sustainability

MDPI

Review

Email Security Issues, Tools, and Techniques Used in Investigation

Esra Altulaihan [1,*], Abrar Alismail [1], M. M. Hafizur Rahman [1] and Adamu A. Ibrahim [2]

[1] Department of Computer Networks and Communications, King Faisal University, Al-Ahsa 31982, Saudi Arabia; 222401435@student.kfu.edu.sa (A.A.); mhrahman@kfu.edu.sa (M.M.H.R.)
[2] Department of Computer Science, KICT, International Islamic University Malaysia (IIUM), Kuala Lumpur 53100, Malaysia; adamu@iium.edu.my
* Correspondence: 221400737@student.kfu.edu.sa

Abstract: The email system is a globally distributed communication infrastructure service that involves multiple actors playing different roles to ensure end-to-end mail delivery. It is an indispensable method of communicating that is changing how people share data and information. As a result, it facilitates effective and efficient communication, especially in business, as well as convenience, accessibility, and replication. Today, email can send more than just text files; it can also send audio, video, photos, and other files with extensions. With email becoming ubiquitous in all aspects of our lives, enhancing its security, operating procedures, and forensic investigation has become essential. The purpose of this paper is to review some real email forensic incidents and the tools and techniques that have been proposed. A discussion of the major threats to email as well as techniques to mitigate them will also be provided. A comparison study was made of several techniques and analysis tools used in email forensics. In addition, this paper compares the available software tools for email forensics based on factors such as language interface, user interface, programming language, creation of image files, calculation of hash value, cost, and advantages.

Keywords: email; security; email forensics; threats; investigation; email analysis; tools; techniques

Citation: Altulaihan, E.; Alismail, A.; Hafizur Rahman, M.M.; Ibrahim, A.A. Email Security Issues, Tools, and Techniques Used in Investigation. *Sustainability* **2023**, *15*, 10612. https://doi.org/10.3390/su151310612

Academic Editors: Ali A. Alwan Al-juboori and Yonis Gulzar

Received: 16 May 2023
Revised: 28 June 2023
Accepted: 29 June 2023
Published: 5 July 2023

1. Introduction

Email is a common way of communicating between two parties. People share their data and information using email as an integral method of communication. As a result, communication is more effective and efficient, especially in business, convenience is offered, and access and replication are made an easier investigation. Email communication has gained popularity in recent years via computers, laptops, and mobile phones [1].

Email is a communication technique that combines flexibility and near-instantaneous information sharing through a digital network of computers (servers) that is now essentially global. V. A. Shiva Ayyadurai built new electronic mail software in 1978, when he was 14 years old, incorporating the features of every future email software "application": Inbox, Memo (To, From, Date, Subject, Cc, Bcc), Outbox, Address Book, Trash, Folders, Attachments, and more. His goal was to replace the pneumatic post system used to carry letters among office staff at a tiny medical institution with email [2].

Shiva Ayyadurai, in an excellent new account on how "experts" have continued to mistakenly expect the death of email from its creation, has described how they "keep confusing email with other media: chat, online bulletin boards, texting, instant messaging, blogs, etc.". But, when one considers the origins of email—the interoffice mail system, which served as the motor of business communications—it becomes evident that as long as businesses, large and small, exist, email will be there for a long, long time [2].

Email, short for electronic mail, was first developed in the 1960s and 1970s as a way for researchers and scientists to communicate with each other over long distances. It quickly became popular among businesses and individuals in the 1990s with the widespread

adoption of the Internet. Today, email remains one of the most commonly used forms of communication worldwide [3].

During this Internet era, email is one of the most important communication tools. There are many benefits to it, including effective communication, convenience, and easy access. Even though email communication makes things simple, efficient, and powerful, it is important to remember that email should not be used as a substitute for face-to-face communication in certain situations, such as delivering bad news or discussing sensitive topics. Additionally, it is crucial to maintain proper email etiquette and professionalism in all correspondence. Nevertheless, fraudsters often use email as a means to commit fraud or collaborate with their accomplices. The use of email for cybercrime can include spoofing, phishing, and bogus offers [4].

There is security when there are no threats, dangers, fears, or anxieties. Thus, email security refers to protecting email accounts and communications against unauthorized access, loss, or compromise [5]. Organizations can enhance their email security posture by establishing policies and using tools to protect against malicious threats such as malware, spam, and phishing attacks [6]. It is also a common method of hacking into a company's network and stealing sensitive information [7]. As a result, we must ensure email accounts, information, and communications are protected from unauthorized access, data loss, and other hostile threats [8].

Email communication between a user and a recipient has much evidence, including who sent it, where it came from, what type of data was attached, what message ID was used to identify the email, when and where it was sent, the type of message, what time and date the email was sent, and what type of message was shared [9]. Therefore, computer forensic investigators apply appropriate tools and techniques to analyze and extract these artifacts from email. All available and related evidence, including electronic or digital data, must be collected by the fraud examiner or investigator. In addition to physical evidence, investigators should be able to obtain and analyze digital data as evidence. One email can contain thousands of emails and contacts. Therefore, digital forensic tools should be used to analyze the contents of the email.

There have been many technologies developed to examine and protect emails over the past few years. Some of these technologies include spam detection, phishing detection, content filtering, and attachment filtering [4]. To correctly identify important information, such as the recipient's name or identity, one of the keys to designing and developing these technologies is conducting forensic investigations on sample emails. The path between the sender and the recipient that transported the email, the client-side application that composed it, the timestamp when a message was generated, the unique message ID, etc.

1. Therefore, this study aimed to review some real incidents in email forensics and the tools or techniques proposed for that;
2. It also presents some email abuse scenarios and reviews and compares some available tools and techniques for email forensics using some perimeters;
3. The study also assists fraud investigators in selecting the best tool for email analysis;
4. The goal of the study is to outline the concept of email forensics and how it may be implemented;
5. This project will also investigate the vulnerabilities, dangers, and hazards connected with the email environment, as well as the protective solutions that may be used in such contexts;
6. In addition, this study will conduct a comprehensive assessment of common email forensics tools and compare them. This study aims to assist fraud investigators in picking the optimal technology for email analysis;
7. The study will also create awareness among individuals and organizations about the best tools for undertaking email forensics. To achieve this purpose, a literature review will be conducted, and relevant research will be selected and analyzed.

Due to the proliferation of digital records in email, and the increase in cyber-attacks in the email environment, email crime investigators should take advantage of the tools and techniques available to investigate email crimes. The increasing cyber-attacks in the email environment require the creation of new technologies to ensure a secure electronic environment. However, this article goes beyond simply advocating for a technical solution to include a more comprehensive understanding of the challenges evaluators may face while assessing email crime, such as finding email users' personal information to demonstrating techniques that contribute to a secure email environment.

Among the crucial scientific contributions this paper presents, a significant feature is exposing what email headers contain. It composes of a variety of data, including the sender's IP address, that can be used to trace the source of the email, making its research an essential scientific endeavor. While certain email providers, like Google, do provide users with the ability to see when and from where an email was accessed, in most cases, a more sophisticated tool is needed. You can use this information to see if your email was actually read.

- Other significant scientific advances in the field of email security, email investigation, and related areas are in analyzing Malware, which can reveal vital information about the attacker by revealing the malware's behavior and where it came from. Email data that has been deleted or altered can sometimes be recovered through the use of digital forensics tools and then used as evidence in court. As a result, research of this nature is essential. Email filtering systems also contribute significantly by minimizing the likelihood of email attacks through the research linked with these tools being used to demonstrate the technique of blocking spam and malicious emails. Finally, the research on two-factor authentication, which may be used to add an extra layer of protection to email accounts and reduce the danger of illegal access, is another key scientific contribution related to email security issues, tools, and techniques utilized in the email investigation. Email encryption is a tool for preventing the contents of emails from being viewed by prying eyes.
- This research has made important contributions to science in a number of ways, including drawing attention to problems with email security and investigating how researchers utilize email to work together effectively. The speed at which scientific discoveries can be made is directly correlated to the ease with which researchers can communicate, discuss ideas, and exchange information with one another. The research has often also revealed the necessity to be security conscious in relation to data sharing, which enables more efficient and collaborative communication. New cybersecurity approaches and technologies, such as encryption and digital signatures, have been developed to secure email communication from illegal access and alteration; this study has exposed those working in these areas. This investigation has also revealed that concerns about email security have prompted studies of online crime and risks like phishing, spam, and virus campaigns, all of which are frequently sent over electronic mail. The findings of this study will be useful in identifying patterns and designing solutions to these security issues. The research has also shown how concerns about email security have advanced the fields of cybersecurity, cybercrime research, data privacy research, and policy development.

This paper is organized as follows. Section 2 discusses the email system organized into three-subsections: security issues in email protocols, email abuse scenarios, and threats in email communication and email security techniques. Section 3 presents the methodology. Section 4 presents and analyzes the empirical studies. Section 5 discusses and compares the mail analysis tools. Section 6 provides discussion and open questions on the prep empirical study studies associated with email issues. Section 7 provides limitations to the research and suggests future work. Section 8 concludes the research.

2. Email System

Today, email is one of the most popular Internet applications and is used by most people to communicate (see Figure 1). A variety of components make up its architecture. We discuss the working and architecture of the email service to better define the email forensics tools and techniques. Initially, email messages were short and usually consisted only of text. Currently, email can also contain audio, video, and text messages.

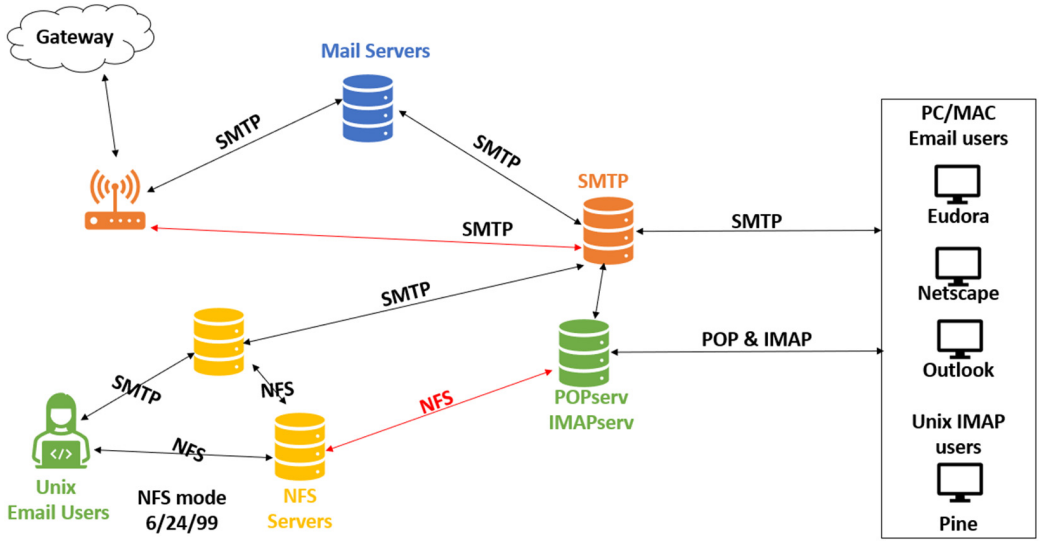

Figure 1. An email system architecture.

The main components of the email system lie with "User Agents", "Message Transfer Agents (MTA): Simple Mail Transfer Protocol (SMTP)", and "Mail Access Agent (MAA): Post Office Protocol (POP3) and Internet Message Access Protocol (IMAP)". In user agents, messages are composed, read, replied to, forwarded, and mailboxes are arranged or handled to make the sending and receiving of messages much easier. There are two types of user agents: command-driven and GUI based. In general, we use GUI-based user agents to make it easy for users to access services and use them. Netscape and Outlook are GUI-based user agents [9]. MTAs are responsible for actual mail transferring. MTA clients and servers are required for sending and receiving emails. SMTP is the protocol used by MTA clients/servers to handle working and proceedings [9–11]. Two stages of SMTP are used, one between the sender and the mail server, and another between the sender's server and the receiver's server. Since SMTP is a push protocol and a pull protocol is required at the receiver's end, either POP3 or IMAP4 can be used between the receiver's mail server and the receiver [10–16]. When sending messages between a client and a server, SMTP uses commands and responses.

