Ca’ Foscari University of Venice

Department of Environmental Sciences, Informatics and Statistics

MSc In Computer Science

Master Thesis

Digital Forensics overview and real scenario

Author:
Davide Gastaldon
ID number 810733

Supervisor:
Prof. Riccardo Focardi
www.dsi.unive.it/~focardi

Academic year 2012/2013

Contents

List of Tables

List of Figures

Acknowledgements

Probably this is the end of my academic career so some thanks are required, either from academic and professional or affective.

I need to say thanks to my supervisor *prof. Riccardo Focardi* who always encourage me and to *dott. Andrea Albarelli* to direct me in computer vision domain.

Without *Yarix* help and support this work would not be born, so a special thanks to my tutor *dott. Stefano Meller* partner and co-founder and to *Mirko Gatto* CEO of Yarix, over and above *Diego Marson* and *ing. Alberto Lorenzon* invaluable sources of knowledge.

But my best thanks are aimed to my parents *Adriana* and *Arcangelo*. These small text would be simplistic to explain what they are for me.

Also my *wrestling team* have to be mentioned since "once you have wrestled everything else it is easier", my masters *Giovan Battista* and *Angelo* teached, and still teach, how to never give up.

To my friends, which bear me, but especially to *Andrea* keen mind and my best reviewer.

Last but not least my brother *Giorgio* who shared more than a sleepless night for whatever i need.

"He who does not prevent a crime when he can,
encourages it."
- Lucius Annaeus Seneca -

"Eliminate all other factors,
and the one which remains must be the truth."
- Sherlock Holmes -

1 Introduction to Digital Forensics

The current period can be seen as the era of the digital revolution, characterized by widespread, easy to purchase, and easy to use various digital devices that are getting smaller and smaller in size, increasingly faster, and have larger storage capacities. We live in a world that is constantly connected to the internet which is increasingly overlapping physical reality, both socially and occupationally. As with all major changes, the new possibilities available can be used for both legitimate and noble purposes, as well as for unlawful and despicable purposes. Like the rest of society, the criminal world benefits from new technological advances, using these new means to support their criminal activities which has substantial repercussions on the work of law enforcement.

Based on Locard's principle,[1] considered to be the founder of forensic disciplines, it was clear that there was a need to complement the usual investigative techniques, developing methodologies for the extraction and analysis of data from digital devices with the aim of it being admissible as evidence in any subsequent legal hearings.

These expert analyses are becoming more and more necessary, even for traditional types of crime such as murder, in which the digital device is not the means or body of evidence, but

> "the information from these things that can be extracted and analyzed, which can reveal clues for understanding and resolving cases such as the suspect's habits, interests, social relationships, technical skills which, at times, allows for the timeline of the crime and the modus operandi to be reconstructed.[1]"

The lack of a shared taxonomy of the various associations that deal with research and development in forensics leads to the fact that there is currently no universally accepted definition of Digital Forensics. In the Anglo-Saxon world, where this Discipline is born, DFRWSm[13], which means Digital Forensics Research Work Shop, provide this definition:

Definition 1 (DFRWS definition) *The use of scientifically derived and proven methods toward the identification, preservation, collection, validation, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations.*

It can be inferred that Digital Forensics is a macro container that contains many branches:

- Computer Forensics, the branch most widely studied and used.
- Mobile Forensics, the fastest growing sector, generally dealing with portable devices, mobile phones, smartphones, and tablets.

[1] "Wherever he steps, whatever he touches, whatever he leaves, even unconsciously, will serve as a silent witness against him. Not only his fingerprints or his footprints, but his hair, the fibers from his clothes, the glass he breaks, the tool mark he leaves, the paint he scratches, the blood or semen he deposits or collects. All of these and more, bear mute witness against him. This is evidence that does not forget. It is not confused by the excitement of the moment. It is not absent because human witnesses are. It is factual evidence. Physical evidence cannot be wrong, it cannot perjure itself, it cannot be wholly absent. Only human failure to find it, study and understand it, can diminish its value."

- Network Forensics, the part that deals with investigations involving computer networks with techniques and methodologies shared with those involved in network security.
- Database Forensics, forensic study and analysis of databases.

Digital Forensics tools and methodologies are widely used also inside Incident Response and E-Discovery activities [19], with the peculiarity that investigators use them after a system compromise in order to ensure full, accurate, uncontaminated and documented examinations of the cyber crime scenes.

In its first decade, Digital Forensics has evolved rapidly thanks to the interest expressed by many researchers and developers who have implemented methodologies and tools for intercepting digital evidence, also ensuring its probative value in court. On the other hand, the continuous improvement of high-tech tools have introduced new problems in the activity of Digital Forensics Investigation including:

- The increasing size of storage devices which means more time in creating a forensic image and processing of all of the data extracted.
- The increasing prevalence of solid state drive storage media and the proliferation of different hardware interfaces.
- The proliferation of operating systems and file formats which are also very different from each other and which greatly increases the requirements and complexity of the instruments to examine the data and the cost of developing them.
- The increase of complex investigations that require analysis of multiple devices followed by the correlation of the data found to identify evidence.
- Pervasive encryption that greatly complicates analysis of the media, making the interpretation of data useless if you are unable to decrypt the encrypted content.
- The use of cloud computing for the provision of computer services that share a virtual hardware and software platform (Software as a Service, Platform as a Service, and Infrastructure as a Service) and have the common denominator of broadband remote access through a VPN user authentication.

Forensic analysis tools, developed to help technicians identify potential digital evidence, become all the more obsolete as technologies quickly evolve and are renewed. In addition, the identification of evidence for case studies of investigation such as murder, terrorism, conspiracy, where the computer tool used is not strictly related to the type of crime, in contrast to cases such as child pornography or stalking, the potential support of these tools is lowered considerably. In these cases, since it is difficult to obtain a chronological reconstruction of the events or actions of the offender, it is the technicians who conduct these activities based on their previous investigation experience. A possible solution to the limitations set out above can be represented by a new methodological approach enabling a greater degree of abstraction and independence from the specific digital device that will be subject to investigation.

Scientific research is also geared to the development of methods and tools that enable a greater degree of abstraction from the type of data to be analyzed, although the efforts in the standardization of methods of representation and processing of information have not achieved the expected results. A similar need for standardization is also seen in the field of architecture.

So, today, we are seeking to develop Digital Forensics tools that are as portable as possible (cross-language and multi-platform tools). The need for guidelines that standardize a forensic analysis process was noticed immediately both in academia and in the operational and investigative sectors.

In 2001, the Digital Forensics Research Workshop (DFRWS), after the above mentioned Definition 1, identified a seven-step process: identification, preservation, collection, examination, analysis, presentation, and decision. In 2004, the U.S. Department of Justice (DOJ) published a guide for law enforcement relating to the Forensics of digital evidence. Three basic principles were represented in the guide for the forensic investigation of digital devices:

- Actions taken to secure and collect digital evidence should not affect the integrity of that evidence.
- Persons conducting an examination of digital evidence should be trained for that purpose.
- Activity relating to the seizure, examination, storage, or transfer of digital evidence should be documented, preserved, and available for review.

Following years have seen the publication of many research works and guidelines to describe and standardize Digital Investigations. The thesis of B.Carrier [2] must be cited since he first describes, in a technical and scientific way concepts as transfer, identification, classification, individualization, conjunction and reconstruction. Also worth mentioning the official digital evidence handbook [4] prepared by ACPO, Association of Chief Police Officers, with the same thread: recover evidence source, save its integrity so as to use inside a court law with full probative value.

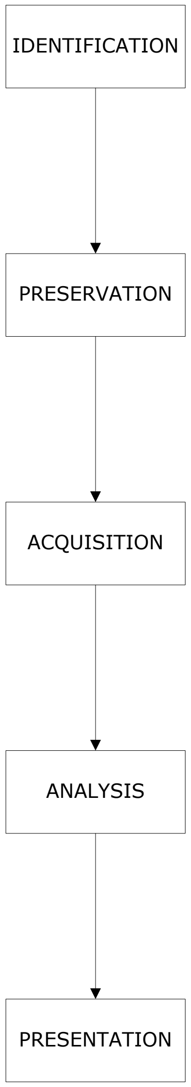
2 Forensic investigative process

The main steps envisaged for handling so-called Digital Evidence (handling steps) in a typical forensic investigation workflow include at least the following steps: identification, preservation, acquisition, analysis and presentation. Figure 1 points out it.
Let's look at each of them in detail in next sections.

2.1 Identification

The first phase in the Digital Forensics Investigation process is aimed at identifying potential media and the data it contains. The increase in storage capacity of data and the functions available today on all digital devices complicates identifying on which devices hypothetical digital data that is useful to the investigation might actually reside. In fact,

Figure 1: Digital Forensic Process Steps



you run the risk of subjecting devices to analyses, which by their nature are complex and lengthy, that are of no relevance to the investigation. Identification must be certain and unambiguous since the primary objective is to indisputably connect the data to the media it was extracted from.

It should however be stressed that the identification process is not limited to only the media (hard disk, CD, usb key, mobile phone, etc...). Evidence is not only understood as all of the data contained in one device, but also extends the same set of evidence extracted from individual devices. Therefore, it seems evident that identification is not a step in itself, but also occurs in the later phases of acquisition and analysis.

2.2 Preservation

Just like the identification phase, preservation is not even a step in itself, but rather it inextricably blends with the phases of acquisition and analysis. The digital evidence is preserved, technically, through the use of devices/software and analysis is only conducted on copies/-forensic images. In addition, in order to ensure the repeatability of digital evidence in court, the technicians keep the so-called chain of custody updated where they document all of the operations performed on the findings and forensic copies, which may also be admitted as evidence to be evaluated by a court of law. Failure to observe these operating procedures and precautions voids the very probative value of the evidence.

2.3 Acquisition

Once the possible media containing information relevant to the investigation have been identified, it is necessary to make a forensic copy of the data on the media, or, in the case of Network Forensics, in transit over a network. This is a delicate phase of the process because any operations carried out by personnel that is not properly trained and/or not competent could lead to the destruction of potentially relevant data or invalidation of the media and/or the data it contains. The acquisition phase consists of a three-step process: the development of an acquisition strategy, the acquisition itself, and verifying the integrity of the data acquired.

Developing a plan to acquire the data is an important first step in most cases because there may be several potentially important sources. The analyst should, therefore, create a plan that determines the priority of the sources, establishing the order in which the data must be acquired. Determining factors of the strategy might be represented, for example, by previous experience in similar situations which should help the analyst estimate the likely value of each data source.

For the acquisition phase itself, there are many commercial software packages today that are designated to carry out this operation automatically which we will discuss in more detail later in this text.

Finally, verifying the integrity of the data acquired constitutes a crucial and necessary action so that the evidence will have a probative value in court proceedings. For this purpose, the message digests of the original and copied data are calculated with dedicated software tools and then compared to ascertain whether there have been any changes.

2.4 Analysis

Once the information which is important for the forensic investigation is extracted, a study and a detailed analysis of the digital evidence is conducted. The techniques that are used in this phase are meant to study and understand all of the attributes and characteristics of the individual data extracted which may be useful to establish the relationships, habits, or technical capabilities of the suspect.

The analysis should lead to the traceability of all possible computer evidence that is useful for evidentiary purposes, and to this end, today's storage devices offer a considerable amount of information. However, often times this process gets complicated, as would be the case of partially deleted data that could provide the most interesting information. Forensic technicians may use some basic guidelines to orient themselves on what and where to look for any computer evidence and forensic tools developed specifically for that purpose.

The most important issues during the analysis phase are due to a variety of causes: the varied types of devices with different operating systems, different ways of storing data, different types of file systems and memory organization. Often, then, forensic technicians are forced to use different toolkits for the same case or even manually perform some delicate analysis operations.

2.5 Presentation

The final step in the forensic investigation process involves the preparation of reports containing the most important details relating to each phase, including references to the operating protocols followed and the methods used to seize, document, collate, preserve, and analyze every single piece of evidence that may have probative value.

The presentation is, therefore, a kind of summary and conclusive description of the entire forensic investigation process that gives visibility and transparency for all who are involved in the case.

3 New outlooks

In this chapter new views and ways of thinking are shown in. This is a necessary step, because some of the following observations are not inside the standard domain of most of the computer scientists and experts.

Actually there is no standard analysis methodology accepted worldwide, mainly for these reasons:

- There is a wide variety that a digital forensic consultant has to cope with.
- Device evolution are exponentially and as a result of security researches, forensic analysis difficult increase day by day, limiting consultant capabilities.
- Legislative disciplines are quite different from one country to another and often there are gaps from law domain to technology evolution.
- Every consultant has his own style and experience.
- There are many products and each of these has different ways to present results, moreover, especially commercial tools, guide user through the creation of a report, so presentation style depends on used tools.
- When a survey must be submitted to a jury or it is requested by a public prosecutor, as far as possible, it has to be adapted in order to be fully understood during the case's proceeding in the court.

From the experience of consultant [12] and law enforcement agents we can derive these mandatory milestones:

- Maximum regard for the evidence.
- Time optimization.
- KISS (Keep it simple stupid!).
- Open minded approach.
- Be able to replicate what has been done.

And the most important, *be chary*.

Following some common context are mentioned.

3.1 Lawsuit

An high level introduction of Italian legislation on cyber-crime and digital evidence can be found in Appendix 10, here some considerations are stated in order to understand the perspective of a computer forensic expert when his work has court as target.

Digital, forensically sound, investigation domain was born with the aim to use its results inside a lawsuit, inter alia in criminal proceedings. As repeatedly stated in this thesis, during most of our time we have with you digital devices that can be useful to testify an alibi or to help the identification of an outlaw. Also keep in mind that trade is finally moving to e-commerce and the most part of the fraud are following it.

A forensic consultant can get the job from law enforcement officers, from the involving court or from defense lawyers. In these situations surveyor has to start from the beginning, acquiring seized devices or acting with an evidence that fill forensic requirement, up to results' presentation that must answer to investigative question.
Theoretically, applicant must clearly express which issue the survey has to address, but often this cannot be considered true because there are various gap between law and information technology evolution.
Generally analysis is requested to notice probative items inside the device, for example:

- find specific files, i.e. child abused videos or pictures.
- reconstruct a usage time-line of a device, e.g. determine if a user was using the device inside a time range.
- profile Internet activity, such as downloading material protected by copyright.
- inspect network traffic.

During consultant's work the lawsuit is suspended to permit exhaustive and proper operations.

3.2 Enterprise

In Italy inside a company, managers are requested to comply the "Labour Code" (regulations that protect workers inside business), especially Article 4 [16] that says:

> "Audiovisual surveillance and any other remote control technique to monitor employees' activities are strictly forbidden".

Also "Privacy Code" [18] regulates inspecting possibilities inside workplaces, since for Italian legislation, is considered a place with social need.

At the same time, however, the law guarantees employer's right to control enterprise tools and to prevent the illicit use of them.
In this context is fundamental an internal act that clearly states utilization policy and explains any measures adopted to monitor devices that workers receive for business reasons.
It must be emphasized that principles of "relevancy" and "not-overbalance" should be met, in other words investigations that affect personal freedoms are not allowed and controls which are prolonged or inspired by prejudices and dislikes are not legal.
This rules book must be accepted by each worker and it is the base to proceed to further investigations.

Investigations should begin on aggregated ways, for example by investigating a business unit in whole and then, if there are founded assumptions, raise detail's threshold, keeping in mind that focusing on a specific individual can be viewed as an harassment.
Once specific device is detected, or rather its use is undue, company legal representative may commission an analysis to a forensic consultant.
Consultant proceeds in the usual way, observing generally framework previously explained, because from the results of his survey a disciplinary actions can be released and perhaps

the outcome can be a compliant sent to the court. It is worth noting that during a business analysis can be uncovered a crime.

An observation is mandatory, based on experiences of digital forensic community: is better to prevent malicious uses than starting litigation between company and worker; for example, stop not allowed internet browsing with a proxy, is a common way to avoid illegal behaviour during working hours.

3.3 Incident Response

In forensic domain incident is defined as

Definition 2 (Incident) *"every action not authorized, not acceptable, lawless that concern a computer or a network."*

Formally Incident Response is defined as "all actions" to stop and contain a cyber incident, in which forensic experts cover an important role to assess actions and responsibilities.

These are good examples of what definition implies:

- theft of corporate data by disloyal employee;
- unauthorized access to an information system;
- breach of a protected wireless network;
- ownership and spread of illegal stuff, such as child pornography.

Such cases are prosecuted by Italian legal system, so first of all it is necessary to detect and prove them doubtless.

It might be that facts inherently disclose the incident, think about a Denial Of Service attack that stops core business services, or most commonly in the case of forensic analysis, incident' author try to hidden, as far as possible, every evidence correlated; according to scientific community forensic is better when it is coupled with prevention.

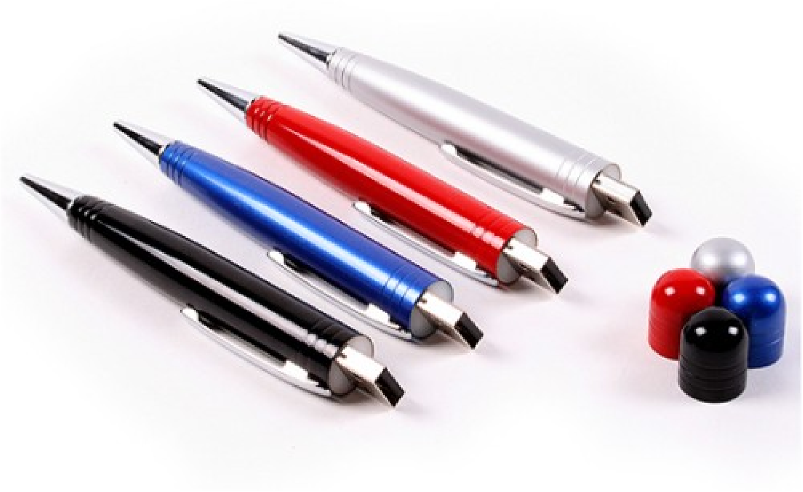
These are reasons behind Incident Response team creation, a growing trend inside sensitive company that permits, if core business is threatened or there are economic damages, to ward enterprise concern and reputation.
It is common in this case to operate most in live mode to avoid loss of volatile data, such as RAM's content, and to help business continuity, think for example to an acquisition of a mail server that obviously cannot be halted without disruptions.

3.4 Tips and mistakes

Here are reported various mistakes occurred in few famous Italian cases which have nullified evidences during a trial and on the other hand some hints are given mainly to convey of how it is necessary act in common investigative situation but out of the box for a computer scientist.

First of all inspection phase, generally follow out by police forces, must be carried out with attention. The more investigator is professional the less can be commented, so it is important to dress gloves to ensure digital fingerprints integrity within keyboards, mouse and electronic devices. This is a mandatory point in cases where there are other kinds of forensic investigations planned. Keyboards requires a detailed analysis particularly when encryption must be faced as a statistical analysis of keys wear can reveal chars used more and steer brute-force attacks. Development of sleuthing is common also for digital forensic consultants and can be useful during this stage to smell out not common devices that have storage capacity but camouflaged to look like anything else, think for example to Figure 2 a USB drive inside a pen.

Figure 2: USB camouflages



After which investigators have to inhibit any action leading to the destruction of digital equipments. It is a well-known fact inside forensic circle the attempt, performed during an inspection, to warp a hard disk by putting it under running water with the hope to induce short-circuit or oxidations. Hard drive was entirely recovered using a specific cleanroom for data recovering.

Another striking event is the attempt made from Israeli border agents to kill a MacBook by shooting it without hit the hard disk in which suspicious photographs were stored[2].

Another stage where it is important to work with caution and attention it is the analysis phase, within it any error occurred implies the overturn of prosecution theory. An error commonly referred to, is what happened during a murder investigation when prime suspect spontaneously delivered personal notebook to use it as a proof of his innocence, but the trouble was that notebook was powered on and off several times by police before perform forensic acquisition[3]. Considering that, in a normal windows environment, boot and shut down modify the system state by changing on average 1000 files, subsequent timeline analysis and files authorship are unacceptable.

[2] http://www.youtube.com/watch?v=ihXtbB-4GWw

[3] http://w3.uniroma1.it/mastersicurezza/images/materiali/SENTENZA_STASI.pdf pag. 37

Figure 3.4 summarizes a common acquisition process for a Personal Computer.*Volatile Data* means the suspect of presence of useful data inside RAM memory, for example a password or other items that usually are kept in RAM.

NOT SHUT DOWN

is emphasized since shutdown operation alters system state, raising the plug the system is frozen. Figure is an example of which kind of photographic documentation is necessary.

