Analytical study of email security systems

ARUNDHATI MENON 20BCY10031

July 2022

1 Introduction

2 Related Work

Eltigani B. Abdelsatir et al. [14] incorporated the Cocks IBE scheme with the current emailing services framework. The study considered Identity-Based Encryption (IBE) without pairing while accounting for a secure email system. The suggested approach offered a secure email environment that does away with PKI difficulties and allowed for simple key generation and administration. To summarise the work on IBE implementation, a quick explanation of the mathematical underpinnings of factoring-based IBEs was presented along with a discussion of the Cocks IBE scheme. The suggested secure email system's layout and structure was then presented.Florian Holzbauer et al. [22] investigated the impact of real-world deployments in the field of email security and SMTP and described how these technologies keep up with the increasing complexity of delivery- and security options in an email system. The paper also elaborated on the measurements and procedures to evaluate core aspects of email delivery, including security features, DNS configuration, and IP version support across different types of providers and based on a data survey conducted by the researchers numerous useful inferences were drawn and presented in the form of statistical data. Jianjun Chen et al. [18] elaborated upon 18 exploits through a series of case studies to help deploy proper mitigation techniques which may prove effective in preventing large-scale email based attacks on organizations and individuals. The paper also highlighted several suggestions at the end-users, email clients and email provider levels to help mitigate the attacks mentioned in the study.

Alex Sumner et al.[38] provided a general summary of social engineering assaults, as well as information on how to spot them, how to educate and train others to avoid them, and how vulnerable people are to them. The mitigation techniques were reviewed upon considering 3 aspects- detection, education and training, and susceptibility. Alessandro Ecclesie Agazzi [20] focused on four distinct levels of defence against social engineering attacks through the use of automated tools and decision-aid tools. The importance of users' expertise and knowledge in addressing potential dangers was also thoroughly covered. To

emphasise and offer a second line of defence against spear phishing and other forms of phishing, the significance of having multi-factor authentication (MFA) was emphasised. Sudarshan S. Chawathe [17] elaborated on a semi-automated, rule-based system tool to improve email security and identify malicious email communications. The design and implementation of a system for enhancing email security using fuzzy rules, as well as an experimental investigation of the effectiveness and efficiency of such a system on a well-known real data set, were the key contributions of the research. The results of the studies revealed that choosing a fuzzy rule-based system has no appreciable disadvantage because its performance was consistently close to that of the best-performing classifier.

Sabah Al-Fadeghi et al. [15] focused on defining the bounds of email system security and creating a representation by developing a diagrammatic representation as a means to describe computer network facilities. The overall goal was to illustrate the security system's representation, which is crucial in making thinking apparent. To do this, a conceptual description of the operational environment, or the area where actual security activities are carried out, was provided. During the investigation, the suggested model for email security was also used to conceptually describe a straightforward email system. Gurpal Singh Chhabra et al. [19] analysed the architecture and operation of the present email system as well as the security protocols commonly used to secure email interactions. The study also exposed the flaws in email systems and used additional email forensics to examine email contents, header data, email transit paths, sender or recipient information, and other information to gather proof. The researchers elaborated upon common email forensic investigation techniques and tools used in the email forensic process. Rusydi Umar et al. [40] provided an insight into the procedures and practices adhered to by the National Institute of Standards and Technology (NIST) to procure and retrieve sensitive data such as e-mail delivery timestamp, e-mail recipient timestamp, sender protocol port, recipient protocol port, and source address IP and destination IP address from Android-based email services and compared the forensic instruments used to obtain the pieces of evidence using two tools- Wireshark and Networkminer. Tianlin Li et al. [29] discussed the analysis conducted on five popular public email systems - Gmail, Hotmail, Yahoo, 163, and Sohu to analyse the security and usability of the mentioned systems. The criteria used in the paper to compare the reliability and usability of the five email systems included password management and the multiple sign-in features of each structure. Detailed studies for Password Composition policies, Secure Password transmissions, Password recovery mechanisms, and Password reset Mechanisms were included for each email system. Under the second criteria, the distinguishing features were recorded based on Adding Two Accounts without the User's consent, Storing files under the wrong google drive, and effects of multiple sign-in on Gmail services completed via the introduction of two algorithms from which numerous inferences were drawn.

Deep semantic analysis, machine learning, and deep learning approaches were used by Sikha Bagui et. al [16] to identify emails as phishing or non-phishing and to capture the inherent properties of the email text. Deep semantic analysis using machine learning and deep learning approaches was used

to arrange emails using a one-hot encoding strategy with and without word-phrasing. Yahao Zhang et al. [42] studied the filtering of information systems, and proposed a method based on Bayesian optimization to filter the mail transmission under the hoop topology to improve the filtering level of phishing mail-in information. The findings were simulated using MATLAB and the Bayesian optimization model developed could successfully meet the needs of phishing email screening in information transmission. S. Srigayathri et al. [37] suggested the use of the SWEET protocol to filter spam messages effectively. At the SWEET server, unsolicited bulk emails were to be filtered out of the server and blocked before they reached the email server. The IP address of that spam message was identified and prevented from receiving spam messages in the future. The study also offered a method for pinpointing the precise spam messages and locating the spammer and guaranteed that using the aforementioned approach can reduce the number of spam messages and altogether reduce the risk of productivity loss, bandwidth and storage usage. Jen-Wei Huang et al. [26] suggested an email security categorization system that extracts data from emails and enables its transformation into a multidimensional vector using an artificial neural network. The suggested system was assessed in a real corporate environment. The outcomes revealed that the suggested feature extraction method represented email data more accurately in true positive rates and F1 scores than existing methods. Shafi'i Muhammad Abdulhamid et al. [36] aimed to assess the effectiveness of various classification algorithms that are used to categorise emails as spam or not spam, such as Bayesian Logistic Regression, Hidden Naive Bayes, RBF Network, Voted Perceptron, Lazy Bayesian Rule, Logit Boost, Rotation Forest, NNge, Logistic Model Tree, REP Tree, Nave Bayes, J48, Multilayer Perceptron, and Random Tree. The research also presented the methods and materials employed for the cause of the study and provided detailed results and inferences as a conclusion.

