DOI: http://dx.doi.org/10.5769/C2011014

Using XSLT Filters to Improve Productivity and Quality on Cell Phone Forensics

Alexandre Vrubel
Computer Forensics Laboratory
Criminalistics Institute of Curitiba - Paraná
Curitiba, Brazil
alexandre.vrubel@ic.pr.gov.br

***Abstract* - The growth of mobile telephony in Brazil has generated a significant increase in the number of cell phones subjected to forensic examinations. Specialized tools aimed to the forensic examination of mobile phones and SIM cards help to tackle this problem, but they usually provide reports with limited customization options. Most tools, however, allow exporting the captured data in XML (eXtensible Markup Language) format. Using the LibreOffice suite and XSLT (eXtensible Stylesheet Language for Transformation) transformation filters, it is possible to automatically convert the data captured by various forensic tools directly into customized and standard reports. The use of these filters allowed an enormous gain in productivity and quality on cell phone forensic reports.**

keywords - cell phones; XSLT filters; LibreOffice; report customization.

1. Introduction

The great expansion of the mobile phone market, combined with the rapid evolution of the mobile phones themselves (smartphones already behave like small computers), has generated a significant increase both in quantity and complexity of the forensic examinations performed in these type of equipments.

The notorious lack of forensic examiners in Brazil requires the use of specialized tools and techniques to accelerate forensic examinations in mobile phones, at the risk of the impossibility to meet the current and future demand.

This paper presents the experiences and tools developed and used in the Computer Forensics Laboratory of the Criminalistics Institute of Curitiba-PR, to face the current reality of cell phone forensics. In particular, a series of XSLT (eXtensible Stylesheet Language for Transformation) filters developed by this paper's author are discussed, which transform the data captured by various forensic tools directly into standard and customized reports, using the LibreOffice suite. These reports are then used as annexes to the resulting forensic examination reports written by the experts.

Section II presents a review of existing equipment and tools for cell phone forensics. Section III presents a brief review of the XML (eXtensible Markup Language) technologies used in the creation of XSLT filters, which are presented, exemplified and evaluated in Section IV. Conclusions and future work are presented in section V.

2. Cell Phone Forensics

Cell phones contain a wealth of information that can be useful for criminal investigations, such as records of made and received calls, phonebook entries, sent and received text messages, audio recordings, photos, videos, web browsing history, GPS records, among others. Unfortunately, the police or judicial authorities do not usually ask specific questions, regularly requiring all content to be transcribed, which increases the amount of work required to prepare the forensic reports.

The major difficulty is the lack of standardization. There are dozens of manufacturers and thousands of models of mobile phones. Each device has its own characteristics, such as the operating system used (usually proprietary of the respective manufacturers), type of communication connector, storage capacity, among others. This diversity makes it necessary to use specific tools for cell phones forensics.

The most basic form of data extraction is the manual transcription, through the manipulation of the device. Under a forensic point of view, this approach has problems because it can not guarantee the preservation of the evidence found on the devices [1]. Moreover, this procedure is extremely unproductive and error prone. However, it is a "last resort" option always available to the expert, if the device is not supported by any other tool.

Most manufacturers of mobile devices usually provide a set of software and communication cables to extract data from them. However, these tools are designed for ordinary users, not for the forensic needs. They require the installation of several different device drivers, sometimes mutually incompatible. They usually do not return all the information present in the devices. If not used carefully, they can easily modify the contents of the device being examined, through the automatic synchronization between phone and computer. Still, these tools save time in examinations, compared with the manual transcription.

The next step on the convenience scale for the expert is the software-based cell phone forensic tools. There are free tools, but support for different models of devices is very limited. Some popular commercial software-based cell phone forensic tools are "*MOBILedit! Forensic*" from Compelson Laboratories, and Paraben's "*Device Seizure*". Both require the installation of device drivers from the cell phone manufacturers, but have the option of providing a package containing several communication cables in addition to the forensic software. They allow the acquisition of information stored in SIM cards used in GSM handsets, and as a great advantage, they ensure the read-only access to the equipment being examined, which prevents unintended changes in them. In addition, they support a wide range of manufacturers and models, generating standardized reports within the tool framework. However, these reports provide only minimal customization, which often makes them unsuitable as annexes to the forensic reports written by the experts.