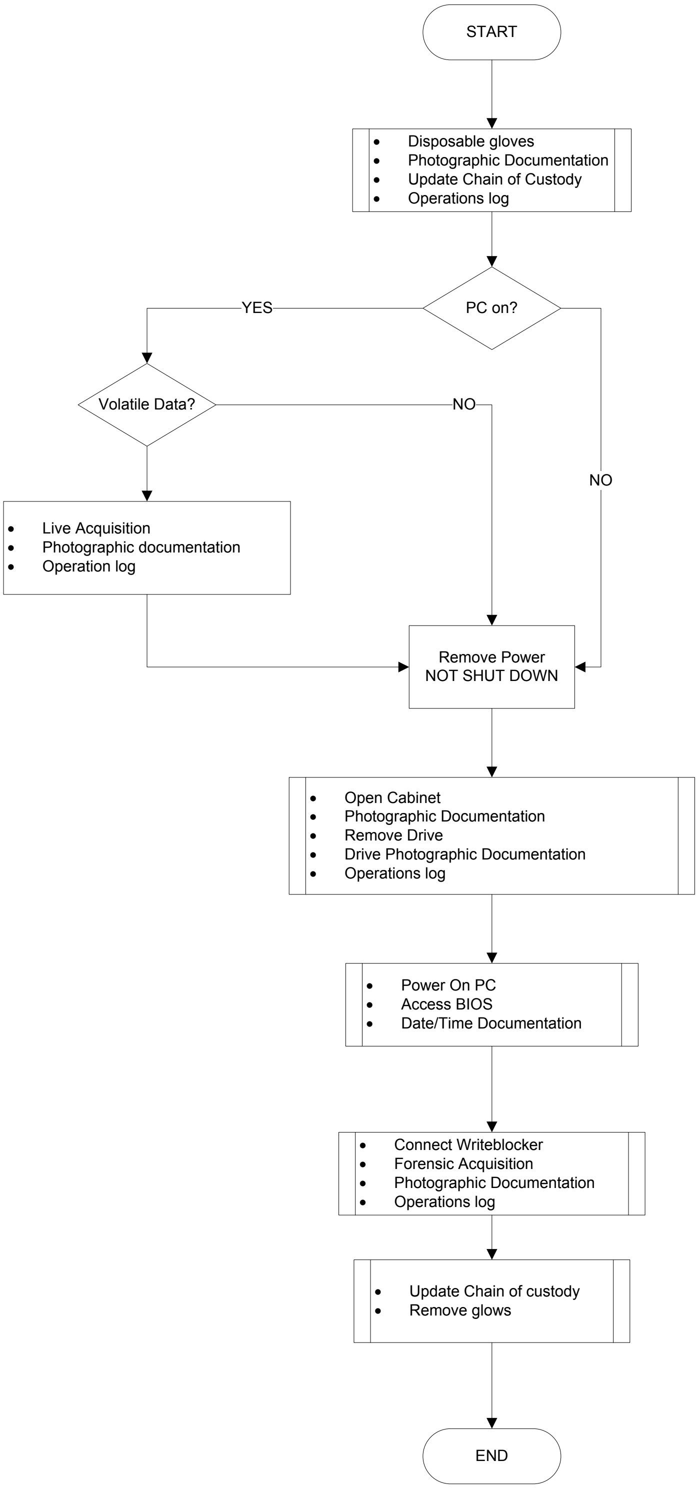
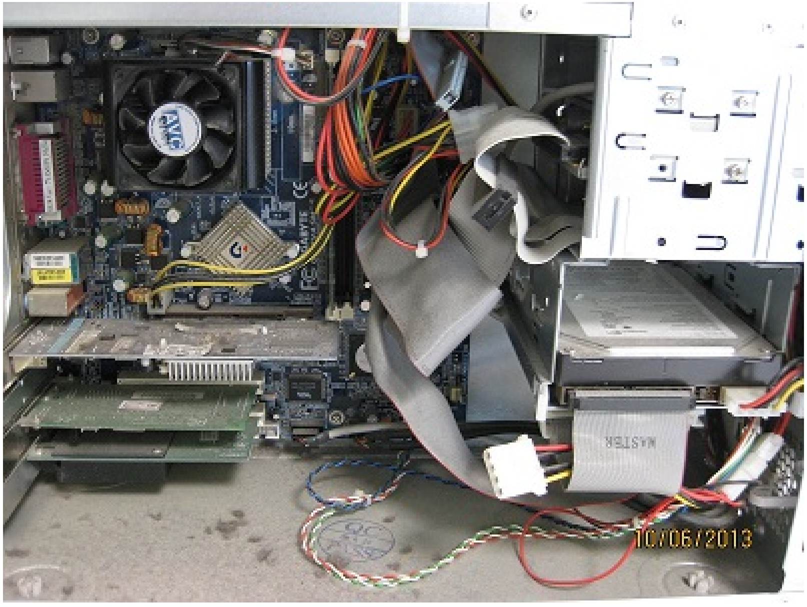


Figure 3: Flowchart of a PC acquisition

Figure 4: Example of photographic documentation



4 A real case

This is a description of my internship activities carried out at Yarix srl [4], under the supervision of dott. Stefano Meller. I follow what a computer forensic expertise do during an analysis requested by law enforcement entity, with the goal to find all the evidence relevant to the court for a criminal proceeding. Obviously all the links to the real case are hidden, what is true is the methodology of work, the problems to face and results achieved.

4.1 The context

Public prosecutor office is investigating in the domain of pedo-pornography. Input was given by a girl, underage, which reveals to Carabinieri a blackmail perpetrated by some "friends" account on Skype, Netlog, Facebook and MSN Messenger. This girl, hereafter called Alice, did some striptease via webcam during a conversation with accounts that she believed were peers, namely with more or less the same age. Alice did not know that indeed these accounts are fake identities and moreover they recorded strip show. With this video they started to demand other show by threatening to disclose the video to all Alice's friends on social network and instant messaging. Alice has also stated that they continually asked to physically meeting with sexual purposes.

Initial investigation, with the help of a forensic consultant and Internet Service Providers, led Carabinieri to identify accounts owners, particularly three men hereafter called Dan, Charlie and Bob. Then prosecutor has authorized the seizure of all digital devices and commissioned an examination to a forensic consultant with the request to analyze the entire device set and report all the evidence useful for the lawsuit.

4.2 Identification

Identification process was made by Carabinieri police force, through perquisition inside suspects' homes and workplaces.
In this phase officers confiscate all digital devices they can found; they inspect also hidden place, like bathroom furnishings, where some devices are founded. Each item is photographed in the context in which is discovered.

There are various kinds of device inside the set of stuff seized; the first operation to do is to list every item so as to uniquely identify during the survey. All the devices are also described by its manufacturer, model and serial number. Device that falls inside mobile equipment category, i.e. every device that can use mobile communication networks like cellular phone, Smartphone and tablet, are also identified by IMEI number. SIM card are described also by phone number.

During this thesis this data are hidden to respect privacy of involving actors.

Tables 1 2 3 report stuff seized.

[4] www.yarix.com

Table 1: Dan's item

Device	Code
Cellular phone	DAN-CELL01
Cellular phone	DAN-CELL02
Cellular phone with Camera	DAN-CELL03
SIM inside cellular phone with Camera	DAN-CELL03-SIM01
iPhone 3	DAN-IPHONE01
iPhone 5	DAN-IPHONE02
SIM inside iPhone5	DAN-IPHONE02-SIM01
iPad 2	DAN-IPAD01
iPod 2nd generation	DAN-IPOD01
External Hard Disk	DAN-HDE01
Notebook	DAN-NB01
Unique Hard Disk Inside Notebook	DAN-NB01-HD01
USB Stick	DAN-USB01
USB Stick	DAN-USB02
Memory card inside Camera	DAN-MEM01
Video tape inside video Camera	DAN-MEM02

Table 2: Charlie's items

Device	Code
Cellular phone with Camera	CHA-CELL01
Cellular phone	CHA-CELL02
SIM inside cellular phone	CHA-CELL02-SIM01
Office Personal Computer	CHA-PC01
Unique Hard Disk Inside Office Personal Computer	CHA-PC01-HD01
Personal Computer	CHA-PC02
Unique hard Disk Inside Personal Computer	CHA-PC02-HD01
USB stick	CHA-USB01
USB stick	CHA-USB02

Table 3: Bob's items

Device	Code
Cellular phone	BOB-CELL01
Smartphone	BOB-CELL02
SIM inside smartphone	BOB-CELL02-SIM01
Tablet	BOB-TAB01
Notebook	BOB-NB01
Unique hard Disk Inside Notebook	BOB-NB01-HD01

4.3 Preservation

All items are kept inside a repository, physical protected from unauthorized access and isolated from outside communication.

Preservation correctness is ensured by a chain of custody module. All investigators assigned to the case are obliged to refresh it whenever they use a find. A typical chain of custody model has the layout of Table 4.

Table 4: Chain Of Custody Model

Investigator Name
Evidence Name
Time/Date of withdrawal from repository
Time/Date of return to repository
Analysis scope and methodology

4.4 Acquisition

4.4.1 Inital consideration

As we can see there are a lot of device to acquire and analysis, so it is necessary a triage technique to optimize time and effort.

In this case are also useful opinions given by law enforcement officers, especially these observations:

- DAN-HDE01 has been found in a hidden location, the same as for CHA-USB02.
- Experience of the older agent applied to Bob's behaviour, suggests his innocence. This thesis is also supported by Bob age; he is younger than the other and not so older with respect to Alice.

Another constraint is given by CHA-PC01, it is located inside company office and cannot be moved and acquired inside Yarix forensic lab.

It is worth noting that the total amount of devices is very high, so the acquisition process is a high time consuming task. Investigators also ask to find relationship between Dan, Charlie, Bob and Alice. According to these assumptions the chosen scheduling is the following:

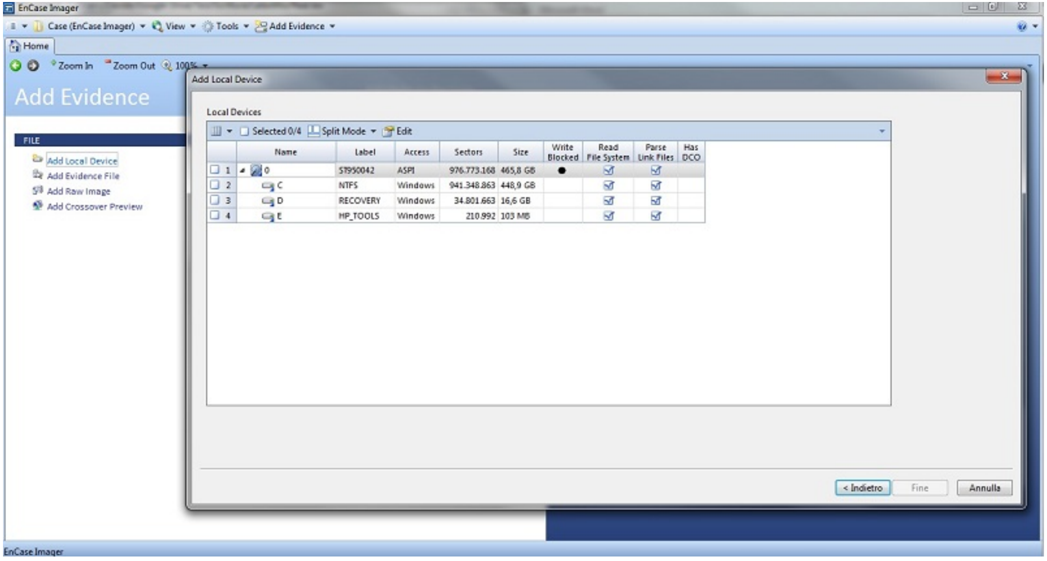
1. Live acquisition of CHA-PC01 under the statement that is a non repeatable process;
2. Acquisition of DAN-HDE01 and CHA-USB02
3. Acquisition of BOB's devices;
4. Acquisition of other storage media at the same time of Smartphone, tablet and cellular phone;

4.4.2 Tools used

The consultant decides to use EnCase Forensic Suite (Figure 5), under Windows environment, to acquire storage media with a hardware write blocker to ensure forensically soundness of the process (Figure . The live acquisition of CHA-PC01 is made under Caine, shown in Figure 7 and then the evidence file is imported in EnCase.

For mobile devices is used Oxygen Forensic Suite and UFED Cellbrite8; the union of these tools supports most of the cellular phone, smartphone and tablet. UFED also permit acquisition of the SIM cards. Internet use, Social Network analysis and Instant Messaging are made by Internet Evidence Finder.

Figure 5: EnCase Imager



Local Devices

		Name	Label	Access	Sectors	Size	Write Blocked	Read File System	Parse Link Files	Has DCO
□	1	0	ST950042	ASPI	976.773.168	465,8 GB	●	☑	☑	
□	2	C	NTFS	Windows	941.348.863	448,9 GB		☑	☑	
□	3	D	RECOVERY	Windows	34.801.663	16,6 GB		☑	☑	
□	4	E	HP_TOOLS	Windows	210.992	103 MB		☑	☑	

Same results can be obtained by using open source tools, the choice of commercial tools is justified by these factors:

- Consultant is a certified examiner inside these environments;
- These tools are widely accepted and well known in Digital Forensic context;
- At the same time EnCase use a proprietary file format, but there are various open source tools that support it, moreover EnCase allows you to export evidence file in raw format.
- Task can be easily automated

4.4.3 Acquisition process

Writeblocker and EnCase ensure a forensically sound acquisition process, for each device is automatically computed MD5 digest, and it is saved in Encase image file format, divided into segments of 4400 MB that can be easily written in DVDs if it is necessary.

Figure 6: Hardware Writeblocker

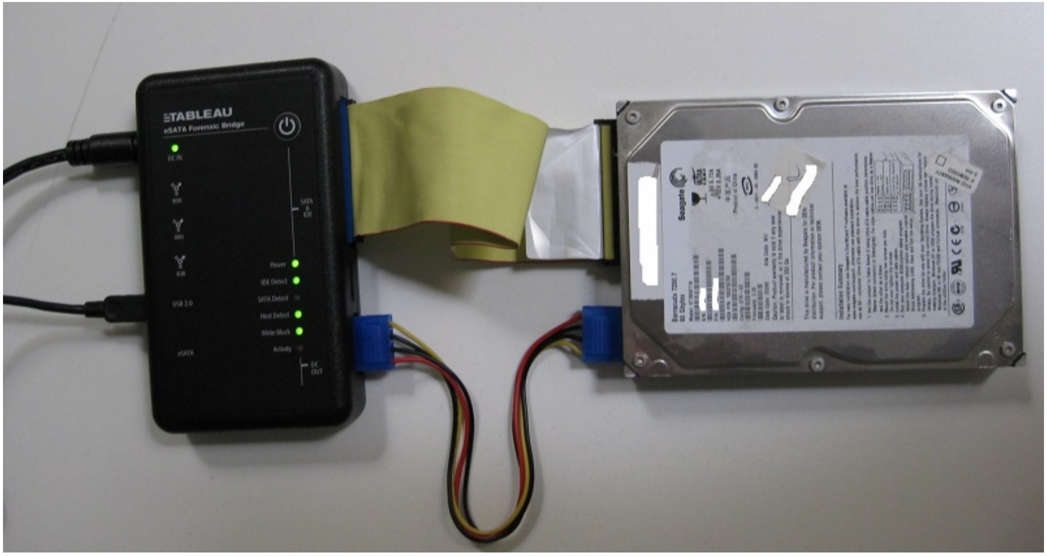


Figure 7: CAINE distro

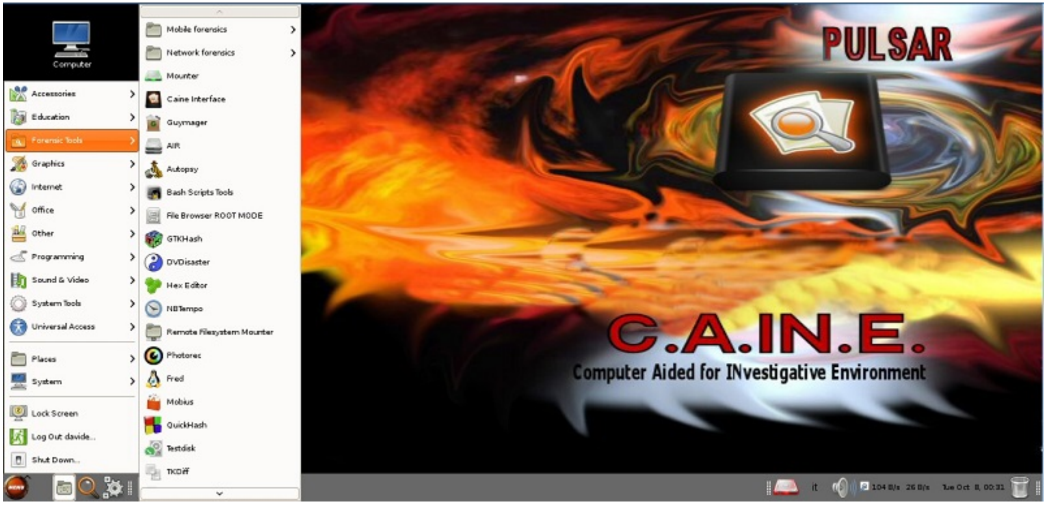


Figure 8: UFED acquisition support

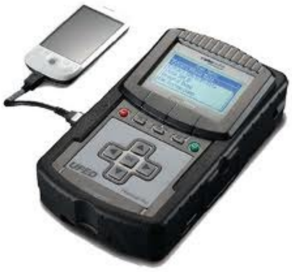
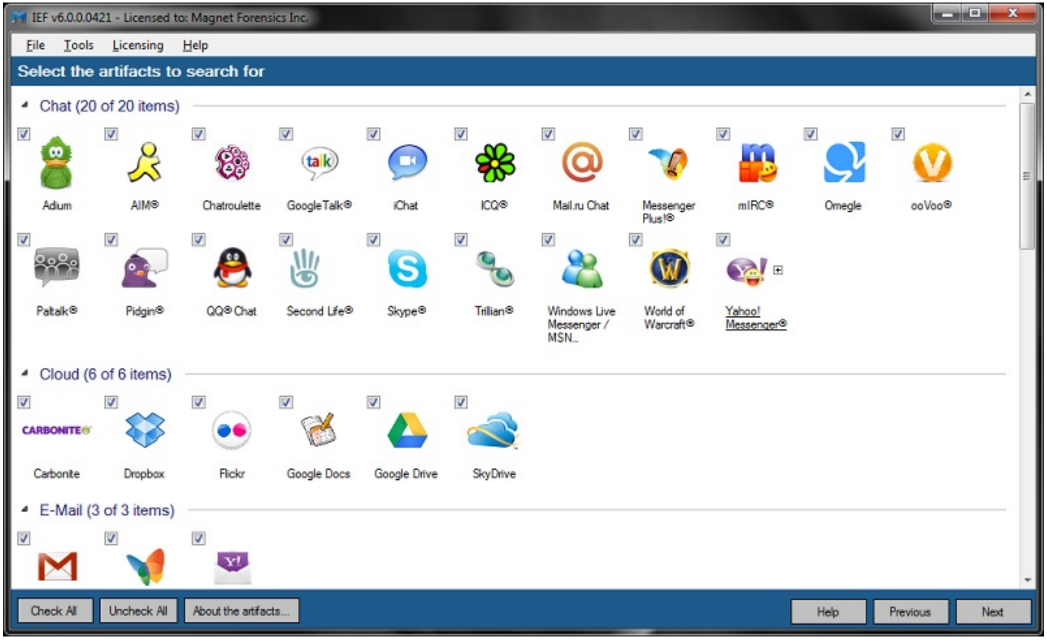


Figure 9: Internet Evidence Finder search options


IEF v6.0.0.0421 - Licensed to: Magnet Forensics Inc.
File Tools Licensing Help
Check All
Uncheck All
About the artifacts...
Help
Previous
Next

Select the artifacts to search for

Chat (20 of 20 items)

☑ Adium ☑ AIM® ☑ Chatroulette ☑ GoogleTalk® ☑ iChat ☑ ICQ® ☑ Mail.ru Chat ☑ Messenger Plus!® ☑ mIRC® ☑ Omegle ☑ ooVoo®

☑ Paltalk® ☑ Pidgin® ☑ QQ® Chat ☑ Second Life® ☑ Skype® ☑ Trillian® ☑ Windows Live Messenger / MSN... ☑ World of Warcraft® ☑ Yahoo! Messenger®

Cloud (6 of 6 items)

☑ Carbonite ☑ Dropbox ☑ Flickr ☑ Google Docs ☑ Google Drive ☑ SkyDrive

E-Mail (3 of 3 items)

☑ ☑ ☑

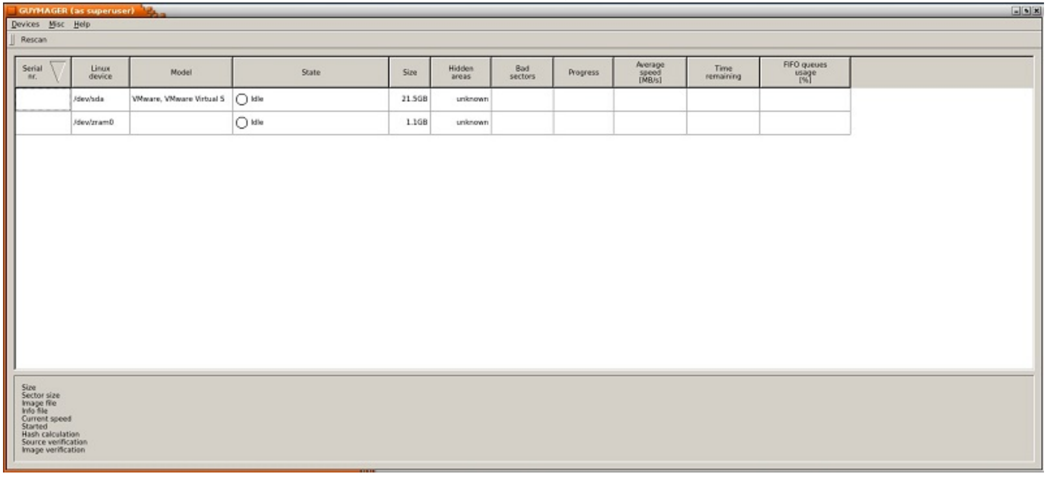
During this phase we have to overcome a small inconvenience: DAN-HDE01 cannot be acquired using writeblocker because with USB, plug alimentation is too lower for an External Hard Disk of this type. Instead of using a software writeblocker we prefer performing acquisition through Caine with Guymager(Figure 4.4.3). As of smartphone, tablet and cellular phone we use Oxygen Forensic Software and UFED Cellbrite. These two tools support most of the devices actually used, including Chinese chip based phone. UFED is very useful because is provided with all original cable to connect phone to workstation (Figure11).

Here the approach is a little bit difficult because we have to ensure that no communication channel can reach the device. Other consultants use jammer, which in Italy is illegal because it fills public radio frequencies, or faraday cage. Yarix has a lab, constructed under the guidelines of the NIST that is completely isolated from outside. In this proceeding all the devices arrived in off-mode, so we can easily remove SIM card.
Problems begin when we realize that some old cellular phones cannot be accessed without SIM card inserted, and also we do not know PIN number of most of the SIM cards. The solution is to insert a SIM that Yarix use as support, it is not active to ensure a further security level in addition to Faraday-cage lab. Meanwhile we communicate with lay entity to retrieve pin numbers, but this is not an operation of primary importance considering limited capacity of SIM cards to store and manage multimedia data.

Also in this step each activity that involves moving and connecting devices is photographically documented.

At the end of each acquisition is performed a comparison between has calculated on the original device and the hash of the evidence, the process is acceptable if the two hash

Figure 10: GuyMager interface



Serial nr.	Linux device	Model	State	Size	Hidden areas	Bad sectors	Progress	Average speed [MB/s]	Time remaining	FIFO queues usage [%]
	/dev/sda	VMware, VMware Virtual S	○ Idle	21.5GB	unknown					
	/dev/zram0		○ Idle	1.1GB	unknown					

Size
Sector size
Image file
Info file
Current speed
Started
Hash calculation
Source verification
Image verification

Figure 11: UFED cable set

match, otherwise acquisition procedure has to be redone, Figure 4.4.3 illustrates an example. Within EnCase this control is automatic executed and for each device gives positive result, so analysis stage can be started.

