A Crime Depended Automated Search and Engine for Digital Forensics

J.P. Fizaine and N.L. Clarke

Centre for Security, Communications and Network Research,
Plymouth University, Plymouth, UK
e-mail: info@cscan.org

Abstract

As the number of cybercrime is rising up, time and accuracy in digital forensic become more complex to handle. This phenomena is due to the large amount of data to be investigate hence requires more time to analyse. In the meantime the technology is more and more complex and it more skills are necessary to be able to investigate such complex system. Existing tools like EnCase or DataAccess could tackle the complexity of new technologies by adding scripts which also allow some automation. But those are not built to face large amount of information in restricted amount of time. During the investigation there is always the same tasks to do. In addition each crime has a pattern. So common pattern where found for the same type of crime. The idea is to build a module that take advantage from the common pattern to automate the analysing process of evidence. This fact would help in model as a response to automate the investigation process. This paper aims to present a simple concept of file identification determined by a crime profile.

Keyword

Automated tool, automated investigation process, automated forensic tool, digital forensics, digital formalization formation, automated forensic model, file carving, data hiding.

1 Introduction

Digital forensic, a part field of forensic science, is specialised in crimes investigation where computers, electronic devices and computer networks are involved in. In response to the increasing threats of crime involving computers and the complexity of information technology, appropriated software are being developed such as Encase (http://guidancesoftware.com/) or FTK(http://accessdata.com/products/forensic-investigation/ftk).

Computers are now involved in various nature of crime such as child pornography, fraud, identity theft, as well as there are weapons of crimes. Spreading computer virus, hacking computers and networks to gained unauthorized access, performing deny of service are some other examples. The U.S Justice Office of Justice Programs (2008) had categorised each crime based on involved information. Where most of researches work about new models and frameworks which is detailed in section 2 they focus on the current approach used by FTK and Encase.

Those tools are widely used in industries, governments, agencies as well for law enforcement. There are highly specialised tools that require experiences and solid knowledge in operating systems and networks. Therefore they can only be used by technical and skilled investigators. Still they are efficient tools with large panel of features required in digital investigation. Their approach is to let the investigator officer manages his own investigation process. As a example Encase has it own script engine to give the opportunity to automated the most common task.

During the first step of the investigation, the investigator is aware of the suspected crime and so he must look for specific information which are liable for the crime or to dismiss the accusation. Lets take as an example a case of child pornography. He will be looking for Images and websites. However in case of financial fraud he would be instead looking for bank account number, financial documents, etc. If it the case of unauthorized access as computer intrusion, program, logs, would then be the information he would look for. At the end of the first step of investigation he will performed a deep analysis of what he found, to prove are not the crime.

This first step of collecting useful information can be automated because each crime is defined differently as regarding the information it involves. Such a process is not performed by Encase and FTK or even other forensic tools. Even if the research in digital forensic is active, must of them focus on models and frameworks to analyse the data more efficiently. Some attempt to automate the investigation process is also part of the research area.

This paper is about a new approach in digital forensic. The tool fits in the first step of investigation. Its purpose is to perform an automated process for collecting information related specifically to one crime. And to prepare this set of information for further deep analysis by appropriated program. Therefore the global architecture and general process would be detailed.

The section 2 regroups relevant models, frameworks and tools and explains why there are interesting. The next section 2.2 quickly discuss about related work and background. Base and requirement are detailed and explained in section 3. Then the novel approach is detailed in the following section 4. The paper continue through the section 5 where it explains the extraction process in more in detailed and with in section 6 the heart of the core with the formalization of the triage. At last future work and improvements are introduced in section 7.

2 Survey on forensic tools, frameworks and models

Novel frameworks, new tools and models are hot topics in computer forensic science. Several sort of technologies and different approaches had being introduced but the community agreed on the necessity to have a suitable tool for an appropriate use. This section first presents previous researches and works that explain necessities and need in computer forensic. It is followed by a second section where a background is given.

