Data Privacy Perceptions About Digital Forensic Investigations in India

Robin Verma, Jayaprakash Govindaraj, Gaurav Gupta

Chapter 2

DATA PRIVACY PERCEPTIONS ABOUT DIGITAL FORENSIC INVESTIGATIONS IN INDIA

Robin Verma, Jayaprakash Govindaraj and Gaurav Gupta

Abstract   A digital forensic investigation requires an investigator to examine the forensic images of the seized storage media and devices. The investigator obtains full access to all the data contained in the forensic images, including private and sensitive data belonging to the individual being investigated that may be entirely unrelated to the case. Unrestricted access to forensic images poses a significant threat to data privacy. No legal or technical structures are in place to prevent abuse.

This chapter presents the results of three surveys, one for each stakeholder group in digital forensic investigations, namely investigators, lawyers and the general public, that sought to capture their data privacy perceptions regarding the investigative process. The survey responses show a lack of professional ethics among some of the investigators, lack of legal support for lawyers to protect data privacy and confusion among the general public regarding their data privacy rights. The results highlight a pressing need for a privacy-preserving digital forensic investigation framework. To this end, a simple, yet efficient, solution is proposed that protects data privacy without hindering forensic investigations.

Keywords: Data privacy, digital forensic investigations, stakeholders, survey

1.   Introduction

Privacy is a very complex term, primarily because there are different definitions of privacy in different contexts. An important aspect of privacy is the ability of an individual to control access to his/her personal space [11]. An individual's personal space in the digital world comprises data in the form of files. These personal files are stored on digital devices or on local or online storage.

A digital forensic investigation attempts to collect and analyze all the digital evidence related to the case at hand. A digital forensic investigator typically gains access to the entire contents of the seized storage media in order to collect and analyze all the evidence pertaining to the case. In addition to potential evidentiary files, the seized storage media also contain private data belonging to the owner, such as personal photographs, videos, business plans, email, medical documents, financial documents, music, movies, games and software. Unrestricted access to files that are unrelated to the case, including the owner's private files, poses a significant threat to data privacy. No well-defined standards or guidelines exist to assist an investigator in deciding when all the important evidence has been gathered. This lack of clarity motivates an investigator to search for more evidence, which tends to increase the possibility and scope of data privacy violations.

Legal assistance is necessary to safeguard the data privacy of suspects and victims during investigations and the subsequent court proceedings. Lawyers should be knowledgeable about all the legal provisions that protect data privacy. Suspects and victims should also be knowledgeable about their data privacy rights.

This research sought to collect the ground truth about the principal data privacy issues related to digital forensic investigations. Three surveys were conducted, one for each stakeholder group, namely investigators, lawyers and the general public. The survey instruments were designed to capture the data privacy perceptions of the three stakeholder groups regarding digital forensic investigations. Note that the general public group corresponded to the investigated entities (suspects and victims) whose storage media would be seized in investigations. The surveys focused on the Indian context and, hence, all the participants were from India. However, the results of the study, including the concerns raised, are relevant in countries around the world.

An analysis of the literature reveals that this is the first study of the perceptions of investigators, lawyers and members of the general public regarding data privacy during digital forensic investigations. The survey responses show a lack of professional ethics on the part of some investigators, a lack of legal support for lawyers to protect data privacy and confusion on the part of the general public regarding their data privacy rights. The results highlight a pressing need for a privacy-preserving digital forensic investigation framework. To this end, a simple, yet efficient, solution is proposed that protects privacy without hindering digital forensic investigations.

2. Research Methodology

Surveys are a well-established research methodology. Researchers in digital forensics have used surveys to understand the opinions of target audiences on a variety of topics [12, 13]. The survey results have helped the researchers gain insights into particular problems and explore possible solutions.

The first step in the survey design involved personal interviews with one candidate each from the investigator and lawyer groups. Simultaneously, five potential candidates were interviewed from the general public group. The answers enabled the researchers to identify closed sets of relevant questions for the surveys of the three groups of individuals.

The second step in the survey design involved the conversion of the subjective questions and responses to objective questions with comprehensive answer options. The initial questionnaires were shown to the interviewed candidates to collect their feedback on question formulation. The feedback enabled the researchers to improve the readability, relevance and comprehensiveness of the survey questionnaires. The three surveys were then posted on the Survey Monkey website (`surveymonkey.com`).

The questionnaire used in the investigator survey incorporated three subsections that focused on:

- Adherence to digital forensic procedures.

- Suitable time to stop gathering evidence in an investigation.

- Access to the private files of the investigated entities.

The questionnaire used in the lawyer survey incorporated four subsections that focused on:

- Minimum amount of evidence required in an investigation.

- Investigation of one case leading to the prosecution of another case.

- Concerns raised by the investigated entities about data privacy.

- Misuse of personal information collected during an investigation.

