conclude, the idea of the approach is attractive, however, the description and the development are immature.

Hitchcock et al. [27] introduce a Digital Field Triage (DFT) model to offload some of the initial tasks performed in the field by forensic examiners to non-digital evidence specialists. The primary goals of the model are twofold: (i) To increase the efficiency of an investigation by providing digital evidence in a timely manner; (ii) To decrease the backlog of files at a forensic laboratory. The proposed model is based on Rogers et al. [2] and it has four phases: planning, assessment, reporting, and threshold. The DFT model has inherent risks associated with it. They are as follows: the management, training, and supporting tools. The management and ongoing training are integral parts of the success of the DFT model. The tools must support the management. For the DFT to work, there are three fundamental concepts:

1. DFT must work with a supervising examiner
2. DFT must maintain the forensic integrity of the digital evidence
3. A DFT assessment does not replace the forensic analysis

Therefore, the DFT model is not a replacement for full analysis, but is part of the overall strategy of handling digital evidence. The first version of the DFT model was implemented in Canada six years ago. The implementation achieved the goals pursued by the model; however, persistent attention needs to be turned to the risks associated with the model.

Leimich et al. [28] propose a variation of cloud forensic methodology tailored to a live analysis of Random-Access Memory (RAM) for Hadoop Distributed File System (HDFS). The aim of the methodology is to minimize the disruption to the data center after data breach. The Hadoop is a Java implemented system developed for UNIX based operating systems. It is a master/slave distributed architecture for storing and processing big data. The HDFS consists of DataNodes (slaves), which store the data, and NameNode (master) that manages the DataNodes. The methodology is oriented to the acquisition of the NameNode contents of to pinpoint the affected DataNodes. The forensic analysis of the DataNodes is out of scope of the proposed methodology. The methodology contains nine phases: preparation, live acquisition of the NameNode, initial cluster reconnaissance, checkpointing via a forensic workstation, live artefact analysis, establish 'suspect' transactions and map to data block, perform targeted dead acquisition of the DataNodes, data reconstruction, and report. To test the validity of the methodology a small HDFS cluster that has one master and three slaves, was configured with a single scenario of deleted data. The phase of data reconstruction is not carried out. The experiment confirms that the methodology enables locating the deleted data blocks. Liemich et al. [28] discuss the ability to implement the proposed methodology in forensic tool in compliance with the National Institute of Standards and Technology (NIST) Computer Forensic Tool Testing criteria.

Montasari [8] extends the Rogers et al.'s [2] model by dividing all phases into two stages and introducing new sub-tasks into the phases. The single planning activity is assigned to the first stage. The planning should be carried out before attending the site. Montasari [8] considers many models of the forensics process, not just triage models, because according to the author, the single model proposed by Rogers et al. [2] exists for the onsite triage process. The author selects activities, which would be appropriate for the triage process, from other models. Therefore, several sub-tasks are added to the model of the forensics field triage process, and the model is presented in a more detailed and categorized way. Additionally, the model is extended by a set of investigative principles joined into a group under the name of "Overriding Principles", which are an additional contribution of the paper. These principles are as follows:

1. To preserve chain of custody
2. To maintain an accurate audit trail
3. To maintain a restricted access control
4. To maintain an effective case management
5. To maintain the information flow

Peersman et al. [29] present an approach that incorporates artificial intelligence and machine learning techniques (support vector machines) to automatically label new Child Sexual Abuse (CSA) media. The approach employs two stages for labelling the unknown CSA files. The first stage uses the text categorization techniques to determine whether a file contains CSA content based on its filename. The text categorization applies the following features: predefined keywords, forms of explicit language use, expressions relating to children and family relations in English, French, German, Italian, Dutch, and Japanese. Additionally, all patterns of two, three, and four consecutive characters are extracted from the filenames. The second stage gets the files from the first level and examines the visual content of images and audio files. The second stage bases the decision on multi-modal features. The multi-modal features consist of the following representations: colour-correlograms, skin features, visual words and visual pyramids, and audio words for audio files. The conducted experiment shows a false positive rate of 20.3% after the first stage. The second stage reduces the false positive rate to 7.9% for images and 4.3% for videos. The approach is implemented into the iCOP toolkit [30] that performs live forensic analysis on a P2P network. Therefore, the proposed approach is designed for a proactive monitoring activity. To label the most pertinent candidates for the CSA media, an examiner can login to the iCOP canvas that automatically arrange the results. Additionally, the approach can be adapted to the identification of the new CSA media during a reactive investigation. The approach is implemented in the Gnutella P2P network.

Quick and Choo [31] develop the idea of data reduction introduced in [32]. The authors present the methodology to reduce the data volume using selective imaging. The methodology suggests to select only the key files and data. Windows, Apple and Linux operating systems and their filesystems are considered. A forensic examiner makes the decision to include or exclude particular file types. The decision is based on the data, contained in these file types, relevance to the case. The other possibility considered for reducing data volume is a thumbnailing of video, movie, and picture files. The thumbnailing significantly reduces large image files. Once the file types are selected and some thumbnails are loaded into the forensics software, the logical image file is created. The presented methodology can be applied using common digital forensics tools. The methodology is applied to test as well as real world data. Many results of the experiments that illustrate the viability of the methodology are provided. In general, time reductions observed are 14 min on average to collect a logical image and process in the Internet Evidence Finder, meanwhile the processing of full forensic image takes 8 h 4 min on average. The presented methodology can be applied to either write-blocked physical media or a forensic image.

## 3. Methods of Post-Mortem Triage

Marturana and Tacconi [33] summarize the research works [34,35] delivered at conferences and present a model intended for both live and post-mortem triage using machine learning techniques.

The presented model consists of the following four steps: forensic acquisition, feature extraction and normalization, context and priority definition, and data classification. For such model, there are two main challenges, the definition of crime-related features and collection of a consistent set of classified samples related to the investigated crimes. The crime-related features are defined for two cases studies, copyright infringement and child pornography exchange. Guidelines for using the classifiers are provided. The attention of the experiment is mostly directed to the comparison of the classifiers used at the last stage of the model. No conclusion is made as to which classifier is best suited for the investigated cases. The presented statistical approach has proven to be valid for ranking the digital evidence related to copyright infringement and child pornography exchange. However, for this approach to be viable, it is necessary to have a deep understanding of possible relations between the crime under investigation and the potential digital evidence.

McClelland and Marturana [36] extend the research presented by Marturana and Tacconi [33]. The authors investigate the impact of the feature manipulation on the accuracy of the classification. The weights are assigned to the features. Two approaches are used for assigning weights to the features, automatic and manual. The automated feature weights are quantified using the Kullback–Leibler measure. The manual weights are determined on the basis of the surveyed digital forensic experts' contribution. The Naïve Bayes classifier is used for the experiment. The only improvement is achieved in the child pornography case.

