Digital Profiling: A Computer Forensics Approach

Clara Colombini[1,*] and Antonio Colella[2,**]

[1] External researcher at University of Milan, Italy
cmcolombini@email.it
[2] Lieutenant Colonel Italian Army, Rome, Italy
colella@acm.org

Abstract. Nowadays investigations have become more difficult than in the past. It is already clear that, in modern crime scene, a vast amount of evidence are in the electronic or digital form and that the computer system or network have a paramount role in researching of indicators and evidence. The correct analysis of log file and the data saved in the system memory, in this new scenario, are crucial for understanding the criminal actions. Moreover, in order to transform these new elements in evidence, it is important, as well, do not lose sight of the goal of the investigative process and namely identify the perpetrator, even in the cases in which the association of the criminal and of the computer, where crime has been committed, is difficult.

This paper, under this prospective, aims to recognize an alternative investigation approach to traditional criminal profiling. Starting from digital evidence left on the computer system, this research suggests an analytic methodology useful to draw a compatible user digital profile in conjunctions to the evidence left on the system.

Keywords: Hacking profiling, modus operandi, data mining, criminal behaviour, hackers signature.

1 Introduction

The development of modern technology has led to an evolution in the role of digital devices that, now, turned from data containers to a sort of "digital diaries".

The software is being implemented on a growing number of digital devices with a high level of personalization: the agendas of meetings, access to chat rooms, blogs, forums, social networks, etc.., have now turned the phone, MP3 player, game console, satellite navigation, in real custodians of the lifestyle of the individual who normally uses it.

The Digital Profiling, in this scenario, offers a new tool to digital investigation. It analyses the digital memory through specific technical and intelligence profiling, in order to obtain information with which it is possible to reconstruct the user fingerprint and description of its *modus operandi*.

---

[*] Digital Forensics Consultant for the Italian Prosecutor's Office, Italy and member of IISFA Italian Chapter, International Information System Forensics Association (http://www.iisfa.it).

[**] Criminologist and Computer Forensics Expert, Professor at Master of Art in Forensics Science, University of Rome La Sapienza and member of IISFA Italian Chapter, International Information System Forensics Association (http://www.iisfa.it).

A M. Tjoa et al. (Eds.): ARES 2011, LNCS 6908, pp. 330–343, 2011.

The process starts from research and analysis of all the information that can be gathered from "digital footprints" left on PC. The computer is a machine and its user tends to customizes the electronic environment, as well as, normally he do in the real world. Thus, user cannot avoid to leave, even unconsciously, evidences that can be detected, recognized and compared.

2 Techniques of Digital Profiling Analysis of a Computer System

The process of Digital Profiling that has been developed in this research includes six steps:

1) identify the goal: what to look for in relation with the type of problem;
2) collect and assess targeted data from mass memory;
3) selection of relevant information and extraction of *indicators*;
4) information matching of data *(indicators*);
5) collection of information (previously compared) and develop a *"digital profile*";
6) interpretation of the result in comparison to the initial goal.

3 The Method

The Digital Profiling is based on a method that includes mining, comparison and recognition of digital profiles of a user digital device. Identification is done through the comparison of a digital basic profile, built with those data collected from PC and directly attributable to subject under investigation, and all possible profiles extracted from other digital devices on which crimes were committed. It should be noted that the method, upon which is based, is a two-way method, that means you can also start from user digital profile "anonymous" of the device, for comparison with profiles of other devices (also not involved in the offence) attributed with certainty to particular subjects. It can also extract a digital profile of a modus operandi (e.g. cyber attack) to compare with others in order to recognize and identify the author.

The method comprises the following steps which describe a cycle that can be repeated whenever new information is added:

- extrapolation of a basic user digital profile established as "standard profile";
- extrapolation of the users profiles from the digital devices in any other analysis;
- comparison of the profiles in order to highlight convergence-divergence;
- quantitative and qualitative analysis of the convergence-divergence among the profiles for identification of the subject.

4 The Model

The creation of model starts from the study of information characterizing the detected files on a PC and the devices based on the memory capacity, and for high degree of customization allowed by all available applications.

The model describes the elements, the profiles, the features and functions of the elements, the sequence of operations to create the digital profile, the comparison, the evaluation of the result.

5 The Characteristics and Functions of the Elements

5.1 D - Digital Device

Digital device "D_i " means:

- any digital device provided with permanent memory capable of storing files. Example: PC, mobile phone, navigation system, etc..;
- data storage device. Example: Hard drive, smart card, USBpen, etc..);
- remote data storage area created by users;
- virtual machine containing an operating system;
- set of data file access. Example: log file.

