ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY (AAiT)

SCHOOL OF INFORMATION TECHNOLOGY AND ENGINEERING (SiTE)

A Structured Framework for Email Forensic Investigations

By Biruk Bekele Tadesse

Advisor: Henok Mulugeta (PHD), Professor (Assistant)

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN CYBER SECURITY WITH A SPECIALIZATION CYBER INTELLIGENCE AND WARFARE

Jan 2025

ADDIS ABABA, ETHIOPIA

ADDIS ABABA INSTITUTE OF TECHNOLOGY (AAiT)

SCHOOL OF INFORMATION TECHNOLOGY AND ENGINEERING (SiTE)

Approved by

Dr. _____Signature _____Date_____

Advisor

Dr. _____Signature _____Date_____

Internal Examiner

Dr. _____Signature _____Date_____

Internal Examiner

Dr. _____Signature _____Date_____

External Examiner

Abstract

Email forensics investigations become vital regarding legal, cybersecurity, and corporate challenges. However, most of the existing frameworks are suffering from inefficiency problems, data integrity, and handling such diverse data sources with complexity, considering encrypted emails and metadata. This thesis applied the Design Science Methodology to develop a structured framework that enhanced efficiency and effectiveness in email forensic investigations. These specifically deal with data quality, diversity in data management, and integrity of evidence. Among others, one key component is case management, which systemizes and keeps track of the investigation from the very outset to the last step in an appropriate manner and ensures each step is conducted methodically. The framework comprises key phases: case management, governance, identification, preservation, classification, analysis, presentation and compliance that address critical challenges such as ensuring data quality, managing diverse data sources, and maintaining evidence integrity. Case management forms the core part of the proposed framework for organizing, tracking the investigation process from start to finish in order ensuring that evidence is handled properly, and all phases are executed in a systematic manner. It integrates open-source tools, case studies of different varieties, and best practices to be relevant to different real-world scenarios. The effectiveness of the artifact can also be demonstrated in practical application, performance being measured in terms of speed of investigation, data quality, accuracy, and user satisfaction, among other metrics. This research underscores that the suggested framework decreases the time of investigation, reduces the rate of errors, increases the quality of data management, and guarantees the effective access of various data sources. This thesis contributes on both practical and theoretical levels, guiding practitioners and researchers comprehensively in the area of digital forensics to bring current email forensic investigations into a more efficient, accountable, and adaptable condition.

Key word: Email, Email Related Crimes, E-mail Forensic Investigation, Email Investigation Framework

Acknowledgement

I want to express my profound gratitude to Henok Mulugeta (PHD), my mentor and advisor, for his excellent, helpful criticism, ongoing direction, and tremendous assistance in making this research a success.

INSA, Federal Police and their staff, particularly those in the Department of Digital Forensic, deserve special recognition for their readiness to take part in the study. Abel Feleke, the manager of the Digital Forensic team, deserves special recognition for his excellent coordination of the data collection process.

Without the help of my wife and her family, I could not have finished this project. I am grateful to them for providing additional room for writing and research.

I also want to express my gratitude to my family, who have supported me during this entire study process. The accomplishment of this research would not have been possible without their constant support, encouragement, and direction.

Lastly, I want to express my gratitude to my friends and coworkers for their support, encouragement, and helpful criticism. Throughout our investigation, their contributions have served as a source of motivation and inspiration.

Table of Contents

Table

Figure

Abbreviations and acronyms

ACPO	Association of Chief Police Officers
DS	Design Science
Email	Electronic mail
ENVID	Envelope Identifier
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
MIME	Multipurpose Internet Mail Extensions
NIST	National Institute of Standards and Technology
OTP	One Time Password
PST	Personal Storage Table
S/MIME	Secure Multipurpose Internet Mail Extensions
SMTP	Simple Mail Transfer Protocol
SQL	Structured Query Language
SOS	Morse code distress signal

CHAPTER ONE

1. Introduction

1.1 Background Information

Email has become an imperative in contemporary communication and a significant medium of personal, commercial, and government transactions. It is imperative in all media where it has been applied for such crucial activities as commercial undertaking, scholarly exchange, social interaction, and dissemination of confidential information. The application of email as a mode of communication has increased immensely with the universal expansion of internet connectivity, and now it forms an integral component of everyday undertakings [1]. Email communication was created to be simple, efficient, and influential, and it can be used for almost anything nowadays, including business, study, connecting to social media, and communicating with people who do not use social media. Email has become a primary communication channel for many official tasks due to the fast growth of internet use around the world. Not just businesses, but also individuals, utilize email for vital business tasks such as banking, sharing official messages, and sharing confidential files. Email communicates via computer networks, principally the Internet, as well as local area networks. [2]

The store and forward concept are used in today's email systems. Messages are accepted, forwarded, delivered, and stored by email servers. Users and computers are not required to be online at the same time; they must connect, often to a mail server or a webmail interface, to send or receive messages or download content [3]. E-mail systems are made up of numerous hardware and software components, such as sender's client and server computers and receiver's client and server computers, each with the necessary software and services installed. Aside from them, it makes use of many Internet systems and services. Both sending and receiving servers are always connected to the Internet, while the sender and receiver's clients only connect when necessary. [4]

However, this kind of communication has become vulnerable to cyber-attacks. Email continues to be the most popular route for opportunistic and targeted attacks. To specified employees within the organization, the adversary sent a targeted email with a malicious attachment that appeared to come from a trusted source. Cyber forensics is the collection and analysis of data from computer systems, networks, wired or wireless communication streams, and storage media using scientifically validated methodologies in a court of law. [5]

Email forensic investigations face several significant challenges that complicate the process of gathering and analyzing evidence. One primary issue is data complexity; modern email systems often utilize encryption methods such as PGP and S/MIME, which enhance security but make it difficult for investigators to access the content without the appropriate cryptographic keys. This requirement can delay investigations and hinder the retrieval of crucial evidence. Additionally, the sheer volume of metadata generated by emails poses another challenge. While metadata, including timestamps and routing information, is essential for establishing timelines and context, its unstructured nature can overwhelm investigators, complicating analysis and interpretation.

Legal compliance is also a major concern. Investigators must navigate various regulations, to ensure that the evidence collected is admissible in court. Many existing forensic frameworks lack robust governance protocols, increasing the risk of evidence being ruled inadmissible due to procedural errors. Moreover, the inefficiency of current forensic processes often relies on manual methods that are time-consuming and prone to error. Federal bureau investigation, Internet Crime Reports indicate In 2023, they received 21,489 business email compromise complaints with adjusted losses over 2.9 billion, It is a sophisticated scam targeting both businesses and individuals performing transfers of funds. The scam is frequently carried out when a subject compromises legitimate business email accounts through social engineering or computer intrusion techniques to conduct unauthorized transfers of funds. [6]

Depending on the industry in which digital inquiry is required, there are various types of computer forensics. E-mail forensic analysis is a sort of cyber forensic analysis that is used to investigate the source and content of an e-mail message as evidence, identifying the true sender, receiver, and timestamp in order to gather credible evidence to bring offenders to justice. As a result, a forensic investigator needs efficient tools and methodologies to execute the analysis with high accuracy and in a timely manner. [6]

1.2 Motivation of the Study

Email forensics is a growing field in digital investigations, the email has gained prime importance, and therefore the limitation of the already existing systems in meeting the challenge that exists nowadays was one of the main reasons why we decided to conduct this study. Emails have become one of the means of communication, most of the time they are involved in cases of fraud, cybercrime, and corporate disputes. The proliferation of email frauds such as phishing, business email compromise, and invoice scams uncover the urgent need for proper forensic methodologies. However, the current forensic approaches are inefficient, not comprehensive; adaptability can have good results Yet Few forensic systems are able to manage the complexities of email evidence efficiently, which include different data formats, encryption, metadata analysis, and the jurisdictional issues.

Those issues render it essential to setup an overall framework that offers forensic professionals the path to information accuracy, data integrity, and legal compliance. Moreover, there is also very high pressure for platforms or rather tools that are open source and can be applied in many real situations, which must support a large number of investigation habits.

One of the reasons that also supports the research of this study is the necessity of a better actually management process which ensures that forensic investigations are transparent, well documented, and legally defensible. By the implementation of a structured framework designed with the Design Science Methodology, this research is anticipated to fill the existing gap and this thesis applied the Design Science Methodology to develop a structured framework that enhanced efficiency and effectiveness in email forensic investigations. Additionally, it extended the body of knowledge by developing a practical solution that complemented the effectiveness and reliability of these investigations. In the final analysis, grappling with these motivations enabled investigators to better combat the burgeoning menace of email fraud.

1.3 Problem statement

Data breaches, fraud, and harassment are among the cybercrimes that have significantly increased because of the quick development of digital communication, particularly via email. Effective forensic techniques are now essential since email is still a major medium for both personal and professional communications [7]. However, most forensic methods used today lack this organized framework for the systematic investigation of email-related problems, creating disjointedness and inconsistent approaches that eventually undermine the effectiveness of investigations. Data governance is a big challenge; most organizations in investigations fail to manage the integrity, confidentiality, and compliance of the email data, probably leading to legal consequences. Other

challenges include the very labor-intensive process of ingestion, which does not effectively collect data around emails coming from several sources like servers, personal devices, and cloud platforms. The result is incomplete data sets, leading to difficulties in deeper analyses. Moreover, such a huge number of emails contributes to the complication in data processing because large datasets may be hard for traditional methods to sift, filter, and arrange efficiently, hence making it hard for investigators to find pertinent patterns or abnormalities.

Without case management utilities that centralize the investigations, an investigator must systematically detail all activities, record his findings, and attempt to coordinate the effort. Obviously, not having centralized case management exacerbates disorganization and inefficiency. Equally important, machine learning methods have barely been applied to e-mail forensics yet; though these methods are bound to greatly extend what might be done in investigations, they are absent in the application of most current frameworks. It is easy to underestimate the value of open-source tools in email forensics investigations. These applications can provide flexible, reasonably priced solutions that foster creativity and cooperation between forensic practitioners. Investigators can practice their skills and create effective methods for performing email forensics using these tools. Linked to e-mails and do not take advantage of sophisticated technological tools.

The lack of such a holistic framework that covers data governance, ingestion, and processing, case management, machine learning involved, and the usage of open-source tools in an effective manner is what keeps forensic investigators from handling the complexities introduced by email-related incidents. This may lead to more legal consequences and risks for an organization. This weakness in the current processes is probably going to result in investigations being unsuccessful and not able to recover vital evidence.

It thus intends to provide a codified set of best practices in key areas, including the use of open-source tools that offered a structured approach toward improving the rigour and reliability of email forensic investigations. Ultimately, it would be able to respond effectively to the threats posed by cybercrimes and enhance the ability of organizations to resist such email-based crimes.

1.4 Research Questions

- How can machine-learning techniques be effectively integrated into email forensic investigations?
- How can data quality and integrity be maintained across diverse data sources and formats within an email forensic investigation?
- What role do open-source tools play in supporting email forensic investigations?

1.5 Objective of the Study

The general objective of the research was developed A Structured Framework for Email Forensic Investigations. To successfully address these requirements, the following specific objectives are also considered.

Specific objectives.

- To develop a structured framework that enhances the accuracy, efficiency, and legal compliance of email forensic investigations.
- To identify and integrate effective tools, techniques, and best practices for analyzing email data, particularly in legal and regulatory contexts.
- To establish strategies for maintaining data quality and integrity across diverse data sources and formats.
- To assess the performance evaluation and validation procedures for the proposed framework

1.6 Expected Contribution of the Study

This study is expected to make significant contributions to the field of email forensics and digital investigations by:

- Framework Development: The framework will offer a systematic approach that improves the accuracy, efficiency, and legal validity of email forensic investigations.
- Access to Tools and Methods: It will ensure that forensic investigators have recommended tools and techniques available for analyzing evidence, especially in legal and regulatory situations.
- Data Quality and Integrity Standards: The framework will set forth standardized practices for maintaining data quality and integrity, which are crucial for preserving email evidence from various sources and formats.
- Accountability and Transparency: It will give a well-documented process to ensure accountability by maintaining the chain of custody and making the investigations resistant to legal challenges.
- Accessibility of Open-Source technologies: The framework will make email forensics more affordable and accessible by focusing on open-source technologies, particularly for enterprises that have tight resources.
- The framework will integrate lessons from the case studies and become a useful tool for both scholars and forensic practitioners. This will further develop the theory by providing

useful advice that can be used in the contexts of both scholarly research and real-world situations.

1.7 Scope/delimitation

The study focused on the various techniques and procedures of email forensics, with the goal of coming up with a structured framework that would enhance rigor and dependability in such processes. The study also touched on legal and regulatory issues regarding email forensics and the suitability and availability of open-source tools for investigators to ensure compliance with relevant laws and regulations. It did not provide the comprehensive analysis of all the existing tools, especially commercial software, and alternative digital channels of communication.

1.8 Structure of the Document

This paper is organized in five chapters. The first chapter is the introduction part of the study and it includes background of the study, statement of the problem, research questions, objective, significance, scope, limitation and finally structure of the document. The second chapter is about review of literature, these literatures are important and bases for the research as a whole in developing theoretical framework and deeper understanding regarding the subject. The third chapter about research design and methodology, which includes research design, population and sampling techniques, types of data and tools which are used for data collection, method of data analysis and finally ethical consideration. The fourth chapter includes data presentation, analysis and interpretation section and the last chapter deals with summary, conclusion and recommendation of the study

CHAPTER TWO

2. Literature review

2.1 Literature review

2.1.1 Email

E-mail systems are made up of various hardware and software components, such as sender's client and server computers and receiver's client and server computers, each with the necessary software and services installed. Aside from these, it makes use of various Internet systems and services. The sending and receiving servers are always connected to the Internet, but the sender and receiver's clients only connect when necessary. [3]

E-mails are composed of two major components: the message header and the message body. The header section contains e-mail routing information as well as other information such as the e-source mails and destination, the sender's IP address, and time-related information. The actual message of the email message subject and body is contained in the message body. The body may also include MIME or S/MIME (Secure/MIME) attachments. [8]

2.1.2 Email Related Crimes

E-mail has become the most popular form of communication in recent years. Every day, a million e-mails are sent around the world. This unprotected transfer of emails from one source to another opens the door for Delinquent minds. A person who is likely to commit a crime sees email as a convenient and quick tool. One may be unaware of the various crimes committed via e-mailing platforms. Some of the major email related crimes are:

I. EMAIL SPOOFING

A spoofed email is one that appears to originate from one source but actually emerged from a distinct source. It is done by using a fake name and/or e-mail address. Usually, the email is sent using the identity of the original or desired sender of the e-mail that the victim feels safe to access. Certain web-based email services like www.SendFakeMail.com, offer a facility wherein a sender can enter the email address of the purported sender of the email. A person can use such services to send viruses, trojans, worms, etc. to persons who would unknowingly download them. [8]

II. THREATENING EMAIL

Anyone with basic knowledge of computers can easily become a blackmailer in order to threaten someone via e-mail. Various incidents like texting i.e., sending persistent text

messages, and sexting i.e., sending sexually explicit photographs/ MMS have emerged as a major cyber offence faced by the victims.

III. EMAIL BOMBINGS

Sending a large number of emails to someone that ultimately crashes the receiver's email account, irrespective of its nature is known as e-mail bombing. Terrorism has hit the internet in the form of mail bombings. The shutdown of the entire system of the victim of cyber-attack leads to the destruction of information. [10]Email bombing, which involves sending a huge number of mails to an individual or organization with the purpose to overwhelm or disrupt, is a kind of harassment that can inflict serious harm.

IV. DEFAMATORY EMAILS

Defamation means, the intentional publication of some false information or statement about someone that can demean or injure his/her reputation in society. When someone sends e-mails containing defamatory information about someone, it would amount to cyber defamation. [8] Defamatory emails can be distributed to numerous sources, either by accident or on design, resulting in a problematic situation as an unintended consequence.

Defamation needs publishing, such as an email sent from a confirmed email address. This implies that when you write a defamatory comment about someone, you authorize the publishing of that email.

V. EMAIL FRAUDS

Hacking of email account using various tools is referred to as email fraud. It can be done by sending spoofed emails, sending spam emails, or attaching emails with malware embedded, or hacking the email using OTP in case of two-factor authentication. Once hacked, email can be used to send SOS emails to the contacts of the victim, sending offensive messages to clients in case of business in order to destroy the reputation, or having unauthorized access to mail to gain access to other accounts like social media, net banking, etc. [9]

VI. SENDING MALICIOUS CODES THROUGH EMAIL

A code that leads to malware or a virus on a computer can be said to be malicious code. It can also create various issues like destroying or crippling various valuable property of the user and making the computer or the set of computers vulnerable to other malware attacks. It can disrupt every connection that the victim makes through the affected computer [10]. Malicious email attachments are files sent with emails that are intended to compromise or harm the recipient's computer system or exfiltrate sensitive information. These harmful payloads can disguise themselves as innocuous items documents, PDFs, images, or audio files but when opened, they

unleash malware, such as ransomware which locks data access until a ransom is paid, spyware which stealthily collects and transmits personal information without consent, or viruses which corrupt systems and spread to other devices.

VII. PHISHING

Phishing scams utilize email spoofing or instant messaging to trick people into entering information on a bogus website that closely resembles the authentic one. Phishing is a social engineering approach that deceives users. Phishing is a type of cybercrime where criminals attempt to trick individuals into revealing sensitive information, such as login credentials or financial information, by impersonating a legitimate organization or individual. [10] Phishing attacks often come in the form of fraudulent emails, text messages, or social media messages that appear to be from a trusted source like a bank, government agency, or company. The goal is to get the recipient to click on a malicious link or attachment, which can then install malware or direct them to a fake website to enter their information. Common phishing tactics include creating a sense of urgency, exploiting current events or news, and using logos/branding to appear legitimate. Preventing phishing involves being cautious of unsolicited messages, verifying the source before clicking links or attachments, and using security software and strong passwords. Reporting phishing attempts to the proper authorities can help combat this growing cybersecurity threat.

2.1.3 E-mail Forensic Investigation

E-mail forensics refers to the study of source and content of e-mail as evidence to identify the actual sender and recipient of a message, data/time of transmission, detailed record of e-mail transaction, intent of the sender, etc. This study involves investigation of metadata, keyword searching, port scanning, etc. for authorship attribution and identification of e-mail scams. Various approaches that are used for e-mail forensic are described in include header analysis, bait tactics, server investigations, and network device investigation. Besides mandatory headers, custom and MIME headers appearing in the body of the message are also analyzed for sender mailer fingerprints and software embedded identifiers [11]

Accordingly, to [8] E-mail forensics refers to the study of email details including: source and content of e-mail, in order to identify the actual sender and recipient of a message, date/time of transmission, detailed record of e-mail transaction as well as the intent of the sender. Therefore, e-mail forensic investigation often involves analysis of metadata, keyword searching as well as port scanning, for authorship attribution and identification of cyber-crime. An email has header and body. The header part includes many important information such as sender’s IP

Address, mail user agents, and servers in transit, message id field, and signatures field. The Received field indicates the date and time when the email was received at the server from and to represent the sender and recipient, respectively. [10]Identities used in E-mail are globally unique and are mailbox, domain name, message-ID and ENVID. Mailboxes are conceptual enteritis identified by e-mail address and receive mail. Email address has become a common identity identifier on the Internet

Digital forensic e-mail analysis, particularly header analysis, is used to gather credible evidence to bring criminals to justice. A detailed header analysis can be used to map the networks that messages travel through. If there is multiple received field information, the origin must be tracked from bottom to top. Received field, with the bottom received address representing the original sender's IP address and the top received address representing the actual receiver's IP address. [13]

E-mail forensics can examine both the email header and body. The first examination is a header examination investigation, which should include the Examining sender's e-mail address, message initiation protocol (HTTP, SMTP), Message ID, and sender's IP address. The second examination is a body examination investigation. Email storage format, Maildir (each email is kept separate in a file for each user), and mbox format are examples of server-side storage formats (all email files are in a single text file). Email is stored on the server in SQL Server databases. Reading various formats for forensics analysis can be accomplished using notepad editor and regular expression-based searches. An email is stored as mbox format (Thunderbird) on the client side. Emails can also be stored on the client side as.PST (MSOutlook) and NSF (Lotus Notes) files. The third point to consider is the availability of an email backup copy. All copies are transferred to the client when checking from the serve side. This necessitates the seizing of the client computer. Webmail copies are always saved on the server. The last one is Protocol, which is used to transport email. Email can be initiated and transported using SMTP or HTTP depending on the email server applications. [12]

2.1.3.1 E-MAIL FORENSIC INVESTIGATION TECHNIQUES

Preservation of Evidence

Before starting any investigation, it is crucial to preserve the integrity of the evidence. Make sure to create a backup or forensic image of the entire email system to prevent any alterations or loss of data.

Metadata Analysis:

Email metadata is a crucial component in the analysis and investigation of email-based evidence. The metadata associated with email communications can indeed provide invaluable insights that help establish the authenticity and origin of the emails under examination. Let me expand on the importance of email metadata analysis within the proposed email forensic investigation framework.

Sender and Recipient Addresses:

Examining the email header information, including the "From," "To," "Cc," and "Bcc" fields, can reveal crucial details about the parties involved in the email communication. Anomalies in the sender or recipient addresses, such as spoofed or forged addresses, can be indicators of fraudulent or malicious activities.

Timestamps:

The date and time stamps associated with email messages, including the "Date" header and any additional time-related metadata, can help establish the chronology of events. Analyzing the temporal patterns and discrepancies in email timestamps can aid in identifying inconsistencies or potential attempts to obscure the true timing of the communications.

IP Addresses:

The email header often contains information about the IP addresses involved in the email's transmission, such as the "Received" headers. Investigating the IP addresses can help trace the email's origination and route, potentially uncovering connections to suspicious or known malicious entities.

Email Header Examination:

The proposed email forensic investigation framework includes an analysis of email headers, which are information pertaining to the delivery path of an email. It is from these details that essential insights to anchor the authenticity and integrity of email evidence can be obtained. [14] Under the framework, the investigation of the email headers forms a key component in the data ingestion and analysis phases. By closely scrutinizing the header information, forensic investigators can find valuable information about the path or journey that an email takes, such as: **IP Addresses and Domains**: It contains IP addresses and domain names of every server that moved the transmission of an email forward. Investigation of these IP addresses and domains will serve to corroborate the origins, find potential connections to known malicious entities, and detect efforts to disguise the real point of origin for this communication [13]

Time Stamps: The time within the "Received" header lines can be cross-referenced to establish a timeline of the email's delivery that would be important in pointing out inconsistencies or attempts to manipulate the timing of the communication. [14]

E-mail Client Information: Certain email headers contain information from "X-Mailer" or "User-Agent" fields identifying the client software generating an e-mail. The nature of this information helps in knowing the technological level a sender possesses, as it often comes useful to look out for automatic or harmful generations of e-mails. Thus, the proposed framework provides for the forensic investigator to verify the delivery path of the email in question and any anomalies or inconsistencies regarding potential tampering or spoofing, establish the timeline and sequence of events surrounding the email communication to uncover connections between the email and known malicious actors or infrastructure. [15] This comprehensive header analysis coupled with other investigative techniques within the framework significantly enhances the ability of forensic professionals to effectively evaluate the authenticity and evidentiary value of email-based information.