Three important problems about email spoofing were addressed by Hang Hu et al. [25]. (1) How do email providers identify and deal with fraudulent emails? (2) Under what circumstances do fake emails manage to get past the user's defence and into their inbox? (3) How do email service providers alert consumers after the fake email is received? Through thorough user research that included both modelled and real-world phishing experiments, end-to-end measurements on 35 leading email providers, and the responses to the questions. Scott Ruoti et al. [34] was the first to use a consistent quantitative metric for cross-system comparisons and do a comparative (i.e., A/B) usability evaluation of three alternative key management schemes. Additionally, the study provided qualitative participant feedback that offered insightful information on user perceptions of each key management strategy and secure email. The study served as a template for future research on encrypted email because it used A/B experiments, common measurements, and a two-person design. A. Malatras et al. [31] analysed the current privacy and security issues in international email exchanges and provided a list of workable solutions based on variations of current standards that can successfully mitigate the risks found. Based on this analysis, the researcher offered a list of technical recommendations for email providers to go by

to improve security while maintaining ecosystem compatibility. The paper also proposed a set of security features to be considered as minimum requirements for a future inter-operable and secure email system. Peter Menegay et al. [32] detailed email services that optionally allowed communications to be routed across a blockchain using well-known client programs like MS Outlook or Thunderbird. Similar issues were raised when discussing chat programs that employed the IRC protocol and a unique web-based protocol based on Socket. IO. The capacity of many parties to sign documents and release payments was also demonstrated in a MIPR application. The integrity of the blockchain was discussed together with improved software to enable speedy recovery from a breach, which was described as inherent in the distributed nature of the technology. Zhuorao Yang et al. [41] aimed to extract 18 features, including psychological features, header-based features, URL-based features, and script-based features from a data set compiling the features of a proposed model of an SVM classifier. The study analysed the email header structure, URL data, and email-script functionality to extract the features. The suggested method had a 9 per cent false-positive rate, a 99 per cent true-positive rate, 91.7 per cent precision, and 95.0 per cent accuracy overall. The analysis also evaluated the usefulness of the suggested psychological traits.

Author, Year	Key contribution	Threats and attacks mentioned	Mitigation techniques discussed	Model/framework proposed	Authentication/encryption algorithm used	ML/DL concept applied	Security of current email systems discussed	Email security protocols discussed	Forensic tools were used
Sabah Al-Fadeghi et al., 2020	Developed a modelling language to build a security foundation for secure email system architecture.	No	No	Yes*	No	No	No	No	No
Alex Sumner et al., 2019	Educated the readers of the growing risk of phishing attacks and provided mitigation techniques	Yes	Yes	No	No	No	No	No	No
Hang Hu et al., 2018	Conducted real-world experiments in order to evaluate the effectiveness of user-level protections against email spoofing	Yes	No	No	Yes	No	Yes	No	No
Sikha Bagui et al., 2019	Classified and differentiated between phishing and non-phishing mails through the application of machine learning and deep semantic analysis.	Yes	No	Yes	Yes	Yes	No	No	No
Gurpal Singh Chhabra et al., 2019	Analyse security protocols in currently used e-mail systems to list out the limitations and highlight e-mail forensic tools used to recover or manage damaged emails	No	No	No	No	No	Yes	Yes	Yes
Florian Holzbauer et al., 2022	Demonstrated protocol support and compliance in email ecosystems and, based on the findings, provided email system operators to decrease the distribution of SPAM	Yes	No	Yes	Yes	No	Yes	Yes	No
Yahao Zhang et al., 2022	Introduced a method based on Bayesian optimization to filter mail transmission under the hoop topology	Yes	Yes	Yes	Yes	No	No	Yes	No
Alessandro Ecclesie Agazzi, 2020	Reviewed the workings and inner mechanisms of Phishing and Spear Phishing attacks through 5 steps to detect and mitigate attacks	Yes	Yes	Yes*	Yes	No	No	No	No
Jianjun Chen et al., 2020	Analysed email security protocols through simulations and case studies to highlight composition issues in existing email security protocols	No	No	Yes	Yes	No	Yes	Yes	No
Rusydi Umar et al., 2019	To obtain digital evidence from tools that are subjected to Android-based email services.	No	No	No	No	No	No	No	Yes
Scott Ruoti et al., 2018	Compared three key management systems from various families to better understand how key management affects the usability of encrypted email during system setup.	No	No	Yes*	Yes	No	Yes	No	No
Eltigani B. Abdelsatir et al., 2020	Designed a concept for a time and space-saving email system that integrates existing encryption systems with Cocks IBE to get rid of key management problems with PKI.	No	No	Yes	Yes	No	Yes	No	No
S. Srigayathri et al., 2018	Discovered strategies to filter out unsolicited bulk email on the server by proposing a practical model for spam email removal.	Yes	Yes	Yes*	Yes	No	Yes	Yes	No
A. Malatras et al., 2018	Analysed the preservation of security and privacy of email communications by reviewing the design of email systems and pertinent communication protocols.	Yes	Yes	Yes*	No	No	Yes	Yes	No
Tianlin Li et al., 2018	Provided better understanding of security and usability of email systems by conducting a comparative analysis of popular email systems.	No	No	Yes*	No	No	Yes	Yes	No
Sudarshan S. Chawathe, 2018	Experimentally studied the effectiveness of a fuzzy rule-based classifier on a real dataset and compare its results with the results from other classifiers	Yes	Yes	Yes*	No	Yes	No	Yes	No
Jen-Wei Huang et al., 2018	Proposed an effective system to assist in the security level classification of corporate emails using text mining and machine learning techniques	No	Yes	Yes	Yes	Yes	Yes	No	No
Peter Menegay et al., 2018	Applied blockchain technology to enhance email and communication security and proposed an encryption technique using blockchain.	No	No	Yes	Yes	Yes	Yes	Yes	No
Shafi'i Muhammad Abdulhamid et al., 2018	Performed a comparative analysis of classification algorithms in the field of email spam detection and filtering to determine which algorithm performs best for any chosen metric	Yes	Yes	No	No	No	No	Yes	No
Zhuorao Yang et al., 2019	Proposed an SVM module based on 18 salient features to train and classify the emails and detect phishing emails effectively	Yes	Yes	Yes	No	No	No	Yes	No