Finally, there are the dedicated equipments to cell phone and SIM card forensics. These are the most complete solutions because they integrate both the hardware and software for data extraction, including the necessary communication cables with the mobile devices without the need to install device drivers from the manufacturers. The best known are the "*Cellebrite UFED*" (Universal Forensics Extraction Device) and "*Micro Systemation XRY*". The UFED, for example, avoids the need to use a computer, being able to capture data from mobile phones directly into a USB flash drive, besides having its own battery which makes it portable for field operations [2]. Being forensic equipments, they ensure the integrity of the evidence present in cell phones and SIM cards, avoiding any change on the examined devices. In addition, some versions are able to extract and analyze physical memory dumps of the devices in order to recover deleted data from them.

However, even the dedicated forensic equipments do not fully meet the forensic needs. Some phone models are supported only partially, e.g. only the phonebook is extracted. For other models, simply there is no support. Despite having report generators, customization is usually restricted to the headers and footers, with the body of the report not being customizable at all. In the case of UFED, Portuguese reports have incorrect terms, as "*Tagged Numbers*" to indicate outgoing calls; some tables are poorly structured, spending a lot of space, or mixing information, e.g. sent and received text messages are reported together. This ultimately makes these reports unsuitable as annexes to the forensic reports written by the experts.

From the critical analysis of the available solutions, some conclusions can be made: the ideal is to have as many tools as possible so that if one does not support a given cell phone model, another can be adopted; and there should be a way to transparently combine the data returned by the various tools in a clear and compact report, suitable for the use as an annex to the forensic report written by the expert.

It was this latter need that led to the development of XSLT filters for the LibreOffice suite. As most of the tools are able to export the gathered information in XML format, it was only needed a way to convert these XML files created from different tools to a single standard representation suitable as a forensic report annex. Section III presents the technologies that enable this.

3. XML Technologies

The XML (eXtensible Markup Language) specification is a set of rules for encoding data into text files, and comes from the W3C (World Wide Web Consortium). This format is very popular on the Internet and for application developers, thanks to its flexibility and broad support in both free and commercial tools [3].

The XML specification, as it is a set of rules, can not be used directly. Each specific application defines the elements and attributes of the XML files that it uses. These settings are usually described in schemas.

In the case of cell phone forensic tools, each one ends up defining a different XML file format. Although not usually having available schemas, the analysis of files generated by them allows inferring the specification adopted in each one.

The great advantage of using the XML format to the specific problem of generating customized and standardized reports from data gathered from mobile phones is the existence of the mechanism XSLT (eXtensible Stylesheet Language for Transformation). XSLT is a declarative language based on XML, and is used to convert an XML representation into another form of representation, e.g. XML, XHTML or text. XSLT was also been specified by the W3C [4].

Conveniently, the LibreOffice suite represents their documents in XML format, following the ODF (Open Document Format) specification. It also provides native support for XSLT filters. This makes the process of converting any source XML file into a formatted document to depend only on the development of a specific XSLT filter that converts the source XML to the internal XML representation of LibreOffice documents.

A. LibreOffice and ODF

The LibreOffice suite is a derivative of the OpenOffice suite from Oracle. It implements the ODF specification to store its documents, and is able to read and write Microsoft Office documents. It is free software, supported on Windows and Linux platforms.

The ODF specification is maintained by a committee of the OASIS (Organization for the Advancement of Structured Information Standards), from the initial specification created by Sun Microsystems [5]. It is an open specification, which means that any manufacturer can adopt it without copyright restrictions. The big advantage of an open specification for documents is to ensure future accessibility of them, regardless of the direction that technology and applications will take.

This format was adopted as the technical standard NBR ISO / IEC 26300:2008 from ABNT.

The ODF documents (which have file extensions ODT, ODS or ODP, among others) are represented by XML files, which are then compressed using the ZIP format. Images and other media are also incorporated into the documents, being part of the compressed file.

LibreOffice, because of its native support for XSLT filters, is suitable for importing XML files containing extracted information from cell phones, as presented in Section IV.

4. XSLT Filters for Cell Phone Forensic Reports

Given data reports extracted from mobile phones, the purpose of XSLT import filters is to convert those reports, originally in proprietary XML formats specific to each tool, to the XML representation of a LibreOffice ODF document. The resulting document should be suitable for use as an annex to forensic reports. In addition, data from different tools should generate standard documents, which allow their combination in a transparent way.