Recovering Deleted Emails:

These deleted emails can further be retrieved from mail servers, local repositories, or backups with the help of relevant tools and techniques. Extract relevant information from the deleted emails retrieved using appropriate tools and techniques. Recovery and analysis of deleted emails is one of the key elements in the proposed framework for email forensics investigation. Most deleted emails often contain critical information for establishing timelines of events, pinpointing questionable activities, and corroborating or refuting claims of the parties involved. The framework is therefore fitted with strong techniques and tools that retrieve and analyze deleted email data from any given source. [15]

Backup and Archive Analysis:

The framework accesses different backup and archiving solutions that identify on-premises and cloud-based possibilities of accessing deleted email data that might have been retained in these secondary storage systems. This includes the analysis of backup files, network-attached storage, and cloud backup services, the recovered deleted email data is integrated into the overall email forensic investigation to provide an in-depth understanding of the history of email communications that may contain evidence. [13]

Data Carving and Reconstruction:

Where the erased email data is not immediately available through server or client-side sources, the framework uses advanced data carving and reconstruction algorithms. Specialized open-

source forensic programs search the available storage media for pieces of erased email data that can be rebuilt and examined. This data carving technique could expand the scope of the investigation by recovering key information in partly deleted or fragmented email data. [3] In email, forensic investigation enhances the capability of the investigators in identifying important evidence that may be overlooked or deliberately destroyed with the incorporation of these exhaustive techniques in recovering and analyzing deleted email data. It is expected that recovered deleted emails would provide vital pieces of information on timeline, communication patterns, and an overall context of the incident under investigation, which could provide credibility and evidential weight for forensic findings.

Link and Attachment Analysis:

Carefully examining links and attachments sent via email can help identify potential malware or phishing attempts. Utilize sandbox environments or other secure means to analyze suspicious files without compromising your system. [17]

Network Device Investigation

This kind of email research uses the logs kept by network hardware like switches, routers, and firewalls to look into the origin of an email message. This type of investigation is complicated and is only utilized when the logs of servers (Proxy or ISP) are unavailable for some reason, such as when an ISP or proxy fails to preserve a log, when ISPs refuse to cooperate, or when the chain of evidence is not maintained. [2]

2.1.4 Email Investigation Frame work

Email forensics, as a sub-discipline of digital forensics, has undergone significant growth, but most of the existing frameworks and methodologies do not address the specific challenges related to email data. Several standards, including the NIST Special Publication 800-86, present a general framework for conducting digital investigations anchored on five phases that may assist an organization in responding effectively to cybersecurity incidents. The first part-Identification-consists of the potential identification of incidents through alarms, anomalies, or user reports. This step is very important because early identification can reduce damages and preserve the capability of quick intervention. The second part-Preservation-deals with securing the incident scene to prevent alterations of data, including isolating the affected systems and preserving logs and data in order to maintain evidence integrity.

Thirdly, Collection: This involves the systematic gathering of data, logs, and files, and system memory in general, in a very structured way, ensuring no relevant data is left out. Next, Examination: The evidence collected should be used with forensic tools to uncover hidden data

and recreate timelines of events associated with the incident. Finally, the result of the examination is interpreted into meaningful data during the Analysis phase, which provides insight into the tactics, techniques, and procedures of the attacker, feeding into mitigation strategies and recovery efforts to strengthen defenses against future incidents. [13]]. However, these guidelines do not go deeply into the methodologies specific to email and hence leave many gaps in practice.

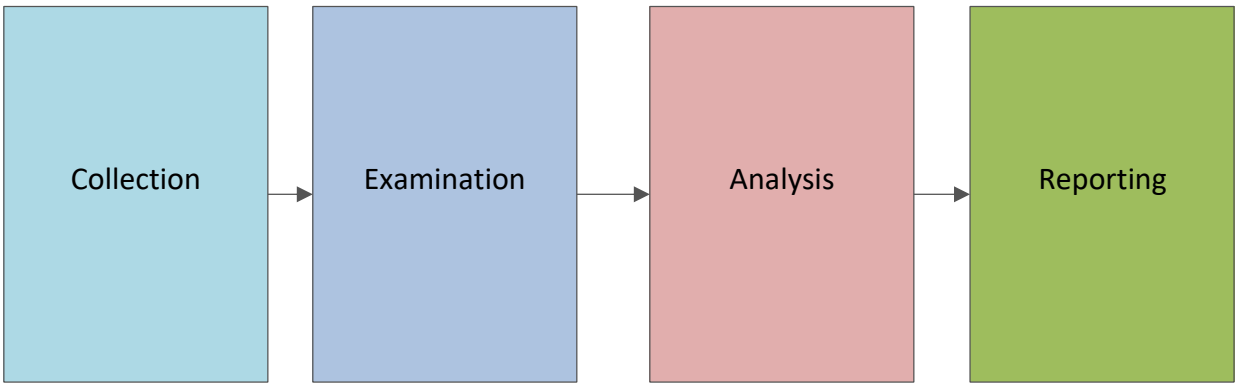


Figure 1 NIST framework

In the same way, the ISO/IEC 27037 standard outlines best practices meant for identifying, collecting, and preserving digital evidence but falls short in addressing nuances of email forensics, especially with regard to metadata and encryption challenges [14]. Although the guidelines laid down by the Association of Chief Police Officers emphasize the legal admissibility of digital evidence, they are bereft of technical details that would be required to explain e-mail communications [15].

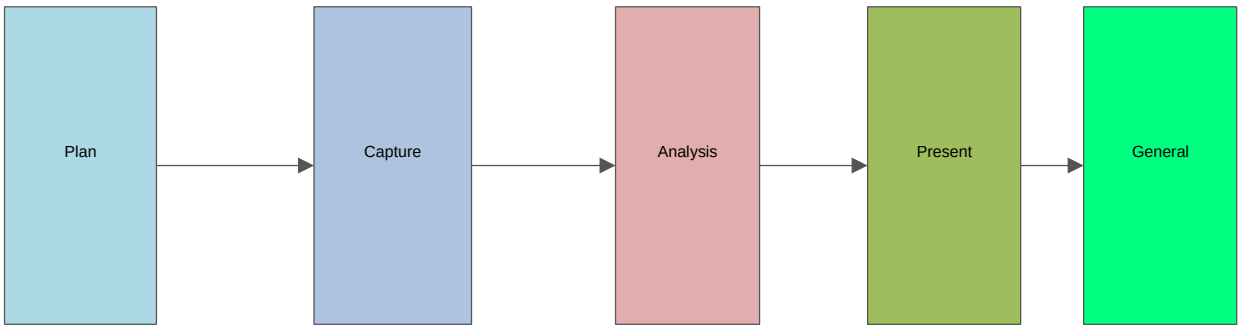


Figure 2 Association of Chief Police Officer

2.2 Related works

Accordingly, M .Taric Brandy [4] did a study on Techniques and tools for forensic investigation of email. The aim of their study was to illustrate e-mail architecture from forensics perspective. It describes roles and responsibilities of different e-mail actors and components, itemizes Meta data contained in e-mail headers, and lists protocols and ports used in it. It further describes various tools and techniques currently employed to carry out forensic investigation of an e- mail message. Another study done by Vamshee Krishna Devendran, [12] on A Comparative Study of Email Forensic Tools. Their research focused on comparing and contrasting five popular open source email forensic tools based on a set of common features. According to their findings, not all email forensic tools are the same and provide different types of functionalities. They are convinced combining analysis tools may allow for the acquisition of detailed information in the field of email forensics.

Rachid Hadjidj, M[21] worked on the project towards an integrated e-mail forensic analysis framework. The primary in this paper, they present an e-mail forensic analysis software tool that they created by combining existing cutting-edge statistical and machine learning techniques with social networking techniques, and they designed and implemented a comprehensive software toolkit called the Integrated E-mail Forensic Analysis Framework.

According to Charalambous Elisavet, [6] they did a paper on Email forensic tools: A roadmap to email header analysis through a cybercrime use case. According to their study, review existing email forensic tools for email header analysis, as part of email investigation, with emphasis on aspects related to online crime while still considering legal constraints. Through their analysis, they investigate a common case of cybercrime and examine the breadth of information one may gain solely through email forensics analysis. Additionally, in this paper presented a roadmap for email forensic analysis, combining features and functionality already available, to assist the process of digital forensic analysis.

Accordingly, to Ahmad Ghafarian, Ash Mady and Kyung Park [2] focus on an empirical analysis of email forensics tools. In this research, they experimentally compare the performance of several email forensics tools and their aim is to help the investigators with the tool selection task, evaluate the tools in terms of their keyword search, report generation, and other features such as, email format, size of the file accepted, whether they work online or offline, format of the reports, and also, they use Enron email dataset for their experiment

[19]Introduced a Digital Forensic Data Reduction Framework, which significantly reduced storage requirements while maintaining the integrity of critical evidence. This work demonstrates how rapid triage methodologies can improve efficiency but is less focused on email-specific

investigations. [20] Explored the application of machine learning and data mining techniques for anomaly detection in digital forensics. While effective in general forensic scenarios, their framework does not explicitly address email-specific challenges such as header analysis or attachment extraction. [21] Emphasized the role of proper data governance to ensure compliance with legal and regulatory standards. They argue that robust governance policies enhance the credibility and admissibility of digital evidence in court.

Table 1 summery of literature review

Title	Author and Citation	Contribution	Method	Limitation
Comparative Study of Email Forensic Tools	VamsheeKrishna Devendran, HossainShahriar, Victor Clincy [12]	This paper Comparing and contrasting five popular open source email forensic tools based on a set of common features. According to their findings, not all email forensic tools are the same and provide different types of functionalities. They are convinced combining analysis tools may allow for the acquisition of detailed information in the field of email forensics.	compare email forensic tools based on a set of desired attributes	Doesn't provide Framework
Techniques and tools for forensic investigation of email	M.TaricBrandy [3])	This article discusses illustrate e-mail architecture from forensics perspective. It describes roles and responsibilities of different e-mail actors and components, itemizes Meta data contained in e-mail headers, and lists protocols and ports used in it. It further describes various tools and techniques currently employed to carry out forensic investigation of	Content Analysis	Discusses the benefits of automation but does not provide Framework

Towards an Integrated mail forensic analysis framework,	RachidHadjidj, Mourad Debbabi, Hakim Lounis, Farkhund Iqbal, AdamSzporer, and Djamel Benredjem [14]	This article explores an e-mail forensic analysis software tool that they created by combining existing cutting-edge statistical and machine learning techniques with social networking techniques, and they designed and implemented a comprehensive software toolkit called the Integrated E-mail Forensic Analysis Framework	statistical analysis, text mining (classification and clustering), and stylometric features analysis	The study Developed by integrating existing state-of-the-art statistical and machine learning techniques complemented with social networking techniques only a single framework's
Email forensic tools:A roadmapto email header analysis througha cybercrime use case	Charalambous Elisavet,Bratskas Romaios, Koutras Nikolaos,Karkas George Anastasiades Andreas [8]	This paper review existing email forensic tools for email header analysis, as part of email investigation, with emphasis on aspects related to online crime while still considering legal constraints. And, presented a roadmap for email forensic analysis, combining features and functionality already available, to assist the process of digital forensic analysis.	review	Doesn't provide framework and they used review method

an empirical analysis of email forensics tools	Ahmad Ghafarian, Ash Mady and Kyung Park [15]	This article experimentally compare the performance of several email forensics tools and their aim is to help the investigators with the tool selection task, evaluate the tools in terms of attribute	experimental	The study focuses on tools comparison, not the email forensic Framework
NIST Special Publication 800-86	NIST	Provides guidelines for digital forensic investigations across various phases.	Tailored approach	Lacks specificity in methodologies tailored to email forensics, such as handling encrypted emails and analyzing metadata.
ACPO Guidelines	ACPO	Emphasizes legal admissibility of digital evidence and appropriate handling procedures.	Tailored approach	Primarily focuses on general digital evidence without specific protocols for email forensics, such as handling attachments or complex communication threads.

Digital Forensics Framework (DFF)		Systematic approach to digital forensic investigations.	Tailored approach	Does not provide tailored methodologies for the unique intricacies of email communications.
Handbook of Digital Forensics and Investigation	Casey (2011)	Highlights the importance of metadata analysis in email investigations and methods for extracting it	Tailored approach	Limited discussion on handling encrypted metadata and the evolving complexities of
Digital Forensics and Cyber Crime	Garfinkel (2013)	Discusses extracting and analyzing metadata, particularly for detecting email tampering		Discusses extracting and analyzing metadata, particularly for detecting email
Forensic analysis of digital evidence	Kessler (2018)	Reviews recovery techniques for deleted emails across various platforms.	Tailored approach	Effectiveness can vary significantly depending on the email system's data retention policies,
A cloud-based framework for digital forensics investigation	Guo et al. (2019)	Proposes using social network analysis to identify insider threats through email communications.	Tailored approach	Requires advanced algorithms and tools to analyze large datasets effectively, which may not be readily

Forensic investigation techniques for email communications	Rahman et al. (2020)	Explores the application of artificial intelligence in automating phishing detection and categorizing emails.		Integration into traditional forensic processes is still in its infancy, limiting practical applications at this
Frameworks and methodologies for email forensic analysis	Venter & Eloff (2020)	Provides practical guidance on forensic investigations specifically targeting email evidence.	Tailored approach	Tools discussed may struggle with encrypted content, necessitating further development of

Table 1 literature review summarization

2.3 Research Gaps

The current body of research reveals several limitations in existing email forensic methodologies, including

- The current body of research indicates several shortcomings in the existing email forensic methodologies, including the following:
- Most of the case management systems presently lack comprehensive frameworks that assure high data quality, thus making the resultant decisions suffer from issues such as accuracy, consistency, and timeliness.
- The rising volume, variety, and velocity of big data are of great challenge to traditional case management approaches, which usually cannot handle such complexities.
- There is a need for effective strategies and tools that integrate all types of data-structured, semi-structured, and unstructured-within case management systems for complete analysis.
- The literature is poor in what pertains to best practices concerning data quality in case management.
- There is also a lack of research into the benefits of using open-source software for email Investigations in terms of cost-effectiveness, flexibility, and customization.

Therefore: This research aims to fill these gaps by:

- by appraising a holistic framework for data governance that captures the challenges of data integration. This new framework addresses data quality, consistency, and

compliance, ensuring the key dimensions, which were largely overlooked by many methodologies.

- Evaluates different tools and techniques for data integration to establish their suitability for large volumes of data, which will address the knowledge gap with respect to the suitability of various tools to integrate efficiently a number of sources.
- The strategies proposed to handle high data volume relate to architecture scalability and scalable data processing techniques. In fact, such specific guidance was lacking in current literature on the handling complexities arising due to large volumes of data.
- Integrates consideration for legal and regulatory compliance of data governance practices in the process, which ensures that data integration efforts are non-conflicting with applicable laws and standards. This is usually absent or underrepresented in previous literature.

.

CHAPTER THREE

3 Methodology

3.1 Research Approach

This study adopted a design science research methodology (DSRM) to develop an effective and efficient email forensic investigation framework using open-source tools. The DSRM is a well-established research approach in the field of information systems that focuses on the creation of innovative artifacts to address real-world problems. [23] The DSRM provides a structured process for

Table 2 Design Science Research Methodology

Research Step	Concerns	Output to Next Step	Entry Point?
Identify Problem & Motivate	Define problem Show importance	Inference	Problem-Centered Initiation
Define Objectives of a Solution	What would a better artefact accomplish?	Theory	Objective-Centered Initiation
Design & Development	Artifact	How-to Knowledge	Design & Development
Demonstration	Find suitable context Use artefact to solve	Metrics, Analysis	Client/Context Initiated
Evaluation	Observe how effective, efficient	Disciplinary Knowledge	
Communication	Scholarly publications Professional publications		

Table 3design science Methodology

A. Identify problem and motivation.

The exact limitations in the existing approaches will be determined at this first stage when the research group carries out a comprehensive literature review and liaises with the experts in email forensics. Among the aspects to be covered are a lack of sufficient frameworks and that many case management systems do not have a structured approach, which tends to guarantee a high quality of data leading to the decisions to be full of inaccuracies. Besides, big data challenges

feature increases in volume, variety, and velocity of information, which creates serious obstacles for classic forensic approaches often unable to handle such complex datasets. There is also an urgent need for strategies and tools that allow the integration of different types of data, including structured, semi-structured, and unstructured data in case management systems. The literature also shows insufficient best practices related to the entire forensic process in maintaining data quality, hence leaving practitioners in a situation where there is no clear guidance. Finally, there is a lack of underexplored open-source solutions; limited research is available on how such tools can enhance email investigations in terms of cost-effectiveness and customization. Dealing with these issues is relevant for the advancement of the field and improving the effectiveness of methodologies related to email forensics

This research aimed to develop an effective and efficient email investigation framework. First, the researcher studied the literature and the real techniques used for email investigation. This was done through a study of many previous related works concerning security in various fields (integrity and confidentiality). Then, the researcher suggested which open-source tool for email investigation to work on. The research proposed the necessary tools and techniques to adopt the resulting model. Finally, an effective and efficient email forensic framework was developed.

B. Define objectives

The clear objectives of the designed artifact will be articulated to ensure that they address the identified problems in email forensic methodologies.

- ❖ providing a detailed case management framework, which systematically handles the quality of data with a structured approach toward an email forensic investigation. This forms a very basic tool that makes the findings more reliable.
- ❖ tackle big data by being able to handle big data in a manner that manages complex datasets to ensure meaningful insights can be derived by an investigator without loss or degradation of critical data.
- ❖ developed with integrations for a variety of data types to support different forms of data, such as structured, semi-structured, and unstructured data, in a holistic manner that supports comprehensive analytics.
- ❖ to determine best practices by setting guidelines through empirical research that enhance data quality across the forensic process.
- ❖ examine open-source tools by investigating and integrating those resources for flexibility and cost-saving to increase the functionality and accessibility of the framework to practitioners in the field. The integration of such approaches would address the lacuna in the existing methodologies and result in successful email forensic investigations.

C. Design & development

The design and development of the email forensic investigation framework followed an iterative process informed by the DSRM. This involved the following key activities:

- the prototyping of initial versions of the framework was undertaken. These advanced prototypes incorporated data integration modules, mechanisms for quality assessment, and case management that were easy to use by investigators.
- Conceptual modeling and architectural design of the framework, incorporating the identified open-source tools and addressing the defined objectives.
- Stakeholder engagement also importantly complemented the design process. The team actively collaborated with forensic experts, practitioners, and direct users to elicit information and comments. This engagement occurred through workshops, focus groups, and interviews, providing stakeholders an avenue to share their input on the design and functionality of the framework. Their expertise proved valuable in identifying challenges and opportunities for improvement.
- Integration of the open-source tools within the framework, ensuring seamless data flow and functionality.
- The experiment was performed on several laptops hosting Microsoft Windows 11 Ultimate 64 bits. The hardware specification includes Intel core 2Duo CPU, 2.5 GHz, 4GB RAM, and 320GB HDD (two partitions). To prepare for the experiment, the researcher first reimaged the laptops. This process ensures that no applications exists on laptops and sound forensics process is followed. Then, the researcher installed an email forensics tool on each laptop, imported the Enron email dataset file to each of the software tool ready for the experiment. In addition, the parameters of evaluation will be discussed.

The process was iterative, whereby the developed framework was continuously refined through feedback received from identified stakeholders. Each iteration ensured the enhancement of the framework's features to closely align with user needs and operational requirements.

D. Demonstration

Email Investigation applies natural and physical scientific methods during an investigation process for solving crimes. Email is used for many different purposes, including contacting friends, communicating with professors and supervisors, requesting information, and applying for jobs, internships, and scholarships. Depending on your purposes, the messages you send will differ in their formality, intended audience, and desired outcomes. Unfortunately, cybercriminals also see the value of data and seek to exploit security holes to put your

information at risk. Email forensics refers to analyzing the source and content of emails as evidence. Investigation of email related crimes and incidents involves various approaches. Through this research project, the researcher will try to find a more effective way to investigate email crime through the appropriate open-source tool. The proposed research was tried to provide a structured email investigation framework

E. Evaluation

An essential step in determining the efficacy of the framework's design was the evaluation phase. Among the essential elements that made up this were:

- Qualitative Assessments: To gather qualitative input from users, usability testing was conducted using actual email forensic cases. This indicates how the framework will truly satisfy an investigator's needs by assessing its usability and practical application.
- Quantitative Indicators: The group then created a few performance indicators that were used to assess the framework's efficacy. These included decision-making precision, data processing speed, and data type integration. The team would be able to objectively analyze the framework's success thanks to these quantitative parameters.
- Comparison Studies: Comparing the new framework with the existing methodologies highlighted improvements in terms of data quality, processing speed, and overall efficiency. The framework gave a clear perception in advancing the state of email forensics and showing the advantages over traditional approaches.

These various evaluation strategies were thus employed by the research team in an attempt to make sure that this framework would indeed be effective and practical to apply in any real-life email forensic investigation.

F. Communication

Communicate the problem, its solution, and the utility, novelty, and effectiveness of the solution to researchers and other relevant audiences. Based on the results obtained from evaluating the artefact, reflect on what worked well and what did not and what can be improved in future iterations. Use the feedback obtained in the reflection phase to go back to step 3 and refine the design theory and iterate again until a satisfactory investigation framework is developed. Hope this helps! Let me know if you would like me to modify anything or if you have any further requests

3.2 Data Collection

The data collection for this study involved a multi-pronged approach:

3.2.1 Literature Review

The researchers conducted a comprehensive review of the existing literature on email forensics, including academic journals, conference proceedings, industry reports, and practitioner publications. The review focused on identifying the current challenges, best practices, and emerging trends in the field of email forensic investigations.

3.2.2 Stakeholder Interviews

The researchers conducted in-depth interviews with email forensic experts, law enforcement officials, and IT security professionals to gather insights into the practical challenges, operational requirements, and organizational barriers faced in conducting effective email forensic investigations.

3.2.3 Case Study Analysis

The researchers analysed several real-world email forensic investigation cases, both from publicly available sources and through collaboration with law enforcement agencies and legal organizations. These case studies provided valuable insights into the practical application of email forensic techniques and the associated challenges.

