Evaluating and Comparing Tools for Mobile Device Forensics Using Quantitative Analysis

Shahzad Saleem, Oliver Popov, and Oheneba Kwame Appiah-Kubi

Department of Computer and Systems Sciences
Stockholm University, Forum 100, Isafjordsgatan 39
SE- 16440 Kista, Sweden
{shahzads,popov,okak}@dsv.su.se

Abstract. In this paper we have presented quantitative analysis technique to measure and compare the quality of mobile device forensics tools while evaluating them. For examiners, it will provide a formal mathematical base and an obvious way to select the best tool, especially for a particular type of digital evidence in a specific case. This type of comparative study was absent in both NIST's evaluation process and our previous work (Evaluation of Some Tools for Extracting e-Evidence from Mobile Devices). We have evaluated UFED Physical Pro 1.1.3.8 and XRY 5.0. To compare the tools we have calculated Margin of Error and Confidence Interval (CI) based on the proportion of successful extractions from our samples in different scenarios. It is followed by hypothesis testing to further strengthen the CI results and to formally compare the accuracy of the tools with a certain level of confidence.

Keywords: Digital Forensics, Mobile Device Forensics and tools, e-Evidence, Evaluation, Confidence Interval, Hypothesis Testing and Quantitative Analysis.

1 Introduction

The digital world as we know it today is becoming increasingly mobile, mostly based on the growing computational and communication capabilities of the small scale digital devices (SSDD) and the associated services. The rate of penetration of these devices is three times faster than the one of personal computers [1] and recent statistical studies by ITU, indicate that 86.7 individuals out of 100 are using a mobile device [2].

Indeed, mobile SSDD have literally become a sort of digital behavioral archives both on individual and collective levels. They are omnipresent recordings of all our activities, even the illicit ones. Hence, during investigations these digital archives can prove crucial in providing the evidence in furthering and/or resolving a potential legal case.

Although every investigation does not end up in a court, even then it is advisable to treat the entire investigative process in a forensically sound manner. Hence, one can produce evidence which is admissible in a court of law, if such a need arises. The term forensically sound and how digital evidence must be handled is stipulated by

M. Rogers and K.C. Seigfried-Spellar (Eds.): ICDF2C 2012, LNICST 114, pp. 264–282, 2013.

many published documents (that contain principles, standards, rules and guidelines) such as IOCE's guidelines [3], RFC 3227 [4], Daubert's Principle [5], and Federal Rules of Evidence [6], [7].

Growth in the number of mobile device forensics (MoDeFo) tools is almost proportional to the volume and variety of mobile devices. These tools are rarely verified and validated by independent third parties. The evaluation results provided by the vendors are the only results available to the investigator for selecting the right tool in a particular situation.

National Institute of Standards and Technology (NIST), as an independent third party, realized the need to evaluate MoDeFo tools to facilitate the selection of a better tool for a particular scenario. Therefore, NIST has developed "Smartphone Tool Specifications [8]" and subsequently formulated "Smart Phone Tool Test Assertions and Test Plan" [9].

NIST has also evaluated some MoDeFo tools and published their results at CFTT-Mobile Devices Project's website [10]. Each tool has been evaluated individually and the results published for each tool separately [10]. Every test case is elaborated in a tabular format where one table represents the data regarding the single case. The outcomes of the evaluation process are presented as either pass or fail with some additional comments on anomalies. Neither a visualization of evaluation results nor a comparative study is conducted to help an investigator in selecting a better tool. The whole process of selection relies on use of heuristics rather than on provable formal procedures.

In the earlier published paper titled "Evaluation of Some Tools for Extracting e-Evidence from Mobile Devices" [11] the visualization of reliability assurance levels is provided for assisting the investigator to compare the tools together in order to select the better one. This paper tries to improve the selection of MoDeFo tool by using formal quantitative analysis methods.

While [11] addresses only the **reliability assurance** levels derived from NIST's specifications, the work presented in this paper deals with **accuracy** and **integrity protection** as discussed in Section 3.1. The tools we have evaluated are XRY 5.0 developed by Micro Systemation[1] and UFED Physical Pro 1.1.3.8 developed by Cellebrite[2]. Mobile phones used to evaluate these tools were Nokia 5800 Xpress Music and Sony Ericsson Xperia X1.

Mathematical foundations by using quantitative analysis are provided to compare the tools for each type of digital evidence. In particular, we have calculated the confidence interval (CI) and the margin of error (MoE) for each tool based on the proportion of successful extractions. Both CI and MoE factors when studied together should help an investigator to select a better tool for a specific investigation.

By using inferential statistics we have further strengthened our findings and made the comparison process more obvious. Based on hypothesis testing we are able to formally compare the tools in order to determine which one performs better for a specific type of digital evidence. Graphical visualization of hypothesis testing results will simplify the comparison and selection process even further.

[1] http://www.msab.com/

[2] http://www.cellebrite.com/

We have organized our paper in five sections. First section is a brief introduction of the overall research and the relevance of the work. Brief discussion concerning digital and mobile device forensics is the subject of the second section. In this section, we have also outlined the forensic process model which has been followed. Third section is about the methodology and the performance measurements. It describes CI and MoE for evaluation of the tools. Finally hypothesis testing is employed to formally compare the tools together. The analysis and discussion of the results is presented in the fourth section, while the last one (fifth) is about conclusions and the direction of future work.

2 Digital and Mobile Device Forensics

Digital Forensics (DiFo) is a relatively new and rapidly evolving discipline of the traditional Forensics Science. Its roots can be traced back to 1984 [12][13]. One of the first definitions of the term came from First Digital Forensics Research Workshop (DFRWS) [14].

DiFo is related to digital evidence or data stored, transformed and transmitted using a computer, which can help to support or refute a theory about an offense or its critical elements [15]. Advancement and evolution in the field of digital systems has spurred the progress in DiFo as well, resulting in the development of four new branches namely:

1. Computer Forensics
2. Network Forensics
3. Database Forensics
4. Mobile Device Forensics.

In this paper, the focus is on the evaluation of MoDeFo tools. MoDeFo tools deal with the digital evidences found in mobile devices. Mobile devices, as indicated in Section 1, have become important archives of the daily human behavior thus making the topic of this research both important and interesting.

2.1 Mobile Device Forensics Process Model

Various organizations, working groups and standardization bodies such as DFRWS, SWGDE, CART, NIJ, TWGDE have tried to build a standardized vocabulary, remove inconsistencies and to formalize the terminologies and the overall process as well as sub-processes [16][17]. As a result some digital forensic process models have also been developed [12], [14], [18–29].