The questionnaire used in the general public survey comprised two subsections that focused on:

- Attitudes regarding the privacy of data and personally-identifiable information.

- Awareness of digital forensics and the investigative process.

Table 1.   Digital forensic cases handled by investigators.

Cases Handled	Responses (Total: 15)	Response Rate
Less than 10	6	40.00%
10 to 29	2	13.33%
30 to 49	3	20.00%
50 to 69	2	13.33%
70 to 99	1	6.67%
100 or more	1	6.67%

The third and final step in the survey involved sending links to the surveys to the target audiences. The investigator and lawyer surveys were posted on Survey Monkey in August 2013. The last response in the investigator survey was received in January 2014 and the last response in the lawyer survey was received in February 2014. The public survey was posted on Survey Monkey in September 2013 and the last response was received in December 2014.

3.   Survey Participant Demographics

The investigator and the lawyer surveys included participants who were experts in their respective fields. All the participating investigators had undergone professional training and had received certifications in digital forensics. The participating lawyers were experts on Indian information technology law who had actively worked on cyber crime and computer fraud cases.

A total of fifteen digital forensic investigators responded to the survey. The investigators had experience working on criminal cases and corporate incidents. All the questions in the surveys were answered by the fifteen investigators. Eleven of the fifteen respondents were from private digital forensic laboratories or companies; the remaining four investigators worked at government forensic laboratories. Ten of the fifteen investigators had degrees in computer science; the remaining five investigators had various other academic backgrounds. Seven of the fifteen investigators had less than two years of work experience in digital forensics; four investigators had two to five years of experience; the remaining four investigators had five to ten years of experience. Table 1 shows the numbers of cases handled by the respondents during their investigative careers.

The lawyer survey respondents worked as cyber lawyers at reputed courts in India, including the Supreme Court of India. Five of the ten

Table 2. Experience levels of cyber lawyers.

Experience (Years)	Responses (Total: 10)	Responses Rate
0 to 2 years	4	40%
3 to 5 years	2	20%
6 to 8 years	2	20%
9 to 10 years	0	0%
More than 10 years	2	20%

respondents were with private firms (one of the five respondents owned a law firm); three respondents were independent legal consultants; the remaining two respondents worked at government agencies. All ten respondents answered all the questions in the survey. Table 2 presents the experience levels of the cyber lawyers who participated in the survey.

Table 3. Age distribution of general public respondents.

Age	Percentage
Up to 18	4.00%
19 to 24	61.10%
25 to 34	17.80%
35 to 44	6.80%
45 and above	10.30%

A total of 1,235 members of the general public completed the demographics section of the survey; 654 individuals quit before completing the demographics section. Male respondents constituted 66.6% of the participants and females constituted 33.4% of the participants. Table 3 shows the age distribution of the respondents.

Table 4 shows the educational qualifications of the survey participants from the general public. A total of 17.2% of the respondents had less than four years of experience using computing devices, 21.5% had four to six years of experience and 61.3% had more than six years of experience. The demographic data reveals that the survey participants were well educated and had adequate experience using computing devices.

A hypothesis was framed that the participants' levels of awareness about privacy issues related to digital documents were high. However, this hypothesis was rejected after a thorough analysis of the public survey results.

Table 4.   Educational qualifications of general public respondents.

Education	Percentage
High School Diploma	5.20%
Undergraduate Diploma	5.80%
Baccalaureate Degree	56.50%
Post-Graduate Degree	29.90%
Doctoral Degree	2.60%

4.   Investigator Perceptions of Privacy

The goal of the investigator survey was to assess how digital forensic investigators handled the personal data extracted from seized devices belonging to the subjects of investigations. Although the number of participants in the investigator survey was limited, the responses were valuable due to the expertise of the participants. The following subsections discuss the three components of the investigator questionnaire.

4.1   Following Forensic Procedures

The chain of custody is a legal document that tracks an exhibit (potential item of evidence) from the time of its seizure until it is presented in court or is returned to the owner after an investigation. The document contains information about the exhibit, along with the names of the individuals who had custody of the exhibit, their designations and periods of custody. The chain of custody is maintained to ensure accountability and fairness of the investigative process and judicial proceedings.

Two questions were framed on chain of custody procedures to assess if the investigators were well versed in the basics of their trade and took their jobs seriously. The first question asked the investigators if they filled out chain of custody forms – fourteen of the fifteen respondents responded in the affirmative; one respondent was unaware of chain of custody documentation.

The second question asked the investigators who filled out chain of custody forms about the frequency of filling out the forms – eleven of the fourteen did this every time; two did this most of the time, but not always; and one did this some of the time.

The next question asked the fourteen respondents who filled out chain of custody forms about when they filled out the forms. Two of the fourteen respondents said that they only filled out the forms for cases that were going to be tried in court. Three respondents did this only for

important cases. The remaining nine respondents filled out the forms for all types of cases.

