EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Digital Forensic Framework for Protecting Data Privacy during Investigation

Suvarna Chaure[1,*], Vanita Mane[2]

[1] Department of Computer Engineering, Ramrao Adik Institute of Technology, Navi Mumbai,India
[1] Department of Computer Engineering, SIES Graduate School of Technology, Navi Mumbai,India
[2] Department of Computer Engineering, Ramrao Adik Institute of Technology, Navi Mumbai,India

Abstract

Rapid technological breakthroughs, a surge in the use of digital devices, and the enormous amount of data that these devices can store continuously put the state of digital forensic investigation to the test. The prevention of privacy breaches during a digital forensic investigation is a significant challenge even though data privacy protection is not a performance metric. This research offered solutions to the problems listed above that centre on the efficiency of the investigative process and the protection of data privacy. However, it's still an open problem to find a way to shield data privacy without compromising the investigator's talents or the investigation's overall efficiency. This system proposes an efficient digital forensic investigation process which enhances validation, resulting in more transparency in the inquiry process. Additionally, this suggested approach uses machine learning techniques to find the most pertinent sources of evidence while protecting the privacy of non-evidential private files.

Keywords: Privacy preservation, Digital Forensics, Machine Learning

Received on 11 June 2023, accepted on 28 August 2023, published on 27 September 2023

doi: 10.4108/eetsis.4002

1. Introduction

Computers are widely used and required in today's world, whether in business, at home, in the office, or in a variety of other settings. The digital world is rapidly developing as many and many people join the cyber globe and benefit from it. However, not all internet users follow the recognized and lawful use of computers. The negative use of computers and the internet for unlawful or unauthorized work, has become increasingly popular in this period. Computer misuse causes a plethora of sub domains that must be avoided. Digital forensics (DF) plays a critical role in recognizing and evaluating these subdomains, acting as a dissuasion to cyber-attacks. Cyber experts utilize digital forensics techniques and technologies to collect and correlate digital evidence to uncover the truth behind any cybercrime [1]. In the current environment, both commercial and public businesses face significant risks from a variety of IT security threats. If an incident happens, a forensic inquiry may become necessary because not all incidents can be fully resolved. Digital forensic investigations can take a lot of time and money. Some of these effects are so severe that they endanger businesses' ability to compete or even remain in business. Data must be collected from a variety of devices, some of which may contain sensitive employee or customer information, as part of a deception within a business or government agency [2]. Additionally, the widespread usage of digital devices makes it difficult for investigators to collect data while maintaining privacy. Data privacy protection

*Corresponding author. Email: suvarnakendre@gmail.com

difficulties develop in these situations. Obtain the consent of anybody who might be in possession of private information; this could be a laborious and time-consuming process. Specially those people who are not involved in the crime. Additionally, laws in the EU, like the GDPR, mandate that businesses implement the necessary organizational and technical safeguards to secure customer data. According to Article 8 of the EU Charter of Fundamental Rights, everyone has the right to have their own personal information protected. As per GDPR company needs to take appropriate measure to protect personal data of employees. However, the preservation of privacy and the thoroughness of the investigation, however, are mutually exclusive. Since there are four categories of evidence that can be relevant in a digital forensics' inquiry, finding the relevant one is crucial. These categories help investigator in effective investigation [3]. Stricter standards, laws, and regulations for organizations dealing with private information create obstacles in the data processing for Digital Forensics investigations. In order to create a hard disc picture that can be used for further analysis, this research focuses on a novel method for integrating privacy protection elements into digital forensics. The suggested method will be used to determine whether a particular file will be sent to a DF investigator during the collecting and inspection procedure phase of a DF investigation.

1.1 Forensic Limitations

Information security has undergone changes from a conventional strategy to an intelligent system. Due to the drawbacks of conventional approaches, such as labor-intensive manual data analysis and intelligence envisage to make the evidence more coherent, more researchers are focusing on presenting intelligent systems and frameworks based on forensic case studies.