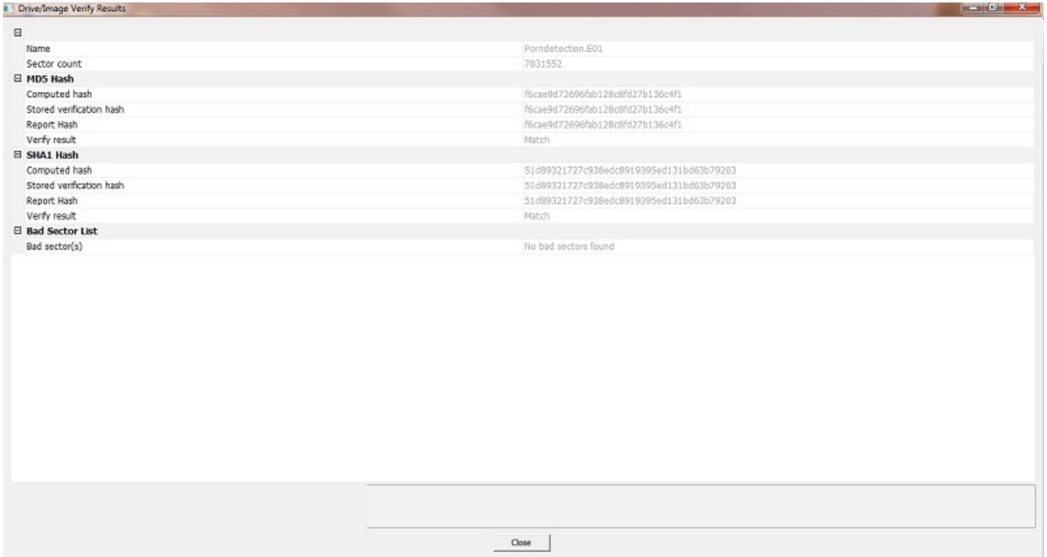
4.5 Analysis

The primary objective that examination has to reach is to determine if Alice pictures or videos are within Dan, Charlie and Bob's files. There are a large amount of bites to investigate, so a manual inspection to all pictures is impossible and not the clever approach. The best way is to use information already available, with permission of the investigating magistrate, we achieve a copy of pictures that Alice previously said to be sent. With these pictures we compute hash digest and histogram, so we can do the following researches:

- Determine if there are the same files inside seized devices;
- Restrict search area with Exif (Exchangeable image file format) metadata[5], especially creation date and time. Table 5 reports some EXIF metadata
- Text software develop in section 8.

[5] http://www.cipa.jp/english/hyoujunka/kikaku/pdf/DC-008-2010_E.pdf

Figure 12: Hash verification step



Drive/Image Verify Results

Name	Porndetection.E01
Sector count	7831552
MD5 Hash	
Computed hash	f6cae9d72696fab128c8fd27b136c4f1
Stored verification hash	f6cae9d72696fab128c8fd27b136c4f1
Report Hash	f6cae9d72696fab128c8fd27b136c4f1
Verify result	Match
SHA1 Hash	
Computed hash	51d89321727c938edc8919395ed131bd63b79203
Stored verification hash	51d89321727c938edc8919395ed131bd63b79203
Report Hash	51d89321727c938edc8919395ed131bd63b79203
Verify result	Match
Bad Sector List	
Bad sector(s)	No bad sectors found

Table 5: Some Exif metadata

tag
Make
Model
ImageSize
CreateDate
ModifyDate
DateTimeOriginal
CreatingApplication
GPS Information
ColorSpace
Contrast
Saturation
Sharpness
FileModifyDate
FileName
FileSize
FileSource
FileType
FirmwareVersion

With Alice's account name we can direct inspections on social network activity done by Dan, Charlie and Bob. At the same time each element denoting an ill-judged behaviour must be reported and further study in deep, focusing attention on sharing and creation on child-abuse material, hence analysis must cover these topics:

- Retrieve all multimedia files and pictures;
- Recover the most possible artefacts left by file sharing and peer-to-peer programs;
- Analyze in depth deleted items;
- Reconstruct the use of social network and instant messaging;
- Explanation of contacts between case's actors.

To do this the help of software like Internet Evidence Finder is invaluable, especially with such amount of evidence, but to fully understand traces meaning the underlying principles must be well learning and logically we face some issues that are reported and unfolded in subsequent section, such as analysis of Internet messaging systems, instant messaging activity, social network use, forensic examination of mobile devices.

In this particular case the timeline aspect does not play an important role.

4.6 Presentation

Results achieved during analysis stage are then presented with respect to investigators requests. In this stage the more the consultant is clear, the more is appreciated, so we decide upon a survey template which clearly sets out all activities by means of the following structure:

1. Introduction part, including a copy of investigative mandate;
2. A list of received items from Carabinieri;
3. For each find:
   - a The chain of custody;
   - b A brief description of appearance to report any damages, tampering or to simply document the condition of discovery. All of these considerations are supported by photographic documentation.
   - c A full description of the acquisition process.
   - d A complete review of items found inside;
   - e A list of items that may be useful for further proceedings by the judicial authority;
   - f A detailed explanation of relevant evidence for this case proceedings;

This report has to be done for each suspected.

We have to keep in mind that forensic consultant has to found its conclusions to technical reasoning and the more results are logically proven, the more conclusions are resistant. Personal opinions have to be eliminated from the finally survey and finally digital forensic consultant should eventually be able to bear his methodology inside a court room.

5 iOS Analysis

HFS+ File system

This is the core of every iOS based device. It is a dynamic file system built above a 512 byte block scheme, each block can be of two types: logical or allocation. Logical block are numbered from the volume starting point to the end and they are a static structure. Allocation block can be clustered to improve HFS efficiency.

File system logical structure is defined as:

Volume header stored inside sectors 0 and 1 it is utilized to contain information about the structure of the HFS volume and it has a backup automatic created in the last 1024 bytes of the volume. This backup is thought to permit automatic restore, but it can be very useful for forensic purposes.

Startup file.

Allocation file maintain a map of which blocks are used or are free. This file is not necessarily stored contiguously within the volume.

Attributes file.

Extents overflow file lists all extents used by a file and its allocation blocks using a B+-tree to guarantee the proper order.

Catalog file is the description of folder and files hierarchy located in the volume, it has a primary forensic relevance since it store all the metadata about all files including modified, access and created times. Each file created has a unique ID number that refers to the B+-tree used by catalog file.

iOS partitions

Basically common iOS devices use two partitions.

The first one include only firmware and it is read-only unless a firmware update is being performed. iTunes has the capability to write inside it in order to upgrade firmware version. Its dimension depend on the size of the memory and usually falls between 0.9 and 2.7 GB and contain only system data, upgrade information and basic applications. No user data are stored inside this space.

The second partition is the most interesting one because it contains all user data provided by applications and user's profile.

SQL lite databases

SQL lite is a relational and transactional database management systems that can be fully contained inside small programming library. It permits most of the SQL-92 standard operations with exception to some forms of multithreading and concurrent access, but it has not a separate server process, and the overhead to retrieve data and perform query is very low.

Hence this kind of databases are commonly employed inside app development, including iOS app. Native iOS features, such as Calendar, SMS messaging system, Notes, Photo and contacts book utilize SQL lite databases.

This is a very important feature for forensic examination because common databases are saved without encryption and data inside are easily available by using a free viewers.

Plists files

Sometimes they are called property file, and every Plists file stores various metadata in an XML way.
They can contain various types such as string, dates, boolean and numbers. Apps that need to store and maintain configuration items use plists files. Sometimes they are saved in a binary format, but various viewers are accessible in the Internet, that convert binary objects to an human readable XML.

5.1 Acquisition

Facing with iOS devices acquisition could be very interesting. Basically one must be able to deal with various scenarios, for example you may not have physical device or it is locked and encrypted. Due to BYOD (Bring Your Own Device) trend, pass locked devices are increasing its presence either for company security policies or for user awareness of theft. Overcome code protection is not trivial and always feasible.

Recommended approach is to retrieve passcode from owner and disable it once devices is unlocked; another tip is to put device in “Airplane mode”, hence examination can proceed without worrying about connections isolation.

There are various way to perform an acquisition, the most widely used are: iTunes backup, logical API type method, jail breaking and physical image of the storage.

iTunes Backup

This is a popular approach when the device is not physically available and a backup is retrieved from other source. It should be noted that iTunes suite perform backup during the synchronization process and do it independently from user’s input unless iTunes configuration is changed, but usually normal users does not modify it because it is more convenient to use. By default backup can be found inside this location:

Table 6: iTunes backup location

O.S.	Path
Windows XP	`%UserProfile%\Application Data\Apple Computer\MobileSync\Backup`
Windows 7	`%UserProfile%\AppData\Roaming\Apple Computer\MobileSync\Backup`
MacOS	`Users/%username%/Library/Application Support/MobileSync/Backup`

An examiner must keep in mind that maybe there are several devices backup stored inside iTunes repository, so he has to check if under examination there is the correct one.

These information are provided by the status, info and manifest plist files located in the root directory:

- status.plist provides data about the latest backup.
- info.plist contains useful items to match backup with device, such as IMEI number and phone number.
- manifest.plist lists all metadata concerning selected backup files, for example it says if the Passcode was set or not.

It is worth noting that iTunes can encrypt backups, which obviously can lead to unfeasible time requirements.There are various tools that can run a password cracking utility versus manifest.plist , alternately, jail breaking approach can deal with some passcode protection but it is more pervasive since it needs to replace certain configuration files.

Logical Methods

Logical acquisition method is becoming day by day more popular, allowing the recovery of the allocated and active files using a iOS-native synchronization method. By such method analyst can gather SMS, call logs, calendar events, contacts, photos, web history, email accounts and social network activities.

The main drawback is that this method does not allow slack space access, so if an evidence is suspected to be deleted or in slack space a physical acquisition is necessary.

Jail Breaking

Jail break basically replace the firmware with a hacked version that allows install and execute tools out of the so-called sandbox that iOS provide to ensure security policies. For example with jail break an examiner can take root privileges and run utility like SSH and terminal.

Most popular jail breaking framework is redSn0w which replace the original firmware with Cydia package through a simple wizard, once the device has completed the process can be accessed by SSH connection via wireless network, simply typing in a terminal:
`ssh root@<device ip>`.

For iOS devices using hardware encryption, such as iPhone4S or iPhone5, the resulting disk image will be encrypted and it is necessary a physical acquisition.

Physical Methods

As for other devices, the best forensic scenario is to obtain a bit-by-bit copy of the storage media in order to act as usual, like in a computer forensic situation.
This is not so simple as it can appear, but is continuously researched since it is the powerful way to recover artefacts.

The challenge to deal with is the impossibility to access and remove the memory medium since it is embedded with different layouts between various models, an iPhone4 has a different design with respect to an iPhone 3 or an iPad. This aspect also applies to security model of iOS operating system. The first method presented [6], actually available only to Law Enforcement for author's mind, is the most pervasive and replace the RAM software with a version that allows execution of a live recovery agent capable of disk imaging.

For users that are not Law Enforcement there are other, priced, tools available.

UFED Cellbrite

This is the most used commercial tool in mobile forensic domain. It provides essentially three acquisition level:

1. Physical Extraction;
2. File System Extraction;
3. Logical Extraction;

Physical extraction acts under the following , overall process:

1. Infiltrate the device to inject the code bypassing lock.
2. Inject code into mobile RAM.
3. Execute the code to read flash memory.
4. Transfer the data from the device to the analysis workstation.
5. Leave no trace.

Obviously code are not public available, but company website says that:

> UFED can extract a bit-by-bit copy of the entire flash memory of a mobile device also data that is hidden or has been deleted.

Second level that is file system extraction gives access to all the files present in storage that are correctly indexed by file system. In this way deleted files are not available except from what is store inside sqlite databases. Recover of database deleted entries is possible due to sqlite management system. Logical extraction exploits well known set of commands whom allow communication between data extraction tools and device's operating system. This approach is the easier and quicker one, but is not available for locked devices and has a limited extraction capacity however is enough in most cases where the focus is on call logs, phonebook entries and SMS.

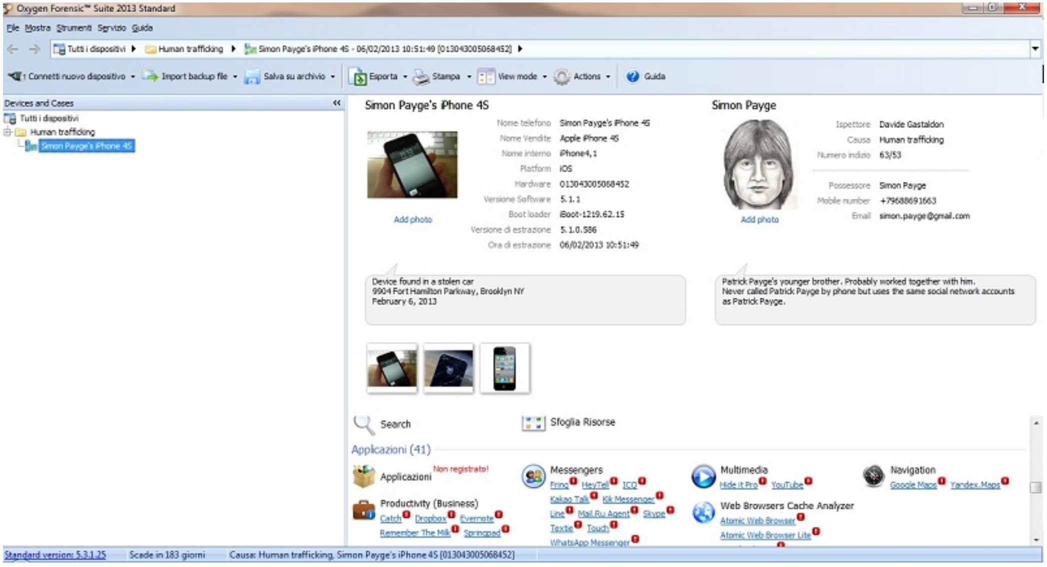
[6] http://www.zdziarski.com/blog/?page_id=150

Oxygen Forensic Suite

Oxygen uses a logical extraction approach improved by the use of a proprietary protocol combined with phone APIs. In particular with iDevice it use iTunes to gain access of device and then it can extract all addressed data.

This suite is widely used thanks to its simply interface and the automation of common tasks via wizard interface that make it suitable for non specialized operators like Law Enforcement units. It is less pervasive than physical extraction but enough powerful.
Figure 13 represent the User Inteface with free license and demo iPhone given by manufacturer to train.

Figure 13: Oxygen Forensic Suite Interface


Oxygen Forensic™ Suite 2013 Standard
File Mostra Strumenti Servizio Guida
Tutti i dispositivi ▸ Human trafficking ▸ Simon Payge's iPhone 4S - 06/02/2013 10:51:49 [013043005068452] ▸
Connetti nuovo dispositivo Import backup file Salva su archivio Esporta Stampa View mode Actions Guida
Devices and Cases
Tutti i dispositivi
Human trafficking
Simon Payge's iPhone 4S
Add photo
Add photo
Search
Sfoglia Risorse
Applicazioni (41)
Applicazioni Non registrato!
Messengers
Fring HeyTell ICQ Kakao Talk Kik Messenger Line Mail.Ru Agent Skype Textie Touch WhatsApp Messenger
Multimedia
Hide it Pro YouTube
Navigation
Google Maps Yandex.Maps
Productivity (Business)
Catch Dropbox Evernote Remember The Milk Springpad
Web Browsers Cache Analyzer
Atomic Web Browser Atomic Web Browser Lite

Simon Payge's iPhone 4S

Nome telefono Simon Payge's iPhone 4S
Nome Vendite Apple iPhone 4S
Nome interno iPhone4,1
Platform iOS
Hardware 013043005068452
Versione Software 5.1.1
Boot loader iBoot-1219.62.15
Versione di estrazione 5.1.0.586
Ora di estrazione 06/02/2013 10:51:49

Device found in a stolen car
9904 Fort Hamilton Parkway, Brooklyn NY
February 6, 2013

Simon Payge

Ispettore Davide Gastaldon
Causa Human trafficking
Numero indizio 63/53
Possessore Simon Payge
Mobile number +79688691663
Email simon.payge@gmail.com

Patrick Payge's younger brother. Probably worked together with him.
Never called Patrick Payge by phone but uses the same social network accounts as Patrick Payge.

Standard version: 5.3.1.25 Scade in 183 giorni Causa: Human trafficking, Simon Payge's iPhone 4S [013043005068452]

5.2 Analysis

iOS provide a complete set of time information such as modified, accessed, changed and born date mainly reported in CF Absolute Time, which means the number of seconds since the reference date set to `1/1/2001`. So it is necessary translate these times by the following formula

$$Localtime = \frac{CfAbsoluteTime}{60*60*24} + Date(2001,1,1)$$

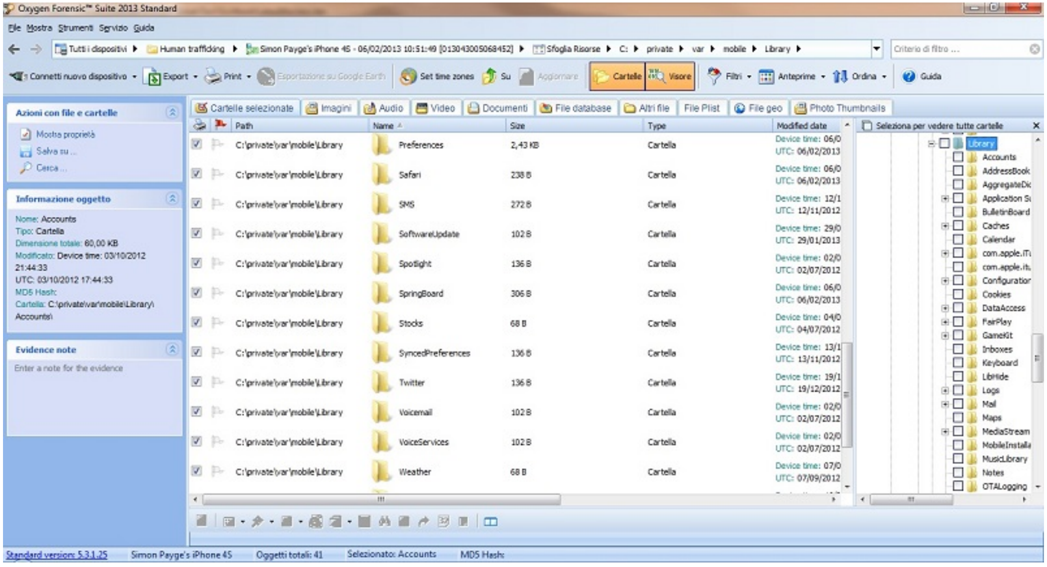
The iOS directory schema is common across al iOS devices and it is derived from UNIX layout so there are directory that are more interesting than other, Figure 14 is an example of how it appears under Oxygen.

Limiting to default application the following location needs attention:

iTunes App

Applications delivered by iTunes store automatically create a new directory located in `private/var/mobile/Application` folder inside which are stored all associated

Figure 14: iOS Logical File System


Oxygen Forensic™ Suite 2013 Standard
File Mostra Strumenti Servizio Guida
Tutti i dispositivi ▶ Human trafficking ▶ Simon Payge's iPhone 4S - 06/02/2013 10:51:49 [013043005068452] ▶ Sfoglia Risorse ▶ C: ▶ private ▶ var ▶ mobile ▶ Library ▶
Criterio di filtro ...
Connetti nuovo dispositivo
Export
Print
Esportazione su Google Earth
Set time zones
Su
Aggiornare
Cartelle
Visore
Filtri
Anteprime
Ordina
Guida
Azioni con file e cartelle
Mostra proprietà
Salva su ...
Cerca ...
Cartelle selezionate
Imagini
Audio
Video
Documenti
File database
Altri file
File Plist
File geo
Photo Thumbnails
Selezione per vedere tutte cartelle

Informazione oggetto

Nome: Accounts
Tipo: Cartella
Dimensione totale: 60,00 KB
Modificato: Device time: 03/10/2012 21:44:33
UTC: 03/10/2012 17:44:33
MD5 Hash:
Cartella: C:\private\var\mobile\Library\Accounts\

Evidence note

Enter a note for the evidence

		Path	Name	Size	Type	Modified date
☑		C:\private\var\mobile\Library	Preferences	2,43 KB	Cartella	Device time: 06/0 UTC: 06/02/2013
☑		C:\private\var\mobile\Library	Safari	238 B	Cartella	Device time: 06/0 UTC: 06/02/2013
☑		C:\private\var\mobile\Library	SMS	272 B	Cartella	Device time: 12/1 UTC: 12/11/2012
☑		C:\private\var\mobile\Library	SoftwareUpdate	102 B	Cartella	Device time: 29/0 UTC: 29/01/2013
☑		C:\private\var\mobile\Library	Spotlight	136 B	Cartella	Device time: 02/0 UTC: 02/07/2012
☑		C:\private\var\mobile\Library	SpringBoard	306 B	Cartella	Device time: 06/0 UTC: 06/02/2013
☑		C:\private\var\mobile\Library	Stocks	68 B	Cartella	Device time: 04/0 UTC: 04/07/2012
☑		C:\private\var\mobile\Library	SyncedPreferences	136 B	Cartella	Device time: 13/1 UTC: 13/11/2012
☑		C:\private\var\mobile\Library	Twitter	136 B	Cartella	Device time: 19/1 UTC: 19/12/2012
☑		C:\private\var\mobile\Library	Voicemail	102 B	Cartella	Device time: 02/0 UTC: 02/07/2012
☑		C:\private\var\mobile\Library	VoiceServices	102 B	Cartella	Device time: 02/0 UTC: 02/07/2012
☑		C:\private\var\mobile\Library	Weather	68 B	Cartella	Device time: 07/0 UTC: 07/09/2012

- ☐ Library
  - ☐ Accounts
  - ☐ AddressBook
  - ☐ AggregateDic
  - ☐ Application Su
  - ☐ BulletinBoard
  - ☐ Caches
  - ☐ Calendar
  - ☐ com.apple.iTu
  - ☐ com.apple.itu
  - ☐ Configuration
  - ☐ Cookies
  - ☐ DataAccess
  - ☐ FairPlay
  - ☐ GameKit
  - ☐ Inboxes
  - ☐ Keyboard
  - ☐ LibHide
  - ☐ Logs
  - ☐ Mail
  - ☐ Maps
  - ☐ MediaStream
  - ☐ MobileInstalla
  - ☐ MusicLibrary
  - ☐ Notes
  - ☐ OTALogging

Standard version: 5.3.1.25 Simon Payge's iPhone 4S Oggetti totali: 41 Selezionato: Accounts MD5 Hash:

files.

Commonly each application creates a new folder subtree typically structured as follows:

1. Document folder for relevant files to that application.
2. Temp folder for temporary runtime files.
3. Library folder for preferences, cached data.

Examiner, if he is lucky, may find username and password data, cookies or pictures that will help provide evidence for the investigation.

Photos

Photos taken rather synced to the device are stored under private
`/var/mobile/media/DCIM` folder, further divided in `100APPLE`, which contains photos taken with device camera.
Examiner should know that images are numbered sequentially starting from `IMG_0001`, without alter the count if a item is deleted or moved; this can be a good way to understand if all the taken photos are available or if some items was previously deleted.

Since iOS devices have the ability to take screenshots of itself, another folder to consider is `999APPLE` in which these files are stored.