Horsman et al. [10] extend the ideas presented in [37] and discuss a Case-Based Reasoning Forensic Triager (CBR-FT) method for retrieving the evidential data based on the location of the digital evidence in the past cases. The CBR-FT maintains a knowledge base for gathering the previous experience. Each location on the system stored in the knowledge base is assigned an evidence relevance rating (ERR), which is used as the prior probabilities in the Bayesian model to determine the priority of a particular location for searching. The model enables calculating a primary relevance figure (PRF) for each location. The search is carried out in two stages: in the first stage, only locations with a PRF above 0.5 are used, while the second stage is optional. If the examiner suspects that additional evidence can exist, s/he proceeds to the second stage. During the second stage, the examiner focuses on identifying similar patterns in cases stored in the CBR-FT knowledge base. The CBR-FT knowledge base must cover enough cases to reflect its target population correctly. That is the first restriction for application of the method. The study focuses on fraud offences and it has constructed a fraud knowledge base from 47 prior investigations. The experiment shows that the CBR-FT is more effective when compared to a commercial application EnCase Portable [38], which uses precision and recall rates. However, an additional shortcoming of this study is that it focuses only on offences of fraud.

Bashir and Khan [39] suggest a triage framework oriented to analyzing and resolving an attack. The framework contains the usual steps that belong to a general investigative process. The term "triage" refers to a certain part of the framework. The main idea of the triage framework is to create a blacklist database that contains a list of the previously known attacks with details on how to resolve. Every attack is characterized by six attributes: identifier, name, description, status, signature, and then counter measures. The key attribute is the signature that is a placeholder to store unique signatures of cyber-attacks in the form of MD5 hashes. If the signature of any of the affected files is found in the blacklist database, then it means that the attack is known. The answer to how to resolve it is in the blacklist database. However, if the attack is unknown, there is no triage process; a detailed analysis follows. The blacklist database is updated periodically on the basis of the new knowledge and new attacks.

Dalins et al. [40] introduce a crawl and search method that can be used for digital triage. The proposed method adopts the Monte Carlo Tree Search strategy that is used in games for the filesystem search, which is called Monte Carlo Filesystem Search (MCFS). The original random selection is leveraged with non-binary scoring to keep guided search. Three file scoring methods are introduced, each built on the previous one: simple scorer, type of interest scorer, and similarity-based scorer. Other customizations are made to deliver better performance: integration of domain knowledge to enhance guided search, use of proprietary Microsoft PhotoDNA algorithm to measure

the similarity of images, and skin tone detection to identify exposed skin that is usual component of child pornography. The experiment is carried out on real data that was obtained from the Australian Federal Police. The data presented as forensics images are related to the possession and online trading of child pornography. The experiment shows that the proposed MCFS is an effective method for larger and complex tree structures of the file system hierarchy. The search efficiency can be improved by around a third compared to uninformed depth-first search. However, the integration of domain knowledge and skin tone detection scoring showed lower results than expected. An additional investigation is necessary to improve these customizations. In general, the improved proposed method is promising, since many performance limitations arise due to the complicated filesystem design [7].

Fahdi et al. [41] investigate the possibility of utilizing the Self-Organising Map (SOM) technique to automatically cluster notable artefacts that are relevant to the case. A SOM is a neural network that generates a mapping from the high dimensional input data into a regular two dimensional array of nodes based upon their similarity in an unsupervised manner. The approach is based on using the metadata from several sources, such as the file system, email, and Internet, as the input into the SOM clustering. Moreover, the approach is oriented at the investigation of the suspects' systems rather than the victims' systems. Several pre-processing options are employed before the application of the approach. These options include the creation of the file list, expanding compound files, data carving, entropy test for encryption, and known file search. The results of data carving are not included into the file list of the SOM. Data carving should not be deployed during triage, since data carving tends to generate a lot of data due to high false positive rates [7]. The experiment shows that the use of the approach as a triage to verify the existence of the notable files allows identifying 38.6% of notable files at a cost of 1.3% of noise files. It is possible to expand the network size to increase the percentage of the notable files, however, at the cost of picking up more noise files. Most of the analysis takes a relatively trivial amount of time for small data sets (several GB); however, it takes an hour on average to process a large data set (0.5 TB). The appeal of the approach is that the only examiner interaction required in this process is when selecting the crime category. The approach can be a building block with further research and refinement to provide a triage tool for investigating simpler and technically more trivial cases that represent a large proportion of the forensic examiners' daily activities.

## 4. Triage of Mobile Devices

Mislan et al. [42] discuss the onsite triage process for mobile devices. The following steps are suggested for an on-scene triage investigation of mobile devices:

1. Initiate the chain of custody
2. Isolate the device from the network
3. Disable the security features
4. Extract the limited data
5. Review the extracted data
6. Preview the removable storage media.

All the steps are discussed in details. The process of the investigation should be well documented in order to validate the results. The mobile device technicians, who are less experienced as technical examiners, should perform the onsite triage. The basic requirements for the automated onsite triage tools are outlined. To present shortly, they are as follows: simplicity of use, audit trail, and access control. The legal allowances of the United States to examine mobile devices are considered as well.

Walls et al. [43] introduce an investigative tool DEC0DE for recovering information from mobile phones with unknown storage formats. The main idea is that the data formats from known phone models can be leveraged for recovering information from the new phone models. The evaluation focuses on feature phones, i.e., phones with less capability than that of smartphones. The DEC0DE takes the physical image of a mobile phone as input. It is the first limitation of the tool, because the image is not its concern. The second limitation is the assumption that the owner of the phone has left

the data in plaintext format. The next shortcoming is that the extracted results are limited to address books and call log records. The contribution of the paper is a technique for an empirical mobile phone data analysis. The used technique consists of two steps—removal of known data and recovering information from the remaining data. The latter step is called an inference process. Block hash filtering accomplishes the first step. The second step adapts the techniques from natural language processing, namely the context-free grammar, and uses probabilistic finite state machines to encode typical data structures. The Viterbi algorithm treats the created finite state machines twice. Finally, the decision tree classifier is used to remove the potential false positive. The development is based on the four following models: Nokia 3200B, LG G4015, Motorola v551, and Samsung SGH-T309. The performance of DEC0DE's inference engine is evaluated against two metrics, recall and precision. The conducted experiment on the phones that have not been seen previously shows an average recall of 93% and precision of 52% for address books, and an average recall of 97% and precision of 80% for call logs.