The first stage (between sender and server) and the second stage (between server and recipient) use SMTP [9]. Due to its push nature, SMTP is not used in the third stage of the process (between the receiver's mail server and the receiver's mail server). On the other hand, the third stage required a pull protocol, which meant that the client had to retrieve messages from the server. The third stage of the process is achieved by using Message Access Agents, which implement either POP3 or IMAP protocols. POP3 stands for Post Office Protocol 3; it is quite simple but limited in functionality. It is necessary to install POP3 client software on the receiver's computer and POP3 server software on the server to which the receiver is connected. Users can download mail from their mailboxes residing on mail servers by connecting to MAA servers on Transmission Control Protocol (TCP) port 110 and verifying their identity by passing a username and password to the server. Afterward, he can retrieve his mail messages from his mailbox. As with POP3, IMAP4 is also an email access protocol but with more features than what POP3 offers. The POP3 protocol has several limitations, including the inability to arrange emails on the server, the

inability to categorize emails in different folders on the server, and the inability to partially download emails from the server before downloading. In addition to overcoming all these shortcomings of POP3, IMAP4 allows users to create mailbox hierarchies in a folder for email storage, search mail before downloading it from the server, and more.

Hyper Text Transfer Protocol (HTTP) is the most commonly used protocol for accessing HTML documents on the World Wide Web (WWW). The protocol is a mix of FTP and SMTP. In comparison to File Transfer Protocol (FTP), it uses only one connection, whereas FTP requires two connections for control and data transmission. Similar to SMTP, HTTP has a similar message format; additionally, MIME headers control the header format [9].

The SMTP server is constantly listening (see Figure 2). The client establishes a TCP connection to the SMTP server. The SMTP server waits for a connection and then establishes one on that port. The link has been formed. The client notifies the SMTP server that it wishes to send an email. Assuming the server is operational, the client sends the message to its mail server. DNS is used by the client's mail server to obtain the IP address of the receiver's mail server. The mail is then transferred from the sender's mail server to the receiver's mail server through SMTP. Once the message reaches the receiver's mail server, it is stored in a mailbox until the recipient retrieves it. The recipient can access their mailbox using an email client or webmail interface provided by their mail server. If the recipient's mail server is down or unreachable, the sender's mail server will attempt to deliver the message at regular intervals until it is successfully delivered or bounced back to the sender [10].

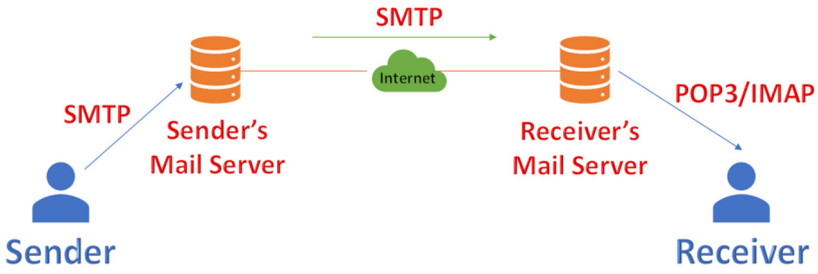

Figure 2. An outline of the SMTP and POP3.

IMAP is the most widely used email protocol and adheres to the client–server architecture (see Figure 3). It is a mix of client and server processes that execute on other computers connected by a network. This protocol communicates via the TCP/IP protocol. Once the link is established, the server defaults to listening on port 143, which is not encrypted; 993 is the safe, encrypted communication port. IMAP allows users to access their email from multiple devices and keeps emails stored on the server, allowing for easy access and synchronization. However, IMAP can be slower than other protocols due to the constant communication between the client and server [11].

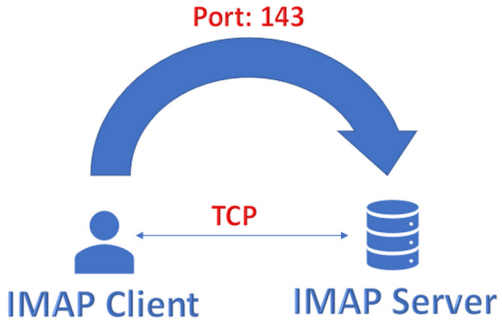

Figure 3. An outline of the IMAP.

2.1. Security Issues in Email Protocols

The SMTP protocol does not encrypt messages. Because SMTP server communications are in plain text, eavesdropping is possible. In SMTP messages, information about the sending computer and software can also be captured and used maliciously. Phishing attacks are also easy to send because it does not check message integrity. The SMTP protocol does not have a mechanism to prevent repudiation, which would force a sender to deny sending emails [12]. The POP and IMAP protocols are pull-type protocols, which means that a request is sent to the mail server for access to the mailbox using the username and password. Unless SSL is used, these details are not encrypted before being sent. We are therefore at risk of losing confidential information. PGP stands for Pretty Good Privacy, and it provides application-layer security. Emails are protected by it by providing confidentiality and authentication. PGP is a free and open-source encryption tool that runs on different platforms and is not controlled by anyone [13]. A PGP encryption process consists of the following phases: digital signature, compression, encryption, digital envelope, and Base64 Encoding. The main threat to PGP is imitation and tampering with the public as, if these keys are lost, all data can be lost.

Emails can be protected with cryptographic security by using Secure/Multipurpose Internet Mail Extensions (S/MIME). As long as the client software installed at the sending and receiving clients supports S/MIME, there is no need to change the sending and receiving MTAs or the email transmission process. In its most basic form, it provides authentication of the sender, non-repudiation of the sender, intertie-gritty of the message, and security of the message using encryption and digital signatures. The sender's privacy is at risk because recipients can forward emails with digital signatures without the sender's consent. With S/MIME, it is also impossible to ensure non-repudiation if keys are lost. In general, S/MIME and Pretty Good Privacy (PGP) do not sign message headers, making it possible for intermediaries to modify them. Since S/MIME and PGP do not require domain owners to be involved, retired employees of the company can continue to use their signatures [14].

2.2. Email Abuse Scenarios and Threats in Email Communication

Email may be used by criminals to gain personal information from consumers (victims). Email abuse may take numerous forms, including fraud email, spamming, bombing, phishing, eavesdropping, message modification, identity theft, repudiation, false messages, and malicious software spread via emails. As email abuse continues to become more prevalent, consumers need to be aware of the potential dangers and take measures to protect themselves. Consumers should be cautious about opening emails from unknown senders, since these emails may contain malicious links or software. They should also be wary of emails that appear to be sent from a legitimate source, such as a bank or business, but contain unexpected requests for personal information.

Fraud is defined as purposeful deceit committed over email for the perpetrator's advantage or to inflict harm to others [15]. The email content comprises an offer needing little capital with a large profit margin or a sale of valuable items at a reduced price. Victims of email fraud are often attracted to the offers due to their promises of large returns for minimal investments. Furthermore, the perpetrator of email fraud will use tactics such as urgency, over-the-top compliments, and the threat of punishment to manipulate the victim into sending their money without hesitation [16]. Spam emails are unsolicited emails that are sent to recipients in quantity and, frequently, the majority of these spam emails are commercial in nature; however, chain emails are also possible. Spam emails can range from advertisements for products, services, or events to malicious scams and threats, making them a very dangerous form of communication. To combat the growing number of spam emails, many email providers and websites have adopted advanced anti-spam technology [17].

Email bombing is a way of sending a large number of emails to overrun the inbox or flood the email server, culminating in a DoS (denial-of-service) attack. It has become an increasingly popular form of cyber-attack, particularly as the cost of bandwidth and storage continues to decrease. Email bombing is a destructive and dangerous form of attack, and can have serious consequences for the recipient [18]. Email spoofing is the process of modifying the contents of the fields in an email header such that it looks to come from a perfectly genuine source. This process is often employed by malicious actors for nefarious purposes, such as sending spam or phishing emails. Email spoofing has several implications, as it can be used to disguise the true source of an email and, consequently, mask its malicious intent [19–22]. Phishing is a form of cybercrime that employs social engineering and technological deception to obtain bank account identities and passwords [20]. The social engineering scheme is carried out through the use of counterfeit emails purporting to be from reputable business organizations, with the goal of trick viding personal information such as email addresses and passwords [23–29].

During eavesdropping, email communications go across networks that are part of a larger picture, such as the Internet, which has many individuals on it. As a result, it is quite simple for someone to trace or capture your communication and read it. In message modification, where the contents of your message are altered by anyone who captures it if it is not encrypted. The message you send can also be modified by anyone with administrative rights on any of the SMTP servers your message visits. This refers to someone pretending to be you on a network [30]. If proper security protocols are not followed, someone could steal your username and password and use them to read and send emails from your account without your knowledge. Messages sent through email can easily be forged, so whoever sends you a message can later deny that it was sent, and it is difficult to prove that. In terms of business communications, this also has implications for the use of emails as contracts, in which messages can be easily sent, masquerading as coming from someone else by fabricating the sender's name. It is also possible to spread viruses, worms, and other malicious software using emails. The attachments are attached to your emails, and when clicked, they attack your computer.

To achieve a secure and protected email environment, there is a set of techniques in place including digital signatures, encryption techniques, and hashing algorithms. Digital signatures verify the authenticity of an email's source, encryption techniques ensure that only authorized individuals can read the message, and hashing algorithms provide a secure way to store passwords and other sensitive information. Digital signatures generate a unique virtual fingerprint for a person or entity and are used to identify users and safeguard information in digital messages or documents. A digital signature added to an email message adds a degree of protection by ensuring the receiver that the originator is not an impostor who signed the email message's contents. The digital Tenderer generates the digital signature, which comprises the sender certificate and public key. But also, that digital ID acts as the sender's unique digital identifier, informing the recipient that the material was not changed in route. Furthermore, it also guarantees that the content of the email has not been tampered with or altered. Digital signatures, then, are a powerful tool in preventing fraud or tampering with emails and their contents [31–33].

Email encryption techniques are important to achieve a secure email environment. Securing the email environment involves the encryption of connections and the encryption of emails. The email connection is critical since other network users may obtain login information and see messages that were being sent or received if the connection to the email provider is not encrypted. Therefore, the entire data exchange between the client PC/browser/mail program on the one side and the server on the other side should only run via SSL (Secure Sockets Layer) or TLS (Transport Layer Security). These are cryptographic protocols that provide secure communication over the internet. SSL and TLS ensure that data transmitted between two systems remains private and cannot be intercepted by unauthorized parties [34]. Furthermore, it is important to pay attention to the address bar: if it begins with https rather than just http, then SSL/TLS encryption is enabled. For this

reason, users should be aware of the potential dangers associated with accessing email on public Wi-Fi networks and take extra steps to secure their accounts. To send encrypted emails, one can either utilize the mail provider's encryption features, install the necessary software, or use a client add-on. Most email encryption techniques, such as S/MIME (Secure/Multipurpose Internet Mail Extensions) and Open PGP (PGP stands for Pretty Good Privacy), are two widely used encryption protocols for securing email communication. S/MIME is supported by most email clients and uses a public key infrastructure to encrypt and sign messages, while Open PGP relies on a web of trust model and can be implemented using various software tools. Both protocols provide end-to-end encryption, ensuring that only the intended recipient can read the message [35]. which need the installation of security certificates on computers and the distribution of "public keys" to contacts, are asymmetric in nature. The message may now be decoded by the receivers. The Microsoft support website explains how to enable the capability in various versions of Outlook. Additionally, encryption for web-based email services is supported through browser add-ons like Gmail S/MIME for Firefox [36–39].