DiFo is applied on cases with varying circumstances, heterogeneous requirements and technologies so creating a single DiFo model that fits all is a challenge in itself. Moreover, we were working in a controlled laboratory environment; with a goal to find a better tool through evaluation and comparison of various available MoDeFo tools. So, we followed a condensed form of “Forensic Investigation Process Model For Windows Mobile Devices” [29], as depicted in Figure 1.

3 Tool Evaluation

Evaluation is a process used to ensure that a tool behaves satisfactorily and it meets the performance requirements. According to Matt Bishop "*Evaluation is a process in which the evidence for assurance is gathered and analyzed against criteria for functionality and assurance*"[31]. Formally, verification and validation are the two different approaches to evaluation.

Verification requires high expertise and knowledge of the source code that is not available in our case due to the commercial nature of the tools. Therefore, in our case, validation approach is selected. According to IEEE glossary, "validation is the process of evaluating a software system or component during, or at the end of, the development cycle in order to determine whether it satisfies specified requirements" [32]. During validation, we have tested whether the tool performs as intended. In addition we also worked to find out some statistical performance measures to provide a formal basis for matching MoDeFo tools together.

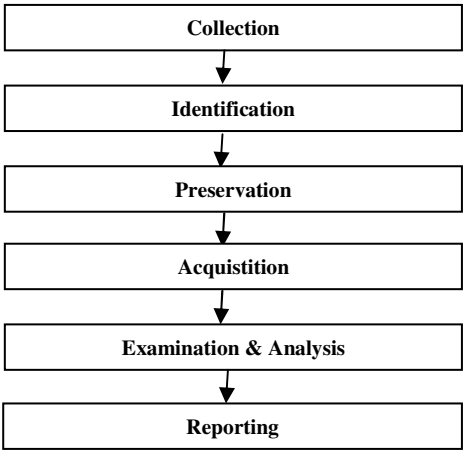

Fig. 1. The condensed form of the Windows Mobile Forensic Process Model [30]

3.1 Measuring Quality of MoDeFo Tools

The objective of the MoDeFo tools evaluation is to identify measures of their performance as criteria of quality. According to Carrier [27] DiFo tools operate by employing layers of abstraction. They transform raw bits and bytes into a presentable format which is human readable at the apex of the abstraction process. An abstraction layer transforms input data to output data by following a certain rule set, and of course, with some margin of error as depicted in Figure 2.

We validated the tools and calculated the individual and cumulative MoE induced by the underlying layers of abstraction for all types of digital evidence. The proportion of successful extractions of digital evidences was used as a base to calculate the performance indicators.

Carrier identified the following requirements, a MoDeFo tool must have: [27]

1. Usability: to address the complexity problem a tool must provide the data at a layer of abstraction that should help the investigator.
2. Comprehensive: the investigator must have access to all the data at the given layer of abstraction.
3. Accuracy: the MoE must be known to solve the "Error Problem" and to interpret the results accurately.
4. Deterministic: tool must produce the same output data when given the same input data and the rule set.
5. Verifiable: the ability to ensure accuracy of the tool by verifying its results either manually or by some independent third party tool.
6. Read Only: the ability to only read and not modify the original contents.
7. Sanity Checks: to detect any modification in the digital evidence.

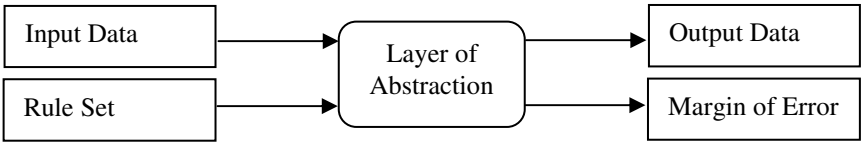

Fig. 2. Abstraction Layer Inputs and Outputs [27]

However, in our case these requirements have been condensed to **Reliability**, **Accuracy** and **Integrity Preservation**. Reliability includes the notions of usability, comprehension, determinism and verifiability. We have already measured and published [11] reliability assurance levels by following NIST smart phone tools specifications [8], smart phone test assertions and test plan [9].

The thrust of this work is to measure the accuracy and the integrity preservation capabilities of the MoDeFo tools by following B. Carrier's requirements [27]. To do so, MoE and the CI were calculated for the proportion of successful extractions by the MoDeFo tools. Hypothesis testing was then used to not only strengthen the results of CI and MoE but also to formalize and automate the comparison process. Additional tests were done to determine the ability of MoDeFo tools in preserving the integrity of digital evidence. All these results will help an investigator to choose a better tool for a specific job.

B. Carrier [27] treats integrity preservation as a recommended feature. However, in case of MoDeFo it is very hard to detach the media from the mobile system. Consequently, the extraction is performed on a live mobile system. So, the extracted copy of the potential digital evidence becomes a snap shot of a particular system in a specific time. Some portions of the original data are eventually modified during the normal operations of the mobile device. Thus, there is no "original data" to compare with the extracted copy for the verification of its integrity. Therefore, preserving the integrity of digital evidence is a must have feature for the MoDeFo tools.

In traditional forensics a trained serologist can comment on the correctness of DNA by using the explanations from molecular biology, genetics and probability

theory [14]. Nevertheless, finding similar analogy is difficult in DiFo, because digital evidence is a transformation, representation and interpretation of reality.

Moreover, digital evidence is very fragile in nature and one can possibly modify it without being detected [33]. So, to avoid any ambiguities in such circumstances the tools should not only extract the data in a forensically sound manner (as explained in Section 1) but they must also preserve its integrity to make its admissibility more plausible.

3.2 Evaluation Methodology

NIST has developed an evaluation methodology in the field of DiFo. The project is called Computer Forensics Tool Testing Project (CFTT) [34]. One of the CFTT's branch is associated with testing of MoDeFo tools [10]. NIST has developed a set of Smartphone Tool Specifications [8] and Smartphone Test Assertions and Test Plan [9] to evaluate MoDeFo tools. We have followed them [8], [9] to measure the Reliability Assurance Level and published in our paper as well [11]. We also classified and published different types of digital evidences associated with mobile devices [11].

In this paper, the same classification as presented in [11] is used to extend our previous research work. All the data, processed while calculating CI, MoE and inferential statistics, comes from our previous work as well [11]. For the sake of reproducibility, we have again explained the procedure used to populate the potential digital evidences in the sample mobile devices.

To further extend the work described in [11], we have used "Quantitative Research Methodology" to evaluate the tools for MoDeFo in terms of their accuracy for retrieving the digital evidences. As discussed by B. Carrier [27] and presented in Section 3.1, we calculated point estimate of the proportion of successful extractions by the MoDeFo tools from our samples. Then we used those proportions to calculate MoE and CI. In our research, CI is an interval within which the success proportion will lie with 95% confidence level. We have used 95% confidence level because it is the number usually used in the scientific research [35].