5.2 Feature - f

For feature "f_i" is defined the single basic hardware or software feature, derived from the files stored inside the device and selected on the basis of objective investigation, describing the "digital behaviour" of the user, that cannot be broken down further more in the context of the study. It may consist of:

- file properties (metadata type);
- content of the file (type of information);
- a file may contain one or more feature: they are considered basic features, depending on the purpose of the investigation:

– *Filename*. Example: texts, photographs, music, movies, videos, etc...;
– *Path.* Example: some files seem identical, this feature indicates if this file has the same location in the folder tree with respect to another one (same folder name or set of folders);
– *MD5* (or other hash algorithm).The features provides the mathematical certainty of coincidences among the same files found on various devices;
– *Date of creation, modification, deletion.* These three features provide a history of saving, editing, deleting the same files found on other devices;
– Any type of information relevant to the target can be taken from its content.

5.3 Area of File - A

The file that can potentially be considered as feature inside device memory called generically D, is divided in specific areas with expression A_i (D), according to typology and in order to better identified them.

$$\bigcup_i A_i(D) \subset D \tag{1}$$

It defines the $A_i(D)$ as the homogeneous subset of D that contains all the different kind of files that may contain features relative to the device D.

5.4 Classification of Areas of File A

Each device has its own specific line mapping file that contains features and available applications. This is a generic classification of the basic areas related to PC. The number of research areas of this feature is flexible and depends on the type of research and applications on the device.

A_1 _ Registry File: system users.
A_2 - Registry File: hardware installation.
A_3 - Registry File: software installations.

It is considered as "personal files" all those files stored on the user device, excluding the installed programs, which may contain information that characterize the "digital behaviour" of the users. The area of personal files has been divided by type of file in the following categories:

A_4 - Text personal file - Text files were written by the user (notes, memoranda, personal letters, etc..) (file doc, docx, txt, rtf, odt, pdf, xls, etc...) that reveal the writing style. Their analysis can highlight several features. In addition to information that can provide through metadata analysis, other features can be detected by the content of the following files: signature, nickname, proper name, password to access, idiom, misspelling, typing mistakes, reference to a specific event, reference to a particular person, reference to a given object, reference to a place, particular phrases, email address, etc.

A_5 - Personal email messages (except for newsletters, advertising, etc.).
A_6 – Chats.
A_7 - Images ((bmp, jpg, tif, etc.) - Photographs taken from cameras, cell phones, etc.
A_8 - Graphic images (jpg, tif, dwg, etc.).- Collections of graphic images, such as DVD covers, CD, thematic collections of pictures, art, comics, etc.
A_9 - Movies video (Mpg avi., etc..) - Movies made by video cameras, cell phones, etc..
A_{10} - Audio files (wav, mp3, etc.) - Collection of audio files stored by the user.
A_{11} – URL – Connection logs to personal webpages, FTP connections, etc.

5.5 Collection of Feature - F

The analysis of the different areas, points out a set of basic features. However, with Feature F is defined a set of all the individual background characteristics analysed in a digital device.

$$F = \{f_1 (A_i)(D_i), f_2(A_i)(D_i), \ldots f_n(A_i) (D_i)\} \quad (2)$$

5.6 Minimum Feature - m

Once you fix the set of the maximum possible feature detectable from the device, it must be reduced to the features actually present on the device under analysis,

according to the specific requirements of the investigation. For a particular device the order is compose an initial selection of features, , which restricts the number to form the minimum set of features. Therefore, the m_i is a consistent feature, which belongs to all the basic features, selected in relation with the specific investigation.

$$m_i(A_i)(D_i) \in F(D_i) \tag{3}$$

The name of this minimal feature is therefore: $\mathbf{m_i(A_i)(D_i)}$ where:
$\mathbf{m_i}$ - identifies the minimum feature;
$\mathbf{A_i}$ - identifies the area belonging to the source file;
$\mathbf{D_i}$ - identifies the digital device from which it was extracted.

5.7 Minimum Set of Features - M

A subset S (D_i) in relation to the individual case under investigation is defined as the minimum set of features.

$$M(D_i) \in F(D_i)$$
$$M(D_i)= \{m_1(A_i)(D_i)\{m_2(A_i)(D_i),... \{m_n(A_i)(D_i)\} \tag{4}$$

The set of features is the minimum set of filters applying to the files for the extraction of characteristic information *(indicators)* that will make the digital profile.

5.8 Indicator - s

The indicator represents the "single information" collected and analysed in the context of study for the purpose of profiling. It is obtained from the files selected by the application of minimum features filter $\mathbf{m_i}$, during the generation of the digital profile. It is defined as $\mathbf{i_i(l_i)(A_i)(D_i)}$ information, in a specific area $\mathbf{(A_i)}$ of a device $\mathbf{D(_i)}$. $\mathbf{(l_i)}$ identifies the file from which the indicator has been extracted.The indicator is a *digital evidence* and can be detected, recognized and compared as well.