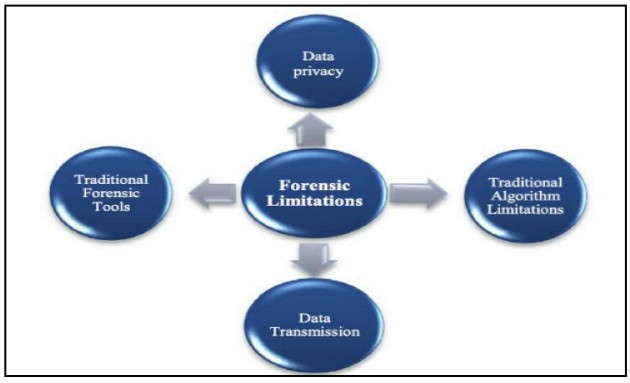

Figure 1. Forensic Limitations [33]

1.2 Challenges in Digital Forensic

Cybercrimes have flooded the cyberspace in the current situation, posing several difficulties for cybersecurity professionals. Digital forensics has become extremely important in the process of investigating an incident connected to cybercrime, as attackers can abuse end users or their businesses due to their lack of awareness. But this has also given specialists and examiners a lot of problems and difficulties, some of which are related to technology or innovation, others to standards and regulations, and still others to the fundamental operation of an inquiry. Three key categories can be used to classify the primary problems and challenges in digital forensics.

- Root Related Issues & Challenges
- Legislation Related Issues & Challenges
- Scientific Issues & Challenges

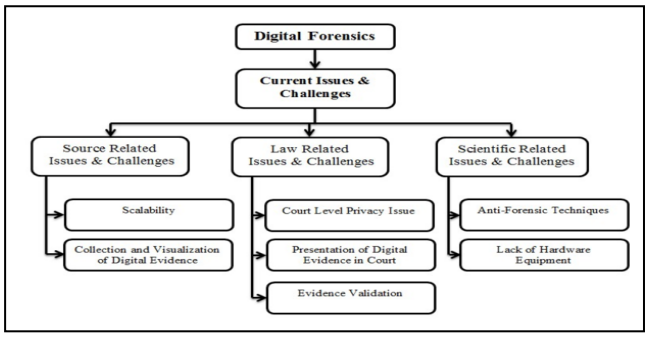

Figure 2. Tree structure of the issues and challenges in Digital Forensics [27]

1.3 Privacy issues in Digital Forensics

Table 1 provides a list of privacy challenges prompted by digital forensics, considering the key aspects of the existing ecosystem of technologies and the field. Certain publications examined in this study only touch on a few of the mentioned difficulties. Additionally, there are some problems that are interconnected in some way. For instance, even with specific approvals, there is a high possibility of a third party's privacy being violated because the consent of a user does not govern that user's storage of records about other users. Also, informed consents must be straightforward and unambiguous, which is getting harder to do because of the various jurisdictions, laws, standards, and users themselves.

Digital forensic investigations require new approaches that include forensic image preprocessing, machine learning-based filtering of the most relevant records, and privacy level assessments to address the above issues [4]. This solution should open new ways to bring state-of-the-art digital forensics research and systems together in one place. Few of the objectives behind taking up this research is mentioned below.

To develop an intelligent framework for preserving data privacy during forensic investigation.

Digital forensic practices	Privacy issues
Arbitrary digital data gathering and acquisition	It is necessary to prevent Third Party Privacy Breach (TPPB).
Full disc images are made, then examined.	Deleted files may result in erroneous allegations.
Information can be gathered from a personal device.	Users must be able to give informed permission that is complete and comprehensible.
The data to be collected during the inquiry may be housed on servers in many nations.	Different legal systems may have different concepts of privacy.
A large group of subject-matter experts must approve of and test digital forensics tools and procedures.	Requirements for privacy must be built into all currently used technologies and methodologies.

Table 1. Digital Forensic Practices and Privacy issues

1 To protect privacy of forensic files.
2 To implement an approach based on machine learning algorithms for identifying potential evidence files.
3 This saves the examiner's time and assists them in preventing unintentional data privacy violations when evaluating privacy levels of evidence.
4 To achieve transparency in forensic investigation process.
5 To create a repository of significant amount of reference files, which will be helpful in making valid conclusions about privacy of a particular file**.**

The rest of the paper is organized as follows: Prior to providing information on the present state of privacy protection, an overview of background knowledge, a literature review, and details on the current implementations are given. Based on that, Section 3 will construct the analysis of digital forensic investigations. Section 4 depicts the pro-posed methodology. Section 5 includes initial results.

2. Literature Survey

Personal computers brought about a kind of revolution in the lives of regular people as they were used more frequently in daily life. While most people used PC power to enhance their quality of life and the welfare of society, a small minority tried to exploit it maliciously. Investigative agencies that were tasked with looking into such computer abuse instances discovered ground-breaking techniques for gathering evidence from them. The processes used to conduct an inquiry on devices and entities that contain digital data are becoming standardized by researchers and practitioners in this domain. A digital forensic process model outlines the steps that should be taken from the initial response to an occurrence until the inquiry is finished. It serves as the investigators' user handbook, instructing them on how to gather and examine potential evidence from gadgets [1].

