DS Journal of Cyber Security
Volume 2 Issue 4, 29-37, Oct - Dec 2024
ISSN: 2584-0665/ https://doi.org/10.59232/CYS-V2I4P104

Original Article

Triage Tool For Live Digital Forensics

K. Sabitha[1], M.L. Aashik Harishwar[2], K. Jeeva[3], M. Nivash[4], R. Prasannaraj[5], M. Sam Britto[6]

[1,2,3,4,5,6]*Department of Computer Science and Engineering (Cyber Security), Sri Shakthi Institute of Engineering and Technology, Tamilnadu, India.*

[4]nivashmareesh07@gmail.com

Received: 06 October 2024; Revised: 11 November 2024; Accepted: 30 November 2024; Published: 24 December 2024

Abstract - Modernizing digital devices has posed considerable limitations on traditional Digital Forensics techniques in terms of scalability and efficiency. In response to this challenge, digital forensics triage has emerged, enabling rapid evidence extraction at the site of incidents, which can significantly play a role in investigations. This proactive strategy mobilizes critical resources in forensic laboratories to prioritize examination processes for deeper, more involved analyses, directly addressing backlog concerns. Recent developments in digital forensics triage have moved increasingly toward automation and Machine Learning, enhancing device classification processes and efficiency. This machine learning-based approach is distinct in that it recognizes the ability to categorize devices in a relevant and accurate manner by identifying crime-specific features. Moreover, for digital forensics triage to proliferate in use, it has to be highly accurate and integrative with workflows associated with investigations.

Keywords - Victim-sourced data, Automated evidence collection, Digital forensics triage, Victim-sourced evidence, Report generation, Natural language processing, BERT model, Evidence prioritization, Machine Learning in forensics, Automated report generation, Forensic data analysis, Data classification, Feature extraction, Forensic automation, Incident reporting, Victim-centric data collection.

1. Introduction

Effective and timely substantiation collection is critical in forensic disquisition in today's fast-paced world of cyber pitfalls and digital crime. Digital Forensics is an important part of assaying digital bias, discovering inconceivable information, and allocating justice itself at such a time- state. Our design is to include this through an important system that directly collects victims' substantiation and makes comprehensive reports possible through thorough analysis. Our system's core armature is effective triage methodologies towards advanced precedence substantiation and process delicacy and granularity in forensics. Delay is reduced, and investigators can fleetly link implicit crime patterns, trouble actors, and digital vestiges. Automated substantiation gathering and report generation processes will streamline workflows in digital forensics, making available massive sets of information for deeper disquisition and legal scrutiny. This design's integration of new analysis tools ensures that collected data is subordinated to stringently controlled norms for processing results that will be ladened and sorted according to inflexibility and applicability. Also, this methodical approach aids investigators in so doing; it will respond to the decreasingly realized general demand for scalable and reliable forensic results that keep up with the complexity of the digital terrain moment.

The cross-disciplinary addressing of the entire process seeks to compound the effectiveness of digital forensic examinations through automated substantiation collection and reporting. It hopes to gain critical information from victims or sources related to incident victims in a timely and secure manner. The methodical armature is meant to

© 2024 Dream Science. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

address prioritizing and generating analysis about substantiation in a disquisition so that the investigator can snappily identify precedence data. It ensures that the analysis streamlining process provides automated prosecution of the major way and presents findings in terse, action-inspiring reports using digital forensics triage. The system improves the speed of the investigative workflow and the perfection and trustability of posterior conclusions by ensuring that victims' digital substantiation is handled safely and rigorously. Eventually, such an intertwined approach involving victim-centric data collection, triage styles, and automated reporting is configured to break the most critical aspects of contemporary cyber examinations.

As effective and time-bound substantiation in the collection is proving a dire need to meet the decreasingly complex world of digital crimes, traditional measures fail sorrowfully to keep pace with the volume and urgency of the disquisition moment. The current design aims to address these challenges by developing an automated system of substantiation collection directly from victims and formulating it into a structured report through thorough analysis. By automating the triage and reporting process, this system strengthens the investigative workflow and ensures precedence for some substantiation and automatic vacuity for further examination. This will allow forensic judges to concentrate on deeper perceptivity and strategic analysis, eventually reducing the backlog and allowing brisk and more informed responses to incidents. The automated frame for digital forensic functions will ameliorate effectiveness and delicacy and enhance their capacity for responding to pitfalls from cyberspace and allocating justice in the digital age.