The responses to the last two questions are inconsistent. Eleven investigators said that they created chain of custody documentation every time, but only nine of them said that they created the documentation for all types of cases.

4.2   Completing the Evidence Gathering Phase

The first question on this topic asked investigators if they stopped gathering evidence after finding potentially relevant evidence or if they explored the forensic images further, increasing the probability of encountering personal files that were not relevant to the case. Eight of the fifteen investigators said that they stopped only after they had gathered all possible – pertinent and irrelevant – evidence. Six investigators stopped after they gathered all possible evidence related to a case. The remaining investigator stopped after collecting the minimum amount of evidence needed to prove or disprove a case.

The next question asked the investigators if they experienced situations where they collected evidence that was not related to the case at hand, but that could be used to make a separate case against the subject. Surprisingly, seven of the fifteen investigators said yes, most of the time; four responded yes, only sometimes; and four did not encounter situations where they collected such evidence.

The responses to the questions indicate that gathering excess evidence is a common practice among digital forensic investigators. The habit of searching for more evidence than required is due to an investigator's indecision about gathering adequate evidence to make a case or an attempt to enhance his/her professional reputation by discovering unrelated evidence that opens a new case against the subject. Both these situations increase the likelihood of data privacy breaches.

4.3   Accessing Private Files

The first question in this part of the survey asked investigators about their reactions after they encountered private files (e.g., personal photographs, videos, songs, business plans or intellectual property) during an investigation. Six of the fifteen investigators said that they viewed private files and copied the files related to the case being investigated as well as files that were not linked to the case, but appeared to be illegal or questionable. Four other investigators said that they viewed and copied private files because these files were more likely to contain evidence relevant to the case at hand as well as to other possible cases.

The remaining five investigators said that they viewed private files, but only copied the files that were relevant to the case at hand. The results reveal that all the surveyed investigators routinely accessed private files that may or may not be associated with the case at hand. Surprisingly, ten of the fifteen investigators would not hesitate to copy private files whether or not they found irregularities related to the case.

Another question asked the investigators if they had seen other investigators copy files such as wallpaper, songs, movies, games or commercial software from case images. Three of the fifteen respondents stated that they had seen their colleagues at their laboratories do such things. Four investigators said that they had seen investigators at other laboratories copy non-malicious personal files belonging to investigated entities. One investigator had not seen anyone copy such files, but she did not see any problem with such copying. The remaining investigators had not seen such copying and they felt that it was inappropriate.

Surprisingly, half the participants had seen investigators in their laboratories or elsewhere copy non-malicious content from forensic images. Such unprofessional behavior poses a serious threat to data privacy. If an investigator is willing to copy wallpaper, songs, movies, games and application software from media belonging to the subject of an investigation, then the security of the subject's private files, including personal photographs, audio and video files and countless other confidential documents, cannot be guaranteed.

The final question asked the investigators if they had heard of or been involved in a situation where a subject reported the misuse of information or evidentiary items being used to threaten him/her. Interestingly, only one of the fifteen investigators was aware of such a situation. Nine said that they had not heard of any misuse during their careers. The remaining five investigators doubted that such abuse could ever occur.

5.   Cyber Lawyer Perceptions of Privacy

The goal of the lawyer survey was to obtain insights into the legal aspects of privacy during digital forensic investigations and court proceedings. A pilot interview was conducted with a lawyer who had argued cases before the Supreme Court of India; the interview helped frame a comprehensive questionnaire for the survey. Although the number of participants in the lawyer survey were limited, the responses are valuable due to the expertise and prominence of the participants. The following subsections discuss the four components of the cyber lawyer questionnaire.

5.1   Completing a Case

The first question asked the respondents when a case of cyber crime or computer fraud was ready for trial. Seven of the ten respondents felt that a case was ready after all the possible evidence – relevant as well as irrelevant to the case – was collected and analyzed; some of this evidence could also be used in a fresh case. Two respondents felt they could stop after gathering all the evidence related to a case. The remaining respondent said that he would stop after collecting the minimum amount of potential evidence.

The next question asked about the minimum amount of evidence sufficient to prove or disprove a case. Four of the ten respondents said that one or two pieces of evidence would be sufficient. Three participants said three to five pieces of evidence would be enough while the remaining three respondents felt that six to ten pieces of evidence would be required.

The responses to this question are significant because they set an upper limit on the amount of evidence required in a typical case. It is interesting that digital devices containing hundreds or thousands of files are typically seized during an investigation; however, all the respondents felt that no more than ten pieces of evidence would be adequate. The remaining files on the seized digital devices would be irrelevant to the case and may well contain personal or private data belonging to the subject of the investigation.

