Mobile Forensics "triaging": new directions for methodology

Authors:

Fabio Marturana, Department of Computer Science, Systems and Production, University of Rome Tor Vergata, Rome, Italy, marturana@libero.it;
Gianluigi Me, Department of Computer Science, Systems and Production University of Rome Tor Vergata, Rome, Italy, me@disp.uniroma2.it;
Rosamaria Bertè, Department of Computer Science, Systems and Production University of Rome Tor Vergata, Rome, Italy, rosamariaberte@libero.it;
Simone Tacconi, Servizio Polizia Postale e delle Comunicazioni Rome, Italy, simone.tacconi@interno.it.

Abstract Over the past few years Mobile Forensics, the branch of Digital Forensics that deals with gathering, retrieving, identifying, storing and documenting mobile phone’s evidence with probative value in court, has become more and more specialized.

Nowadays, specific extraction tools have been developed in order to acquire and store phone’s content and digital evidence, in compliance with forensic methods.

A new approach to Mobile Forensics could therefore take advantage of mixing up features of the aforementioned extraction tools with capabilities of “Data Mining” and “Machine Learning” theory with the aim of defining a methodology to quickly analyze the extracted data and provide a classification.

This paper aims at explaining some interesting results based on the Mobile Forensics “Triaging” concept and the adoption of self-knowledge classification algorithms for predicting and classifying device usage profiles (i.e. base, medium or expert).

In order to give new perspectives to the actual work procedures of the Italian Police cybercrime unit, the adopted methodology has been extensively discussed with specialists, aiming to find a viable methodology to identify the most interesting mobile devices from an investigative point of view by analyzing the device owner’s usage profile, a relevant parameter to consider during forensic investigations.

Keywords: Triaging, Mobile Forensics, Data Mining, Machine Learning.

Introduction

Nowadays cell phones, PDAs and new generation smartphones are widespread all over the world. A 2010 official report, issued by ITU (International Telecommunication Union), indicates a cell phones global market penetration rate above 50% in general and almost 100% with regards to the most developed countries [1].
Cell phones and mobile handsets commercial success is basically the result of two factors: ever-decreasing selling prices, on one side and ever-increasing storage capacity and supported features, on the other.

As a consequence, Cybercrime Police Units have to deal with an ever-increasing number of crimes related to mobile phones usage [2] and therefore they consider cell phones, PDAs and smartphones important items to analyze during investigations. These devices can provide indeed investigators with some useful intelligence regarding suspect habits, interests, social relations and technical skills and let them also reconstruct the crime timeline and modus operandi by collecting and analyzing the evidence stored in the devices [3].

As a result of interviews to the Italian Police cybercrime specialists, we noticed an increase of complexity in today’s forensic investigations due to the ever growing number of new mobile phones released every year, each with its own Operating Systems among Android, RIM, Windows Phone, Apple iOS etc. and a different File System and memory organization. For this reason, in order to fill the gap, some highly specialized Companies developed complex extraction and analysis toolkits in order to provide investigators with all the needed instruments to conduct a technical inquiry [4, 5].

Due to such proliferation, in 2007 the NIST (National Institute of Standards and Technology) issued a report summarizing the results of a comparative analysis among different toolkits available in the marketplace, such as Paraben Device Seizure [6], FTK [7], XRY [8], UFED [9], Mobil-Edit [10], CellDek [11] and reporting their strengths and weaknesses in detail [12].

Moreover, we realized that the 4-steps mobile forensic methodology based on device identification, acquisition, analysis and reporting, resulted to be inadequate since all the sized phones are flatly analyzed by forensics analysts in sequence with the same priority. In fact, during the device identification step, a unique identifier based upon case number, device manufacturer and model is given to every cell phone which is then processed by one or more extraction tools in order to create a device image. Later on the image is analyzed in order to find evidence concerning the crime under investigation and finally the list of evidence is summarized in a report with probative value in court.

We decided therefore to modify the classic forensic procedure introducing an intermediate step, called triage and located between acquisition and analysis which could help investigators to limit the area of interest and reduce the number of relevant devices to focus on. As an example, during a complex investigation involving different persons, crimes and mobile equipment, adopting the triage method, i.e. a “divide-and-conquer” approach which splits up relevant and less important aspects of the case, it is possible to assign a priority to every involved device, person and crime [13, 14]. By means of a quick memory search on the whole set of seized devices, in fact, it is possible to create a list, ordered by probative value, of phones containing the most useful evidence and requiring additional processing at Forensic Lab.

The triage method is based on the adoption of “Data Mining” and “Machine Learning” algorithms, in a fashion that imitates the hospital “triage” protocol by
which first aid personnel assign disease seriousness priority codes to persons requiring medical treatment [15].