Table 1: Summary of research papers

3 Methodology

3.1 Background

3.1.1 Definition

Email security refers to the contemporary security technologies and techniques adopted to safeguard private information and account details surrounding a user's email account. Frequent use of email involves disseminating and distributing malware, spam, and phishing assaults. Threat agents and attackers often use deceptive communications to convince users to provide sensitive information, open attachments, or open hyperlinks that install malware on the victim's computer. Emails are a popular access point for attackers attempting to gain entry to a network and steal valuable company data. Since email is a crucial part of today's business environment, firms have introduced and established policies to govern the data flow across the network. Viewing the contents of emails coming via their email servers is one of the first regulations most corporations set. It is critical to comprehend the entire email to behave effectively. Subsequently, an organization can enact various security measures on those emails.

3.1.1.1 How secure is email?

Emails are the most prevalent mode of communication among organizations. Communications between employees of an organization and between employees and vendors are all carried out using email as a medium of propagation. Despite the use of emails as a frequent mode of online communication, it is not advisable to share sensitive information and documents through emails. The use of encryption techniques can help safeguard the message's body, but it takes both the sender and receiver to set it up ahead of time and some additional technical knowledge. While encrypting an attachment is simpler, mail systems may discard them as malicious files since inspection of these attachments for safety is a tedious task. Attackers can compromise email accounts through phishing attempts or other means, exposing your email conversations to cyber criminals. Interception of email messages and attachments while travelling via the email network is frequently observed. Emails are not encrypted by default as they travel from your email servers to the recipient.

3.1.1.2 Common email security practices

This section describes standard and effective practices that a user can adopt to safeguard emails and personal information from the prying eyes of malicious hackers.

- **Strong Passwords:** Passwords are the forefront vanguard defence against illegal access to users' emails and personal information. To ensure the

safety of user messages, the use of strong and complex passwords has become mandatory. A password is impenetrable in nature if it satisfies all the requirements given below:

  - Contains at least 8-12 characters.
  - Combination of lowercase, numbers, uppercase, and special characters(*, @, !, etc.)
  - Use different passwords for different accounts or services.
  - Do not use personal information (DOB, address, name, SSN, etc.) as passwords

- **Two-Factor Authentication:** Two-Factor Authentication (2FA) adds an extra layer of security to users' email accounts.[8] In addition to the username and password, Two-Factor Authentication aims to provide the user email account additional security through user-defined pin codes, passwords or secret questions. It is an email security technique adopted by all email service providers across all platforms.

- **Phishing emails:** Phishing is a form of social engineering attack where the attacker sends out fraudulent messages, pretending to be from legitimate sources, in bulk to a victim, intending to gain unauthorized access to the user's confidential information. Emails are the most common attack vector threat agents follow to spread malicious links and codes among email users, thus, making it crucial for users to be vigilant and meticulous about spam emails.

- **Logging out:** Logging out of the email account can help user safeguard their emails and messages from unwanted modifications and prevent passive attacks like spoofing and sniffing.

3.1.2 Email Security Protocols

Maintaining the confidentiality and integrity of the email content is of utmost priority to email service providers. The adoption of various email security protocols ensures secure transmission of messages. Email security protocols refer to a defined set of rules and guidelines that an email server must adhere to protect the emails sent via the server from unauthorized access and modification. In the case of email, the source, receivers, and servers may all be different, but they must all receive the data, interpret the information, and render it in the same manner as the sender. Email protocols describe how an email message must be encoded, delivered, received and rendered, thus making protocols vital. While email protocols complicate sending emails, they ensure that email is a standard, dependable, and ubiquitous form of communication. Contemporary email servers and service providers adopt numerous email security protocols to ensure the safety and integrity of the data shared through emails.