The term "hashing" refers to the act of encoding such common email addresses with a cryptographic hashing algorithm. To represent the email going forward, this method generates a hashed string of characters. Each hash has a predetermined number of characters, depending on the type of hash function used. The two most important hashing algorithms utilized by digital forensics practitioners are Message Digest (MD5) and Secure Hash (SHA1). These are cryptographic hash functions used for data integrity and authentication. MD5 produces a 128-bit hash value while SHA1 produces a 160-bit hash value [40]. In email forensics investigations, the MD5 and SHA1 hashing algorithms are commonly used. These algorithms enable forensic investigators to preserve digital evidence from the time it is obtained until it is presented in court. Hash values are also significant since electronic documents are shared with legal experts and other parties throughout the inquiry. Therefore, ensuring that everyone has identical copies of the files is crucial. Hash values are generated to ensure that the integrity of the data is not compromised and that it remains unaltered during the investigation. These hash values act as "fingers" for the electronic documents, and any minor alteration to the file will result in a different hash value being generated [41–44].

2.3. Email Security Techniques

Email is designed to be as open and accessible as possible. Using it, people within and outside organizations can communicate with each other. The problem with email security is that it is not reliable on its own. Email can be used by attackers to cause problems to make money. The lack of security of emails makes it easy for attackers to conduct sophisticated targeted attacks, spam campaigns, malware distribution, or business email compromise (BEC). Since most organizations use email for business, attackers exploit email to steal sensitive information [45].

Since email is an open format, anyone can intercept it and view it, causing concerns about email security. In the past decade or so, organizations have begun sending confidential and sensitive information via email. The contents of an email can easily be read by an attacker if the email is intercepted. To protect sensitive or confidential information from attackers, individual organizations have increased email security measures over the years [46].

To achieve a secure and protected email environment, there is a set of techniques in place including digital signatures, encryption techniques, and hashing algorithms.

- Digital signatures

  Digital signatures verify the authenticity of an email's source, encryption techniques ensure that only authorized individuals can read the message, and hashing algorithms provide a secure way to store passwords and other sensitive information. Digital signatures generate a unique virtual fingerprint for a person or entity and are used to identify users and safeguard information in digital messages or documents. A digital

signature added to an email message adds a degree of protection by ensuring the receiver that the originator is not an impostor who signed the email message's contents. The sender's digital ID generates the digital signature, which comprises the sender certificate and public key. But also, that digital ID acts as the sender's unique digital identifier, informing the recipient that the material was not changed in route. Furthermore, it also guarantees that the content of the email has not been tampered with or altered. Digital signatures, then, are a powerful tool in preventing fraud or tampering with emails and their contents [47–49].

- Email Encryption

  Email encryption techniques are important to achieve a secure email environment. Securing the email environment involves the encryption of connections and the encryption of emails. The email connection is critical since other network users may obtain login information and see messages that were being sent or received if the connection to the email provider is not encrypted. Therefore, the entire data exchange between the client PC/browser/mail program on the one side and the server on the other side should only run via SSL (Secure Sockets Layer) or TLS (Transport Layer Security). Furthermore, it is important to pay attention to the address bar: if it begins with https rather than just http, then SSL/TLS encryption is enabled. For this reason, users should be aware of the potential dangers associated with accessing email on public Wi-Fi networks and take extra steps to secure their accounts. To send encrypted emails, one can either utilize the mail provider's encryption features, install the necessary software, or use a client add-on. Most email encryption techniques, such as S/MIME (Secure/Multipurpose Internet Mail Extensions) and Open PGP (PGP stands for Pretty Good Privacy), which need the installation of security certificates on computers and the distribution of "public keys" to contacts, are asymmetric in nature. The message may now be decoded by the receivers. The Microsoft support website explains how to enable the capability in various versions of Outlook. Additionally, encryption for web-based email services is supported through browser add-ons like Gmail S/MIME for Firefox and Chrome. It is important to note that while encryption can help protect the contents of emails, it does not guarantee complete security as other vulnerabilities such as phishing attacks and malware can still compromise email accounts. Therefore, it is recommended to use encryption in combination with other security measures [50–55].

- Hashing Algorithm

  The term "hashing" refers to the act of encoding such common email addresses with a cryptographic hashing algorithm. To represent the email going forward, this method generates a hashed string of characters. Each hash has a predetermined number of characters, depending on the type of hash function used. The two most important hashing algorithms utilized by digital forensics practitioners are MD5 and SHA1. In email forensics investigations, the MD5 and SHA1 hashing algorithms are commonly used. These algorithms enable forensic investigators to preserve digital evidence from the time it is obtained until it is presented in court. Hash values are also significant since electronic documents are shared with legal experts and other parties throughout the inquiry. Therefore, ensuring that everyone has identical copies of the files is crucial. Hash values are generated to ensure that the integrity of the data is not compromised and that it remains unaltered during the investigation. These hash values act as a "fingerprint" for electronic documents, and any minor alteration to the file will result in a different hash value being generated [56–60].

Table 1 provides a comprehensive analysis of the available security techniques to achieve a secure email environment based on a set of criteria including the technique's main idea, authentication mechanism, purpose, validation, and security.

Table 1. Analysis of techniques toward a secure email environment.

Criteria/Technique	Digital Signature	Email Encryption Techniques	Hashing Algorithm
Main Idea	Generate a unique virtual fingerprint for a person or entity to safeguard information in digital.	Securing the email environment involves the encryption of connections and the encryption of emails.	Act of encoding such common email addresses with a cryptographic hashing algorithm.
Authentication Mechanism	Generated digital certificate based on the user ID.	The address bar must begin with https rather than just http.	Provide a secure way to store passwords and other sensitive information.
Used Algorithm	DSA, RSA	SSL/TLS	MD5 and SHA-1
Used for	Verify the authenticity of an email's source.	Ensure that only authorized individuals can read the message.	Provide a secure way to store passwords and other sensitive information.
Validation	Performed by a trustworthy certificate authority.	Validated by the certificate authority.	Trusted by a third-party tool, compare them to the data received from the source. Great, if they match. If they do not, there is a problem, and the email should be disregarded or retrieved from a trusted source directly.
Security	Highly secure	Vulnerable to alteration by attackers.	Highly secure

According to this comparison, email signatures and hashing algorithms provide good email security. While email encryption ensures the confidentiality of messages, it does not guarantee that the sender is who they claim to be. In addition, they do not protect against a man-in-the-middle attack, regardless of whether the message is encrypted. The digital signature ensures these two properties, which can be seen as the digital equivalent of the manuscript signature. The purpose of signature protocols is not only to authenticate the sender of a message but also to provide the non-repudiation property, which ensures that no mistake can be made when signing a message with a signature key. Hashing algorithms provide a secure method of storing passwords and other sensitive information in an email.

Digital forensics is the process of using scientific approaches to locate, collect, authenticate, and analyze digital evidence to fulfill legal requirements. The forensic examination of emails and their contents to establish the authenticity, source, date, time, real sender, and recipients. Electronic forensics, also known as digital forensics, is the process of using specialized computer science techniques to examine emails and other digital evidence. The goal is to create digital evidence that is admissible in civil or criminal courts [1]. By utilizing electronic forensics, law enforcement, and legal professionals can establish important elements of digital evidence that will be essential in any criminal or civil proceeding. Email forensics is a subset of digital forensics that focuses on collecting digital evidence for cybersecurity attacks and security events through the forensic examination of email [4]. Through the analysis of emails and their associated metadata, investigators can provide legal evidence that can be used in criminal or civil court cases. By collecting and analyzing evidence in this way, law enforcement and legal professionals can build a strong digital case against perpetrators of crime. Email forensics is a powerful tool for law enforcement and legal professionals, helping them establish the facts in digital criminal cases and build a compelling digital case against those accused.

3. Research Methodology

PRISMA guides the search, which is divided into three parts. The Saudi digital library and Google Scholar databases were searched during the identification stage using the following inclusion criteria: papers describing email forensics tools, cybersecurity threats, and the email environment, and papers published between January 2018 and December

2022, as well as papers published in an academic journal or conference paper, are listed as the source type. Table 2 lists four exclusion criteria: publications that do not address email cybersecurity threats or digital forensics tools, papers that are not published in English, and papers that are not directly related to email cybersecurity threats or email forensics tools. In addition, we excluded papers that are not available online.

Table 2. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Papers describing cybersecurity threats, digital forensics tools in the email environment.	Papers do not address risks and digital forensics tools in the email environment.
Papers published between January 2018 and December 2022.	Papers that are not written in English.
Papers published in academic journals or conference papers.	Papers that are not available online.

Figure 4 illustrates that during the identification step, a total of 33,200 articles were identified, with 33,200 papers remaining after duplication was deleted. At the screening stage, the articles were then screened based on their titles and abstracts, resulting in 4350 out of the 33,200 papers being eligible for full-text review assessed for title and abstract being disbanded for not closely meeting the standards. At the eligibility stage, 4350 studies are qualified to move on to the final round. The full-text review step involved a thorough assessment of the eligibility criteria, resulting in the final selection of 4350 articles for inclusion in the systematic review; 4334 were discarded, leaving 16 for review.

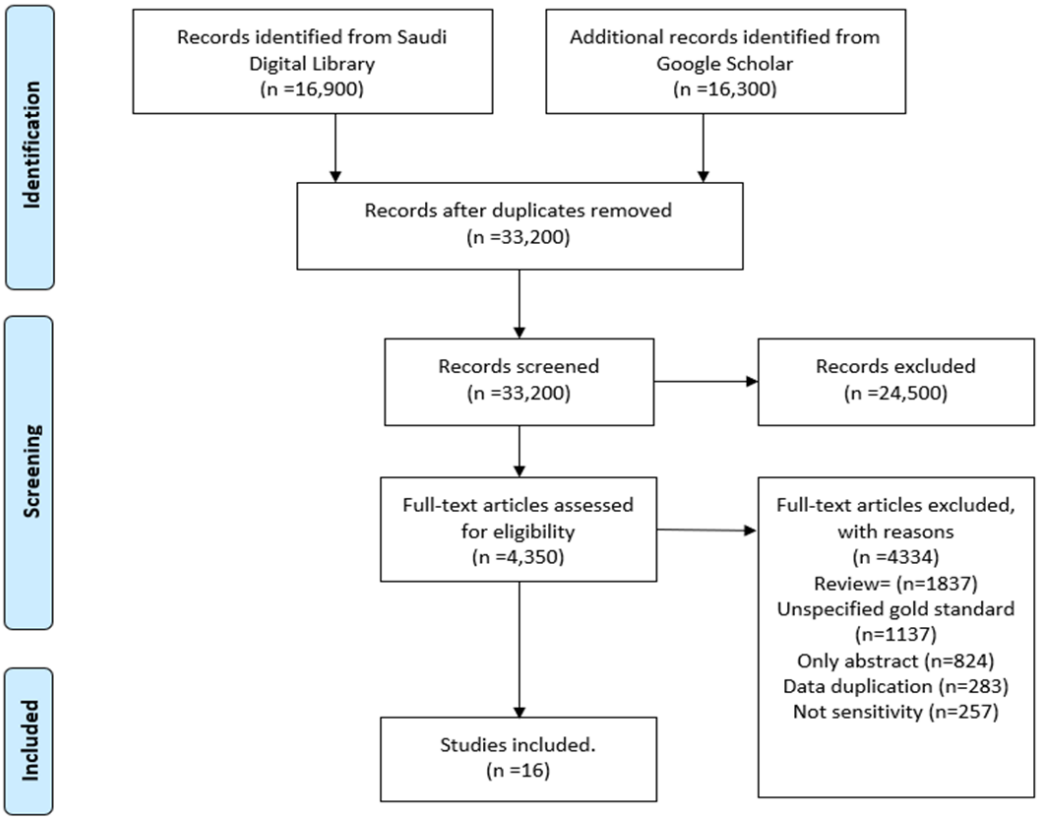

Figure 4. Prisma research methodology.

4. Empirical Studies

This section aims to present previous empirical studies associated with the analysis of email, previous empirical studies on email systems associated with use cases, and previous empirical studies associated with email abuse investigation scenarios. The presentation of these studies will provide a comprehensive understanding of the various approaches and methodologies used in analyzing email data. This knowledge can be useful in developing effective strategies for email management, security, and investigation.