3.3 Data Analysis

The data collected from the literature review, stakeholder interviews, and case study analysis was synthesized using qualitative data analysis techniques, such as thematic analysis and coding. The researchers identified the recurring themes, patterns, and pain points related to email forensic investigations, which informed the design of the proposed framework.

3.4 Framework Design and Development

Based on the insights gathered from the data collection and analysis, the researchers designed a comprehensive email forensic investigation framework. The framework was developed iteratively, with constant feedback and validation from the stakeholders and subject matter experts.

3.5 Prototype Implementation

To demonstrate the practical applicability of the proposed framework, the researchers developed a functional prototype, which included the key components and capabilities identified during the framework design phase. The prototype was designed to address the identified technical, operational, legal, and organizational hurdles.

3.6 Evaluation and Validation

The designed framework and the developed prototype were evaluated using a set of predefined metrics, including simplicity, completeness, consistency, integrity, security, and usability. The

evaluation involved both quantitative and qualitative assessments, such as user studies, expert reviews, and performance benchmarking.

3.7 Communication and Dissemination

The findings and outcomes of this research, including the designed framework, developed prototype, and evaluation results, are being communicated through this chapter and other relevant outlets to share the knowledge and insights with the academic and practitioner communities.

3.8 Ethical Concerns:

In this research using the DSR approach, the ethical considerations to ensure integrity and fairness and adhere to legal and professional standards become important. Among the primary concerns in conducting an email, forensic investigation is privacy and confidentiality in data. Since most of the email data contains sensitive personal or corporate information, all data used in this study have been anonymized to protect against unauthorized access or disclosure. Moreover, the research design follows the relevant data protection regulations, such as GDPR and HIPAA, to keep it compliant with legal frameworks.

Forensic experts who took part in the evaluation of the structured framework developed the informed consent. All participants were clearly informed of the objectives of the research, their role in the study, and how their output would be put to use to ensure that the process was voluntary and transparent. Another ethical issue is to avoid bias, so that the framework may be non-discriminatory; testing has thus considered diversified case studies in different jurisdictions to ensure that none of the various tools, datasets, and investigative methods may be favoured.

Of particular interest has been the decision-making transparency since automation and AI are becoming part of forensic investigations. It has integrated automated tools that are able to present explainable results, ensuring the investigators interpret and verify the results instead of relying on black-box algorithms. This research also ensures compliance with legislation and regulations because the framework was developed in line with international digital forensic standards, considered the challenges of jurisdictions, and was based on methodology applicable in a wide range of legal environments.

Lastly, due care was taken to ensure responsible disclosure and security with a view not to misuse forensic tools and methodologies. Forensic procedures and access to tools are restricted to authorized professionals to ensure that the knowledge and technology developed through this research are used ethically and lawfully. By considering these ethical concerns, this study ensures that its contribution to the field of email forensics will not only be effective but also responsible in nature, hence promoting the advancement of forensic practices upholding legal and ethical integrity.

CHAPTER FOUR

4 Proposed Solution

4.1 Proposed new Framework

The email forensic framework – see Fig. 1. For establishing a standard email investigation forensic architecture, The Proposed Email Forensic Investigation Framework is a contemporary, structured, and effective method to cope with the evolving challenges of email-based investigations in the modern technological environment. By the rising complexity of cybercrimes as such as phishing, fraud, email spoofing, malware dissemination, and insider threats, conventional investigative techniques tend to be inadequate for dealing with the enormous scale and intricacy of contemporary email systems. This framework builds on best-practice forensic methodologies, strong data control measures and emerging technologies to offer a complete, legally admissible and defensible model of email investigation.

One of the distinguishing features of the framework is its focus on scalability. Email investigations, as used to be in traditional methods, are usually not able to handle a massive number of emails, attachments, and metadata effectively. This framework integrates automated data ingestion, indexing, and filtering techniques that allow investigators to handle vast datasets systematically, reducing noise and enabling them to focus on the most relevant evidence.

A major contribution of the framework is its strong focus on compliance with legal, regulatory, and organizational policies. Investigators are obliged to process restricted personally identifiable info (PII), confidential communications, and trade secrets. This framework has the effect of embedding data governance requirements at each stage, such as role-based access control (RBAC), data reduction, encryption and compliance with privacy laws. Upholding transparency, accountability, and ethical treatment of data makes the framework such that all findings are admissible at trial and comply with international standards.

The framework also emphasizes iterative analysis, an approach that allows investigators to revisit and refine their findings as new evidence emerges. Iterative analysis guarantees that information extracted during metadata review, content review, or as a function of an interaction with external intelligence sources are embedded back into earlier phases of the investigation. This iterative process deepens and refines the process of investigation and can lead to the identification of hitherto unexpected relationships or aberrant behavior. This adaptive strategy is necessary when involved in complicated cases in which the attackers use advanced obfuscation strategies.

Moreover, the framework integrates advanced technological tools to enhance the investigative process. These tools enable the automation of labor-intensive tasks, such as parsing email headers, analyzing attachments for malware, and identifying suspicious communication patterns.

Visualization tools help investigators to model communication networks, to trace transmission pathways, and to identify abnormalities in email traffic. By leveraging these technologies, the framework ensures efficiency without compromising forensic rigor.

Beyond technical improvements, the framework provides advanced reporting infrastructures to enable transparent and actionable dissemination of the results. Reports of forensic analysis produced within this framework comprise comprehensive logging of protocols, time scales, metadata, and evidence, in a format, which is both technical and non-technical, including all relevant provenance. Visual aids like timelines, metadata flow charts, and communication flow charts enhance clarity and allow judges and other participants to understand the results easily.

The proposed email forensic investigation framework goes beyond the conventional approaches in that it addresses key demand for scalability, demand for compliance, and iterated analysis. Its incorporation of contemporary processes, governance standards, and advanced tools ensures modern complexities in email investigations can be handled effectively by the investigator. This framework allows investigators to provide results that are accurate, swift, and compliant with the Constitution, commensurate with ethical law and effectively responsive to a changing digital world.

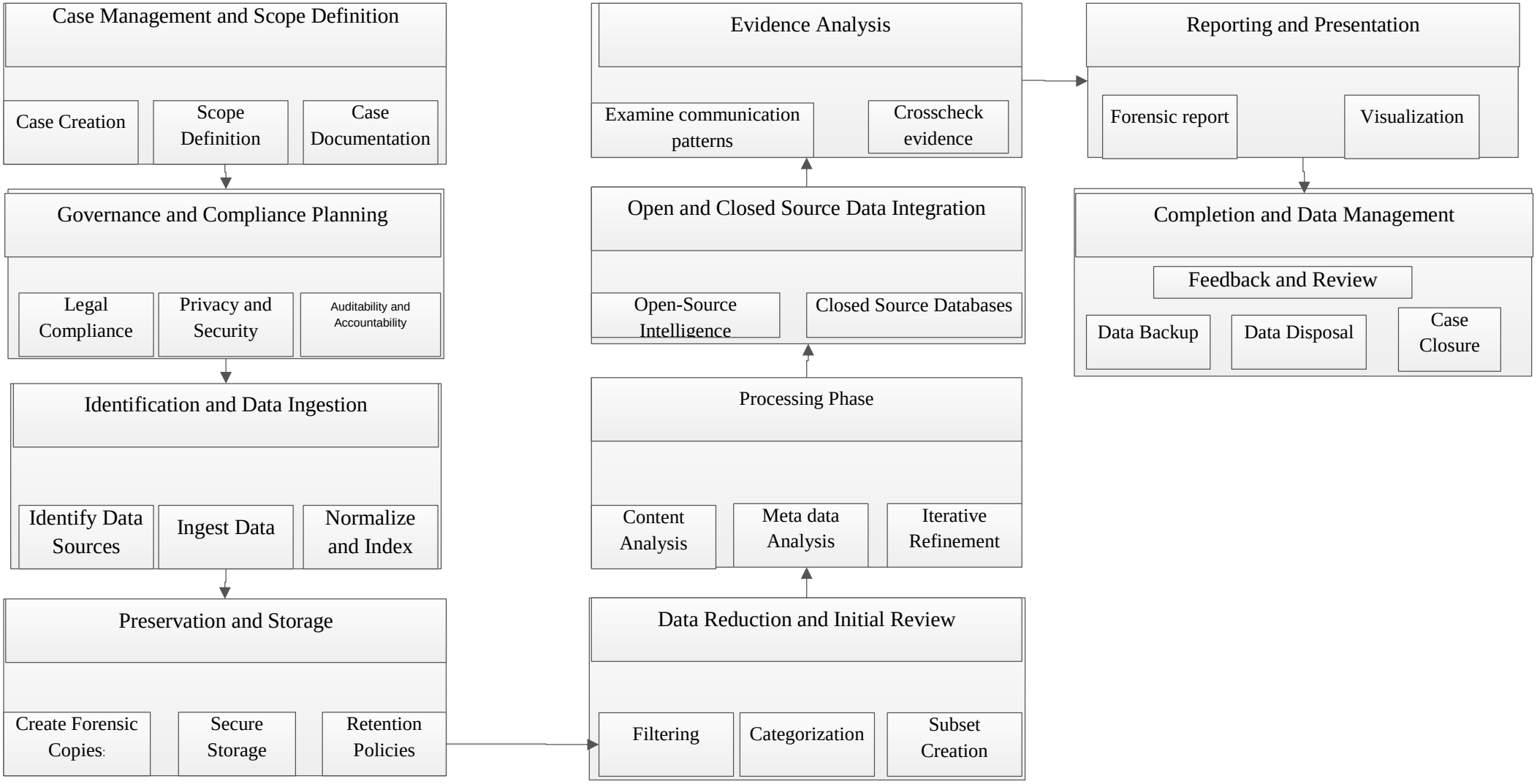


Figure 3structured email investigation Framework

Standard Email Investigation Framework

The standard email investigation framework will, therefore, provide a formalized and structured step-by-step process for the forensic investigation of emails to ensure compliance, accuracy, and legal defensibility. It covers major aspects: case management, governance, data acquisition, processing, storage, and analysis.

Phase 1: Case Management and Scope Definition

It identifies the scope of the inquiry through definition of the investigation, statement of its objectives, and establishing an adequate governance structure within which a case manager puts it onto systems supported and a unique identifier assigned. Information critical to the target accounts, devices, timeframe of analysis is documented; stakeholder investigators, counsel legal, and compliance officers are clearly pointed. During this phase, the framework also aligns the investigation to relevant laws and organizational policies, ensuring accountability and legal admissibility of findings.

Phase 2: Governance and Legal Preparation

Governance is central in ensuring compliance, privacy, and transparency throughout the investigation process. Legal permissions are reviewed down to the last detail, such as search warrants or organizational consent, before any collection of data will take place to make sure all investigative actions are lawful and justified. Clear data governance policies are laid down to define how sensitive information is to be treated, including encryption of data in transit and at rest; role-based access control, where the access rights of investigators depend on their roles; and data minimization, where only that information is collected which has a relevance to the investigation. All activities within the investigation are recorded with due diligence for traceability, and measures for auditing performed actions are also put in place to ensure transparency and, when relevant, subject to review by the authorities. Also, all governance actions are based on the respective regulatory frameworks pertinent to securing data protection and privacy through local, national, and international laws. Regular reviews against governance policies are done to meet the change in legislation or best practices, undertaken in trusting stakeholders that the investigations will be done in an ethical and transparent manner.

Phase 3: Identification and Data Ingestion

This is a very important stage in email investigation that created a solid basis for the analysis process, narrowing down sources of evidence and ingesting email data into a forensic platform. Investigators identify diverse sources of evidence that can include email servers, mobile devices, laptops, computers, cloud storage platforms, and backup archives, all of which may contain

essential email communications. Having identified these sources, extraction of data and ingestion is allowed into the forensic platform where normalization is done on all types of different formats into one and then into standardized consistent formats. At that instance, a searchable index would have been created which allowed for efficient analysis: search by keywords, or even subject matter, filtering results according to sender, receiver, or by date. To preserve the integrity of the data, cryptographic hashes are created for each email or set of data, thus providing a digital fingerprint that will show if the data has been modified or tampered with in any way. This stage will ensure that all available evidence is systematically acquired and rigorously prepared for review, while comprehensive documentation of the collection process maintains a clear chain of custody for possible legal procedures. In so doing, investigators will be able to carry out extensive analyses leading to valid and reliable results.

Phase 4: Preservation and Storing

Preservation is paramount to maintaining the integrity of evidence within the process of investigation through email. In ensuring this integrity, all collected data is forensically copied bit-by-bit images using write-protected tools to prevent any alteration of original data during the copying process. These forensic copies are then hashed to verify their authenticity; unique hash values are generated that act like digital fingerprints. Copies are then taken to tamper-proof facilities for further avoidance of unauthorized access or changes. Also, with logical evidence containers, refined datasets will have .E01 or .AFF format, which can easily be managed and analyzed. These data are retained in an encrypted secure storage system; access is controlled so that only authorized persons can deal with it. It institutes retention policies on how long the evidence needs to be retained and when it should be securely deleted, thus balancing complete investigations with demands for data protection. This very critical stage needs to ensure the integrity, security, and admissibility of evidence into court, so that standards of legal compliance and forensic integrity are maintained throughout.

Phase 5: Data Reduction and Initial Review

This phase therefore makes the investigation organized through elimination of data not relevant to the case, hence effectiveness and concentration. Investigators use keyword searches, date ranges, and filters on sender and recipient details as ways to extract the relevant material from the bulk. During this stage, investigators should identify and eliminate duplicate e-mails or non-essential messages that are not related to the investigation. After refining, the emails are then tagged and categorized according to their themes, whether suspicious activities, phishing attempts, or normal communications. It allows investigators to focus on a targeted subset of data in an organized manner that can be used to conduct detailed analyses with much more ease. This

will help the investigation move forward expeditiously by ensuring resources are focused on relevant information leading to the most pertinent findings and conclusions.

Phase 6: Processing Phase

The processing stage is a vital step for structuring and analyzing the refined data to extract useful information that may be essential for the investigation. This phase begins with metadata analysis, whereby investigators look at the headers of emails to trace transmission routes, detect spoofing attempts, and verify the authenticity of senders. Through the analysis of metadata, investigators are able to pick out critical details on the origin and path of the emails. At the same time, email bodies and attachments are scanned for malicious content, such as phishing attempts, embedded URLs, or malware, making sure that any potential threats are identified and assessed. In order to remove redundant data and further streamline the dataset, giving a finer focus on communications that are unique and relevant to the investigation, de-duplication techniques are used. Iterations of refinement enable investigators to uncover deeper patterns and connections that may well not be immediately obvious. Such extended processing will eventually lead to deepened insight from the evidence to inform decisions and act on the issues identified

Phase 7: Open and Closed Source Data Integration

Value addition from this investigative finding is done with external intelligence sources that extend both the analytical scope and contextual environment. Investigators leverage email evidence using OSINT platforms to uncover connections to known threats; the platform identifies relationships and associations not obvious directly from the email itself. Examples of these closed-source databases include, but are not limited to, law enforcement records or organizational threat intelligence critical for offering background insight. These are findings from those sources, carefully documented and integrated back into earlier phases of the investigation in order to iteratively refine them. This feedback allows investigators to revisit and revise their analyses as new information emerges, thereby supporting an integrated investigation that is continuously revised as more data become available. This phase, by applying both open and closed-source intelligence, significantly enhances the investigation process with more robust conclusions and informed decision-making.

Phase 8: Evidence Analysis

Detailed forensic analysis shall be conducted to reveal communication patterns, hidden relationships, and possible malicious activities across the email data. Investigators map sender-receiver networks with visualization tools in order to spot anomalies and unusual interactions indicative of suspicious behavior. Attachments will be scrutinized for malware, macros, or hidden payloads; many times, sandbox environments allow for safe analytics without putting at risk the

integrity of the systems involved. These findings are further correlated with threat intelligence databases to strengthen the investigation, enabling the investigators to attribute connections to known attack patterns and tactics by cyber adversaries. This in-depth analysis builds up a strong narrative of evidence, where the insights are interwoven to support conclusions and actionable recommendations. The presentation of such a strong case, while systematically analyzing the data for any relevant correlations with external intelligence, would enhance understanding and response strategies pertaining to the identified threats.

Phase 9: Reporting and Presentation

In this stage, the forensic findings would be prepared as actionable reports and presentations, communicating results. Reports would be prepared to document methodologies, findings, and conclusions in a clear and concise manner that presents all information available. Timelines, communication diagrams, metadata flowcharts, and other visual aids are used in helping stakeholders understand the complexity of evidence and relationships among several data points. Wherever necessary, redaction is applied to protect sensitive information unless law compels disclosure. Findings are presented to stakeholders, legal teams, or courts in both technical and nontechnical formats, catering for diverse audiences and ensuring that the information is understandable and actionable. This cautious revelation of evidence not only supports the conclusions of the investigation but also furthers and strengthens informed decision-making from those findings to effective responses, including possible legal outcomes.

Phase 10: Completion and Data Management

The final phase is one of case closure and investigation data management. Evidence, forensic images, and reports are safely archived according to retention policies. Irrelevant or temporary data are safely deleted in order to minimize risks. A post-investigation review helps document lessons learned to enhance governance policies and refine forensic processes. The feedback from the stakeholders will also be considered for the improvement of future investigations. This phase ensures that the investigation is complete in a systematic and ethical manner.

It embodies the investigation workflow with steps involved in processing and storing and therefore may be employed to formalize a trusted and compliant approach. In this chain, every stage addresses specified needs within an investigation for data integrity, privacy, and governance. These steps constitute a structured process through which fast yet legally defensible forensic investigations can be carried out over email.

CHAPTER FIVE

5 Experiments, Results, and Analysis

This studied used a set of in-depth case studies to demonstrate the efficiency of the open-source tools developed within this study's framework. These case studies will show how this framework can satisfy forensic needs for a wide variety of devices and applications that are closely aligned with the principles and guidelines provided by the DS community. These case studies are not intended, therefore, for exhaustive analysis of any single data set but as showcases for the application of the proposed email forensic investigation process. By walking through these real-world examples, this research tries to construct a strong, versatile framework that will be useful in the evaluation of usefulness and potency regarding suggested email forensic tools.

These case studies in this work range from corporate email servers to personal email accounts and extend into investigating emails communicative across different file formats and storage locations. Each case study represents a detailed workflow, as laid out in the email forensic investigation framework, showing smooth integration with the Autopsy Digital Forensics Platform and its suite of specialized email analysis capabilities.

These case studies demonstrate research on how the proposed framework effectively collects, processes, and analyzes email-based evidence in a manner that guarantees integrity and admissibility during an investigation. Case studies also portray flexibility in how the framework handles diversity on the sources of email data; the efficiency in extracting and interpreting critical metadata, identification of suspicious communication patterns, and potential evidence of fraudulent or malicious activities is warranted.

The research documented in detail case studies that were useful references among forensic investigators, cybersecurity professionals, and researchers in the area of email forensics. Such a set of case studies will provide good opportunities for lessons and insights that could be used in refining and further developing the proposed email forensic investigation framework to be better in applicability and relevance during real investigative scenarios.

5.1 OPEN SOURCE E-MAIL FORENSIC TOOLS

Open-source software has many benefits that can allow an organization to choose it. it is almost universally free to use, drastically lowering the purchase price and thus making it available to people who have tight budgets. Users are able to take advantage of the flexibility and tailoring that open-source solutions enable, because of the ability to tailor the source code to reflect the particular organizational requirements. [22] Open-source software transparency helps build trust because by viewing the code, users can check for security weaknesses and can thereby comply with security standards. Furthermore, continuous community support supports the development

of the software, because the community members can support each other, share good experiences, and participate in continuous improvements. In contrast to proprietary tools, open-source tools obviate vendor lock-in, therefore providing the organizations with a flexibility in selection and adaption of the software without dependence risks. Interoperability is one of the main benefits, because most open-source plugins are very compatible with existing tools, so it is easy to exchange data.

Email forensic tools that have their source code made freely available for anyone to view, use, alter, or distribute are known as open source Email forensic tools. Researchers, digital forensic practitioners, and other stakeholders can better comprehend the tool's operation thanks to this degree of transparency, which can boost confidence in the tool's output. [17] Several open source email forensic tools are available to help with the analysis of the source and content of email messages so that an attack or the malicious intent of intrusions may be looked into. These tools enable to identify the origin and destination of the message, trace the path taken by the message, and identify spam and phishing networks, among other capabilities, while also giving an easy-to-use browser interface and automatic reports. Some of these tools are introduced in this section.

Mbox Parser

Mbox Parser is an open-source Python library that allows you to parse and extract data from mbox-format email files. Mbox is a common file format used to store multiple email messages in a single file. In our investigation, one of the key tools we are utilizing is the Mbox Parser - an open-source Python library that allows us to parse and extract data from mbox-format email files. Mbox is a common file format used to store multiple email messages in a single file. This is particularly useful for our research, as many of the email datasets we are working with are provided in the mbox format.

The Mbox Parser library enables us to efficiently read and process these mbox files, extracting important metadata such as sender, recipient, timestamps, subject lines, message bodies, and attachments. This allows us to build a structured representation of the email data that can be further analyzed and correlated with other forensic artifacts.

One of the key advantages of using the Mbox Parser is its ability to handle complex email structures, including nested attachments and embedded content. This is crucial for our investigation, as we often encounter email messages with multiple layers of forwarded content or rich media elements that need to be properly extracted and examined.

Additionally, the Mbox Parser's open-source nature allows us to extend and customize its functionality as needed to meet the specific requirements of this research. We have been able to integrate the parser with other components of our email forensics investigation framework, such

as the network traffic analysis and threat intelligence modules, to provide a more comprehensive and streamlined investigative workflow.

Overall, the Mbox Parser has proven to be an invaluable tool in this email forensics research, enabling us to efficiently process and analyze large email datasets from a variety of sources. Its robustness, flexibility, and integration capabilities have been instrumental in helping us uncover valuable insights and develop a more effective investigation framework.

ETL

In the context of email data analysis, the ETL (Extract, Transform, and Load) process is essential for preparing data for meaningful insights. This process begins with extraction, where data is sourced from various formats, typically CSV files that include key fields such as sender, recipient, subject, and date. The next stage, transformation, involves cleaning and filtering the data to improve its quality; this may include removing missing values, converting date formats, and applying filters to focus on relevant timeframes, such as emails sent after a specific date. Additionally, transformations can involve extracting domains from email addresses to enhance analytical capabilities. Finally, in the load phase, the cleaned and transformed data is saved into a new format, such as a CSV file or a database, making it accessible for further analysis. By implementing this ETL process, researchers can ensure that their analyses of email communication patterns are based on high quality, relevant data, ultimately leading to more accurate and insightful conclusions about communication trends and behaviors. This ETL process implemented in real practice using Python. The code in the appendix provides a snippet to extract email data from some source, transforms with necessary cleaning and normalization, and then loads the result to a Mongodb database. This piece of code could be more specific with modifications and customized additions to cater for particular needs by an organization intending to integrate emails.