2.1 Needs in digital computer forensics

The fist attempt to automated forensic is found 2004 with the work from Slay et al. (2004) and Gladyshev & Patek (2004). The work from Slay et al. (2004) expresses the necessity of developing new tools. Despite the fact that the work is about Australian usage and requirement, they argue that the most widely used software, Encase (from Guidance software system) have an expensive cost and it requires skilled investigators. They also claim, the tool is not available for police officer. But Gladyshev & Patek (2004) had another approach based on finite state machine. Their work consist of automate the reconstruction process whereas Slay et al. (2004) really focus on a whole automated model and tool.

Peisert et al. (2007) has expressed some principles and qualities a good forensic model should have. The goal of their work is to build a rigorous approach based on attack profiles formalise by graph. Practical forensic is based on logged data from various tool which do not constitute a rigorous model. The authors had highlighted five principles:

- The whole system must be considered;
- To log as much as possible without considering the attack or the failure;
- The effect of events must be considered and not only the action;
- The interpretability and the understandability of events must suit the context.
- Work on conditions from before and from after the event.
- The event must be presented in a way, they can be analysed and be understandable by the forensic investigator.

The authors also spotted other requirements, forensic tool must have. The data should be logged from different layer of abstraction. And a limit must be set to prevent collecting to much data.

Andrew (2007) had establish in his work a model to perform rigorous analysis of digital devices and media storage. His paper focuses on requirements to perform such analysis. The basis of his model are the principle of consistent result and the principle of static storage. Forensic software are put on the top. followed by the concept of *individualization* from Dr Paul Kirk and *identification*.

2.2 Background and related work

EnCase from guidance software, is the most spread forensic tool. Garber (2001) in his case of study about the tool concluded that it is a tool for who know what he is doing. It is a very complete tool. Even if it as a script engine to give the capability to used automated process, the tool is still for skilled people. This point is argued by Slay et al. (2004) and they claimed the fact that it is still an expensive solution reserved to forensic laboratory.

Since a few years, research on model, tool and framework is a hot topic and significant work as been published. Slay et al. (2004) has argued the need of a more

simple tool then Encase to give the ability to perform forensic investigation in the industries Marrington et al. (2010) expressed in there work the necessity of an automated tool to handle the quantity problem and the complexity problem developed by Carrier (2003a).

Most of the new models focus on the analysing aspect as Bhat et al. (2010) and Marrington (2010). Bhat et al. had developed a new approach based on data mining but Marrington based his work on an accurate framework. Bhat et al. had a non-forensic approach and a non-rigorous approach as Hunton (2011a) defined it. And the work of Marrington has a general and rigorous approach with significant features. Hunton (2011b) has noticed the existence of technical challenged and a gap between examiner and non-technical people. As a response he developed a model composed of several layer to process cybercrime investigation.

As the response to give the ability to perform forensic investigation, with the lowest cost, by non-technical and non-skilled people and for the world of industry, the challenge is to develop a tool with automated process investigation. Such a process first have an extraction process and make the extracted data available for further deep analysis by appropriated tools. The automated process is based on the fact that cybercrime can be ordered in a taxonomy Casey (2004).

The tool would be based on a rigorous approach with a scientific approach and based on science forensic concepts. Therefore evidences which are exposed in court can't be argued. Works from Andrew (2007), Carrier (2003a), Marrington et al. (2010) and Hunton (2011a) would be the foundation of a such rigorous approach.

3 Requirements Analysis

The research consist of developing a new forensic tool. Therefore the tool must fit some requirements and it must follow a rigorous approach. The tool would fit the general investigator process described by Casey (2004). The investigation process starts with acquisition of data, a process to prevent the integrity of data to change or be altered. It consists of make an exact duplication of the digital media which results in a file called image. It is simply a file contained raw data from digital media. The image is the basis to perform an investigation.

In order to perform deep analysis by other tools, data must be collected and stored in a secure location. A database is the best solution. Because it has security feature and can handle a huge amount of data. Every software with a database interface and the specification of the table can access easily to data. The design of the storage must take in consideration the concept of individualization from Dr. Paul Kirk Andrew (2007). Individualization is an important concept because it is the basis of re-construction. To perform the re-construction process, relevant information must be collected and extracted. The quantity problem orders to take care during the process of extraction, but there must be enough information to allow the re-construction.