In the second step, "hypothesis testing" is used to formally compare the MoDeFo tools by using one tailed tests. It assisted us in choosing the tool which performs better in terms of accuracy with 95% confidence level. Hypothesis testing is a concept related to CI so this test will strengthen our CI results as well. Towards the end we have tested the ability of the two tools to preserve the integrity of digital evidence.

Confidence Interval and Margin of Error. Estimating some point estimator for the population while dealing with a sample is merely a maximum likelihood estimator for the actual parameter of the population under consideration. For instance sample mean $\overline{X}$ is a maximum likelihood estimator of the population mean μ. We know that $\overline{X}$ will not exactly be equal to μ but it will be close. Basically, finding out the point estimator is of interest along with the determination of the interval within which the actual population parameter will lie (with a certain level of confidence) [36].

In the case being considered we are finding out both the estimations of MoE and CI based on the proportion of successful extractions. These measures will be useful in deciding the level of confidence in a specific tool. The higher the point estimates for the proportion of successful extractions, the lower the margin of error, and the narrower the range between upper and lower bounds of confidence interval, the better the tool is with respect to its performance and accuracy.

The equations to calculate CI are given below:

$$CI = p \pm MoE \quad \text{.....Equation 1}$$

$$MoE = z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}} \quad \text{When n≥30……Equation 2}$$

$$MoE = t_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}} \quad \text{When n < 30…….Equation 3}$$

$$p = x/n \quad \text{.....Equation 4}$$

Whereas:

CI = Confidence Interval
MoE = Margin of Error
p = Proportion of successful extractions
x = number of objects retrieved successfully
n = total number of objects populated
$z_{\alpha/2}$ = 1.96 for 95% confidence level when n≥30 [37]
$t_{\alpha/2}$ = 2.05 for 95% confidence level when n<30 [37]

Hypothesis Testing and One Tailed Test: CI and MoE are calculated for each tool individually. Based on these performance measures, the investigator will still have to compare and eventually select a better tool manually. In order to overcome this problem, hypothesis testing as a formal comparison method is employed.

Testing a particular hypothesis concerning the unknown parameters of a population by using the sample data [38] was more interesting as compared to the explicit estimation of the unknown parameters. Hypothesis test is a “one tailed test” if the set of values lesser or greater than the critical value lies only on one side of the probability distribution, as shown in Figure 3 [39][40].

Rejection region, in case of left tailed test, lies below -1.645, hence z-score lesser than -1.645 will have enough evidence to not to accept H_o with 95% confidence level. In this case we will have 0.05 probability of Type I Error [40]. Similarly, in the case of right tailed test the critical region lies on the right hand side of the probability distribution, with all the z-scores greater than 1.645. These values were used to interpret the hypothesis testing results.

The tests were done to compare the tools together for each category of digital evidence. This type of individual comparison is useful when an investigator has many tools at his disposal. This way, he can select different tools for different categories of digital evidences during the same investigative process e.g. UFED for call logs and XRY for SMS.

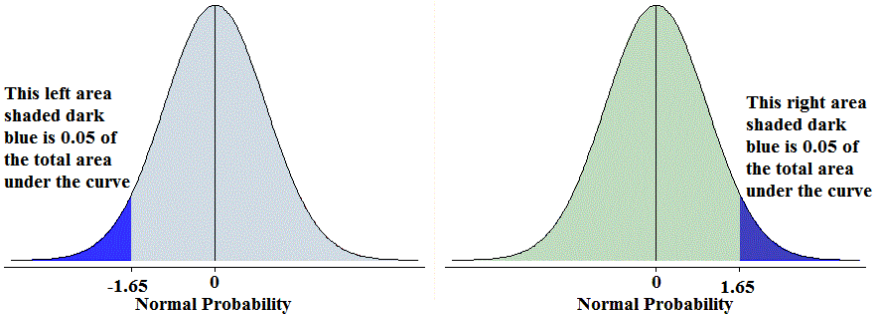

Fig. 3. One Tailed Hypothesis Testing [39]

The cumulative result for each tool is a combination of all the results in every category of digital evidence (assuming that every type of digital evidence is equally important and relevant). The hypothesis testing on the combined results comprehensively compared the tools. This type of analysis can help an investigator to select one tool for the entire investigative process based on the accuracy criterion.

Equations relevant to these statistics are the following:

$$z = \frac{p_1 - p_2}{S_{p_1 - p_2}} \quad \text{.....Equation 5}$$

$$S_{p_1 - p_2} = \sqrt{p * q\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad \text{.....Equation 6}$$

$$q = (1 - p) \quad \text{.....Equation 7}$$

$$p = \frac{(x_1 + x_2)}{(n_1 + n_2)} \text{ .....Equation 8}$$

$$p_1 = \frac{x_1}{n_1} \text{ .....Equation 9}$$

$$p_2 = \frac{x_2}{n_2} \text{ .....Equation 10}$$

Whereas:

x_1, is the total number of objects retrieved by XRY

x_2, is the total number of objects retrieved by UFED

$n_1 = n_2$, is the total number of objects populated in the mobile device

We have tested the following hypothesis:

Right Tailed Test:

$H_0 : p_1 \leq p_2$ Null hypothesis i.e. XRY does not perform better than UFED

$H_1 : p_1 > p_2$ Alternate hypothesis i.e. XRY performs better than UFED

Left Tailed Test:

$H_0 : p_1 \geq p_2$ Null hypothesis i.e. XRY performs better than UFED

$H_1 : p_1 < p_2$ Alternate hypothesis i.e. XRY does not perform better than UFED

With the theoretical aspects of the work outlined, the next step is to populate the mobile devices with potential digital evidences.

3.3 Population of Data Objects

The specifics of the data population process relative to mobile devices are well described in [30] and [11]. For the sake of reproducibility we will reiterate them in the following section.

Three different methodologies were used to populate data objects in the mobile devices.[41]

1) *Manual:* Using the normal handset interfaces only e.g. sending and receiving SMS via normal handset operations and a network of a mobile operator.
2) *Semi Manual:* Copying or moving data from a similar mobile device.
3) *Automatic:* Automated population of data objects e.g. with a tool or a software.

Since, timeline is critical in forensic science, so first of all the date and time was set in the sample mobile devices. The initial states were extracted and saved as “control states” to detect and eventually avoid possibility of any errors during the entire process.