3.1.2.1 Common email security protocols

To enhance email security and message confidentiality, security companies and service providers have introduced numerous email security protocols. There are several factors to email security. Email security technologies such as SMTP, S/MIME, OpenPGP, and others encrypt data in transit to prevent domain spoofing. These security methods also authenticate delivered messages from valid domains to guarantee content security and protect email message integrity. Other commonly used email security protocols include:

- STARTTLS
- SMTP MTA-STS
- SSL/TLS for HTTPS
- DMARC
- Digital Certificates

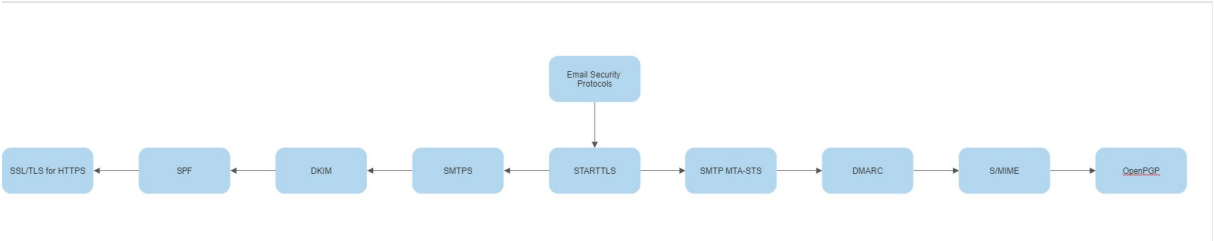

Figure 1: Email security Protocols

The upcoming sections briefly describe 3 of the most popularly used email security protocols.[10]

3.1.3 Simple Mail Transfer Protocol (SMTP)

SMTP is an application layer protocol that governs the delivery of messages[13]. The client that wishes to send the email establishes a TCP connection with the SMTP server and then transmits the email over the connection[13]. The SMTP process initiates a connection through port 25 as soon as it detects a TCP connection from any client[28]. The client sends the email immediately after successfully establishing a TCP connection[13]. The working of the SMTP is shown graphically in Figure 2.

3.1.4 Secure/Multipurpose Internet Mail Extension (S/MIME)

S/MIME is an industry standard for email encryption and signature technique used by enterprises to increase and enhance email security[12]. It defines how

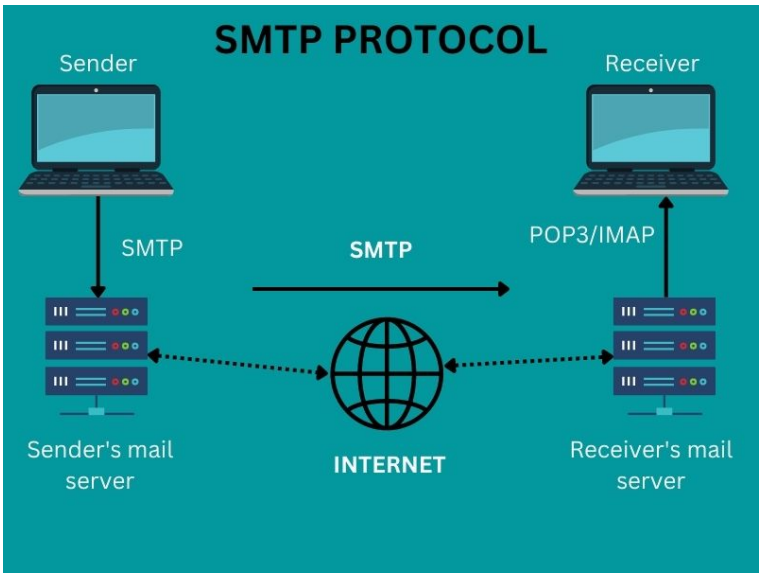

Figure 2: Working of SMTP Protocol

communications are formatted. It also encrypts and digitally signs emails to ensure their authenticity. The working of the S/MIME is shown graphically in Figure 3.

3.1.5 Post Office Protocol 3 (POP3)

Post Office Protocol, or POP, is an application-layer protocol that email clients utilise to retrieve email from a mail server[4]. POP version 3 (POP3) is the most often used[7]. This protocol specifies how email clients recover messages from SMTP servers. POP3 clients connect to the server, retrieve all messages, store them on the client computer, and then proceeds to delete them[7].

3.2 Problem Statement

Email is a prominent target for cyber attackers because of its vulnerabilities firmly based on human mistakes and the characteristics of the data communicated. The annual Data Breach Investigations Report (DBIR) published by Verizon in 2020 identifies email and direct installs as the top two attack vectors for malware infections[11]. According to data published by PurpleSec in their 2021 Cyber Security Statistics report, a staggering 92 per cent of all malware is delivered via email.[11]

According to the report by Statista Research Department, Jul 7, 2022,[5] phishing emails were the most common source of ransomware infection, according to 54 per cent of responding MSPs. The statistic is represented in a graphical format by Figure 4.