4.1. Previous Empirical Studies Associated with the Analysis of Email

Many earlier research studies have been carried out in connection with a key component of email and email systems. The research conducted by Hamdi et al. [15] demonstrates that email communications constitute a form of digital evidence that is increasingly being used in courts. This system also provides a hybrid English lexical Dictionary, which can be utilized for the r forensic examination of email (SentiWordNet 3.0). The proposed system can learn in settings including a substantial amount of data from a variety of sources.

Zakiyaturrahma et al. [16] demonstrated that phishing is a tactic used by attackers to obtain email users' social network accounts. The purpose of this study is to undertake forensics on phishing email assaults. Wireshark and NetworkMiner are used to determine the perpetrator's IP and IP address. Salhi et al. [17] conducted research based on a novel approach of email clustering to extract the poor and excellent emails in this publication. The study's goal is to create an automated screening and detecting mechanism on servers to separate harmful emails from good ones. Ghafarian et al. [18] demonstrated that email forensics is the examination of email detail and content as evidence to identify the true sender and receiver of a message. Email forensics entails the study of metadata, keywords, searching, report production, and other elements such as email format, file size acceptable, report format, and so on. Baroto et al. [19] recognized that mail has become a significant aspect of human connection because it has changed the way individuals transfer data and information. Fraudsters may employ modern technologies to aid in their deception. This study uses an email body and content to undertake digital forensics in an inquiry. The findings suggest that digital forensics may assist investigators in analyzing emails while maintaining the integrity of the overall investigative process.

Devendran et al. [20] conducted a review of five source email forensic tools, namely, MainXaminr, Add4Mail, Digital Forensic Framework eMailTrackerPro, and Paraben Email Examiner. Based on nine criteria, the tools were compared: input file on disk, search option, the information provided, recovery appealable quality supported, visualization format supported, operating system supported, export format, and extended device support. They found that Add4Mail is the only tool that can analyze email hard drives (offline analysis) as well as remote servers (online analysis). Add4Mail has the highest level of capability for gathering information, compared to other tools, when it comes to the search option. Additionally, only a few tools support extended devices like USB memory sticks (Add4Mail and Digital Forensic Framework). Hamdi et al. [15] presented a unique method of classifying emails based on data processing and mining, trimming and refinement, and adapting several algorithms. For email forensic analysis, they use the SWARM algorithm, as well as the hybrid English lexical dictionary SentiWordNet30 and machine learning. This system is capable of learning from large and varied datasets. To test the proposed system, a set of Enron data was selected as the available data. Compared with previous research papers, this study achieved a higher classification accuracy (95%).

Baroto [22] analyzed the email fraud investigation using the design science research methodology. Based on the process of research and demonstration, it was discovered that data integrity in emails (metadata and content) can be maintained using a sound forensic procedure. The email body can be extracted for further analysis (keyword searches) as well as more advanced analyses such as sensitivity analysis. It is important to extract several useful data from the header of an email so that further analysis can be performed. A network theory can help investigators find suspects, eliminate unnecessary data, and

visualize relationships. Finally, email forensic frameworks that combine digital forensics and social network analysis are needed. In Hina et al. [23], to organize emails, a multi-label classification system was proposed. This paper proposes a method for efficiently classifying massive email data for forensic investigation (e.g., an image of an email server). Investigators could use this method when investigating crimes related to email. Based on a comparative analysis of machine learning algorithms, logistic regression was found to be the most accurate method over naïve Bayes, stochastic gradient descent, random forest, and support vector machine. According to their experiments conducted on benchmark datasets, logistic regression performs best, with an accuracy of 91.9%.

Shalin et al. [24] presented a forensic system for analyzing and retrieving email artifacts stored on digital media. The researchers attempted to recover email artifacts from digital evidence in a forensically sound manner. It has been discovered that the Paraben Email Examiner and OS Forensic both have limitations that only allow them to extract a few types of files from evidence. In the first place, it places a mental burden on the examiner to determine whether this specific type of file exists in evidence or not. Examiners may have difficulty extracting artifacts from digital evidence if they cannot locate these files. Using the Forensic Toolkit tool, you can mount entire images of digital evidence. The tool automatically searches email files and lists them in their respective controls. In this study, they found that tool selection has a great deal of importance in investigation and gives forensic investigators a better way to retrieve and analyze emails, being a great asset to them.

Wiliam et al. [25] stated that as digital records proliferate, archives must adjust their ways of gathering, assessing, and giving access to materials. Digital records need the creation of novel techniques to ensure that archived information is gathered, protected, and preserved. This article goes beyond simply advocating a technical solution to include a more comprehensive understanding of the challenges that archivists face. Umar et. al [26] revealed that email is a means of communication that may be used to share information, data, and so on. The developing technology of greater cybercrime, such as email fraud, must be filed before a court in a row. To collect proof, technologies like Wireshark and NetworkMiner can be used to examine network traffic on live networks. Minchev et al. [27] demonstrated that future civilization is projected to provide computers with life-like features, which will introduce several uncertainties and relativities into the next complex data handling. The research investigates the issue by building a system of system-transcending future modeling vision and assessment from the standpoint of data relativities.

Armknecht et al. [28] developed a novel approach to privacy-preserving email forensics by allowing for non-interactive threshold keyword searches on encrypted emails (or text more generally). Essentially, a user can search the encrypted text autonomously for keywords. When the investigator is searching for t keywords, the search process reveals the content of an email. Without this information, the investigator will not be able to determine whether the selected keywords are contained in the email. An encrypting software prototype was implemented as a proof of concept to encrypt email inboxes of different formats. They also developed a plugin for the well-known open source forensic framework Autopsy to enable it to handle encrypted containers to demonstrate the feasibility of their approach. The authors concluded that their approach to improving privacy in forensic investigations is novel, secure, and practical and that it can be applied not only to email analysis but also to other areas of digital forensics.

Banday [29] presented a proposal for researching the design and development of new and improved email security protocols and forensic tools. The research project was proposed in four areas relating to email: security protocols, security procedures, forensic tools, and legal resolutions. As part of the study, improvements were made to existing protocols and procedures, as well as new protocols and procedures for detecting email date spoofing, filtering, and forensic terrorist emails, ad filtering multilingual contaminated email messages. Email flow and security problems in email flow were briefly introduced in the proposal. Afterward, it reviewed the literature on email security protocols. Following that, it outlined the importance of the study and its objectives. As part of the proposal, a proposed methodology and work plan was provided.

Riadi et al. [29] analyzed spamming emails through header analysis. This study uses a method developed by the National Institute of Standards and Technology (NIST), which consists of four stages: collection, examination, analysis, and reporting. According to the results of the NIST tests, examining the email headers was an important aspect of the investigation since the headers contained metadata and other information. By analyzing email headers, you can identify the sender's IP address, the sender's source, and more. Applications could be used to track IP addresses to find out who sent an email. The tracking of IP addresses made it easy to discover routes, geographic locations, network providers, etc. Based on the results of this study, it is possible to determine the spam sender's email address, the spam sender's IP address, and other information.

4.2. Previous Empirical Studies on Email Systems Associated with Use Cases

The proposed system [10] is divided into many phases, each with its function: data preparation, feature extraction, clustering, feature selection, optimization, classification, and prediction outcomes. The four machine learning algorithms utilized in this paper are K-means for clustering and naïve Bayes for feature extraction, particle swarm optimization, and support vector machine for classification. Clustering is a way of arranging data into groups. Clustering's fundamental purpose is to partition the entire dataset into numerous clusters. Clustering techniques that are often employed include k-means, k-medoids, hierarchical, density-based, and numerous others. Current feature selection strategies in classification procedures pick characteristics that are acceptable for 0–1 categorization. They do not consider the precision of the class probability estimations supplied by the classifier. Current feature selection approaches use naïve Bayesian classifiers to provide correct class probabilities. Particle swarm optimization (PSO) is a swarm-based intelligence approach that mimics the behavior of flocks of birds and fish to discover the best solution. It works by creating a population of random solutions and then updating them to find the best option.

The support vector machine (SVM) algorithm is a supervised machine learning technique used to categorize and analyze data. Data is usually distributed in the nonlinear support vector machine classifier. Non-linear classification may be achieved for this problem by applying a kernel to hyperplanes with the greatest margin. Enron Corpus is being used for research purposes. Enron Corpus was produced during the legal inquiry of Enron Corporation and revealed several integrity issues. This information is useful. It is the only major bunch of public "actual" emails that I am aware of. Other datasets are not publicly accessible due to privacy concerns. Another reason the Enron dataset has thousands of categories and samples is that data is deemed to represent actual communications. Data processing is an important step in the data mining process. This is done to prepare data for use in the next steps. Document representation can take several forms, including vector model, graphical mode, and so on. Many metrics are also used to weigh documents. Data preprocessing is sometimes the most time-consuming component of a machine learning project. If there is a lot of redundant and unnecessary information, or data confusion and untrustworthy data, learning will be more difficult during the training stage. Tokenization and stop word removal are the two phases of data preprocessing. The proposed system employs a java tokenizer to turn each email message into different words or tokens.

The primary purpose of the tokenization stage is to break down the message text into smaller components. This enables feature extraction, which is the process of extracting all features from a dataset. The forward feature extraction approach decreases the original feature set by removing unnecessary text characteristics. The email dataset was examined for forensic terms, and POS was found. Tagging categories include nouns, verbs, adverbs, and adjectives, and the score for each phrase is generated using SentiWordNet 3.0 communications to be separated into text streams based on their constituent meaning, as units known as tokens. Stop words are frequent terms present in nearly all text scripts. They include no valuable information that may be used to assess whether an email message belongs to a given categorization. The feature space dimensions will be reduced if stop words are

removed from the email content. If the contents of an email meet a specified cybercriminal categorization, it is categorized as malicious. There are several terms for various sorts of crimes and the criminals that perpetrate them. After loading the Forensic terms dictionary datasets, forensic words are searched in the email dataset. The Part-of-Speech tagger is a tagging tool used to tag each word and allocates parts- of- speech to each "word" and another for the "token". It distributes documents or sentences and assigns a part of speech to each term. SentiWordNet3.0 dictionary is a WordNet database opinion lexicon mining tool. Each token is associated with numerical ratings that indicate positive and negative sentiment data. The goal of this stage is to examine the information in the email dataset and assign a score to each phrase.

4.3. Previous Empirical Studies Associated with Email Abuse Investigation Scenarios

During this period, the volume of digital documents obtained from government and private contributors has steadily increased. One of the most difficult issues we had when we originally initiated the digital preservation program was our inability to examine the records due to software obsolescence. As the email was transmitted to the archives as an a.pst file, certain further considerations were necessary. While we had the option of opening, searching, and seeing the information in Microsoft Outlook, there was worry about the potential of changing the record. There were further worries about storing the files in an a.pst container due to the recognized hazards connected with this file type. Even though an a.pst file can contain hundreds of email messages, the loss of a single.pst file can be disastrous. As an early solution, we acquired an application called Aid4Mail, which was used to unpack the.pst file into multiple.msg files. This enabled us to search the email using Windows Explorer. This method of searching and retrieving information, however, proved to be extremely slow and required hours to accomplish. In at least one example, a slip of the hand led to communications being relocated to another place, prompting the archives to collect the email from government backups to guarantee no information was lost. Few Canadian provincial archives have established a strategy for methodically selecting digital material. The advantages of employing digital forensics to select and arrange digital documents prompted NPAANB to reconsider its position in the archival process. These difficulties prompted us to conduct a more detailed examination of the techniques used by other archive organizations [31].

This section aims to provide various scenarios for the investigation of email abuse scenarios including fraud email, spamming email, bombing email, spoofing email, and phishing email investigation scenarios. Investigating email abuse can involve a wide range of activities such as analyzing logs and headers, deciphering malicious code, decrypting data, and messages, collecting evidence, and even testifying in court.