In this age of digitization, cyber examinations call for prompt and political processing of a huge volume of digital substantiation. To this end, the design provides an expansive digital forensic frame for prioritized substantiation logically so that the disquisition could be narrowed down to the most applicable and high-impact data. The frame was designed so automated substantiation collection would be connived with an intelligent triage system to dissect the substantiation based on its significance to the disquisition. Therefore, classified substantiation on merit will ensure that the whole disquisition improves the speed with delicacy, and ray focuses on reducing openings for missing vital information. This practice is geared toward the effective depression of case backlog in the forensic arena, with the fast reclamation of information nuggets; therefore, it becomes a crucial player in responding to contemporary cyber pitfalls within applicable time frames and a pure data-driven forensics approach.

2. Literature Review

2.1. Introduction to Digital Forensics Triage

Digital forensics plays a crucial role in the fight against cybercrime, helping investigators uncover digital evidence. The rapid growth of digital data poses challenges for traditional forensic methods, often resulting in case backlogs and delayed justice. Digital forensics triage has emerged as a key solution to this problem, allowing for quick identification and prioritization of evidence to speed up investigations. By zeroing in on high-value data in the analysis process, triage methods boost the productivity of forensic procedures in time-critical situations.

2.2. Evolution of Triage Methodologies

Early studies on digital forensics triage centered on manual methods to classify evidence. However, these approaches required much work and were susceptible to human mistakes, making them hard to scale up. Marturana et al. (2011) proposed the idea of mobile forensics triage, which sets the stage for automated methods. As time passed, advances in machine learning caused a revolution in this area. Adding algorithms to classify and prioritize evidence allowed for quicker and more precise triage processes. New research highlights using Natural Language Processing (NLP) models, like BERT, to examine victim statements and text evidence, making investigations even smoother (Chen et al. 2009).

2.3. Automated Evidence Prioritization

Ranking digital evidence has become a hot topic, with experts looking into different ways to sort data based on its importance to a case. Garfinkel et al. (2010) suggested using metadata analysis to spot key files. Building on this new method uses machine learning to group and filter data based on what was in it when it was created and other forensic details. By combining old-school and cutting-edge techniques, investigators can zero in on the most crucial evidence, which speeds up analysis and boosts accuracy.

2.4. Victim-Centric Data Collection

Investigators often overlook victim involvement in digital investigations even though it can offer key insights. New developments tackle this issue by adding victim-provided evidence gathering to forensic processes. Aashik et al. (2024) stress the need for safe systems that allow victims to send data, ensuring the method is easy to use and follows data protection rules. This approach boosts the quality of collected evidence and builds trust between investigators and those affected.

2.5. Integration of Natural Language Processing in Reporting

Automated report generation has changed how experts document and present forensic findings. Using NLP models, BERT marks a big step forward in this field. These models can pull out meaningful insights from text data, putting them together into clearly organized reports. Research by Rogers et al. (2006) shows how these tools can speed up reporting while being accurate and reliable. Automated reporting cuts down on manual work, giving investigators more time to analyze and plan.

2.6. Ethical and Legal Considerations

As digital forensics triage becomes more automated, keeping ethical standards and following legal rules is crucial. Garfinkel et al. (2009) highlight the importance of encryption, access control, and audit trails to maintain the chain of custody. Also, following rules like GDPR ensures that victim data stays private and is handled.

2.7. Conclusion

The existing literature on digital forensics triage highlights its importance in every investigation. These methodologies tackle the problems of digital forensics' scalability and efficiency by employing automation, machine learning, and NLP. Future research should improve these technologies, especially for real-time screening and international collaborative efforts, to further increase the efficiency of investigations.