Keystrokes

All users during normal utilization insert via keyboard words that not always exists inside preconfigured dictionary.
ny word prompted in Notes, Safari, Messages, Facebook or any application that permits text input will be captured and stores inside a text file named `dynamic-text.dat` located

in private/var/mobile/Library/Keyboard.
The original goal would be to facilitate user input activity but from a forensic point of view this can be fashionable to narrow the set of frequently typed strings and can be thought as a starting point for keywords searches. In addition to this a good analysis can uncover interesting or special words, like acronyms or technical terms, that could be useful for the inquiry.
Examiner should consider that timestamps are not recorded, so any word could have been typed at any time during device life.

Passwords

Apple natively provides a keychain management system defined as[7]

> "A keychain is an encrypted container that holds passwords for multiple applications and secure services."

Further deepening states:

> "Each keychain item contains data plus a set of attributes. For a keychain item that needs protection, such as a password or private key (a string of bytes used to encrypt or decrypt data), the data is encrypted and protected by the keychain. For keychain items that do not need protection, such as certificates, the data is not encrypted."

iOS support this mechanism by using the key-chain-2.db file located at
/private/var/Keychains
always up to date with passwords and accounts used such as wireless access point key phrases, login passcodes and son on.

Not always items are stored in clear, more often they are encrypted by default with iOS procedure but there are several, non open source, tools that can reconstruct original password.

Notes

As everyone knows the easy way to remember useful information, think about passwords, credit cards number, PIN and so on, is write them.
This usually leads to the creation of a set of notes containing such items that in iOS are stored inside /private/var/mobile/Libaray/Notes with a sqlite database.

Text Messages

SMS are one of the most important things to recover and in iOS environment this can be easily done by accessing sms.db sqlite database, stored at /private/var/mobile.
This database, despite what is commonly believed, can hold also deleted messages and conversation.

[7] http://www.apple.com/iphone/business/it/security.html

Whatsapp Messenger

Whatsapp is becoming the most used messaging application for smartphone in the European area. It can be run on Android, iOS and Blackberry. Also there are porting projects for Symbian and other Nokia based Operating Systems.

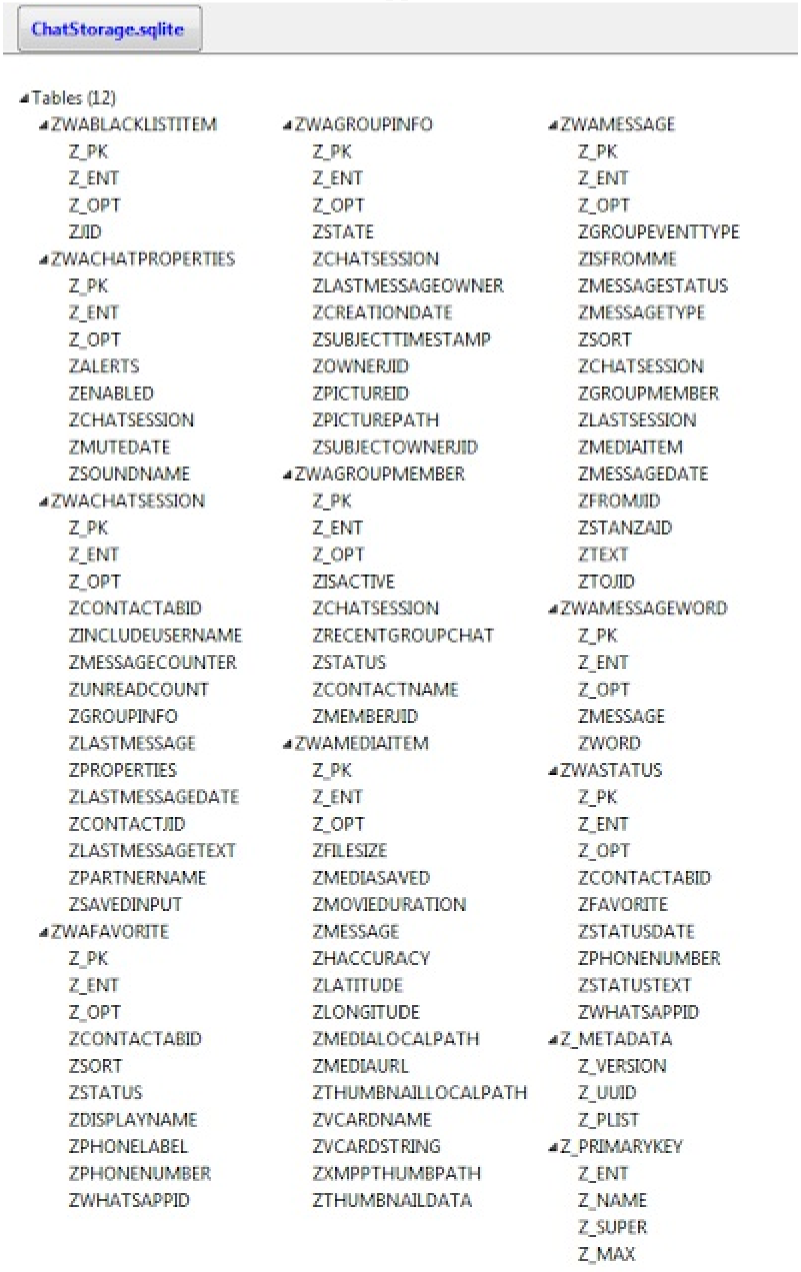
On iOS interesting files can be found inside `Application/net.whatsapp.WhatsApp/Documents/ChatStorage.sqlite`. Database is made of 12 tables explained in Figure 15.

Interesting evidence can be found by browsing these tables: `ZWACHATSESSION` and `ZWASTATUS` have the contacts, while `ZWAMESSAGE` and `ZWAMEDIAITEM` store messages and attachments.

Unlike Android version, iOS Whatsapp version does not encrypt this database, it only use hardware encryption provide by manufacturer, so it is possible to extract all logs after a physical acquisition, moreover, there is a python free tool,`Whatsapp_Xtract` [8], that automatically create a report.

[8] `https://code.google.com/p/hotoloti/downloads/detail?name=Whatsapp_Xtract_V2.0_2012-05-02.zip&can=2&q=`

Figure 15: Whatsapp database structure



```
ChatStorage.sqlite

◢Tables (12)
  ◢ZWABLACKLISTITEM
     Z_PK
     Z_ENT
     Z_OPT
     ZJID
  ◢ZWACHATPROPERTIES
     Z_PK
     Z_ENT
     Z_OPT
     ZALERTS
     ZENABLED
     ZCHATSESSION
     ZMUTEDATE
     ZSOUNDNAME
  ◢ZWACHATSESSION
     Z_PK
     Z_ENT
     Z_OPT
     ZCONTACTABID
     ZINCLUDEUSERNAME
     ZMESSAGECOUNTER
     ZUNREADCOUNT
     ZGROUPINFO
     ZLASTMESSAGE
     ZPROPERTIES
     ZLASTMESSAGEDATE
     ZCONTACTJID
     ZLASTMESSAGETEXT
     ZPARTNERNAME
     ZSAVEDINPUT
  ◢ZWAFAVORITE
     Z_PK
     Z_ENT
     Z_OPT
     ZCONTACTABID
     ZSORT
     ZSTATUS
     ZDISPLAYNAME
     ZPHONELABEL
     ZPHONENUMBER
     ZWHATSAPPID
  ◢ZWAGROUPINFO
     Z_PK
     Z_ENT
     Z_OPT
     ZSTATE
     ZCHATSESSION
     ZLASTMESSAGEOWNER
     ZCREATIONDATE
     ZSUBJECTTIMESTAMP
     ZOWNERJID
     ZPICTUREID
     ZPICTUREPATH
     ZSUBJECTOWNERJID
  ◢ZWAGROUPMEMBER
     Z_PK
     Z_ENT
     Z_OPT
     ZISACTIVE
     ZCHATSESSION
     ZRECENTGROUPCHAT
     ZSTATUS
     ZCONTACTNAME
     ZMEMBERJID
  ◢ZWAMEDIAITEM
     Z_PK
     Z_ENT
     Z_OPT
     ZFILESIZE
     ZMEDIASAVED
     ZMOVIEDURATION
     ZMESSAGE
     ZHACCURACY
     ZLATITUDE
     ZLONGITUDE
     ZMEDIALOCALPATH
     ZMEDIAURL
     ZTHUMBNAILLOCALPATH
     ZVCARDNAME
     ZVCARDSTRING
     ZXMPPTHUMBPATH
     ZTHUMBNAILDATA
  ◢ZWAMESSAGE
     Z_PK
     Z_ENT
     Z_OPT
     ZGROUPEVENTTYPE
     ZISFROMME
     ZMESSAGESTATUS
     ZMESSAGETYPE
     ZSORT
     ZCHATSESSION
     ZGROUPMEMBER
     ZLASTSESSION
     ZMEDIAITEM
     ZMESSAGEDATE
     ZFROMJID
     ZSTANZAID
     ZTEXT
     ZTOJID
  ◢ZWAMESSAGEWORD
     Z_PK
     Z_ENT
     Z_OPT
     ZMESSAGE
     ZWORD
  ◢ZWASTATUS
     Z_PK
     Z_ENT
     Z_OPT
     ZCONTACTABID
     ZFAVORITE
     ZSTATUSDATE
     ZPHONENUMBER
     ZSTATUSTEXT
     ZWHATSAPPID
  ◢Z_METADATA
     Z_VERSION
     Z_UUID
     Z_PLIST
  ◢Z_PRIMARYKEY
     Z_ENT
     Z_NAME
     Z_SUPER
     Z_MAX
```

6 Windows

This is a short introduction given by a forensic perspective, obviously before a real investigation knowledge must be improved [3].

6.1 NTFS overview

Windows develops NTFS file system to overcome inherent limitations in FAT, as for many other products is a proprietary technology which has undergone an extensive reverse engineering analysis that leads to the development of many free open source drivers.

NTFS supports 255 unicode chars long name file, giving up DOS 8.3 approach. Theoretical maximum size is declared to be 16TB against FAT limit to 4GB and also each volume can reach 256TB with a maximum number of file plus than 4 billion.

Reliability of NTFS is ensured by the use of a journaled technique, which assure that in case of crash only current operations are lost, with no effect to the integrity of the remaining file system.

It permits a major level of granularity in control police, each file and folder has its own access policy as well as different permissions in read, write and execute.

Other changes introduced by NTFS are the use of index to speed-up search operations, the support to cryptography and compression.

It uses a linux-based approach in which every file system's item is a file and the only mandatory sector is the first which contains boot information.

6.2 Pro and Con

As for any system windows has a set of pro and con, in this chapter the analysis of a windows environment is generally presented and later deeper issues are exploited basically based on problems faced during the investigations carried out for the resolution of the real case.

From the point of view of a forensic consultant we can state these pro items:

- Windows is the most used operating system so it is well documented.
- It is well supported by tools.

At the same time Microsoft leading operating system has some drawbacks:

- Default logging level is very low, this implies manually research inside artefacts left by user operations;
- Core evidences are saved in binary format so it is necessary some reverse engineering or sometimes the use of ad-hoc software with not known reliability degree.

- If the file system in use is FAT and in the systems there are many users it is difficult to determine a correspondence between actions and users. This is because there are few metadata in the FAT architecture. Although FAT is an old file system, is not so remote the likelihood analysis.
- The great part of windows machines has account with administrative levels, especially those for private use. From the user side this consideration leads to security issues, for the forensics analyst implies further work to establish that operations are done with user acknowledgment and not as a result of security holes activities, i.e. virus, spyware, Trojan, malware and so on. Since Vista release, Microsoft try to address this problems with UAC (user account control), which ask the user to confirm operations whenever a system-level change is made, but the majority of users disable this control because it is too much pervasive.
- Need of third part security suite. This dispute is linked to the previous point, almost every windows environment has a third part antivirus or antispyware that inside scanning operations reset file's Access Time field and this is reflected upon timeline reconstruction.
- Large number of available commercial tools. Windows has a lot of developers producing tools for all user needs, starting from system maintenance arriving to Computer Aided Design environment. This leads to a lot of proprietary file formats requiring a lot of viewers and in the worth case the hexadecimal analysis. In the case of judicial proceedings this is one of the most important challenge to face, because at the end a consultant has to persuade law actors of the result's correctness.

According to [12] in a Windows environment common places to search evidence are the following:

- Non persistent data inside kernel structures;
- Slack space , where you can obtain information about previously deleted files;
- Obviously the logical file system;
- As trivially what is marked as free space;
- What system save to log event;
- Registry that is the core of windows;
- Logs of specific applications that are not managed by Windows Event Log Service;
- Pagefile.sys that is the swap file for RAM on the active partition;
- Application-levels files, such as browsers' cache and cookies or locally stored databases;
- Temporary files created by many applications;
- Recycle Bin or rather the hidden logical file structure where system send recently deleted items;
- Printer spool;

- Backup or replica copy of email clients.

Obviously not always it is necessary a complete examination of aforementioned point, survey strategy depends on what investigative mandate says rather than examination deadline.

At least there are three common insights to do:

1. Registry analysis;
2. File carving including hidden and deleted items;
3. Browsers history.

6.3 Registry

Windows registry is the core structure in which operating system store settings and options and it is structured as a hierarchical database designed as a binary tree with this layout:

- **HIVE**
  - **Subtrees**
    - **Subkeys**
      - **Keys**

HIVE are the following:

Table 7: Registry Hive

HKEY_CLASSES_ROOT	contains two kind of data: Association between file type and application aside from configuration data for core level applications and components, like COM, scripting, programming languages.
HKEY_CURRENT_USER	that is a link to the currently logged-in user located in \HKEY_USERS\[SID OF CURRENT USER].
HKEY_LOCAL_MACHINE	tores information about computer configuration, including hardware data, operating system state, system bus, drivers, boot and startup parameters.
HKEY_USERS	ontains users' profile subtrees, indexed by SID number.
HKEY_CURRENT_CONFIG	is a pointer to a subtree stores in HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet \Hardware Profiles\Current.

Registry is physically stored between different file, most interesting are listed in Table 8.

When the system create a Restore point it saves also the registry state.
Such copies can be found in folders called `RPxx` inside
`%MainDrive%\System Volume Information\_restorexx`.

Properly analyzed registry can reveal a lot of useful information such as:

- Recent files.

Table 8: Registry Files

NTUSER.DAT	located in `%UserProfile%` directory and mapped to `HKEY_CURRENT_USER`.
SECURITY	located in `%SystemRoot%\Sytem32\config\` and linked to `HKEY_LOCAL_MACHINE\SECURITY`.
SYSTEM	located in `%SystemRoot%\Sytem32\config\` and mapped to `HKEY_LOCAL_MACHINE\SYSTEM`.
SOFTWARE	located in `%SystemRoot%\Sytem32\config\` and linked to `HKEY_LOCAL_MACHINE\SOFTWARE`

- User settings.
- Application path.
- USB drive previously connected.

The last point is interesting since it allows to make an inventory of previously connected devices, but also to compare available devices with previously connected to check if a specific USB drive was previously linked. It is worth noting that, each registry key has associated the last modified value and this can be used to reconstruct registry utilization in order to make a reliable timeline.

6.4 Data Carving

It is a common task the recovering of the whole set of a given file type, for example retrieve all the pictures, also hidden and deleted, from a finding so it is important the full understanding of files management system to perform correct operations of carving.

Usually in a Windows Environment to each file is associated a name and extension, a type and a dimension. Name identifies the file and permit its referencing meanwhile type explicit its content kind summarized in extension that was designed to help association with correct viewer. Modifying extension implies that viewer association change, and it is an easy task very often performed by ordinary users when they try to hidden something.

Again is well known that logical delete does not remove file content, but only references inside file system structures, so until physical sectors are not overwritten content is still available also if the file is not indexed. Also formatting methods simply deletes file system structures without changing or removing content. There are various methods to permanently delete contents, both hardware and software based. Hardware based deletion makes the support unserviceable and needs specific tools, like degausser or punchings while software based secure deletion uses algorithms to overwrite with random data, which however is a high time consuming tasks, in average with a 500GB space it is necessary more or less 8 hours.

For these reasons data carving could lead to interesting results as it does not operate on extension or filename but on the so-called magic number, or rather the starting values representing the sign of the file, a concept inherited from Unix world in which there are no extension.

For example whatever JPEG picture starts with this signature: `0xFF 0xD8 0xFF`, or png file format with `0x89 0x50 0x4E 0x47 0x0D 0x0A 0x1A 0x0A`.

Sign is used as the search parameters, also in non allocated or slack space, and to discover trivial camouflages attempts.

More depth techniques use statistical analysis to recognize a known file structure and then perform content part analysis.

There are various tools to perform data carving that act under evidence mounted as a physical disk; the most used are foremost, photorec and scalpel. They use a config file in which are stated header and footer structure of file formats. The file can be enhanced with new signature.

6.5 Browser History

Inside a Windows environment it is common to detect various browsers belonging to the following set: Internet Explorer, Mozilla Firefox and Google Chrome.

Statistics claim that Internet Explorer is widely used, this is because is pre-installed and in enterprise environment many times is the only allowed browser. Mozilla Firefox and Google Chrome are commonly used in private environment, with a widespread circulation of Chrome, many thanks to Android mobile devices explosion.

All these ones use cache principle, they locally save web pages elements so that the next time the user visit the same page they do not need to download again the same graphics and pages. This can be a source of treasures for profile user web behaviour. Also cookies are very interesting to understand session, login and user stay inside specific website.

Internet Explorer

Microsoft browser makes use of `.dat` files, which are binaries formatted and they manage both chronology and cache. Paths are listed in table 9.

Table 9: Internet Explorer History paths

Item	Path
Histoy	`%UserProfile%\AppData\Local\Microsoft\Windows\History`
Cookies	`%UserProfile%\AppData\Roaming\Microsoft\Windows\Cookies`
Cache	`%UserProfile%\AppData\Local\Microsoft\Windows\Temporary Internet Files`

Conversely Forms AutoComplete and Password AutoComplete are stored inside the registry `HKEY_CURRENT_USER\Software\Microsoft\Internet Explorer\IntelliForms\` as explained by Table 10

Table 10: Internet Explorer Registry keys

Type	Key
Forms	`Storage1`
Passwords	`Storage2`

Please note that these are Windows 7 path, for a Windows XP analysis interesting items are stored under `%UserDirectory%`, that usually is `C:\Documents and settings\%username%`, under `Local Settings`.

Mozilla Firefox

This open source browser saves a lot of useful information, in the form of .sqlite databases, inside these directories.
The root folder is the following:
`%UserDirectory%\AppData\Local\Mozilla\Firefox\Profiles` and then there are subfolders listed in Table 11.

Table 11: Mozilla Firefox History paths

Item	Path
Cached Pages	`%RootFolder%\%some profile number%\Cache`
Form History File	`%RootFolder%\%some profile number%\formhistory.sqlite`
Password File	`%RootFolder%\%some profile number%\signons.sqlite`
Cookies	`%RootFolder%\%some profile number%\cookies.sqlite`

Google Chrome

Also Google browser uses .sqlite database to store data and they are located in:
`%UserDirecotry%\AppData\Local\Google\Chrome\User Data\Default\`.
Figure 16 shows what is store inside such folder.

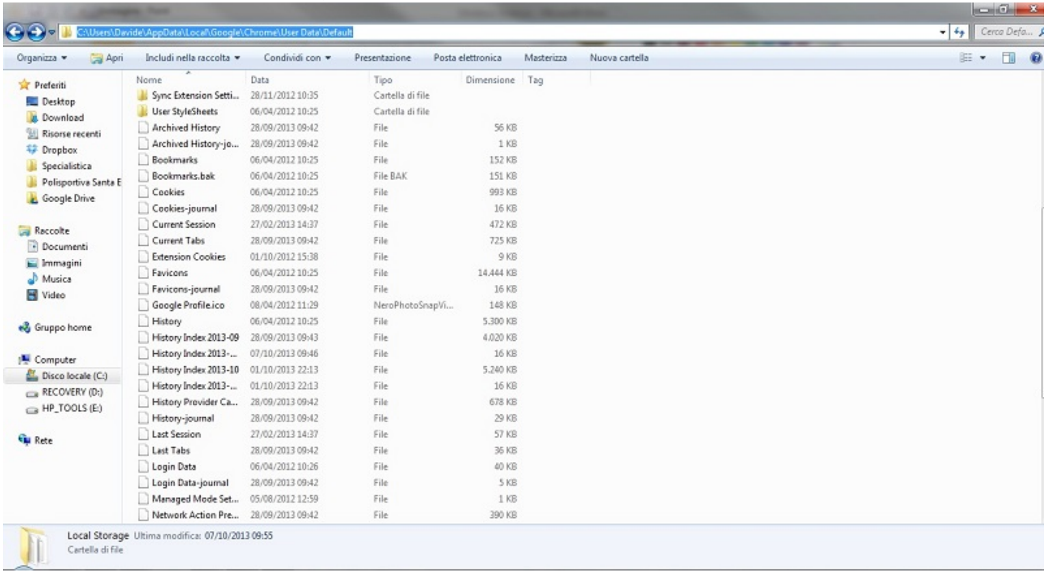
In Private Browsing features

All three browser offer the possibility to start a session in the so-called private mode. Internet Explorer InPrivate browsing was introduced in Internet Explorer 8 and official documentation states the follow[9]

> "InPrivate Browsing in Internet Explorer 8 helps prevent one's browsing history, temporary Internet files, form data, cookies, and usernames and passwords from being retained by the browser, leaving no easily accessible evidenceof browsing or search history. InPrivate Filtering provides users an added level of control and choice about the information that third party websites can use to track browsing activity. InPrivate

[9] `http://en.wikipedia.org/wiki/Internet_Explorer_8`

Figure 16: Google Chrome Temp Folder



C:\Users\Davide\AppData\Local\Google\Chrome\User Data\Default

Nome	Data	Tipo	Dimensione	Tag
Sync Extension Setti...	28/11/2012 10:35	Cartella di file		
User StyleSheets	06/04/2012 10:25	Cartella di file		
Archived History	28/09/2013 09:42	File	56 KB	
Archived History-jo...	28/09/2013 09:42	File	1 KB	
Bookmarks	06/04/2012 10:25	File	152 KB	
Bookmarks.bak	06/04/2012 10:25	File BAK	151 KB	
Cookies	06/04/2012 10:25	File	993 KB	
Cookies-journal	28/09/2013 09:42	File	16 KB	
Current Session	27/02/2013 14:37	File	472 KB	
Current Tabs	28/09/2013 09:42	File	725 KB	
Extension Cookies	01/10/2012 15:38	File	9 KB	
Favicons	06/04/2012 10:25	File	14.444 KB	
Favicons-journal	28/09/2013 09:42	File	16 KB	
Google Profile.ico	08/04/2012 11:29	NeroPhotoSnapVi...	148 KB	
History	06/04/2012 10:25	File	5.300 KB	
History Index 2013-09	28/09/2013 09:43	File	4.020 KB	
History Index 2013-...	07/10/2013 09:46	File	16 KB	
History Index 2013-10	01/10/2013 22:13	File	5.240 KB	
History Index 2013-...	01/10/2013 22:13	File	16 KB	
History Provider Ca...	28/09/2013 09:42	File	678 KB	
History-journal	28/09/2013 09:42	File	29 KB	
Last Session	27/02/2013 14:37	File	57 KB	
Last Tabs	28/09/2013 09:42	File	36 KB	
Login Data	06/04/2012 10:26	File	40 KB	
Login Data-journal	28/09/2013 09:42	File	5 KB	
Managed Mode Set...	05/08/2012 12:59	File	1 KB	
Network Action Pre...	28/09/2013 09:42	File	390 KB	

Local Storage Ultima modifica: 07/10/2013 09:55
Cartella di file

> Subscriptions allow you to augment the capability of InPrivate Blocking by subscribing to lists of websites to block or allow." “As with other private browsing modes there are ways that information about a browsing session can be recovered.”