The author of this paper has developed XSLT import filters for the "*Motorola Phone Tools*" software, "*MOBILedit! Forensic*" software, and the "*Cellebrite UFED*" equipment. The latter is the filter used as an example in the following explanations, since the UFED (see Fig. 1) is the main tool used in cell phone and SIM card forensics in the Criminalistics Institute of Curitiba-PR. The adopted procedure in the examinations is as follows:

1. Using the UFED equipment the information from mobile phones and SIM cards are extracted and saved to USB flash drives. By default, the UFED saves the captured data in HTML and XML formats, creating a folder for each device or SIM card captured, identified with the final four digits of the IMEI or ICC-ID. Media files such as images, videos and audio files are stored in subfolders.
2. The data from the flash drive are copied to the expert's computer. Inside the LibreOffice Writer application, the expert uses the "*File / Open*" option, selects the file type called "*Cellebrite XML*", and opens the "*report.xml*" file present in each folder copied from the flash drive. Data captured from the cell phone or SIM card are sorted, formatted and listed in items and tables that are automatically numbered.
3. Each cell phone or SIM card examined generates a different document, when opened in LibreOffice Writer. To avoid this multitude of files, other files are selected entirely by choosing "*Edit / Select All*", copied and pasted at the end of the first file. The items and tables are automatically numbered sequentially. The resulting single file is annexed to the expert report.
4. In parallel to importing each captured XML file, the expert will prepare the main report itself, which contains the information obtained only by manual means, such as description of the equipments, component identification numbers, and references to each device's corresponding items in the annex. Answers to specific questions are presented after the description of the examined materials. A series of auto texts and a document template are used to ease the preparation of the main forensic report.
5. According to the authorities' request, media files (photos, videos, audio recordings, etc.) are stored in a CD-R or DVD-R that is attached to the report. Such files, along with their MD5 hashes and other metadata, have already been listed in the annex generated by the XSLT import filter during step 3.

Figure 1. Cellebrite UFED.

It is clear that through XSLT filters the import of data captured from mobile devices is made easy and simple. The filters add new file types to the list of known LibreOffice files, such as "*Cellebrite XML*", "*MOBILedit XML*" or "*Motorola XML*". No matter which tool was used, the generated reports have the same format, making their combination seamless.

The XSLT filters themselves are stored in files with XSL extension, and must be configured on LibreOffice to enable the new file types. For this, the "*Tools / XML Filters*" option of LibreOffice is used, through the specification of the name of the import XSL file and the description of the new file type that it implements, e.g. "*Cellebrite XML*".

The development of XSLT filters is nontrivial. Initially, the resulting XML files from cell phone forensic tools should be analyzed. Samples from acquisitions containing all possible types of information that may be returned are necessary, so that the specific XML elements and attributes can be discovered. Then, the resulting document template should be produced in LibreOffice, and the corresponding XML files should be interpreted. Finally, for each type of information contained in the input file, an XSLT template should be created to convert it to the desired LibreOffice XML format. Important operations, such as the chronological ordering of calls and text messages should also be encoded into the XSLT template. Failure cases, such as information that are not available, should also be properly treated. For that, any

"blank" fields should be tagged and adequate remarks added. As an example, the XSLT filter for the Cellebrite UFED has over 3,700 lines of code and 300 KiB in size.

The UFED allows the physical capture (memory dump) of the data present in the internal memory of some cell phone models, and through the "*UFED Physical Analyzer*" software, the result of the capture is interpreted. Data (including deleted records) such as calendar entries, call logs, text messages and files can be recovered. Although the retrieved data can be exported in XML format, some omissions have been detected; e.g. there is no indication whether the phonebook entry, call or text message was deleted or not. To overcome this flaw, a Python script (language supported by the Physical Analyzer) was developed which exports the recovered information to a file using the native XML format used by the UFED on logic acquisitions, with some additional XML entities to support deleted data. This XML file exported by the Python script is then processed by the standard XSLT filter for the UFED.

B. Examples

In Fig. 2 and Fig. 3 partial examples of the main report and the annex generated by the XSLT filter for Cellebrite UFED can be seen.