Forensic science orders to analyse deeply the evidence. In our case the evidence is a digital media storage which stores files in organised structured. Such a concept is called file system, a feature of a modern operating system. But in order to solve the

quantity and the complexity problem from Carrier (2003b), the tool is based on layer of abstracted. This approach would cover only files from the file system and thus introduces errors. Slack space analysing and data carving are important feature as well. Some old deleted or hidden file can be find in slack space. It is hence important to detect their location within the evidence in order to apply data carving techniques to extract deleted file. Detecting and extracting hidden file is a far more complex problem.

Sometime criminals can be very skilled in technology. They are aware of anti-forensic techniques. They would use cryptography, steganography techniques and even more advanced data hiding techniques. The tool would consider possibility with in two points. A profile of the suspect would give an average idea of its skills. This profile would be a measurement of the skill. The value is calculated by analysing all the software found in the evidence.

In all the possible crimes involving computers, networks, or electronic devices, the investigator should be able to perform keyword research. This feature is based on engine that first collect all text pattern. This is an indexing process of words found in the devices. And a list of keyword is given by the investigation. Such keyword could be names, phone number, ip address.

A preliminary report would give a summary of what was found. The investigator knows hence which files request it attention. It would be on those flagged files that he must performs the deep analysis. Those new results would be summarized in final report. This last document is specially to be present to the court. But the preliminary report is not suitable for the court because the individualization concept is not applied. Moreover the facts are not linked together to re-construct the crime. So any piece of evidence are linked together regarding the action and the time. Thus the crime cannot be proven.

EnCase is an expensive and a proprietary solution. It has the consequence that the tool can only be extended by the presence of the scripting engine which brings some flexibility. It has not the ability to communicate properly with other software. From the view of the development, an open-source solution would give the tools lots of possible contribution in the future. And more a deep review by the community of the sources would gives trust in the tool. The essence of opensource software gives the capacity to fully evaluate the software. Carrier (2002) opensource explains in his work that open sourced software are more likely to suit *Daubert*'s principals than closed source tool would.

4 The crime taxonomy

A cybercrime tree taxonomy is the foundation of the triage process. It defines which files must be look for during the triage process. But it needs the help of layer of abstraction to identify properly different nature of data. The last point explain the need of a database to store data.

4.1 The foundation of the extraction process

In our knowledge all the existing automated process of forensic software, models and framework concentrates on the final analysis. EnCase software has no native automated process, despite the fact it could be added with the script engine. Actually it is the script engine that give the capability of automating some tasks. But the software remains mainly conducted by the investigator. To give our tool this capacity to automate an investigation process with a minimal interaction of the investigator, the automated engine is based on the taxonomy of cybercrime. Casey (2004) explained than each crime first affects different files. With the *Locard Exchange Principal* from Dr. Edmund Locard itself, the criminal would interact with the environment, here in operating system Andrew (2007), causing some changes in it. In a case of crime involving computer devices, it would alter some files and the environment, the operating system.