Marturana et al. [34] discuss the application of machine learning algorithms for digital triage of mobile phones. The triage stage is introduced between the stages of acquisition and analysis. The extracted data are firstly preprocessed in order to clean data, remove redundant attributes, and normalize data. Several classification algorithms are used to show the ability to classify whether a mobile phone was used to commit a pedophilia crime. The attention is devoted to the performance of the classification algorithms. The research is the first step towards the post-mortem forensic triage of mobile phones.

Varma et al. [44] present a system, called LIFTR, for prioritizing the information recovered from Android phones. The initial data for the system is a forensic image extracted by a recovery engine. Three recovery engines—DEC0DE [43], Bulk Extractor [45], and Strings, a common UNIX utility for identifying stings of printable characters in a file —are used as the suppliers of the forensic images. Therefore, the LIFTR should operate in concert with the recovery engine, as it augments the results obtained by the engine. The basic idea is that the recovery engine returns many unrelated items to the investigated crime results, since it does not consider the semantics behind the recovered content. Varma et al. [44] explore the filesystem of the Android phones and learn the rules of storing the information. These rules learnt and the feedback from the examiner form the basis for information prioritizing. The examiner labels the relevant information units of the investigated crime at the page level. The labeling takes several times and it is performed in the cycle. All the information is ranked based on a combination of the examiner's feedback, the actual content, and the storage system locality information. To test the validity of the approach, the open-source prototype of the system LIFTR is implemented. The LIFTR's ranking algorithm is evaluated against 13 previously owned Android smart phones. Moreover, the set includes nine phones with the Yaffs filesystem [46]. To improve the results, the authors wrote a special Yaffs parser to identify the expired pages that are important to the information relevance. The experiment shows that the LIFTR ranking improves the score of standard information retrieval metric from 0.0 to an average 0.88.

Guido et al. [47] introduce a differential acquisition technique that can be used for forensic image acquisition of mobile devices for triage purposes. The advantage of the technique introduced is its runtime that is several times faster than other compared commercial tools or techniques. The main idea is to use the precomputed baseline hashes. Therefore, the hashes of the unknown blocks are only sent to the server. The prototype named Hawkeye is implemented. The Hawkeye uses MD5 algorithm for hashing. Several other improvements are implemented to obtain less runtime. They are as follows: threading (10 threads by default) and comparison function of the zero block. The Hawkeye runs on Android devices in the recovery mode. The experiment is performed with 16 GB Samsung Galaxy S3 smartphone (Samsung, Seoul, South Korea). The acquisition techniques of the tool can be applied to other platforms, such as iOS (Apple Inc., Cupertino, CA, USA) as well.

## 5. Triage Tools

We only review the tools that are presented in research papers. Such tools are usually not commercial.

Roussev and Quates [12] assume that digital triage is conducted in the laboratory environment. The purpose of digital triage is to identify the most relevant artifacts. The speed and reliability are critical for digital triage; these requirements are imperative for building triage tools. The authors demonstrate the use of tools sdhash and sdhash-dd to generate similarity digests for the purposes of digital triage. The first tool, sdhash, is an open source tool [48]. The similarity digests are intended to solve the two following problems of content correlation: resemblance and containment. The overall process of the application is quite simple; the similarity digests are generated for all targets and all queries. The queries are then systematically applied to the targets. The resemblance of data is detected at the bit-stream level, and parsing or understanding of the data being processed is not necessary. For the experiment, the authors use 1.5 TB of raw data that consist of many disk images, RAM snapshots, network captures, and four Universal Serial Bus (USB) devices. The investigation of Windows machines is carried out and three cases are investigated. It takes 180 min to generate the similarity digests. The total triage time for the investigated cases is 110 min and extra 90 min for the additional file hashing to prepare the similarity digests of the targets. As we see, the preparation for the triage takes significantly more time that the triage itself. However, the digest formation can proceed in parallel with the acquisition process. The problem of the method is to have the identified targets as informative as possible, since the method is based on the content resemblance and containment.

Cantrell and Dampier [9] present the implementation of the automated phases in the partially-automated digital triage process model [13]. The implementation is carried out on the basis of series of scripts comprised of original and open source tools written in Perl. The Linux distribution CAINE [49] installed to a USB drive is chosen as the development and testing environment in order to provide some form of boot media and to incorporate full onsite capability. The Windows registry is obtained by using the open source tool RegRipper [50]. The final report is provided in the form of HyperText Markup Language (HTML) pages. The tool is implemented to search the Web browser history for Internet Explorer only. The initial testing is done on a series of 300 GB drives. The runtimes are not provided.

Lim et al. [51] introduce a Live Data Forensic System (LDFS) designed to collect and analyze live data for Microsoft Windows-based systems. The LDFS consist of two separate tools, LDFS collection and LDFS analysis. The LDFS collection system gathers volatile and non-volatile data such as: memory dump, page file, web browser artifacts, instant messaging services clients, Windows Registry, and file system metadata. The distinctive feature of the LDFS collection system is that it can decode encoded BuddyBuddy, Yahoo, and MissLee messenger clients' chat logs. The physical memory dump and dump of all active processes are performed by means of third-party applications. The focus of choosing these applications is based on the least changes to the investigated system by the tool. The XML collection report holds all the collected items with their MD5 and Secure Hash Algorithm 1 (SHA1) hash values. The LDFS collection system is tested against five different types of Windows OSs (Microsoft, Redmond, WA, USA). Several experiments are conducted to test the performance of the system. The largest collection time does not exceed 49 min. The LDFS analysis module has the capabilities for analyzing all the collected data; however, it has not been fully implemented yet. Lim et al. [51] argue that the input data and its trustworthiness are of paramount importance in the live forensics analysis. However, it is not clear whether any defense against the subversion of the collection process is implemented in the LDFS collection system.

Casey et al. [52] discuss the need for and possibilities of honing the digital forensic processes to obtain the timely results. Many tasks in the forensic processes are not resource limited, and rethinking the overall organization of the forensic processes can assure greater improvements than considering the tasks separately. Therefore, improving the complete forensic process is oriented towards two areas, namely, dismantling the barriers between the tasks of the forensic process and providing useful information to support the key decisions. The efforts discussed in this paper focus on processing data from three primary sources: (i) filesystems, (ii) malware, and (iii) network traffic. Many triage tools analyze the filesystems. The analysis reveals that the main bottleneck in this process is the disk

Input/Output (I/O) speeds. Using the results of the analysis, Casey et al. [52] provide the following guidelines for the triage or forensic data extraction tools to improve efficiency:

1. A tool can simultaneously deliver data into multiple extraction operations and create the forensic duplicate
2. A tool can store extracted information in both, the XML format and SQLite database
3. A tool should provide a user-friendly interface to facilitate the viewing, sorting, and classification of files

Additionally, tool developers have to consult about each step of the development with their customers. For the malware, the main suggestion is that the tool should firstly determine whether the file has been seen before. Next, the automatic malware processing tool developed by Defense Cyber Crime Center (DC3) is presented as an illustrative example. However, no suggestions are provided for the network traffic tools. The suite of tools PCAPFAST, developed by DC3, is provided as the example of the right network traffic tool.