1) *Sony Ericsson Xperia X1*

a. *PIM Entries:* A total of 631 PIM (phonebook, calendar, note and task) entries were populated. Fifteen of them were manually deleted. We not only used both the mobile devices collaboratively to populate each other but also synchronized them with MS Office Outlook 2007. Contact entries include:
   i. Special characters
   ii. Blank entries
   iii. Associated email addresses
   iv. Associated picture or image

b. *Message Entries:* Xperia X1 uses its internal memory to store all the types of messages. A total of 339 message entries were populated while 21 of them were manually deleted. We used two SIM cards by Lycatel[3] and Tele2[4] to manually populate the messages.
   i. Lycatel provides free services for both SMS and EMS. So it was used to populate SMS (comprising both ASCII and Non-ASCII characters) and EMS entries (having both smileys and emoticons).
   ii. Tele2 sim was used to populate MMS entries (containing audio, video and graphics).

c. *Call Log:* A total of 295 call log entries were populated while 14 of them were manually deleted. Moreover, we also noted that switching off and then removing the SIM card does not affect the call logs in Xperia X1.

d. *Emails:* A total of 444 email entries were populated while 399 of them were manually deleted. To populate emails, we connected Xperia X1 to our university WLAN and synchronized it with an existing email account (automated approach). We also used the mobile devices to create email entries via mobile operator's network (manual approach).

e. *Internet History:* A total of 500 internet history entries were populated while 10 of them were manually deleted. We connected our mobile device to our university WLAN for accomplishing this task.

f. *Standalone Files:* A total of 1629 standalone file entries, including audio, video and picture/graphic files, were populated while 386 of them were manually deleted (manual approach)

g. *Application Files:* A total of 448 application file entries (including word, excel, power point, one note and pdf files) were populated while 5 of them were manually deleted.

h. *GPS Entries:* GPS entries are also associated with pictures. So we used them to measure the performance of MoDeFo tools for GPS entries. We captured the pictures after enabling location services. These pictures were subsequently saved in the mobile device (manual approach) as standalone files of graphics type.

[3] `http://www.lycatel.com/`
[4] `http://www.tele2.se/`

2) *Nokia 5800 Express Music*
The approach to populate data objects in Nokia phone is the same as for SonyEricsson Xperia X1 with some minor difference in the total number of objects. The actual numbers are presented in Section 4.

3) *SIM Card:*
   a. *PIM:* A total of 246 PIM entries were populated. These entries were populated manually and also copied from the internal memory of our mobile devices.
   b. *Message:* A total of 30 message entries were populated while 10 of them were manually deleted.
   c. *Call Log:* A total of 11 call log entries were populated.

4 Results and Discussion

This section is about the results of the evaluation process. Initially, it deals with the results of CI and MoE. Then it proceeds with the formal comparison of the two tools by using hypothesis testing.

4.1 Margin of Error and Confidence Interval

The numbers, showed in four tables (1 through 4) depict:

1. Individual performance measures for each type of data objects.
2. Performance measure of MoDeFo tools for each class of data objects obtained by joining individual measures.
3. All the classes are also merged together to determine the cumulative performance measure of MoDeFo tools. It should be note that, merging the results in bullets 2 and 3 is based on an assumption that every object is equally important and relevant.

The tool with higher proportion of success, smaller MoE and thus higher confidence level is considered to be better and hence more appropriate to be used in a specific case.

Table 1 is about the evaluation results of both the MoDeFo tools for SonyEricsson Xperia X1 mobile device. Similarly, Table 2 is about the results of MoDeFo tools when applied on Nokia 5800 Xpress Music. The numbers in both the tables indicate that, in most of the cases examined, XRY is performing better than UFED.

4.2 Hypothesis Testing

The performance in terms of accuracy of the two MoDeFo tools is easily determined by comparing MoE and CI results. However, this type of comparison is not obvious, and it still has to done manually. Hypothesis testing as a formal method has a clear potential for automatic execution.

Tables 3, shows the hypothesis testing results for Xperia X1 with both the MoDeFo tools. It has a column with remarks showing whether we have sufficient evidence to reject the null hypothesis and to conclude that XRY performs better for a specific type

of data objects with 95% confidence level. Table 4, (just like Table 3) shows the hypothesis testing results for Nokia Xpress Music with both the MoDeFo tools.

Both the tables also show combined results for a specific class of data objects with an assumption that every type of data object is equally relevant and important. Similarly, all the classes are also joined together to perform hypothesis testing on all the data objects when seen together, again with same assumption as above.

Table 1. MoE and CI with Sony Ericsson Xperia X1 for both MoDeFo tools

Sony Ericsson Xperia X1												
		Total Populated	Accurately Retrieved		Success Proportion (p)		Margin of Error		Confidence Interval			
									XRY		UFED	
			XRY	UFED	XRY	UFED	XRY	UFED	Lower	Upper	Lower	Upper
PIM Entries	Phonebook/Contacts	307	292	292	0.951	0.951	0.024	0.024	0.927	0.975	0.927	0.975
	Calendar Entries	107	107	0	1.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000
	Memo/Notes	117	113	116	0.966	0.991	0.033	0.017	0.933	0.999	0.975	1.008
	Tasks/To-Do-Lists	100	95	0	0.950	0.000	0.043	0.000	0.907	0.993	0.000	0.000
	Total	**631**	**607**	**408**	**0.962**	**0.647**	**0.015**	**0.037**	**0.947**	**0.977**	**0.609**	**0.684**
Messages	SMS	185	185	179	1.000	0.968	0.000	0.026	1.000	1.000	0.942	0.993
	EMS	36	36	20	1.000	0.556	0.000	0.162	1.000	1.000	0.393	0.718
	MMS	118	30	0	0.254	0.000	0.079	0.000	0.176	0.333	0.000	0.000
	Total	**339**	**251**	**199**	**0.740**	**0.587**	**0.047**	**0.052**	**0.694**	**0.787**	**0.535**	**0.639**
Call Logs	Voice Calls	215	215	115	1.000	0.535	0.000	0.067	1.000	1.000	0.468	0.602
	Video Calls	80	80	80	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
	Total	**295**	**295**	**195**	**1.000**	**0.661**	**0.000**	**0.054**	**1.000**	**1.000**	**0.607**	**0.715**
Emails	**Total**	**444**	**438**	**0**	**0.986**	**0.000**	**0.011**	**0.000**	**0.976**	**0.997**	**0.000**	**0.000**
Internet History	URLs Visited	250	250	250	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
	Bookmarks/Favourites	250	150	0	0.600	0.000	0.061	0.000	0.539	0.661	0.000	0.000
	Total	**500**	**400**	**250**	**0.800**	**0.500**	**0.035**	**0.044**	**0.765**	**0.835**	**0.456**	**0.544**
Standalone Files	Audio	568	568	568	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
	Video	446	346	446	0.776	1.000	0.039	0.000	0.737	0.814	1.000	1.000
	Graphics/Pictures	615	515	615	0.837	1.000	0.029	0.000	0.808	0.867	1.000	1.000
	Total	**1629**	**1429**	**1629**	**0.877**	**1.000**	**0.016**	**0.000**	**0.861**	**0.893**	**1.000**	**1.000**
Application Files	Word	146	146	146	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
	Excel	42	42	42	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
	PowerPoint	130	130	130	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
	PDF	130	130	130	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000
	Total	**448**	**448**	**448**	**1.000**	**1.000**	**0.000**	**0.000**	**1.000**	**1.000**	**1.000**	**1.000**
Grand Total		**4286**	**3868**	**3129**	**0.902**	**0.730**	**0.009**	**0.013**	**0.894**	**0.911**	**0.717**	**0.743**