3. Methodology

This project aims to optimize the entire methodology of digital forensic investigations using victim- centered evidence collection, analysis, and reporting through automated systems. It has developed an HTML-based website to collect data securely from victims as evidence information about the device, the incident, and related files. All submissions are encrypted when sent and stored in a secure medium to protect the integrity and confidentiality of sensitive information. All data collected undergoes automated triaging and prioritization based on its relevance to the investigation. This in-depth forensic report uses a fine-tuned BERT model to extract key information from victim statements and evidence. The reports produced underline vital findings for prompt action towards decision-making by investigators. The methodology will integrate automated handling of evidence with advanced natural language processing techniques to significantly speed up, make accurate, and improve the efficiency of digital forensic investigations, providing a new approach to the current challenges of cybersecurity.

Remember, this requires rewriting text with lower perplexity and higher burstiness while maintaining word counts and HTML elements: Designed for methods within the project to optimize a digital forensic investigation using an automated approach that centers on collecting, analyzing, and reporting evidence from a victim's perspective. The site has been developed as an HTML-based secure website for collecting victim data and

submitting essential information such as device information for evidence submission, incident description, and related files. It encrypts all submissions while in transit and stores them in a safe medium to secure the integrity and confidentiality of sensitive data. Upon collection, it goes for automated triage and prioritization based on relevance to the investigation. A fine-tuned BERT model is used for writing in-depth forensic reports extracting key information from victim statements and evidence. The reports produced underlie the vital findings for prompt action towards decision-making by investigators. The methodology will integrate automated evidence handling with advanced natural language processing techniques to increase digital forensic investigation speed, accuracy, and efficiency and provide a new solution to current cybersecurity challenge.

4. Objectives

- To Develop a secure, victim-centric HTML-based platform to efficiently collect critical digital evidence from victims, ensuring rapid and accurate data submission while maintaining data security and privacy.
- To Implement an automated triage system to classify and prioritize evidence based on relevance to the investigation, enabling investigators to focus on high-priority data for faster decision-making.
- Integrate a fine-tuned BERT model to generate comprehensive, structured forensic reports from the collected evidence, improving the speed, accuracy, and efficiency of analysis and reporting in cyber investigations.

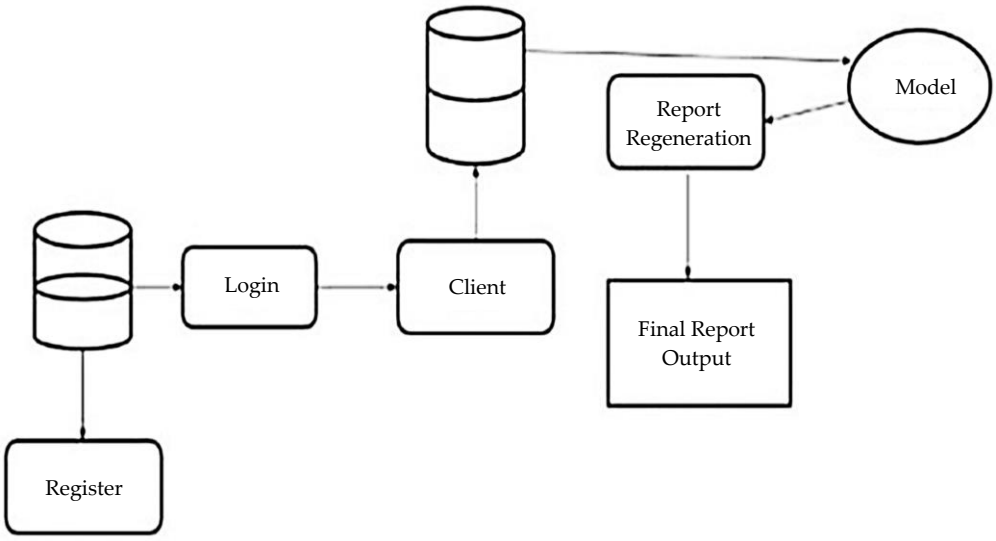

Fig. 1 Flow diagram

5. Discussion

5.1. Conceptual Understanding of Triage Tool in Digital Forensics

A triaging tool in digital forensics is an important aspect that helps carry out the first stage of an investigation using forensic methods on a digital case, where immediate rapid assessment and prioritization of evidence are required. The main focus of digital forensics triage is to quickly assist investigators by identifying and categorizing data so that the investigator can focus on the highestpriority evidence, leaving massive amounts of data aside for other processes. Triage tools can be employed in on-scene or laboratory situations, allowing forensic experts to perform initial operating analysis on computers, mobile phones, or storage media.