5.9 Set of Indicators - I

It is defined as the set of indicators $\mathbf{I(D_i)}$:

$$I(D_i)= \{i_1(l_i)(A_i)(D_i) \ i_2(l_i)(A_i)(D_i) \ ... \ i_n(l_i)(A_i)(D_i)\} \tag{5}$$

The set of indicators that characterize all the information is collected from the files. It describes the user device "*digital behaviour*" under analysis.

5.10 File That Contains Indicators - k

$\mathbf{k_i(A_i)(D_i)}$ uniquely identifies every file that contains one or more indicators, when:

- $\mathbf{(A_i)}$ identifies the area where you found the file;
- $\mathbf{(D_i)}$ identifies the device.

The file that contains one or more indicators is the *"source of digital evidence*" confirming the source of the indicator.

5.11 Set of Files Containing the Indicators - K

K (D_i) defines the set of files that contain information related to a specific device (D_i).

$$K(D_i) = \{k_1(A_i)(D_i)\ \ k_2(A_i)(D_i)\ \ ...\ k_n(A_i)(D_i)\} \quad (6)$$

6 The Sequence of Operations Useful for the Creation of Digital Profile

The sequence of operations includes the following extrapolation of five profiles from a PC:

(1) the profile obtained from the log files,
(2) the profile obtained from the files in the user folder ,
(3) the profile obtained from the files in the remaining areas of memory .

From which are derived :

(4) the user profile, formed by their union;
(5) the model profile, which matches with the user profile, but refers to a device selected as the "*sample*" for the comparison with others. From the sample profile are drawn: the indicators or the information characterizing to be used for comparison with other profiles for user identification and the file containing them (test wells).

Having in mind that a PC can identify the presence of multiple users, the above mentioned explanation of the method, presents an example of the digital profile of a personal computer referred to a single user, in relation to a Microsoft Windows operating system.

6.1 Profile System - Ps

Starting point is the log files (Area A_1), providing all the information *(indicators)* about the user machine configuration. They will form the profile of system $\mathbf{Ps_i(D)_i}$, where (D_i) identifies a specific device.

$$Ps_i(D_i) = I(Ps_i)(D_i) \cup K(Ps_i)(D_i) \quad (7)$$

where:

- the set of indicators collected from the log files, called **I (Ps_i)(D_i)** where:
  - **I** - all the indicators measured;
  - **Ps_i** - identifies the specific profile of system**;**
  - **D_i** - identifies the specific device.

- the set of files that contains them, called **K (Ps_i)(D_i)** where:
  - **K** - set of files;
  - **Ps_i** - Identifies the specific profile of system;
  - **D_i** - Identifies the specific device.

  in which:
  - each indicator consists of a single information which cannot be further decomposable;
  - each indicator refers to one or more files;
  - each file can contain one or more indicators.

6.2 User Profile Folder - Pc

The Second step is the analysis of files stored in folders created by the operating system for each user. In fact, they contains the most "personalized files" made by the user. This creates a profile called PC (D_i) (user profile folder), based on the analysis of files in the folder you created on the operating system of the device D_i..There is a PC for every user folder found in the PC memory.

(e.g, D_1:PC OS Windows XP: c: \ Documents and Settings \ SuperPippo \ ...).

If there are multiple operating systems (including OS contained in virtual machines), each of them should be treated as a separate device. The user profile folder **$Pc_i(D_i)$** is defined as:

$$Pc_i(D_i) = I(Pc_i)(D_i) \cup K(Pc_i)(D_i) \quad (8)$$

where:

- **$I(Pc_i)(D_i)$** is the set of indicators collected by the files in your user folder, where:
  - **I_i** - set of indicators collected
  - **Pc_i** - identifies the user profile folder
  - **D_i** - identifies the device
- **$K(Pc_i)(D_i)$** is the set of files that contains them, in which
  - **K** - set of file
  - **Pc_i** - identifies the user profile folder
  - **D_i** - identifies the device

  in which:
  - each indicator consists of a single information which cannot be further decomposable,
  - each indicator refers to one or more files,
  - each file can contain one or more indicators.

6.3 Device Profile - Pd

The creation of the user profile folder is not sufficient to delineate the entire profile of the user machine, since other features can be detected from files stored in areas not included in the user folders. The Device Profile includes those files, for example, contained in directory on other partitions, on additional hard disks, including also deallocated files, etc. .. A second round is done so that, using the feature of all M

(minimum feature), which aims to highlight all those feature file containing stored outside the user folders. The Device Profile $\mathbf{Pd_i(D_i)}$, is defined as:

$$Pd_i(D_i) = I\,(Pd_i)(D_i) \cup K\,(Pd_i)(D_i) \quad (9)$$

where:

- the set of indicators drawn from the files contained in your user folder, called $\mathbf{I_i}$ $\mathbf{(Pd_i)(D_i)}$, where:
  - **I** - all the indicators measured
  - $\mathbf{Pd_i}$ **-** identifies the device profile
  - $\mathbf{D_i}$, - identifies the device
- all the file that contains them, called $\mathbf{K_i(Pd_i)(D_i)}$, where:
  - **K** **-** set of files
  - $\mathbf{Pd_i}$ **-** identifies the device profile
  - $\mathbf{D_i}$, - identifies the device

  in which:
  - each indicator consists of a single information which cannot be further decomposable;
  - each indicator refers to one or more files;
  - each file can contain one or more indicators.