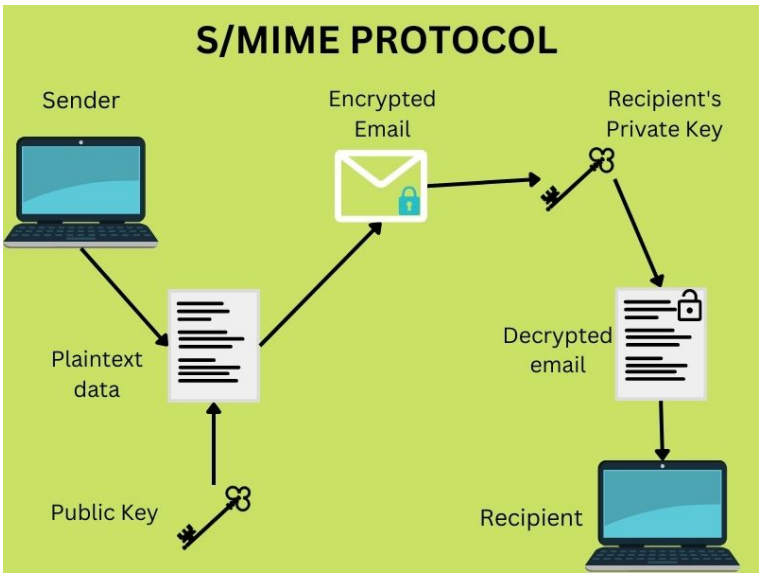

Figure 3: Working of S/MIME Protocol

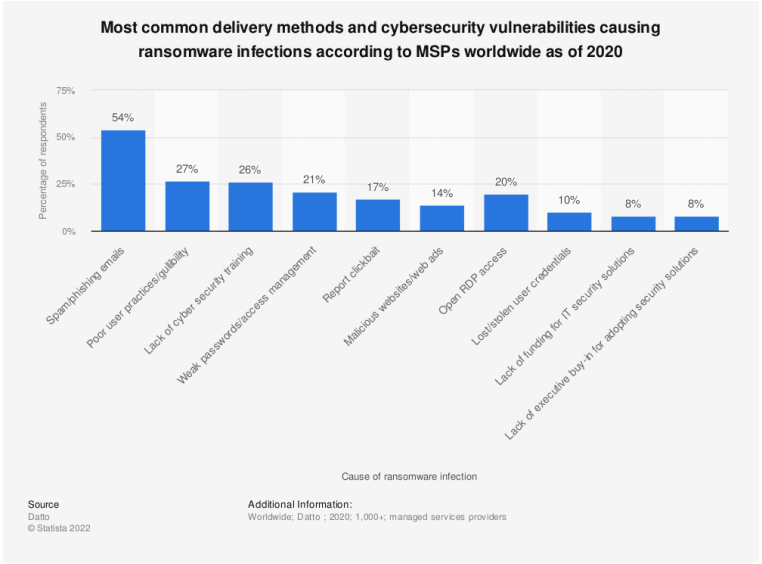

Figure 4: Most common delivery methods and cybersecurity vulnerabilities causing ransomware infections according to MSPs worldwide as of 2020

In today's digital age, email has become a vital component of our lives, as practically all business communication takes place through it. Despite the ease with which emails are utilised, it is critical for an organisation not to get complacent about protecting the data transmitted via emails since this can do catastrophic damage to its business. With the increasing threat of hackers, viruses, spam, phishing, identity theft, and ransomware attacks, businesses have an increased obligation to secure their corporate data and prioritise email security.[27]

While the internet and email provide several corporate benefits, they also represent a range of possible security risks. Malware, spam and phishing, social engineering, and destructive unauthorised access are some prevalent email security concerns.[3]

Recent studies and analytical approaches have revealed potential and probable solutions for the above-mentioned issues and were presented in great detail in Table 1. Through a practical approach, researchers proposed a solution to the issue of malicious emails and phishing acts in[17][20][38]. Research papers [22][18][14] briefly describe upon the need for email authentication and security protocols like SMTP in order to overcome the issues of email authenticity and security. Upon further analysis it is established that [16][42][37][26][36] focus on the parameter 5 of Table 1,i.e., machine learning, deep learning, and ANNs but fail to adhere to parameters 3,7,8. Though numerous authors presented distinct processes, the possibility of a hybrid mechanism with the intention of integrating email security protocols and deep learning techniques was not investigated, and it is one area that writers could investigate in the near future.

3.3 Discussion of Existing Solutions

Under this section solutions discussed in mentioned research papers and literature have been compared and analysed to establish and recognize potential solutions that will help secure email systems and servers.

3.3.1 Comparison

Author, year	Cryptographic Method/Protocols Used	Effectiveness of Cryptographic method
Hang Hu et al., 2018	SPF/DMARC on Alexa 1 million domain	w/ SPF 493,367 (49.3%), w/ valid SPF 449,848 (45.0%), w/ DMARC 46,159 (4.6%), w/ valid DMARC 45,580 (4.6%) on all domains
Rusydi Umar et al., 2019	PKD and Identity-Based Encryption(IBE)	SUS Percentage for IBE-81% and for PKD-76%
Zhuorao Yang et al., 2019	Hybrid feature proposed in paper compared to 17 existing features and protocols.	Suggested method has a 9 percent false-positive rate, a 99 percent true-positive rate, 91.7% precision, and 95.0% accuracy overall

Table 2: Analysis of cryptographic methods

Author, Year	Mitigation Technique Used	RESULTS
Alex Sumner et al., 2019	Distraction(DIS) and Authority persuasion(AUTH)	In the top six ranked emails displayed, there were a total of 54 elements listed. Out of those 54, 47 (87%) were associated with DIS and 27 (50%) were associated with AUTH. This means that phishing emails using Distraction and Authority persuasion principles have higher success rates in terms of phishing susceptibility.
Alessandro Ecclesie Agazzi, 2020	MFA and Anti-phishing protocols	Anti-phishing e-mail tools to block Phishing content such as link, contained in mails, at server and client levels with 89.5% of all infected mails being detected by the tool

Table 3: Analysis of mitigation techniques against phishing attacks

Author, Year	Ml-DL technique proposed	RESULTS
Sikha Bagui et al., 2019	Naïve Bayes;Support Vector Machines (SVM); Decision Tree;Long Short Term Memory (LSTM); Convolutional Neural Networks (CNN);Word Embedding	Naïve Bayes- 93.27% SVM- 82.35% Decision Tree -97.50% LSTM -96.64% CNN -97.20% Word Embedding -98.89%
Yahao Zhang et al., 2022	Improved Bayesian Algorithm	MATLAB simulation results show that the consistency between the amount of data sent by e-mail and the amount received is good, the consistency rate reached 92.3%. the data security level is 95%, encryption proportion/data proportion ratio under Bayesian optimization are higher than those of unfiltered method which is up to 97.2%
S. Srigayathri et al., 2018	SWEET protocol; Naive Bayes algorithm,	The proposed approach is based on the naïve Bayes algorithm which leads to accurate identification of spam messages and is reliable.
Jen-Wei Huang et al., 2018	Artificial Neural Network; bi-PVDBOW	True positive rates obtained using under-sampling: security level III email. Method Security level I Security level II Security level III Uni-BoW/100D 0.7916 0.83565 0.85838 Uni-LDA/100 topics 0.07832 0.61035 0.64247 Uni-PVDM/100D 0.51468 0.71379 0.71136 Uni-PVDBOW/100D 0.91748 0.92579 0.91263 Bi-BoW/100D 0.14405 0.80588 0.44381 Bi-LDA/100 topics 0.20419 0.74787 0.76437 Bi-PVDM/100D 0.8 0.85822 0.85963 Bi-PVDBOW/100D 0.97482 0.97846 0.97741