2.1. Data Privacy in Digital Forensics

The majority of the current digital forensic frameworks are focused on forensic analysis, but there are still no integrated solutions for privacy protection, evidence mining, or evidence storage. However, most early forensic research models lacked a distinct idea for how to handle each phrase. One of the key issues is privacy protection, particularly when detectives take bit-by-bit photographs of digital devices, examine the complete image data, and present it in court. Such actions may be against the law on privacy and human rights. [3,4,5] argued in favour of completely automated data selection to pick the private and/or relevant data and suggested different privacy levels for private information computer forensics. According to the author in [7], DF's capability is limited by privacy-preserving actions. This is the justification behind Srinivasan's proclamation of 10 rules that the DF inquiry procedure must follow. Given the central role that individuals and their personal information will play in present and future digital research, authors in [6,7] want to shed light on this intriguing, highly demanding topic. They also examined the role of anonymity in digital forensics. The authors have also thoroughly examined the state of the art in the aforementioned privacy-aware digital forensic methodologies. The difficulties and problems that forensic investigators currently confront have been discussed by the authors of [8]. A list of the various categories of digital forensics is also included. The relevance and necessity of future study into digital forensic methods is highlighted in the paper's later sections. The authors pointed out the necessity for clear and defined global cybercrime and forensic investigation regulations and standards because there is too much uncertainty about the laws and territorial procedures. The identified research gap thus motivated the study's author to think about creating a new digital evidence paradigm with safeguarding data privacy as a fundamental component.

Clustering may help a DF investigation's extraction process run more efficiently. It can be used as a filtering algorithm by grouping only useful data objects. Therefore, several pertinent document clustering techniques have been investigated and examined. According to [13], the purpose of document clustering is to separate a sizable number of documents into more manageable groupings that are related to subjects. To decrease the amount of time required to recover meaningful information, it is crucial to break up the information overload into smaller pieces. Also pointed that the clustering process is strongly dependent on user-specific external information that must first be input into the clustering algorithm. Document clustering in DF was used by [14] to group documents kept on a person’s computer. Semantic document clustering technique can group documents into subjects. It searches through a long collection of synonyms to locate words that are frequently used in illegal activity. [15] provide a comprehensive assessment of the

domains that are currently thought to be privacy-aware along with their level of DF achievement. The authors draw attention to unresolved issues as well as the competing demands of privacy and conducting a DF investigation. Comparing existing security-aware techniques for DF in databases, computers, servers, applications, networks, cloud mobile and IoT highlights the urgent need for adaptation. The authors in [16] propose the implementation of a privacy-preserving digital forensics skeleton, which would aid investigations by safeguarding suspect confidentiality through a variety of privacy rules and procedures. It features an access control mechanism that restricts access to private data and identifying digital evidence to only authorized investigators. The authors in [17] integrated several methodologies with the purpose of systematically detecting and excluding files containing critical information. This privacy-preserving technology can be implemented into a Digital Forensics assessment process to provide an image that is free of both private and vital information. This method shows how investigations in businesses may be aided and enhanced by adapting current algorithms and methods from adjacent fields to incorporate privacy-preserving features into the investigation process. Authors [37] attempt to close this gap by using a case study of a PIA (Privacy Impact Assessment) on a large data forensic platform. They have addressed the issue of how a PIA should be conducted for extensive digital forensic operations and outlines the privacy dangers and threats we discovered while carrying it out.

2.2 Data Privacy Solutions

Authors in [9] suggest a method that uses homomorphic and commutative encryption to shield the investigator from the privacy of data stored on a third-party service provider's storage facility. The process also makes sure that the service provider is kept in the dark regarding the investigator's questions. A comparable solution on a distant server is mentioned here. Authors in [10] implemented Identity Based Encryption is used to analyses network traffic data while maintaining anonymity. Offer general privacy guidelines for network forensic investigations as well. The authors in [11] examined current methods for incorporating data privacy into the DF investigation process. In the context of DF, they noticed a distinction between preservation and protection. The goal of preservation is to preserve the integrity of digital evidence during the whole inquiry while also recording every step of the process. Human rights are intended to be protected, and protection makes sure that they are upheld throughout the investigation. The authors in [12] proposed a method for eliminating pointless personal information from the analytic process to improve privacy. With this, it is possible to rule out even a remote chance of unintentionally disclosing personal information to unapproved parties. Because the cloned image's contents can be changed or removed, this technique poses a serious issue with protecting the integrity of digital evidence. According to Philip [19], Selective Imaging in DF is viewed as a decision not to capture all information feasible and was not approved as a result. It has, however, recently grown in popularity as a replacement for application areas where full coverage is no longer feasible due to enormous volumes of data. Selective imaging was divided by Philip into three categories:

- Manual Selective Imaging
- Semi-automatic Selective Imaging
- Automatic Selective Imaging

2.3 Intelligent Frameworks for DF

Clustering may help a DF investigation's extraction process run more efficiently. It can be used as a filtering algorithm by grouping only useful data objects. Therefore, several pertinent document clustering techniques have been investigated and examined. According to [13], the purpose of document clustering is to separate a sizable number of documents into more manageable groupings that are related to subjects. To decrease the amount of time required to recover meaningful information, it is crucial to break up the information overload into smaller pieces. Also pointed that the clustering process is strongly dependent on user-specific external information that must first be input into the clustering algorithm. Document clustering in DF was used by [14] to group documents kept on a suspect's computer. Semantic document clustering technique can group documents into subjects.
It searches through a long collection of synonyms to locate words that are frequently used in illegal activity. [15] provide a comprehensive assessment of the domains that are currently thought to be privacy-aware along with their level of DF achievement. The authors draw attention to unresolved issues as well as the competing demands of privacy and conducting a DF investigation. A strong necessity for adaption is demonstrated by contrasting existing privacy-aware approaches for DF in database, computer, server, networks, applications, mobile, cloud, and IoT. They suggest the potential reuse of several strategies from other DF fields. The authors in [16] propose the implementation of a privacy-preserving digital forensics framework, which would aid investigations by safeguarding suspect confidentiality through a variety of privacy rules and procedures. It features an access control mechanism that restricts access to private data and identifying digital evidence to only authorized investigators. The authors in [17] integrated several methodologies with the purpose of systematically detecting and excluding files containing sensitive information. This privacy-preserving technology can be implemented into a Digital Forensics assessment process to provide an image that is free of both private and vital information. This method shows how investigations in businesses may be aided and enhanced by adapting current algorithms and methods from adjacent fields to incorporate privacy-preserving features into the investigation process. A framework that respects user privacy and ensures a forensically sound digital investigation has been proposed by the authors [36] after they analyzed perceptions about data privacy preservation and issues related

to digital forensics investigation from three perspectives, namely, user, forensic investigator, and technology.

3. Privacy Preservation and Protection in Digital Forensics(P3DF)

In current scenario Digital forensic is challenged by significant amount of private data generated using digital devices. Such data may only be processed with the permission of the data subject and for specific purposes. Violations of these rules can result in hefty fines. As a result, these policies have an impact on DF and Digital Forensics Readiness (DFR), particularly in the industry and government sectors. An efficient investigation procedure that protects data privacy without limiting the investigator's powers should exist to avoid this problem.

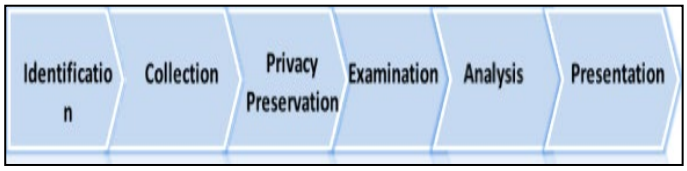

Figure 3. Phases of Privacy-preserving Digital Forensic Framework

The proposed method describes an approach that can meet the necessary concerns to conduct a privacy-preserving Digital Forensics investigation process. This work provides a new technique for pre-filtering sensitive information from examination materials before investigators get access to them in order to create forensically sound photos for a DF investigation. . This approach can be used to determine whether a specific file can be forwarded to a Digital Forensic (DF) investigator between the collecting and examination process phases of a DF investigation.

Figure 3 shows various phases of Digital forensic investigation process which includes added privacy preservation phase. The identification phase of the proposed Privacy-preserving DF framework is the first and is initiated by a crime being reported or a suspected occurrence. Here, the characteristics and nature of the offences are also listed.

Using synchronization and evidence analysis, the gathering phase of this system tries to preserve the digital murder scene for eventual validation. In this step, the contents of the original data are duplicated bit by bit. Imaging should be done in a method (i.e., write protected mode) that can stop any unintentional or accidental alterations made by an investigator when copying. The focus of this research is on Privacy Preservation phase which aims at an automated process to find the most correct data that are related to the issues under investigation. The relevancy of data is checked using machine leaning algorithms. For differentiation between non-private and private data advanced machine leaning can be used. If the data is private and non-relevant, then it can be removed. And if the data is private and relevant then encryption technique can be applied to prevent the evidence from unauthorized access, it can be helpful in legal rules to ensure the confidentiality. This helps accidental privacy violation by the investigators. The last step of this approach is to prepare a hard drive image without private information. This image can be sent for further examination and analysis phase.