Mongo db

MongoDB is an ideal storage solution for email forensic investigations due to its flexible, document-oriented database model, which allows for the storage of unstructured or semi structured data such as emails, attachments, and metadata. Unlike relational databases that rely on predefined schemas, MongoDB enables the storage and querying of large volumes of diverse data, including email content, headers, timestamps, sender/recipient information, and attached files. Given the variety and complexity of forensic evidence, MongoDB’s NoSQL structure is well suited to adapt as data needs evolve during investigations. This flexibility and scalability make MongoDB a compelling choice for managing the growing amounts of forensic data in modern digital investigations. To enhance the performance of investigations, particularly when

dealing with large volumes of email data, MongoDB's indexing capabilities are crucial. Indexes can be created on common query fields such as sender, recipient, timestamp, and subject, allowing for fast retrieval of relevant emails. Additionally, MongoDB supports text indexes, enabling full-text search within email bodies, subjects, and headers. This feature is especially valuable for identifying key phrases or suspicious keywords. Geospatial indexes can also be applied if metadata such as IP addresses or geolocation data is collected, allowing investigators to track the geographic origin of suspicious emails or activity.

Maintaining the integrity of evidence throughout the forensic process is crucial. MongoDB offers features that help ensure integrity, such as write concerns and read concerns, which guarantee that all database operations meet specified consistency and durability requirements. Replication can maintain copies of data across multiple servers, protecting against loss or corruption. Additionally, audit logs stored in a separate Logs Collection track every action taken on the data, helping maintain the chain of custody and ensuring that the evidence handling process is traceable.

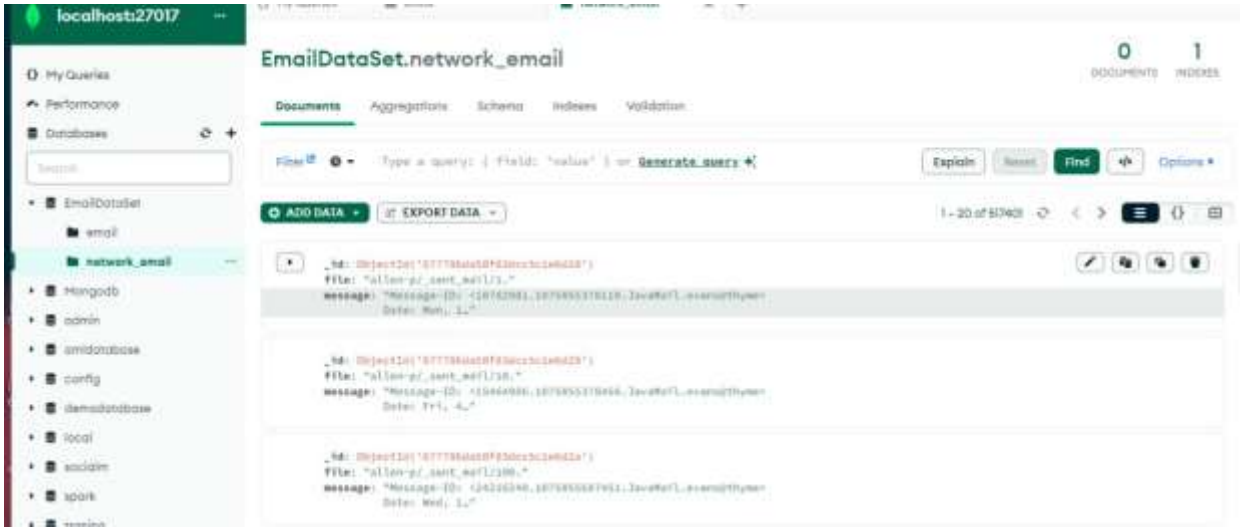


Figure 4mongodb database

Figure 5 mongodb data structure

Researcher would select MongoDB over SQL databases during the case of email forensic analysis due to its flexibility, scalability, and fault tolerance, which is critical in a quest to analyze varied and changing forms of information typical in a case of an email forensic. MongoDB document database easily accommodates complex email data, i.e., attachments, various types of forms, and metadata, without laying down any schema in advance. It is this scalability that makes it easy to change data models with new evidence or requirements arising, which helps speed up the response to evolving investigation needs. Horizontal scalability in MongoDB also makes large collections of email data simple to process and analyze, rendering it a great choice for managing huge email archives and real-time analysis of data. Preferably, its built-in fault tolerance features such as replica sets and auto-failover ensure that data remains available and sound in the event of hardware crashes or abrupt disruption. In most cases, the qualities ensure a more efficient and effective investigation through faster identification of patterns and abnormalities necessary for generating useful insights at the same time maintaining data integrity and availability.

Criteria	**mongoDB**	**SQL**
Schema Flexibility	Schema-less	Rigid schema
Horizontal Scalability	Horizontally using sharding	scale vertically by adding more resources (CPU, RAM)
Embedded Documents and Arrays	Supports embedded documents and arrays natively	Requires complex joins between tables
High Availability and Fault Tolerance	Supports built-in replication through replica sets	Requires complex , additional setup and configuration for replication
Faster Development and Agile Iteration	Rapid application development	Often require migrations that can slow down development.

JSON Data Format	Highly compatible with modern web applications	Requires mapping to make it compatible with modern web applications
Performance in Big Data and Real-Time Analytics	Optimized for large-scale data processing and can handle massive amounts of data	struggle with performance at scale, especially when dealing with complex queries and joins on very large datasets

Figure 6 database comparison

Apache Atlas

The integration of Apache Atlas into a basic Email Forensic Framework extends increased metadata management and data governance for an organization interested in digital forensics. Bearing in mind that the volume of email exchange is still growing, adequate facilities to handle and explore the data are of prime importance. Apache Atlas allows for the creation of comprehensive metadata models that define entities such as emails, attachments, senders, and recipients, thus enabling an organization to classify email data based on sensitivity and compliance requirements. This classification is of the essence in maintaining local and International regulatory compliance. Furthermore, Apache Atlas enhances data lineage by providing the capability for an organization to visualize email data flow across systems; this, in essence, documents the transformations occurring on the data and maintains a clear audit trail. This type of transparency is important in forensic investigations since it provides context and history for analysts operating on the data in question. The integration of Apache Atlas into an organizational email system enables it to automatically ingest metadata, enhancing efficiency and accuracy across forensic processes. Further, search and discovery enable forensic analysts to uncover, in a few clicks, emails relevant to a case for effective investigation. Essentially, the adoption of Apache Atlas in an email forensic framework extends its capabilities in not only providing effective data governance but also positions the organizations to better respond to incidents involving emails, thereby reassuring them about data integrity and compliance.

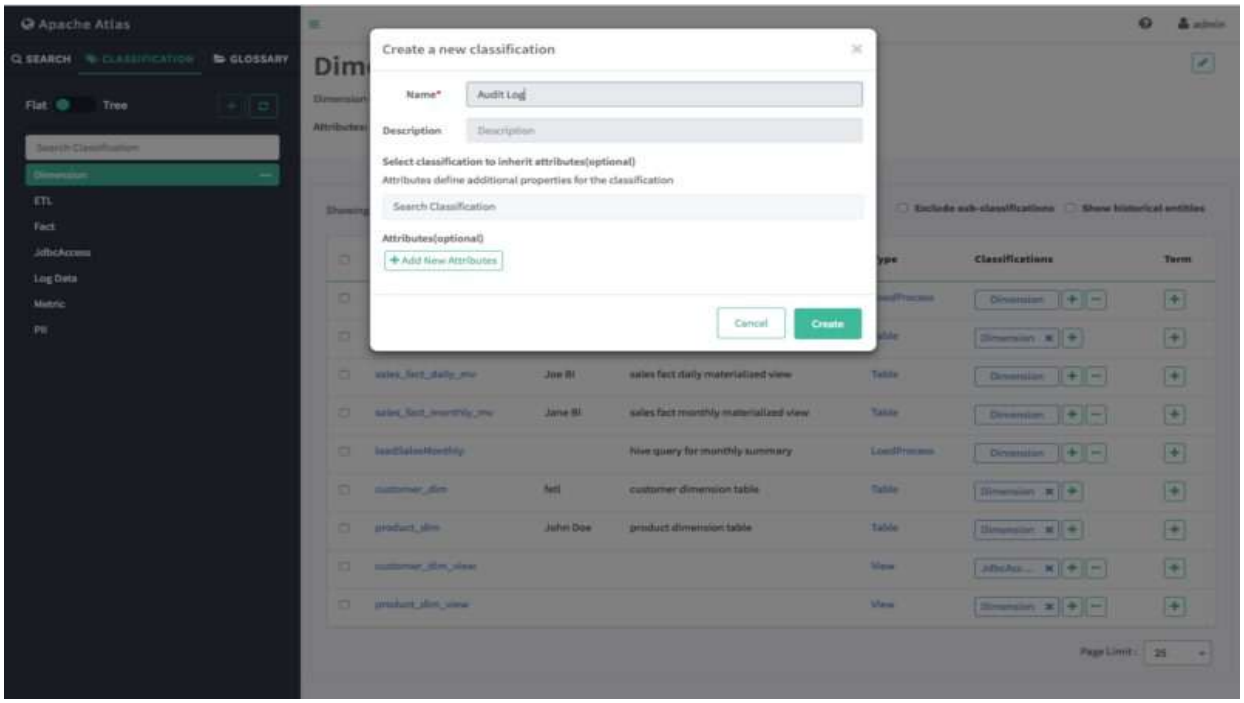


Figure 7Apache atlas

Machine learning algorisms

In this framework, machine learning gives email investigators a strong tool for effectively identifying and investigating threats. In order to accurately classify emails as malicious or lawful within an email forensic framework, the researcher employed logistic regression. Logistic regression algorithms learn to recognize patterns and associations linked to malicious emails by examining important data such as sender and recipient addresses, subject lines, email body text, and the presence of particular keywords or URLs. This makes it possible to accurately categorize email types.

Importing the Dependencies

In [4]:
```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression, LogisticRegressionCV
from sklearn.metrics import accuracy_score
```

Data Collection & Pre-Processing

In [6]:
```
# loading the data from csv file to a pandas Dataframe

# for file in the content folder of the drive => used for colab
raw_mail_data = pd.read_csv('C:/Users/SUPER USER/Desktop/mail_data.csv')

# if file is present in local directory
# raw_mail_data = pd.read_csv('./spam_ham_dataset.csv')
```

In [7]:
```
raw_mail_data.head()
```

Out[7]:

	Category	Message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

Figure 8ml import dependency

In [8]:
```
# replace the null values with a null string
mail_data = raw_mail_data.where((pd.notnull(raw_mail_data)),'')
```

In [9]:
```
# printing the first 5 rows of the dataframe
mail_data.head()
```

Out[9]:

	Category	Message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

In [16]:
```
# checking the number of rows and columns in the dataframe
mail_data.shape
```

Out[16]: (5572, 2)

Label Encoding

In [18]:
```
# Check the actual column names
print(mail_data.columns)

# If the column name is 'Category' and you want to assign 1 for spam and 0 for ham
mail_data.loc[mail_data['Category'] == 'spam', 'Category'] = 1
mail_data.loc[mail_data['Category'] == 'ham', 'Category'] = 0
```

```
Index(['Category', 'Message'], dtype='object')
```

In [160]:
```
# label spam mail as 1;  ham mail as 0;

mail_data.loc[mail_data['Category'] == 'spam', 'Category',] = 1
mail_data.loc[mail_data['Category'] == 'ham', 'Category',] = 0
```

In [161]:
```
# Print the column names to verify available columns
print(mail_data.columns)
```

```
Index(['Category', 'Message'], dtype='object')
```

Figure 9 data cleaning

```
In [161]: # Print the column names to verify available columns
          print(mail_data.columns)
```

```
Index(['Category', 'Message'], dtype='object')
```

spam - 0

ham - 1

```
In [162]: # If the text data is in the 'Message' column
          X = mail_data['Message']  # Text data
          Y = mail_data['Category']  # Assuming 'label_num' is your target variable
```

```
In [163]: print(X)
```

```
0       Go until jurong point, crazy.. Available only ...
1                           Ok lar... Joking wif u oni...
2       Free entry in 2 a wkly comp to win FA Cup fina...
3       U dun say so early hor... U c already then say...
4       Nah I don't think he goes to usf, he lives aro...
                              ...
5567                                                     
5568                                                     
5569                                                     
5570     The guy did some bitching but I acted like i'd...
5571                                 Rofl. Its true to its name
Name: Message, Length: 5572, dtype: object
```

```
In [164]: print(Y)
```

```
0       0
1       0
2       1
3       0
4       0
       ..
5567     
5568     
5569     
5570    0
5571    0
Name: Category, Length: 5572, dtype: object
```

Splitting the data into training data & test data

Figure 10data training

```
In [38]: # Replace NaNs or empty strings with a default value (e.g., 0)
         Y_train = Y_train.fillna(0).replace('', 0)
         Y_test = Y_test.fillna(0).replace('', 0)

         # Convert Y_train and Y_test to integers
         Y_train = Y_train.astype(int)
         Y_test = Y_test.astype(int)

         # Check the result
         print(Y_train.head())
         print(Y_test.head())
```

```
28    0
45    0
12    1
74    0
50    0
Name: Category, dtype: int32
42    1
86    0
23    0
19    1
32    0
Name: Category, dtype: int32
```

```
In [39]: # transform the text data to feature vectors that can be used as input to the Logistic regression

         feature_extraction = TfidfVectorizer(min_df = 1, stop_words='english', lowercase='True')

         X_train_features = feature_extraction.fit_transform(X_train)
         X_test_features = feature_extraction.transform(X_test)

         # convert Y_train and Y_test values as integers

         Y_train = Y_train.astype('int')
         Y_test = Y_test.astype('int')
```

Figure 11data train

Logistic Regression

```
In [44]: model = LogisticRegression()

In [45]: print(X_train_features.shape)  # Should output (num_samples, num_features)
         print(Y_train.shape)           # Should output (num_samples,)

         (4457, 547)
         (90,)

In [46]: from sklearn.model_selection import train_test_split

         X = mail_data['Message']  # Assuming the text data is in 'Message'
         Y = mail_data['Category']  # Assuming 'category' contains the labels (0 or 1)

         # Split the data into training and testing sets
         X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)

In [47]: from sklearn.feature_extraction.text import CountVectorizer

         # Vectorize the text data (X_train and X_test)
         vectorizer = CountVectorizer()
         X_train_features = vectorizer.fit_transform(X_train)  # X_train is the raw text
         X_test_features = vectorizer.transform(X_test)  # X_test is the raw text

In [48]: print(X_train_features.shape)  # Check number of rows
         print(Y_train.shape)           # Check number of rows

         (4457, 668)
         (4457,)

In [51]: # Assuming 'spam' -> 1 and 'ham' -> 0
         Y_train = Y_train.replace({'spam': 1, 'ham': 0})
         Y_test = Y_test.replace({'spam': 1, 'ham': 0})

In [53]: print(Y_train[Y_train == ''])  # This will show you rows with empty strings

         1978
         3989
         3935
         4078
         4086
              ..
         3772
         5191
         5226
         5390
         860
         Name: Category, Length: 4373, dtype: object
```

Figure 12 vectored the data

```
In [54]: print(Y_train.isna().sum())  # Check for NaNs
         print((Y_train == '').sum())  # Check for empty strings

         0
         4373

In [55]: Y_train.replace('', 0, inplace=True)  # Replace empty strings with 0
         Y_test.replace('', 0, inplace=True)   # Similarly for Y_test

In [56]: Y_train.replace('', 'ham', inplace=True)  # Replace empty strings with 'ham'
         Y_test.replace('', 'ham', inplace=True)   # Similarly for Y_test

In [57]: Y_train = Y_train.astype(int)  # Now it should work without errors
         Y_test = Y_test.astype(int)    # Similarly for Y_test

In [58]: print(Y_train.head())  # Verify the target variable after modification

         1978    0
         3989    0
         3935    0
         4078    0
         4086    0
         Name: Category, dtype: int32
```

Figure 13 column data replacing

The performance of the Logistic regression algorithms models in detecting Spam email was evaluated using a comprehensive set of classification metrics that offer insights into different aspects of model performance. Each metric provides a different lens through which the effectiveness of the model can be assessed, particularly in their ability to correctly classify fraud and non-fraud cases, minimize misclassifications, and balance the trade-off between identifying fraudulent activities and avoiding false positives. For this purpose, the evaluation process was made in relation to the following metrics.

```
object
['0' '1' '']
Shape of X_train_features: (4457, 534)
Shape of Y_train: (4457,)
Accuracy: 0.9775784753363229
Classification Report:
              precision    recall  f1-score   support

                   0.98      1.00      0.99      1087
           0       1.00      0.14      0.24        22
           1       0.00      0.00      0.00         6

    accuracy                           0.98      1115
   macro avg       0.66      0.38      0.41      1115
weighted avg       0.97      0.98      0.97      1115

Confusion Matrix:
 [[1087    0    0]
 [  19    3    0]
 [   6    0    0]]
```

Figure 14 logistic regression performance measurement

MxToolBox Email Header Analyzer

Its free web based email investigation tool, when reporting spam that slips past the filters, it is essential that the researcher receive the full message headers from a message. Additionally, sometimes our Support department may request the full headers from an email message in order to troubleshoot mail delivery problems.

Every single Internet e-mail message is made up of two parts the header and the message body of the email. Every single email you send or receive on the Internet contains an Internet Header; a full and valid e-mail header provides a detailed log of the network path taken by the message between the mail sender and the mail receiver(s) (email servers). Email client program will usually hide the full header or display only lines, such as From, To, Date, and Subject, see below for more information on pulling headers for your email client:

In this research case study, the researcher examine (figure 1) The MxToolBox Email Header Analyzer provides access to the complete email header, which contains a detailed log of the network path taken by the message. This information is crucial for troubleshooting mail delivery problems and analyzing potential phishing attempts. The full header data can help the research team and the PhishTank support department identify the origin of the message, trace its routing, and gather other metadata that may reveal insights about the source and nature of the spam or phishing attempt.

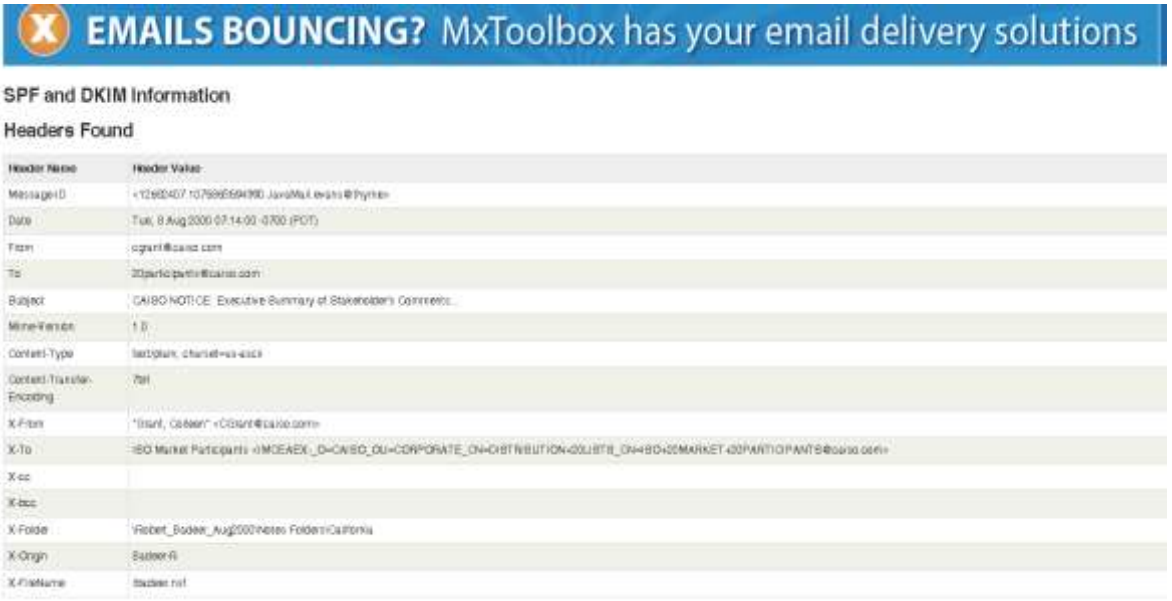


Figure 15mxtoolbox analysis result

PhishTool

PhishTool automatically retrieves all of the relevant metadata from a phishing email, providing the most comprehensive technical view of a phishing email possible. This combined with OSINT and heuristic detection, makes PhishTool one seriously powerful tool. PhishTool can analyse .eml, .msg and .txt message format. This is experiment demonstrated that the tool has the following major features.

- Easily reverse engineer attachments and URLs
- Integrates with third-party APIs
- Automatically detects how a phishing email defeated security controls
- Capture a detailed forensic report

In this research case study, the researcher examine (figure 2) The PhishTool Analyzer providesIdentification of common tactics, techniques and procedures (TTPs) used by phishers Analysis of URL structures, domain registrations, and other indicators of compromise, Detection of email authentication failures that enabled the spoofing attempts, Tracking of network paths and infrastructure used to host the malicious content

.