6.4 User Profile - Pu

The profiles that are extrapolated so far (see **Fig. 1)** consist of all the elements necessary for creating the user profile called $Pu(D_i)$. It is the digital behavioural model that describes the user interaction with the digital device under analysis. It is therefore composed of:

- all the characterizing information *(indicators)* that are recognized on the entire machine during the analysis,
- all files that contain them.

The user profile Pu (D_i) is then defined by:

$$Pu\,(D_i) = I\,(Pu)\,(D_i) \cup K\,(Pu)\,(D_i) \quad (10)$$

where:

- **I (Pu) (D_i)** - derive from the union of the three sets of indicators reported:
  $$I\,(Ps)\,(D_i) \cup I\,(Pc)\,(D_i) \cup I\,(Pd)\,(D_i)$$
- **K (Pu) (D_i)** - all derive from the union of three sets of files:
  $$K\,(Ps)\,(D_i) \cup K\,(Pc)\,(D_i) \cup K\,(Pd)\,(D_i).$$

in which each indicator is no further information from a single piece:

- each indicator refers to one or more files;
- each file can contain one or more indicator.

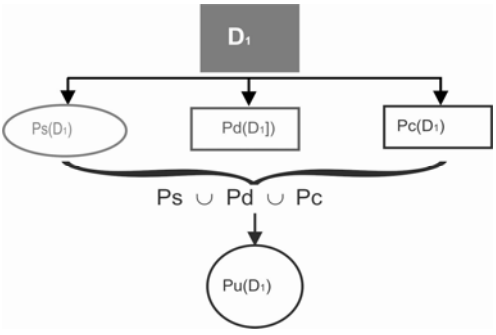

Fig. 1. - The user profile Pu

The follow table summarize the coincident indicators detected by the comparison of profiles.

Table 1. Sample summary of coincident indicators detected by the comparison of profiles.