5.2   Using Evidence in Other Cases

The first question in this subsection was designed to verify the results of the investigator survey, where the participants who collected evidence about activities not related to the case at hand used the evidence to open new cases. Asking the same question to cyber lawyers made sense because evidence collected by investigators is compiled and used by cyber lawyers in legal proceedings. One of the ten lawyer respondents always encountered situations where some of the evidence was used to start fresh cases. Five respondents experienced such situations most of the time while the remaining three respondents experienced these situations some of the time. One respondent had never encountered such a situation.

5.3   Protecting Data Privacy

The questions in this subsection focused on three privacy-related laws in the Constitution of India and the Information Technology Act of 2000 and its 2008 amendment [14]. The first question asked the lawyers

about the numbers of cases they handled in which the investigated entity requested the right to privacy by referring to the freedom of speech and expression provided by Article 19(1)(a) or the right to life and personal liberty provided by Article 21 of the Constitution of India, or both. Five of the ten lawyers handled less than ten such cases and three encountered ten to 29 such cases. The remaining two respondents observed 30 to 49 and 50 or more such cases.

The second question asked about instances of investigated entities complaining about data privacy breaches under Section 72A of the (Indian) Information Technology Act of 2000. This section provides protection against the access and disclosure of private information belonging to an investigated entity that is irrelevant to the case at hand. An example is the access and/or disclosure of personal or family photographs and videos, when the owner of the material is being investigated for financial fraud. Six of the ten lawyers answered in the affirmative, with two to five instances of such cases. The remaining four lawyers answered in the negative.

The third question asked about instances of investigated entities complaining about data privacy breaches under Section 43A of the (Indian) Information Technology Act of 2000. This section provides protection against the improper or negligent handling of an individual's sensitive personal information during an investigation. Six of the ten lawyers answered in the affirmative, with one to five instances of such cases. The remaining four lawyers answered in the negative.

A subsequent question asked the lawyers about the numbers of cases they had worked on or had knowledge about, where an investigated entity requested the court to protect the private data or files residing on his/her seized digital devices. Three of the ten lawyers encountered up to ten such cases while two others encountered ten to 20 such cases. Interestingly, one respondent had knowledge of more than 90 cases. The remaining four respondents had never encountered such a case.

5.4   Misusing Personal Information

The last question asked the participants if they had heard of incidents where an investigated individual reported the misuse of personal information (especially, evidence being used after the completion of the investigation) as a threat or for purposes of intimidation. Two of the ten respondents knew about such cases; one respondent reported two cases of evidence mishandling while the other reported one case. Three of the ten respondents had never encountered such a case. Two respondents

opted not to answer the question. The remaining three respondents were skeptical if such an abuse of evidence could ever occur.

6. General Public Perceptions of Privacy

After acquiring images of the computing devices involved in a case, a digital forensic investigator has full access to the contents of the images. The owner of the devices has no way of ensuring that the investigator does not access private data unrelated to the case at hand. For example, if a person is suspected of financial fraud, then his family holiday photographs and videos – which are not related to the case – should not be accessed during the investigation.

Half of the investigators reported seeing fellow investigators copy private data belonging to investigated individuals that were completely unrelated to the cases at hand. Two respondents in the lawyer survey knew of instances where an investigator used the data gathered during a case to threaten the investigated individual. These reports raised serious privacy concerns and prompted the researchers to survey the general public to obtain insights into their sensitivity about data privacy. A hypothetical question was framed that asked the surveyed individuals if the seizure of their digital devices would affect their perceptions about data privacy. The following subsections discuss the two components of the general public questionnaire.

6.1 Attitudes Towards Privacy

The questions in this subsection focused on how people handled their private data. Specifically, the types of files that people considered to be private and where these files were stored. The protection of personally-identifiable information is another dimension of privacy in the digital world. Thus, the survey instrument created for the general public incorporated some questions related to personally-identifiable information.

Storage of Personal Information. The first question in this subsection asked the survey participants about the frequency with which they stored private data on digital devices. Table 5 summarizes the responses. Note that the percentage values in the table were obtained by summing the values corresponding to three responses: (i) sometimes; (ii) usually; and (iii) always.

Since considerable amounts of private data are stored on digital devices, the loss of a device could pose a serious privacy threat to its owner. Therefore, the next question asked the participants if they had lost any of their digital devices during the past five years. Table 6 summarizes

Table 5.   Devices used to store private data.

Devices	Users
Mobile Phones	70.3%
Laptops	75.1%
Desktops	54.9%
Portable Hard Drives	45.4%
USB Drives	58.1%

Table 6.   Devices lost during the past five years.

Devices	Users
Mobile Phones	33.0%
Tablets	0.7%
Laptops	3.1%
Portable Hard Drives	3.3%
USB Drives	39.9%
None	41.5%

the responses. The table shows that 59.5% of the respondents lost at least one digital device during the past five years. This high number implies that the device owners were not cautious about their devices or their devices were stolen at some point in time. Valuable items such as smartphones and laptops are attractive targets for thieves, but the loss of low-cost devices such as USB drives may be the result of careless behavior on the part of their owners. Indeed, 39.9% of the survey participants reported that they had lost at least one USB drive during the past five years.