While Google official help says that[10]

> "Browsing in incognito mode only keeps Google Chrome from storing information about the websites you’ve visited. The websites you visit may still have records of your visit. Any files saved to your computer or mobile devices will still remain. For example, if you sign into your Google Account on http://www.google.com while in incognito mode, your subsequent web searches are recorded in your Google Web History. In this case, to prevent your searches from being stored in your Google Account, you’ll need to pause your Google Web History tracking."

At the same time Mozilla documentation [11]

> "What does Private Browsing not save?
>
> - Visited pages: No pages will be added to the list of sites in the History menu, the Library window’s history list, or the Awesome Bar address list.
> - Form and Search Bar entries: Nothing you enter into text boxes on web pages or the Search bar will be saved for Form autocomplete.
> - Passwords: No new passwords will be saved.
> - Download List entries: No files you download will be listed in the Downloads Window after you turn off Private Browsing.
> - Cookies: Cookies store information about websites you visit such as site preferences, login status, and data used by plugins like Adobe Flash. Cookies can also be used by third parties to track you across websites. For more info about tracking, see How do I turn on the Do-not-track feature?

[10] https://support.google.com/chrome/answer/95464?hl=en

[11] https://support.mozilla.org/en-US/kb/private-browsing

- Cached Web Content and Offline Web Content and User Data: No temporary Internet files (cached files) or files that websites save for offline use will be saved."

It is worth noting that these are not real antiforensic techniques, as a very interesting report([10] and [9]) in IEF (Internet Evidence Finder) official website states, it is possible to recover a lot of useful information by examination of particular Windows files such as Pagefile.sys, and Hibernation.sys if the device is a notebook, joined to RAM dump. Furthermore for Internet Explorer case the InPrivate mode simply conceals session history from user access but it still store some elements.

6.6 Instant Messaging

There are various Instant Messenger programs and also many websites and applications allow chat sessions so it is unfeasible to enumerate all such tools, but statistics says that Windows Live Messenger and Skype are the most used ones.

Windows Live Messenger

It is the evolution of MSN messaging system, which was pre installed in Windows build. From a forensic point of view it is important to consider the following:

- User can choose to store all the conversations, depending on the version they can be found inside these location:
  - `%UserProfile%\Documents\My Received Files\PassportName######\History` where ###### is a number sequence provided by a hash function.
  - `%UserProfile%\Documents\My Chat Logs\Month Year`.
- Every user can manage multiple messaging accounts so it is compulsory to enlighten account activity:
  - This can be done by examining windows event logs: after each successful login or logout two items are written inside Application Event log file stored in `%SystemDirectory%\system32\config\AppEvent.Evt`.
    Login entry has this description:
    `MsnMsgr(<process_ID>)\\.\C:%UserProfile%\AppData\Local\Microsoft\Messenger\<WLM_account>\SharingMetadata\Working\database_<unique_computer_ID>\dfsr.db: The Database engine started a new instance(0)` while a logout has the same description but instead of `started` is `stopped`.
    Both entries have `ESENT` as source.
  - Also registry stores useful information.
    During a login an new pair key is inserted in
    `HKEY_CURRENT_USER\Software\Microsoft\MSNMessenger\PerPassport-Settings\` with the MSN Passport ID as name and user's settings and preferences as values.
    When the login attempt fail the only value is a binary item named
    `DefaultSignInState`.
- Every account creates a series of folders:

  - %UserProfile%\Contacts.
  - %UserProfile%\AppData\Local\Microsoft\Windows Live Contacts\.
  - %UserProfile%\AppData\Local\Microsoft\Messenger\.

- Transmitted files are stored by default inside %UserProfile%\Documents\Received Files.
- Inside notebook or netbook hiberfil.sys could contain MSN protocol traces together with Content.IE5 inside Temporary Internet Files folder.

Files containing contacts, shared files and profile pictures must be interpreted and decrypted as it is explained in [6].

Skype

Skype is one of the most used VOIP applications, which permits text and video chats, single and conference calls and file sharing. Architecture behind this program is out of the scope of this thesis, since it is explaining a context of post-mortem analysis. An excellent understanding of Skype architecture is required in cases of interceptions and network forensics[12].

Artefacts left by Skype are located in %UserProfile%\AppData\Roaming\Skype\%AccountName% using a sqlite database format. In figure 17 folder content is shown.
Table 12 summarize database schema which is completely shown in Figure 18. There exist also useful tool, like SkypeLogView [13], which can simplify data browsing by automating some SQl statement.
Figure 19 report a complete history summary.

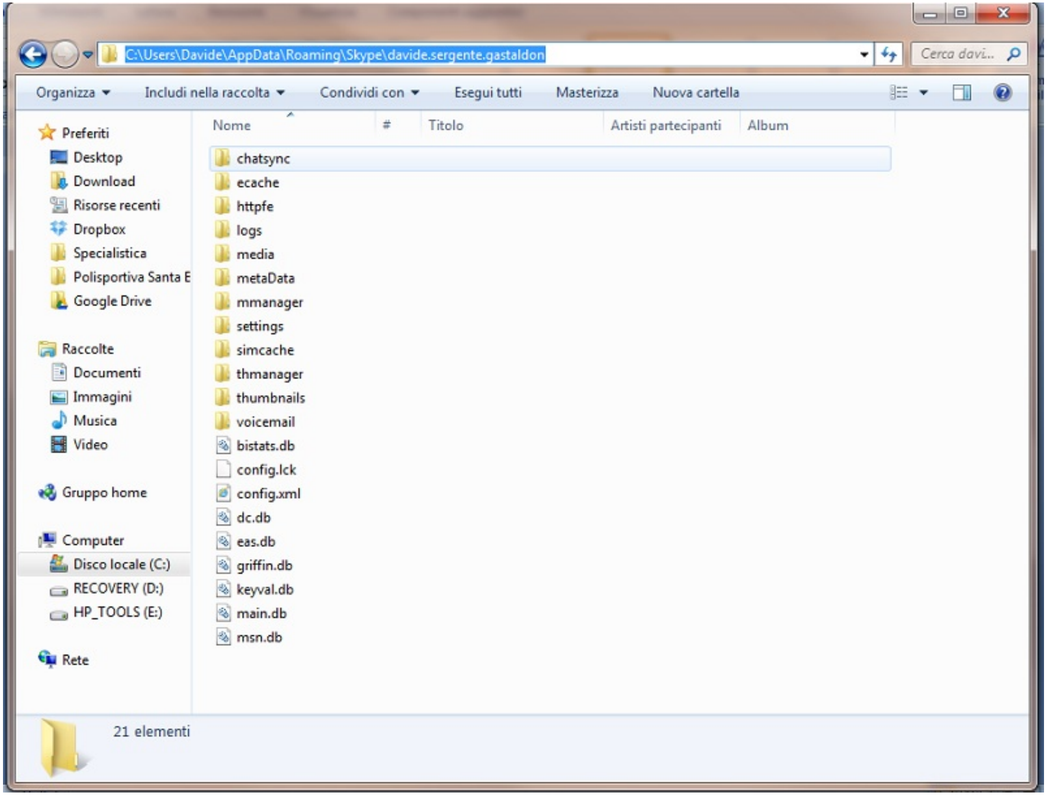
Table 12: Skype database schema

Table	Content
Calls	contains VOIP calls detail
Videos	stores information about camera devices
VideoMessages	include detail of Video calls
Chats	maintains records about chat sessions
Messages	stores text chats
Contacts	contains information about linked contacts
Accounts	stores detail about used accounts
Transfers	lists all transferred files

[12] see for further details [5]

[13] http://www.nirsoft.net/utils/skype_log_view.html

Figure 17: Skype folder

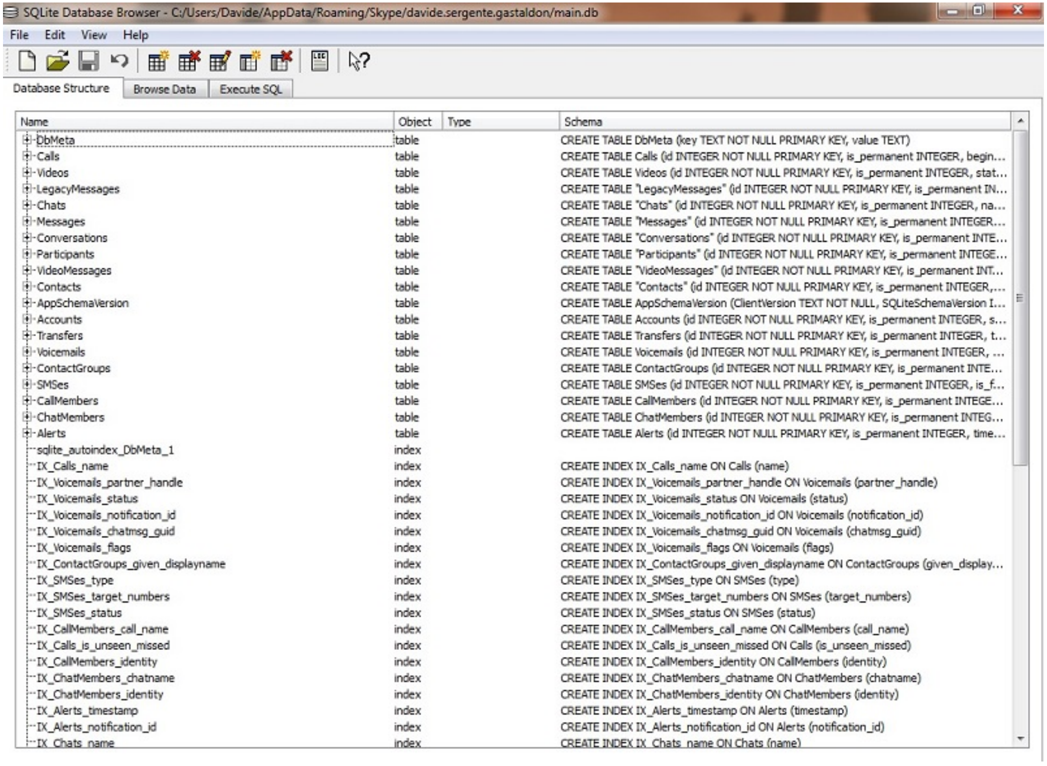


C:\Users\Davide\AppData\Roaming\Skype\davide.sergente.gastaldon

```
chatsync
ecache
httpfe
logs
media
metaData
mmanager
settings
simcache
thmanager
thumbnails
voicemail
bistats.db
config.lck
config.xml
dc.db
eas.db
griffin.db
keyval.db
main.db
msn.db
```

21 elementi

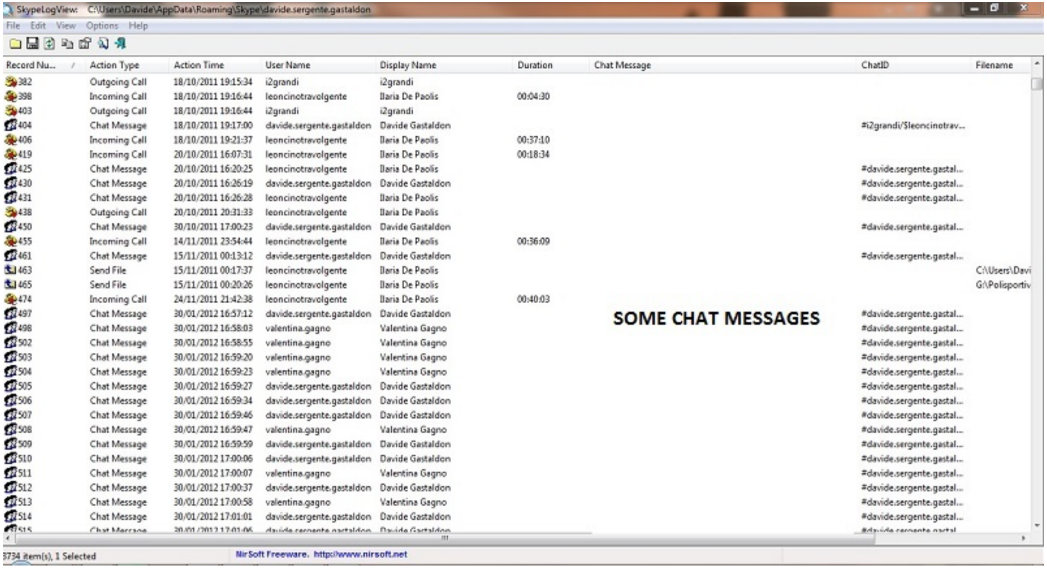
Figure 18: Main.db schema



SQLite Database Browser - C:/Users/Davide/AppData/Roaming/Skype/davide.sergente.gastaldon/main.db

Name	Object	Type	Schema
DbMeta	table		CREATE TABLE DbMeta (key TEXT NOT NULL PRIMARY KEY, value TEXT)
Calls	table		CREATE TABLE Calls (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, begin...
Videos	table		CREATE TABLE Videos (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, stat...
LegacyMessages	table		CREATE TABLE "LegacyMessages" (id INTEGER NOT NULL PRIMARY KEY, is_permanent IN...
Chats	table		CREATE TABLE "Chats" (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, na...
Messages	table		CREATE TABLE "Messages" (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER...
Conversations	table		CREATE TABLE "Conversations" (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTE...
Participants	table		CREATE TABLE "Participants" (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGE...
VideoMessages	table		CREATE TABLE "VideoMessages" (id INTEGER NOT NULL PRIMARY KEY, is_permanent INT...
Contacts	table		CREATE TABLE "Contacts" (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER,...
AppSchemaVersion	table		CREATE TABLE AppSchemaVersion (ClientVersion TEXT NOT NULL, SQLiteSchemaVersion I...
Accounts	table		CREATE TABLE Accounts (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, s...
Transfers	table		CREATE TABLE Transfers (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, t...
Voicemails	table		CREATE TABLE Voicemails (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, ...
ContactGroups	table		CREATE TABLE ContactGroups (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTE...
SMSes	table		CREATE TABLE SMSes (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, is_f...
CallMembers	table		CREATE TABLE CallMembers (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGE...
ChatMembers	table		CREATE TABLE ChatMembers (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEG...
Alerts	table		CREATE TABLE Alerts (id INTEGER NOT NULL PRIMARY KEY, is_permanent INTEGER, time...
sqlite_autoindex_DbMeta_1	index		
IX_Calls_name	index		CREATE INDEX IX_Calls_name ON Calls (name)
IX_Voicemails_partner_handle	index		CREATE INDEX IX_Voicemails_partner_handle ON Voicemails (partner_handle)
IX_Voicemails_status	index		CREATE INDEX IX_Voicemails_status ON Voicemails (status)
IX_Voicemails_notification_id	index		CREATE INDEX IX_Voicemails_notification_id ON Voicemails (notification_id)
IX_Voicemails_chatmsg_guid	index		CREATE INDEX IX_Voicemails_chatmsg_guid ON Voicemails (chatmsg_guid)
IX_Voicemails_flags	index		CREATE INDEX IX_Voicemails_flags ON Voicemails (flags)
IX_ContactGroups_given_displayname	index		CREATE INDEX IX_ContactGroups_given_displayname ON ContactGroups (given_display...
IX_SMSes_type	index		CREATE INDEX IX_SMSes_type ON SMSes (type)
IX_SMSes_target_numbers	index		CREATE INDEX IX_SMSes_target_numbers ON SMSes (target_numbers)
IX_SMSes_status	index		CREATE INDEX IX_SMSes_status ON SMSes (status)
IX_CallMembers_call_name	index		CREATE INDEX IX_CallMembers_call_name ON CallMembers (call_name)
IX_Calls_is_unseen_missed	index		CREATE INDEX IX_Calls_is_unseen_missed ON Calls (is_unseen_missed)
IX_CallMembers_identity	index		CREATE INDEX IX_CallMembers_identity ON CallMembers (identity)
IX_ChatMembers_chatname	index		CREATE INDEX IX_ChatMembers_chatname ON ChatMembers (chatname)
IX_ChatMembers_identity	index		CREATE INDEX IX_ChatMembers_identity ON ChatMembers (identity)
IX_Alerts_timestamp	index		CREATE INDEX IX_Alerts_timestamp ON Alerts (timestamp)
IX_Alerts_notification_id	index		CREATE INDEX IX_Alerts_notification_id ON Alerts (notification_id)
IX_Chats_name	index		CREATE INDEX IX_Chats_name ON Chats (name)

Figure 19: Skype Log View



SkypeLogView: C:\Users\Davide\AppData\Roaming\Skype\davide.sergente.gastaldon
File
Edit
View
Options
Help

Record Nu... /	Action Type	Action Time	User Name	Display Name	Duration	Chat Message	ChatID	Filename
382	Outgoing Call	18/10/2011 19:15:34	i2grandi	i2grandi				
398	Incoming Call	18/10/2011 19:16:44	leoncinotravolgente	Ilaria De Paolis	00:04:30			
403	Outgoing Call	18/10/2011 19:16:44	i2grandi	i2grandi				
404	Chat Message	18/10/2011 19:17:00	davide.sergente.gastaldon	Davide Gastaldon			#i2grandi/$leoncinotrav...	
406	Incoming Call	18/10/2011 19:21:37	leoncinotravolgente	Ilaria De Paolis	00:37:10			
419	Incoming Call	20/10/2011 16:07:31	leoncinotravolgente	Ilaria De Paolis	00:18:34			
425	Chat Message	20/10/2011 16:20:25	leoncinotravolgente	Ilaria De Paolis			#davide.sergente.gastal...	
430	Chat Message	20/10/2011 16:26:19	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
431	Chat Message	20/10/2011 16:26:28	leoncinotravolgente	Ilaria De Paolis			#davide.sergente.gastal...	
438	Outgoing Call	20/10/2011 20:31:33	leoncinotravolgente	Ilaria De Paolis				
450	Chat Message	30/10/2011 17:00:23	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
455	Incoming Call	14/11/2011 23:54:44	leoncinotravolgente	Ilaria De Paolis	00:36:09			
461	Chat Message	15/11/2011 00:13:12	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
463	Send File	15/11/2011 00:17:37	leoncinotravolgente	Ilaria De Paolis				C:\Users\Davi
465	Send File	15/11/2011 00:20:26	leoncinotravolgente	Ilaria De Paolis				G:\Polisportiv
474	Incoming Call	24/11/2011 21:42:38	leoncinotravolgente	Ilaria De Paolis	00:40:03			
497	Chat Message	30/01/2012 16:57:12	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
498	Chat Message	30/01/2012 16:58:03	valentina.gagno	Valentina Gagno			#davide.sergente.gastal...	
502	Chat Message	30/01/2012 16:58:55	valentina.gagno	Valentina Gagno			#davide.sergente.gastal...	
503	Chat Message	30/01/2012 16:59:20	valentina.gagno	Valentina Gagno			#davide.sergente.gastal...	
504	Chat Message	30/01/2012 16:59:23	valentina.gagno	Valentina Gagno			#davide.sergente.gastal...	
505	Chat Message	30/01/2012 16:59:27	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
506	Chat Message	30/01/2012 16:59:34	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
507	Chat Message	30/01/2012 16:59:46	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
508	Chat Message	30/01/2012 16:59:47	valentina.gagno	Valentina Gagno			#davide.sergente.gastal...	
509	Chat Message	30/01/2012 16:59:59	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
510	Chat Message	30/01/2012 17:00:06	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
511	Chat Message	30/01/2012 17:00:07	valentina.gagno	Valentina Gagno			#davide.sergente.gastal...	
512	Chat Message	30/01/2012 17:00:37	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
513	Chat Message	30/01/2012 17:00:58	valentina.gagno	Valentina Gagno			#davide.sergente.gastal...	
514	Chat Message	30/01/2012 17:01:01	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal...	
515	Chat Message	30/01/2012 17:01:06	davide.sergente.gastaldon	Davide Gastaldon			#davide.sergente.gastal	

SOME CHAT MESSAGES

3734 item(s), 1 Selected

NirSoft Freeware. http://www.nirsoft.net

7 Child Pornography

Digital forensic is always more frequently needed by the police and the judicial body for cases concerning child pornography. An Italian statistic claims that over 30% of total case involve Child Pornography detection.[14]

That's why the development of knowledge and tools in order to obtain photographic and multimedia materials containing child pornography is very important. There are 2 important benefits concerning this research:

- Investigators don't waste time, they can filter a huge quantity of data, focusing their attention only on significant material.
- In the ideal case of an "almost" perfect tool, the strain caused by the vision of emotionally stressing material can be avoid, or at least operators can be prepared in time to deal with the close examination. According to **??** exposure to child abuse material can lead to health problems an to resign.

7.1 Definition

The definition from the English Dictionary [15] is:

Definition 3 (Child Pornography) *"the illegal use of children in pornographic pictures or films."*

The problem arising from this definition is the restriction on using automatic recognition during a forensic procedure, aiming at the production of evidences in a legal debate.
The definition of child pornography is very abstract and it depends on the emotion appeal of the examiner. What someone can define as *art*, can be considered *porn* by someone else.

- The nudity of the subject is not always a necessary and sufficient condition to speak of child pornography.
- The age estimation is an open issue and is a subject matter of research. It must be mentioned that every country has different standards to judge these circumstances.

The sexual assault against a child is recognized by everyone as child pornography.
Furthermore it represents a legal base to punish the production, detention or circulation of child pornography.

7.2 Considerations

Shown below some observations arising in case of a post mortem analysis, i.e. the expertise made by a technical adviser sent by a judge:

- In any case the final result has to be evaluated and validated by an operator.

[14] IISFA suvey 2010, lo stato dell'arte della computer foreniscs in Italia http://blog.clusit.it/sicuramente/2010/05/

[15] http://www.freethesaurus.org/dictionary/child+pornography

- The context in which the material is generated is essential. In particular it is important to define if:
  - The elements are made by the person under investigation or if they have been found on the web;
  - Videos or images are locally saved or they belong to any application cache;
  - They belong to material, which has been shared through file sharing and/or peer to peer programs;
  - It is possible to trace back the e-mail addresses of the subjects receiving this documents per e-mail

  That's why it become important to preserve peculiarities and metadata of the original evidence.