Reference files (sources)	FEATURE (filter applied)	Indicator
$k_1(A_1)(D_1)$ - SAM	$m8(A_1)$ – computer name	$i_1(k_1)\ (A_1)(D_1)$ - PC_SuperPippo
	$m9(A_1)$ – user system name	$i_2(k_2)\ (A_1)(D_1)$ - SuperPippo
$k_2(A_1)(D_1)$ - SYSTEM.DAT	$m10(A_1)$ - name of installed hardware	$i_3(k_3)\ (A_1)(D_1)$ - USBpen Trust
	$m14(A_1)$ - hardware installed-serial	$i_4(K_4)\ (A_1)(D_1)$ - A01234567
$k_3(A_1)(D_1)$ - SOFTWARE.DAT	$m13(A_1)$ - software installed: nome	$i_5(K_5)\ (A_1)(D_1)$ - AVAST v1.34
	$m14(A_1)$ - software installed: serial	$i_6(K_6)\ (A_1)(D_1)$ - AD1234DC1234
$k_4\ (A_2)(D_1)$ - XXX.DOC	$m_1\ (A_2)$ - Nome file	$i_7\ (k_4)(A_2)(D_1)$ - xxx.doc
	$m_6\ (A_2)$ - Path	$i_8\ (k_4)(A_2)(D_1)$ - c:\Documents and Settings\SuperPippo\Desktop\XXX\
	$m_{16}(A_2)$ - nickname	$i_9\ (k_4)(A_2)(D_1)$ - ilgiaguaro
	$m_7\ (A_2)$ - MD5	B1E5CBE1E019E12E5B73EB4AFB619B5A
$k_5\ (A_2)(D_1)$ - NOTAMIA.TXT	$m_1\ (A_2)$ - Nome file	$i_{10}\ (k_5)(A_2)(D_1)$ - Notamia.txt
	$m_{16}(A_2)$ - nickname	$i_{11}\ (k_5)(A_2)(D_1)$ - superpippo
	$m_6\ (A_2)$ - Path	$i_{12}\ (k_5)(A_2)(D_1)$ - c:\Documents and Settings\SuperPippo\Desktop\XXX\
	$m_7\ (A_2)$ - MD5	$i_{13}\ (k_5)(A_2)(D_1)$ - C1E5CBE1E019E12E5B73EB4AFB619B5A
$k_6\ (A_3)(D_1)$ - message01.eml	$m_{28}\ (A_3)$ – email address	$i_{14}\ (k_6)(A_3)(D_1)$ - superpippo@lamiaposta.com
	$m_{28}\ (A_3)$ - email address	$i_{15}\ (k_6)(A_3)(D_1)$ - ilgiaguaro@jahoo.com
$k_7\ (A_3)(D_1)$ - message02.eml	$m_{28}\ (A_3)$ - email address	$i_{16}\ (k_7)(A_3)(D_1)$ - superpippo@lamiaposta.com
	$m_{28}\ (A_3)$ - email address	$i_{17}\ (k_7)(A_3)(D_1)$ - ilgiaguaro@jahoo.com
$k_8\ (A_3)(D_1)$ - message03.eml	$m_{28}\ (A_3)$ - email address	$i_{18}\ (k_8)(A_3)(D_1)$ - superpippo@lamiaposta.com
	$m_{28}\ (A_3)$ - email address	$i_{19}\ (k_8)(A_3)(D_1)$ - ilgiaguaro@jahoo.com
$k_9\ (A_4)(D_1)$ - 0261f112b3f57021.dat	$m_{19}(A_4)$ - idiomatic expression	$i_{20}\ (k_9)(A_4)(D_1)$ - ola hombre
	$m_{16}(A_4)$ - Nickname	$i_{21}\ (k_9)(A_4)(D_1)$ - ilgiaguaro
	$m_{16}(A_4)$ - Nickname	$i_{22}\ (k_9)(A_4)(D_1)$ - superpippo
	$m_{27}(A_4)$ - particolar phrase	$i_{23}\ (k_9)(A_4)(D_1)$ - non mi hai lasciato le sigarette nel solito posto ieri
	$m_{24}(A_4)$ – reference to an object	$i_{24}\ (k_9)(A_4)(D_1)$ - sigarette
	$m_{25}(A_4)$ - reference to a place	$i_{25}\ (k_9)(A_4)(D_1)$ - solito posto
	$m_{26}\ (A_4)$ - reference to a data	$i_{26}\ (k_9)(A_4)(D_1)$ - 24/12/2009
	$m_{22}(A_4)$ - reference to an event	$i_{27}\ (k_9)(A_4)(D_1)$ - mancata consegna
	$m_{23}(A_4)$ - reference to a person	$i_{28}\ (k_9)(A_4)(D_1)$ - giaguaro
$k_{10}\ (A_5)(D_1)$ - DSC_0001.jpg	$m_1(A_5)$ - file name	$i_{29}\ (k_{10})(A_5)(D_1)$ - DSC_0001.jpg
	$m_6(A_5)$ - path	$i_{23}\ (k_{10})(A_5)(D_1)$ - c:\Documents and Settings\SuperPippo\101ND040\
	$m_{32}(A_5)$ – image of a specific object	$i_{30}\ (k_{10})(A_5)(D_1)$ - yellow car with palate nr. MI01234567
	$m_{24}(A_5)$ - reference to an object	$i_{31}\ (k_{10})(A_5)(D_1)$ - yellow car
	$m_{24}(A_5)$ - reference to an object	$i_{32}\ (k_{10})(A_5)(D_1)$ - targa MI01234567
	$m_7(A_5)$ - MD5	$i_{33}\ (k_{10})(A_5)(D_1)$ - D1E5CBE1E019E12E5B73EB4AFB619B5A
$k_{11}\ (A_6)(D_1)$ - Dvd01.tif	$m_1(A_6)$ - nome file	$i_{34}\ (k_{11})(A_6)(D_1)$ - Dvd01.tif
	$m_6(A_6)$ - path	$i_{35}\ (k_{11})(A_6)(D_1)$ - Dvd01.tif c:\Documents and Settings\SuperPippo\Desktop\XXX\copertine dvd\
	$m_7(A_6)$ - MD5	$i_{36}\ (k_{11})(A_6)(D_1)$ - A2E5CBE1E019E12E5B73EB4AFB619B5A

Table 1. (*continued*)