Common Passwords for Different Accounts. The survey revealed that 32.6% of the participants used common passwords for multiple accounts, whereas 45% of the participants used different passwords. The remaining 22.4% opted not to reveal any information about their passwords. The results show the casual behavior of people with regard to password security and data security as a whole.

Storage of Passwords on Devices. The survey revealed that 24.6% of the participants stored their passwords on smartphones or tablets and 25.6% stored their passwords on laptops or desktops. Although the majority of the participants (63.9%) did not store passwords on their devices, one in three did, in fact, store their passwords on their devices. Thus, one can assume that one in three devices seized in investigations

Table 7.   Personal files and documents stored on digital devices.

File/Document	Personal Computer	USB/Portable Hard Drive	Tablet	Smartphone
Photographs	81.50%	33.90%	6.70%	30.30%
Video Files	69.10%	23.10%	3.80%	20.50%
Audio Files	62.90%	20.70%	3.70%	22.00%
Bank Statements	43.60%	7.10%	1.50%	4.80%
Travel Bookings	50.40%	9.80%	3.10%	12.80%
Transcripts/Admit Card	67.30%	15.00%	3.10%	7.20%
Resume	71.90%	20.40%	4.10%	10.60%
Medical Reports	36.30%	6.30%	1.80%	3.10%
Job Offers	58.30%	10.80%	2.30%	5.90%
Passport	49.20%	10.70%	2.40%	5.00%
PAN Card	52.50%	10.20%	2.40%	4.90%
Aadhar Card	44.10%	8.60%	2.00%	4.40%
License	42.90%	8.10%	1.90%	4.80%
Voter ID	46.10%	8.40%	1.80%	4.40%
Birth Certificate	45.30%	8.20%	1.70%	3.10%
Credit/Debit Card Data	32.60%	5.20%	1.20%	3.70%

would contain stored passwords and that one in three investigated individuals would have common passwords for all their accounts.

Storage of Personal Files on Devices.   This question was framed to make the survey participants aware of the private files stored on the digital devices they own or use. This would enable them to appreciate the risk they would incur if their devices were to be seized in digital investigations. The question asked the survey participants to specify the devices on which they stored certain types of private files. The participants were required to answer this question for all the listed private files. The responses provided a relative ranking of the devices on which individuals prefer to store various types of private files.

A total of 1,474 individuals answered this question. For reasons of space, only the notable findings are discussed. The survey revealed that 81.50% of the respondents stored their personal photographs on their laptops or desktops, and 30% to 35% stored personal photographs on USB drives, portable hard drives, online accounts and smartphones. Digital photographs were the most ubiquitous type of personal files stored across digital devices and online storage services.

Table 7 shows the percentages of survey participants who stored various personal files and documents on digital devices. Note that the PAN card is issued by the Income Tax Department of India, the Aadhar card

Table 8.   Rankings of personal files/documents and PII stored on digital devices.

File/Document	Percentage	PII	Percentage
Credit/Debit Card Data	76.90%	Phone Number	74.60%
PAN Card	73.10%	PAN Card Data	72.30%
Transcripts/Admit Card	72.00%	Email Address	72.20%
Voter ID	71.40%	Full Name	70.39%
Passport	68.60%	Bank Data	69.90%
License	68.30%	Biometrics	69.10%
Aadhar Card	65.60%	Date of Birth	68.60%
Photographs	63.00%	Home Address	68.50%
Job Offers	62.80%	Passwords	68.30%
Birth Certificate	62.70%	Father's Name	67.66%
Resume	61.90%	Passport Data	65.60%
Bank Statements	61.20%	Aadhar Card Data	64.00%
		License Details	63.60%
		ATM PIN	62.20%
		Mother's Maiden Name	61.53%

is a biometric identity card issued by the Government of India and the Voter ID is issued by the Election Commission of India. For every type of private file or document specified in the question, the largest percentage of survey participants stored the file or document on their laptops or desktops. This finding supports the hypothesis that an individual's laptop and desktop tend to contain large amounts of personal data and that the individual's privacy is at risk when these devices are seized in digital forensic investigations.

Ranking Personal Files/Documents and PII.   Two questions were framed to obtain the relative rankings of personal files/documents and personally-identifiable information (PII). The participants were asked to rank the entries on a scale of 1 to 5, where 1 corresponded to least important and 5 corresponded to most important. The motive was to have each participant assign relative priorities to personal files/documents and personally-identifiable information before the participant was introduced to the processes involved in evidence seizure and digital investigations. After obtaining the preliminary rankings for the first question, the second question asked the participants if they would change their rankings if their devices were seized for investigative purposes.

A total of 1,474 individuals responded to the first question. Upon counting only rating values of 4 and 5, a total of 63% of the respondents rated personal photographs as being important. Table 8 provides the corresponding percentages for other types of files/documents.