Table 2. MoE and CI with Nokia 5800 Xpress Music for Both MoDeFo Tools

Nokia 5800 Xpress Music												
		Total Populated	Accurately Retrieved		Success Proportion (p)		Margin of Error		Confidence Interval			
									XRY		UFED	
			XRY	UFED	XRY	UFED	XRY	UFED	Lower	Upper	Lower	Upper
PIM Entries	Phonebook/Contacts	277	272	267	0.982	0.964	0.016	0.022	0.966	0.998	0.942	0.986
	Calendar Entries	107	107	0	1.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000
	Memo/Notes	66	66	0	1.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000
	Tasks/To-Do-Lists	105	95	0	0.905	0.000	0.056	0.000	0.849	0.961	0.000	0.000
	Total	**555**	**540**	**267**	**0.973**	**0.481**	**0.013**	**0.042**	**0.959**	**0.986**	**0.440**	**0.523**
Messages	SMS	147	142	142	0.966	0.966	0.029	0.029	0.937	0.995	0.937	0.995
	EMS	37	32	32	0.865	0.865	0.110	0.110	0.755	0.975	0.755	0.975
	MMS	133	133	0	1.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000
	Total	**317**	**307**	**174**	**0.968**	**0.549**	**0.019**	**0.055**	**0.949**	**0.988**	**0.494**	**0.604**
Call Logs	Voice Calls	238	235	0	0.987	0.000	0.014	0.000	0.973	1.002	0.000	0.000
	Video Calls	45	39	0	0.867	0.000	0.099	0.000	0.767	0.966	0.000	0.000
	Total	**283**	**274**	**0**	**0.968**	**0.000**	**0.020**	**0.000**	**0.948**	**0.989**	**0.000**	**0.000**
Emails	**Total**	**389**	**389**	**0**	**1.000**	**0.000**	**0.000**	**0.000**	**1.000**	**1.000**	**0.000**	**0.000**
Internet History	URLs Visited	250	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Bookmarks/Favourites	250	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Total	**500**	**0**	**0**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Standalone Files	Audio	626	621	112	0.992	0.179	0.007	0.030	0.985	0.999	0.149	0.209
	Video	462	412	412	0.892	0.892	0.028	0.028	0.863	0.920	0.863	0.920
	Graphics/Pictures	621	521	521	0.839	0.839	0.029	0.029	0.810	0.868	0.810	0.868
	Total	**1709**	**1554**	**1045**	**0.909**	**0.611**	**0.014**	**0.023**	**0.896**	**0.923**	**0.588**	**0.635**
Application Files	Word	145	143	0	0.986	0.000	0.019	0.000	0.967	1.005	0.000	0.000
	Excel	40	36	0	0.900	0.000	0.093	0.000	0.807	0.993	0.000	0.000
	PowerPoint	130	129	0	0.992	0.000	0.015	0.000	0.977	1.007	0.000	0.000
	PDF	152	151	0	0.993	0.000	0.013	0.000	0.981	1.006	0.000	0.000
	Total	**467**	**459**	**0**	**0.983**	**0.000**	**0.012**	**0.000**	**0.971**	**0.995**	**0.000**	**0.000**
Grand Total		**4220**	**3523**	**1486**	**0.835**	**0.352**	**0.011**	**0.014**	**0.824**	**0.846**	**0.338**	**0.367**

It is evident from Table 3 that for most of the data objects, we have sufficient evidence (z-score > 1.645) to reject the null hypothesis and to conclude that XRY performs better when Xperia X1 is used as a source of digital evidences. The tools (XRY and UFED) are equally good/bad in the case of:

1. Phonebook/Contacts, equally good.
2. Video Calls, equally good with 100% success proportion for both the tools.
3. URLs visited, equally good with 100% success proportion for both the tools.
4. Audio Files, equally good with 100% success proportion for both the tools.
5. Both the tools are equally good for all the types of application files with 100% success proportion.

Table 3. Hypothesis Testing with Xperia X1 for Both MoDeFo Tools

SonyEricsson Xperia X1												
		Total Populated	Accurately Retrieved		Success Proportion		Weighted Proportion		Standard Deviation	Z-Score	Remarks	
			XRY	UFED	XRY	UFED					H_0 Rejected	
		$n = n_1 + n_2$	x_1	x_2	$p_1 = \frac{x_1}{n_1}$	$p_2 = \frac{x_2}{n_2}$	$p = \frac{(x_1 + x_2)}{(n_1 + n_2)}$	$q = (1 - p)$	$S_{p_1 - p_2} = \sqrt{p * q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$	$z = \frac{p_1 - p_2}{S_{p_1 - p_2}}$	XRY is Better	
PIM Entries	Phonebook/Contacts	307	292	292	0.951	0.951	0.951	0.049	0.017	0.000	Equally Good	
	Calendar Entries	107	107	0	1.000	0.000	0.500	0.500	0.068	14.629	Yes	
	Memo/Notes	117	113	116	0.966	0.991	0.979	0.021	0.019	-1.356	No	
	Tasks/To-Do-Lists	100	95	0	0.950	0.000	0.475	0.525	0.071	13.452	Yes	
	Total	**631**	**607**	**408**	**0.962**	**0.647**	**0.804**	**0.196**	**0.022**	**14.119**	**Yes**	
Messages	SMS	185	185	179	1.000	0.968	0.984	0.016	0.013	2.470	Yes	
	EMS	36	36	20	1.000	0.556	0.778	0.222	0.098	4.536	Yes	
	MMS	118	30	0	0.254	0.000	0.127	0.873	0.043	5.863	Yes	
	Total	**339**	**251**	**199**	**0.740**	**0.587**	**0.664**	**0.336**	**0.036**	**4.227**	**Yes**	
Call Logs	Voice Calls	215	215	115	1.000	0.535	0.767	0.233	0.041	11.415	Yes	
	Video Calls	80	80	80	1.000	1.000	1.000	0.000	0.000	NA	Equally Good	
	Total	**444**	**438**	**0**	**0.986**	**0.000**	**0.493**	**0.507**	**0.034**	**29.399**	**Yes**	
Emails	**Total**	**389**	**389**	**0**	**1.000**	**0.000**	**0.500**	**0.500**	**0.036**	**27.893**	**Yes**	
Internet History	URLs Visited	250	250	250	1.000	1.000	1.000	0.000	0.000	NA	Equally Good	
	Bookmarks/Favourites	250	150	0	0.600	0.000	0.300	0.700	0.041	14.639	Yes	
	Total	**500**	**400**	**250**	**0.800**	**0.500**	**0.650**	**0.350**	**0.030**	**9.945**	**Yes**	
Standalone Files	Audio	568	568	568	1.000	1.000	1.000	0.000	0.000	NA	Equally Good	
	Video	446	346	446	0.776	1.000	0.888	0.112	0.021	-10.613	No	
	Graphics/Pictures	615	515	615	0.837	1.000	0.919	0.081	0.016	-10.433	No	
	Total	**1629**	**1429**	**1629**	**0.877**	**1.000**	**0.939**	**0.061**	**0.008**	**-14.597**	**No**	
Application Files	Word	146	146	146	1.000	1.000	1.000	0.000	0.000	NA	Equally Good	
	Excel	42	42	42	1.000	1.000	1.000	0.000	0.000	NA	Equally Good	
	PowerPoint	130	130	130	1.000	1.000	1.000	0.000	0.000	NA	Equally Good	
	PDF	130	130	130	1.000	1.000	1.000	0.000	0.000	NA	Equally Good	
	Total	**448**	**448**	**448**	**1.000**	**1.000**	**1.000**	**0.000**	**0.000**	**NA**	**Equally Good**	
Grand Total		**4380**	**3962**	**2934**	**0.905**	**0.670**	**0.787**	**0.213**	**0.009**	**26.836**	**Yes**	