k_{12} (A_6)(D_1) - Dvd02.tif	m_1(A_6) - nome file	i_{37} (k_{12})(A_6)(D_1) - Dvd02.tif
	m_6(A_6) - path	i_{38} (k_{12})(A_6)(D_1) - c:\Documents and Settings\SuperPippo\Desktop\XXX\copertine dvd\
	m_7(A_6) - MD5	i_{39} (k_{12})(A_6)(D_1) - A3E5CBE1E019E12E5B73EB4AFB619B5A
k_{13} (A_6)(D_1) - Dvd03.tif	m_1(A_6) - nome file	i_{40} (k_{13})(A_6)(D_1) - Dvd03.tif
	m_6(A_6) - path	i_{41} (k_{13})(A_6)(D_1) - c:\Documents and Settings\SuperPippo\Desktop\XXX\copertine dvd\
	m_7(A_6) - MD5	i_{42} (k_{13})(A_6)(D_1) - B6E5CBE1E019E12E5B73EB4AFB619B5A
k_{14} (A_8)(D_1) - La cumparsita.mp3	m_1(A_8) - nome file	i_{43} (k_{14})(A_8)(D_1) - La cumparsita.mp3
	m_6(A_8) - path	i_{43} (k_{14})(A_8)(D_1) - c:\Documents and Settings\SuperPippo\Desktop\XXX\miomp3\
	m_7(A_8) - MD5	i_{44} (k_{14})(A_8)(D_1) - C3E5CBE1E019E12E5B73EB4AFB619B5A
k_{15} (A_8)(D_1) - El dindondero.mp3	m_1(A_8) - nome file	i_{45} (k_{14})(A_8)(D_1) - El dindondero.mp3
	m_6(A_8) - path	c:\Documents and Settings\SuperPippo\Desktop\XXX\miomp3\
	m_7(A_8) - MD5	i_{46} (k_{14})(A_8)(D_1) - E6E5CBE1E019E12E5B73EB4AFB619B5A
k_{16} (A_9)(D_1) - History.dat	m_{40}(A_9) - URL	i_{47} (k_{16})(A_9)(D_1) - http://www.facebook.com/superpippo2345cdk0945.php
	m_{40}(A_9) - URL	i_{48} (k_{16})(A_9)(D_1) - http://www.ilmiosito.com/ superpippo234sdfgoap43.php
	m_{37}(A_9) - URL	i_{49} (k_{16})(A_9)(D_1) - http://www.lamiaposta.com/superpippo3456asdf567.php
K_{17} (A_2)(D_1) - carved[123456789].doc	m_{16}(A_2) - nickname	i_{50} (k_{17})(A_2)(D_1) - superpippo
	m_{18}(A_2) - password	i_{51} (k_{17})(A_2)(D_1) - piùvelocedellaluce
	m_{28}(A_2) - indirizzo email	i_{52} (k_{17})(A_2)(D_1) - superpippo@lamiaposta.com
	m_{26}(A_2) - riferimento a dato	i_{53} (k_{17})(A_2)(D_1) - 339123456
	m_7 (A_2) - MD5	i_{54} (k_{17})(A_2)(D_1) - D1E9ABE1E009E12E5B23EB4DFB689B5E
K_{18} (A_5)(D_1) - carved[123456749].jp	m_{32}(A_5) – image of an object	i_{55} (k_{17})(A_5)(D_1) - credit card Bankamericard
	m_{26}(A_2) - reference to an object	i_{56} (k_{17})(A_5)(D_1) - Bankamericard
	m_{26}(A_2) - reference to a data	i_{57} (k_{17})(A_5)(D_1) - 4935 1500 4556 5784
	m_7 (A_2) - MD5	i_{58} (k_{17})(A_5)(D_1) - A1E5CBE1E019E12E5B73EB4AFB619B5A
K_{18} (A_9)(D_1) - carved[123451049].3gp	m_{30} (A_2) – person image	i_{59} (k_{17})(A_9)(D_1) - Rossi Mario
	m_{23} (A_2) - reference to a person	i_{60} (k_{17})(A_2)(D_1) - Rossi Mario
	m_7 (A_2) - MD5	i_{61} (k_{17})(A_2)(D_1) - B1E5CBE1E019E13E5B73EB4AFB619B5D
ORGANIZATION FILES: Organization of personal files and folders of user "SuperPippo" in D_1: c:\Documents and Settings\SuperPippo\Desktop\XXX\ c:\Documents and Settings\SuperPippo\Desktop\XXX\copertine dvd\ c:\Documents and Settings\SuperPippo\Desktop\XXX\miomp3\ c:\Documents and Settings\SuperPippo\101ND040\		

6.5 User Profile Sample - Puc

The user profile sample Puc (D_i) matches with the user profile Pu (D_i), which differs only by definition because it is set as a benchmark for comparison with other devices. In fact , the indicators collected will be used as filters to search for information within the overlapping memories of other devices.

7 The Comparison

Once you have the sample profile Puc (D_1) from a device, the indicators collected are used as filters for the detection of the same profile on other devices, to detect connections and/or differences. The follow describe the comparison of the coincident indicators in the two different devices.

The final step, if necessary, is the comparison between the dates of creation / modification / deletion of files extracted by the two devices in order to reconstruct the history of user actions on devices over time. The example of **Table 2** illustrates how the search for indicators, are extrapolated from the device D_1. The files stored in the device D_2 have 30 information characterizing the user, share (75% of filters applied). They show that both devices were used by the same subject. However, this type of comparison is one way: the search characteristic information is performed based on the indicators found in a single device, called "sample", leaving out the analysis and therefore the search for possible indicators on other devices. To work around this problem you can take an additional step of refining the profiles through the cross referencing, which is based on the contents of memory to all devices.