Table 9. Ratings after digital devices were hypothetically seized.

Private Data	No Effect	Increase	Decrease
Personal Files/Documents (1,304 responses)	47.3%	43.8%	8.8%
Personally-Identifiable Information (1,304 responses)	46.7%	46.6%	6.8%

A total of 1,287 individuals rated various personally-identifiable information on a scale of 1 to 5, where 1 corresponded to least important and 5 corresponded to most important. The results reveal that 70.39% of the survey participants felt that their full names were important, and 67.66% rated their father's name and 61.53% rated their mother's maiden name as important personally-identifiable information. Table 8 provides the corresponding percentages for other important types of personally-identifiable information.

6.2 Awareness of Investigations

The questions in this subsection were designed to understand how an individual's ratings of personal information would change if his/her digital devices were to be seized for investigative purposes. A drastic shift was anticipated in the privacy ratings in such a situation and the shift was expected to be inversely proportional to the trust that an individual had in investigative agencies. The change in attitude was also expected to depend on the individual's awareness of the digital forensic investigation process and the fact that most digital forensic tools can locate and extract hidden and deleted data.

Trust in Investigative Agencies. The survey participants tended to believe that law enforcement and other investigative entities would not misuse their personal data if their devices were to be seized for investigative purposes. Table 9 shows that 56.1% (= 47.3% + 8.8%) and 53.5% (= 46.7% + 6.8%) of the participants said that there would either be no effect on their previous privacy ratings for personal files/documents and personally-identifiable information, respectively, or their privacy ratings would decrease. It is especially interesting that the participants, who said that their privacy ratings would decrease, were actually less concerned about the privacy of their data after a hypothetical device seizure.

Awareness of Digital Forensics. Another question asked the survey participants if investigative agencies had tools to recover hidden or deleted data. The survey results indicate that 32.21% of the participants were not sure if this was possible and 20.25% believed that such data could not be recovered. Only 47.4% of the participants were aware that hidden or deleted data could be recovered. Despite the fact that nearly half of the survey participants knew that hidden or deleted data was recoverable, when answering the next question, 40.95% of the participants said that they temporarily stored their personal information on their office devices.

7.   Proposed Data Privacy Solution

The survey results indicate that the privacy of an investigated individual is at risk during a digital forensic investigation and that there is an urgent need to incorporate data privacy measures into the investigative process. A data privacy solution should protect the investigated individual while ensuring that neither the completeness of the investigation nor the integrity of the digital evidence are compromised. It is also highly desirable that the solution enhance investigator efficiency and save time and effort. Dehghantanha and Franke [4] have highlighted the importance of privacy-respecting digital investigations. The next section briefly discusses the research literature that addresses privacy in the context of digital forensic investigations. Following this discussion, the proposed data privacy solution is presented.

7.1   Privacy and Investigations

Aminnezhad et al. [1] have noted that digital forensic investigators find it difficult to strike the right balance between protecting the privacy of investigated individuals and performing complete investigations. They also observe that the general lack of awareness about data privacy on the part of digital forensic investigators could result in unintentional abuses.

Several researchers have attempted to use cryptographic mechanisms to protect data privacy during digital forensic investigations. Law et al. [10] have proposed a technique that encrypts data in an email server and simultaneously indexes case-related keywords. The investigator provides keywords to the server owner who has the encryption keys and uses them to decrypt emails containing the keywords, following which the emails are sent to the investigator.

Hou et al. [6, 9] have proposed mechanisms for protecting data residing at service provider storage centers using homomorphic, commutative encryption. The mechanisms also ensure that the service provider does

not have any knowledge of the queries issued by an investigator. Hou et al. [7, 8] also present a similar solution for a remote server.

Shebaro and Crandall [15] have used identity-based encryption to conduct a network traffic data investigation in a privacy-preserving setting. Gou et al. [5] have specified generic privacy policies for network forensic investigations. Croft and Olivier [3] have proposed a technique that compartmentalizes data into layers of sensitivity, where less private data is in the lower layers and more private data is in the higher layers. Investigator access to private information is controlled by initially restricting access only to the lower layers. The investigator is required to demonstrate his knowledge and behavior in the lower layers to obtain access to information in the higher layers.

Van Staden [16] has proposed a privacy-enhancing technology framework for protecting the privacy of third parties during digital forensic investigations. The framework requires an investigator to write focused queries when searching for potential evidence. The framework evaluates whether or not the query results cause a privacy breach. If a breach is deemed to occur, then the investigator is asked to submit a more focused query. If an investigator overrides the query results and attempts to access private data, the framework logs the investigator's actions in a secure location.

7.2   Privacy Solution

The proposed data privacy solution does not interfere with the outcomes of a digital forensic investigation. The solution, which is presented in Figure 1, brings more transparency to the investigative process and increases investigator accountability.