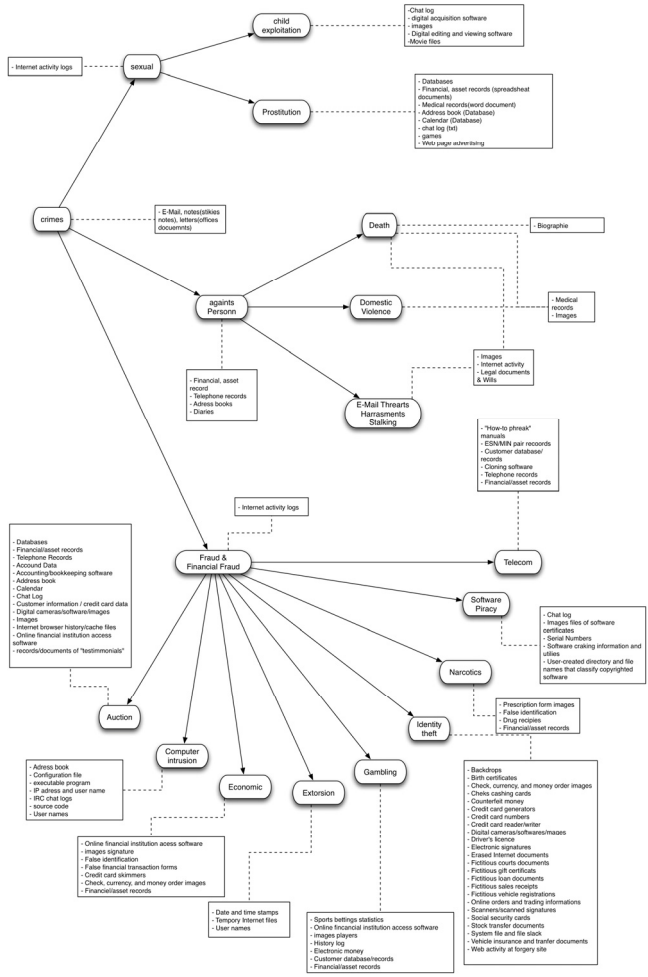

Figure 1: Cybercrime tree taxonomy

From this, a taxonomy of cyber crime is built. Figure 1 shows the result of the classification based on work from U.S. Office of Justice programs (2008).

The nodes represent nature of crime and the leaves represent the type of crime. The tree is enriched at each node and leave, with file extension names and type of data. The extraction process would only focus on files extension name and data that characterized one crime. In fact the file signature would be used because it is more relevant than file extension name which can be easily modified.