The collecting and forensic analysis of evidence concerning email hacking, phishing assaults, tracking, and recovery of stolen monies are known as email fraud investigations. Email fraud is intentional deceit used for personal benefit or to harm another person via email. Almost as soon as email became popular, it began to be exploited to deceive individuals. Email fraud might appear like a "con game" or scam. Investigating email fraud encompasses all aspects of cybercrime, from recovering monies sent to a fraudster's bank account to conduct a forensic study to establish how fraudsters gained access to email accounts. Digipos's Email Fraud Investigation Team consists of qualified digital forensic professionals and fraud examiners that can assist with all email scams and fraud cases [16].

Email spam, also known as junk email, refers to unsolicited email messages, usually sent in bulk to a large list of recipients. Email spam is typically used by companies to advertise their products, but it can also be used for malicious purposes such as phishing scams and identity theft. Email spam senders, or spammers, regularly alter their methods and messages to trick potential victims into downloading malware, sharing data, or sending money. Email spam has become a pervasive problem due to its low cost of distribution and the relative anonymity of the sender. A payload is the one element that all phishing and emails have in common. It is usually an infected file or a link to a fraudulent website that

seeks login credentials and other sensitive information such as passwords or credit card data. Legitimate companies would never ask for sensitive information through emails or login links. Never open any attachments unless you are certain that the communication is from a trusted source. To be sure, contact the sender and ask them to confirm its authenticity [17].

Email bombing is an illegal form of cyber-attack that can have serious consequences for both the attacker and the recipient. The FBI investigates the willful damage or destruction of property utilized in interstate or foreign commerce with explosives. The bombing or attempted bombing of college or university premises is one of these issues. In addition, the FBI supports US Attorneys in preparing evidence or exhibits for trial [18]. Scammers will employ email spoofing to impersonate a supervisor, professor, or financial institution to deceive people into doing some kind of action. Scammers adopt this type of deception because they know that if a person knows who sent the message, they are more likely to engage with the content of the email. In order to identify the spoofed email address, we need to check the email header information. The email headers contain a lot of tracking information that shows where the message has been on the Internet. These headers are shown differently in different email systems. To access the header information, we need to open the message and view its source. The source code of the message contains all of the email headers, including IP addresses that are used to trace the location of the email sender [19].

This section aims to provide a comprehensive analysis of the available email analysis tools, email forensics software tools, and email investigation techniques. The analysis will include a detailed examination of the features and functionalities that the tools offer, as well as their associated costs. There are several techniques that are useful for conducting an email investigation during this step, including header investigation, server investigation, investigation of software-embedded details, investigation and discovery of hidden emails, and investigation of anti-forensic activity [20,21]. During the header investigation, the header is critical for inquiry and evidence collecting. It includes information on the sender/receiver, the path, and the message. This metadata/control information is sometimes manipulated or falsified. As a result, during header analysis, the authenticity of the information included in the header is also evaluated. To investigate the email header, we need to look for four important info informal pieces including "From" as a source of email address, "To" as a destination ion where the email will arrive, "Date and Time" as a date and time where the email where an email where sent, and the "Subject" that represents the title of the email [22]. During the header investigation, it was discovered that the company's website had been hacked. The hackers had gained access to sensitive customer information, including names, addresses, and credit card numbers. The company immediately took steps to notify affected customers and enhance its cybersecurity measures to prevent future breaches [23].

During the server investigation, mail servers keep copies of our emails even after we remove them from our mailboxes. On request, mail servers can be investigated, or suitable legal procedures can be followed. Servers also keep logs, which might be useful for tracing the computer/server where the transaction occurs. These logs can be examined to trace the source of the malicious activity or any suspicious behavior that occurred during the period in question [24]. During the investigation of embedded software details, software used to create emails or process emails on the server might contain sensitive information about the sender's identity and preferences. This sensitive information can include information like the sender's IP address, the time of day they sent the email, the device used to send the email, as well as account information associated with the sending address. This sensitive data can be used to identify the sender, even if they have taken steps to disguise their identity by using a pseudonym or proxy [25]. Server investigation can reveal important information about security breaches and potential threats to a company's network. By analyzing server logs and monitoring network traffic, IT professionals can identify suspicious activity and take measures to prevent further damage. It is crucial for companies to regularly conduct server investigations to ensure the safety of their sensitive data and protect against cyber-attacks [26].

During the investigation and discovery of hidden emails, a message is considered a hidden email when it is an original email that has been quoted in at least one email in a folder but cannot be found in the same folder because it was destroyed intentionally or inadvertently. Many objectives need the reconstruction and search for concealed emails, notably in forensics [27]. In addition to forensic procedures, anti-forensic operations must be considered. Once an investigation approach for cyber or email forensics is discovered, a new defensive strategy is created to fight it. Some perpetrators also employ anti-forensic tactics to thwart cyber forensic investigations. Anti-forensic operations can range from using specialized encryption algorithms and steganography to hide information, to changing or deleting timestamps of files to distort the temporal context of a criminal investigation [28]. X-headers are email headers that are appended to messages in addition to regular headers such as the Subject and To fields. These are frequently provided for spam filter content, and authorization results, and can be used to identify the email client program, such as Outlook or Opera Mail. Furthermore, the x-originating-IP header can be utilized to locate the original sender, the sender's computer's IP address. The importance of x-headers lies in the fact that they can provide very useful information to email administrators and owners. X-headers can be used to trace the source of an email, thereby allowing email administrators to prevent malicious attacks or spam from reaching their user's inboxes [29].

Message ID is a unique identifier that aids in the forensic investigation of emails all over the world. It is made up of a lengthy string of characters that concludes with the fully qualified domain name (FQDN). Client applications that send emails, such as mail user agents (MUA) or mail transfer agents (MTA), create message IDs (MTA). A message ID is composed of two components. One component comes before @, and the other comes after @. The first component of the message ID comprises data such as the timestamp of the message. This data represents the time at which the message was sent. The second component of the message ID comprises FQDN-related information. The second component of the Message ID contains information such as the domain name of the sending server. This data helps in identifying the exact source of the message and verifies the authenticity of the message. Email administrators and users need to be aware of this data, as it can help them identify spam emails [30].

In some cases, the sender's email program may include additional information about the message and associated files in the email. It can be found, for example, as a transport neutral encapsulation format (TNEF) or custom header in Multipurpose Internet Mail Extensions (MIME) content. An in-depth examination of these areas can disclose important information about the sender, such as the MAC address, the sender's Windows login username, and the PST file name. This information can help identify the source of the email and also provide clues to other accounts or files used by the sender. This type of information can be invaluable to forensic investigators trying to trace the source of an email, as it can give them a better understanding of who sent the email, what type of computer and operating system they used, and even what other accounts or files they used [31].

Large collections of mailboxes are frequently reviewed, studied, and utilized as evidence in court situations. As a result, in many circumstances, legal practitioners must deal with huge mailboxes. Most email service apps, such as Outlook and Gmail, have a dashboard with various useful functionalities. However, employing only keywords in the interface may not yield the required results. Date and time are two characteristics of emails that are required when they are offered as evidence in a lawsuit. Emails, like physical papers, may be falsified, and hackers may tamper with their characteristics. Furthermore, because an email does not travel directly from the sender to the receiver, accurately tracking its course is a difficult task. To prevent the possibility of tampering with emails, the characteristics of emails, such as date and time, need to be verified through some external source, other than keywords alone. That is why legal systems often require that dates and times of emails be verified from an external source such as the sender's email account or the server's record of the transmission. The development of digital signatures and encryption techniques also provides an extra layer of security to ensure that emails remain confidential [32].

Table 3 provides a comprehensive analysis of the various email investigation techniques. This analysis is based on a set of characteristics in terms of technique generality, comprehensiveness, accuracy, effort, and supported platforms.

Table 3. Email forensics investigation techniques.

Technique	Generic	Comprehensive	Accuracy	Effort	Supported Platform
Header Investigation	Specific	Comprehensive	Accurate	Easy to be conducted	Outlook Gmail Opera Mail
Server Investigation	Specific	Comprehensive	Accurate	Easy to be conducted	Outlook Gmail Opera Mail
Software Embedded Details Investigation	Some of these details are generic, others are specific	Not Comprehensive	Somewhat accurate	Time-consuming	Outlook Gmail Opera Mail
Discovery of Hidden Emails Investigation	Specific	Not Comprehensive	Somewhat accurate	Time-consuming	Outlook Gmail Opera Mail
Anti-forensic Activity Investigation	Some of these activities are generic, others are specific	Comprehensive	Accurate	Time-consuming	Outlook Gmail Opera Mail
Sender Mailer Fingerprints	Specific	Comprehensive	Accurate	Easy to be conducted	Outlook Gmail Opera Mail
Message ID	Specific	Comprehensive	Accurate	Easy to be conducted	Outlook Gmail Opera Mail
Embedded Software Identifier	Specific	Comprehensive	Accurate	Easy to be conducted	Outlook Gmail Opera Mail
Bulk Email Forensics	Specific	Comprehensive	Accurate	Time-consuming	Outlook Gmail Opera Mail

5. Email Analysis Tools

There is a variety of tools that can be utilized to analyze emails such as header analysis tools, Wireshark, NetworkMiner, clustering, and information gain.

5.1. Header Analysis

The header analysis is performed with a focus on the translation of the Received field to answer the questions of what, who, when, and how. Who: Who is the sender of the email? What: What is the email's subject? What is the email's Attachment file? When will the email be sent? When did you get the email? Where: What is the IP address of the Email Sender's server? and where exactly is it? How: How does the process of sending an email from sender to recipient work? [33].

5.2. Wireshark

The IMAP server email is read to record email data packets on Wireshark. The results of data packet capture on Wireshark contain information in the form of emails received with IP address 203.161.184.94, on the follow TCP stream result there is information that there is spam software running on the breaks.id.web.the host system, and finally the results of Wireshark data packet analysis are analyzed using the NetworkMiner [34].

5.3. NetworkMiner

The Packet Capture (PCAP) file analysis of Wireshark data packets was retrieved using NetworkMiner based on the follow TCP stream analysis findings acquired from the Wireshark monitoring results. PCAP file successfully extracted using NetworkMiner [34].

5.4. Clustering

Clustering is a process that involves arranging items into groups (clusters) that have two qualities. On the one hand, they are found during the procedure rather than being preset by the analyst. The class to which each item belongs is unknown in advance. Email categorization is a well-known field for distinguishing between excellent and poor emails [35].

5.5. Information Gain

The attribute selection is measured by the information gain [36]. This metric is founded on information theory. The attribute with the biggest information gain is chosen as the splitting attribute [37]. This characteristic reduces the amount of information gained required to categorize the tuples in the scores acquired. Gain (A) indicates how much we gain by branching on A. This is a predicted decrease in demand due to the information knowing OFA value. The attribute A with the biggest information gain is chosen as the dividing element. This is comparable to saying that we are looking for partition qualities that will result in the best ranking [38]. Information gain is considered an automatic detecting system based on the mathematical approach, and this method aids in calculating the importance of qualities that compose emails to classify them to locate the bad ones (forensics) [39].

Figure 5 provides an overview of the email analysis tools. These tools are header analysis, Wireshark, network minter, clustering, and information gain.

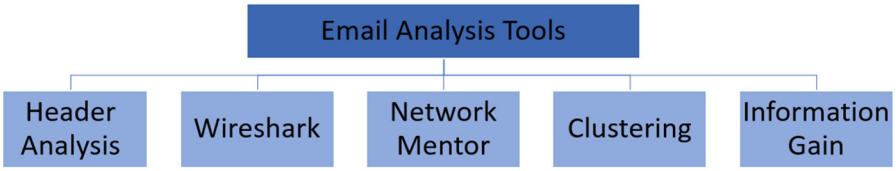

Figure 5. Email analysis tools.

5.6. Email Forensics Software Tools

This section aims to provide an overview of the most popular software tools that can be used to perform email forensics. These software tools include Aid4Mail v3.8, eMailTrackerPro V10, MailXamine V4, Paraben EMX V8.6.5277, Autopsy V4.13.0, and OSForensics V7, Encase, and FTK.