- The operating speed has a minor importance, which means that the process doesn't need to be optimized in order to work in real time context.

- The profiling of subjects linked to this kind of material has shown that only a few subjects change the file name and in most cases they contain an explicit reference to the content itself. In addition to the primary field of the research concerning the computer vision, it should be considered also the above mentioned perspective.

There are three different attitudes that can be applied to an automatic identification software:

i. Filtering: the software works as a filter, deleting the non-pornographic material. In this way the operator will examine only the elements with a highly probably "dirty" content.

ii. Researching: the software must operate as a maximum precision investigation tool. The material that for sure concerns child pornography must be recognized by the software, afterwards the author can decide to expand the research.

iii. Mixed: which means using a researching attitude at the beginning and a filtering attitude at a later stage in order to obtain as much material as possible.

7.3 State of the art

Although the researchers are always operating in both fields of recognizing pornographic contents and age estimation, a commercial or open-source product solving the problems mentioned above does not exist.
They mainly analyze images, considering videos as a string of pictures, following a filtering method, they only search for images having a high probability of nude content. Among the most used tool there are:

- FTK Explicit Image Detection;
- EnCase Image Analyzer App;
- Paraben Porn Detection Stick;

- Videntifier Forensic;
- Adroit Photo Forensics;
- RedLight Pornography Scanner;
- Internet Evidence Finder;

FTK Explicit Image Detection

It is a additional feature of Access Data forensic Toolkit since 3.0 version. There is no demo version to test but official documentation states [16]

> - Image recognition is performed by instant comparison with a database of about 30000 explicit pictures, as well as detecting skin tone changes (Flesh Color) .
> - Recognition process assigns a mark to inspected picture, ranking it in a pornographic probability scale from 1 to 100.
> - It is possible to choice from four level of matching: Explicit Level (default), Explicit Level (fast), Explicit Level (zero false negatives), Explicit Level (zero false positives).

Company website states that inside analysis stage advanced features extraction techniques are employed, such shapes recognition, and that product was trained with a database of 30000 explicit pictures. The training activity is still performed. A brief description of operating level is given: [17]

> - X-DFT Default (XS1): This is the most generally accurate. It is always selected.
> - X-FST Fast (XTB): This is the fastest. It scores a folder by the number of files it contains that meet the criteria for a high likelihood of explicit material. It is built on a different technology than X-DFT and does not use “regular” DNAs. It is designed for very high volumes, or real-time page scoring. Its purpose is to quickly reduce, or filter, the volume of data to a meaningful set.
> - X-ZFN Less False Negatives(XT2): This is a profile similar to X-FST but with more features and with fewer false negatives than X-DFT. You can apply this filter after initial processing to all evidence, or to only the folders that score highly using the X-FST option. Check-mark or highlight those folders to isolate them for Additional Analysis.

EnCase

By default Guidance Software leading product does not include such feature. However there exists the possibility to enhance Encase functionalities by EnScript, a built-in programming language. This chance is exploited in partnership with Image Analyzer, a company who develop a computer vision SDK, by provide an application that has a free trial limited to 100 pictures.

[16] `http://www.sistemieservizi.net/Prodotti/Sku.aspx?XRI=22`

[17] `http://www.accessdata.com/products/digital-forensics/ftk`

Paraben Porn Detection Stick

Originally thought as a parental control tool , it is commonly used also in forensic analysis. It is a USB standalone drive that search for picture inside a mounted filesystem classifying them depending on their content. There is no trial or demo version, also after an explicit request to manufacturer, and official documentation is lacked.

VIdentifier Forensic

It was developed as a forensic image toolkit with the goal to detect fake elements and tampering traces. Its main scope is not to detect pornography, but it is widely used also for this reason as it is useful to discover picture history and it provide good results in terms of time and detection rate. It is one of the most scientifically documented [18].

Adroit Photo Forensic

It is developed inside New York University [19] and is one the few products that include age estimation techniques. Official documentation says that it has three modes, a fast one which skims and the other two goes more in depth. It allows to train built-in classifier by human supervision. It has a free trial version.

Red Light Pornography Scanner

Based on Sean P. Alvarez PhD work on University of Rhode Island [20] it is thought to be helpful and usable by Law Enforcement officers. Developers ensure that it employs advanced techniques such as Region of Interest research and Human Body shapes recognition. Website claims that it has a detection rate of 80% with a speed that is from 5 to 10 times better than other tools like FTK.

Internet Evidence Finder

This is product has become a standard-de-facto in detection of artefacts linked to Internet browsing. It detect pictures related to Social Networks and Instant Messaging locally stored, with the possibility to filter them based on skin tone pixel.

7.4 Academic Literature

A lot of research has been done in this field, in particular it has been searched for pornographic material in order to prevent minors from using this kind of material. Here below one of the suggested approaches to detect child sexual abuse images [22].

1. Localising every face in the image, filtering the skin tone regions thanks to facial recognition techniques;
2. Using an age classifier to elaborate the material found in step 1
3. Analysing the results in order to localise the related skin regions

[18] http://www.videntifier.com/papers/

[19] http://digital-assembly.com/technology/

[20] http://dfcsc.uri.edu/research/redLightFeatures

4. Using pattern recognition techniques on human structures, when faces belong to a naked body.

The research has followed different paths to reach the goal of filtering explicit material and reveal its presence in a web context. A different point of view arise compared to the post mortem identification, which means after the sequestration of the device containing the saved material. Furthermore the attention was focused on recognizing pornographic material instead of child pornographic material.

The identification of pornographic material is normally handled as a classification problem. Images can be analysed through the following steps:

1. Identifying the proportions of images representing skin;
2. Separating this regions from the rest of the image;
3. Extracting some features from these regions, such as colours, texture and shapes
4. Using the extracted features to distinguish between benign and non-benign pictures.

It is also possible to use content-based image retrieval method. Taking an image as model, other similar images are taken from a specific database. If most of these images are considered pornographic, that will means that also the starting image is pornographic, otherwise it will be classified as innocuous.

First studies about this theme date back to 1996 through a paper [8] where filters were used in order to identify skin and shapes attributable to human beings. It was recommend to use information about both colour and texture to find regions containing skin. The identification of human shapes is based on biometric characteristics obtained through a geometrical analysis.

Later, a research named WIPE [23] introduced a shape matching algorithm based on wavelet and texture and colour histogram filters.

The statistical approach [15] defined a colour model separating skin regions from non-skin regions. Among the features used by this method we can mention: the percentage of pixel representing skin, the mean probability of a pixel to represent skin, the size of the biggest region of contiguous pixel containing skin, used to train a neural network connection whereon a classifier bases.

Another work [7] uses different color spaces, YOV and YIQ in order to detect areas containing skin, then it employs some filters, Sobel and Gabor operator, in order to delete non-significant regions.

The use of regression trees and SVM (Support Vector Machines) has been analysed and compared [21]. The approach based on SVM reaches better performances achieving 90.4% accuracy.

The statistic characterization has been further investigated [25] with IASCM (Illumination Adaptive Statistical Color Model), useful to detect skin region in variant illumination environment. The typical uniformity of skin textures is a discriminating factor to delete spurious regions. Colours, other textures and shapes are extracted from validated regions and elaborated through SVM to detect adult images.

Nonlinear classifiers are used to detect images representing nudity [24]. The first step of this method consists in the extraction of some ROI (Region Of Interest) containing a lot of pixels within the skin tone range. This extraction is performed through image partition and a greedy approach to grew regions connected through the skin tone. Once CBT (Contour of Body Trunk) is defined, the features concerning the shapes used to establish the classifier can be extracted.

Entropy can improve the color model [26] and create a multilevel neural network to create the classifier.

An adaptative model to detect skin colour can be used to filter images [27]. At the beginning a generic skin model is used to detect pixel that could represent skin. All this is used to train a GMM (Gaussian Mixture Model) with several parameter together with a SVM.

These works investigate two big problems:

- Selecting an appropriate training set
- Choosing parameters and thresholds to an appropriate classification

The training set is normally composed by a selection of possible images representative of thresholds conditions for recognition algorithm. Parameters and thresholds classifications are set through empiric methods and recently using machine learning approaches.

7.5 A Color Space approach

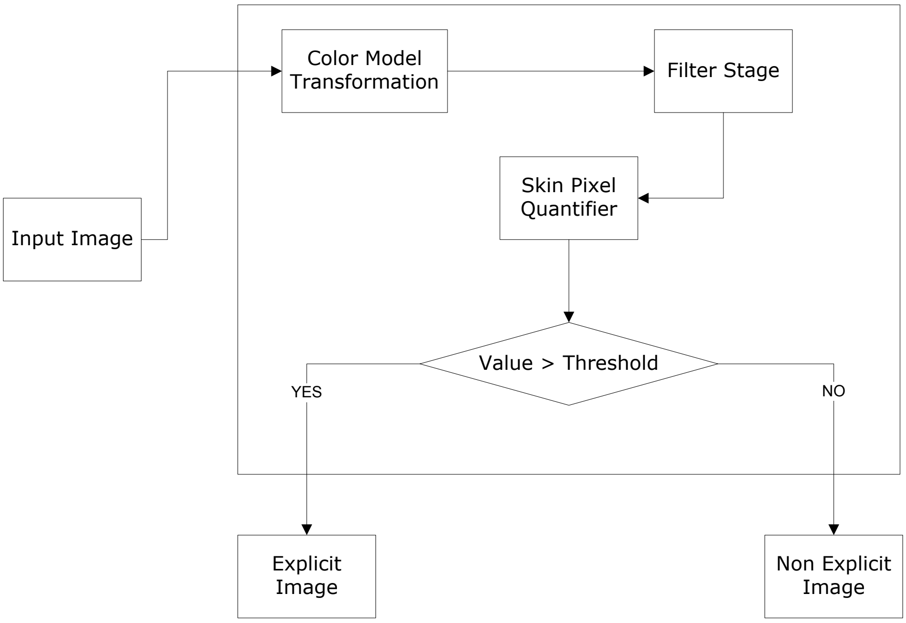
A blending between already mentioned tools and academic studies leads to this algorithm which is studied and tested in depth in [17].

These are mandatory concepts abstracting implementation details:

1. Color model transformation to YCbCr or HSV which permits object identification; in this way properties of interest are isolated. The model choice is justified to decrease drawbacks of image luminance;
2. A filter stage to extract human body shapes. An empirical skin tone act as a threshold;
3. Probability to be a nudity picture is computed:

Picture 7.5 illustrates the flowchart.

Figure 20: A Color Space Approach Flowchart



Color Model transformation

Normally pictures are saved and distributed with RGB-based format, such as JPEG, PNG, TIFF and so on. Here algebraic relations between color models are unfolded:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.996 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Conversely HSV is a non-linear transformation of RGB space settled by:

$$H = \arccos \frac{\frac{1}{2} \times (R-G) + (R-B)]}{\sqrt{[(R-G)^2 + (R-B)(G-B)}}$$

$$S = 1 - 3 \times \frac{min(R,G,B)}{R+G+B}$$

$$V = \frac{1}{3} \times (R+G+B)$$

Filter Stage

After color model transformation, next step requires identification of human pixels. This is done by provide a range in which pixels component must fall, this can be seen as a skin-range calculated empirically.
A lot of researches in pornography detection domain perform such computation, in this section some of these are shown.
For YCbCr [17] shows this result:

$$\begin{bmatrix} Y > 80 \\ 80 \leq Cb \leq 120 \\ 133 \leq Cr \leq 173 \end{bmatrix}$$

Meanwhile in [11] Cb and Cr must meet these constraints, Y-dependent:

$$\beta_1 = \left\{ \begin{matrix} 14 - \frac{Y}{16}, & if \quad Y > 128 \\ 6, & otherwise \end{matrix} \right\}$$

$$\beta_2 = \left\{ \begin{matrix} 4 + \frac{Y}{16}, & if \quad Y > 128 \\ 12, & otherwise \end{matrix} \right\}$$

$$\beta_3 = \left\{ 6, \quad if \quad Y > 1282 + \frac{Y}{32}, \quad otherwise \right\}$$

$$\beta_4 = \left\{ -8, \quad if \quad Y > 128 - 16 + \frac{Y}{16}, \quad otherwise \right\}$$

$$\left\{\begin{matrix} Cr \geq -2Cb + 336 \\ Cr \geq -Cb + 239 \\ Cr \geq -4Cb + 512 \\ Cr \geq 2.5Cb + 2.5\beta_1 - 192 \\ Cr \geq \beta_3 + 128 \\ Cr \geq -0.5Cb + 0.5\beta_4 + 192 \\ Cr \leq \frac{-Cb + 1116}{6} \\ Cr \leq \frac{4(-Cb + \beta_2) + 896}{3} \end{matrix}\right\}$$

Where Y,Cb and Cr range in interval [0;255].
As long as HSV range, normalized in [0,1] ,is computed as follow:

$$\begin{bmatrix} 0 < H < 0.25 \\ 0.15 < S < 0.9 \\ 0.2 < V < 0.95 \end{bmatrix}$$

On the basis of what has been obtained by the search, [11] performances are better, but during this work [17] approach has been chosen to use to ensure lower computational complexity and the chance to trim range trim during the testing phase.

Skin Pixel Quantifier

A simple percentage can be used to fix if an image is made by pornographic content or not.

$$SkinPercentage = \frac{SkinColorPixels}{ImagePixelsTotal} \times 100$$

Threshold is set to 50%

7.6 Image Zoning

A more recent work [20] present an algorithm for detecting nudity in images filtering through skin tone and moreover it introduces the concept of image zoning.

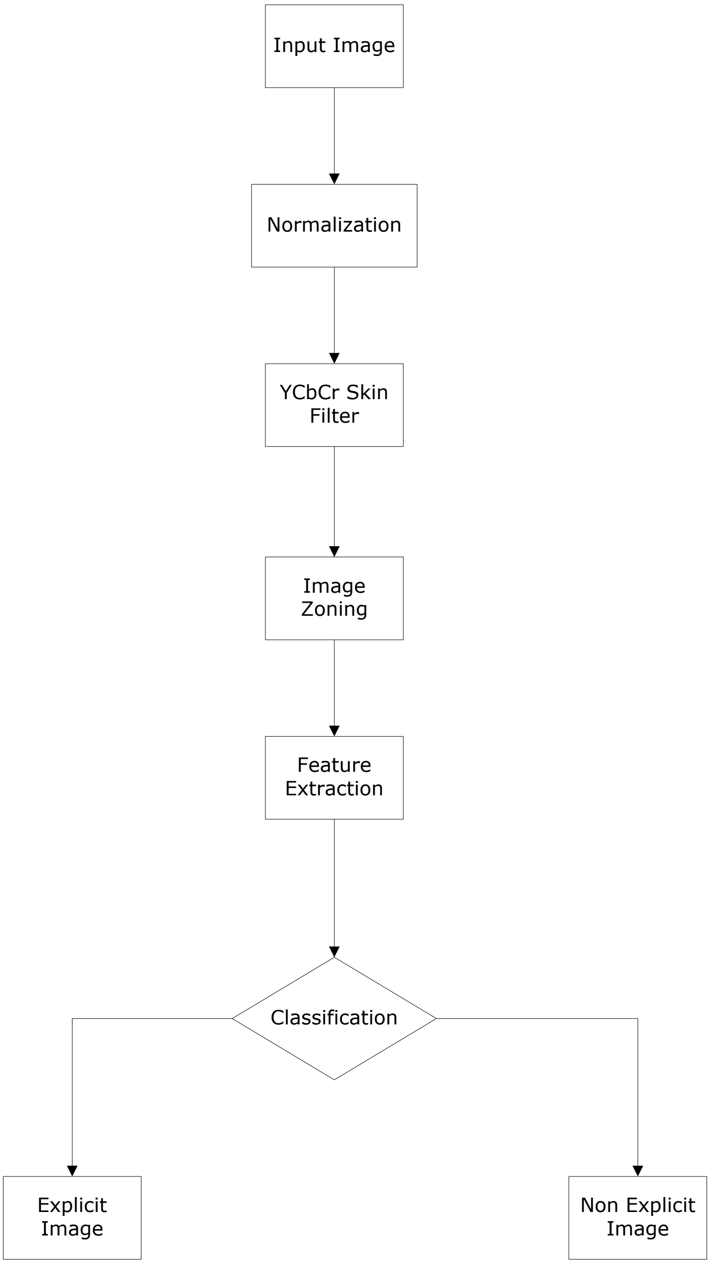
The main assumption, confirmed by the results of the research, states that the major part of skin tone pixel is localized in the central part of an image; hence the image will be divided into independent zone, from each six features can be extracted, these features will be combined and used by SVM in order to be classified. The authors of this work highlight the fact that the term nudity is defined as the state where the person is fully undressed or wearing little clothing, i.e the image of a child on a beach is considered nude picture.

Figure 21 schematizes the approach.

Normalization

In order to solve the problem of images provided with different format and resolutions, every image is normalized to 256*526 and converted to JPEG format. This dimension has been chosen empirically among different configurations.

Figure 21: Image Zoning Flowchart



Skin Filter

According to previous studies, best features to classify images are based on colours. This is the reason why this approach uses the YCbCr color space because:

- It can mitigate luminosity effects, which is one of the key issue in images containing human skin.
- Human skin ranges within this colorspace
- Previous studies reveal that this approach achieve high skin detection rates while reducing computational costs.

The range defined by Cb and Cr normalized on [0,255] is slightly different from the previous algorithm and it is defined in this way:

$$\begin{bmatrix} 77 \leq Cb \leq 127 \\ 133 \leq Cr \leq 173 \end{bmatrix}$$

Zoning Images

The fundamental hypothesis states that the most important region in nude pictures is the central one, therefore there will be the major part of skin tone pixel. According to image zoning we can define independent regions in the image and adopt feature extraction algorithms to locally analyse each zone. Following this approach images are divided into K concentric rectangular zones of proportional size.
The algorithm is defined in this way:

```
NZ(I)=K
For all image I do:
    divide I into NZ;
    for all zone(Z) do
        FE(Z);
        FE(I)=FE(I)+FE(Z);
    end;
end;
CP(FE(I));
```

Where:

- I is a generic image;
- NZ(I) is the number of zones in I;
- Z is a specific zone of I;
- FE(Z) is the local feature extraction process in zone (Z);
- FE(I) is the feature vector obtained combining every FE(Z);
- CP is the classification process.

Feature Extraction

The selection of features is the key point for building a robust learning model to employ during the classification process. Six features are extracted for each zone, namely two color-based and four texture-based features.

Color based:

1. The number of skin pixel connected in the zone
2. The proportion of skin pixel to the total number of pixel in the image.

Texture based extracted from GLCM (Grey-Level Co-Occurrence Matrix)

1. Contrast
2. Correlation
3. Energy
4. Homogeneity

Classification

A SVM is used to accomplish the classification, showing that features extracted from central zone provide to the classification process the most significant information. Furthermore the results obtained and described in [20], show that by reducing the feature dimension to six it is possible to achieve remarkable results, since - the time spent being equal - it reduces computational cost using a big sample during the training phase.

8 Experiments on Child Abuse Detection

In this section first detection stage are investigated, that is the skin detection problem. Goal of work done is not to provide a reliable and final solution to child abuse detection, but only to provide a first framework for further researches. Mentioned solution are tested inside a real digital forensic investigation.

8.1 Data set

In order to evaluate performance of the already mentioned tools, a testing data set is constructed. These conditions are met:

- To test free Image Analyzer for EnCase we can use only 100 items, which must cover most of the possible real cases.
- To reconstruct, as far as possible, a real scenario images must have different sizes.

The first testing set constructed has 100 items which are divided in this way:

- 50 are pornographic images;
- 50 are non-adult images, further divided in 25 in non offensive nudity images, such as "bikini" images like 22 , and 25 other kinds of images, such as dressed people, animal, plants, cars, landscapes and so on.

Figure 22: Bikini Image

In Graph 23 distribution of size of first data set is represented.

The second data set has a bigger size, 800 items. This choice is motivated to compare results with [17], at least in the number of items seeing paper data set is not publicly available. Images' typology division is the same of the first data set:

- 400 are pornographic images;
- 400 are non-adult images.