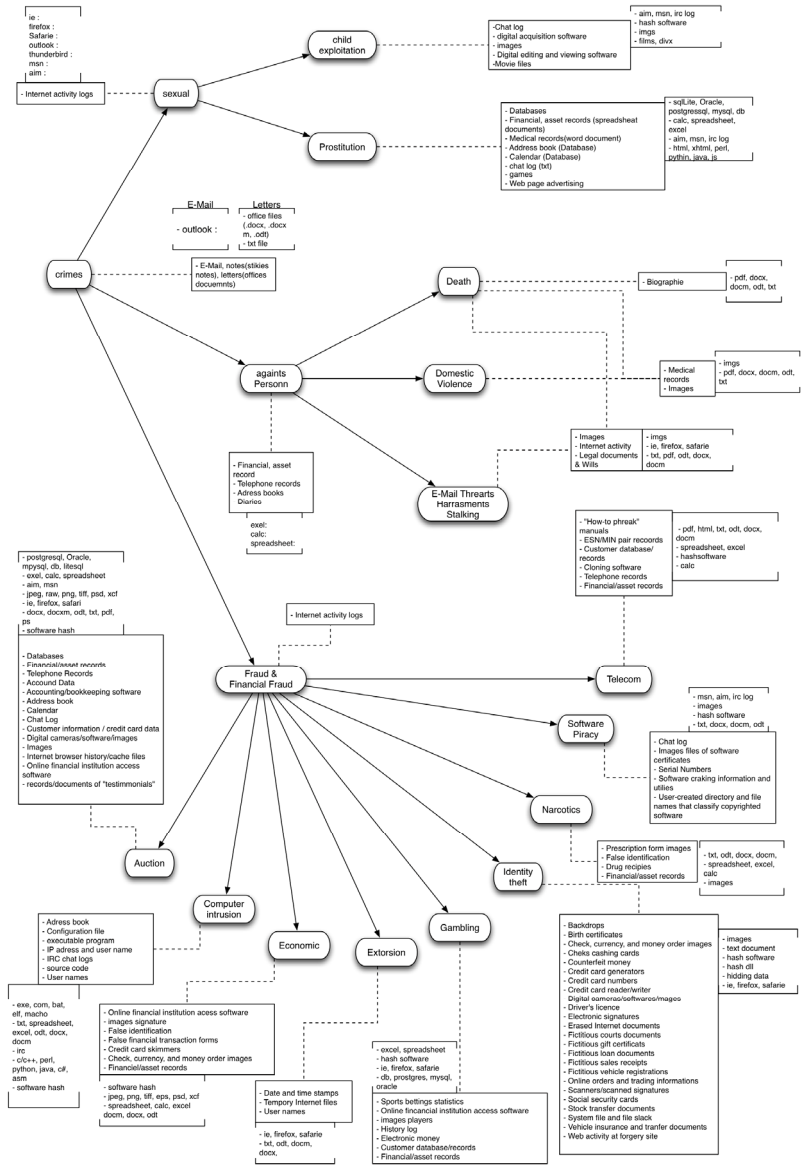

Figure 2: Enriched taxonomy tree

The U.S Department of Justice Office of Justice programs (2008) details each sort of information which caracterised one crime. Pedo-pornography requires image whereas financial fraud requires documents like financial report, bank account number or credit card number. For example images are stored in jpeg, bmp, or png file, and user documents are files created with office software like Microsoft Office or Openoffice. Therefore each information is listed with the exact file extension. From this fact we enriched the previous taxonomy tree with file extension name. Figure 2 presents the results.

4.2 Layer of Abstraction

The input of the tool is raw data, produced from the acquisition process. A tool such as *dd* would performed such task. There are no logical reading, which means if data are directly read they are no structure. Files and directories or operating system information as the registry in windows case cannot be accessed. Moreover the tools must handle various file-system ands the most spread one, linux's , macintosh's and window's file system. The tool must keep the ability to access different layer, from the hardware to the operating system. Blocks, cluster and file and directory are all the different layer it needs.

Sleuth kit offer this capacity. It is a set of open-source tools developed by Carrier to give investigator the ability to work at various abstraction layer Carrier (2003b). It handles and retrieves information and file from NTFS, FAT, EXT2-3 and HFS+ file-system. The set of tools make also possible the extraction of information at the partition level. Therefore different type of slack space can be define and deep analyse can be accomplish to detect hidden data, deleted file or encrypted data.

4.3 Storing the data and case management

Sleuth-kit is set of tools which makes possible to deal with abstracted layer of data. In a matter of accuracy the software needs to store all the abstracted layer as well as all the information which were extracted. The first objective is to limit the quantity problem. The size of extracted data and information is function of the size of the digital media. A database system would accomplish this task. It is an excellent solution to share the data.

The database system would be organised databases. Each case would be stored in it own database because they are no relation between data of several cases. So each case would be described by a sets of abstracted sets of data. Only relevant information would be stored to prevent huge size of data, especially when the investigation is performed on tetra-bytes amount of data. Generally offset and size are stored for partitions as well as files. In the last case, metadata of files are stored because they contain relevant information as last access time, creation time, etc... Such information are useful to perform the chronological analysis or data correlation analysis where it would use file owner, hash of file.

The abstraction layer has a layer for hidden and encrypted data. The database would stored as much information as possible. A special part is dedicated to slack space information. Those information are significant to detect hidden data. In case of

supposed encrypted data, only the offset and the length are stored but the database would have the ability to stored information found after a deep analysis.

In order to make possible an analysis of digital media by several programs, Alink et al. (2006) had developed a integrated approach named XIRAF. It uses a XML and a database system to wrap together different tools. Alink et al. focus their work on the analysing process where their developed an automated process. The XIRAF approach is an interesting approach for our software.

5 The working flow

The tool fit in a classic investigation process explained and detailed in the working flow as a iterative process. The next point gives the details on how to make the process efficient by selecting and avoiding defined sets of information. Then the paper focuses on the filter as a recursive function.

5.1 The general workflow, an iterative process

The tool does not follow a defined model from various work. It follows the digital forensic approach, acquisition, extraction, deep analysis and report as explain by Casey (2004). Still the extraction process follows a model which takes into consideration hidden and deleted data and the crime profile. The model is drawn out in the figure 3.