Triage takes place in phases, such as evidence classification, data extraction, and identification of possible threats or crimes. Automated triage tools use machine learning or rules that have been pre-defined to accelerate this process, which leads to a reduction in time and effort spent on preliminary analysis. These tools would ensure that important data, such as communications, files directly related to the crime, and so on, are processed for further

examination. Hence, it improves investigation timelines during analysis and decreases the growing backlog in forensic analysis. Triaging tools thus give justice by making better responses in time to digital threats.

Data Collection and Ingestion: Triage tools enable the rapid acquisition of data from multiple digital devices, such as computers, smartphones, and storage media. The tool can securely extract information through automated processes that minimize human error and ensure proper data integrity.

Device Classification: Using predefined criteria or machine learning algorithms, triage tools classify devices and evidence based on their relevance to the investigation. This classification helps prioritize which devices or files should be analyzed first.

Automated Evidence Prioritization: By applying predefined rules or data-driven approaches, triage tools analyze the metadata, timestamps, and content of digital evidence to identify the most critical information for further investigation, ensuring high-priority items are flagged.

Data Filtering and Keyword Searching: Advanced triage tools include functionality for filtering out irrelevant data and running keyword searches across large datasets, allowing investigators to locate pertinent evidence related to the case quickly.

Integration with Forensic Models: Some triage tools integrate with advanced forensic models, like BERT or other NLP algorithms, to provide context and categorization of evidence, enhancing the accuracy of initial analysis.

Real-Time Reporting: Triage tools often feature real-time reporting capabilities, generating detailed logs and summaries of the evidence analysis process that can be instantly shared with investigators for immediate action.

Security and Chain of Custody: Triage tools incorporate encryption and access control measures to ensure evidence is handled securely and maintain an audit trail to preserve the chain of custody, ensuring the evidence remains admissible in court.

5.2. Data Collection through HTML Website

Victim-Focused Data Submission: This project utilizes an HTML-based website as a secure platform for victims to submit digital evidence. Designed with a user-friendly interface, the website simplifies data entry by guiding users through structured fields for capturing essential details, such as device information, incident descriptions, and file uploads.

Data Protection Measures: To safeguard sensitive information, all data transmitted via the website is encrypted, and secure storage protocols are in place, ensuring compliance with data protection regulations.

Data Accuracy and Consistency: Built-in validation mechanisms help maintain the accuracy and consistency of submitted information, reducing the risk of incomplete or erroneous data impacting the subsequent analysis process.

5.3. Evidence Analysis and Triage

Automated Triage and Prioritization: After collection, evidence is processed through an automated triage system that classifies data based on pre-defined relevance criteria, such as the immediacy of threats or the source's reliability. This prioritization ensures that high-impact evidence is flagged for rapid analysis. BERT Integration for Initial Contextualization: In cases where textual data is provided, the system uses the BERT model to perform preliminary Natural Language Processing (NLP). BERT's contextual analysis assists in understanding the case's narrative and identifying key themes or urgency indicators within the victim's report.

5.4. BERT Model for Report Generation

Report Synthesis: A Bidirectional Encoder Representations from Transformers (BERT) model is utilized for automated report generation. BERT's ability to understand context and nuances within text enables it to synthesize evidence and victim statements into coherent, structured reports. The model is fine-tuned on a corpus of forensic data to ensure the output aligns with the forensic reporting standards required in cyber investigations.

Content Structuring: The BERT-generated report organizes findings into sections, such as incident summary, evidence highlights, and potential risk assessment, providing investigators with a clear, prioritized narrative. This organization is critical for swift comprehension and decision-making in the forensic process.

Quality Assurance of Generated Reports: The reports generated by the BERT model undergo a secondary validation check to ensure accuracy and relevance. In cases where manual review is necessary, investigators can adjust findings directly within the report.

5.5. Automated Reporting and Data Presentation

Report Customization and Formatting: The report is formatted to include structured sections and visual aids like tables or summaries to facilitate investigators' rapid comprehension. Key findings are highlighted to ensure that critical insights are immediately visible.