Table 4: Analysis of proposed ML-DL models

3.3.2 Analysis

Through the above comparison tables it is established that Machine Learning and Deep learning models are the best alternatives to existing email security protocols in terms of detection and mitigation techniques. Table 2 shows comparison between different cryptographic methods that were used in survey papers. The analysis establishes that models incorporating Identity-Based Encryption and hybrid features provided better results on terms of encryption and security implementations.

Table 3 describes upon the phishing mails mitigation techniques that were discussed in the research papers. MFA and anti-phishing tools provide the required amount of detection of malicious links in the email content.

Table 4 elaborates on different machine learning, deep learning and artificial neural networks that have been introduced by researchers to successfully automate email security procedures. Results provided by the improved Bayesian Optimization Algorithm in terms of TPR, TNR and safety rates prove that the proposed method is optimum for filtering through malicious mails for devices and networks connected under hoop topology.

Figure 5 shows the general architecture of the model proposed in this paper.

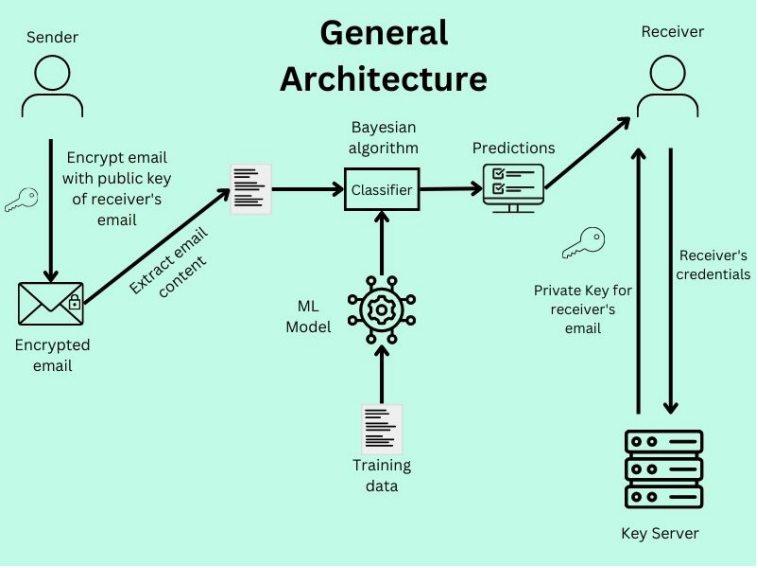

Figure 5: General Architecture of proposed model

The terms used in the general architecture are defined as follows:

- Cryptographic Techniques: In case of an adversary, cryptographic procedures are used to ensure data secrecy and integrity. Various cryptographic

approaches, such as symmetric key cryptography or public key cryptography, can be utilised during data transportation and storage depending on the security requirements and dangers involved[1].

- Encryption: Encryption is a method of encrypting data such that only authorised parties can decipher it[35]. It is the process of transforming human-readable plaintext to incomprehensible text, also known as ciphertext, in technical terms[21]. To put it simply, encryption modifies readable data so that it appears arbitrary.[21]

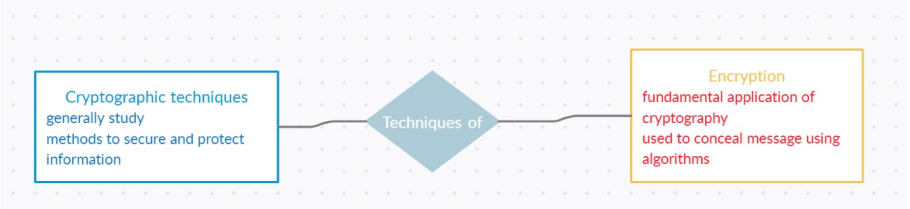

Figure 6: ER Diagram between cryptographic techniques and encryption

- SPF: The Sender Policy Framework (SPF) is an email authentication technology that is used to prevent phishing attempts.[9]
- DMARC: DMARC is an email authentication standard that assists mail administrators in preventing hackers and other attackers from faking their company and domain.
- PKD: Key distribution for public keys is accomplished by public key servers. When someone generates a key-pair, they keep one key private and upload the other, known as the public-key, to a server where anyone may use it to send the user a private, encrypted message.[33]
- Identity-Based Encryption: ID-based encryption, also known as identity-based encryption (IBE), is a vital primitive in ID-based cryptography. As such, it is a type of public-key encryption in which a user's public key is some unique information about the user's identity (for example, an email address).[2]
- MFA: MFA is an authentication mechanism that requires the user to give two or more verification factors in order to get access to a resource such as an application, an online account, or a VPN.[6]
- Anti-phishing protocols: Anti-phishing refers to the security measures that individuals and organisations can implement to either prevent or lessen the consequences of a successful phishing assault. Certain anti-phishing protection may prevent phishing emails from accessing a company's email system at all.