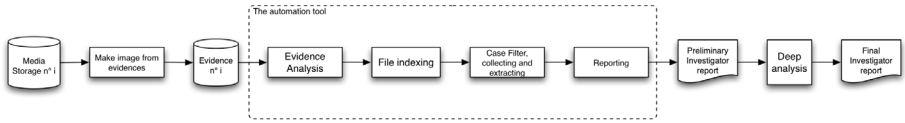

Figure 3: General working flow

The investigator enters all the digital media as evidence to the case and gives the type of crime. Then the process starts by analysing the evidence to determine the general structure of the evidence. The operating system would try to be guessed if it was not given by the investigator. For Each partition on the digital media, it performs file indexing. It is the second level of abstraction where it enables the software to sort out files by their nature. In the following the tool filters files that would be reliable for the case according to the crime profile. At the end of the process it produces a preliminary report, a starting point for the investigator to build the deep analysis.

5.2 A more efficient process

The filtering is based on the cyber crime tree taxonomy but it still need to look through entire evidences. To prevent time consuming, during the analysis step of each evidences, the environment, in fact files issued the operating system, are not analysed in filtering process. The reason is that not all cybercrime requires such a deep analysis of the evidence. The level of analysis is function of the suspect profile and the installed tools.

We need to understand what to analysed and how. For that purpose we developed a an formalization of the storage media. It is a model to reprensation how the data are organized on the disk. It helps on the process of the triage.

The structure of an image can be formalized with the help of the theory of set. The difference between sets are conceptual differences.

$$E=FS \cup MBR \cup PT \cup fAS \cup uAS$$

where

- E = $\{ci,\ i \geq 0$ and $i \leq EvidenceSize/ClusterSize)\}$ and $E \neq \emptyset$ is the image, is the set of all clusters. Size are in bytes.
- MBR = $\{b_j\ ,\ 0 \geq j \geq 1\}$, where *b* is a block of size 512 bytes. It defines the set of clusters allocated to the master boot record.
- *FS* is the set of clusters used to store file system information, $FS = \{fs_i,\ i \in \{1, 2, 3, 4\}\}$, where fs_i is a file
- *PT* is the set of cluster to store information about the partition table, $FS \cap PT = \emptyset$
- *fAS* is the set of file allocated space from file system, e.g. allocated clusters and $(FS \cup PT) \cap fAS = \emptyset$
- *uAS* is the set of unallocated space, e.g. unallocated clusters.

Each set can be process by an appropriate function. We need for each of sets to define functions processing them and their domain of result. Analysers of the Evidence are formalise as function:

- Evidence analyser:
$$EA: E \times MBR \rightarrow FS$$
- File indexing:
$$FI: FS \rightarrow fAS$$
- Slack space analyser:
$$SSA: E \cdot \times MBR \cdot \times fAS \rightarrow uAS$$
- Data analyser:
$$DtAn: sE \in (fAS \cup uAS) \rightarrow St$$

  where *sE* is the subset of clusters build from the allocated and unallocated cluster. It can also be view as a partition of the union of allocated and unallocated cluster. And *St* is the set of possible status defines as $St=\{encrypted,fileFragment,file,unknown\}$

- Triage:
$$T: CR \times \cdot fAS \rightarrow SF$$

  where *CR* is the set of crime profile, *IF* is the set of indexed file and $SF \in E$ is the set of suspected data.

- Deep analysis:

$$DA: SF \rightarrow CL$$

where $CL \in E$ is a set of clues.

We cannot consider file slack space in the modelization. The reason is data extracted from there are unstructured data. But need to be able to index them for future analysis. As indeed to keep the formalization simple it was not considered.

We first start to define file slack space as:

$$fAS = RD \cup FSS$$

where

- RD is the set of bytes of the file and
- FSS is the bytes that are not used by the file.

A file is defined regarding a given file system $fs/inFS$. File can be formalise as a function over fAS:

$$file : fAS_{fsi \in FS} \rightarrow \mathrm{F}$$

where F is the set of all file inside a file system. We do not need to considerer directory as there are either a file or a meta data for the file system.

Now we can formalized other kinds of abstraction as file produced by the user, file from the operating system and file from program. The set F of all files from the file system is therefore define as:

$$F = uF \cup sF \cup pF$$

where

- $uF \in F$ is the set of files produced by the user.
- $sF \in F$ is the set of files from the operating system.
- $pF \in F$ is the set of files from programs.