The role of triage methods and techniques and the advantage of applying “Data Mining” techniques are extensively discussed in the paper. A methodical approach to triage is presented and the adopted classification algorithms are evaluated and compared by means of key performance indicators.

**Proposed methodology foundations**

Unlike other approach to Mobile Forensics Triaging [13, 14], where the attention is focused on standardizing methods and training cell phone specialists in order to perform on-scene inspection and examination, we decided to focus the attention on a cold data analysis method.

This paragraph describes the followed 3-stage methodology applied on a set of Police reports and seized mobile phone’s extracted data, provided by Servizio Polizia Postale e delle Comunicazioni (the Italian Cybercrime Police Unit) and regarding different crimes such as pedophilia and pedopornography, homicide, non-disclosure agreement violation, human trafficking and extortion.

The extracted data are first normalized, in order to remove the misalignments produced by different extraction tools and then processed in order to create the “dataset”. Finally normalized data are elaborated by means of “Knowledge Analysis” algorithms in order to predict the device owner’s usage profile (base, medium or expert usage). The following figure summarizes the process:

**Fig. 1. Process workflow**
The first stage of the process is the data collection and it’s called crime reports collection (forensic acquisition). It is based on a classical forensic acquisition of a mobile phone memory with the appropriate extraction tool [1, 3, 6]. It can be carried out both on crime scene or in a forensics lab.

The second stage of our workflow is called data normalization and feature extraction; i.e. the extracted data, regarding common/serious and cybercrimes, are firstly normalized, in order to remove the misalignment produced by different extraction tools and then loaded into databases.

This stage is fundamental because of the large, often incomplete, noisy due to outliers, inconsistent and redundant amount of data. According to the so called “Knowledge Discovery Process” [16], before submitting the “dataset” to the classification process, it is necessary to perform some preprocessing operations consisting of data cleaning (i.e. noise reduction), relevance analysis (i.e. redundant attributes elimination) and data transformation (i.e. normalization).

After the aforementioned two stages we are able to create the following data structure, called input matrix:

![Input matrix](image)

<table>
<thead>
<tr>
<th>Attribute name</th>
<th>Case Galaxy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone_model</td>
<td>Smartphone</td>
</tr>
<tr>
<td>Number_phonebook_contacts</td>
<td>11</td>
</tr>
<tr>
<td>Number_received_calls</td>
<td>35</td>
</tr>
<tr>
<td>Number_dialed_calls</td>
<td>133</td>
</tr>
<tr>
<td>Number_missed_calls</td>
<td>17</td>
</tr>
<tr>
<td>Mean_duration_received_calls</td>
<td>38</td>
</tr>
<tr>
<td>Mean_duration_dialed_calls</td>
<td>38</td>
</tr>
<tr>
<td>Number_read_sms</td>
<td>High</td>
</tr>
<tr>
<td>Number_sent_sms</td>
<td>Low</td>
</tr>
<tr>
<td>Percentage_read_sms</td>
<td>High</td>
</tr>
<tr>
<td>Percentage_sent_sms</td>
<td>Low</td>
</tr>
<tr>
<td>Number.URIvisited</td>
<td>73</td>
</tr>
<tr>
<td>user_class</td>
<td>Medium usage</td>
</tr>
</tbody>
</table>

Attribute’s name is indicated in the left column (i.e. the parameter called Number_phonebook_contacts etc.) while, in the right column shows an instance sample of a mobile phone (i.e. the collection of extracted and normalized data), where each attribute has its own correspondent value that can be either nominal or numeric. The matrix in our research consists of 114 rows or attributes and 23 columns, corresponding to the number of analyzed phones.

In particular, we focus our attention on the following parameters/data:

– Phone model (Smartphone, GSM);
– Number of phonebook contacts (stored both on SIM and phone);
– Number of dialed/received/missed calls;
– Percentage of dialed/received/missed calls (with regards to the specific time slot: Morning, Afternoon, Evening and if generated or received from phonebook contacts or not);
– Average duration of dialed/received calls (with regards to the specific time slot: Morning, Afternoon, Evening and if generated or received from phonebook contacts or not);
– Number of received/sent SMS/MMS;
– Percentage of received/sent SMS/MMS (with regards to the specific time slot: Morning, Afternoon, Evening and if they are sent or received from phonebook contacts or not);
– Number and percentage of visited URLs, with regards to the specific time slot: Morning, Afternoon, Evening and if they are bookmarked or not;
– Number and percentage of downloaded images/videos/audio files stored on the device or created by means of the device camera and microphone;
– Number of sent/received E-mail;
– Number of stored Notes.