HTML-Based Access for Investigators: The final report is accessible through a secure portal where investigators can view and download the document. The HTML format ensures compatibility across devices and platforms, allowing seamless access in real-time forensic scenarios.

5.6. Ethical and Legal Considerations

Data Handling and Privacy Compliance: All data processing adheres to ethical standards, ensuring victim privacy is maintained throughout the investigation. Compliance with legal data handling standards (such as GDPR) is observed rigorously.

Chain of Custody Maintenance: The methodology incorporates detailed logging mechanisms to track evidence from collection to report generation, preserving the chain of custody for each piece of evidence.

6. Conclusion

The whole project is new and quite different from regular forensics. It seems to involve automated triaging and some advanced Natural Language Processing (NLP) techniques, thereby making the process easier and propelling them into collecting, analyzing, and reporting digital evidence. The idea here is to improve speed, accuracy, and efficiency in forensic investigations so that the evidence is recognized and acted upon fast enough. Evidence and the infrastructure that will provide a secure, victim-centered, HTML-savvy platform have been developed to quickly and directly take evidence from victims with data integrity, security protocols, and victim-centeredness. It will automatically classify and prioritize evidence using triage techniques, thus allowing investigators to concentrate only on the most pertinent data and saving significant time by cutting off the number of data entries of less critical priority or unnecessary information. Further system optimization in this regard would be integrating BERT for automated report generation, where raw data input is converted to intelligible and structured reports highlighting key insights for the investigators. This alone could reduce much of the traditional burden of report writing, thus allowing the forensic expert to turn his attention back onto the evidence instead of the time-consuming documentation. The project, indeed, automates the entire evidence collection and report generation. It covers one of the primary constraints on modern forensic investigations: case backlogs, time constraints, and a growing mass of complexity in digital evidence. It focuses well on data security and privacy, as all victim-uploaded data must remain encrypted and stored.

Let us not overestimate this project. We should acknowledge that this is an innovative project that brings new approaches to digital forensics: automated triaging and advanced natural language processing techniques that fast track the collection, analysis, and reporting of digital evidence. The main target is to enhance speed, accuracy, and efficiency in forensic investigations whereby concerned evidence can be recognized and acted upon fast enough. With this system development based on a secure, victim-centric HTML-savvy platform, victims can collect evidence directly. Data integrity, security protocols, and victim-centeredness can be ensured.

Evidence will automatically be categorized and triaged for priority to the most relevant evidence that the investigators could have to save much time spent analyzing less critical priority or irrelevant data entries. Further optimization in this domain would be through the integrated BERT for automated report creation, where raw data input is converted to intelligible, structured reports highlighting key insights for investigators. This alone may save a considerable part of the traditional burden of report writing, allowing the forensic expert to turn his gaze back to the evidence instead of the time-consuming task of documentation. The project has automated the entire Evidence Collection Process through records and Case Reports. This one answers one of the main constraints on modern forensic investigations: case backlog, time constraints, and increasing complexity regarding digital evidence. Again, it focuses on data security and privacy; every data uploaded by the victim must remain encrypted and kept securely.

It is difficult to retrieve reliable data as it tends to shrink, while data validation techniques also help ensure the accuracy and consistency of the information, bringing the chances of errors that could tarnish the outcome of any investigation down to minimal levels. In short, this project tends towards creating a scalable solution against changing investigations in the field of digital forensics. It offers a framework that decreases the time needed in the investigative process, increasing the quality and reliability of the analysis. Indeed, the combination of automation, machine learning, and security features makes it a promising tool in today's cyber investigations, especially where it helps law enforcement and forensic experts solve cases faster and better.