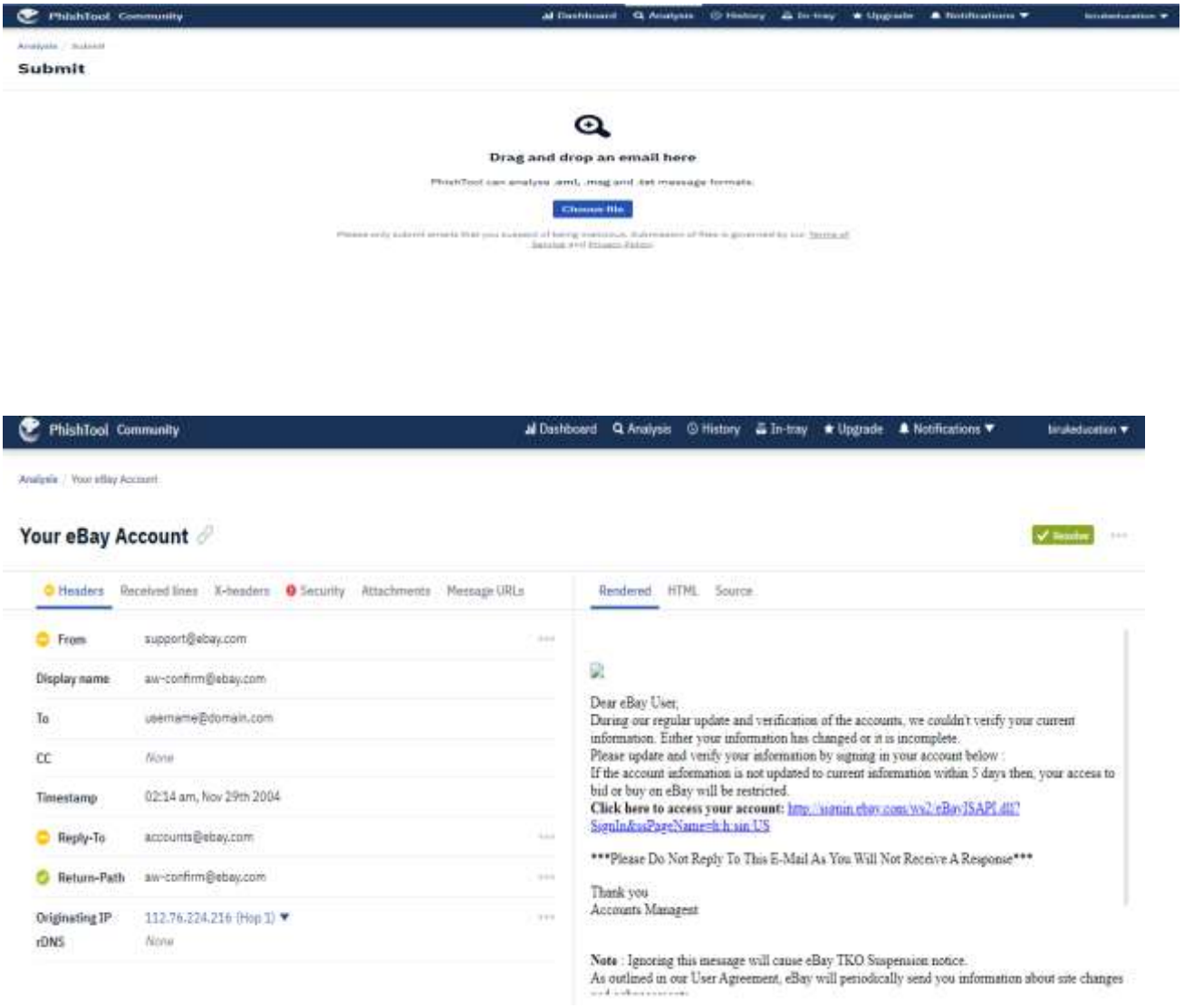


Figure 16 phisihtool examination result

PhishTank

PhishTank is a free community site where anyone can submit, verify, track and share phishing data, collaborative clearinghouse for data and information about phishing on the Internet. In addition, PhishTank provides an open API for developers and researchers to integrate anti-phishing data into their applications at no charge. In this research case study, the researcher examine the PhishTank platform, a free community site where anyone can submit, verify, track and share phishing data, serving as a collaborative clearinghouse for information about phishing on the Internet. PhishTank also provides an open API for developers and researchers to integrate anti-phishing data into their applications at no charge.

In this research case study, figure 3 shows the investigative team conducts an in-depth examination of the PhishTank platform. As a free, community-driven resource, PhishTank serves

as a collaborative hub where individuals and organizations can submit, verify, and share information about phishing attacks and tactics.

The researchers leverage the open API provided by PhishTank, integrating the platform's anti-phishing data and capabilities into their own investigation tools and analytical workflows. This allows the team to thoroughly analyze reported phishing attempts, gather insights on emerging threats, and contribute their findings back to the broader PhishTank community.

By studying the PhishTank model and exploring the ways in which researchers can leverage its resources, the investigative team aims to uncover best practices and innovative approaches for enhancing the collective defense against phishing and other cyber threats.

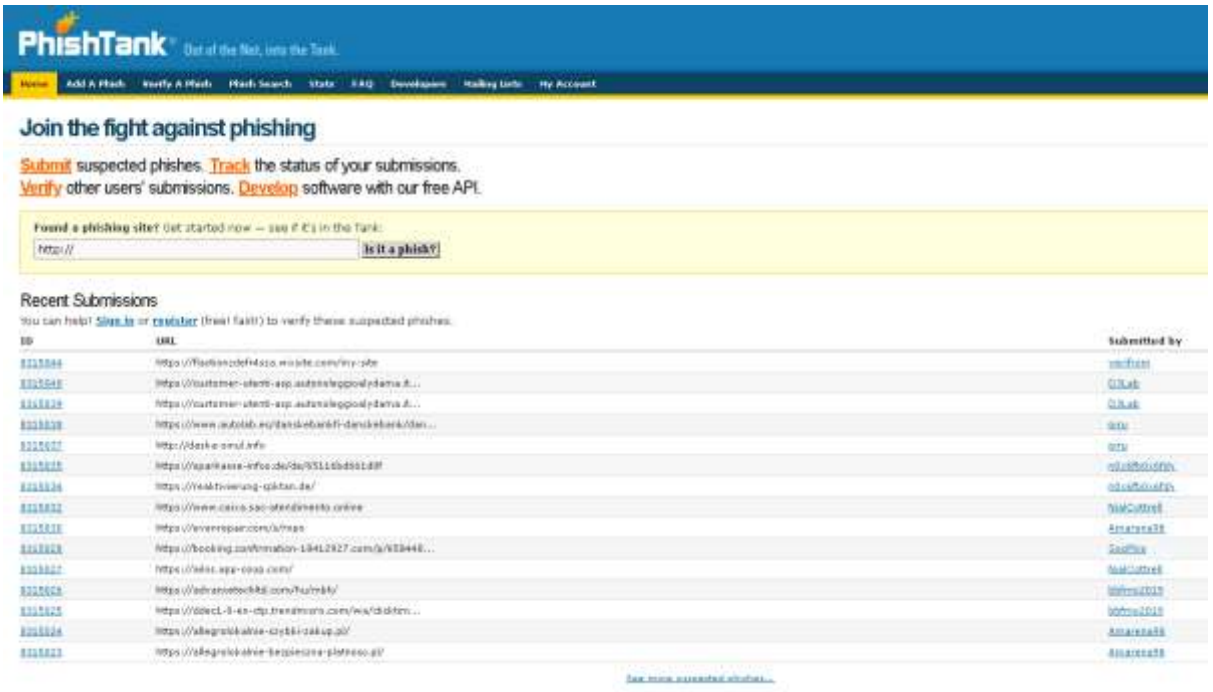


Figure 17 phish tank examination result

Autopsy

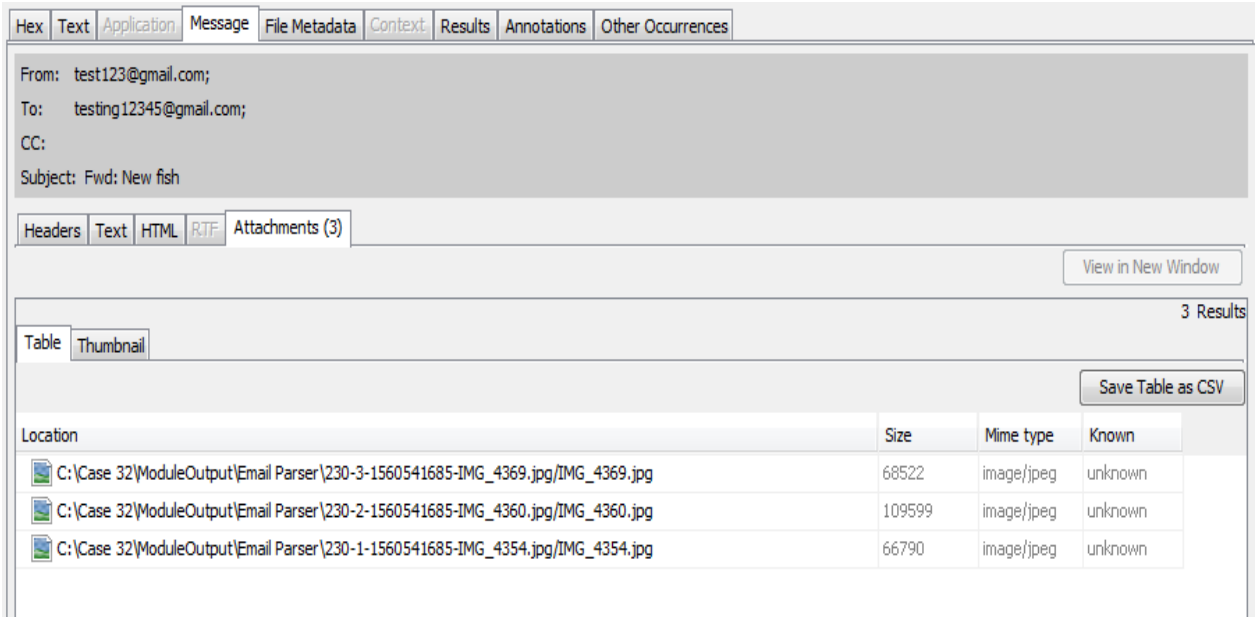
The Email Parser module identifies MBOX, EML and PST format files based on file signatures, extracting the e-mails from them, adding the results to the Blackboard. This module skips known files and creates a Blackboard artifact for each message. It adds email attachments as derived files. This allows the user to identify email-based communications from the system being analyzed.

In this research case study, the researcher examine the utilization of Autopsy, a widely adopted open source digital forensics platform, in the context of email forensics investigation. Autopsy serves as a graphical user interface (GUI) that facilitates the investigative process by integrating various open source tools, notably including The Sleuth Kit. With its user-friendly environment, Autopsy offers a comprehensive solution for analyzing email artifacts, encompassing Mbox files, PST files, and individual email messages. Key functionalities provided by Autopsy include

keyword searching, timeline analysis, email threading, attachment viewing, and metadata extraction. By streamlining the investigation process, Autopsy empowers digital forensics practitioners to gain a holistic and in-depth view of email-related evidence, thereby enhancing the effectiveness and efficiency of email forensics examinations. Integrating MongoDB with Autopsy involves using MongoDB to store data extracted from forensic investigations and then utilizing Autopsy for analysis. Although Autopsy lacks native functionality for directly importing data from MongoDB, this can be fixed by writing a Python script.

```python
import pandas as pd
from pymongo import MongoClient

# Connect to MongoDB
client = MongoClient('mongodb://localhost:27017/')
db = client['forensic_data']
collection = db['extracted_data']

# Fetch data from MongoDB
data = pd.DataFrame(list(collection.find()))

# Save to CSV for Autopsy
data.to_csv('path/to/export_for_autopsy.csv', index=False)
print("Data exported to CSV for Autopsy.")
```

Figure 18 python script for fetch data from mongo dB



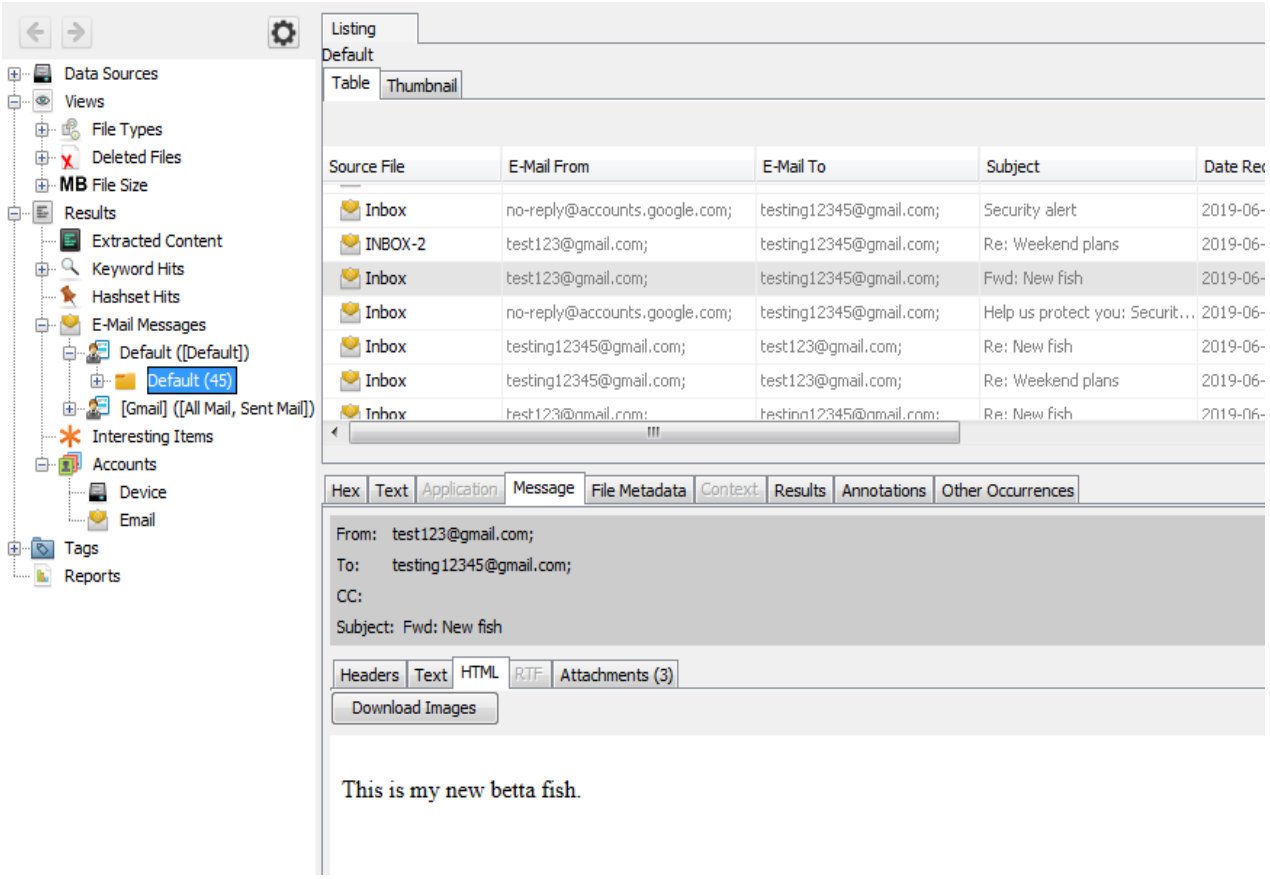


Figure 19 autopsy examination result

NetworkMiner

NetworkMiner is an open source network forensics tool that extracts artifacts, such as files, images, emails and passwords, from captured network traffic in PCAP files. NetworkMiner can also be used to capture live network traffic by sniffing a network interface. Also NetworkMiner parsers for POP3 and IMAP as well as an improved SMTP parser. These are the de facto protocols used for sending and receiving emails Not only does NetworkMiner show the contents of emails within the tool, it also extracts all attachments to disk and even saves each email as an .eml file that can be opened in an external email reader in order to view it as the suspect would.

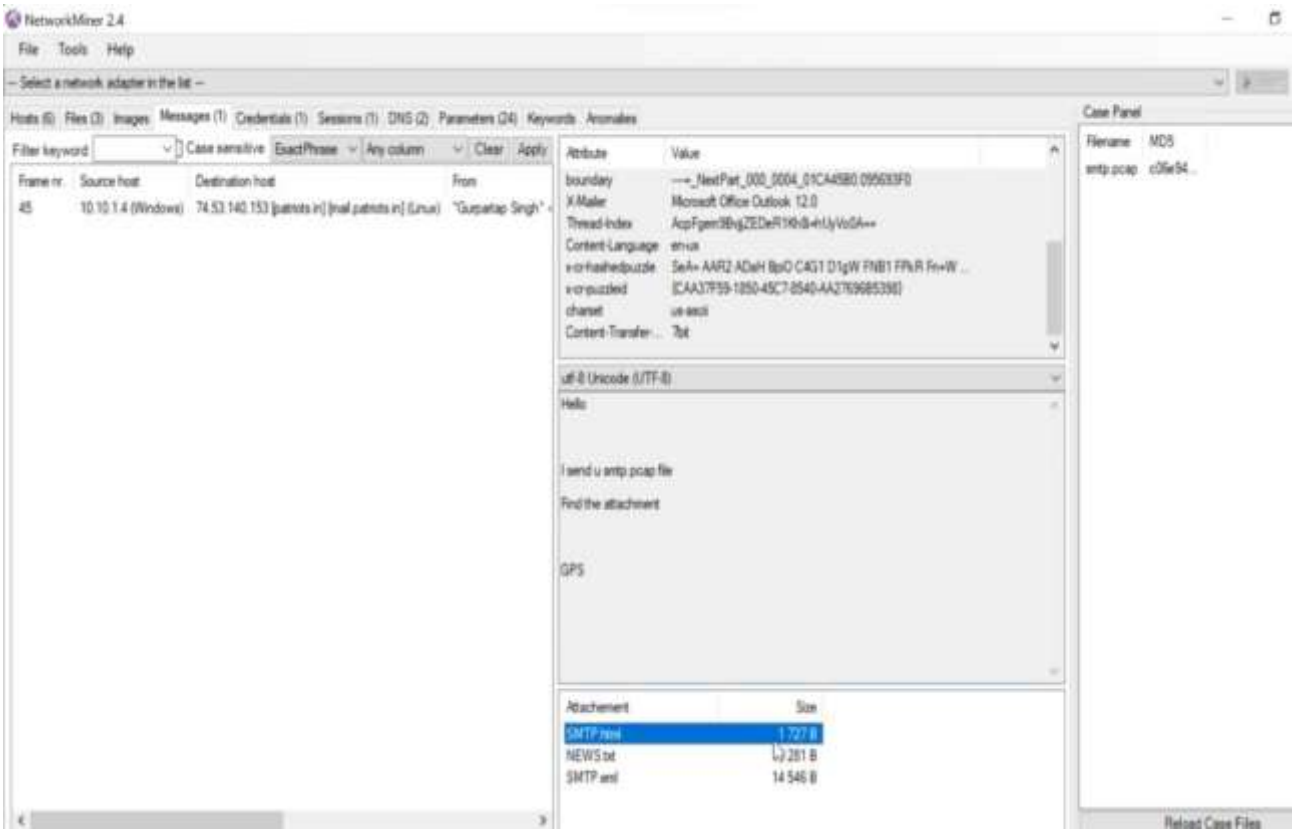
In this research case study, the researcher explore the application of NetworkMiner, an open source network forensic analysis tool, in the context of email forensics investigation. NetworkMiner offers a user-friendly interface and a range of features that aid in the analysis of network traffic and the extraction of relevant information.

In this investigation, NetworkMiner was employed to capture and parse network traffic data related to email communications. The tool facilitated the reconstruction of TCP/IP streams, allowing us to analyze the content of individual packets and gain insights into the email exchanges taking place on the network.One of the significant capabilities of NetworkMiner is its ability to extract various types of data from captured packets, including email messages, attachments, and other associated artifacts. This feature proved invaluable in our email forensics examination, as it enabled the retrieval and analysis of email content that may contain crucial evidence.

Additionally, NetworkMiner facilitated the extraction of metadata from network traffic. The tool automatically extracted metadata such as IP addresses, MAC addresses, domain names, and user agents. This metadata played a crucial role in identifying the origin and destination of network communications, establishing timelines, and linking communications to specific entities or individuals involved.

The file reconstruction feature of NetworkMiner further enhanced our investigation. It enabled the identification and extraction of email attachments transmitted over the network, allowing for a comprehensive analysis of the content within those attachments. This capability proved particularly useful in uncovering potential evidence that may have been concealed within email attachments. Moreover, NetworkMiner offered additional analysis features such as DNS analysis, host and domain name resolution, and keyword searching within captured data. These features assisted in identifying patterns,

relationships, and relevant information within the network traffic, contributing to the overall effectiveness of our email forensic examination.



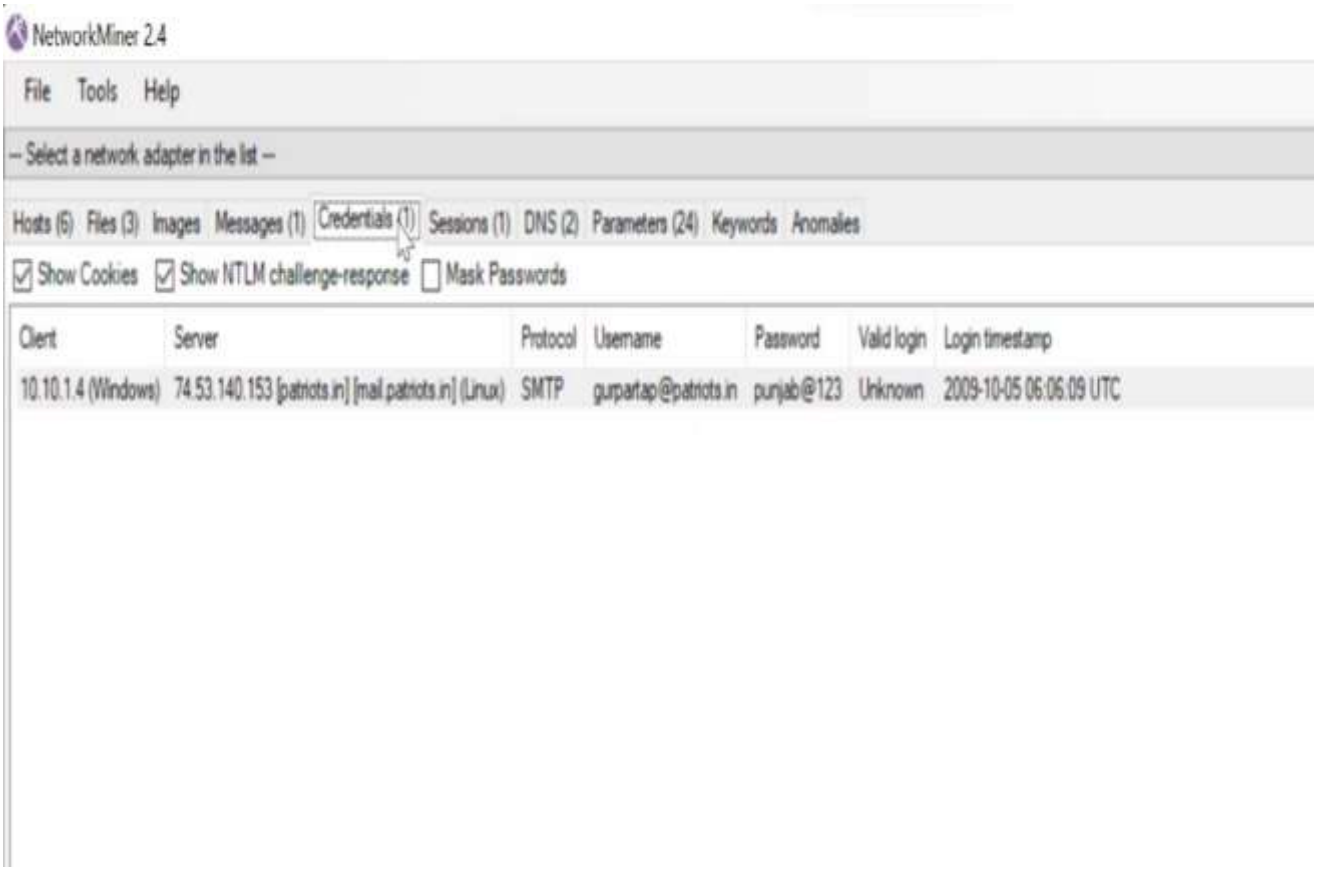


Figure 20network miner examination result

```
import pandas as pd
from pymongo import MongoClient

client = MongoClient('mongodb://localhost:27017/')
db = client['network_data']
collection = db['extracted_info']

data = pd.DataFrame(list(collection.find()))

data.to_csv('path/to/export_for_networkminer.csv', index=False)
print("Data exported to CSV for NetworkMiner.")
```

Wireshark

Wireshark is available free, is open source, and it is used to capture and analyze network traffic. Wireshark captures the bits from the NIC card of system and process them to show us in standard TCP/IP referenced layer model. WireShark will the packet captures analyzed through email protocol (SMTP (- port 25, SMTPs - port 587, POP3 - port 110, POP3s - port
995, IMAP - port 143, IMAPs- port 993) analyzer display information such as IP addresses, directional flow of data, data amounts being transferred, protocols being used and plain text communications. To view SMTP traffic, enter the SMTP filter in

Wireshark. In this example, the researcher can see: Sender email address, Recipient email address, Sender first and last name, Subject line of the email, Body of the email

In this research case study, the researcher investigate the use of Wireshark, an open source network protocol analyzer, in the context of email forensics investigation. Wireshark offers powerful capabilities for capturing, analyzing, and interpreting network packets, providing valuable insights into email communications and associated artifacts. In this investigation, Wireshark was utilized to capture network traffic containing email exchanges. By capturing and analyzing SMTP, POP3, IMAP, and other email-related protocols, Wireshark allowed us to reconstruct email conversations and examine the underlying network packets.