The third stage of our workflow is called data classification & triaging and it is based on techniques enabling the quick evidence classification in order to highlight the device owner’s usage profile which can be categorized as base, medium or expert.

The classification process can be either supervised or unsupervised; in the first case the set of classes is defined “a priori” and the goal is to identify the class to assign to each sample, called pattern, based on the vector of measurements, called feature; In the second case the set of classes is unknown and the classification is performed by means of clustering techniques which automatically reveal structured data. We adopt the supervised approach creating a training set, i.e. a collection of representative patterns each of them with a known class, in order to train the classifier and a test set, i.e. a collection of real patterns with unknown class, in order to evaluate the accuracy of the method.

We make the assumption that the usage profile is a valuable information because, as investigators’ experience suggests, the more complex the device is, the higher is, likely, the number of interactions between owner and equipment and, as a result, the higher is the probability to find evidence inside it. The goal is therefore to build a priority list among the seized devices where cell phones with an “expert usage” profile are located on top and therefore examined with higher priority than the others.

In order to choose the best performing classifier, we analyzed the WEKA suite [17], making a detailed comparative analysis among the supported algorithms and evaluating their performance on our dataset by means of key performance indicators such as Precision, Recall and F-measure that will be described later. As a result, we compared the following: Bayesian Networks, Decision Tree and LWL-Locally Weighted Learning and, further, we built two custom Decision Tree classifiers, specifically designed for our two training set (numeric and nominal-numeric). In order to build an efficient decision tree and reduce its complexity, we decided to
adopt a stopping rule [18] by setting the maximum tree depth to 6. As a result, with a total of 23 instances to classify, we have a custom tree with depth 5 and 9 leaves and the other with depth 6 and 8 leaves.

All the aforementioned algorithms train themselves by inspecting all the instances included in the training-set and create different statistical models.

With **Bayesian Networks**, what is being estimated is the conditional probability distribution of the values of the class attribute (i.e. the usage class) given the values of the other attributes. Ideally, the classification model represents this conditional distribution in a concise and easily recognizable way. Bayesian network are drawn as a network of nodes, one for each attribute, connected by directed edges in a directed acyclic graph [11].

Adopting a **Decision Tree** algorithm, we apply a “divide-and-conquer” approach to the problem of learning from a set of independent instances; the corresponding binary tree representation consists of nodes implying a test on one or more attributes and leaves giving a classification to all the instances that reach it. For example, in order to classify an unknown instance, the tree is walked according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to that leaf [12].

**Locally Weighted Learning** is a general algorithm and can be applied with any learning technique that can handle weighted instances. In particular, it can be used for classification. It assigns weights using an instance-based method and builds a classifier from the weighted instances; **LWL** only assumes independence within a neighborhood, not globally in the whole instance space [13].

At the end of the “training” process, WEKA creates a statistical model; every classifier is able to assign an usage class to all the instances within the trainingset, comparing the predicted class with the original one. Moreover, classifiers’ effectiveness (i.e. the ratio between performance and acquired knowledge) is evaluated by means of the following indicators:

- **Precision** i.e. the ratio of true positive and the sum of those with false positive (note: if the number of false positive is low the Precision is close to 1);
  \[
  \text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}
  \]

- **Recall** i.e. the ratio of true positive and the sum of those with false negative (note: if the number of false negative is low the Recall is close to 1);
  \[
  \text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
  \]

- **F-Measure** i.e. the harmonic mean of recall and precision:
  \[
  \text{F-measure} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}
  \]

In the former approach we used the entire dataset to calculate the error rate by adopting an iterative and predictive method called **10 folds cross-validation** where the training-set is split into ten approximately equal partitions, each in turn used for testing and the remainder for training. That is, use nine-tenth of the data for training and one-tenth for testing, and repeat the procedure ten times so that in the end, every instance has been used exactly once for testing [17].

As a result, applying **10 folds cross-validation** to both Bayesian Networks and WEKA Decision Tree (called J48) we have the following results:
Table 1. Comparative results for complete dataset

<table>
<thead>
<tr>
<th>Model Classifier</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>BayesNet</td>
<td>0.737</td>
<td>0.739</td>
<td>0.724</td>
</tr>
<tr>
<td>J48</td>
<td>0.825</td>
<td>0.826</td>
<td>0.823</td>
</tr>
</tbody>
</table>

We noticed that performance could be improved reducing the number of attributes, with special regards to those with more missing values (i.e. called outliers) in the training instances. To give an idea of the redundancy of attributes, WEKA Decision Tree (J48), during the training stage, uses only 4 to 5 attributes out of 114 to build the tree.