Table 2. Comparison of the coincident indicators in the two devices

Nr.	Feature	Indicator	D_1	D_2
1	organization folders	... \ SuperPippo \ Desktop \ XXX \	●	●
2	organisation folders	... \ SuperPippo \ Desktop \ XXX \ dvd covers \	●	●
3	organization folders	... \ SuperPippo \ Desktop \ XXX \ miomp3 \	●	●
4	path file	... \ SuperPippo \ Desktop \ dvd covers \ Dvd01.tif	●	●
5	path file	... \ SuperPippo \ Desktop \ dvd covers \ Dvd02.tif	●	●
6	path file	... \ SuperPippo \ Desktop \ XXX \ miomp3 \ The cumparsita.mp3	●	●
7	path file	... \ SuperPippo \ Desktop \ XXX \ miomp3 \ dindondero.mp3	●	●
8	personal file	Dvd01.tif	●	●
9	personal file	Dvd02.tif	●	●
10	personal file	The cumparsita.mp3	●	●
11	file personal	dindondero.mp3	●	●
12	sender email	superpippo@lamiaposta.com	●	●
13	email recipient	ilgiaguaro@jahoo.com	●	●
14	nickname sender skype	SuperPippo	●	●
15	skypenickname	recipient'sfriend jaguar	●	●
16	skype password	piùvelocedellaluce	●	●
17	idiomatic expression	ola hombre	●	●
18	nickname	ilgiaguaro	●	●
19	particular sentence	you left me no cigarettes in the same place yesterday	●	●
20	in reference cigarettes	sigarette	●	●
21	to risereference	usual place	●	●
22	referenceat the date	24/07 / 2010	●	●
23	object reference	Bankamericardns 4935 1500 4556 5784	●	●
24	reference no phone	339123456	●	●
25	refers to vehicles	with yellow number plate Car MI01234567	●	●
26	url	http://www.facebook.com/superpippo2345cdk0945.php	●	●
27	url	http://www.ilmiosito.com/superpippo234sdfgoap43.php	●	●
28	url	http://www.lamiaposta.com/superpippo3456asdf567.php	●	●
29	hardware	USBpen Trust sn A01234567	●	●
30	software	v1.34 sn AVAST AD1234DC1234	●	●

8 Cross Comparison

The step consists of crossing the analysis of all the information gathered for each device (see **Fig. 2**). Its implementation involves the following steps:

1. Puc sample extrapolation of user profiles of all devices in the analysis, each of which will consist of:
   - union of the three sets of indicators reported: $I(Ps)(D_i) \cup I(Pc)(D_i) \cup I(Pd)(D_i)$;
   - union of three sets of files: $K(Ps)(D_i) \cup K(Pc)(D_i) \cup K(Pd)(D_i)$.
2. extraction of the indicators **I (Pu) (D_i)** and its files **K(Pu) (D_i)** from each profile;
3. the application of each set of filters drawn from the indicators **I(Pu)(D_i)** to each of the devices;
4. update individual profiles to new indicators identified.

The procedure, though having the disadvantage of lengthening lead times, may prove useful in cases where the information obtained from the analysis of a single device are not very significant because it allows to analyse the data in all devices, increase the number of indicators obtained and make more consistent the user profiles. It's also allows to detect any additional users.

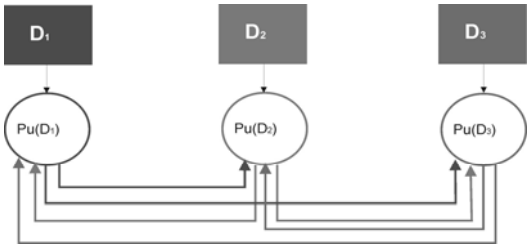

Fig. 2. Comparison of cross

9 Multi-user Devices

A more complex case (see **Fig. 3**) can occur if the same device D_i is used by more than one person (e.g. PC). Then, a profile for each user can be extrapolated according following rules:

1. build a profile Pc for each user (ie Pc_1,Pc_2,etc..) on each of the user folders;
2. build a system profile Ps (eg, Ps_1,Ps_2,etc..) for each user;
3. build a unique profile Pd;
4. cross compare each Pc and Pd that produces so many profiles as there are user devices Pu Pc User Profiles folder;
5. each user profile $Pu_i(D_{i)}$ will be defined as:

$$Pu_i(D_i) = Pc_i(D_i) \cup Pd(Pc_i)(D_i) \cup Ps(Pc_i)(D_i) \quad (11)$$

The comparison between different user profiles folder Pc device profile and Pd are designed to:

- identify their own indicators in the areas of memory included in the device profile;
- extract the files containing them and add them to your **Pdu,** formed by:
  - the set of indicators in common with the PC;
  - all files that contain them.
- create many profiles $\mathbf{Pu_i(D_i)}$ how many user folders (not empty) comprising:

$$Pc_i(D_i) \cup Pdu_i(Pc_i)(D_i) \cup Ps_i(D_i) \quad (12)$$

- decrease the size of the Pd profile that will ultimately be composed of these indicators (and related files) are not included in different user profiles.