The Environment checker can than be defines as:

$$Ec: (pF \cup sF)\cdot \times NSRL \rightarrow \{sane, unsane\}$$

where $NSRL$ is the set of checksum of program file and from operating system file. This database allows to gain trust on some sets of data. We can therefore avoid any analysis on that space. But in the case of unauthorized access this set must be analysed to find any rootkits, malware and other malicious tools.

6 Formalization of the core and basic of the triage process

The filter and the extraction are recursive functions where one of the argument is the profile of the suspect and the crime profile. The profiling would be first initiated by the investigator. It first analyses program which are installed to detect known tools and software for securing data, performing hacks or penetration computer system. It also involves cryptographic techniques as steganography.

If no such tools are found the software does not try to look for encrypted or hidden data more deeply. In the opposite it would try to make available such data by looking for password and try then to decrypt the data. This part of the process depends on the operating system and on the user. For example an accurate acknowledged user in information security would not stored his passwords in the environment but a less accurate user would effectively stored them in his account. For the moment any procedure nor algorithm have been yet developed whereas detection of slack space, therefore possible hidden data, in windows environment exist Carrier (2005).

The triage process can be seen as a function to selection particular data within a set. The selection is based on some condition. A logical approach could be used but we preferred a vectorized approach because it gives the ability to tune and to give weight on some parameter. Let recall the definition of the triage function:

$$T{:}CR{\cdot}\times fAS \rightarrow SF$$

Now let give a formal definition of the set of crime profile CR. We define an element $cr \in CR$ as a vector of n independent vectors c_i, $0 \leq i \geq n$, hence we have $c_r = (c_0, c_1, \ldots, c_n)$. A vector c_i is one characteristic of the crime. The value 0 means the parameter is not considered whereas the value 1 considered it in the triage process. As an example of an arbitrary crime profile is defined by the following vector $crime$=(1,1,0,0,1,0,1). The number of field is not yet defined here.

The triage process is based on the characteristic of crime. The following function maps the file and the crime profile. The definition is:

$$p{:}F \rightarrow CR$$

For any files there is a mapping with the type of information it represents. Table 1 gives some examples.

Extension name	Type of information
jpeg	image
bmp	image
png	image
xls	spreadsheet
docx	office document
odf	office document
exe	applications
com	applications
pdf	office document
text	office document

Table 1: Examples of extension file maps to type of information

The definition of the triage function is now very simple. It needs two inputs, the crime profile and the file. But the file is mapped before the triage process beginnings. The result is either the value 0 or 1. The value 1 express the file is selected.

$$T(cr, p(f)) = \begin{cases} f \in SF & \text{if } p(f) \times cr^T > 0, \\ f \notin SF & else. \end{cases}$$

In fact vectors can be translated into matrix with allow us to describe the operation.

If the multiplication of matrix of file type by the transpose matrix of crime profile then the file is considered as suspected. We need to use the transpose of the matrix of crime profile so the multiplication of matrix makes sense.

Here some example of triage process in case the file is not suspected and in the case it is. Let defines an arbitrary crime profile cr=(0,0,1,1,0) and two files f_1 and f_2. Their profile is hence $p(f_1) = pf_1$ = (0, 0, 1, 0, 0) and $p(f_2) = pf_2$ = (0, 1, 0, 0, 0). File f_1 is suspected because

$$p_{f_1} \times cr^T = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$= 1$$

whereas f_2 is not because we obtain

$$p_{f_2} \times cr^T = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$= 0$$

7 Future work and improvements

The interface of this tool was not yet discussed yet which is still in development. But here are some guideline to design the software interface. It must remain as simple as possible and the more intuitive as possible. So the design would care about what is the more suitable for human.

Cryptographic aspects were introduced in the process but it represents a big part of development. As we meant previously, this aspect would first start by look stored password. An appropriate algorithm must be developed to extract and validate automatically the password. It represents another big challenge.

The keyword search engine was also briefly mentioned. This is one of the next point to work on.