Figure 23: 1st Data set size distribution

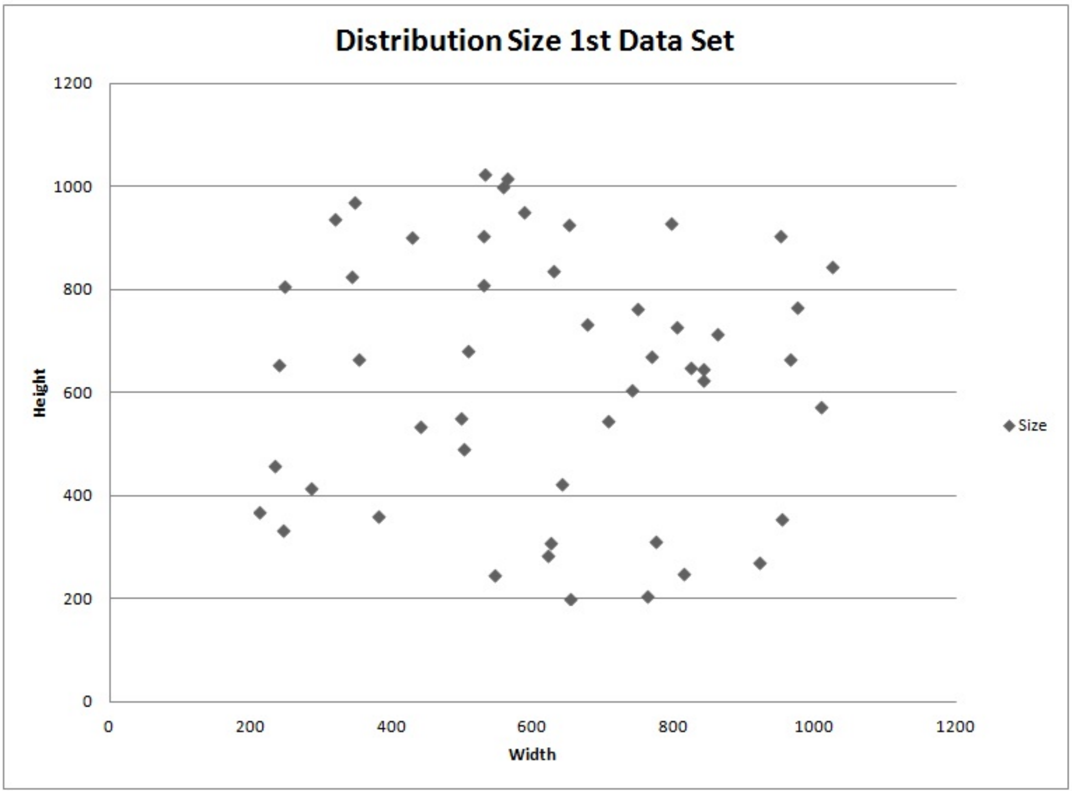
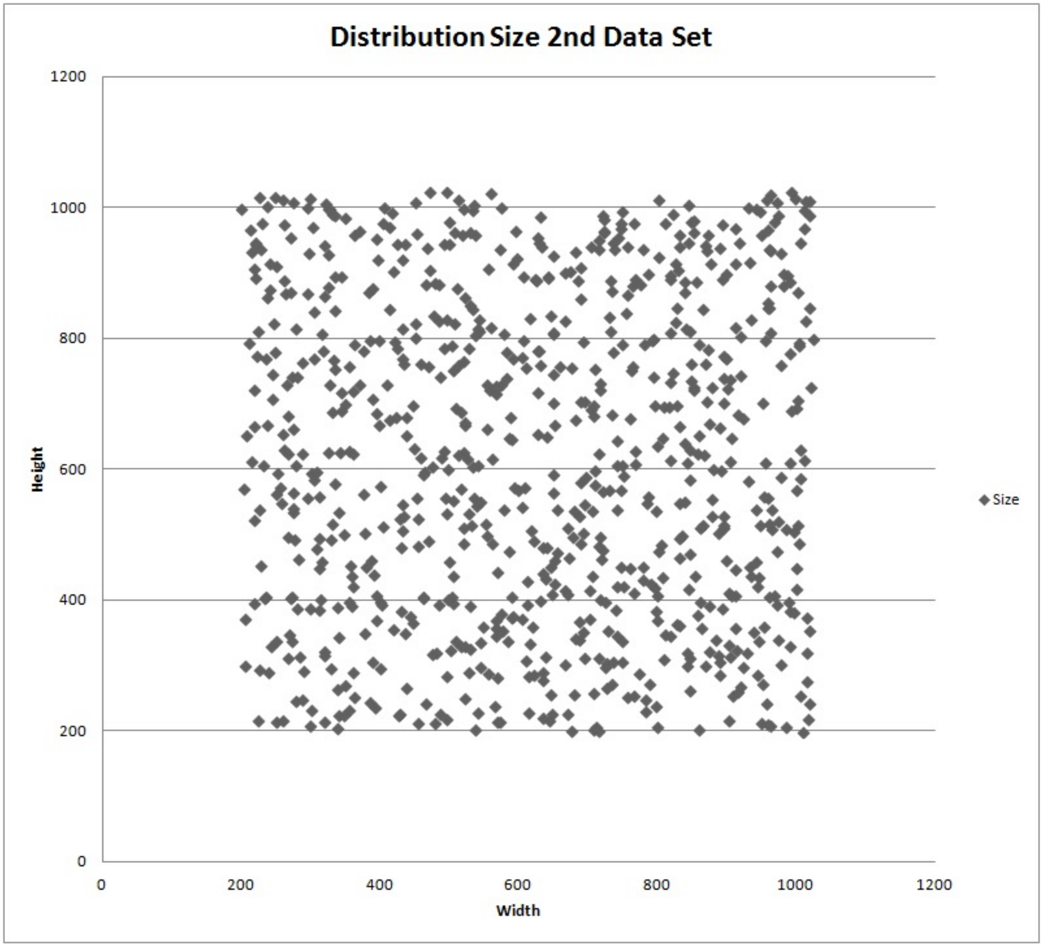


Figure 24 represents sizes distribution within second data set.

Figure 24: 2nd Data set size distribution



Research results are further improved by inserting inside the aforementioned sets a sampling of images founded by examination of evidence during previously explained investigations. In author's opinion this can be useful to provide a real framework of comparison and results can be used to improve further research in the domain of child-abuse material detection.

The random sampling follow what are stated in [14] and use the EnScript provided by the authors that is based on what is reported in Section 11.1.

This dataset obviously cannot be made public.

8.2 OpenCV for Skin Detection

OpenCV (Open Computer Vision) is the leading free library on Computer Vision [21]

[21] http://opencv.org/

> OpenCV is released under a BSD license and hence it's free for both academic and commercial use.
>
> It has C++, C, Python and Java interfaces and supports Windows, Linux, Mac OS, iOS and Android.
>
> OpenCV was designed for computational efficiency and with a strong focus on real-time applications. Written in optimized C/C++, the library can take advantage of multi-core processing. Adopted all around the world, OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 6 million.
>
> Usage ranges from interactive art, to mines inspection, stitching maps on the web or through advanced robotics.

One of its countless tutorials explains how to implement a basic skin detector[22]:

```
#include <opencv2/imgproc/imgproc.hpp>
#include <opencv2/highgui/highgui.hpp>

using namespace cv;

int main()
{
    Mat src = imread("people.jpg");
    if (src.empty())
        return -1;

    Mat hsv;
    cvtColor(src, hsv, CV_BGR2HSV);

    Mat bw;
    inRange(hsv, Scalar(0, 58, 89), Scalar(25, 173, 229), bw);

    imshow("src", src);
    imshow("dst", bw);
    waitKey(0);

    return 0;
}
```

Code 1: OpenCv Tutorial Code

By inspecting the code we can derive these interesting step

Colorspace Transformation

```
cvtColor(src, hsv, CV_BGR2HSV);
```

transform src (source) image to HSV colorspace.

Threshold the image

```
inRange(hsv, Scalar(0, 58, 89), Scalar(25, 173, 229), bw)
```

[22] http://bsd-noobz.com/opencv-guide/60-4-skin-detection

set pixel in `bw` to 255 if the corresponding pixel in `hsv` falls inside the range which lower boundary is `Scalar(h, s, v)` and upper boundary is `Scalar(H,S,V)`.

In Code 1 lower boundary is set to (0,23,25) and upper boundary is set to (50,68,90).

This range is normalized with respect to OpenCV HSV model in which component are inside these intervals:

$$\begin{bmatrix} 0 \leq H \leq 180 \\ 0 \leq S \leq 255 \\ 0 \leq V \leq 255 \end{bmatrix}$$

As explained in Section 7.5 threshold range consider to be optimal is:

$$\begin{bmatrix} 0 < H < 0.25 \\ 0.15 < S < 0.9 \\ 0.2 < V < 0.95 \end{bmatrix}$$

but in range[0;1], so a normalization is necessary by this rule:

$$OCVvalue = value \times \frac{OCVUpperBoundary}{valueUpperBoundary}$$

where `OCVvalue` stands for OpenCv value, `OCVUpperBoundary` stands for OpenCV HSV range upper boundary and `valueUpperBoundary` is equal to 1 in such case.

Previously stated values become

$$\begin{bmatrix} 0 < H < 45 \\ 38.25 < S < 229.5 \\ 51 < V < 242.25 \end{bmatrix}$$

So tutorial code become

```
#include <opencv2/imgproc/imgproc.hpp>
#include <opencv2/highgui/highgui.hpp>

using namespace cv;

int main()
{
    Mat src = imread("people.jpg");
    if (src.empty())
        return -1;

    Mat hsv;
    cvtColor(src, hsv, CV_BGR2HSV);

    Mat bw;
    inRange(hsv, Scalar(0, 38.25, 51), Scalar(45, 229.5, 242.25), bw);

    imshow("src", src);
    imshow("dst", bw);
    waitKey(0);

    return 0;
}
```

Code 2: OpenCv Refined Code

8.3 Experiment description

Previously mentioned data sets are analyzed by these tools:

i. OpenCV implementation in 2

ii. EnCase Image Analyzer trial, only for the first data set.

iii. Adroit Photo Forensic free trial version.

iv. RedLight trial version.

v. Internet Evidence Finder.

vi. A C++ skin detector foound in [23]; code is listed in Section 11.2.

In first experiment, OpenCV implementation is setted to recognize porn if more than 50% of its pixel fall inside skin region, this value is called Skin Percentage Threshold.
After the percentage has changed until it reaches 90% by increasing 5% each step.

Figure 8.3 illustrates results obtained. In the graph *Detection Rate* means the percentage of pictures correctly classified as pornography meanwhile False Positive rate means pictures wrongly classified as pornography, i.e. normal images that OpenCV implementation recognize as explicit.
This test is performed with another database,formed by 20 explicit images and 20 images as far as possible from pornographic content. In subsequent tests it is setted to 75% percent, as a trade-off between detection rate and false positives rate.

Regarding Internet Evidence Finder, it has the possibility to manually set the Skin Percentage Threshold. Obviously detection algorithm detail are not available, so threshold is set to 75%, the same as OpenCV implementation.

Results obtained from comparisons of the tools set are summarized in next sections.

8.4 Experiments results

8.4.1 First Data Set

Table 13 reports this two kind of measure:

Detection Rate Images correctly classified as pornographic with respect to total explicit elements

False Negatives Pictures wrongly classified as pornographic with respect to total inexplicit elements

Figure 25: Rates trends with respect to Skin Percentage threshold

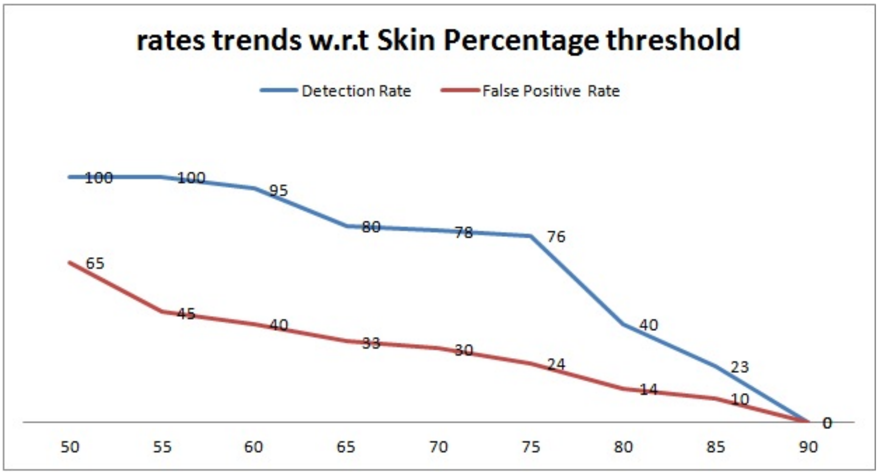


Table 13: Results obtained with 1st Data Set

Tool Detection Rate	False Negatives	
OpenCV	65%	27%
EnCase	80%	16%
Adroit	34%	N.A.
RedLight	70%	5%
C++ detector	60%	35%
Internet Evidence Finder	68%	25%

It is worth noting that trial version of Adroit Photo Forensic is very limited with respect to categorization task, but is the only tool that has built-in carving. It gives only a simply presentation of total features and its deduction rules are based also from human supervision that is blocked in demo version. For these reasons it is deleted from tool set in the subsequent tests.

False negatives rates is more meaningful if it is further divided into "porn-similar" and "total-different" rates because this is the way in which data set has been built. It is less significant a wrong classification of images to non offensive nudity fields than a wrong classification of a completely different pictures. Details are illustrated in Table 14.

Table 14: False Negatives Splitting

Tool Nudity	Total Different	
OpenCV	75%	25%
EnCase	85%	15%
RedLight	90%	10%
C++ detector	65%	35%
Internet Evidence Finder	70%	30%

8.4.2 Second Data Set

Results are presented in the same way of previous section. EnCase are not inserted inside toll set since its free Image Analyzer trial is limited to 100 pictures.

Table 15 and 16 show what is obtained.

Table 15: Results obtained with 2nd Data Set

Tool Detection Rate	False Negatives	
OpenCV	63%	30%
RedLight	75%	7%
C++ detector	58%	40%
Internet Evidence Finder	72%	24%

Table 16: False Negatives Splitting

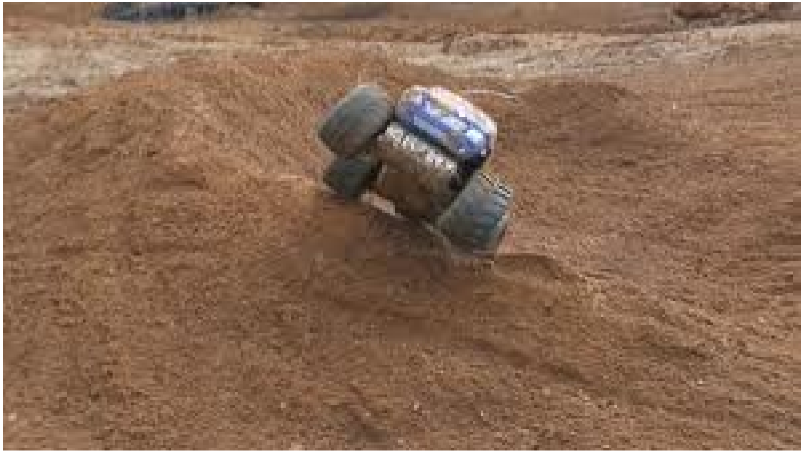
Tool Nudity	Total Different	
OpenCV	80%	20%
RedLight	83%	17%
C++ detector	62%	38%
Internet Evidence Finder	78%	22%

[23] http://cis.ait.asia/course_offerings/148

8.5 Analysis of the results

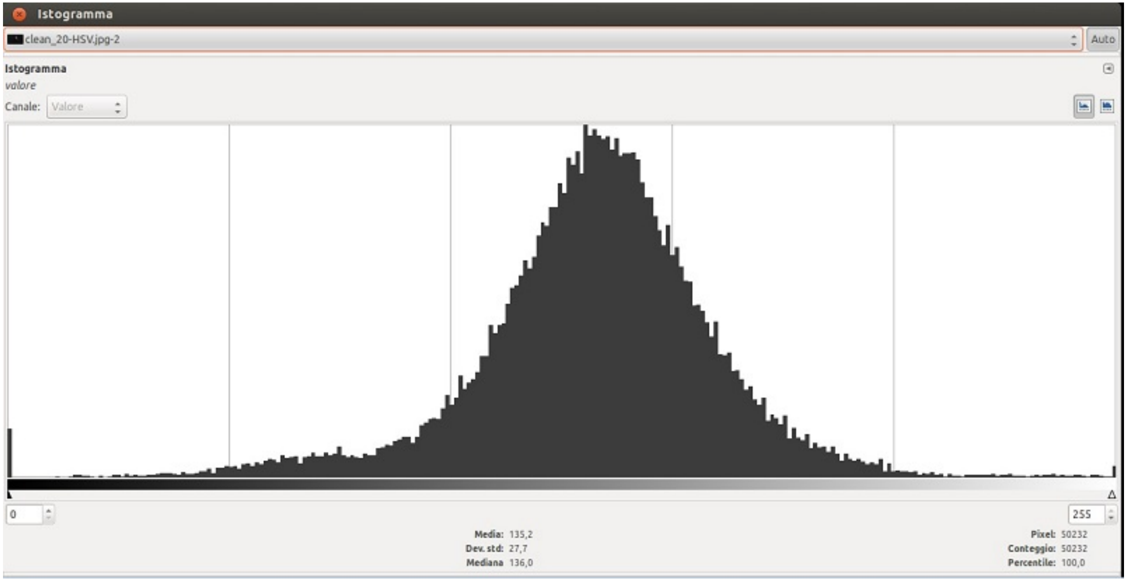
It is interesting to report picture 26 that is always incorrectly classified by all the tools.

Figure 26: Significant false negative picture



By provide its histogram, in Figure 27, it immediately follow all the limits of simply skin detection based on skin tone.

Figure 27: False Negative Histogram



clean_20-HSV.jpg-2

Istogramma
valore

Canale: Valore

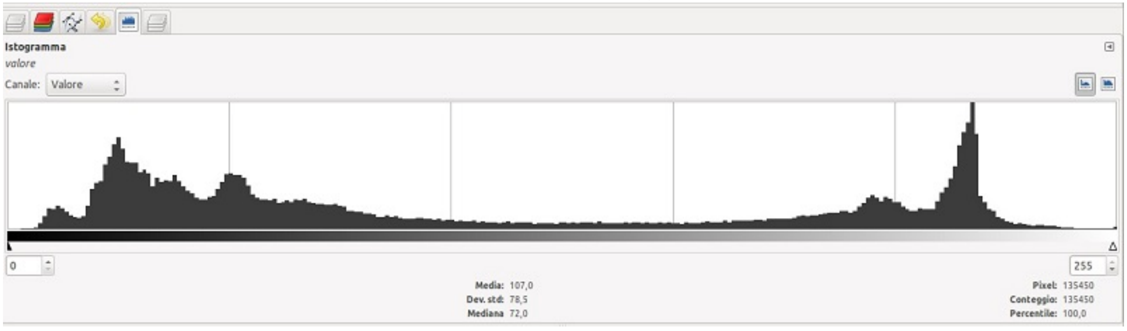
0
255

Media: 135,2
Dev. std: 27,7
Mediana 136,0

Pixel: 50232
Conteggio: 50232
Percentile: 100,0

In such case most of the picture has pixels that fall inside skin tone range, but is not explicit. This is the reason for which skin detection must be enhanced with shape recognition and features extraction.
However detection rates are in according with results obtained by [20] to demonstrate that HSV transformation is a good way to decrease lightning problems in skin detection step.

Figure 28: Ad-Hoc histogram



Istogramma

valore

Canale: Valore

0

255

Media: 107,0
Dev. std: 78,5
Mediana 72,0

Pixel: 135450
Conteggio: 135450
Percentile: 100,0

Further researches have to deepen the use of Artificial Intelligence techniques and the implementation of advanced features extraction techniques.

9 Ad-Hoc Solution

In the context of real case already explained in 4 a more targeted setting can be exploited in order to retrieve Alice's photos.

From previous investigations a sample of pictures portraying Alice is available. So we can derive specific information on color space values. Histogram of HSV color space is reported in Figure 28.

It is possible to fairly set range value in OpenCV implementation as: `inRange(hsv, Scalar(0, 60, 93), Scalar(23, 165, 212), bw)`.

This solution allows to skim results.

Furthermore the number of pixel in skin tone are computed and this reveal that, within sample items, the so-called Skin Percentage ranges from 65% to 80%.

It is worth noting that this a feasible shortcut only if it is assumed that no changes occur to pictures after Alice send them.

In the specific case this approach has permitted to detect pictures that had been sent via instant messaging and e-mail. Note that pictures were not stored inside default location, so search was not so trivial. If the suspect modifies the picture, for example by editing actions, or if communication changes intrinsically pictures, think about facebook that automatically compress and convert in jpg images, this ad-hoc solution could not be useful.

10 Appendix - The legal landscape in Italy

A complete treatment of Digital Forensics related principles of Italian legislative framework lies outside this thesis purposes and this matter does not come within author's competence, therefore this Appendix discusses only basics that are useful for a computer scientist in the throes of forensic analysis, in order to operate in a fruitful way. Moreover, there are various legislative gaps due to the different rates in which operate Information Technology evolution and legislative adjustment; they are usually filled by suprem Court jurisprudence.

With respect to both Civil and Criminal environment, there are several potential actors interested in to be able to use during trial digital evidence; consider, for example:

- A public prosecutor that must support the charge when digital evidence are the medium or the purpose of a criminal activity.
- A defendant which has to show that had nothing to do with his charge.
- A person or a corporate body who is seeking a compensation provided by civil law in case of injury.
- Consider also labour law.

10.1 Digital evidence

The term source of evidence is anything that is able to provide results that are relevant to the court's decision. It may be something such as a murder weapon, a footprint, or a person, such as a witness. In this regard, we differentiate the various stages leading to the creation of evidence:

- Means of proof. The instrument through which an item is used for a court's decision such as a witness, a comparison, an identification of persons or things, an expert report, a document: in compliance with Articles. 194-243 of the Italian Code of Criminal Procedure. In modern indirect surveys, often happens to create photographic dossiers containing items collected from Social Media like Facebook, twitter, Flick and so on.
- Evidence. The information that is obtained from the means of proof such as raw data, not yet assessed by the Judge: in compliance with Article 65 paragraph 1 of the Italian Code of Criminal Procedure. We can express a functional relationship by this equation:

  Means of proof(evidence source)=Evidence

- Probative evidence. The evidence once assessed by the judge according to the criteria of credibility and reliability which allows the ascertainment of the act under prosecution: Article 192 paragraph 1 of the Italian Code of Criminal Procedure. Also in this case a functional relationship is valid and give a good resume:

  Magistrate Evaluation(Evidence)=probative result

Based on these elements, Digital Evidence can be defined as binary data and the respective data having probative value, obtained from a technical computer assessment carried out as part of a criminal or civil proceedings.
The evidence is acquired when data or electronic devices are seized and secured for examination.
The digital evidence often comes in the form of a file or more generally a computer document which:

- Is volatile, like the residue of gunpowder.
- Is latent, like fingerprints or DNA evidence.
- Crosses jurisdictional borders quickly and easily. Not necessarily all of the data in use by the party must be stored on the mass storage devices that belong to the party.
- Is easily altered or damaged due to the inherent fragility of the data contained therein.
- Can be time/machine dependent.

A file or any form of digital data is considered just like unusual evidence covered in Article 189 of the Italian Code of Criminal Procedure, dedicated to "evidence that is not regulated by law", and therefore, so that it may be admitted by the judge as expressly provided for by Article 190 of the Italian Code of Criminal Procedure - The Right to Evidence, must ensure the following features:

- It must be admissible, that is, the method must comply with the dictates of the Italian Code of Criminal Procedure, which is likely to be evaluated during the trial.
- It must be authentic, that is, the acquired data must be accurately traced back to its original form and content, i.e., the one on the machine under investigation.
- Complete, that is, provide all the information related to the acquisition method, availability and location of the data, and not be restricted solely to the description of its existence on a given mass storage device.
- It must be reliable, that is, provide sufficient evidence to provide a solid chain of custody of the data in order to ensure its authenticity and integrity.
- It must be understandable, that is, provide a justification relevance to the investigation of the electronic trace detected, namely to provide a logical-deductive link that is understandable to people who do not have an extensive IT knowledge.

10.2 Inspection

Inspection (pursuant to Article 354 paragraph 2 of the Italian Code of Criminal Procedure) means the entirety of operations that are methodical, scientific, and direct to identify, collect, and secure all the elements useful for reconstructing the event and identifying the offender.

To conduct a good inspection the following tasks must be performed:

- Description: closely observe the place where the crime occurred using scientific methods and knowledge;
- Technical description: establish everything that is observed with irrefutable findings (photos, video, layout, models, and fingerprints).
- Evidence collection: direct activities for taking and preserving any object that is connected to the crime.
- Preservation: every exhibit must be characterized by elements that ensure its integrity and uniqueness, ensuring the chain of custody.
- Documentation: putting into writing the work carried out through the use of reports or technical reports in the event that the person who wrote it is not part of the Criminal Investigation Department.