The solution focuses on the analysis phase of a digital forensic investigation, during which an investigator analyzes images of the storage media in the seized digital devices. In addition to the images, the solution methodology takes two additional inputs, namely the learned knowledge of similar cases from a case profile database and the details of the case at hand. The case profile database is a collection of case-specific features that may be used to predict potential pieces of evidence for a particular case. The database contains a feature list based on the contents and metadata of evidence files and investigator reviews obtained from historical cases studies. The feature list selection for the database also requires taxonomic information about private data and files that exist on computer systems.

All the inputs are processed by a privacy-preserving forensic tool that identifies the pieces of evidence relevant to the case at hand. The forensic
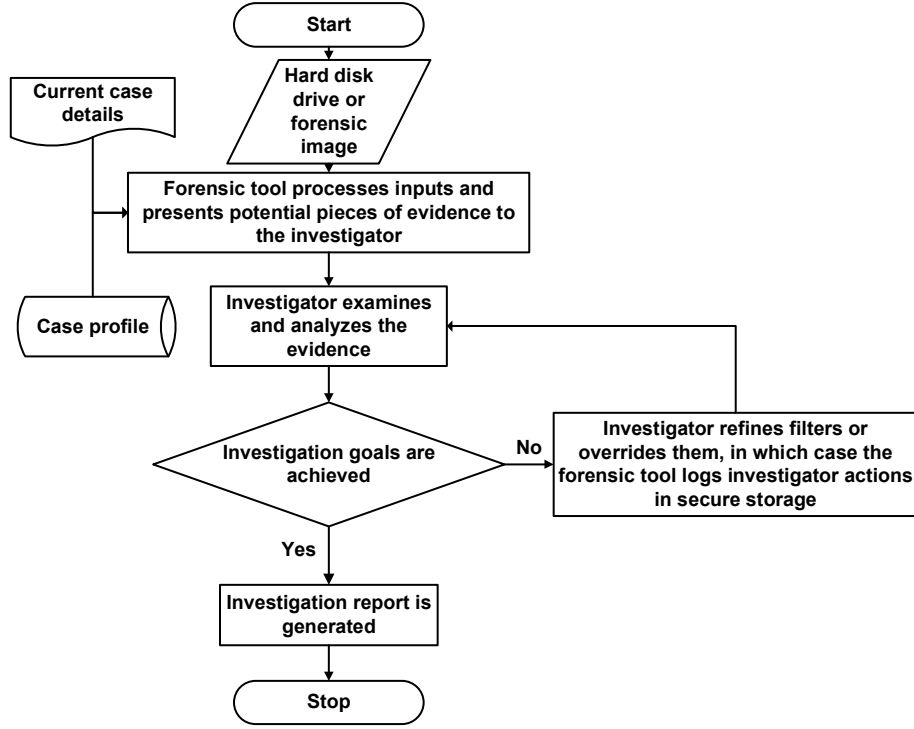
Figure 1. Privacy solution.

tool needs to ensure the completeness of an investigation. The tool may generate false positives, but it should never report a false negative. The system proposed by van Staden [16] requires an investigator to submit focused queries to obtain potential evidence from an image. If the tool determines that the investigator's query results violate privacy, then the investigator has two options. The first option is to submit a fresh query that does not violate privacy. The second option is to override the privacy-filtering functionality and conduct the investigation in a conventional manner; in this case, the tool logs all the investigator's actions in secure storage to prevent tampering. As such, the tool adds an extra layer of search without any gain in knowledge or efficiency.

The proposed privacy-preserving forensic tool simplifies the investigator's task by providing advice regarding potential evidence for the case at hand. If the investigator finds the results to be insufficient, then the investigator could mark the existing evidence and fine-tune the tool predictions by adding more information. Or, the investigator could override the prediction results and continue the investigation in a conventional manner; all the investigator's actions and their timestamps would then

be logged in a secure manner. The authenticity of the logged actions are strengthened by obtaining the modification access change date and time-stamps corresponding to the actions directly from the operating system kernel [2, 17]. The logs are vital to resolving complaints about potential privacy breaches. After the investigator collects sufficient evidence from the output list, the investigative process is complete and the case report is generated.

The proposed privacy-preserving solution would not infringe on the powers of the investigator; it simply brings more accountability and transparency to the investigative process. The investigator would have a clear idea about the responsibilities with regard to data privacy and the performance of the investigator would not be compromised.

8.   Conclusions

The surveys of investigators, lawyers and the general public provide valuable insights into their data privacy perceptions with regard to digital forensic investigations. The survey results reveal a lack of professional ethics on the part of some investigators, a lack of legal support for lawyers with regard to data privacy protection and confusion among the general public regarding their data privacy rights. While the numbers of participants in the investigator and lawyer surveys were limited, the survey responses were valuable due to the levels of expertise and experience of the participants. A total of 654 out of 1,889 (34.6%) participants in the general public survey did not complete the survey; their principal complaint was that the survey was very comprehensive and time-consuming. Nevertheless, the results and the concerns raised are relevant in India as well as in countries around the world.