We cannot reject the null hypothesis and thus conclude that XRY performs better for just three types of data objects:

1. Memo/Notes: Here, UFED is actually performing a bit better. Nevertheless, we cannot conclude (by using left-tailed test) that UFED performs better than XRY with 95% confidence level.
2. Video: Using left-tailed test, we can conclude with 95% confidence level that UFED performs better than XRY for this type of data objects.

3. Graphics/Pictures: Using left-tailed test, we can conclude with 95% confidence level that UFED performs better than XRY for this type of data objects.
4. Standalone Files: We can conclude with 95% confidence level that UFED performs better than XRY for this class of digital evidences including audio, video and graphics files.

For the rest of the eight types of data objects, XRY performs better with 95% confidence level. We can reject the null hypothesis, with 95% confidence level, and conclude that XRY performs better than UFED for the combined performance measures of all the types of data objects in Table 3, with an assumption that every data object is equally important and relevant.

Similarly results in Table 4 provide enough evidence to reject the null hypothesis, with 95% confidence level, for most of the types of digital evidences, and to conclude that XRY performs better than UFED. If we merge the results of all the objects of a class together, with an assumption that all of them are equally important, then we can see that we still have enough evidence to reject the null hypothesis for all the classes of digital evidence (except Internet History) with 95% confidence level. Thus we can conclude that XRY performs better than UFED for all the classes of objects except "Internet History". Both the tools did not extract even a single digital evidence of the "Internet History" class. This amounts to "equally bad" performance by both tools for this data class of digital evidence.

Table 4. Hypothesis Testing with Nokia 5800 for Both MoDeFo Tools

Nokia 5800 Xpress Music											
		Total Populated	Accurately Retrieved		Success Proportion		Weighted Proportion		Standard Deviation	Z-Score	Remarks
			XRY	UFED	XRY	UFED					H_0 Rejected
		$n = n_1 + n_2$	x_1	x_2	$p_1 = \frac{x_1}{n_1}$	$p_2 = \frac{x_2}{n_2}$	$p = \frac{(x_1 + x_2)}{(n_1 + n_2)}$	$q = (1-p)$	$S_{p_1 - p_2} = \sqrt{p * q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$	$z = \frac{p_1 - p_2}{S_{p_1 - p_2}}$	XRY is Better
PIM Entries	Phonebook/Contacts	277	272	267	0.982	0.964	0.973	0.027	0.014	1.309	No
	Calendar Entries	107	107	0	1.000	0.000	0.500	0.500	0.068	14.629	Yes
	Memo/Notes	66	66	0	1.000	0.000	0.500	0.500	0.087	11.489	Yes
	Tasks/To-Do-Lists	105	95	0	0.905	0.000	0.452	0.548	0.069	13.171	Yes
	Total	**555**	**540**	**267**	**0.973**	**0.481**	**0.727**	**0.273**	**0.027**	**18.394**	**Yes**
Messages	SMS	147	142	142	0.966	0.966	0.966	0.034	0.021	0.000	Equally Good
	EMS	37	32	32	0.865	0.865	0.865	0.135	0.079	0.000	Equally Good
	MMS	133	133	0	1.000	0.000	0.500	0.500	0.061	16.310	Yes
	Total	**317**	**307**	**174**	**0.968**	**0.549**	**0.759**	**0.241**	**0.034**	**12.345**	**Yes**
Call Logs	Voice Calls	238	235	0	0.987	0.000	0.494	0.506	0.046	21.544	Yes
	Video Calls	45	39	0	0.867	0.000	0.433	0.567	0.104	8.296	Yes
	Total	**283**	**274**	**0**	**0.968**	**0.000**	**0.484**	**0.516**	**0.042**	**23.046**	**Yes**
Emails	**Total**	**389**	**389**	**0**	**1.000**	**0.000**	**0.500**	**0.500**	**0.036**	**27.893**	**Yes**
Internet History	URLs Visited	250	0	0	0.000	0.000	0.000	1.000	0.000	NA	Equally Bad
	Bookmarks/Favourites	250	0	0	0.000	0.000	0.000	1.000	0.000	NA	Equally Bad
	Total	**500**	**0**	**0**	**0.000**	**0.000**	**0.000**	**1.000**	**0.000**	**NA**	**Equally Bad**
Standalone Files	Audio	626	621	112	0.992	0.179	0.585	0.415	0.028	29.200	Yes
	Video	462	412	412	0.892	0.892	0.892	0.108	0.020	0.000	Equally Good
	Graphics/Pictures	621	521	521	0.839	0.839	0.839	0.161	0.021	0.000	Equally Good
	Total	**1709**	**1554**	**1045**	**0.909**	**0.611**	**0.760**	**0.240**	**0.015**	**20.397**	**Yes**
Application Files	Word	145	143	0	0.986	0.000	0.493	0.507	0.059	16.796	Yes
	Excel	40	36	0	0.900	0.000	0.450	0.550	0.111	8.090	Yes
	PowerPoint	130	129	0	0.992	0.000	0.496	0.504	0.062	16.001	Yes
	PDF	152	151	0	0.993	0.000	0.497	0.503	0.057	17.321	Yes
	Total	**467**	**459**	**0**	**0.983**	**0.000**	**0.491**	**0.509**	**0.033**	**30.042**	**Yes**
Grand Total		**4220**	**3523**	**1486**	**0.835**	**0.352**	**0.593**	**0.407**	**0.011**	**45.142**	**Yes**

Moreover, both the tools performed equally good/bad for:

1. SMS, equally good.
2. EMS, equally good.
3. URLs visited, equally bad (0% success proportion for both the tools)
4. Bookmarks/Favorites, equally bad (0% success proportion by both the tools)
   URLs visited and Bookmarks constitute Internet History class, so we can say that for this class of digital evidence both the tools performed equally bad.
5. Videos, equally good.
6. Graphics/Pictures, equally good.