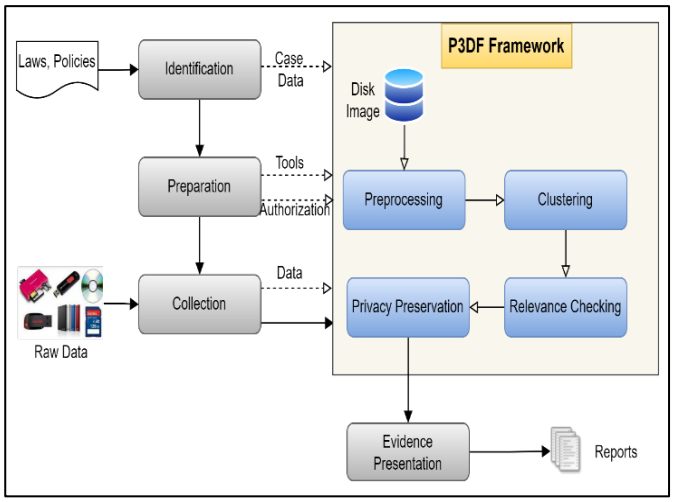

Figure 4. System Architecture of Privacy Preservation and Protection in Digital Forensics (P3DF) Framework

This paper suggests a method to preserve data privacy in digital forensic investigation procedures that doesn't affect the process' effectiveness or results. The proposed method shown in Figure 4 promotes accountability of the investigators and increases process transparency. The functional aspects of the recommended P3DF framework, which may be used for data collection, analysis, and preservation of digital data in a coordinated system without jeopardizing user privacy, are detailed. A P3DF architecture shows the many components and how they are interdependent. Identification, preparation, collection, pre-processing, clustering, privacy checking and privacy-preservation of evidence files and preservation of the evidence, presentation of the evidence are the phases of the P2DF framework.

4. Methodology

The framework accepts as inputs network logs, memory dumps, forensic exhibits and photographs of computers, laptops, cell phones, tablets, and other devices that hold data. The investigator enters all case-related information into a document as the inputs pass on to the following stage of pre-processing. The document is made up of forensically significant data that is specific to the case being investigated, including individual keywords, timings, and other helpful data. Every input image is processed to exclude forensically irrelevant information. The forensic image structure is also changed without sacrificing the input's integrity to enable

speedy and parallel operations in subsequent investigation phases. To enable concurrent data reading by forensic tools, the pre-processing seeks to reorganize and combine the data present in all the provided forensic images.