One of the key functionalities of Wireshark is the ability to inspect and analyze email headers. By examining the header information within captured packets, the researcher were able to gain critical details such as sender and recipient addresses, subject lines, timestamps, and other relevant metadata. This information played a crucial role in establishing the context and timeline of email communications. Wireshark's packet-level analysis capabilities were instrumental in our email forensics investigation. The researcher were able to dissect individual packets and extract email content, including message bodies and attachments. This enabled us to retrieve and examine the actual content of the emails, thus uncovering potential evidence and insights.

Furthermore, Wireshark facilitated the identification of potential malicious activities or data leaks through network-based email channels. By analyzing the network traffic, the researcher were able to detect anomalies, identify suspicious patterns, and pinpoint any unauthorized or unauthorized access to email systems. This helped us in uncovering potential security breaches and incidents related to email communications. Wireshark's graphical interface and filtering capabilities were advantageous in managing and analyzing large volumes of network data. The ability to apply filters based on specific criteria, such as IP addresses, email addresses, or keywords, allowed us to narrow down our focus and extract relevant information efficiently. In conclusion, this research case study demonstrates the significant role of Wireshark in email forensics investigations. By capturing and analyzing network packets, Wireshark provides valuable insights into email exchanges, metadata, message content, and potential security issues. Its packet-level analysis capabilities, filtering options, and intuitive interface make it a powerful tool for digital forensic practitioners investigating email-related incidents. By leveraging Wireshark's capabilities, investigators can

enhance their ability to uncover evidence, detect anomalies, and gain a comprehensive understanding of email communications within a network environment.

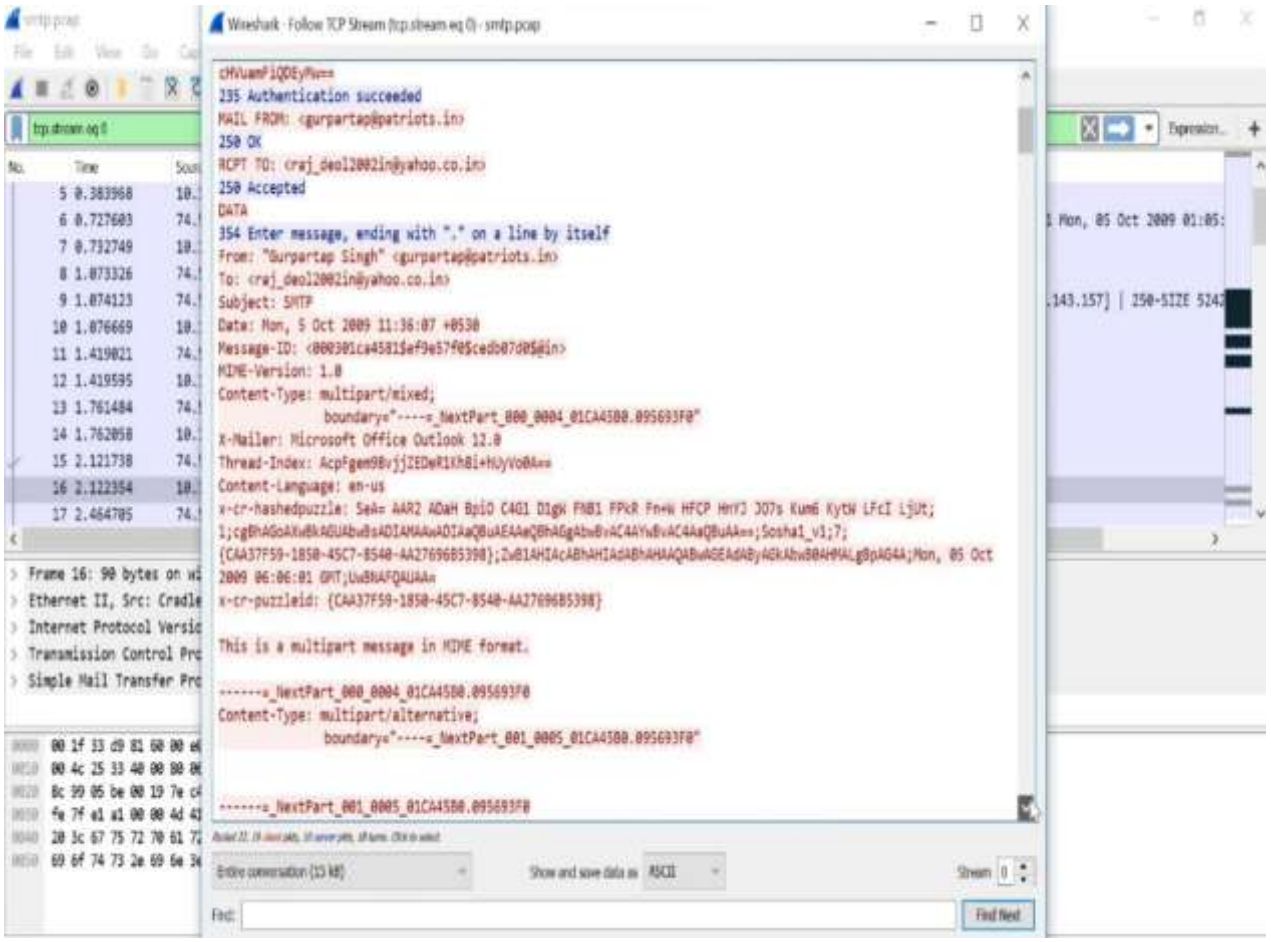


Figure 21 wireshark examination result

```
import pandas as pd
from pymongo import MongoClient

data = pd.read_csv('path/to/extracted_data.csv')

client = MongoClient('mongodb://localhost:27017/')
db = client['network_data']
collection = db['packet_metadata']  # Define your collection name

collection.insert_many(data.to_dict('records'))
print("Metadata successfully inserted into MongoDB.")
```

Figure 22 store data to mongodb

Kernel Outlook PST Viewer

Kernel for Outlook PST Viewer is a free tool that scans, opens and displays corrupt as well as healthy PST file items such as emails, calendars. Features of tool include Open and View Mutiple PST files Data Open Email attachments View deleted items No file size limitation Supports all

Outlook versions The tool generates File analysis reports, where you can see the details of the PST, like Interaction between users, Total item types, & Mail flow density (by date/senders). So, you can analyze the content of PST files easily

In this research case study, the researcher explore the integration of Kernel Outlook PST Viewer, a specialized software tool, within an email forensics investigation framework. Kernel Outlook PST Viewer offers advanced capabilities for analyzing and examining PST files, a common file format for storing email data in Microsoft Outlook.In this investigation, Kernel Outlook PST Viewer played a pivotal role in the analysis and examination of PST files obtained from a target email system. The tool provided an intuitive and user-friendly interface; allowing forensic examiners to navigate through the PST file structure and access its contents with ease.

One of the key functionalities of Kernel Outlook PST Viewer is its ability to extract and display various email artifacts contained within PST files. This includes emails, attachments, contacts, calendars, tasks, and other related information. By utilizing the viewer's features, the researcher were able to examine the individual emails and their associated metadata, such as sender and recipient details, timestamps, subject lines, and message content.

Kernel Outlook PST Viewer also facilitated the extraction and export of email attachments from PST files. This feature was instrumental in retrieving and analyzing file attachments that may contain critical evidence for the investigation. The ability to export attachments in their original format or convert them to a different format enhanced the flexibility of analysis process.

Furthermore, Kernel Outlook PST Viewer provided search and filtering capabilities, enabling us to locate specific emails or keywords within the PST file. This functionality proved valuable in narrowing down our focus and identifying relevant email communications that were pertinent to the investigation. The viewer's ability to generate reports and export selected data further enhanced this email forensics examination. The researcher were able to generate comprehensive reports containing details such as email metadata, attachments, and other relevant information. These reports served as valuable documentation for this investigation findings.

In conclusion, this research case study highlights the significance of Kernel Outlook PST Viewer in an email forensics investigation framework. By leveraging its capabilities for PST file analysis, email extraction, attachment examination, search and filtering, and report generation, Kernel Outlook PST Viewer played a crucial role in uncovering evidence and providing valuable insights into email communications. The tool's user-friendly interface and comprehensive feature set make it a valuable asset for digital forensic practitioners engaged in email forensics examination

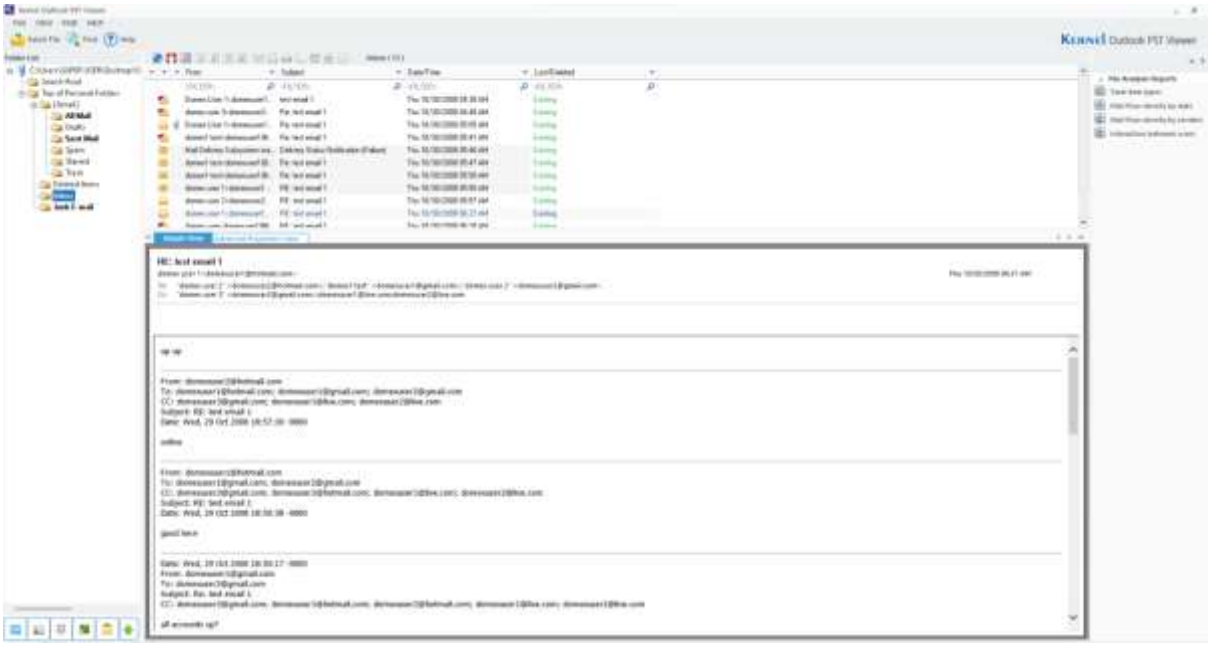


Figure 23Kernel for Outlook PST Viewer xamination result

4n6 Email Forensics Tool

The 4n6 Email Forensics Tool is a robust approach for all major types of email applications dealing with mail data files. The Email forensic converter deals with different types of files that are compliant with more than 60 email clients. PST, OST, EML, MBOX, and MSG include popular email forms provided by 4n6 Email Analyser tools. In individual or bulk mode, users can conveniently export them to other suitable formats. Selective emails or bulk emails are conveniently exported to any other file format, based on the user's preference. It includes different forms of exports such as PST, EML, MSG, TXT, PDF, VCF, vCard, CSV, ICS, HTML, and Gmail, Office 365 and IMAP email services. Customers can scan, check and examine email addresses, attachments, phone numbers, and retrieve them at once. To open it and save it, you can right-click the attachment. Get details like the header of the email, topic, address, start address, date, time, cc, bcc, etc. Click the Save All option if the email includes several attachments. The user can quickly locate or scan for unique word or text content in email data elements by using this optimal email forensic software. An additional option to scan for similar types of messages based on addresses, contacts, schedules, assignments, documents, or this email forensic wizard offers much more.

In this research case study, the researcher explore the utilization of the 4n6 Email Forensics Tool within an email forensics investigation framework. The 4n6 Email Forensics Tool is a specialized software solution designed to assist digital forensic examiners in the analysis and examination of email-related artifacts.

In our investigation, the 4n6 Email Forensics Tool played a crucial role in the comprehensive analysis of email data and associated evidence. The tool provided a wide range of features and functionalities that facilitated the extraction, analysis, and interpretation of email artifacts.

One of the key capabilities of the 4n6 Email Forensics Tool is its ability to process various email file formats, including PST, OST, EDB, MBOX, and EML. This versatility allowed us to handle different types of email data encountered during the investigation, ensuring a comprehensive coverage of potential evidence sources.

The tool's advanced parsing algorithms enabled the extraction of email metadata, such as sender and recipient details, timestamps, subject lines, and message content. This information was crucial in establishing the context, timeline, and participants involved in email communications.

Furthermore, the 4n6 Email Forensics Tool provided advanced search and filtering functionalities. The researcher were able to conduct keyword searches within email bodies, attachments, and other metadata fields, enabling us to pinpoint specific evidence and relevant communications.

The tool also offered data visualization capabilities, allowing us to analyze email relationships, communication patterns, and email thread structures. This visualization aided in understanding the flow of email conversations and identifying key individuals or entities involved.

In addition to email content, the 4n6 Email Forensics Tool facilitated the extraction and analysis of email attachments. The researcher were able to examine file attachments, including their metadata and content, which proved valuable in uncovering potential evidence or identifying hidden information within attachments.

The tool's reporting capabilities were instrumental in documenting these investigation findings. The researcher were able to generate comprehensive reports containing email metadata, message content, attachment details, and other relevant information. These reports served as valuable documentation for legal proceedings or further analysis.

In conclusion, this research case study demonstrates the effectiveness of the 4n6 Email Forensics Tool in an email forensics investigation framework. By utilizing its capabilities for email artifact extraction, metadata analysis, search and filtering, data visualization, attachment examination, and reporting, the tool provided valuable insights into email communications and assisted in uncovering critical evidence. The 4n6 Email Forensics Tool proves to be a valuable asset for digital forensic examiners engaged in email forensics investigations, enhancing their ability to conduct thorough and comprehensive examinations of email-related artifacts

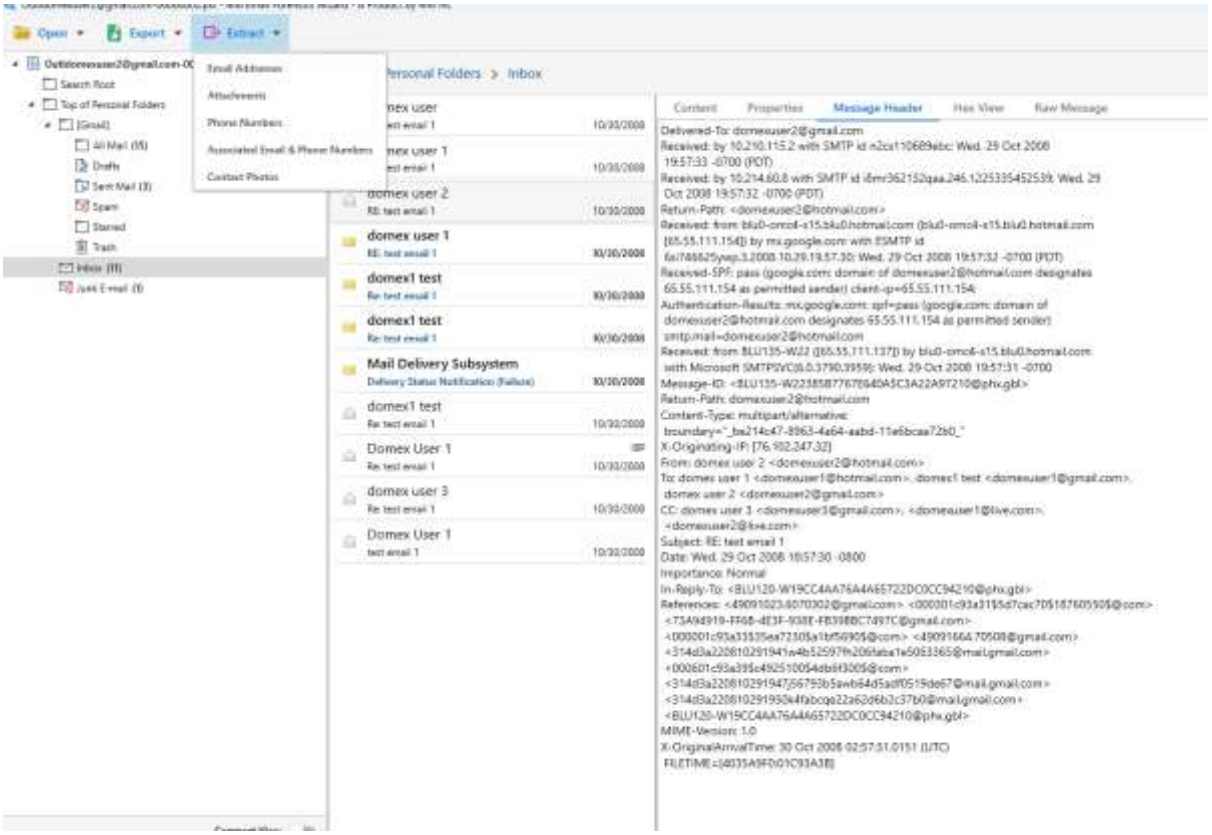


Figure 24 *4n6 Email Forensics Tool examination result*

Link Analysis

In this case study, NetworkX was applied to an email forensic investigation within a corporate environment, where the goal was to uncover fraudulent email communications between a suspected employee and clients. The investigation began with the collection of relevant email data, including metadata and communication logs. Using NetworkX, investigators created a directed graph where email addresses were represented as nodes and the email exchanges as directed edges. This graph allowed for the visualization of communication patterns between the suspect and clients, as well as the identification of central figures through centrality measures like degree and betweenness centrality. By analyzing these metrics, investigators could identify the suspect as a key player within the network, as well as detect any unusual or suspicious communication patterns, such as spikes in activity during odd hours. Link analysis revealed hidden relationships between the suspect, clients, and other employees, uncovering potential coconspirators. The investigation also utilized temporal analysis to detect anomalies, such as sudden bursts of communication that could indicate fraudulent activity. Finally, the findings, including key relationships, suspicious patterns, and visualizations, were presented in a comprehensive report, which included a network diagram that illustrated the connections between

the suspect and other entities. This helped clarify the suspect's role in the fraudulent scheme and provided clear evidence that could be used in legal proceedings or internal action

```
In [ ]: import networkx as nx
        import matplotlib.pyplot as plt
        from networkx.algorithms import community

In [ ]: raw_mail_data = pd.read_csv('E:\thesis\spam_mail_classification-master\dataset.csv')

In [ ]: emails = [
            {"from": "alice@example.com", "to": "bob@example.com", "date": "2023-01-01"},
            {"from": "bob@example.com", "to": "carol@example.com", "date": "2023-01-02"},
            {"from": "alice@example.com", "to": "carol@example.com", "date": "2023-01-03"},
            {"from": "carol@example.com", "to": "dave@example.com", "date": "2023-01-04"},
            {"from": "dave@example.com", "to": "alice@example.com", "date": "2023-01-05"},
        ]

In [ ]: # Initialize a directed graph
        G = nx.DiGraph()

In [ ]: # Add edges based on the email data
        for email in catagory:
            G.add_edge(email["from"], email["to"])

In [ ]: # Display basic information about the graph
        print("Number of nodes:", G.number_of_nodes())
        print("Number of edges:", G.number_of_edges())

In [ ]: # Calculate degree centrality
        degree_centrality = nx.degree_centrality(G)
        print("Degree Centrality:", degree_centrality)

In [ ]: # Identify the user with the highest degree centrality
        most_influential = max(degree_centrality, key=degree_centrality.get)
        print("Most influential user:", most_influential)

In [ ]: # Find communities using the Girvan-Newman method
        comp = community.girvan_newman(G)
        top_level_communities = next(comp)
        print("Detected communities:", list(top_level_communities))

In [ ]: # Visualize the email communication network
        plt.figure(figsize=(10, 6))
        pos = nx.spring_layout(G)  # positions for all nodes
        nx.draw(G, pos, with_labels=True, node_color='lightblue', font_weight='bold', arrows=True)
        plt.title("Email Communication Network")
        plt.show()
```

Figure 25 Network X library

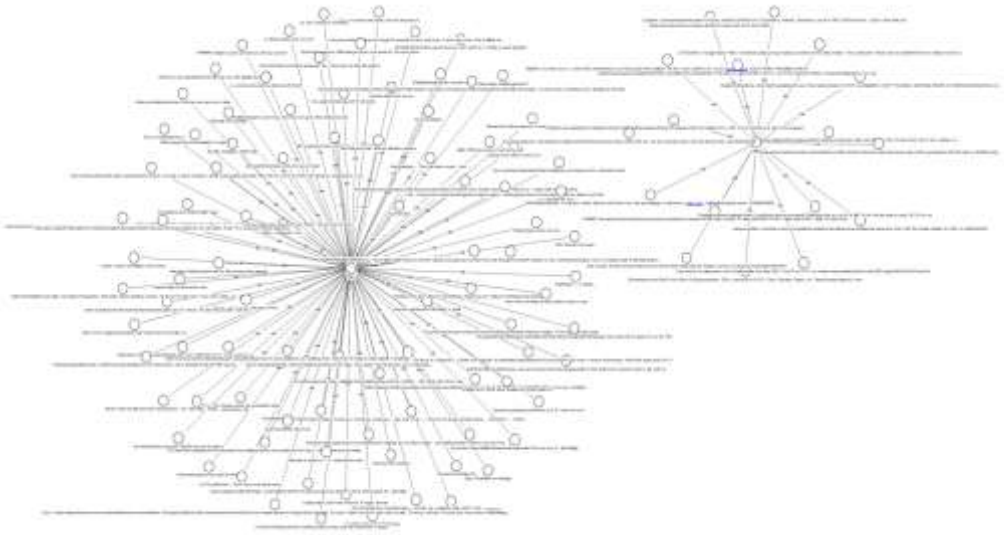


Figure 26 email link graph

Open source intelligence

In the email forensic framework, the insights gained from the Review stage can be significantly enhanced by using third-party tools like VirusTotal. For example, investigators can take an email link and enter it into VirusTotal to check its safety and identify potential threats. This platform enables the analysis of files and URLs, compiling data from various antivirus engines and other sources. By integrating VirusTotal into the investigation, along with both open and closed source data, investigators can gain a deeper understanding of the potential threats linked to email content. The results from VirusTotal, combined with the insights from the Review stage, can then be incorporated into the Evidence Analysis step. This method improves the knowledge base needed to evaluate the evidential value and relevance of the information, ultimately leading to more informed investigative outcomes.