Garfinkel [45] extends the research work presented by Garfinkel et al. [53] and introduces a forensic tool bulk_extractor devoted to the initial part of an investigation. The base of the bulk_extractor is the analysis of bulk data. The bulk_extractor scans raw disk images or any data dump for useful patterns (emails, credit card numbers, Internet Protocol (IP) addresses, etc.). It uses multiple scanners tailored to the certain patterns and heuristics to reduce false positive results and noises. The identified patterns are stored in feature files. When processing is complete, the bulk_extractor creates a feature histogram for each feature file. To improve the speed of processing, the bulk_extractor takes advantage of available multi-core capabilities. It detects and decompresses the compressed data. A lot of attention is devoted to the decompression of data. This feature is not usual for triage tools, because it consumes a lot of processing time. However, the feature is very useful for the forensic tool. The performance of the bulk_extractor is compared to the commercial tool EnCase. The results indicate that the bulk_extractor extracts email addresses from the forensic 42 GB disk image 10 times faster than EnCase, and it takes 44 min. The processing time of the bulk_extractor is between 1 and 8 h per piece of media, depending on the size and complexity of the subject data. The processing time does not meet the triage requirements. The bulk_extractor is successfully applied to 250 GB hard disk drives in two real cases. The processing time is 2.5 h for the first case and 2 h for the second. In general, the bulk_extractor is nice-to-have; however, it is not a triage tool.

Koopmans and James [11] introduce an automated network triage (ANT) solution designed for client-server environment. The purpose of the solution is to sort the analyzed systems by their likely relevance to the investigated case. The ANT is developed on the basis of the Preboot eXecution Environment (PXE) protocol and is composed of a network server that runs various services, and the clients, which are the systems to be analyzed, in a physically isolated network. The ANT server boots a suspected computer via a network. The authors provide many technical details that explain the specific steps—what software to use and how to boot the seized computers. The interface is developed in Personal Home Page (PHP) programming language. The data for triage are as follow:

1. A list of keywords to search for
2. A list of preferred file names or extensions
3. A list of preferred directories
4. A hash database that contains the hashes of files of interest
5. A hash database index file

Three real cases of the likelihood that the suspicious computers actually pose threat are very successfully investigated; the runtimes of the three cases are within 10 min. The runtimes are very short, however, it is not clear why they are so short, and an explanation is not provided. Moreover, Horsman et al. [10] state that hashing and keyword searching approaches can limit the effectiveness of digital triage because they are too restrictive. The limitations of the ANT solution are the following: there is no possibility to boot from the external source and encrypted data could not be analysed.

Moser and Cohen [1] discuss the use of triage in quite a different context than the traditional criminal case investigation—an incident response. The authors consider the use of the GRR Rapid

Response (GRR) system. It is an agent-based open source distributed enterprise forensics system. Moser and Cohen [1] overview the components of the GRR system. A more detailed description of the GRR system is available elsewhere [54]. This method lowers the total time cost of triage analysis by distributing this task to the system agents. The main attention is directed towards the reliability of agents. Constant monitoring of used resources—memory and central processing unit (CPU)—ensures the reliability of agents. The investigation consists of three phases: planning, collection, and analysis. The experiment is carried out on many corporate workstations and laptops. The GRR agents are installed on these computers. The goal of the experiment is to examine the representative cases of a typical enterprise investigation performed by an incident response team. Four cases are analyzed. The majority of agents pick up artifacts in the first few minutes after the start. Nevertheless, the GRR continues running to 24 h so, if the missing machines come back online later, the artifacts will still be detected. The case of the autorun key comparison required an extensive manual analysis, therefore, improvement is necessary for such cases.

Shaw and Browne [15] argue that a digital forensic triage has been conducted on an informal basis for several years. The authors introduce the concepts of administrative and technical triage. The administrative triage assesses the circumstances of a new case before starting an examination of the evidence. Shaw and Browne [15] discuss and summarize the weaknesses of digital triage. The enhanced previewing is suggested as an alternative to digital triage. The Linux forensic distribution CAINE [49] installed on a compact disc (CD) is chosen as a base for the implementation. The bootable CD is remastered to include the existing open source forensic tools and to add new analysis software. A high-level overview of system work is presented. The possibilities to deploy the enhanced previewing in the digital forensic laboratory are analyzed. The weaknesses of the enhanced previewing are as follows: the case management becomes more complicated and the system is not suitable to the field use at all. The authors doubt "whether the Enhanced Previewing process is a subset of technical triage or whether it is a distinct process only loosely related to technical triage". We are inclined to state that the enhanced previewing is not a subset of technical triage, because the processing time of the enhanced previewing would be quite long. We base our conclusion on the provided description of the system.

Shiaeles et al. [55] review three open source triage tools and suggest the ways to improve them. The TriageIR, TR3Secure, and Kludge tools are tested for various Microsoft Windows versions. There is currently no mature framework for practically testing and evaluating triage tools, however, the authors do not suggest a framework and evaluate the tools in their best way imagined. The first principle to assess is the access to volatile data. The next principle to assess is the adherence of tools to forensic principles ensuring the admissibility of the collected evidence to the court. An experiment shows that no single considered tool is better than others. All the tools have their strengths and weaknesses. The solution is to preferably have several tools and maintain a profile of the tool capabilities. The recommendations for improving the tools are as follows:

1. The tools should be made more adaptable, either dynamically or manually
2. Disabling Prefetch on Windows systems will result to less system alterations
3. The tools should record and undo all registry changes, which they perform to the examined system
4. The tools should collect the Internet activity artifacts that belong to all known browsers

Woods et al. [56] present an open source software for automated analysis and visualization of disk images created as part of the BitCurator project [57]. The goal of the presented software is to assist in triage tasks. The data for analysis is obtained from open source forensic tools fiwalk [58] and bulk_extractor [45]. The fiwalk tool recognizes and interprets the content of filesystems that are contained in disk images, and produces an XML report. The bulk_extractor tool reads the raw contents of the disk image and reports on various features. The BitCurator reporting tools produce Portable Document Format **(**PDF) reports on filesystem and for each feature separately. If data entry datasets are large, it is possible to configure the reporting tools to produce the report for a subset of the filesystem or a subset of features. The time required to manage a given disk image with forensic tools fiwalk and bulk_extractor is within the range of tens of minutes. The limiting factor in terms of

time is the BitCurator reporting tools that may have to process an extremely large XML filesystem report and text feature reports. The BitCurator project freely distributes these reporting tools in a variety of ways for the practitioners and researchers to use.