5.6.1. Aid4Mail v3.8

Add4Mail is a proprietary tool package that includes email migration, email discovery, and email archiving. It enables users to process email data for a variety of investigative objectives. We discovered that Add4Mail erased email topics from all reports and the conversion process during the testing. One of the important features of Aid4Mail is the ability to utilize scripts or import new scripts into the software for filtering reports. Some of these scripts include exporting only emails from a Gmail account, emails containing JPG/MOV/AVI files, processing deleted emails, and skipping duplicates on export [40].

5.6.2. eMailTrackerPro V10

This proprietary technology may trace an email using its header and filter spam emails. This application differs from the others in that it does not allow you to import emails from a file, folder, or database [41].

5.6.3. MailXamine V4

MailXaminer is a piece of digital forensic software that allows you to examine email messages from web and application-based email clients. The demo version enables the export of fifty emails, which was sufficient for our study. We began the experiment by importing the email file into MailXaminer like we would with other computer forensics tools. Another feature of MailXaminer is the ability to choose a date range for finding emails. This capability will assist investigators in narrowing the number of emails that must be investigated. This application has a predefined tab for searching for emails using the regular expressions search algorithm. This search function assists the investigator in detecting email message trends by using category and subcategory searches for phone numbers, URLs, addresses, postal codes, and countries. Furthermore, the program can bookmark a record of case-related evidence [42].

5.6.4. Paraben EMX V8.6.5277

Paraben E3 EMX can be used to examine message headers, contents, and attachments. Reports are generated after searching all email files or whole database folders. Maildir email format may be converted to various formats such as Electronic Mail (EML), Enterprise Message Exchange (EMX), MIME HTML (MHT), and Personal Folders file (PST) [43].

5.6.5. Autopsy V4.13.0

An autopsy is a free general-purpose digital forensics program including an email analysis capability. The program imports both plaintext and attachments from the email. It may create reports in a variety of application formats, including PDF, CVS, XML, and others [44].

5.6.6. OS Forensics V7

OS Forensics is a computer forensics program with an email forensics component. The package's trial version has a maximum index restriction of 2500 files. It may search for, and filter indexed emails based on the following fields: keyword, from, to/cc/bcc, and date (from/to). OS Forensics provides two methods for reading indexed emails as text: double-clicking on search Windows and utilizing an email-viewing application [45].

5.6.7. Encase

EnCase has long been used in forensics to retrieve evidence from confiscated hard drives. It enables the investigator to acquire evidence such as papers, images, internet history, and Windows registry information by conducting an in-depth examination of user data. EnCase is particularly useful in crime investigation because it allows investigators to conduct a thorough search of hard drives and collect digital evidence, while at the same time maintaining the integrity of that data. EnCase is considered to be the gold standard for digital forensics because of its ability to find and preserve evidence that could otherwise be overlooked [46].

5.6.8. Forensics Tool Kit

For forensic specialists, FTK provides an easy-to-use interface for email analysis. This includes the ability to scan emails for specific terms, conduct header analysis for the source addresses, and so on. File decryption, a key function of FTK, is likely the software's most prevalent application. This application provides an invaluable tool for law enforcement in helping to uncover evidence for criminal cases, often providing the missing link between a suspect and the crime. However, it is important to note that FTK is only as accurate as the user operating it, meaning it requires an experienced technician to accurately analyze and interpret the results [47].

5.7. Analysis of Email Forensics Tools

Tables 4 and 5 provide a comparison and analysis of email forensics tools based on several criteria, including language interface and user interface, creation of an image file, and calculations of the h value, cost, and advantage. The analysis of the tools in Tables 4 and 5 can assist investigators in selecting the most appropriate tool for their specific needs. It is important to consider factors such as cost, user interface, and language interface when choosing an email forensics tool. The comparison and analysis presented in these tables can assist investigators and forensic analysts in selecting the most suitable email forensics tool for their needs. It is important to consider all the criteria listed to ensure that the chosen tool meets the requirements of the investigation.

Table 4. Analysis of email forensics tools 1.

Criteria/Tool	Aid4Mail v3.8	eMailTrackerPro V10	MailXamine V4	Paraben EMX V8.6.5277
Language Interface	English	English	English	English
User Interface	Ease of use	Must have proper training	Ease of use	Ease of use
Programming Language	Python	Bash script (Linux) PowerShell (windows)	Java	Java
Creation of Image File	Support	Support	Support	Support
Calculation of Hash Value	MD5, SHA-1	RSA	MD5	MD5
Cost	Expensive	Open source software	Open source software	Open source software
Advantage	Allow users to get access to delegated, shared, and public accounts, as well as hidden MS Exchange folders.	Allows users to examine email headers to discover the genuine sender's origin.	Provides an agile keyword search engine that allows for quick finding of evidence from email data, giving users with entire information within the suspicious file via a graph-based dashboard view.	Provides powerful searching and filtering features, as well as multi-encoding support.

Table 5. Analysis of email forensics tools 2.

Criteria/Tool	Autopsy V4.13.0	OS Forensics V7	Encase	FTK
Language Interface	English	English	Chinese	Chinese
User Interface	Ease of use	Ease of use	Must have proper training	Ease of use
Programming Language	Java	Python	Java	Java
Creation of Image File	Support	Support	Support	Support
Calculation of Hash Value	MD5	MD5 and SHA	Support	Support
Cost	Expensive	Free	Expensive	Expensive
Advantage	Forensic autopsies are carried out on people who have died because of sudden, unexpected, suspicious, mysterious, unexpected, obscure, or litigious deaths.	With sophisticated file searching and indexing, it is possible to retrieve forensic evidence from computers and properly handle this data.	Taxonomic classification of digital evidence.	Various evidence searches are supported.

All of these tools have advantages that can help email investigators in finding evidence, such as MailXamine V4, Paraben EMX V8.6.5277, OS Forensics V7, and FTK tools, which provide powerful techniques for finding evidence from email data quickly. For inspecting email headers, the eMailTrackerPro V10 tool can be used to retrieve the genuine sender. Encase tool also provides a taxonomic classification of email evidence.

6. Discussion and Open Questions on Previous Empirical Study Associated with Email Issues

An essential piece of research is needed to investigate the effects of different levels of politeness modification in written English emails sent to readers [48]. This study is essential because it satisfies the need to have a better grasp of the effect readers have on the writing of emails. Szpyrka et al. [49] developed a common concept for evaluating email campaigns based on user and mail server responses, which can be seen in the article. In collaboration with FreshMail, the market leader in email marketing in Poland, the authors of this research created and tested this strategy. The plan was shown to be quite successful after being validated using the actual data from the company. It is essential to have a knowledge base representation of emails that makes use of ontology for spam filtering, and this effect has been researched to evaluate its significance [50]. The study demonstrates an innovative method for gleaning information from electronic mail and arranging it in a hierarchical ontological structure.

Improving malicious email detection with unique designated deep learning architectures utilizing whole emails has been the subject of an important study that has been undertaken to show how [51]. Extensive testing reveals that the suggested system outperforms the state-of-the-art methods for detecting malicious emails by a TPR of 5%, with an AUC of 0.993. This indicates that the proposed system is superior to the state-of-the-art approaches (including human-expert feature-based machine learning models). A study looked at the email marketing practices of the major franchise chains in the United States [52]. the study took into account multiple industries. The findings provide a fascinating comparison of the email marketing practices of large US franchise organizations operating in several different industries. Additionally, they highlight various questions and ideas for digital marketers to consider. Partridge [53] delves into the research around the evolution of the technological aspects of Internet email. Both the rules for message formatting and the protocols used to carry email between systems have, at the very least, undergone a complete overhaul at least once throughout the course of Internet history. This article tracks that development and discusses how and why things have evolved to their current condition. Specifically, it focuses on why things have developed to their current state.

In Singh et al.'s [54] paper, the authors propose using an RNN-Survival model to determine the optimal times to send emails. In this particular investigation, RNN-S was put to use to compute the probabilities of each recipient opening the subsequent email at each possible sending time. The pace at which an email is opened within a given amount of time is exactly proportional to the risk that it poses, and this rate steadily increases as more time passes. In Bahgat et al.'s [55] study, the authors offer a strategy for the effective classification of emails that is predicated on semantic methodologies. The enormous amount of retrieved textual features can be reduced by using the proposed technique, which makes use of the WordNet ontology in addition to other semantically based methodologies and similarity measures. As a result, the space and time complexity of the process can be simplified. In addition, approaches for feature selection such as principal component analysis (PCA) and correlation feature selection (CFS) have been utilized in conjunction with feature dimensionality reduction to give the shortest collection of ideal features. Experiments performed on the gold-standard benchmark Enron dataset demonstrated that the suggested semantic filtering technique, when combined with the feature selection, significantly improves the efficiency of computers by reducing the amount of space required for storage and the amount of time required for processing. Greater than 90% accuracy can be achieved across the board by using the techniques that have been implemented for feature reduction.

According to the findings of a study titled "Indicators of Employee Phishing Email Behaviours on Elaboration, Attention, and Email Typology," [56] many types of effects can be caused by phishing emails. A clustering and categorization of email contents have been provided in Alsmadi and Alhami [57], which takes a similar approach to the one described here. In addition, a thorough investigation into the identification of spam in emails through

the application of bio-inspired optimization techniques has demonstrated a high level of detection accuracy [58]. In a similar vein, an email classification system that makes use of artificial neural networks has demonstrated very high levels of performance accuracy [59].

Email problems have been employed in a variety of applications; one approach lies with the usage of email in contemporary research methodologies for the hotel and tourism industry [60]. On the other hand, it has been argued that such emails can be categorized based on the numerous functions they perform. Analysis of emails sent using techniques derived from machine learning is essential to achieving that goal [61]. Another study looks into ways to improve the effectiveness of email marketing campaigns. Emojis can be used as visual stimuli to alter the level of engagement a customer has [62]. This is very necessary to provide a direct domain for the utilization of email and, to be more specific, the application of email in marketing. It has been demonstrated that email marketing as a method for implementing strategic persuasion is of the utmost significance [63].

The use of email encompasses a wide variety of facets. A study on the use of machine learning in the problem of email spam filtering examined its various methodologies and identified several unanswered research questions [64]. Having said that, when taking into account the email statistics report for the years 2020–2024, it has been established that its implementation is strongly on the rise, and security concerns are the most essential aspect [65]. Durumeric et al. [66] conducted an empirical study on the safety of email delivery. In addition, Imşek and Aydemir's [67] research presents the categorization of unwanted electronic mail (Spam) that contain Turkish content using a variety of various algorithms. The issue of "security by any other name," which is related to the effectiveness of provider-based email security, was presented in Foster et al.'s research paper [68]. In light of the recent revelations regarding security flaws, another study [69] offers some practical suggestions for bolstering the safety of electronic communication via email. This brings up safety concerns, which, when combined with the influence of email marketing's benefits, drawbacks, and improving techniques, raises new questions [70]. As a result, an exhaustive survey for intelligent spam email detection provides a variety of problems associated with the security of email [71]. This topic has also been the subject of more research in the article phishing emails detection techniques benefits [72].

Sinha et al. [49] suggested the modeling of the time it takes for emails to be opened together with a latent state for the level of user engagement. In addition, a multi-industry, longitudinal study of the email marketing practices of the major franchise chains in the United States concluded that these practices are very important [73]. The impact of spam advertisement through email demonstrated that it is highly necessary to conduct a study to examine the influence of anti-spam software on email marketing [74]. It was discovered that machine learning techniques for spam identification in email and IoT platforms and analysis were quite essential [49]. This was determined in conjunction with the research issues. In addition, a machine learning strategy that is open and works well for filtering spam from email has been provided [75]. In a similar vein, a hybrid method that is based on machine learning is essential for the identification of spam in email [76]. In a study on methods for preventing spam in emails containing images, researchers found multiple approaches to filtering [66]. It is essential to have hybrid features that combine visual and textual information to increase the performance of spam filtering [77]. As a result, it was discovered that integrated SPAM detection for multilingual emails was also essential [78]. This is also comparable to the classification of multi-language spam phishing by the email body text towards automated security incident investigation [79]. During COVID-19, it was discovered that a critical attitude toward library marketing that was related to sending text messages and emails to online library customers was highly significant [80]. This is related to the study that was conducted for software as a service design and construction of email marketing with lower usage costs for the hospitality industry [81–83].