For demonstration purposes, researchers can take an email link, input it into VirusTotal, and conduct a thorough analysis to check for safety and potential threats. After performing this test, the researcher should present the results, highlighting any detected risks or security issues associated with the link. This process not only enhances the investigation but also provides concrete evidence for assessing the email's credibility. By incorporating findings from VirusTotal alongside both open and closed source data, investigators can deepen their understanding of potential threats linked to the email content.

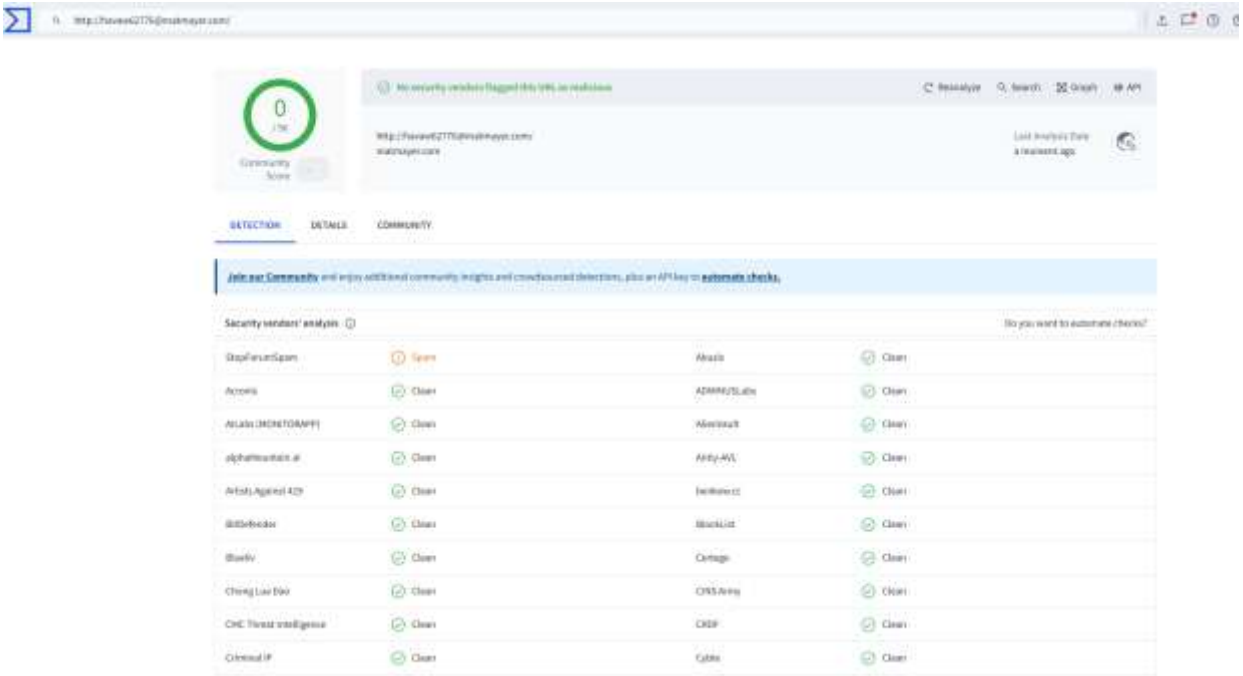


Figure 27 virus total

Case box

Case box is an open-source platform for case management and workflow automation. It provides a web-based interface that allows users to manage and organize cases, evidence, documents, and tasks. Here is a general overview of how Case box works:

Start by installing Case box on a web server or a local machine. Case box is based on the Drupal content management system, so Investigator will need to set up a web server environment with PHP and a MySQL database. The installation process involves downloading the Case box files, configuring the server environment, and running the installation script .and then Creating Cases: Once Case box is set up; you can create cases to manage your investigations. A case represents a specific investigation or project. You can assign a case number, title, description, and other relevant information to each case.

Managing Evidence: Within each case, you can manage evidence related to the investigation. Case box allows you to upload and associate various types of files, such as documents, images, videos, and emails, with the corresponding case. You can organize evidence into folders and attach metadata to provide additional context.

Task Management: Case box offers task management features to track and assign tasks related to the investigation. You can create tasks, assign them to team members, set due dates, and track their progress. This helps in coordinating the investigation and ensuring that important actions are completed. Collaboration and Communication: Case box facilitates collaboration among team members working on the investigation. Users can communicate through comments, discussions, and notifications within the platform. This enables seamless sharing of information, updates, and insights related to the case.

Customization and Workflow Automation: Case box allows customization to adapt the platform to the specific needs of your investigations. You can create custom fields, forms, and templates to capture and organize case-specific data. Additionally, you can automate workflows by defining rules and triggers to streamline processes and ensure consistency.

Reporting and Analysis: Case box provides reporting capabilities to generate summaries, statistics, and analysis based on the data stored in the platform. You can create custom reports and visualizations to gain insights into the investigation, monitor progress, and present findings.

In this research case study, the researcher explores the integration of Case box as an email forensics tool within an email forensics investigation framework. Although Case box is primarily known as a platform for case management and workflow automation, the researcher investigates its applicability and effectiveness in supporting email forensics investigations.

Case box, with its web-based interface and robust features for case management, evidence organization, and task management, can be adapted to support email forensics investigations. Here is an overview of how Case box can be utilized in an email forensics investigation framework:

Case Creation and Management: Investigators can create a case within Case box dedicated to an email forensics investigation. They can provide relevant case details, assign case numbers, and categorize the case based on its nature or subject.

Evidence Collection and Organization: Case box allows investigators to upload and manage email-related evidence within the created case. This includes email artifacts, such as PST, OST, MBOX, or EML files, which can be securely stored and organized within the platform's document management system.

Metadata Extraction and Analysis: Investigators can utilize Case box's capabilities to extract and analyze email metadata. The platform can parse and display important metadata fields, such as sender and recipient details, timestamps, subject lines, and message IDs, allowing investigators to examine the communication patterns and relationships within the case.

Keyword Search and Filtering: Case box's search and filtering functionality can assist investigators in identifying relevant emails or specific keywords within email bodies and attachments. This feature streamlines the process of locating pertinent evidence and facilitates efficient analysis.

Collaboration and Task Management: Case box's collaboration features enable multiple investigators to work together on the same case. Investigators can assign tasks, set deadlines, and track progress within the platform, promoting efficient teamwork and ensuring the investigation proceeds smoothly.

Reporting and Documentation: Case box offers reporting capabilities, allowing investigators to generate comprehensive reports summarizing the findings of the email forensics investigation. These reports can include email metadata, extracted content, attachments, and any other relevant information. The platform's document management system ensures the secure storage and version control of these reports.

It is important to note that while Case box can provide a framework for managing and organizing email artifacts and associated metadata, it may require customization and integration with specialized email forensics tools for in-depth analysis and extraction of email content, forensic artifacts, and advanced analysis techniques.

In conclusion, this research case study demonstrates the potential of Case box as a supportive tool within an email forensics investigation framework. Leveraging its features for case

management, evidence organization, metadata extraction, search and filtering, collaboration, and reporting, Case box can enhance the efficiency and organization of email forensics investigations. However, it is advised to complement Case box with specialized email forensics tools to address the more advanced analysis requirements of email content and forensic artifacts

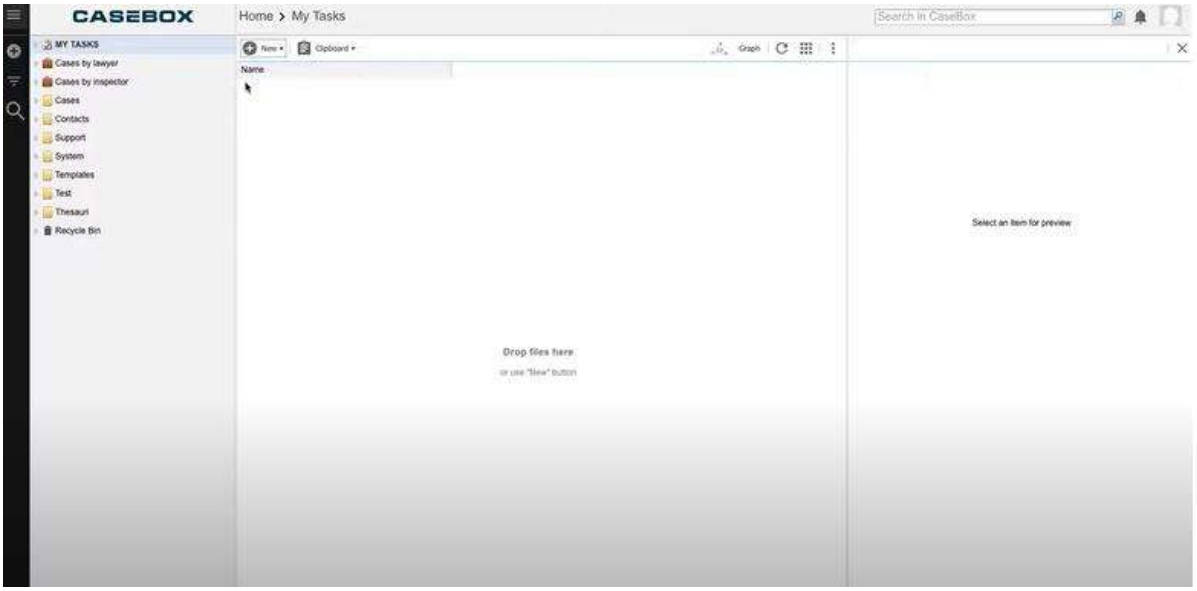


Figure 28 case study

Elasticsearch

Elastic search is integral to the proposed email forensics framework, enhancing the ability to search, retrieve, and analyze email evidence. By providing advanced search capabilities, Elasticsearch allows forensic analysts to quickly locate relevant emails based on various criteria such as sender, recipient, date, keywords, and content. Its scalable architecture is designed to efficiently handle large volumes of data, ensuring optimal search performance even as email datasets grow. The real-time indexing feature enables new email data to be indexed immediately, allowing analysts to access the most current information immediately, which is critical during active investigations.

Integrating the system in this study entails integrating processed email data into Elastic search, usually after it has been extracted and processed using tools like Pandas and MongoDB. The data retrieval process can be greatly accelerated by analysts using sophisticated queries to retrieve particular emails once they have been indexed. To speed up decision-making, they can, for example, search for emails from specific senders or that contain particular terms. Elastic search offers more advantages than just search; it facilitates data aggregation and analytics, allowing analysts to provide insights like detecting repeat senders or examining patterns over time.

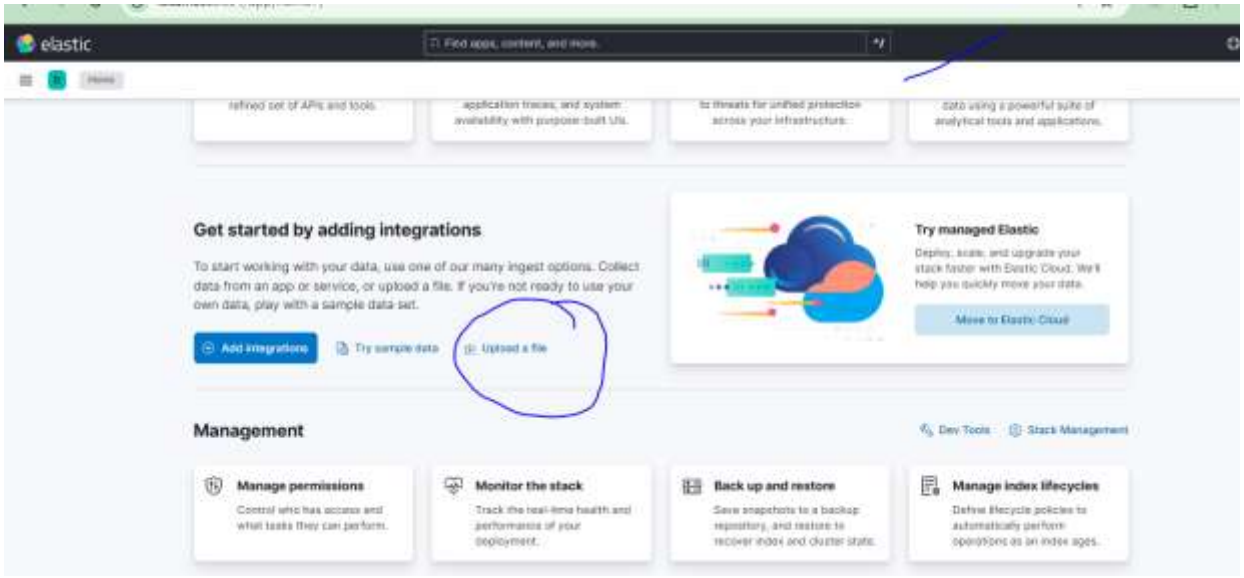


Figure 29 elastic search configeration

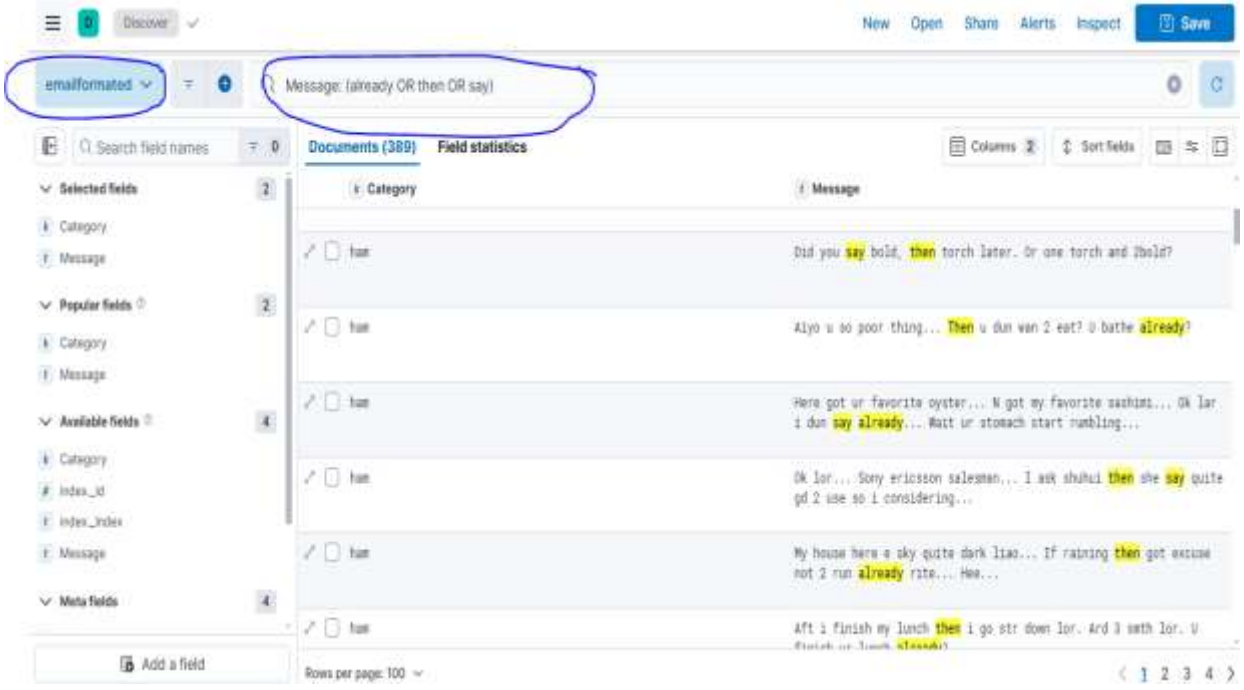


Figure 30elastic search searching capability

Figure 31 dashboard

5.2 Proposed Framework Evaluation

DS-based frameworks may be assessed based on several criteria, including plausibility, efficacy, practicality, predictability, comprehensiveness, scalability, and simplicity of use [17]. Digital forensic specialists using an interview checklist based on certain criteria examined the suggested framework. Experts were given "Yes" or "No" alternatives to evaluate the framework based on the evaluation criteria.

No	Evaluation Criterions	Description of Criteria	Experts evaluation %
1	Effectiveness	❖ Ability to comprehensively extract and process email data from various sources and formats ❖ Accuracy and completeness of the extracted email metadata, message content, attachments, and embedded artifacts ❖ Effectiveness in detecting and analyzing email-borne threats, such as phishing attempts and malware	95%
2	Efficient	❖ Throughput and processing speed for ingesting and analyzing large email datasets ❖ Optimization of resource utilization (CPU, memory, storage) to ensure scalable and high-performance operations ❖ Streamlined and automated investigative workflows to minimize manual intervention	89%
3	Feasibility	❖ Can be operationalized or implemented as described	87.5%
4	Extensibility and Adaptability	❖ Modular and loosely coupled architecture to enable easy integration of new tools and capabilities ❖ Flexibility in accommodating changes in email data formats, threat patterns, and investigative requirements ❖ Ease of deployment, configuration, and maintenance to ensure long-term sustainability	87.5%

5	Reliability	❖ Fault tolerance and high availability: The framework should be designed to minimize single points of failure, with mechanisms to ensure continued operation in the event of component or system failures. ❖ Data integrity and consistency: Robust data validation, error handling, and transaction management mechanisms should be in place to maintain the integrity and consistency of email forensic data throughout the investigative process.	87.5%
6	Functional requirements	❖ Data ingestion and processing ❖ Investigative capabilities ❖ Reporting and knowledge sharing ❖ Case management	100%
7	Scalability	❖ Horizontal scalability: The framework should be designed to scale out by adding more compute, storage, and network resources to handle growing email data volumes and investigative workloads. ❖ Vertical scalability: The framework should be capable of efficiently utilizing increased hardware resources, such as more powerful CPUs and larger memory capacities, to enhance its overall performance. ❖ Adaptive resource allocation: Intelligent resource management mechanisms should be implemented to dynamically allocate and optimize the utilization of available computing, storage, and network resources based on the changing investigative demands. ❖ Distributed and parallel processing: The framework should leverage distributed computing techniques and parallel processing algorithms to parallelize email data ingestion, processing, and analysis tasks,	87.5%

		thereby improving overall throughput and reducing latency.	
8	User Experience and Usability	❖ Intuitive and user-friendly interface to facilitate efficient and effective investigative workflows ❖ Contextual help, documentation, and training resources to enable smooth onboarding and adoption ❖ Customization options to accommodate the preferences and requirements of different user roles (e.g., researchers, incident responders)	87.5%

Table 4 evaluation table

If the average assessment checklist value exceeds 80%, the system is very useable [18]. Respondents scored an average of 94.2% on the evaluation criteria, indicating that the proposed email forensic framework is effective.

CHAPTER SIX

5 Result and discussion

This section presents the evaluation of the proposed Structured Email Forensic Investigation Framework developed by using the Design Science Methodology and how it can significantly contribute to enhancing efficiency, accuracy, and effectiveness during an email forensic investigation. The testing of the framework using different scenarios, open-source tools, and different case studies was performed. After this, the evaluation results using metrics such as speed of investigation, data quality, user satisfaction, and accuracy of evidence were analyzed. The results prove that it has the potential for solving some crucial problems present in traditional approaches, including inefficiency, data integrity, and handling diverse and encrypted data sources.

A key output of the evaluation is the ability of the framework to significantly reduce the duration of investigation without affecting the quality of evidence. The framework made use of data reduction techniques and targeted identification, preservation, and analysis phases, therefore reducing unnecessary overheads in the processing and storage of data. Thus, using targeted data subsets to triage email reduced the time required for indexing and analysis of the data by a great margin compared to working with the full forensic images. This is further valued in real-life investigations where datasets are usually so large that indexing and analysis conventionally take hours or days.

Another important finding is that the case management aspect of the framework brought order and accountability to the investigation process. The structured presentation of evidence and monitoring of phases in the investigation ensured that data were treated properly and the possibility of mistakes or omission was minimized. Investigators reported increased user satisfaction due to the clarity of the framework, logical workflow, and ease with which it adapted to different types of cases. Further improvements in usability came with the incorporation of open-source tools into the framework, making it as cost-effective and accessible to even low-resourced organizations.

The framework performance was also very effective in handling encrypted emails, metadata, and attachments from various data sources. Evidence extraction and analysis were easily performed through the use of different opensource tools such as dd, MongoDB, Virus total, phishtool, utilities for data mining while ensuring integrity of data. This proves critical for maintaining authenticity in evidence, which is crucial to its legal admissibility. The framework could resolve

issues of data integrity and tampering by embedding techniques for robust encryption and hash verification.