Baggili et al. [59] present a five-phase, multi-threaded bootable tool Forensics2020 for forensics triage. The tool is loaded from a bootable Windows Pre-installation Environment using a USB stick. Phases proceed in sequence, however, while the tool is working, the examiner can interact with the tool to see the results up to that point and to request certain types of data. The first phase collects logical files and their metadata. The second phase analyses every image for the Exchangeable Image File Format (EXIF) data. The third phase explores and classifies each file based on its header. The fourth phase parses executable files for audit and threat purposes. The fifth phase hashes each file and takes the longest time of all the phases. The experiment is carried out to assess the efficacy and a forensic soundness of Forensics2020. In sum, 26.33 TB of data from 57 computers are analyzed. The total time required to complete the process is 10,356 s. The tool makes certain changes to the hard drive; however, the changes are greater in number than those of similar Linux-based tools. Two lessons can be learned from the development of Forensics2020. Firstly, a multi-threaded, multi-stage tool allows the examiner to interact with the evidence while the system is performing the forensics processing. Secondly, the mounting of the hard drive by a bootable tool has influence over the perception of the forensic soundness.

Haggerty et al. [60] propose an approach to automate the visualization of quantitative and qualitative email data to assist the triage of digital evidence during a forensics investigation. The quantitative information, which is retrieved from the email, refers to the network events and actor relationships. The qualitative information refers to the body of the emails themselves. The authors have developed a TagSNet software to implement the proposed approach. The software provides two views—a network of the actors and a tag of keywords that are found in the email bodies. Both views are interactive in that the forensics examiner may move the actors and text around. The experiment is carried out on the Enron email data. The average time to process and visualize email data is about 10 min. However, the visualization is not aimed at answering the investigative questions; it only aids the forensics examiner to triage email data more quickly than in the manual mode.

Vidas et al. [61] describe a free forensic tool, OpenLV, which can be deployed in the field and in the laboratory. It is noteworthy that over the past years it has been used under the name of "LiveView". The interface of the tool is oriented to the examiners with little training. The OpenLV asks for configuration and creates a virtual machine out of a forensic image or physical disk. The virtual machine enables booting up the image and gains an interactive environment without modifying the underlying image. The tool natively supports only the dd/raw image format. Other formats require third party software that can be integrated into the tool, which is Windows centric, and a limited Linux support is added. Additionally, the OpenLV aids to remove the barrier of passwords for Windows users. The authors claim that "OpenLV aims to meet the demand for an easy-to-use triage tool", however, neither an example nor a reference is provided for how OpenLV is used for triage purposes.

Conway et al. [62] discuss a development of a Virtual Crime Scene Simulator (VCSS) that can perform a live triage of digital devices. Training is important for the law enforcement officers; therefore, the tool will have a field of its application. The VCSS is an open source project, and it is implemented as game playing, where Unity3D [63] is chosen as the base platform. The virtual environment includes a three-dimensional (3D) representation of a house with four rooms, a hallway, and outside scenery. The crime scenery has a set of the following items: furniture, various hardware devices, and an avatar for interrogation. The following in-game actions are possible: live examination of the various digital devices, interrogation of the avatar, and other actions related to the crime scene. The full device interaction is implemented on Windows version only. The trainer can add new logic by modifying the existing JavaScript. The law enforcement officers from a developing country used the VCSS for training. The participants highly evaluated the educational purpose of the application.

Hegarty and Haggerty [64] present the SlackStick approach to identify files of interest for forensic examiner on the live system. The approach is based on the signatures of the files. To create

the signature of the file, a block within the original file is chosen, which may be from anywhere within a file, except for the first and the last blocks. Several predetermined bytes are chosen to represent the file. The number of bytes can be chosen by balancing the tradeoff between the false positives and false negatives. The higher number of bytes decreases the likelihood of false positives. The SlackStick software written in Python under Slax operating system (Software Manufacturer, City, State, Country). runs from an external device. SlackStick reads the memory blocks on the target machine sequentially to generate block signatures for comparison with the signature library. If a match is found, a report that includes the matched signature and the physical location of the file in the storage media is generated. They conducted an experiment in which it took a dozen of seconds to analyze 1 GB partition that has 2 194 JPEG images. Signatures are generated by selecting 11 bytes within the second block of each target files. Neither false positives nor false negatives are found. As the number of signatures increases, no measurable impact on performance is observed.

Further, van Beek et al. [66] introduce a development of the distributed digital forensic system HANSKEN [67] that is the successor of the operating digital forensic system XIRAF [68]. The goal of HANSKEN is to speed-up the computations of big data. The three forensic drivers for the system are as follows: minimization of the case lead time, maximization of the trace coverage, and specialization of the people involved. These drivers justify the building of the distributed big data forensic platform. To mitigate the threats associated with a big data platform, the development of the system HANSKEN is based on eight design principles. They are enumerated in the order of the priority: 1. Security, 2. Privacy, 3. Transparency, 4. Multi tenancy, 5. Future proof, 6. Data retention, 7. Reliability, and 8. High availability. The first three principles are sociological; meanwhile the other five are business principles and define the system boundaries. The system uses its own forensic image format. The authors justify the need for its own format; however it could be the limitation of the system, especially for the future development. The system HANSKEN stores the data compressed and encrypted. The encryption of data ensures a restricted access to it. The process of extracting data from a forensic image starts as soon as the first bits of the image are uploaded to the system. Such approach acknowledges the right organization of the forensic processes to improve the efficiency of the forensic investigation. The authors admit that triage is a valuable approach for ordering the processing of images, not for leaving images unprocessed. Such form of triage is planned to be included into the system HANSKEN. The system is implemented on the Hadoop realization of MapReduce. The system HANSKEN was planned to be put into production at the end of the year 2015.