- SVM: Support Vector Machine, or SVM, is a prominent Supervised Learning technique that is used for both classification and regression issues. However, it is mostly utilised in Machine Learning for Classification difficulties. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorising n-dimensional space so that we may easily place fresh data points in the correct category in the future.[24] [39] [30]

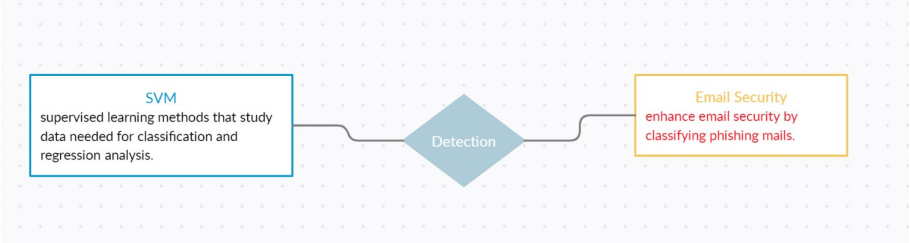

Figure 7: ER Diagram between SVM classifiers and email security

- Bayesian algorithm: We can encode our prior assumptions about what those models should look like using Bayesian machine learning, regardless of what the evidence tells us. This is very important when we don't have a large enough set of data to confidently train our model.
- ANN: An artificial neural network (ANN) is a computer model made up of many processing components that receive inputs and output results depending on predetermined activation functions.
- TPR: Proportion of correct guesses in positive class predictions. In simple terms, refers to actual and anticipated positive values.
- TNR: Values that are actually negative and predicted to negative or the ratio of true negative and total negative

4 Future Directions

4.0.1 Solution

The proposed hybrid model is a compilation of Identity-Based Encryption cryptographic method, Multi-Factor Authentication techniques and Improved Bayesian Optimization algorithm. This hybrid model is capable of successfully encrypting email content and protecting the user's sensitive information by using Bayesian algorithm to filter out spam messages and MFA and anti-phishing tools to detect phishing emails received by the user. Figure 5 shows the visualisation of the architecture of proposed model.

4.0.2 Future work

The proposed method is a theoretical model the implementation of which shall revolutionize email security. A practical approach to implement improved Bayesian optimization algorithm is required and further research upon this topic will provide with a viable solution to existing security problems in email protocols.

Unfortunately, email is inadequate for today's risks because it was built nearly 40 years ago, when its eventual global reach and security difficulties were imagined. Despite decades of work by the email industry to reduce spam, phishing and email-based malware remain significant risks, with email being implicated in more than 90% of all cyberattacks, according to various estimates. Email vulnerabilities have even disrupted elections, such as the 2016 hack of the Democratic National Committee's email (through spear phishing email) and the 2018 attacks on Florida election authorities.[23]

5 Conclusion

References

[1] Cryptocurrency: The death of paper money QSI Kosovo Student News — qsikosovonews.com. https://qsikosovonews.com/?p=219. [Accessed 04-Oct-2022].

[2] Identity-based encryption - Wikipedia — en.wikipedia.org. https://en.wikipedia.org/wiki/Identity-based_encryption. [Accessed 04-Oct-2022].

[3] Internet and email security issues — nibusinessinfo.co.uk — nibusinessinfo.co.uk. https://www.nibusinessinfo.co.uk/content/internet-and-email-security-issues. [Accessed 30-Sep-2022].

[4] Ldapwiki: POP — web.archive.org. https://web.archive.org/web/20211026221323/https://ldapwiki.com/wiki/POP. [Accessed 04-Oct-2022].

[5] Leading cause of ransomware infection 2020 — Statista — statista.com. https://www.statista.com/statistics/700965/leading-cause-of-ransomware-infection/#statisticContainer. [Accessed 30-Sep-2022].

[6] MFA in salesforce — sathishsfdc.com. https://sathishsfdc.com/2021/12/18/mfa-in-salesforce. [Accessed 04-Oct-2022].

[7] Post Office Protocol - Wikipedia — en.wikipedia.org. https://en.wikipedia.org/wiki/Post_Office_Protocol. [Accessed 04-Oct-2022].

[8] Setting Up Two-Factor Authentication — support.verivest.com. https://support.verivest.com/en/articles/5849790-setting-up-two-factor-authentication. [Accessed 04-Oct-2022].

[9] SPF,DKIM DMARC for Message Hygiene Services — informationprotection.ie. https://informationprotection.ie/2019/12/30/spfdkim-dmarc-for-message-hygiene-services. [Accessed 04-Oct-2022].

[10] What are the most important email security protocols? — techtarget.com. https://www.techtarget.com/searchsecurity/answer/What-are-the-most-important-email-security-protocols. [Accessed 04-Oct-2022].

[11] What Is Email Security? — heimdalsecurity.com. https://heimdalsecurity.com/blog/email-security/. [Accessed 30-Sep-2022].

[12] What is Secure/Multipurpose Internet Mail Extensions (S/MIME)? — tutorialspoint.com. https://www.tutorialspoint.com/what-is-secure-multipurpose-internet-mail-extensions-s-mime. [Accessed 04-Oct-2022].

[13] What is the SMTP protocol? - Tutorial — takeuforward.org. https://takeuforward.org/computer-network/what-is-the-smtp-protocol/. [Accessed 04-Oct-2022].

[14] E. B. Abdelsatir and M. H. Alrashdan. On the implementation of a secure email system with id-based encryption. In *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*, pages 1–4. IEEE, 2020.

[15] S. Al-Fedaghi and H. Alnasser. Modeling network security: Case study of email system. *arXiv preprint arXiv:2003.13509*, 2020.