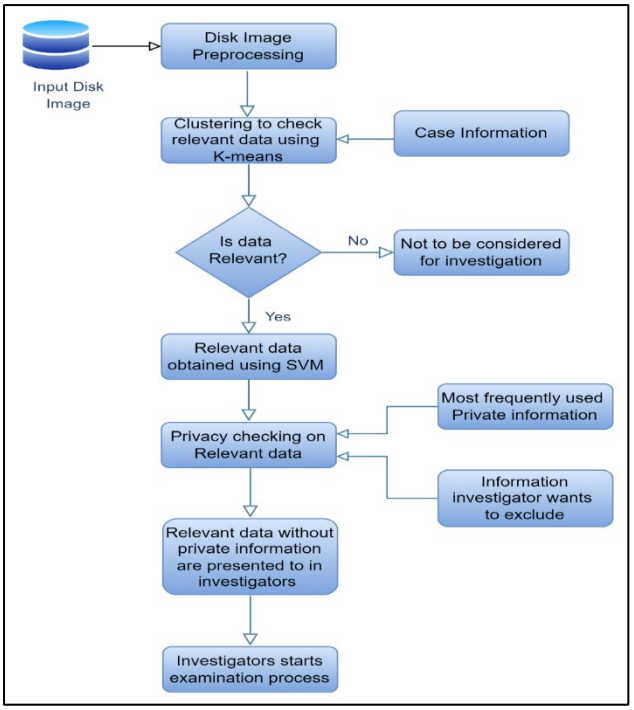

Figure 5. Flowchart of P3DF framework

The results of Pre-processing are then passed on to the next step of Clustering. Clustering may help a DF investigation's extraction process run more efficiently. This method concentrates on a physical device's file-based level. All text related data items, including emails, must first be recovered, recognized, and grouped using clustering algorithms in order to do this. Here, a relevance score for all possible evidence files collected through pre-processing is computed with the use of machine learning algorithms. The investigator can examine these files to determine whether they include enough evidence to support or refute the case. The investigator may continue to ask for the subsequent batch of the most pertinent documents for additional review, until all possible sources of evidence have been considered. Additionally, the framework determines whether a file is private and whether it contains any Personally Identifiable Information (PII) about the suspect. Correlation between each file's data privacy information and the relevant evidence rating is the main aim. The ability of the forensic examiner to conduct investigations will not be restricted by the privacy information of any given file. However, if a data privacy infringement occurs while an investigation is underway, the privacy quotient of each individual file would allow the criminal and the appropriate authorities to evaluate the severity of the violation. This, research also aims to create a repository of significant amount of reference files, which will be helpful in making valid conclusions about privacy of a particular file

5. Results

The 'Hacking-Case' was chosen by the authors for the prototype implementation because there wasn't a real-world digital forensic case to use as a model. On the website for the Computer Forensics Reference Data Sets (CFReDS) project run by NIST, the HackingCase files are accessible. The EnCase images (E01 files) were retrieved from the CFReDS website's Hacking Case page. Analysis of the dataset done using Autopsy tool shown in figure.

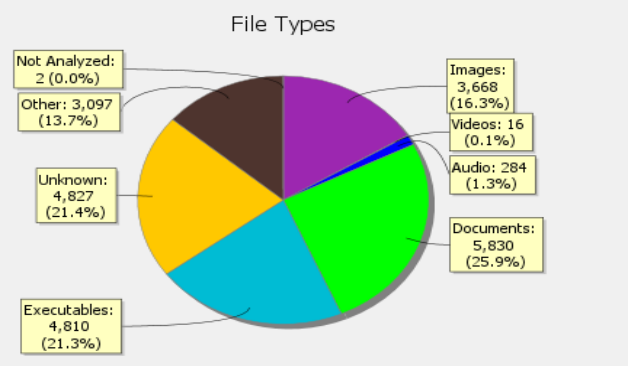

Figure 6. Analysis of Case dataset CFReDS website

For our initial analysis, we tried on text records. We collected the same from Encase. Here, we have chosen a text file which was available as a part of Forensic evidence. For all these disk images we have extracted the text and image files using AUTOPSY. After this we implemented average perceptron tagger for POS tagging. This was helpful when we wanted to extract noun. We have also used NLTK package and have extracted stop words.

Figure 7. Average Perceptron Tagger

The above figures show that, we were able to successfully separate the sensitive information using this technique.

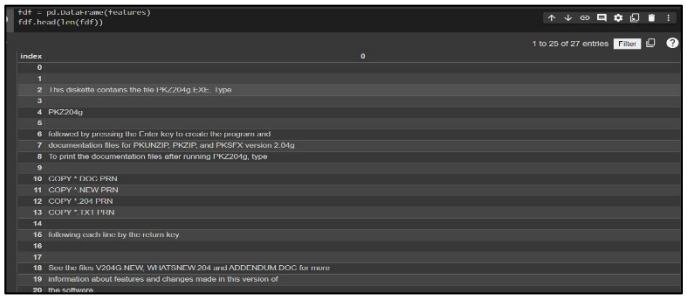

Figure 8. Extraction of Nouns as Sensitive Information

We compared the results of our model with the existing models. The proposed model has seemingly higher accuracy as compared to the current systems used for forensic investigations. The accuracy of the model is approximately near to 90%. Moreover, the proposed model had a better precision as well as recall value.

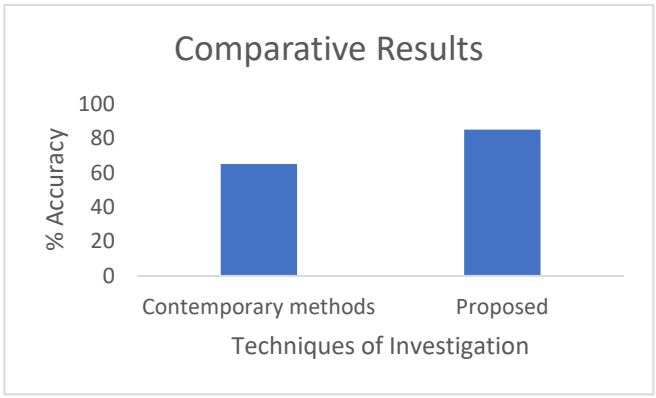

Figure 9. Comparative Results

6. Conclusion

In this work, a privacy-preserving method for digital forensics procedure is put forward. This method may be incorporated into the digital forensic examination process to give investigators access to a picture that is devoid of sensitive information. This suggested strategy aims to create effective machine learning algorithms to extract significant digital evidence from large and diverse amounts of data. This approach offers a fresh viewpoint on how to recognize, filter out certain files, and concentrate solely on pertinent data.