10.3 Regulatory framework

At the same time as the evolution of digital devices there has been an emergence and proliferation of many new forms of crime and criminal aggression sometimes committed via computer systems and/or electronic means, other times against them, no longer deemed as a means to accomplish such offenses, but as material subjects to said offenses. Based on these considerations, we can define cybercrime as an illegal human act (act or omission) committed against or by means of a computer or electronic system. In the last two decades, the protection of computer and electronic systems in general has been the subject of attention by Italian lawmakers who often acted in the wake of the European or international cooperation obligations, such as the laws listed below:

- Law of July 5, 1991, no. 197 - L.197/1991: Rules for preventing the use of the financial system for money laundering purposes.
- Law of December 23, 1993 no. 547 - L.547/1993: Changes and additions to the rules of the Criminal Code and the Criminal Procedure Code on the subject of computer crime.
- Law of August 3, 1998 no. 269 - L.269/1998: Provisions against the exploitation of prostitution, pornography, and sex tourism involving children, also conducted by electronic means.
- Law of August 18, 2000 no. 248 - L.248/2000: Amendments to Law 633/1941, on the subject of copyright.
- Legislative Decree of June 30, 2003 no. 196 - Legislative Decree 196/2003: Code regarding the protection of personal data as amended by Law no. 45/2004.
- Legislative Decree of March 7, 2005 no. 82 - Legislative Decree no. 82/2005: Digital Administration Code; rules on computer documentation and electronic and digital signature.
- Law of February 6, 2006 no. 38 - L.38/2006: Provisions relating to the fight against sexual exploitation of children and child pornography also via the Internet.

- Article 55 of the Legislative Decree of November 21, 2007 no. 231: Use of improper payment orders.
- Law of March 18, 2008 no. 48: Ratification and implementation of the Council of Europe Convention on Cybercrime.
- Law 12/2012 Rules on measures to combat the phenomena of cybercrime.

11 Appendix - Code

11.1 EnScript Random Sampling

This code provides a framework for sampling inside EnCase environment[24].

```
/*
Helps select a sample size and generate random subsets of files

Script author: Geoff Black
               geoff@geoffblack.com
*/

class MainClass;

class SampleSizeDialogClass: DialogClass {
  MainClass         main;
  CheckBoxClass     SelectedItems;
  StaticTextClass   SelectedItemsText;
  RadioButtonClass  ConfidenceLevelSelection;
  RadioButtonClass  MarginOfErrorSelection;
  ChartClass        SSChart;
  ImageWindowClass  ChartImageWindow;
  StaticTextClass   SampleSizeText;
  PathEditClass     SampleOutputLefPath;

  SampleSizeDialogClass(MainClass m):
    DialogClass(null, "Sample Size Selector"),

    SelectedItems(this, "Calculate based only on selected items
      (folders are always ignored)", START, START, 50, DEFAULT, 0, m.SelectedItems),

    SelectedItemsText(this, m.SelectedItemsText, START, NEXT, 391, DEFAULT,
     WindowClass::BORDER | WindowClass::CENTER),

    ConfidenceLevelSelection(this, "Confidence Level", START, NEXT,
     DEFAULT, DEFAULT, 0, m.ConfidenceLevelSelection, "99%\t95%\t90%"),

    MarginOfErrorSelection(this, "Margin of Error",  NEXT, SAME, DEFAULT,
     DEFAULT, 0, m.MarginOfErrorSelection, "± 2%\t± 5%\t± 10%"),

    ChartImageWindow(this, "", NEXT, SAME, 200, 57, 0),

    SampleSizeText(this, m.SampleSizeText, START, NEXT, 391, DEFAULT,
     WindowClass::BORDER | WindowClass::CENTER),

    SampleOutputLefPath(this, "Sample Output Path (LEF)", START, NEXT, 391,
     DEFAULT, 0, m.SampleOutputLefPath, WindowClass::REQUIRED + WindowClass::FILECREATE,
          "Logical Evidence Files\t*.L01"),

    main = m
  {
  }

  void ShowChart(double sampleSize, double population) {
    SSChart = new ChartClass(null, "", ChartClass::ChartTypes::CHARTPIE);
    SSChart.AddDataPoint("Sample Size", sampleSize);
    SSChart.AddDataPoint("Remainder", population-sampleSize);
    SSChart.SetChartType(ChartClass::ChartTypes::CHARTPIE);
    SSChart.SetHorizontal(false);
    ChartImageWindow.SetImage(SSChart);
  }
```

[24] It is based on this work http://www.geoffblack.com/

```
  uint GetConfidenceLevel() {
    if (ConfidenceLevelSelection.GetValue() == 0) {
      return 99;
    }
    else if (ConfidenceLevelSelection.GetValue() == 1) {
      return 95;
    }
    else if (ConfidenceLevelSelection.GetValue() == 2) {
      return 90;
    }
    return 0;
  }

  uint GetMarginOfError() {
    if (MarginOfErrorSelection.GetValue() == 0) {
      return 2;
    }
    else if (MarginOfErrorSelection.GetValue() == 1) {
      return 5;
    }
    else if (MarginOfErrorSelection.GetValue() == 2) {
      return 10;
    }
    return 0;
  }

  virtual void CheckControls() {
    main.Population = main.CountItems(main.Case.EntryRoot(), SelectedItems.GetValue());

    main.SampleSize = StatsClass::GetSampleSize(GetConfidenceLevel(),
      GetMarginOfError(), main.Population);

    SampleSizeText.SetText("Sample Size: " + String::FormatInt(main.SampleSize,
      int::DECIMAL,String::COMMAS));

    SampleSizeText.Update();

    ShowChart(main.SampleSize, main.Population);
  }

  virtual void ChildEvent(const EventClass& event) {
    DialogClass::ChildEvent(event);
    if (SelectedItems.Matches(event)) {
      SelectedItemsText.SetText(main.CountItemsString(main.Case.EntryRoot(),
       SelectedItems.GetValue()));
    }
  }
}

class StatsClass {

  StatsClass() {}

  static double GetZValue(int ConfidenceLevel) {
    /* z
    90% 1.645
    95% 1.960
    99% 2.576
    */
    if (ConfidenceLevel == 90) {
      return 1.645;
    }
    else if (ConfidenceLevel == 95) {
      return 1.960;
    }
    else if (ConfidenceLevel == 99) {
      return 2.576;
    }
    return 0.0;
  }

  static ulong GetSampleSize(uint confidenceLevel, uint marginOfError, ulong population) {
    /*ConfidenceLevel & MarginOfError should be
       passed as a percent value (i.e. 90[%] or 2[%]) */
    double z = GetZValue(confidenceLevel);
    double ci = (1.0*marginOfError)/100;

    double d1 = z * z * 0.50 * 0.50;
    double d2 = ((1.0*population) - 1.0) * (ci*ci) + d1;
    if (ci > 0) {
      return double::Trunc(double::Ceil(((1.0*population) * d1)/d2));
    }
    else {
      return 0;
    }
  }

  static String GetConfidenceLevelString(int choice) {
    if (choice == 0)
      return "99%";
    else if (choice == 1)
      return "95%";
    else if (choice == 2)
      return "90%";
    return "";
  }

  static String GetMarginOfErrorString(int choice) {
    if (choice == 0)
      return "± 2%";
    else if (choice == 1)
      return "± 5%";
    else if (choice == 2)
      return "± 10%";
    return "";
  }
}

class MainClass {

  bool        SelectedItems;
  String      SelectedItemsText,
              SampleSizeText,
              SampleOutputLefPath;
  int         ConfidenceLevelSelection,
              MarginOfErrorSelection;
  ulong       SampleSize,
              Population;
  CaseClass   Case;
  typedef EntryClass[]  EntryArrayClass;
  EntryArrayClass       EntryArray,
                        SampleArray;

  MainClass():
    EntryArray(),
    SampleArray()
  {
  }

  ulong CountItems(EntryClass entryRoot, bool selected) {
    ulong eCount;
    String ret;
    forall (EntryClass e in entryRoot) {
      if (!e.IsFolder()) {
```

```
      if (!selected || e.IsSelected()) {
        ++eCount;
      }
    }
  }
  return eCount;
}

String CountItemsString(EntryClass entryRoot, bool selected) {
  ulong eCount,
        eSize;
  String ret;
  forall (EntryClass e in entryRoot) {
    if (!e.IsFolder()) {
      if (!selected || e.IsSelected()) {
        ++eCount;
        eSize += e.PhysicalSize();
      }
    }
  }
  ret = "Population:  " + String::FormatInt(eCount, int::DECIMAL, String::COMMAS)
     + " entries (" + String::FormatInt(eSize, int::DECIMAL, String::COMMAS) + " Bytes)";
  return ret;
}

void DlgSettings(uint storageOptions=0) {
  StorageClass storeSettings("Sample Size Selector", storageOptions);
  storeSettings.Value("SelectedItems",             SelectedItems);
  storeSettings.Value("ConfidenceLevelSelection",  ConfidenceLevelSelection);
  storeSettings.Value("MarginOfErrorSelection",    MarginOfErrorSelection);
  storeSettings.Value("SampleOutputLefPath",       SampleOutputLefPath);
}

String GetAdjustedPath(const String& path){
  String adjPath = path;
  adjPath = adjPath.SubString(adjPath.Find("\\")+1, -1);  //remove Case name

  return adjPath;
}

void ProcessEntries(EntryClass root, LogicalEvidenceFileClass lef,
  LogicalEvidenceFileClass::DataClass item) {
  //Add all entries to an array
  forall (EntryClass entry in root) {
    if (!entry.IsFolder() && (!SelectedItems || entry.IsSelected())) {
      EntryArray.Add(entry);
    }
  }
  //Random selection based on SampleSize
  for (int i=0; i < SampleSize; ++i) {
    int rnd = SystemClass::Random(EntryArray.Count());
    SampleArray.Add(EntryArray[rnd]);
//delete from array after added so we don’t get duplicates
    EntryArray.Delete(rnd);

  }

  //sanity check
  Console.WriteLine("Sample array contains " + SampleArray.Count() + " entries");
  SystemClass::StatusRange("Saving sample to LEF: " + SampleArray.Count(),
      (1.0*SampleArray.Count()));
  ulong statCount = SampleArray.Count();


  //send sample array entries to LEF
  forall (EntryClass e in SampleArray) {
    SystemClass::StatusInc();
```

```
      SystemClass::StatusMessage("Saving sample to LEF: " + (--statCount));
      item.SetTarget(e, LogicalEvidenceFileClass::DataClass::MODEENTRY);
      if (!lef.Add(item))
        Console.WriteLine("Could not add to Logical evidence file: " +
            GetAdjustedPath(e.FullPath()));
    }
  }

  void Main(CaseClass c) {
    Case = c;
    DlgSettings();
    SelectedItemsText = CountItemsString(Case.EntryRoot(), SelectedItems);
    SampleSizeDialogClass dialog(this);

    if (dialog.Execute() == SystemClass::OK) {
      SystemClass::ClearConsole();
      String subComments = "Selected Confidence Level: " +
         StatsClass::GetConfidenceLevelString(ConfidenceLevelSelection) + "  " +
             "Selected Margin of Error: " +
                  StatsClass::GetMarginOfErrorString(MarginOfErrorSelection) + "  " +
                         "Population: " + Population + "  Sample Size: " + SampleSize;


      Console.WriteLine("Selected Confidence Level: " +
         StatsClass::GetConfidenceLevelString(ConfidenceLevelSelection));

      Console.WriteLine("Selected Margin of Error: " +
          StatsClass::GetMarginOfErrorString(MarginOfErrorSelection));

      Console.WriteLine("Outputting random sample to LEF: " +
          SampleOutputLefPath);

      Console.WriteLine("Population: " + Population + "\nSample Size: " +
          SampleSize);

      DlgSettings(StorageClass::WRITE);
      LogicalEvidenceFileClass LEF();
      if (LEF.Open(SampleOutputLefPath)) {
        LogicalEvidenceFileClass::DataClass comments(null, "", 0);
        comments.Subject = new SubjectClass(null, "Sample Size Selector");
        comments.Subject.SetComment(subComments);
        ProcessEntries(c.EntryRoot(), LEF, comments);
        LEF.Close();
        Console.WriteLine("Completed output to LEF");
      }
      else {
        Console.WriteLine("Could not open LEF: " + SampleOutputLefPath);
      }
    }
  }
}
```

11.2 Cpp Skin Detector

```
#include "opencv2/core/core.hpp"
#include "opencv2/highgui/highgui.hpp"
#include <iostream>
#include <stdio.h>

using namespace std;
using namespace cv;

Mat SkinDetect( Mat matImageIn );

int main(int argc, const char **argv)
{
  Mat matImage, matImageSkin;
  char *pSzFilename;

  /* Get input image file name from command line */

  if ( argc != 2 ) {
    cerr << "usage: " << argv[0] << " <imagefile>" << endl;
    return -1;
  }
  pSzFilename = _strdup( argv[1] );

  /* Read image */

  matImage = imread(pSzFilename, CV_LOAD_IMAGE_UNCHANGED);

  if ( !matImage.data ) {
    cerr << __FILE__ << " " << __LINE__ << ": Cannot load image "
         << pSzFilename << endl;
    return -1;
  }

  /* Detect skin */

  matImageSkin = SkinDetect( matImage );
  if ( !matImageSkin.data ) {
    cerr << __FILE__ << " " << __LINE__ << ": skin detection failed "
         << endl;
    return -1;
  }
  imwrite( "skin.jpg", matImageSkin );

  /* Display the image and wait for user to press a key */

  imshow( "Original", matImage );
  imshow( "Skin pixels", matImageSkin );
  waitKey(0);

  return 0;
}


/* Histogram data structure */

#define HISTSIZE 4096
```

```
typedef struct {
  int cBins;
  double aHist[HISTSIZE];
} tHistogram;

/* Read a histogram data structure from a file */

tHistogram *HistogramRead( char *pSzFilename ) {
  tHistogram *pHistogram;
  FILE *pFile;
  int cObj;

  /* Allocate memory */

  pHistogram = (tHistogram *)malloc( sizeof( tHistogram ));
  if ( !pHistogram ) return NULL;

  /* Open the file */

  pFile = fopen( pSzFilename, "rb" );
  if ( !pFile ) {
    cerr << __FILE__ << " " << __LINE__ << ": cannot open file "
         << pSzFilename << endl;
    free( pHistogram );
    return NULL;
  }

  /* Read the histogram data */

  cObj = fread( &pHistogram->cBins, 1, 4, pFile );
  if ( cObj == 4 ) {
    cObj = fread( &pHistogram->aHist, sizeof( double ), HISTSIZE, pFile )
        ;
  }
  if ( cObj != HISTSIZE ) {
    cerr << __FILE__ << " " << __LINE__
         << ": could not read sufficient data from " << pSzFilename <<
             endl;
    free( pHistogram );
    fclose( pFile );
    return NULL;
  }

  /* Clean up */

  fclose( pFile );
  return pHistogram;
}


/* Free a histogram data structure */

void HistogramFree( tHistogram **ppHistogram ) {
  if ( ppHistogram && *ppHistogram ) {
    free( *ppHistogram );
    *ppHistogram = NULL;
  }
}
```

```

#define MAX3(r,g,b) ((r)>(g)?((r)>(b)?(r):(b)):((g)>(b)?(g):(b)))
#define MIN3(r,g,b) ((r)<(g)?((r)<(b)?(r):(b)):((g)<(b)?(g):(b)))

Mat SkinDetect( Mat matImageIn ) {
  Mat matImageOut;
  Size size;
  int iRow, iCol;
  uchar *aPixelIn, *aPixelOut;
  tHistogram *pHistSkin;
  tHistogram *pHistNonSkin;
  double fHistvalSkin, fHistvalNonSkin;

  pHistSkin = HistogramRead( (char *)"./poshist_hsv16bins.dat" );
  pHistNonSkin = HistogramRead( (char *)"./neghist_hsv16bins.dat" );

  if ( pHistSkin == NULL || pHistNonSkin == NULL ) {
    cerr << __FILE__ << " " << __LINE__
         << ": error: cannot read skin histogram data" << endl;
    return matImageOut;
  }

  if ( matImageIn.channels() != 3 ) {
    cerr << __FILE__ << " " << __LINE__
         << "error: input image is not RGB" << endl;
    return matImageOut;
  }

  size.width = matImageIn.cols;
  size.height = matImageIn.rows;
  matImageOut = cv::Mat( size, CV_8UC1 );

  aPixelIn = (uchar *)matImageIn.data;
  aPixelOut = (uchar *)matImageOut.data;

  for ( iRow = 0; iRow < size.height; iRow++ ) {
    for ( iCol = 0; iCol < size.width; iCol++ ) {

      int R, B, G, F, I, X, H, S, V;

      /* Get RGB values -- OpenCV stores RGB images in BGR order!! */
      B = aPixelIn[ iRow * matImageIn.step + iCol * 3 + 0 ];
      G = aPixelIn[ iRow * matImageIn.step + iCol * 3 + 1 ];
      R = aPixelIn[ iRow * matImageIn.step + iCol * 3 + 2 ];

      /* Convert RGB to HSV */
      X = MIN3( R, G, B );
      V = MAX3( R, G, B );

      if ( V == X ) {
        H = 0; S = 0;
      } else {
        S = (int)((float)(V-X)/(float)V * 255.0);
        F = ( R==V ) ? (G-B) : (( G==V ) ? ( B-R ) : ( R-G ));
        I = ( R==V ) ? 0 : (( G==V ) ? 2 : 4 );
        H = (int)((I+(float)F/(float)(V-X))/6.0*255.0);
        if ( H < 0 ) H += 255;
        if ( H < 0 || H > 255 || V < 0 || V > 255 ) {
          fprintf( stderr, "%s %d: bad HS values: %d,%d\n",
              __FILE__, __LINE__, H, S );
          exit( -1 );
        }
      }

      /* Look up this H/S value in the skin/non-skin histogram tables */

      H = H >> 4;
      S = S >> 4;
      fHistvalSkin = pHistSkin->aHist[H*16+S];
      fHistvalNonSkin = pHistNonSkin->aHist[H*16+S];

      /* Set output pixel based on whether this is skin or not */

      if ( fHistvalSkin > fHistvalNonSkin ) {
        aPixelOut[ iRow * matImageOut.step + iCol ] = V;
      } else {
        aPixelOut[ iRow * matImageOut.step + iCol ] = 0;
      }
    }
  }

  HistogramFree( &pHistSkin );
  HistogramFree( &pHistNonSkin );

  return matImageOut;
}
```

Code 3: CPP skin detector

References

[1] Rosamaria Berté. "Tecniche di Triage applicate alla Digital Forensics". PhD Thesis. Università degli Studi di Roma Tor Vergata, 2010.

[2] Brian Carrier. "Defining Digital Forensic Examination and Analysis Tools". In: *Digital Forensic Research Workshop*. Ed. by DFRWS. 2002.

[3] Brian Carrier. *File System Forensic Analysis*. Addison-Wesley, 2005.

[4] Association of Chief Police Officers. *ACPO Good Practice Guide for Digital Evidence*. 2012.

[5] Ronald C Dodge. "Skype Fingerprint". In: *Proceedings of the 41st Hawaii International Conference on System Sciences*. 2008.

[6] Wouter S. van Dongen. "Forensic Artefacts Left by Windows Live Messenger 8.0". In: *Digital Investigation* (2007).

[7] jiao Feng et al. "Detecting adult image using multiple features". In: *Proceedings of the ICII 2001*. Vol. 3. 2001, pp. 378–383.

[8] Margaret M. Fleck, David A. Forsyth, and Chris Bregler. "Finding Naked People". In: *Proceedings of the 4th European Conference on Computer Vision*. 1996.

[9] Magnet Forensics. *How does Chrome's 'incognito' mode affect digital forensics?* 2013. URL: `http://www.magnetforensics.com/how-does-chromes-incognito-mode-affect-digital-forensics/`.

[10] Magnet Forensics. *How Private is Internet Explorer's InPrivate Browsing?... First define "private"*. 2013. URL: `http://www.magnetforensics.com/how-private-is-internet-explorers-inprivate-browsing-first-define-private/`.

[11] Yanjun Fu and Weiqiang Wang. "Fast and Effectively Identify Pornographic Images". In: *Computational Intelligence and Security (CIS), 2011 Seventh International Conference on*. IEEE. 2011, pp. 1122–1126.

[12] Andrea Ghirardini and Gabriele Faggioli. *Computer Forensics new edition*. Milano: Apogeo, 2009.

[13] Digital Forensic Research Group. *DFRWS*. URL: `http://www.dfrws.org/`.

[14] Brian Jones, Syd Pleno, and Michael Wilkinson. "The use of random sampling in investigations involving child abuse material". In: *Digital Investigation* 9 (2012), S99–S107.

[15] Michael Jones et al. "Statistical Color Models with Application to Skin Detection". In: *International Journal of Computer Vision*. 1999, pp. 274–280.

[16] *Legge 20 maggio 1970 n.300 e successive disposizioni*. 1970. URL: `http://www.wikilabour.it/Default.aspx?Page=statuto%20dei%20lavoratori`.

[17] Jorge A Marcial-Basilio et al. "Detection of pornographic digital images". In: *International journal of computers* 2 (2010), pp. 298–305.

[18] Garante Privacy. *linee guida del Garante per posta elettronica e internet*. 2007. URL: `http://www.garanteprivacy.it/garante/doc.jsp?ID=1387522`.

[19] Chris Prosise, Kevin Mandia, and Matt Pepe. *Incident Response and Computer Forensics, Second Edition*. United States: McGraw Hill, 2003.

[20] Clayton Santos, Eulanda M dos Santos, and Eduardo Souto. “Nudity detection based on image zoning”. In: *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*. IEEE. 2012, pp. 1098–1103.

[21] Raimondo Schettini et al. “On the detection of pornographic digital images”. In: vol. 5150. 2003, pp. 2105–2113.

[22] Glen Thompson. “Automatic detection of child pornography”. In: *Proceedings of the 7th Australian Digital Forensics Conference*. Ed. by Edith Cowan University. 2009.

[23] James Ze Wang, Gio Wiederhold, and Oscar Firschein. *System for Screening Objectionable Images Using Daubechies’ Wavelets and Color Histograms*. 1997.

[24] Jinfeng Yang et al. “A novel approach to detecting adult images”. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 4. IEEE. 2004, pp. 479–482.

[25] Wei Zeng et al. “Image guarder: An intelligent detector for adult images”. In: *Asian conference on computer vision*. 2004, pp. 1080–1084.

[26] Huicheng Zheng, Hongmei Liu, and Mohamed Daoudi. “Blocking objectionable images: adult images and harmful symbols”. In: *Multimedia and Expo, 2004. ICME’04. 2004 IEEE International Conference on*. Vol. 2. IEEE. 2004, pp. 1223–1226.

[27] Qiang Zhu et al. “An adaptive skin model and its application to objectionable image filtering”. In: *Proceedings of the 12th annual ACM international conference on Multimedia*. ACM. 2004, pp. 56–63.