There is just one object type of phonebook/contacts where there is not enough evidence to reject the null hypothesis. In this case, XRY is performing slightly better than UFED and its z-score is 1.309 which is slightly lesser than 1.645, thus we cannot reject the null hypothesis with 95% confidence level. However, when the confidence level is reduced to 90% then there is enough evidence to reject the null hypothesis, and conclude that XRY performs better than UFED.

With 95% confidence level, one can conclude for the rest of the twelve types of objects that XRY performs better than UFED. If we combine all the objects together, again with an assumption that every type of object is equally important and relevant, then we can say, on cumulative base, that XRY performs better than UFED with 95% confidence level.

Regarding SIM Card analysis, both tools had 100% success proportion for extracting the Contacts, SMS/EMS and Call Logs entries. This leads to just one conclusion i.e. both XRY and UFED are 100% accurate in this area. So, there was neither a need to calculate CI and MoE nor to perform hypothesis testing for SIM card analysis.

4.3 Integrity Preservation

The central ideas behind integrity preservation are (1) to preserve the data, and (2) to report on data modifications (if any). The procedure to examine integrity preservation feature in both tools is outlined below:

1. The images from the mobile devices with both the MoDeFo tools were extracted.
2. The contents of the images were modified by using WinHex 15.6. An entry in contacts was modified by changing a contact name from “Shamm” to “55ura”.
3. We reopened both image files with XRY and UFED.

XRY could not identify the modification and opened the file with the modified contact name appearing in its contact entries report window. On the other hand, UFED successfully identified the modification and reported with a “File Corrupted” error message. So in this regard, UFED came on the top.

The use of a secure platform in the form of smartcards to preserve the integrity of digital evidence is proposed as one of the plausible solutions for the above problem [33].

5 Conclusions and Future Work

Two mobile devices, Xperia X1 and Nokia 5800, are used to evaluate two MoDeFo tools i.e. XRY 5.0 and UFED Physical Pro 1.1.3.8.

5.1 Conclusion

The first step translated to computing MoE and CI in order to compare the performance of both tools. The results indicated that XRY is better than UFED for most of the object types, which we studied. But the comparison was neither obvious nor formal. The investigator has still to retain and manually compare the numbers to select a better tool with lesser margin of error, greater success proportion and better confidence level.

Finally, hypothesis testing was used to make the comparison process more obvious. The results of this process helped to conclude, with 95% confidence level, that XRY performs better than UFED for most of the object types. If we assume that all the object types are equally important and relevant than we can also reject the null hypothesis and make an overall conclusion that XRY performs better than UFED with 95% confidence level.

Comprehensive visualization of hypothesis testing results is provided in Figure 4. It shows that most of the vertical bars are above the threshold z-score value of 1.67 for the right tailed test. It provides enough evidence to reject the null hypothesis ($H_0 : p_1 \leq p_2$, XRY does not perform better than UFED) and therefore to accept the alternate hypothesis ($H_1 : p_1 > p_2$, XRY performs better than UFED) for most types of digital evidences found in the mobile devices.

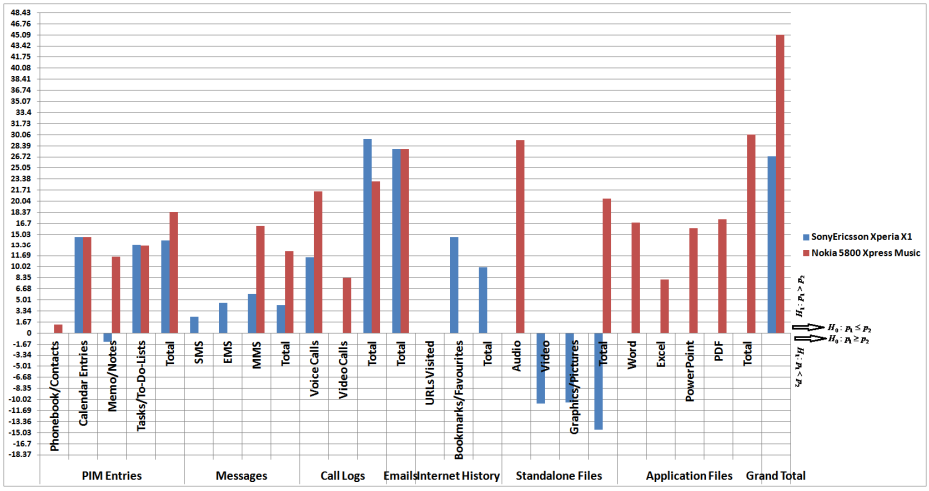

Fig. 4. Visualization of Hypothesis Testing Results

Figure 4 helps in a rapid selection of the appropriate tool for a particular type of digital evidence involving a specific type of mobile device. Another important observation in Figure 4 is that, if a tool performs better for a specific type of digital evidence for one mobile phone then it will also perform better for the same type of digital evidence with other mobile phone. However, at this stage, this rule cannot be generalized as it has an exception as well – the "Memo" type digital evidence.

UFED performs better than XRY as far as preserving the integrity of digital evidence is concerned.

In a nut shell, this paper is about a generic technique, which can be extended both vertically and horizontally. It means that any number of mobile devices and MoDeFo tools can be studied by this technique. Therefore, it will help in selecting the most appropriate MoDeFo tool for any specific incident.

5.2 Future Work

Although the results of CI, MoE and hypothesis testing can help in selecting a better tool for a particular type of digital evidence, in some way we may consider that cumulative comparison is somewhat false. This type of cumulative comparison asks for combining all the types of digital evidences with an assumption that all the types of digital evidences are equally important and relevant, which may not be true in most of the real life scenarios.

Despite the possible fallacy in the assumption the comparison results are necessary when an investigator has to choose just one tool for the entire investigative process. Especially in the circumstances when the relevance of different types of digital evidences in solving or furthering a particular case is known in advance. Hence, there must be a way to compare the MoDeFo tools by considering both performance and relevance of different types of digital evidences as two different criteria of quality. For these criteria, there must also be a mechanism to represent real life scenarios by mapping various degrees of importance and relevance.