7. Future Work

The authors would like repeat this process in the future and take data from picture and PDF files. This will assist us in removing sensitive data from other records as well. Also, this model can be tested on real life dataset and can be refined as per investigators feedback.

References

[1] Pollitt, M. (2010). A History of Digital Forensics. In: Chow, KP., Shenoi, S. (eds) Advances in Digital Forensics VI. Digital Forensics 2010. IFIP Advances in Information and Communication Technology, vol 337. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15506-2_1

[2] Donn B. Parker. 2003. Computer crime. Encyclopedia of Computer Science. John Wiley and Sons Ltd., GBR, 349–353.

[3] Mark M. Pollitt. Computer Forensics: an Approach to Evidence in Cyberspace. In Proceedings of the National Information Systems Security Conference, volume 2, pages 487–491, 1995.

[4] Gary Palmer. A Road Map for Digital Forensic Research. In First Digital Forensic Research Workshop, Utica, New York, pages 27–30, 2001.

[5] Brian D. Carrier and Eugene H. Spafford. Getting Physical with the Digital Investigation Process. International Journal of Digital Evidence, 2(2):1–20, 2003.

[6] Mocas, Sarah. (2004). Building theoretical underpinnings for digital forensics research. Digital Investigation. 1. 61-68. 10.1016/j.diin.2003.12.004.

[7] Gong Ruibin, T. Yun, and M. Gaertner. Case-relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework. International Journal of Digital Evidence, 4(1):147–67, 2005.

[8] Marcus K. Rogers, James Goldman, Rick Mislan, Timothy Wedge, and Steve Debrota. Computer Forensics Field Triage Process Model. Journal of Digital Forensics, Security and Law, 1(2):19–38, 2006

[9] Wynand van Staden. Protecting Third Party Privacy in Digital Forensic Investigations. In Gilbert Peterson and Sujeet Shenoi, editors, Advances in Digital Forensics IX, pages 19–31. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013

[10] Ali Dehghantanha and Katrin Franke. Privacy-respecting Digital Investigation. In 2014 Twelfth Annual International Conference on Privacy, Security and Trust, pages 129–138, July 2014.

[11] Waleed Halboob, Ramlan Mahmod, Nur Izura Udzir, and Mohd. Taufik Abdullah. Privacy levels for computer forensics: Toward a more efficient privacy-preserving investigation. Procedia Computer Science, 56:370– 375, 2015.

[12] Asou Aminnezhad, Ali Dehghantanha, and Mohd Taufik Abdullah. A survey on privacy issues in digital forensics. International Journal of Cyber-Security and Digital Forensics, 1(4):311–324, 2012.

[13] S. Srinivasan. Security and privacy vs. computer forensics capabilities. Information Systems Control Journal, 4:1–3, 2007.

[14] Bilal Shebaro and Jedidiah R. Crandall. Privacy-preserving Network Flow Recording. Digital Investigation, 8:S90 – S100,