[16] S. Bagui, D. Nandi, S. Bagui, and R. J. White. Classifying phishing email using machine learning and deep learning. In *2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pages 1–2. IEEE, 2019.

[17] S. Chawathe. Improving email security with fuzzy rules. In *2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*, pages 1864–1869. IEEE, 2018.

[18] J. Chen, V. Paxson, and J. Jiang. Composition kills: A case study of email sender authentication. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 2183–2199, 2020.

[19] G. S. Chhabra and D. S. Bajwa. Review of e-mail system, security protocols and email forensics. *International Journal of Computer Science & Communication Networks*, 5(3):201–211, 2019.

[20] A. Ecclesie Agazzi. Phishing and spear phishing: examples in cyber espionage and techniques to protect against them. *arXiv e-prints*, pages arXiv–2006, 2020.

[21] Y. Gandhi. Difference Between Encryption and Decryption — Analytics Steps — analyticssteps.com. `https://www.analyticssteps.com/blogs/difference-between-encryption-and-decryption`. [Accessed 04-Oct-2022].

[22] F. Holzbauer, J. Ullrich, M. Lindorfer, and T. Fiebig. Not that simple: Email delivery in the 21st century. In *2022 USENIX Annual Technical Conference (USENIX ATC 22)*, pages 295–308, 2022.

[23] https://www.darkreading.com/author/seth blank. New Standards Set to Reshape Future of Email Security — darkreading.com. `https://www.darkreading.com/application-security/new-standards-set-to-reshape-future-of-email-security`. [Accessed 04-Oct-2022].

[24] https://www.facebook.com/theaisummary/. Learn Mobile Price Prediction Through Four Classification Algorithms - AI Summary — ai-summary.com. `https://www.ai-summary.com/summary-learn-mobile-price-prediction-through-four-classification-algorithms`. [Accessed 04-Oct-2022].

[25] H. Hu and G. Wang. Revisiting email spoofing attacks. *arXiv preprint arXiv:1801.00853*, 2018.

[26] J.-W. Huang, C.-W. Chiang, and J.-W. Chang. Email security level classification of imbalanced data using artificial neural network: the real case in a world-leading enterprise. *Engineering Applications of Artificial Intelligence*, 75:11–21, 2018.

[27] V. Kanade. What Is Email Security? Definition, Benefits, Examples and Best Practices — Spiceworks — spiceworks.com. `https://www.spiceworks.com/it-security/network-security/articles/what-is-email-security/#_002`. [Accessed 30-Sep-2022].

[28] A. Kumar. What is SMTP and Working of the SMTP - Webkul Blog — webkul.com. `https://webkul.com/blog/what-is-smtp-and-working-of-the-smtp`. [Accessed 04-Oct-2022].

[29] T. Li, A. Mehta, and P. Yang. Security analysis of email systems. In *2018 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, pages 91–96. IEEE, 2018.

[30] P. Majumder. Learn Mobile Price Prediction Through Four Classification Algorithms — analyticsvidhya.com. https://www.analyticsvidhya.com/blog/2022/02/learn-mobile-price-prediction-through-four-classification-algorithms. [Accessed 04-Oct-2022].

[31] A. Malatras, I. Coisel, and I. Sanchez. Technical recommendations for improving security of email communications. In *2018 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1381–1386. IEEE, 2018.

[32] P. Menegay, J. Salyers, and G. College. Secure communications using blockchain technology. In *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, pages 599–604. IEEE, 2018.

[33] B. Noah. Often asked: How public keys are distributed? - October 2022 Vintage Kitchen — vintage-kitchen.com. https://vintage-kitchen.com/faq/often-asked-how-public-keys-are-distributed. [Accessed 04-Oct-2022].

[34] S. Ruoti, J. Andersen, T. Monson, D. Zappala, and K. Seamons. A comparative usability study of key management in secure email. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 375–394, 2018.

[35] U. C. Security. DOES USING THE CLOUD MEAN THAT I DON'T NEED CYBERSECURITY — UK Cyber Security Ltd — ukcybersecurity.co.uk. https://www.ukcybersecurity.co.uk/blog/news-advice/does-using-the-cloud-mean-that-i-dont-need-cybersecurity. [Accessed 04-Oct-2022].

[36] M. A. Shafi'i, S. Maryam, O. Oluwafemi, I. Ismaila, and K. A. John. Comparative analysis of classification algorithms for email spam detection. 2018.

[37] S. Srigayathri, V. U. Lakshmi, G. Rathiya, and G. Sangeetha. Enhancing secure email communication using tracking and blocking of suspicious email user.

[38] A. Sumner and X. Yuan. Mitigating phishing attacks: an overview. In *Proceedings of the 2019 ACM Southeast Conference*, pages 72–77, 2019.

[39] tutorialforbeginner.com. Support Vector Machine Algorithm — tutorialforbeginner.com — web.archive.org. http://web.archive.org/web/20211208103554/https://tutorialforbeginner.com/support-vector-machine-algorithm. [Accessed 04-Oct-2022].

[40] R. Umar, I. Riadi, and B. F. Muthohirin. Live forensics of tools on android devices for email forensics. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 17(4):1803–1809, 2019.

[41] Z. Yang, C. Qiao, W. Kan, and J. Qiu. Phishing email detection based on hybrid features. In *IOP Conference Series: Earth and Environmental Science*, volume 252, page 042051. IOP Publishing, 2019.

[42] Y. Zhang, J. Pang, and H. Yin. The optimization analysis of phishing email filtering in network fraud based on improved bayesian algorithm. *International Journal of Circuits, Systems and Signal Processing*, 16:504–508, 2022.