7. Limitations and Future Research

Several contributions have been made to science through this research, including exposing problems with email security and investigating how researchers use email to collaborate effectively. Researchers' ability to communicate, discuss ideas, and exchange information is directly related to the speed at which scientific discoveries can be made. It has also been revealed in research that data sharing must be carried out with security in mind, enabling more efficient and collaborative communication.

This study exposes those working in the field of cybersecurity to new approaches and technologies, such as encryption and digital signatures, that are aimed at securing email communication from illegal access and alteration. According to the results of this study, not all email forensics tools have the same features, so combining analysis tools might allow detailed information about email forensics to be provided. Moreover, each of these tools has advantages and disadvantages, so the choice will be determined by the investigator's needs. However, we recommend looking for the following features in an email forensic tool: the tool should be able to support more than one email format (this is because many companies use different email clients), and demonstrate speed and efficiency (the tool should be fast enough to get results, as time is of the essence when it comes to solving crimes). The tool should be capable of recovering deleted emails from all mailbox files (it should be able to identify and recover deleted emails). Moreover, the tool must produce customized litigation reports, and the results of your searches must be saved in informative and accurate reports so that these documents may be submitted in court in legally acceptable formats such as MSG, HTML, PDF, and EML.

We found that email communications are generally insufficiently protected. Globally, most email communications are subject to serious privacy and security risks. It is common for the content transmitted by email to be intercepted by third parties, putting the confidentiality, integrity, and availability of the information exchanged at risk, such as the message's text and the attached files. Although standards, protocols, and techniques exist that can enhance the security of email communications, they are not always used or implemented appropriately.

Although there is no single countermeasure that has proven to be effective against all security and privacy risks, there are mature technological solutions that, when combined and appropriately implemented, can more effectively mitigate the email risks identified in this report. It may not be enough to use one email forensics tool, but combining several tools may provide a piece of detailed information.

Several email forensic tools have been reviewed and compared. There are, however, more tools available that can be reviewed and compared for this purpose. In addition, the installation and use of these tools are not described in detail. Therefore, we recommend the following future research directions. While email security is becoming increasingly important, few studies have been conducted on email forensics tools.

To prevent the possibility of tampering with emails, email data can be stored in a ledger database such as LedgerDB. It is a centralized ledger system that provides strong auditability, tamper-evidence, and non-repudiation features similar to a blockchain. With LedgerDB, you can achieve much higher throughput than you can with blockchains. It provides greater auditability by adopting a TSA two-way peg protocol, which protects users as well as service providers against malicious behavior. In addition to ensuring verifiable data removals, LedgerDB supports the removal of obsolete records required by many real-world applications, as well as the hiding of records for regulatory purposes. As a result, ledgerDB will be more widely used in future email forensics to provide tamper-evidence and non-repudiation [84].

Further research on email forensics tools is recommended in the future. Secondly, we recommend other researchers compare different tools based on a variety of criteria, such as their disadvantages. Also, other researchers can examine more closely how email forensics tools are installed and used. Furthermore, we recommend researchers discuss the standards, protocols, and techniques for enhancing the security of email communications, and how they can be appropriately implemented.

8. Conclusions

In recent decades, email has been a major means of transporting spam and malicious content over the Internet. In addition, email is one of the most common sources of criminal activity on the Internet. The process of computer forensics involves retaining and analyzing saved emails as part of court proceedings and other civil disputes. There are a variety of fields in emails that can be altered by hackers or malicious users, as well as the flexibility of using offline email applications (e.g., Microsoft Outlook) or online email applications (e.g., Gmail). Several real email forensic incidents are reviewed, along with some proposed tools and techniques in this paper. We also discussed the major threats to email and methods for mitigating them. Several email forensic analysis techniques and tools were compared. Further, the paper compares the available software tools for email forensics according to their language interfaces, user interfaces, programming languages, the creation of image files, and hash value computations. Study results indicate that not all email forensic tools offer similar features and that by combining analysis tools, it may be possible to gather detailed information about email forensics. The goal of this study is to guide and increase awareness among users of a secure email environment and to assist fraud investigators in selecting the best email analysis tool. Furthermore, each of these tools has advantages and disadvantages, and the decision will depend on a company's or individual's needs. We recommended some general features when choosing the tool, such as support of multiple file formats, speed, efficiency, and ability to recover deleted emails and produce customized litigation reports. However, a successful investigation requires careful consideration in selecting the most suitable email forensic tool based on the organization's or individual's specific needs. The continuous advancement of technology and the diverse needs of individuals and organizations necessitate careful consideration when selecting tools. In conclusion, the study emphasizes the importance of selecting the appropriate email forensic tool based on the specific requirements of a case. It also highlights the need for continuous research and development in this field to keep up with emerging threats and evolving technologies. It is imperative that we keep up to date with emerging threats in email forensics and adapt to changes in technology so that the field will remain effective and relevant as it relates to fighting crimes committed through email.

Author Contributions: All authors equally contributed. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was funded by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia under the [GRANT 3,745].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [GRANT 3,745]. The authors would like to thank the anonymous reviewers for their insightful scholastic comments and suggestions, which improved the quality and clarity of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kumari, N.; Mohapatra, A.K. An insight into digital forensics branches and tools. In Proceedings of the 2016 IEEE International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), New Delhi, India, 11–13 March 2016; pp. 243–250.
2. Pagliaro, M. Enhancing the use of e-mail in scientific research and in the academy. *Heliyon* **2020**, *6*, e03087. [CrossRef] [PubMed]
3. Kumbhar, P.C.; Ghante, P.B. *E-mail Based Library Services: An Overview*; Zenodo (CERN European Organization for Nuclear Research): Genève, Switzerland, 2022.
4. Barik, K.; Abirami, A.; Konar, K.; Das, S. Research Perspective on Digital Forensic Tools and Investigation Process. In *Illumination of Artificial Intelligence in Cybersecurity and Forensics*; Springer Nature: Cham, Switzerland, 2022; pp. 71–95.

5. Chinnasamy, P.; Deepalakshmi, P. Improved key generation scheme of RSA (IKGSR) algorithm based on offline storage for cloud. In *Advances in Big Data and Cloud Computing*; Springer: Singapore, 2018; pp. 341–350.
6. Chinnasamy, P.; Deepalakshmi, P. Scalable multilabel-based access control as a service for the cloud (SMBACaaS). *Trans. Emerg. Telecommun. Technol.* **2018**, *29*, e3458. [CrossRef]
7. Chinnasamy, P.; Deepalakshmi, P. HCAC-EHR: Hybrid cryptographic access control for secure EHR retrieval in the healthcare cloud. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *13*, 1001–1019. [CrossRef]
8. Karim, A.; Azam, S.; Shanmugam, B.; Kannoorpatti, K. Efficient clustering of emails into spam and ham: The foundational study of a comprehensive unsupervised framework. *IEEE Access* **2020**, *8*, 154759–154788. [CrossRef]
9. Ahlborg, A. *How Mail Components on the Server Side Detects and Process Undesired Emails: A Systematic Literature Review*; DiVA portal: London, UK, 2021.
10. Vidya, K. An Overview on E-mail and Protocols Included with the IP and SMTP. *Anveshana's Int. J. Res. Eng. Appl. Sci.* **2020**, *5*. Available online: http://publications.anveshanaindia.com/wp-content/uploads/2020/03/AN-OVERVIEW-ON-E-MAIL-AND-PROTOCOLS-INCLUDED-WITH-THE-I-P-AND-SMTP.pdf (accessed on 27 February 2023).
11. Mueller, R.S.; Man With, A.C. *Report on the Investigation into Russian Interference in the 2016 Presidential Election*; US Department of Justice: Washington, DC, USA, 2019; Volume 1.
12. Caropeboka, R.M.; Effendi, J.; Wijayani, I.; Zinaida, R.S. Encrypted Email for Local Government Information Security in South Sumatra. *JINAV J. Inf. Vis.* **2022**, *3*, 109–120. [CrossRef]
13. Chhabra, G.S.; Bajwa, D.S. Review of the e-mail system, security protocols, and email forensics. *Int. J. Comput. Sci. Commun. Netw.* **2015**, *5*, 201–211.
14. Banday, M.T. Effectiveness and limitations of e-mail security protocols. *Int. J. Distrib. Parallel Syst.* **2011**, *2*, 38–49. [CrossRef]
15. Hamdi, S.D.; Radhi, A.M. Digital Cyber Forensic Email Analysis and Detection Based on Intelligent Techniques Investigation. *Iraqi J. Inf. Commun. Technol.* **2020**, *3*, 11–25.
16. Zakiyaturrahma; Riadi, I. Email Forensic from Phishing Attack Using Network Forensics Development Life Cycle Method. *Int. J. Comput. Appl.* **2022**, *183*, 36–42.
17. Salhi, D.E.; Tari, A.; Kechadi, M.T. Email classification for forensic analysis by information gain technique. *Int. J. Softw. Sci. Comput. Intell.* **2021**, *13*, 40–53. [CrossRef]
18. Ghafarian, A. An Empirical Analysis of Email Forensics Tools. 2020. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3624617 (accessed on 3 March 2023).
19. Baroto, W.A.; Prasetyo, A.H. Digital Forensic Process in Fraud Investigation: A Case Study on Email Analysis. *Int. J. Sci. Eng. Sci.* **2020**, *2*, 36–40.
20. Devendran, V.K.; Shahriar, H.; Clincy, V. A comparative study of email forensic tools. *J. Inf. Secur.* **2015**, *6*, 111–117. [CrossRef]
21. Dakheel Hamdi, S.; Merhej Radhi, A. Digital Cyber Forensics contribution for email analysis. *J. Eng. Sustain. Dev.* **2020**, *24*, 9–19. [CrossRef]
22. Baroto, W.A. Email analysis in Fraud Investigation: Digital Forensic and Network Analysis Approach. *Asia Pac. Fraud. J.* **2022**, *6*, 265. [CrossRef]
23. Hina, M.; Ali, M.; Javed, A.R.; Srivastava, G.; Gadekallu, T.R.; Jalil, Z. Email classification and Forensics Analysis Using Machine Learning. In Proceedings of the 2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI), Atlanta, GA, USA, 18–21 October 2021. [CrossRef]
24. Singh, V. Forensic Investigation of Email ARTEFACTS by using various Tools. *Int. J. Sci. Res. Develop.* **2015**, *2*, 2321–2613.
25. Vinh-Doyle, W.P. Appraising email (using digital forensics): Techniques and challenges. *Arch. Manuscr.* **2017**, *45*, 18–30. [CrossRef]
26. Umar, R.; Riadi, I.; Muthohirin, B.F. Live forensics of tools on android devices for email forensics. *TELKOMNIKA Telecommun. Comput. Electron. Control.* **2019**, *17*, 1803–1809. [CrossRef]
27. Minchev, Z. Data Relativities in the Transcending Digital Future. In Proceedings of the 10th International Conference on Business Information Security (BISEC-2018), Belgrade, Serbia, 20 October 2018; pp. 6–9. [CrossRef]
28. Armknecht, F.; Dewald, A. Privacy-preserving email forensics. *Digit. Investig.* **2015**, *14*, S127–S136. [CrossRef]
29. Banday, M.T. Design and Development of E-mail Security Protocols and Forensic Tools: A Research Proposal. In Proceedings of the International Conference on Recent Advances in Electronics and Computer Engineering, Himachal Pradesh, India, 17–18 December 2011.
30. Mustafa, M.; Riadi, I.; Umar, R. Header investigation for spam email forensics using the framework of the national institute of standards and technology. *ILKOM J. Ilm.* **2021**, *13*, 163–167. [CrossRef]
31. Purwiantono, F.E.; Tjahyanto, A. Classification Model for Detection of Phishing Sites in Indonesia. *J. Theor. Appl. Inf. Technol.* **2017**, *95*, 4181–4191.
32. Mandowen, S.A. Forensic Analysis of Computers on Network Traffic. *MIPA Dan Pengajarannya* **2016**, *16*, 14–20.
33. Suryana, A.L.; El Akbar, R.; Widiyasono, N. Investigation of Email Spoofing with the Digital Forensics Research Workshop (Dfrws) Method. *J. Inform. Educ. Res.* **2016**, *2*, 111–117. [CrossRef]
34. Sayal, M.A.; Alameady, M.H.; Albermany, S.A. The Use of SSL and TLS Protocols in Providing a Secure Environment for e-commerce Sites. *Webology* **2020**, *17*, 503–523. [CrossRef]