The survey results demonstrate that there is an urgent need for a privacy-preserving digital forensic investigation framework that protects data privacy without compromising digital forensic investigations. The simple, yet efficient, privacy-protecting solution proposed in this work ensures the privacy of the subjects of investigations without compromising the completeness and efficiency of investigations. The solution also does not infringe on the powers of investigators; it simply brings more accountability and transparency to the investigative process.

Future research will focus on implementing the privacy-preserving digital forensic tool and evaluating its performance in real-world investigations. A key issue when using the tool is to choose the correct filters and parameters so that the probability of finding all possible evidence (before the filters are overridden) is maximized. Future research will also focus on methods for selecting appropriate filters and parameters.

References

[1] A. Aminnezhad, A. Dehghantanha and M. Abdullah, A survey of privacy issues in digital forensics, *International Journal of Cyber-Security and Digital Forensics*, vol. 1(4), pp. 311–323, 2012.

[2] M. Barik, G. Gupta, S. Sinha, A. Mishra and C. Mazumdar, An efficient technique for enhancing forensic capabilities of the Ext2 filesystem, *Digital Investigation*, vol. 4(S), pp. S55–S61, 2007.

[3] N. Croft and M. Olivier, Sequenced release of privacy-accurate information in a forensic investigation, *Digital Investigation*, vol. 7(1-2), pp. 95–101, 2010.

[4] A. Dehghantanha and K. Franke, Privacy-respecting digital investigation, *Proceedings of the Twelfth Annual International Conference on Privacy, Security and Trust*, pp. 129–138, 2014.

[5] H. Guo, B. Jin and D. Huang, Research and review in computer forensics, in *Forensics in Telecommunications, Information and Multimedia*, X. Lai, D. Gu, B. Jin, Y. Wang and H. Li (Eds.), Springer, Berlin Heidelberg, Germany, pp. 224–233, 2011.

[6] S. Hou, R. Sasaki, T. Uehara and S. Yiu, Verifying data authenticity and integrity in server-aided confidential forensic investigations, *Proceedings of the International Conference on Information and Communication Technology*, pp. 312–317, 2013.

[7] S. Hou, T. Uehara, S. Yiu, L. Hui and K. Chow, Privacy preserving confidential forensic investigation for shared or remote servers, *Proceedings of the Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 378–383, 2011.

[8] S. Hou, T. Uehara, S. Yiu, L. Hui and K. Chow, Privacy preserving multiple keyword search for confidential investigations of remote forensics, *Proceedings of the Third International Conference on Multimedia Information Networking and Security*, pp. 595–599, 2011.

[9] S. Hou, S. Yiu, T. Uehara and R. Sasaki, Application of secret sharing techniques in confidential forensic investigations, *Proceedings of the Second International Conference on Cyber Security, Cyber Peacefare and Digital Forensics*, pp. 69–76, 2013.

[10] F. Law, P. Chan, S. Yiu, K. Chow, M. Kwan, H. Tse and P. Lai, Protecting digital data privacy in computer forensic examinations, *Proceedings of the Sixth International Workshop on Systematic Approaches to Digital Forensic Engineering*, 2011.

[11] A. Moore, Defining privacy, *Journal of Social Philosophy*, vol. 39(3), pp. 411–428, 2008.

[12] K. Ruan, J. Carthy and T. Kechadi, Survey of cloud forensics and critical criteria for cloud forensic capability: A preliminary analysis, *Proceedings of the Sixth Annual Conference on Digital Forensics, Security and Law*, pp. 55–69, 2011.

[13] K. Ruan, J. Carthy, T. Kechadi and I. Baggili, Cloud forensics definitions and critical criteria for cloud forensic capability: An overview of survey results, *Digital Investigation*, vol. 10(1), pp. 34–43, 2013.

[14] K. Seth, IT Act 2000 vs. 2008 – Implementation, challenges and the role of adjudicating officers, *Proceedings of the National Seminar on Enforcement of Cyberlaw*, 2010.

[15] B. Shebaro and J. Crandall, Privacy-preserving network flow recording, *Digital Investigation*, vol. 8(S), pp. S90–S100, 2011.

[16] W. van Staden, Protecting third party privacy in digital forensic investigations, in *Advances in Digital Forensics IX*, G. Peterson and S. Shenoi (Eds.), Springer, Berlin Heidelberg, Germany, pp. 19–31, 2013.

[17] R. Verma, J. Govindaraj and G. Gupta, Preserving dates and time-stamps for incident handling in Android smartphones, in *Advances in Digital Forensics X*, G. Peterson and S. Shenoi (Eds.), Springer, Berlin Heidelberg, Germany, pp. 209–225, 2014.