Output:
Forensic Report

Incident Report Details	
ID No:	1
Incident Reported By:	MLA
Incident Title:	Network Exploitation
Report Date:	2024-09-18
Date & Time of Incident:	2024-09-04
IP Address(es):	8.8.8.8
Hostname(s):	Vegeta
Purpose of the System(s):	To Manage and Analyze Network Traffic
Operating System & Version:	Windows 11
Port of communication utilized:	80
Physical Location:	Coimbatore
Attack Vector utilized/exploited:	Exploited through Port Scanning
Additional Details:	Exploited the Weakest Link

Technical Details:
nmap ports, ports scanned, network ports, ports using, regarding ports
Immediate Actions:
Immediately Tracked the Exploiter by tracking his IP and found out where he was
Steps taken to prevent Recurrence:
Address tracker, tracker tracked, tracked analyzed, tracker, tracked
Description of attachments (if applicable, such as logs, reports, or screenshots):
None

Download as PDF

References

[1] Leopoldo Sebastian M. Gomez, "Triage in-Lab: Case Backlog Reduction with Forensic Digital Profiling," *Proceedings of the Argentine Conference on Informatics and Argentine Symposium on Computing and Law*, 2012. [Google Scholar]

[2] Fabio Marturana et al., " A Quantitative Approach to Triaging in Mobile Forensics," *2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, Changsha, China, pp. 582-588, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[3] Fabio Marturana, and Simone Tacconi, "A Machine Learning-based Triage Methodology for Automated Categorization of Digital Media," *Digital Investigation*, vol. 10, no. 2, pp. 193-204, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[4] Simson L. Garfinkel, "An Automated Solution to the Multiuser Carved Data Ascription Problem," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 868-882, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[5] Simson Garfinkel et al., "Bringing Science to Digital Forensics with Standardized Forensic Corpora," *Digital Investigation*, vol. 6, pp. S2-S11, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[6] Marcus K. Rogers, "Computer Forensics Field Triage Process Model," *Journal of Digital Forensics, Security and Law*, vol. 1, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[7] Fabio Marturana et al., "Mobile Forensics "Triaging": New Directions for Methodology," *Proceedings of VIII Conference of the Italian Chapter of the Association for Information Systems (ITAIS)*, 2011. [Publisher Link]

[8] W.A.J.J. Wiegerinck et al., "Approximate Inference for Medical Diagnosis," *Pattern Recognition Letters*, vol. 20, no. 11-13, pp. 1231-1239, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[9] Sadeghi Sarmad, Afsaneh Barzi, and Neda Zarrin-Khameh, "Automated Medical Decision Making Utilizing Bayesian Network Knowledge Domain Modeling," *Google Patents*, 2004. [Google Scholar]

[10] Antonio Grillo et al., "Fast User Classifying to Establish Forensic Analysis Priorities," *2009 Fifth International Conference on IT Security Incident Management and IT Forensics*, Stuttgart, Germany, pp. 69-77, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[11] Jingnian Chen et al., "Feature Selection for Text Classification with Naïve Bayes," *Expert Systems with Applications*, vol. 36, no. 3, part 1, pp. 5432-5435, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[12] C.H. Lee, F. Gutierrez, and D. Dou, "Calculating Feature Weights in Naive Bayes with Kullback-Leibler Measure," *2011 IEEE 11th International Conference on Data Mining*, Vancouver, BC, Canada, pp. 1146-1151, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[13] Usama M. Fayyad, and Keki B. Irani, "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning," *International Joint Conference on Artificial Intelligence*, vol. 93, no. 2, 1993. [Google Scholar] [Publisher Link]

[14] XRFF, Weka. [Online]. Available: http://weka.wikispaces.com/XRFF

[15] Ian H. Witten, Eibe Frank, and Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*, The Morgan Kaufmann Series in Data Management Systems, 2011. [Google Scholar] [Publisher Link]

[16] M.V. Zelkowitz, and D.R. Wallace, "Experimental Models for Validating Technology," *Computer*, vol. 31, no. 5, pp. 23-31, 1998. [CrossRef] [Google Scholar] [Publisher Link]

[17] Robert Kennedy, Reigning in Fully Autonomous 'Killer Robots', 2013. [Online]. Available: http://www.aljazeera.com/indepth/features/2013/04/201344132214594527.html

[18] Gary Cantrell et al., "Research toward a Partially-Automated, and Crime Specific Digital Triage Process Model," *Computer & Information Science*, vol. 5, no. 2, pp. 29-38, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[19] Karen Kent et al., "Guide to Integrating Forensic Techniques into Incident Response," *Nation al Institute of Standards and Technology*, 2006. [Google Scholar] [Publisher Link]