35. Müller, J.; Brinkmann, M.; Poddebniak, D.; Böck, H.; Schinzel, S.; Somorovsky, J.; Schwenk, J. "Johnny, you are fired!"-Spoofing OpenPGP and S/MIME Signatures in Emails. In *USENIX Security Symposium*; USENIX: Santa Clara, CA, USA, 2019; pp. 1011–1028.
36. Susanto, B.M. *Identification of Phishing Websites with Attribute-Based Selection*; Foundation of Computer Science (FCS): New York, NY, USA, 2016; pp. 18–19.
37. Kurniawan, A. Application of Owasp Framework and Network Forensics for Analysis, Detection, and Prevention of Injection Attacks on the Host-Based Side. *Jurnal Telematika* **2019**, *14*, 9–18.
38. Liu, E.; Akiwate, G.; Jonker, M.; Mirian, A.; Ho, G.; Voelker, G.M.; Savage, S. Forward Pass: On the Security Implications of Email Forwarding Mechanism and Policy. *arXiv* **2013**, arXiv:2302.07287.
39. Akanksha, K.; Utkarsha, Z.; Sneha, K.; Sanika, C.; Kazi, K. Email Security. *J. Image Process. Intell. Remote Sens.* **2022**, *2*, 23–31. [CrossRef]
40. Najib, A.F.; Rachmawanto, E.H.; Sari, C.A.; Sarker, K.; Rijati, N. A comparative study MD5 and SHA1 algorithms to encrypt REST API authentication on mobile-based application. In Proceedings of the 2019 IEEE International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 July 2019; pp. 206–211.
41. Hoiriyah, H.; Sugiantoro, B.; Prayudi, Y. Investigasi Forensik pada E-mail Spoofing menggunakan Metode Header Analysis. *Data Manajemen Dan Teknologi Informasi (Dasi)* **2016**, *17*, 20–25.
42. Sah, A.; Riadi, I.; Prayudi, Y. Deteksi Bukti Digital Online Gambling Menggunakan Live Forensik Pada Smartphone Berbasis Android. *Cyber Security Dan Forensik Digital.* **2018**, *1*, 14–19. [CrossRef]
43. Hazel, P. *Exim: The Mail Transfer Agent*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2001.
44. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*, 2nd ed.; University of Illinois at Urbana Champaign, Morgan Kaufmann: San Mateo, CA, USA, 2006.
45. Williams, G.J.; Simoff, S.J. (Eds.) *Data Mining: Theory, Methodology, Techniques, and Applications*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3755.
46. Kodratoff, Y. Technical and scientific issues of KDD (or: Is KDD a science?). In *International Workshop on Algorithmic Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 261–265.
47. Garfinkel, S.L.; Margrave, D.; Schiller, J.I.; Nordlander, E.; Miller, R.C. How to make secure email easier to use. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; ACM Press: New York, NY, USA, 2005; pp. 701–710.
48. Kapadia, A. A case (study) for usability in secure email communication. *IEEE Secur. Priv.* **2007**, *5*, 80–84. [CrossRef]
49. Karim, A.; Azam, S.; Shanmugam, B.; Kannoorpatti, K.; Alazab, M. A comprehensive survey for intelligent spam email detection. *IEEE Access* **2019**, *7*, 168261–168295. [CrossRef]
50. Hendriks, B.; van Meurs, F.; Kakisina, B. The effects of L1 and L2 writers' varying politeness modification in English emails on L1 and L2 readers. *J. Pragmat.* **2023**, *204*, 33–49. [CrossRef]
51. Szpyrka, M.; Suszalski, P.; Obara, S.; Nalepa, G.J. Email Campaign Evaluation Based on User and Mail Server Response. *Appl. Sci.* **2023**, *13*, 1630. [CrossRef]
52. Bindu, V.; Thomas, C. Knowledge Base Representation of Emails Using Ontology for Spam Filtering. *Adv. Intell. Syst. Comput.* **2021**, *1133*, 723–735.
53. Muralidharan, T.; Nissim, N. Improving malicious email detection through novel designated deep-learning architectures utilizing entire email. *Neural Netw.* **2023**, *157*, 257–279. [CrossRef] [PubMed]
54. Abrahams, A.; Chaudhary, T.; Deane, J. A multi-industry, longitudinal analysis of the email marketing habits of the largest United States franchise chains. *J. Direct Data Digit. Mark. Pract.* **2010**, *11*, 187–197. [CrossRef]
55. Partridge, C. The technical development of internet email. *IEEE Ann. Hist. Comput.* **2008**, *30*, 3–29. [CrossRef]
56. Singh, H.; Sinha, M.; Sinha, A.R.; Garg, S.; Banerjee, N. An RNN-Survival Model to Decide Email Sends Times. *arXiv* **2020**, arXiv:2004.09900.
57. Bahgat, E.M.; Rady, S.; Gad, W.; Moawad, I.F. Efficient Email Classification Approach Based on Semantic Methods. *Ain Shams Eng. J.* **2018**, *9*, 3259–3269. [CrossRef]
58. Buckley, J.; Lottridge, D.; Murphy, J.G.; Corballis, P.M. Indicators of employee phishing email behaviors: Intuition, elaboration, attention, and email typology. *Int. J. Hum. Comput. Stud.* **2023**, *172*, 102996. [CrossRef]
59. Alsmadi, I.; Alhami, I. Clustering and classification of email contents. *J. King Saud Univ. Comput. Inf. Sci.* **2015**, *27*, 46–57. [CrossRef]
60. Batra, J.; Jain, R.; Tikkiwal, V.A.; Chakraborty, A. A Comprehensive Study of Spam Detection in E-Mails Using Bio-Inspired Optimization Techniques. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100006. [CrossRef]
61. Alghoul, A.; Al Ajrami, S.; Jarousha, A.G.; Harb, G.; Abu-Naser, S. Email classification using artificial neural network. *IJAER* **2018**, *2*, 8–14.
62. Cobanoglu, C.; Nanu, L.; Ciftci, O.; Berezina, K.; Cavusoglu, M.; Ali, F. *Contemporary Research Methods in Hospitality and Tourism*; Emerald Publishing Limited: Bingley, UK, 2022.
63. Iqbal, K.; Khan, M.S. Email classification analysis using machine learning techniques. *Appl. Comput. Inform.* 2022; *ahead-of-print*.
64. Valenzuela-Gálvez, E.S.; Garrido-Morgado, A.; González-Benito, Ó. Boost your email marketing campaign! emojis as visual stimuli to influence customer engagement. *J. Res. Interact. Mark.* **2022**, *3*, 337–352. [CrossRef]

65. Thomas, J.S.; Chen, C.; Iacobucci, D. Email Marketing as a Tool for Strategic Persuasion. *J. Interact. Mark.* **2022**, *57*, 377–392. [CrossRef]
66. Dada, E.G.; Bassi, J.S.; Chiroma, H.; Abdulhamid, S.M.; Adetunmbi, A.O.; Ajibuwa, O.E. Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems. *Heliyon* **2019**, *5*, e01802. [CrossRef]
67. The Radicati Group. Email Statistics Report, 2020–2024—Executive Summary. 2020. Available online: https://www.radicati.com/wp/wp-content/uploads/2019/12/Email-Statistics-Report-2020-2024-Executive-Summary.pdf (accessed on 3 March 2023).
68. Durumeric, Z.; Adrian, D.; Mirian, A.; Kasten, J.; Bursztein, E.; Lidzborski, N.; Thomas, K.; Eranti, V.; Bailey, M.; Halderman, J.A. Neither snow nor rain nor MITM: An empirical analysis of email delivery security. In Proceedings of the ACM Internet Measurement Conference, New York, NY, USA, 28–30 October 2015; Association Computing Machinery: Washington, DC, USA, 2015; pp. 27–39.
69. Şimşek, H.; Aydemir, E. Classification of Unwanted E-Mails (Spam) with Turkish Text by Different Algorithms in Weka Program. *J. Soft Comput. Artif. Intell.* **2022**, *3*, 1–10. [CrossRef]
70. Foster, I.D.; Larson, J.; Masich, M.; Snoeren, A.C.; Savage, S.; Levchenko, K. Security by any other name: On the effectiveness of provider based email security. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS), New York, NY, USA, 12 October 2015; Association Computing Machinery: Washington, DC, USA, 2015; pp. 450–464.
71. Malatras, A.; Coisel, I.; Sanchez, I. Technical recommendations for improving the security of email communications. In Proceedings of the 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 30 May–3 June 2016; pp. 1381–1386.
72. Fariborzi, E.; Zahedifard, M. E-mail Marketing: Advantages, Disadvantages and Improving Techniques. *Int. J. e-Educ. e-Bus. e-Manag. e-Learn.* **2012**, *2*, 232. [CrossRef]
73. Muneer, A.; Ali, R.; Al-Sharai, A.; Fati, S. A Survey on Phishing Emails Detection Techniques. In Proceedings of the 2021 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 9–10 November 2021; pp. 1–6.
74. Sinha, M.; Vinay, V.; Singh, H. Modeling Time to Open of Emails with a Latent State for User Engagement Level. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Los Angeles, CA, USA, 5–9 February 2018; pp. 531–539.
75. Mostafa, R.; Norizan, M.Y.; Gazi, M.A. Impact of spam advertisement through e-mail: A study to assess the influence of the anti-spam on the e-mail marketing. *Afr. J. Bus. Manag.* **2010**, *4*, 2362–2367.
76. Ahmed, N.; Amin, R.; Aldabbas, H.; Koundal, D.; Alouffi, B.; Shah, T. Machine Learning Techniques for Spam Detection in Email and IoT Platforms: Analysis and Research Challenges. *Secur. Commun. Netw.* **2022**, *2022*, 1862888. [CrossRef]
77. Bansal, C.; Sidhu, B. Machine Learning based Hybrid Approach for Email Spam Detection. In Proceedings of the 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 3–4 September 2021; pp. 1–4.
78. Dhanaraj, S.; Karthikeyani, V. A study on e-mail image spam filtering techniques. In Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, Salem, India, 21–22 February 2013; pp. 49–55.
79. Nam, S.G.; Jang, Y.; Lee, D.G.; Seo, Y.S. Hybrid Features by Combining Visual and Text Information to Improve Spam Filtering Performance. *Electronics* **2022**, *11*, 2053. [CrossRef]
80. Iyengar, A.; Kalpana, G.; Kalyankumar, S.; GunaNandhini, S. Integrated SPAM detection for multilingual emails. In Proceedings of the 2017 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 23–24 February 2017; pp. 1–4.
81. Rastenis, J.; Ramanauskaitė, S.; Suzdalev, I.; Tunaityte, K.; Janulevicius, J.; Cenys, A. Multi-Language Spam/Phishing Classification by Email Body Text: Toward Automated Security Incident Investigation. *Electronics* **2021**, *10*, 668. [CrossRef]
82. Rysavy, M.D.; Michalak, R.; Daly, B. Library marketing: Sending text messages and emails to online library users during COVID-19. *J. Libr. Adm.* **2021**, *61*, 358–365. [CrossRef]
83. Sukarsa, I.M.; Buana, P.W.; Arynasta, I.P.K. Software as a Service: Design and Build Lower Usage Cost Email Marketing for Hospitality Industry. *Sci. J. Inform.* **2020**, *7*, 189–202.
84. Yang, X.; Zhang, Y.; Wang, S.; Yu, B.; Li, F.; Li, Y.; Yan, W. LedgerDB: A centralized ledger database for universal audit and verification. *Proc. VLDB Endow.* **2020**, *13*, 3138–3151. [CrossRef]