Koven et al. [69] further explore and develop the idea of email data visualization [60]. The authors present a visual email search tool InVEST. Firstly, the tool preprocesses the email data to create indexes for various email fields. The duplicate information and junk data are excluded from indexing. Next, the user starts the search process with defined keywords. The search results are presented in five different visual views. The visual views enable better understanding and interpreting of the search results as well as finding the relationships between the search entities. The diverse views show different relationships between search entities and present the contextual information found within these results. All the views support the possibility to refine the search results using filtering and expanding. The process of filtering and expanding is iterative until the search is successful. An experiment is carried out on the Enron email data set. Two case studies are successfully investigated. Koven et al. [69] used the term "triage" in the title of the paper. The term "triage" is used in the sense of a tool, which allows selecting a subset of the emails that are related to a particular subject from the whole email set. However, the time spent to select can be quite long. The process of selecting the subset of the email is interactive heavily involving the user. The authors present an example that "the time to make the discovery and exploration including the skimming of at least 30 of the discovered emails was approximately 1 h". Therefore, the use of the tool in triage process is quite unlikely, unless the data captured is only in form of email.

## 6. Lessons Learned from the Review

To summarize the field of live triage, the noteworthy research focusses are as follows:

1. The stress of a real-time computation problem having allotted limited time and resources for triage, presented by Roussev et al. [7]. The idea is that an increase in the performance can be achieved if acquisition and processing start and complete at almost same time. The implementation of the forensic system HANSKEN [66] proves the appropriateness of the presented idea
2. The selective imaging approaches to reduce data volume, presented by Grier and Richard III [22] and Quick and Choo [31,32]. The difference between the approaches is in selecting the regions that have a forensic value. Grier and Richard III [22] state that the profiles must be created and stored in a library. Moreover, Quick and Choo [32] suggest the idea of thumbnailing video, movie, and picture files
3. The introduction of triage template pipelines into the investigative process for the most popular types of digital crimes, presented by Overill et al. [20]. However, the authors do not enumerate these types of crimes and provide only the DDoS and P2P template diagrams without the discussion of the details
4. The artificial intelligence approaches presented by Turnbull and Radhava [26] and Peersman et al. [29]. Turnbull and Randhawa [26] describe an approach to assist a less technically intrinsic user to run a triage tool. Peersman et al. [29] present an approach to automatically label new child sexual abuse media

To summarize the field of post-mortem triage, the noteworthy research focusses are as follows:

1. Storing and using the knowledge of the past cases, presented by Horsman et al. [10,37] and Bashir and Khan [39]
2. The use of machine learning techniques, presented by Marturana and Taconi [33-35], McCleland and Marturana [36], and Fahdi et al. [41]. The trend is promising because such techniques are indeed valuable in many research areas; however, the presented research works are immature

To summarize the field of triage of mobile devices, the noteworthy research achievement is only single one:

1. The information recovery engine DEC0DE, offered by Walls et al. [43] and the information prioritization system LIFTR, which uses the data obtained from DEC0DE, offered by Varna et al. [44]

To summarize the field of triage tools, the noteworthy research achievements are as follows:

1. The method of similarity digests, offered by Roussev and Quates [12]
2. The online GRR Rapid Response system used for incident response, offered by Moser and Cohen [1]
3. The multi-threaded bootable tool Forensic2020, which allows interaction of the examiner, while the tool is processing data, offered by Baggili et al. [59]
4. The visualization of email data offered by Haggerty et al. [60]. Koven et al. [69] presented an approach of email data visualization, as well. However, the provided runtimes are quite long and, therefore, the tool is not suitable for triage purposes
5. The SlackStick approach to identify the files of interest, when several predetermined bytes are chosen to represent the file, offered by Hegarty and Haggerty [64]
6. The distributed digital forensic system HANSKEN that works on a big data platform, offered by van Beek et al. [66].

## 7. Conclusions and Future Directions

The evolution of modern digital devices is outpacing the scalability and effectiveness of the digital forensic techniques. Digital triage is one of the solutions to this problem, as it can extract intelligence quickly at the crime scene and provides valuable information to the forensic examiner. This form of triage is known as the live triage. In a similar way, such methodology can be used in a laboratory to prioritize the analysis of digital media and to alleviate the examination backlog. This

form of triage is known as the post-mortem triage. The term "forensic" should be used carefully with digital triage, because the process of digital triage does not always adhere to the rules of the forensic process. Moreover, the legitimacy of the process depends on the jurisdiction system of a specific country. Therefore, the digital triage model must be adjusted individually according to the legal system of a specific country. Live triage raises important legal concerns. Sometimes, the process of digital triage in the forensic context is an admission of failure, since important evidence can be overlooked. However, a better approach does not exist for today. To solve this problem, we have to consider digital triage as a technical process that provides information for the forensic examination and does not involve the evaluation of digital evidence.

To increase the performance of digital triage when data cloning procedure is involved, data cloning and data processing should start and complete almost at the same time. Such approach is advanced and it is applied in some forensic tools (for example, the digital forensic system XIRAF). Moreover, it means that data processing should be as fast as data cloning.

The number of mobile devices is increasing quite rapidly. Digital triage of mobile devices is harder than that of desktop computers. It also appears that due to the nature of mobile devices, many forensics procedures must inevitably involve live forensics as the device needs to be powered on. To address this problem, more effective methods for live triage and tools for mobile devices are necessary.

High-level training is required from the examiners on the field. It would be cost-effective to hire a less technically fluent specialist for such job. One possibility to accomplish this is to alter the software to make it friendlier to less skilled technical examiners or to construct specific tools that can incorporate the expert knowledge. The core concept of this approach is that the expert systems can locate and interpret the low-level computing artefacts and provide higher-level concepts. Reducing the need to locate and interpret the low-level artefacts is a still lesser-explored method for conducting digital triage.

Another direction of the research in digital triage could be incorporating intelligent technologies, i.e., techniques from artificial intelligence, computational modelling, and/or social network analysis. Almost no research works of digital triage that are directed at the vast area of the social networks are presented. Furthermore, a new developing area of the Internet of Things is left almost without attention as well.

Large volumes of data are available for forensic examination. Since every desktop computer has several processors, available resources can be applied to speed-up the computations. The methods of parallel processing are already applied; however, only in small quantities so far.

Moreover, a challenging research direction is awaiting the researchers' attention. More precisely, it is the implementation of the algorithms into hardware, because the hardware inherently performs the computations faster than the software.