When a cell phone model is not supported by UFED, other tools can be used, and the resulting reports combined. However, there is always the possibility that no tool supports the device. In this case, there are two options: to indicate that the device is not supported by forensic tools, so their information could not be obtained; or transcribe its data manually in the annex, consigning such procedure as a note that follows the manually transcribed data.

C. Results

The adoption of the forensic equipment Cellebrite UFED and the corresponding XSLT filters allowed an enormous gain in productivity and quality of the forensic reports on cell phone examinations in the Computer Forensics Laboratory of the Criminalistics Institute of Curitiba-PR.

The gain in quality is due to the standardized reports, and the large decrease in the amount of reports that require manual transcription, which is very error prone. In addition, the report format of the data captured with the forensic tools was customized, aiming the generation of clear and compact annexes.

The gain in productivity can be observed on Fig. 4. By the end of 2010, the reports were prepared primarily through manual transcription, supplemented with data obtained from the cell phone manufacturers' software, when possible. This produced an average of 50 cell phones examined by month,

PAGE 3
REPORT Nº xxx.xxx-x

3.2. EQUIPMENT 2

The examination was held on December 27th, 2010, verifying that:

TABLE 3

Phone	Manufacturer:	Sony Ericsson
	Model:	W380i
	IMEI(1):	352704xxxxxxxxx
	Manufacture Country:	Brasil
Battery	Manufacturer:	Sony Ericsson
	Model:	BST-39
	Identification:	054257TMMEPT 08W13
SIM Card(2)	ICC-ID(3):	89550531xxxxxxxxxxxx
	IMSI(4):	72405400xxxxxxx
	Operator:	Claro BR
Memory Card	Manufacturer:	SanDisk
	Model:	M2
	Capacity:	512 MB
	Identification:	AX0805613075

(1) IMEI (International Mobile Equipment Identity): International identification number of the GSM device.
(2) SIM Card (Subscriber Identity Module Card): Subscriber card, associated with a phone number.
(3) ICC-ID (Integrated Circuit Card ID): Unique identifier printed on the SIM card, that is internationally valid.
(4) IMSI (International Mobile Subscriber Identity): Identification number of the subscriber with the operator.

a) This unit is equipped with a camera to record photos and videos;
b) The date and time recorded on the device do not conform with the current date and time of Brasilia, as it was necessary to remove the battery to charge it;
c) It was not possible to verify the phone number through the resources available to the expert, because the line has been disabled by the operator;
d) The data from the device memory was captured by the forensic equipment "*Cellebrite UFED*" and are listed under **Item 2.1** of the annex to this report;
e) The data from the SIM card was captured by the forensic equipment "*Cellebrite UFED*" and are listed under **Item 2.2** of the annex to this report;
f) Information or any other data not obtained through this examination can be supplied by the mobile operator, upon court request.

Figure 2. Partial example of a main forensic report.

PAGE 9
ANNEX TO REPORT Nº xxx.xxx-x

2. EQUIPMENT 2
2.1. Cell Phone:
2.1.1. General Capture Information:

TABLE 16

Parameter	Value
Selected Manufacturer:	Sony Ericsson
Selected Model:	SE W380
Detected Manufacturer:	Sony Ericsson
Detected Model:	AAB-1022111-BV - Sony Ericsson W380
Revision:	R9BB001 080212 1247 1201-4421_GENERIC_LE
IMEI:	352704xxxxxxxxx
IMSI:	724054xxxxxxxxx
Start of Extraction:	27/12/2010 13:58:23
End of Extraction:	27/12/2010 14:00:00
Phone Date / Time:	"99/12/31,21:01:09-12"
Connection Type:	USB Cable
UFED Version:	1.1.5.6 UFED; S/N:5534651

2.1.2. "**Phonebook**" data records present in the device's internal memory, the way it was found during the examination:

TABLE 17

Item	Registered Phone Number	Identification
1.	32580*7410852#963	32580*7410852#9
2.	xxxx4880	GABRIEL
3.	xxxxxxx4600	Jane
4.	xxxx0003	ORTOMASTER
5.	xxxx5034	Paüla
6.	xxxx3531	PICINA

2.1.3. "**Outgoing Calls**" stored in the device's internal memory, the way it was found during the examination:

TABELA 18

Item	Outgoing Call	Date	Time	Identification
1.	xxxxxxx5817	31/12/1999	21:53:00	(not available)
2.	xxxx5817	31/12/1999	21:55:00	(not available)
3.	xxxx5037	31/12/1999	23:51:00	(not available)
4.	xxxx6330	01/01/2000	02:31:00	(not available)
5.	xxxx9734	01/01/2000	02:32:00	(not available)

Figure 3. Partial example of a forensic report annex, generated using the XSLT filter for Cellebrite UFED.

while the average incoming requests in the same period were 140 devices per month. In December 2010, the use of UFED has begun, while the XSLT filter was developed for it. From January 2011, the average examination rate rose to 170 devices per month, a productivity increase of 240% compared to 2010.

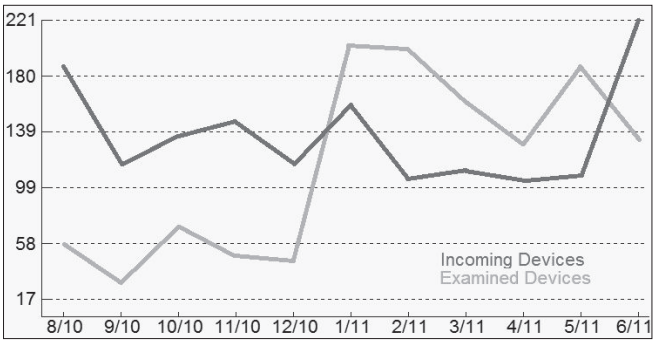

Figure 4. Statistics on cell phone examinations. Vertical axis is the number of devices, and horizontal axis is the month / year.

5. Conclusion

This paper presented the existing tools for assistance in cell phone forensics, as well as the use of these tools with XSLT filters in LibreOffice suite, aiming at standardization and customization of the reports of the various tools. To the best of the author's knowledge, this use of LibreOffice XSLT filters to import data from forensic tools was never used before.

Despite the effort required for the development of the XSLT filters, the benefits in increased productivity and quality were highlighted through the analysis of the cell phone forensic examinations performed by the Computer Forensics Laboratory of the Criminalistics Institute of Curitiba-PR.

Some useful future works can be cited: the creation of XSLT filters for other cell phone forensic tools, such as Paraben's Device Seizure and Micro Systemation XRY; and to check the possibility to deploy XSLT filters or some similar transformation for the Microsoft Office suite.

The XSLT filters already developed, the report templates and the Python script used in conjunction with the UFED Physical Analyzer (see Section IV) are freely available. Just contact the author by the e-mail *alexandre.vrubel@ ic.pr.gov.br*, requesting access to the filters, scripts and report templates. It is hoped this will contribute to improve the productivity in cell phone forensics.

Acknowledgment

The author would like to thank: the managers of the Computer Forensic Laboratory for their support to the creation and improvement of the filters and templates; all the forensic experts of the Computer Forensic Laboratory, for their suggestions and help testing and improving the filters; and Débora Ampessan, forensic expert from the Brazilian Federal Police for presenting to the author the tools and experiences of the Brazilian Federal Police regarding cell phone forensics.

References

[1] W. Jansen and R. Ayers, "Guidelines on cell phone forensics - Recommendations of the National Institute of Standards and Technology", NIST Special Publication 800-101, 2007.

[2] Cellebrite, "Universal forensic extraction device - User manual", June 2009. Available in: <http://www.cellebrite.com/forensic-products/ufed-support-center/downloads.html> Accessed in: July 21st, 2011.

[3] E. R. Harold, "Effective XML", Addison-Wesley Professional, 2003. ISBN 978-0321150400.

[4] S. Mangano, "XSLT cookbook: solutions and examples for XML and XSLT developers, 2nd edition", O'Reilly Media, 2005. ISBN 978-0596009748.

[5] OASIS, "Open document format v1.0 (second edition) specification", Committee Specification1, July 19th, 2006. Available in: <http://www.oasis-open.org/committees/download.php/19274/ OpenDocument-v1.0ed2-cs1.pdf> Accessed in: July 13th, 2011.

Alexandre Vrubel holds a Master's degree in Computer Science from the Federal University of Paraná, and is also a graduate in Informatics from the Federal University of Paraná. Currently works as a forensic expert in the Computer Forensics Laboratory of the Criminalistics Institute of Curitiba - Paraná.