2011. The Proceedings of the Eleventh Annual DFRWS Conference.
[15] Alec Yasinsac and Yanet Manzano. Policies to Enhance Computer and Network Forensics. In Proceedings of the 2001 IEEE Workshop on Information Assurance and Security. 2001
[16] Shahzad Saleem, Oliver Popov, and Ibrahim Bagilli. Extended abstract digital forensics model with preservation and protection as umbrella principles. Procedia Computer Science, 35:812–821, 2014.
[17] Shuhui Hou, Tetsutaro Uehara, S. M. Yiu, Lucas C. K. Hui, and K. P. Chow. Privacy preserving multiple keyword search for confidential investigation of remote forensics. In Proceedings of the 2011 Third International Conference on Multimedia Information Networking and Security, MINES '11, page 595–599, USA, 2011. IEEE Computer Society.
[18] Shuhui Hou, Siu-Ming Yiuy, Tetsutaro Ueharaz, and Ryoichi Sasakix. A privacy-preserving approach for collecting evidence in forensic investigation. International Journal of Cyber-Security and Digital Forensics (IJCSDF), 2(1):70–78, 2013.
[19] Philip Turner. Selective and intelligent imaging using digital evidence bags. Digital Investigation, 3:59–64, 2006.
[20] Johannes Stuttgen, Andreas Dewald, and Felix C. Freiling. Selective ¨ imaging revisited. In Holger Morgenstern, editor, Seventh International Conference on IT Security Incident Management and IT Forensics (IMF), 2013, pages 45–58, Piscataway, NJ, 2013. IEEE.
[21] Jonathan Grier and Golden G. Richard. Rapid forensic imaging of large disks with sifting collectors. Digital Investigation, 14:34–44, 2015.
[22] Christian Zoubek and Konstantin Sack. Selective deletion of nonrelevant data. Digital Investigation, 20:92–98, 2017.
[23] Susan Dumais, John Platt, David Heckerman, and Mehran Sahami. Inductive learning algorithms and representations for text categorization. In Niki Pissinou, editor, Proceedings of the seventh international conference on Information and knowledge management, pages 148–155, New York, NY, 1998. ACM.
[24] Han-Joon Kim and Sang-Goo Lee. A semi-supervised document clustering technique for information organization. In Arvin Agah, editor, Proceedings of the ninth international conference on Information and knowledge management, pages 30–37, New York, NY, 2000. ACM.
[25] Gaby G. Dagher and Benjamin C.M. Fung. Subject-based semantic document clustering for digital forensic investigations. Data & Knowledge Engineering, 86:224–241, 2013
[26] Ana Nieto, Ruben Rios, Javier Lopez, Wei Ren, Lizhe Wang, KimKwang Raymond Choo, and Fatos Xhafa. Privacy-aware digital forensics. Computing. Institution of Engineering and Technology, 2019.
[27] Pandey, Abhishek Kumar,et al. "Current Challenges of Digital Forensics in Cyber Security." Critical Concepts, Standards, and Techniques in Cyber Forensics, edited by Mohammad Shahid Husain and Mohammad Zunnun Khan, IGI Global, 2020.
[28] Advances in Digital Forensics XV, 15th IFIP WG 11.9 International Conference, Orlando, FL, USA, January 28–29, 2019
[29] Ferguson, I., Renaud, K., Wilford, S., Irons, A., (2019) PRECEPT: A Framework for Ethical Digital Forensics. Investigations. Journal of Intellectual Capital, 20(7), ISSN : 1469-1930
[30] M. Bas Seyyar, Z.J.M.H. Geradts, Privacy impact assessment in large-scale digital forensic investigations, Forensic Science International: Digital Investigation, Volume 33, 2020, 200906, ISSN 2666-2817.
[31] Abulaish, Muhammad,et al. "P2DF: A Privacy-Preserving Digital Forensics Framework." IJDCF vol.13, no.6 2021: pp.1-15.
[32] Ludwig Englbrecht and Günther Pernul. 2021. A Combined Approach for a Privacy-Aware Digital Forensic Investigation in Enterprises. In Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES '20). Association for Computing Machinery, New York, NY, USA, Article 58, 1–10. DOI:https://doi.org/10.1145/3407023.3407064.
[33] Hussein Ismael Sahib, Mustafa Qahatan AlSudani, Mohammed Hasan Ali, Haydar Qassim Abbas, Kohbalan Moorthy, Myasar Mundher Adnan, Proposed intelligence systems based on digital Forensics: Review paper, Materials Today: Proceedings,2021
[34] Victor R. Kebande, Richard A. Ikuesan, Nickson M. Karie, Sadi Alawadi, Kim-Kwang Raymond Choo, Arafat Al-Dhaqm, Quantifying the need for supervised machine learning in conducting live forensic analysis of emergent configurations (ECO) in IoT environments, Forensic Science International: Reports, Volume 2, 2020.
[35] Nampoothiri, A. P., & Madhavu, M. L. (2015). *Email forensic analysis based on k- means clustering. 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI).* doi:10.1109/icacci.2015.7275710
[36] Ayush Kumar Verma, Krishnan Ramanathan; Data privacy preservation in digital forensics investigation. *AIP Conference Proceedings* 3 October 2022; 2519 (1): 030051. https://doi.org/10.1063/5.0109813
[37] M. Bas Seyyar, Z.J.M.H. Geradts, Privacy impact assessment in large-scale digital forensic investigations, Forensic Science International: Digital Investigation, Volume 33, 2020, ISSN 2666-2817.