Among the strengths, it identified various challenges-for instance, a steep learning curve with the use of open-source tools and the need to manually set up processes such as data storage configuration or the execution of decryption tasks. Such challenges are indicative of points that need future development, such as automation of routine activities and refinement of training material for investigators without prior experience with particular tools.

The framework also made it possible to garner better intelligence through cross-case analysis and trend identification due to the capability for rapid triaging and analysis of subsets of data. Such speed in identifying patterns, linking cases, and extracting valuable intelligence was possible. Similarly, other relationships among email accounts-automated metadata and communication network analysis-may not have surfaced either; this also illustrates the full extent of the potential of this approach in wider intelligence applications.

Thus, generally speaking, the validation of the presented framework can efficiently help in mitigating problems or challenges that may arise from the investigations related to e-mail forensics. This framework is functional, adaptive, and easily scalable to the needs required for modern digital forensic investigations; such needs reduce investigation time by improving data quality through evidence management. Future research should be directed at fine-tuning automation, broadening the compatibility to cover more tools and data formats, and undertaking long-term research studies that assess the applicability of the framework to a wider range of forensic contexts. Such a position will further establish this framework as a one-stop guide for both practitioners and researchers, furthering efficient, accountable, and adaptive practices in the field of email forensics.

CHAPTER SEVEN

6 Summary, future work

6.1 Summary

The study addresses significant challenges in modern email forensics, including inefficiency, data integrity issues, and the handling of various data sources, such as encrypted emails and metadata, by providing a structured framework for email forensic investigations, designed using the Design Science Methodology. The framework provides an integrated case management system to allow for systematic implementation and proper handling of evidence. It comprises ten phases. The framework is also used with open-source technologies, case studies, and best practices, proving to be flexible in many real-world scenarios. From the review, significant improvements are identified in the happiness of Investigators and data quality, efficacy of inquiry. For instance, the data reduction procedure greatly shortened inquiry times through rapid triaging and prioritization of relevant evidence against specific sub-sets of data. Free solutions such as Email Examiner, Autopsy, dd, and MongoDB allowed handling of big data sets and whatever form of different data format existed; case management warranted accountability and sped up most processes. Besides, through the use of robust hashing and encryption techniques, the framework avoids issues regarding data integrity and ascertains the evidence is valid and admissible in court.

It provides an enabling framework to enhance data quality and access, which provides better information with minimal time required in research at low rates of errors. It is also adaptive in a broad scale of business capacities that range from small-scale start-up businesses up to large investigation agencies. This paradigm advances the status of investigations and encourages effective, accountable, and flexible procedures by offering both theoretical insights and practical answers to the field of email forensics.

6.2 Future Work

The proposed framework provides a broad, practical approach to conducting e-mail forensics. There is considerable scope for future enhancement and development of the original in a number of particular ways. One of these regards the use of artificial intelligence and machine learning to take further automation of analyses within an e-mail dataset. With integrated advanced analytics capabilities, it is possible that the framework will dig deeper into patterns that are more sophisticated, relationships, and anomalies in e-mail communications.

This can further be extended to target the compatibility and interoperability of the framework with an extended set of open-source and proprietary forensic tools to extend flexibility in the choice of various tools and to leverage specialized capabilities. This would therefore imply

continuous benchmarking, hence performance evaluation of different components in the framework, to show ways of optimization and further improvements.

Lastly, further studies may consider exploring how the open-source email forensic framework applies to different organizations, be it corporate, government, or law enforcement agencies, so that there would be a chance to acquire empirical feedback and actual real-world case studies to support and help refine the framework toward greater relevance and effectiveness over an enormous range of email-based investigations.

7. References

[1] Darren Quick, Kim-Kwang Raymond Choo, "Impacts of increasing volume of digital forensic data," *digital invetigation,* vol. 11, no. 4, pp. 273-294, 2024.

[2] Ahmad Ghafarian, Ash Mady and Kyung Park, "AN EMPIRICA0L ANALYSIS OF EMAIL FORENSICS TOOLS," *network security and its application,* 2020.

[3] M. Al-Zarouni, "Tracing E-mail Headers," in *Australian Computer,Network & Information*, 2004.

[4] M. T. Banday, "Analysing E-Mail Headers for Forensic Investigation," *Digital Forensics,* vol. 6, no. 2, pp. 49-64, 2011.

[5] M. T. Banday, "Technology Corner: Analysing E-Mail Headers for Forensic Investigation," *Network Security & Its Applications,* vol. 3, no. 6, 2011.

[6] Charalambous Elisavet, Bratskas Romaios,Koutras Nikolaos,Karkas George,Anastasiades Andreas, "Email forensic tools: A roadmap to email header analysis through," *Safety and Reliability Association Summer Safety and Reliability Seminars,* vol. 7, 2016.

[7] Charalambous ElisavetBratskas, Romaios,Koutras Nikolaos,Karkas George,Anastasiades Andreas, "Email forensic tools: A roadmap to email header analysis through a cybercrime use case," Safety and Reliability Association," *Polish Safety and Reliability Association,* vol. 7, pp. 21-28, 2016.

[8] D. Q. a. K.-K. R. Choo, "Data reduction and data mining framework for digital forensic evidence: Storage, intelligence,," *crime and criminal justice,* vol. 408, 2014.

[9] Darren Quick , Darren Quick a, "Digital forensic intelligence: Data subsets and Open Source Intelligence," vol. 78, pp. 558-567, 2018.

[10] Gauri Manglik and Vandana Pai, "internet crimes: effectiveness of the laws in force," *cyber crime,* vol. 1, pp. 45-46, 200.

[11] R. Proposal, "Design and Development of E-mail Security Protocols and Forensic Tools," in *International Conference on Recent Advances in Electronics and Computer Engineering*, Eternal University, Himachal Pradesh, 2011.

[12] vamshee Krishna Devendran, Hossain Shahriar.,Victor A. Clincy,, "A Comparative Study of Email Forensic Tools," *information security,* vol. 6, no. 2, pp. 111-117, 2015.

[13] M. T. Banday, "Techniques and Tools for Forensic Investigation of E-mail," *Network Security & Its Applications,* vol. 3, no. 6, 2011.

[14] W. A. Baroto, "Email Analysis in Fraud Investigation: Digital Forensic and Network Analysis Approach," *Asia Pacific Fraud,* vol. 6, no. 2, p. 17, 2011.

[15] Manjeet Singh, Jacob Anwar Husain, Navneet Kumar Vishwas, "A Comprehensive Study of Cyber Law and Cyber Crimes," *Engineering and Applied Sciences Research,* vol. 3, no. 2, pp. 20-24, 2014.

[16] Vibhuti Narayan Singh1,shalini, "Forensic Investigation of Email ARTEFACTS by using various Tools," *Scientific Research & Development,* vol. 2, no. 12, 2015.

[17] Vamshee Krishna Devendran, Hossain Shahriar, Victor Clincy, "A Comparative Study of Email Forensic Tools," *Information Security,* vol. 6, no. 2, pp. 111-117, 2015.

[18] NIST, NIST Special Publication 800-86, NIST, 2023.

[19] ISO, "ISO/IEC 27032:2023," p. 28, 2023.

[20] D. J. W. QPM, "ACPO Good Practice Guide," 2012.

[21] Rachid Hadjidj, Mourad Debbabi*, Hakim Lounis, Farkhund Iqbal, "Towards an integrated e- mail forensic analysis framework,," *Computer Security Laboratory,* vol. 5, no. 3-4, pp. 124-137, 2009.

[22] K. Reddy, H. Venter, "The architecture of a digital forensic readiness management system," *computer science ,* vol. 32, pp. 79-83, 2013.

[23] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger and Samir Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Management Information Systems,* vol. 24, pp. 45-77, 2008.

[24] Isa Ismail, Khairul Akram Zainol Ariffin, "Open Source Tools for Digital Forensic Investigation: Capability, Reliability, Transparency and Legal Requirements," *KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS,* vol. 18, no. 9, pp. 2692-2712, 2024.

[25] 2. a. K. A. Z. A. Isa Ismail1, "Open Source Tools for Digital Forensic Investigation: Capability, Reliability, Transparency and Legal Requirements," *KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS ,* vol. 18, p. 2692, 2024.

[26] Stacie Clarke Petter, Deepak Khazanchi, John D. Murphy , "A Design Science Based Evaluation Framework for Patterns," *INFORMATION SYSTEMS AND QUANTITATIVE ANALYSIS,* vol. 41, no. 3, pp. 9-26, 2010.

[27] S. P. Goffnett, "High performance quality management systems and work-related outcomes: Exploring the role of," 2007.

[28] Thomas Jackson†*, Ray Dawson† and Darren Wilson,, "The Cost of Email Interruption,”," *Computer Science Department,* vol. 5, pp. 81-92, 2019.

[29] Sumanjit Das and Tapaswini Nayak, "IMPACT OF CYBER CRIME: ISSUES AND CHALLENGES," *Engineering Sciences & Emerging Technologies,* vol. 6, no. 2, pp. 142-153, 2013.

[30] Saiprethi and 2Aswathy Rajan, "A Critical Study on Email Related Crimes," *Pure and Applied Mathematics,* vol. 119, no. 17, pp. 1781-1795, 2018.

[31] M. C. W. W. Cohen, "https://www.cs.cmu.edu," CALO Project, 8 May 2015. [Online]. Available: https://www.cs.cmu.edu/~enron/. [Accessed 8 May 2015].

[32] U. M. Mbanaso, PhD and E.S. Dandaura, PhD,, "The Cyberspace: Redefining A New World," *IOSR Journal of Computer Engineering,* vol. 17, no. 3, pp. 17-24, 2015.

8. APPENDEX

Functionality		Mbox paris	MX Toolbox	Phish tool	Phish thank	autopsy	Network miner	Wireshark	Kernel outlook
Forensic Capabilities:	Email header parsing and analysis		x						x
	Email attachment extraction and analysis			x	x	x			x
	Support for common email file formats (e.g., .eml, .msg)	x	x	x	x	x			x
	Integration with Network forensics						x	x	
Usability and Automation:	Intuitive user interface and ease of navigation		x	x	x	x	x	x	x
	Availability of scripting or automation features	x	x	x	x	x	x	x	x
	Quality of documentation and community support	x	x	x	x	x	x	x	x
Performance and Scalability:	Processing speed and efficiency for large email datasets	x	x	x	x	x	x	x	x
	Resource utilization (CPU, memory, storage)	x	x	x	x	x	x	x	x
	Ability to handle complex email structures and attachments					x			
Reporting and Visualization:	Comprehensive reporting capabilities					x			
	Visualization of email forensic data and findings					x			
	Export and sharing options for collaboration	x	x	x	x	x	x	x	x
Maintenance and Security:	Active development and community engagement	x	x	x	x	x	x	x	x

	Availability of security updates and bug fixes	x	x	x	x	x	x	x	x
	Compliance with relevant standards and best practices	x	x	x	x	x	x	x	x

Appendix 2

Framework Evaluation Checklist

To evaluate the framework using the evaluation criterions, “Yes” and “No” options were presented for the Digital forensic experts. If the framework satisfies the criteria, the respondent tick (√) for Yes and (X) for No

Digital Forensic Expert	Effectiveness	Efficient	Feasibility	Extensibility and Adaptability	Reliability	Functional requirements	Scalability	User Experience and Usability
1	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	x	✓	x	✓
4	✓	✓	✓	x	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	✓	✓
6	✓	✓	✓	✓	✓	✓	✓	✓
7	✓	✓	✓	✓	✓	✓	✓	✓
8	✓	✓	✓	✓	✓	✓	✓	x

Appendix 3

Data collection Code mail Box

```
import mailbox
import pandas as pd

# Function to parse an MBOX file and extract emails
def parse_mbox(mbox_file):
    mbox = mailbox.mbox(mbox_file)
    emails = []

    for message in mbox:
        if message.is_multipart():
            subject = message.get("subject", "")
            sender = message.get("from", "")
            recipient = message.get("to", "")
            date = message.get("date", "")
            body = ''.join(part.get_payload(decode=True).decode() for part in message.walk() if part.get_content_type() == 'text/
        else:
            subject = message.get("subject", "")
            sender = message.get("from", "")
            recipient = message.get("to", "")
            date = message.get("date", "")
            body = message.get_payload(decode=True).decode()

        emails.append({
            "subject": subject,
            "sender": sender,
            "recipient": recipient,
            "date": date,
            "body": body
        })

    return pd.DataFrame(emails)

# Example usage
mbox_file_path = 'path/to/your/mboxfile.mbox'  # Path to your MBOX file
emails_df = parse_mbox(mbox_file_path)

# Display the first few rows of the DataFrame
print(emails_df.head())
```

Figure 32 mailbox code

```python
import pandas as pd
from pymongo import MongoClient

def extract(file_path):
    return pd.read_csv(E:\thesis\spam_mail_classification-master)

# Transform: Clean and prepare the data
def transform(data):
    # Remove rows with missing values in critical columns
    data.dropna(subset=['sender', 'recipient', 'subject', 'date'], inplace=True)

    # Convert date column to datetime type
    data['date'] = pd.to_datetime(data['date'], errors='coerce')

    # Filter emails sent after a specific date
    filtered_data = data[data['date'] > '2023-01-01']

    # Optionally, extract domain from sender's email
    filtered_data['sender_domain'] = filtered_data['sender'].apply(lambda x: x.split('@')[-1])

    return filtered_data

# Load: Save the transformed data into MongoDB
def load(data, mongo_uri, database_name, collection_name):
    client = MongoClient(mongo_uri)
    db = client[database_name]
    collection = db[collection_name]

    # Convert DataFrame to dictionary and insert into MongoDB
    collection.insert_many(data.to_dict('records'))
    print(f"Inserted {len(data)} records into {collection_name} collection.")

# ETL Process
def etl_process(input_file, mongo_uri, database_name, collection_name):
    # Extract
    data = extract(input_file)

    # Transform
    transformed_data = transform(data)

    # Load
    load(transformed_data, mongo_uri, database_name, collection_name)

# Example usage
input_file_path = 'emails.csv'  # Path to the input CSV file
mongo_uri = 'mongodb://localhost:27017/'  # MongoDB connection URI
database_name = 'email_database'  # Name of the database
collection_name = 'emails'  # Name of the collection

etl_process(input_file_path, mongo_uri, database_name, collection_name)
```

Figure 33 etl code

Mongo DB configurations

Step 1:- configure config server as replica set

Primary

Secondary

Secondary

Step 2:-configure shard nodes

Shard 1

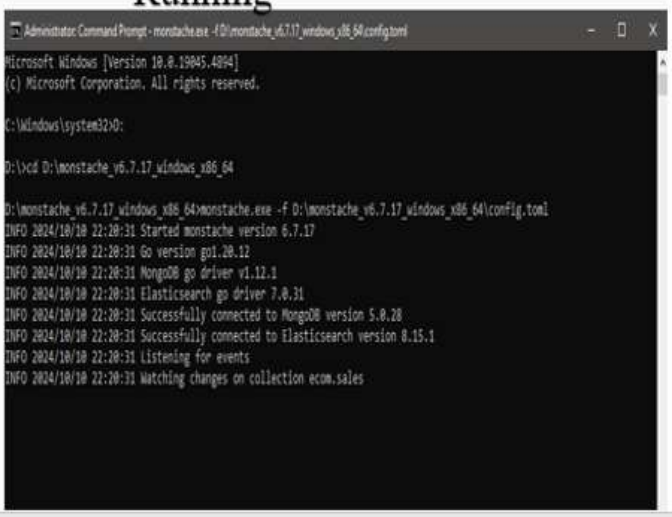
Shard 2

write python script /java code /install monstache to integrate mongoDB and elasticSearch(I use monstache).

Config..

```
# MongoDB settings
mongo-url = "mongodb://localhost:27017"
# Define the MongoDB database and collection to monitor
direct-read-namespaces = ["ecom.sales"]

# Elasticsearch settings
elasticsearch-urls = ["http://elastic:amoral237.@localhost:92
#elasticsearch-urls = ["http://localhost:9200"]

# Optional: Set index settings and mappings
# Uncomment and modify as needed
# index = "myindex"
# type = "_doc"

# Enable change streams (if using MongoDB 3.6+)
change-stream-namespaces = ["ecom.sales"]

# Optional: Configure logging
#log-level = "info"
# log-file = "monstache.log"

# Optional: Set up a bulk size
#max-bulk-size = 1000

# Optional: Run a command on the MongoDB side after the sync
```

Running



Email Forensic Framework Questionnaire

General Information

- Name of Investigator:
- Case Number:
- Date of Investigation:
- Client/Organization Name:
- Contact Information:
- Email

Data Sources

- What Email Forensic System are in use?
- Are their archived email systems? If so, specify
- Time period of Emails which is relevant to this case?
- Is there more than one email account involved? If yes, then list them
- Approximate Number of Emails to analyze?

Data Collection

- What is the email data collection methodology: direct extraction or using third-party tools?
- Are there specific email accounts or users of interest?

What steps will you take to ensure the chain of custody is maintained?

- How will you ensure the integrity of the data collected, such as by using hashing methods?
- Will metadata associated with the emails be collected? If so, what specific metadata does one want to focus on?

Analysis

- What type of email data will be analyzed, such as headers, body content, and attachments?
- What email analysis tools will be utilized?
- Are there any specific keywords, phrases, or patterns you will search for during the analysis?
- How will you handle encrypted or password-protected emails?
- What methods will you use to analyze email attachments?

Governance

- What governance frameworks are in place to guide the investigation?
- How would you ensure that the investigation meets data governance standards?
- How will sensitive or classified information be handled?

Data Integration

- How would email data be integrated with other data sources, for example, databases or systems?
- What are the different kinds of formats in which data can be integrated, for example CSV, JSON?
- Is there an existing suite of tools or platforms currently utilized for data integration? If yes, what?

- How would you ensure accuracy and consistency in integrated data?

  Big Data Considerations
- Will you be dealing with large volumes of email data that may qualify as big data?
- What big data technologies will be leveraged in the analytics, for example Hadoop, Spark?
- How would you ensure scalability in the processing and analysis of large datasets?

  What strategies will you put into place to manage the storage and retrieval of data effectively?
- Are there any big data-specific algorithms or analytics techniques that will be utilized?

  Case Management
- What case management system will be utilized in support of the investigation tracking?
- How will the investigation actions be documented within the case management system?
- What procedures are to be used for notifying stakeholders of case progress?
- How will case-related communications (notes, meetings, etc.) be managed?

  Documentation
- How will the findings be documented throughout the investigation?
- What form-technical report, full report, and others-the final report will take, and what will be the major information to be contained therein?
- Will there be visualizations such as charts and timelines of the data to be presented in your report?
- How many draft versions of your work will you keep while writing?

  Compliance and Legal Considerations
- To what legal or regulatory considerations may the investigation be subject?
- How can you ensure applicability in running the investigation under the various involved laws?
- Will you be interfacing with legal representation? If so, then when?

  Analysis and Correlation
- What will your processes be for peer analysis of the findings?
- Are there any validation techniques that you will conduct to corroborate your findings?
- How will you control for possible bias in your interpretation of the information?
- What are the criteria for establishing credibility in the evidence being collected?

  Incident Response
- What would you do if you find indications that the criminal activity is still active?
- How would you report your findings to the relevant parties, for instance, IT, legal, and management?

  Post Investigation
- What recommendations do you have regarding the investigation findings?
- How would the outcome be presented to the stakeholders?
- Are there any follow-up actions or audits that will result from your findings?

  Additional Comments

- Is there anything else relevant you would like to add or context to this matter?
- What challenges do you anticipate during the investigation?

Email Forensic Framework Questionnaire Response

The recent investigation of the email forensics highlighted the complicated landscape of challenges and methodologies of data collection and analysis by the investigators. The approaches employed ranged from directly extracting information from the email servers to using commercially available forensic tools in data gathering, even manual data gathering from devices, such as network equipment, mobile phones, and desktop computers. These commercially available tools came with a level of functionality at an affordable cost, although purchasing and upgrading these solutions often made many financial strains and continuous costs for maintenance.

This not only strained their budget and resources but also was becoming demanding as an upgrading challenge to keep pace with evolving technologies and security threats. This challenge hindered their ability to sustain an effective and current forensic infrastructure and thus might have compromised the integrity of the investigations. The investigators also felt that the absence of an integrated data integration approach led to fragmented data sets that made the analysis of the data a nightmare. Since there was no single view of data, the correlation of findings across sources became an arduous and time-consuming task, thus delaying the investigation.

Another significant gap that was reported in the data governance practices was in the area of data governance. The investigators suggested that there was a lack of specific tools that could guide them to adhere to the standards for data governance, thus making the data analyzed not credible and compliant. This absence put at risk the credibility of the investigation, as the team struggled to maintain a clear chain of custody and ensure the accuracy of the information. The absence of governance tools created not only legal and regulatory risks but also made the use of commercial tools difficult since they could not fully meet their needs for governance.

Also, fragmentation in data collection made the task of arriving at a standard analytical methodology very challenging. Each investigator approached his source of data uniquely, and there was great variability in how findings were interpreted and reported. This inconsistency further complicated the overall case management whereby the team found synthesizing insights from disparate data streams very difficult.

In the light of these findings, the investigators asserted that there was a dire need to develop improved tools and strategies for enhancing data governance and hence managing cases more efficiently. This included recommendations such as adopting a centralized data management system, which would provide the ability to integrate and analyze information from different sources. Besides, standardization of protocols related to data collection and governance would greatly enhance the efficiency and effectiveness of the investigation.

The investigation has highlighted that a well-defined process and good data-managing tools are needed for future email forensic investigations to be more precise and conform to the law. The

challenges of acquiring commercial tools and their upgrades will be overcome for sustaining the forensic capability over time.

The investigators understood the need to adapt open-source solutions in order to alleviate some of the challenges that were being presented with their existing commercial tools and some of the issues regarding data management. Open-source forensic tools could provide a cheap way to create capabilities sans licensing costs and the frequent upgrading required by commercial software. For example, by using open-source platforms, the team would be able to tailor tools to suit their very specific needs in investigations and ensure their systems are always adept at emerging technologies. Another nice thing about open-source software is that it fosters continuous improvement in security and functionality due to communities working on an ongoing basis. By allowing open-source solutions into their existing toolkits, investigators would be able to put in place a more robust, flexible, and economically sustainable forensic infrastructure to meet their operational needs with increased productivity and efficiency in email forensic investigations. The result is going to be a structured email investigation framework improving investigative efficiency, consistency, and collaboration within the team. It would also provide clear lines along which future investigations could be conducted, thereby making it easier to adapt to changes in the fast-evolving field of email forensics.