In order to simplify the problem and create therefore two reduced datasets, respectively with numeric attributes only (numeric training-set) and with both numeric and nominal attributes (nominal training-set); thus we applied some linear algebra concepts to our input matrix $M$ (rif. fig.2), a rectangular matrix $m \times n$ where $m = 114$ and $n = 23$. Firstly, we recall the rank or maximum number of linearly independent rows and columns of a generic $m \times n$ matrix cannot be greater than $m$ nor $n$; thus $\text{rank}(M) \leq \min(m,n)$. In our case $\text{rank}(M)$ is equal or less than 23 thus there are no more than 23 linearly independent rows or attributes in the problem. Therefore, in the first simplified dataset, we reduce manually the number of rows first eliminating the nominal attributes, obtaining a $m' \times n$ matrix with $m' = 63$ and $n = 23$. We apply then the Gauss Elimination algorithm to reduce the $m' \times n$ input matrix to a row-echelon form, finding out that the value of $\text{rank}(M)$ is exactly 23. In the second simplified scenario it is not possible to apply the Gauss Elimination algorithm since we have both numeric and nominal attributes that cannot be elaborated thus we assume that the result is the same and select only 23 attributes.

In both the aforementioned simplified models, we applied a different approach than 10-folds cross validation; in particular first we use the entire dataset to train the classifiers and then we create a test-set with 3 phones/instances, each with a different usage profile in order to evaluate the training results.

Based on numeric attributes, our correspondent custom Decision Tree Classifier performs better than WEKA Decision Tree (called J48), Bayesian Networks and LWL while with nominal/numeric attributes both our custom Decision Tree and Bayesian Networks perform very well. The results are summarized in the two tables below:

Table 2a. Comparative results for numeric set

<table>
<thead>
<tr>
<th>Model Classifier</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>BayesNet</td>
<td>0.333</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>J48</td>
<td>0.167</td>
<td>0.333</td>
<td>0.222</td>
</tr>
<tr>
<td>LWL</td>
<td>0.111</td>
<td>0.333</td>
<td>0.167</td>
</tr>
<tr>
<td>Custom Decision Tree</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2b. Comparative results for numeric/nominal set

<table>
<thead>
<tr>
<th>Model Classifier</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>BayesNet</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J48</td>
<td>0.5</td>
<td>0.667</td>
<td>0.556</td>
</tr>
<tr>
<td>LWL</td>
<td>0.5</td>
<td>0.667</td>
<td>0.556</td>
</tr>
<tr>
<td>Custom Decision Tree</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Conclusion: a new approach to a well-known problem

This paper deals with a new methodological approach to Mobile Forensics, the branch of Digital Forensics that deals with evidence extracted from mobile phones, PDA’s and smartphones according to well defined and standardized methods and techniques with probative value in court.

Interviewing experts and Police cybercrime investigators, we realized indeed that the actual forensics analysis process, based on the rigid 4-steps workflow (Identification, Extraction, Analysis and Reporting) doesn’t fit to the actual scenario made of hundreds new mobile phone models released every year and chronic lack of standardization among manufacturers. For this reason investigators have to deal with a consistent backlog of data extracted from lots of seized phones by means of different forensic toolkit which needed to be processed in Lab in order to find evidence.

The idea behind the research is thus to redefine the forensics workflow, introducing a further step called triage, between extraction and analysis, with the aim of reducing the amount of time spent by cybercrime investigators to find crime evidence in all the sized mobile devices.

The research’s aim can be thus summarized as “a way to help cybercrime investigators to focus their search on the most relevant evidence first, by providing a quick answer to the following question: which is the device owner’s usage profile?”. This result can be achieved by giving a priority based on it to every device after a quick phone-by-phone analysis of the extracted image.

To do so, the backlog of data were processed by means of Knowledge Management classification algorithms such as Bayesian Networks, Decision Trees, and LWL (Locally Weighted Learning) in order to predict the device owner’s usage profile; What we observed is that apparently heterogeneous data could provide very useful information regarding the human-handset interaction. As a conclusion, this work can be considered a first step towards new triage scenarios such as elaborating data under different prospective in order to find, for example, associations between the extracted data and a criminal behavior, such as pedopornography or stalking crimes.

References

4. Access Data, "AD Triage 1.0.0". Released in April 2, 2011.
7. FTK – AccessData.
8. XRY - Micro Systemation (MSAB).
9. UFED - Cellebrite Universal Forensics Extraction Device.
10. MOBILedit! Forensic - COMPELSON Labs.
12. NIST. (March 2007). "Cell Phone Forensic Tools: An Overview and Analysis Update".
13. Robert J. Walls, Erik Learned-Miller and Brian Neil Levine. "Forensic Triage for Mobile Phones with DEC0DE".