**Acknowledgments:** We express gratitude to anonymous reviewers for the valuable comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Moser, A.; Cohen, M.I. Hunting in the enterprise: Forensic triage and incident response. *Digit. Investig.* **2013**, *10*, 89–98.
2. Rogers, M.K.; Goldman, J.; Mislan, R.; Wedge, T.; Debrota, S. Computer forensics field triage process model. *J. Digit. Forensic Secur. Law* **2006**, *1*, 27–40.
3. Casey, E.; Ferraro, M.; Nguyen, L. Investigation delayed is justice denied: proposals for expediting forensic examinations of digital evidence. *J. Forensic Sci.* **2009**, *54*, 1353–1364.
4. Casey, E. Triage in digital forensics. *Digit. Investig.* **2013**, *10*, 85–86.
5. Casey, E. Differentiating the phases of digital investigations. *Digit. Investig.* **2016**, *19*, A1–A3.
6. Venčkauskas, A.; Jusas, V.; Paulikas, K.; Toldinas, J. A methodology and tool for investigation of artifacts left by the BitTorrent client. *Symmetry* **2016**, *8*, 40.
7. Roussev, V.; Quates, C.; Martell, R. Real-time digital forensics and triage. *Digit. Investig.* **2013**, *10*, 158–167.

8. Montasari, R.A. Formal Two Stage Triage Process Model (FTSTPM) for Digital Forensic Practice. *Int. J. Comput. Sci. Secur.* **2016**, *10*, 69–87.
9. Cantrell, G.; Dampier, D.A. Implementing the automated phases of the partially-automated digital triage process model. *J. Digit. Forensics Secur. Law* **2012**, *7*, 99–116.
10. Horsman, G.; Laing, C.; Vickers, P. A case-based reasoning method for locating evidence during digital forensic device triage. *Decis. Support Syst.* **2014**, *61*, 69–78.
11. Koopmans, M.B.; James, J.I. Automated network triage. *Digit. Investig.* **2013**, *10*, 129–137.
12. Roussev, V.; Quates, C. Content triage with similarity digests: The M57 case study. *Digit. Investig.* **2012**, *9*, S60–S68.
13. Cantrell, G.; Dampier, D.; Dandass, Y.S.; Niu, N.; Bogen, C. Research toward a partially-automated, and crime specific digital triage process model. *Comput. Inf. Sci.* **2012**, *5*, 29–38.
14. Hong, I.; Yu, H.; Lee, S.; Lee, K. A new triage model conforming to the needs of selective search and seizure of electronic evidence. *Digit. Investig.* **2013**, *10*, 175–192.
15. Shaw, A.; Browne, A. A practical and robust approach to coping with large volumes of data submitted for digital forensic examination. *Digit. Investig.* **2013**, *10*, 116–128.
16. Pollitt, M.M. Triage: A practical solution or admission of failure. *Digit. Investig.* **2013**, *10*, 87–88.
17. Lopez-Rojas, E.A.; Axelsson, S. Using the RetSim fraud simulation tool to set thresholds for triage of retail fraud. In *Lecture Notes in Computer Science*, *Proceedings of the 20th Nordic Conference on Secure IT Systems*, *Stockholm*, *Sweden*, *19–21 October 2015*; Springer: Cham, Switzerland 2015; Volume 9417, pp. 156–171.
18. Perumal, S.; Norwawi, N.M.; Raman, V. Internet of Things (IoT) digital forensic investigation model: Top-down forensic approach methodology. In Proceedings of the Fifth International Conference on Digital Information Processing and Communications (ICDIPC), Sierre, Switzerland, 7–9 October 2015; pp. 19–23.
19. Auberson, M.; Baechler, S.; Zasso, M.; Genessay, T.; Patiny, L.; Esseiva, P. Development of a systematic computer vision-based method to analyse and compare images of false identity documents for forensic intelligence purposes—Part I: Acquisition, calibration and validation issues. *Forensic Sci. Int.* **2016**, *260*, 74–84.
20. Overill, R.E.; Silomon, J.A.M.; Roscoe, K.A. Triage template pipelines in digital forensic investigations. *Digit. Investig.* **2013**, *10*, 168–174.
21. Lim, K.-S.; Lee, C. A framework for unified digital evidence management in security convergence. *Electron. Commer. Res.* **2013**, *13*, 379–398.
22. Grier, J.; Richard III, G.G. Rapid forensic imaging of large disks with sifting collectors. *Digit. Investig.* **2015**, *14*, S34–S44.
23. Garfinkel, S.; Malan D.; Dubec K-A.; Stevens C.; Pham C. Advanced forensic format: an open extensible format for disk imaging. In *Proceedings of 2nd International Conference on Digital Forensics, Orlando, FL, January 29-February 01 2006*, Advances in Digital Forensics II. IFIP Advances in Information and Communication, vol. 222. Springer, Boston, MA, 2006, pp.13-27.
24. The Sleuth Kit. Autopsy. Available online: http://www.sleuthkit.org/autopsy/ (accessed on 25 March 2017).
25. Penrose, P.; Buchanan, W.J.; Macfarlane, R. Fast contraband detection in large capacity disk drives. *Digit. Investig.* **2015**, *12*, S22–S29.
26. Turnbull, B.; Randhawa, S. Automated event and social network extraction from digital evidence sources with ontological mapping. *Digit. Investig.* **2015**, *13*, 94–106.
27. Hitchcock, B.; Le-Khac, N.-A.; Scanlon, M. Tiered forensic methodology model for Digital Field Triage by non-digital evidence specialists. *Digit. Investig.* **2016**, *16*, S75–S85.
28. Leimich, P.; Harrison, J.; Buchanan, W.J. A RAM triage methodology for Hadoop HDFS forensics. *Digit. Investig.* **2016**, *18*, 96–109.
29. Peersman, C.; Schulze, C.; Rashid, A.; Brennan, M.; Fischer, C. iCOP: Live forensics to reveal previously unknown criminal media on P2P networks. *Digit. Investig.* **2016**, *18*, 50–64.
30. iCOP. iCOP Toolkit. Available online: http://scc-sentinel.lancs.ac.uk/icop/?q=content/icop-toolkit (accessed on 25 March 2017).
31. Quick, D.; Choo, K.-K.R. Big forensic data reduction: Digital forensic images and electronic evidence. *Clust. Comput. J. Netw. Softw. Tools Appl.* **2016**, *19*, 723–740.
32. Quick, D.; Choo, K.-K.R. Data reduction and data mining framework for digital forensic evidence: Storage, intelligence, review and archive. *Trends Issues Crim. Crim. Justice* **2014**, *480*, 1–11.