The end result is:

- *n* user profiles - the set of characteristic information that describes the behaviour of digital users found on the machine;
- No 1 anonymous Pda Device Profile (if any) - that is, a set of information characterizing not related to those users, which will also include that information on configuring the system does not give users found.

This last profile is not deleted, but is listed as anonymous profile because it contains information that may be useful for the identification of other entities by comparison with other devices in subsequent analysis.

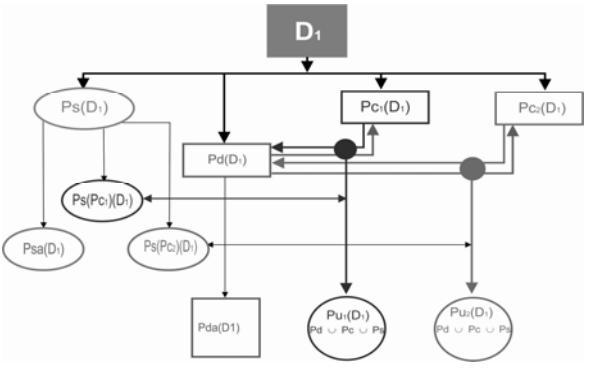

Fig. 3. Process of detection of n. 2 profiles in a multi-user device

10 Evaluation of Results

The evaluation of the result (operation of close relevance investigator), is carried out both in a quantitative sense (ie, considering the number of coincident indicators measured), which in a qualitative sense (ie, the veracity of information), as even a only information can be found as the solution of the problem posed by the analysis goal.

10.1 The Quantitative Assessment

It is carried out in statistical way by calculating the percentage of coincident indicators found by comparing the total of those used as a filter.

EXAMPLE: Quantitative assessment of the results obtained by simple comparison (on the case presented in Table 1):

Task 1 - Create User Profile sample PCU (D_1):
N. filters applied: ....44 RESULT: indicators extracted.....40
Task 2 - Research using filters of the indicators on the device D_2:
N. filters applied: ….40 RESULT: coincident indicators found.....30

On the total of nr. 40 indicators/filter applied by simple comparison has been detected Nr 30 coincident indicators, which is 75%.

10.2 The Qualitative Assessment

This analysis gives to the information obtained (indicators) a value of "relevance" based of the individual indicators in relation to their degree of usefulness for the target. In consideration of the digital nature of the analysis, the sources are not assessed: if properly extracted and verified by hashing algorithms, they are to be considered *"completely reliable."*

With regard to the *information* obtained in the case presented here, it shows no qualitative assessment (under responsibility of the investigator, in the survey), as the specific research described by the example given here was aimed solely to collecting

coincident (ie in possession of only two values: *match / no match*), which could bring with certainty the identity of the same subject in question.

11 Conclusions

The digital profiling is a new computer investigation tool with the aim of extracting information from memory of digital devices and assist computer investigator in their analysis and help them to identify a possible user/criminal digital profile. This type of analysis is suitable to all the devices: to all personal computers, mobile phones, smartphones, tablets etc.

However, embedded devices are not excluded of this methodology: to give just one example, a GPS navigator, even though it may seems at first glance that may not contain data useful to find a solution of a crime, can provide valuable information on the movements of a subject, such as places where has gone, the usual route that, if compare with the position of his home, may help to delineate the aim of its activities.

Digital profiling techniques can also be applied to the contents of storage areas provided in remote provider and data streams selected for example in a certain time on a computer attack.

At the end, this technique is particularly useful in operations against organized crime, anti-terrorism operations, intelligence operations, where it can be interfaced with the statistical study in the prediction and prevention of criminal events.

Acknowledgments. This work was supported by IISFA Italian Chapter, International Information System Forensics Association (http://www.iisfa.it), and does not reflect the official policy or position of the University of Milan and Italian Army General Staff.

References

1. Casey, E.: Digital evidence & computer crime, 2nd edn. Elsevier Academic Press, Amsterdam (2004)
2. Loia, V., Mattiucci, M., Senatore, S., Veniero: Computer Crime Investigation by Means of Fuzzy Semantic Maps, M. Web Intelligence and Intelligent Agent Technologies (2009)
3. Picozzi, M., Zappalà, A.: Criminal Profiling, dall'analisi della scena del delitto al profilo psicologico del criminale. McGraw-Hill, New York (2002)
4. Turvey, B.: Deductive Criminal Profiling: Comparing Applied Methodologies Between Inductive and Deductive Criminal Profiling Tecniques. Knowledge Solutions Library (January 1998)