The interaction with other tools was mentioned. It his challenging to give the ability to the software to communicate and interact with external program on shared data. The constraint is to be able to nicely add, update a external tool without modifying the architecture. As it was late introduced in the research, XIRAF solution is not yet integrated even if it remains a nice solution. If not, a XML based protocol sound like a nice alternative. The challenge here remains in the ability to extend, change, update easily parts of the software without having the necessity to re-develop it from scratch. It is to give an interface to allow such manipulation.

8 Conclusion

This paper introduces foundation and requirements for our novel automated tool. It aims to fill the need of a complete tool for non-technical investigator. Therefore the investigator could focus only on the case without having to matter about technical difficulties. The software has not the purpose to replace a power tool such as EnCase but it is an alternative tools for non-technical and for industries and police officer for example.

The first priority is to develop the search engine and the interface. Still many aspect of the tool are not yet covered in this paper they are planned in the core of the workflow. The core of the process would be developed in Python. It was chosen for its portability and for its fast coding capability it gives. The first step would be to check and to validate the workflow process.

As based-mobile device treats are growing such issue should be surveyed and potential modification may be added to the software. The nature of mobile technology is different from computer even if they share same foundations and principles. The functionality and the behavior would get closer to computer but still differences would remain and it would give new challenges in digital forensic.

9 References

Alink, W., Bhoedjang, R. A. F., Boncz, P. A. & de Vries, A. P. (2006), 'Xiraf - xml-based indexing and querying for digital forensics', Digital Investigation 3(Supplement-1), 50–58.

Andrew, M. W. (2007), Defining a process model for forensic analysis of digital devices and storage media, in Huang & Frincke (2007), pp. 16–30.

Bhat, V., Rao, P. G., V, A. R., Shenoy, P. D., R, V. K. & Patnaik, L. M. (2010), 'A novel data generation approach for digital forensic application in data mining', IEEE Computer Society pp. 86 – 90.

Carrier, B. (2002), 'Open source digital forensics tools: The legal argument'.

Carrier, B. (2003a), 'Defining digital forensic examination and analysis tools using abstraction layers', International Journal of Digital Evidence 1(4).

Carrier, B. (2005), File system forensic analysis, Addison-Wesley.

Carrier, B. D. (2003b), 'Defining digital forensic examination and analysis tool using abstraction layers', IJDE 1(4).

Casey, E. (2004), Digital evidence and computer crime, forensic science, computers and the internet, Elsevier.

Garber, L. (2001), 'Computer forensics: High-tech law enforcement', IEEE Computer 34(1), 22–27.

Gladyshev, P. & Patel, A. (2004), 'Finite state machine approach to digital event reconstruction', Digital Investigation 1(2), 130–149.

Huang, M.-Y. & Frincke, D. A., eds (2007), Second International Workshop on Systematic Approaches to Digital Forensic Engineering, SADFE 2007, Seat- tle, Washington, USA, April 10-12, 2007, IEEE Computer Society.

Hunton, P. (2011a), 'A rigorous approach to formalising the technical investigation stages of cybercrime and criminality within a uk law enforcement environment', Digital Investigation 7(3-4), 105 – 113.

Hunton, P. (2011b), 'The stages of cybercrime investigations: Bridging the gap between technology examination and law enforcement investigation', Computer Law & Security Review 27(1), 61 – 67.

Marrington, A., Mohay, G. M., Morarji, H. & Clark, A. (2010), A model for computer profiling, in 'ARES', IEEE Computer Society, pp. 635–640.

U.S Department of Justice Office of Justice Programs, U. D. (2008), 'Electronic crime scene investigation: A guide for first responders, second edition', a Accessed Online on 15/07/2010, http://www.ncjrs.gov/pdffiles1/nij/219941.pdf.

Peisert, S., Bishop, M., Karin, S. & Marzullo, K. (2007), Toward models for forensic analysis, in Huang & Frincke (2007), pp. 3–15.

Slay, J., Hannan, M., Broucek, V. & Turner, P. (2004), Developing forensic computing tools and techniques within a holistic framework: an australian approach, in 'Information Assurance Workshop, 2004. Proceedings from the Fifth Annual IEEE SMC', pp. 394 – 400.