33. Marturana, F.; Tacconi, S. A machine learning-based triage methodology for automated categorization of digital media. *Digit. Investig.* **2013**, *10*, 193–204.
34. Marturana, F.; Me, G.; Berte, R.; Tacconi, S. A quantitative approach to triaging in mobile forensics. In Proceedings of the International Joint Conference of IEEE TrustCom-11, Changsha, China, 16–18 November 2011; pp. 582–588.
35. Marturana, F.; Bertè, R.; Tacconi, S.; Me, G. Triage-based automated analysis of evidence in court cases of copyright infringement. In Proceedings of the First IEEE International Workshop on Security and Forensics in Communication Systems (SFCS 2012), Ottawa, ON, Canada, 14–15 June 2012; pp. 6668–6672.
36. McClelland, D.; Marturana, F. A Digital Forensics Triage methodology based on feature manipulation techniques. In Proceedings of the International Conference on Communications, Sydney, Australia, 10–14 June 2014; pp. 676–681.
37. Horsman, G.; Laing, C.; Vickers, P. A case based reasoning framework for improving the trustworthiness of digital forensic investigations. In Proceedings of the 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom-2012), Liverpool, UK, 25–27 June 2012; pp. 682–689.
38. EnCase Portable. Available online: http://www.guidancesoftware.com/encase-portable (accessed on 25 March 2017).
39. Bashir, M.S.; Khan, M.N. A triage framework for digital forensics. *Comput. Fraud Secur.* **2015**, *2015*, 8–18.
40. Dalins, J.; Wilson, C.; Carman, M. Monte-Carlo filesystem search: A crawl strategy for digital forensics. *Digit. Investig.* **2015**, *13*, 58–71.
41. Al Fahdi, M.; Clarke, N.L.; Li, F.; Furnell, S.M. A suspect-oriented intelligent and automated computer forensic analysis. *Digit. Investig.* **2016**, *18*, 65–76.
42. Mislan, R.P.; Casey, E.; Kessler, G.C. The growing need for on-scene triage of mobile devices. *Digit. Investig.* **2010**, *6*, 112–124.
43. Walls, R.J.; Learned-Miller, E.; Levine, B.N. Forensic triage for mobile phones with DEC0DE. In Proceedings of the 20th USENIX Conference on Security, San Francisco, CA, USA, 8–12 August 2011; pp. 1–14.
44. Varma, S.; Walls, R.J.; Lynn, B.; Levine, B.N. Efficient smart phone forensics based on relevance feedback. In Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices (SPSM '14), Scottsdale, AZ, USA, 3–7 November 2014; pp. 81–91.
45. Garfinkel, S.L. Digital media triage with bulk data analysis and bulk_extractor. *Comput. Secur.* **2013**, *32*, 56–72.
46. Yaffs. Available online: http://www.yaffs.net/ (accessed on 25 March 2017)
47. Guido, M.; Buttner, J.; Grover, J. Rapid differential forensic imaging of mobile devices. *Digit. Investig.* **2016**, *18*, S46–S54.
48. sdhash home. Available online: http://roussev.net/sdhash/sdhash.html (accessed on 25 March 2017).
49. CAINE. Computer Forensics Linux Live Distro. Available online: http://www.caine-live.net/ (accessed on 27 March 2017).
50. KALI Tools. RegRipper. Available online: http://tools.kali.org/forensics/regripper (accessed on 25 March 2017)
51. Lim, K.-S.; Savoldi, A.; Lee, C.; Lee, S. On-the-spot digital investigation by means of LDFS: Live Data Forensic System. *Math. Comput. Model.* **2012**, *55*, 223–240.
52. Casey, E.; Katz, G.; Lewthwaite, J. Honing digital forensic processes. *Digit. Investig.* **2013**, *10*, 138–147.
53. Garfinkel, S.; Nelson, A.; White, D.; Roussev, V. Using purpose-built functions and block hashes to enable small block and sub-file forensics. *Digit. Investig.* **2010**, *7*, S13–S23.
54. Cohen, M.; Bilby, D.; Caronni, G. Distributed forensics and incident response in the enterprise. *Digit. Investig.* **2011**, *8*, S101–S110.
55. Shiaeles, S.; Chryssanthou, A.; Katos, V. On-scene triage open source forensic tool chests: Are they effective? *Digit. Investig.* **2013**, *10*, 99–115.
56. Woods, K.; Lee, C.A; Misra, S. Automated analysis and visualization of disk images and file systems for preservation. In Proceedings of the 10th IS & T Archiving Conference, Washington, DC, USA, 2–5 April 2013; pp. 239–244.
57. BitCurator. Available online: https://www.bitcurator.net/ (accessed on 25 March 2017).
58. Fiwalk. Available online: http://www.forensicswiki.org/wiki/Fiwalk (accessed on 25 March 2017).

59. Baggili, I.; Marrington, A.; Jafar, Y. Performance of a logical, five-phase, multithreaded, bootable triage tool. Advances in Digital Forensics X. In *IFIP Advances in Information and Communication Technology*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 279–295.
60. Haggerty, J.; Haggerty, S.; Taylor, M. Forensic triage of email network narratives through visualisation. *Inf. Manag. Comput. Secur.* **2014**, *22*, 358–370.
61. Vidas, T.; Kaplan, B.; Geiger, M. OpenLV: Empowering investigators and first-responders in the digital forensics process. *Digit. Investig.* **2014**, *11*, S45–S53.
62. Conway, A.; James, J.I.; Gladyshev, P. Development and Initial User Evaluation of a Virtual Crime Scene Simulator Including Digital Evidence. In *Lecture Notes of the Institute for Computer Sciences Social Informatics and Telecommunications Engineering*, *Proceedings of the 7-th International ICST Conference on Digital Forensics and Cyber Crime* (*ICDF2C*), *Seoul*, *South Korea, October 6–8, 2015*; Springer: Cham, Switzerland 2015; Volume 157, pp. 16–26.
63. Unity. Available online: https://unity3d.com/ (Accessed on 27 March 2017).
64. Hegarty, R.; Haggerty, J. SlackStick: Signature-based file identification for live digital forensics examinations. In Proceedings of 2015 European Intelligence and Security Informatics Conference, Manchester, UK, 7–9 September 2015; pp. 24–29.
65. Slax Linux. Available online: https://www.slax.org/ (Accessed on 27 March 2017)
66. Van Beek, H.M.A.; van Eijk, E.J.; van Baar, R.B.; Ugen, M.; Bodde, J.N.C.; Siemelink, A.J. Digital forensics as a service: Game on. *Digit. Investig.* **2015**, *15*, 20–38.
67. Hansken. Available online: https://www.forensicinstitute.nl/products_and_services/forensic_products/hansken.aspx (Accsessed on 27 March 2017).
68. Bhoedjang R. A. F.; van Ballegooij A.R.; van Beek H. M. A., van Schie J.C.; Dillema F.W.; van Baar R.B.; Ouwendijk F.A; Streppel M. Engineering an online computer forensic service. *Digit. Investig.* **2012**, *9*, 96-108.
69. Koven, J.; Bertini, E.; Dubois, L.; Memon, N. InVEST: Intelligent visual email search and triage. *Digit. Investig.* **2016**, *18*, S138–S148.