In future, we will try to carve a generic model. The model will accommodate multiple criteria to obtain an overall ranking of the available MoDeFo tools by combining the results with varying degrees of importance and relevance. It will help in selecting the most appropriate MoDeFo tool, which may lead to the generation of better digital evidence.

References

[1] Techsling, Personal Computers Outnumbered by Mobile Phones (2010), http://www.techsling.com/2010/10/personal-computers-outnumbered-by-mobile-phones/ (accessed March 28, 2012)

[2] International Telecommunication Union (ITU), ICT Data and Statistics (IDS) (2011), http://www.itu.int/ITU-D/ict/statistics/material/excel/2011/Mobile_cellular_01-11_2.xls (accessed March 28, 2012)

[3] International Organization on Computer Evidence, IOCE - Guidelines for Best Practice in the Forensic Examination of Digital Technology (2002)
[4] Brezinski, D., Killalea, T.: RFC 3227: Guidelines for Evidence Collection and Archiving (2002)
[5] Daubert v. Merrell Dow Pharmaceuticals (92-102), 509 U.S. 579 (1993), http://www.law.cornell.edu/supct/html/92-102.ZS.html (accessed February 29, 2012)
[6] Weissenberger, G., Duane, J.J.: Federal Rules of Evidence: Rules, Legislative History, Commentary, and Authority (2004)
[7] Federal Evidence Review, Federal Rules of Evidence 2012 (2012), http://federalevidence.com/downloads/rules.of.evidence.pdf (accessed June 10, 2012)
[8] National Institute of Standards and Technology (NIST), Smart Phone Tool Specification, Version 1.1 (2010)
[9] National Institute of Standards and Technology (NIST), Smart Phone Tool Test Assertions and Test Plan, Version 1.1 (2010)
[10] National Institute of Standards and Technology (NIST), CFTT- Mobile Devices, http://www.nist.gov/itl/ssd/cs/cftt/cftt-mobile-devices.cfm (accessed June 6, 2012)
[11] Kubi, A., Saleem, S., Popov, O.: Evaluation of some tools for extracting e-evidence from mobile devices. Application of Information and Communication Technologies (10), 603–608 (2011)
[12] Baryamureeba, V., Tushabe, F.: The enhanced digital investigation process model. In: Proceedings of the 4th Annual Digital Forensic Research Workshop, pp. 1–9 (2004)
[13] Noblett, M.G., Church, F., Pollitt, M.M., Presley, L.A.: Recovering and Examining Computer Forensic Evidence. 2(4) (October 2000)
[14] Palmer, G.: A Road Map for Digital Forensic Research, Utica, New York (2001)
[15] Casey, E.: Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet, 3rd edn. Academic Press (2011)
[16] United States Computer Emergency Response Team, Computer Forensics US-CERT (2008)
[17] Meyers, M., Rogers, M.: Computer forensics: the need for standardization and certification. International Journal of Digital Evidence 3(2), 1–11 (2004)
[18] Kent, K., Chevalier, S., Grance, T., Dang, H.: Guide to integrating forensic techniques into incident response, pp. 80–86. NIST Special Publication (August 2006)
[19] Carrier, B.: An event-based digital forensic investigation framework. In: Proceedings of Digital Forensic Research Workshop (2004)
[20] Ieong, R.S.C.: FORZA – Digital forensics investigation framework that incorporate legal issues. Digital Investigation 3, 29–36 (2006)
[21] Beebe, N.L., Clark, J.G.: A hierarchical, objectives-based framework for the digital investigations process. Digital Investigation 2(2), 147–167 (2005)
[22] Agarwal, A., Gupta, M., Gupta, S., Chandra, S.: Systematic Digital Forensic Investigation Model. International Journal of Computer Science and Security 5(1), 118–131 (2011)
[23] Carrier, B.: Getting physical with the digital investigation process. International Journal of Digital Evidence 2(2), 1–20 (2003)
[24] Reith, M., Carr, C., Gunsch, G.: An Examination of Digital Forensic Models. International Journal of Digital Evidence 1(3), 1–12 (2002)
[25] National Institute of Justice, Electronic crime scene investigation: A guide for first responders (2001)

[26] Noblett, M.G., Pollitt, M.M., Presley, L.A.: Recovering and examining computer forensic evidence. Forensic Science Communications 2(4), 102–109 (2000)
[27] Carrier, B.: Defining digital forensic examination and analysis tools using abstraction layers. International Journal of Digital Evidence 1(4), 1–12 (2003)
[28] Shin, Y.-D.: New Model for Cyber Crime Investigation Procedure. Journal of Next Generation Information Technology 2(2), 1–7 (2011)
[29] Ramabhadran, A.: Forensic Investigation Process Model For Windows Mobile Devices. Tata Elxsi Security Group, pp. 1–16 (2007)
[30] Appiah-Kubi, O.K.: Evaluation of UFED Physical Pro 1.1.3.8 and XRY 5.0: Tools for Extracting e-Evidence from Mobile Devices. Stockholm University (2010)
[31] Bishop, M.: Evaluating Systems. In: Computer Security: Art and Science, p. 571. Addison-Wesley Professional (2002)
[32] Radatz, J., Geraci, A., Katki, F.: IEEE standard glossary of software engineering terminology. IEEE Standards Board, New York, Standard IEEE std (1990)
[33] Saleem, S., Popov, O.: Protecting Digital Evidence Integrity by Using Smart Cards. Digital Forensics and Cyber Crime 53, 110–119 (2011)
[34] National Institute of Standards and Technology, "Computer Forensics Tool Testing (CFTT) Project, http://www.cftt.nist.gov/ (accessed: February 26, 2012)
[35] Attia, A.: Why should researchers report the confidence interval in modern research. Middle East Fertility Society Journal 10(1), 78–81 (2005)
[36] Ross, S.M.: Interval Estimates. In: Introduction to Probability and Statistics for Engineers and Scientists, 3rd edn., pp. 240–241. Elsevier Academic Press (2004)
[37] University of Leicester, Online Statistics (2000), http://www.le.ac.uk/bl/gat/virtualfc/Stats/ttest.html (accessed: June 16, 2012)
[38] Ross, S.M.: Hypothesis Testing. In: Introduction to Probability and Statistics for Engineers and Scientists, 3rd edn., p. 291. Elsevier Academic Press (2004)
[39] UCAL Academic Technology Services, What are the differences between one-tailed and two-tailed tests? http://www.ats.ucla.edu/stat/mult_pkg/faq/general/tail_tests.htm (accessed: June 16, 2012)
[40] Easton, V.J., McColl, J.H.: Statistics Glossary V1.1 (1997), http://www.stats.gla.ac.uk/steps/glossary/index.html (accessed: June 16, 2012)
[41] Jansen, W., Delaitre, A.: Mobile forensic reference materials: A methodology and reification. US Department of Commerce, National Institute of Standards and Technology (2009)