Chapter – 1
Introduction

1.1 Introduction to Electronic Mail

Email, or electronic mail, is a method of communicating messages between computers via the Internet. It is most commonly used in document interaction, technical communication, business, and education. It allows you to communicate with individuals anywhere in the world without bothering them [1].

Email has become one of the most significant and popular methods of communication for information exchange over the internet, with over billion users currently using it. E-mail has been a public media for more than fifty years. An email address can be used to contact someone fast [2].

Email is a real-time communication tool that can reach people across large geographic distances with relevant information. Emails are kept on the organization's server and serve as a record of past discussions. When typing out a message, extreme caution is required. It is data that is electronically transferred over a network between two or more parties. There is a sender and a recipient (s). An email is a digital communication or piece of data that is sent electronically between computers. An active internet connection, an email service provider, a personal email address, and the email address of the recipient are required in order to send an email [1].

An email address identifies an electronic inbox. It is divided into two parts: local and domain. The local component normally comprises the user's username, but the domain name is determined by the email service provider [3]. The local component might include words, numerals, or periods, but the domain name is determined by the kind of service provider. For instance, in the email id "*prashantchauhan25@gmail.com*", the username is "*prashantchauhan25*", and the domain name of the email service provider is "*gmail.com*".

Email was designed without security and is still essentially insecure today, despite its widespread use. Even though several methods have been developed throughout time to address the continuous issue of email security, not all of the issues that remain today have been totally overcome. Malicious emails seriously endanger both individuals and businesses [4]. Attackers have been using phishing emails, malware attachments, and URLs that link to malware, exploitation, or other malicious content as a powerful means of infecting the victim's machine. Email sender spoofing is a hostile activity in which the original sender, who is really an attacker, is modified so that the email seems to be from the intended sender. To prevent the entry of attackers, organizations and service providers need to bolster their email security and recognize email risks efficiently and pro-actively [4].

The first email from ARPANET came in 1971. As of 2017, there were 6.3 billion email accounts used by 3.7 billion people, who transmitted over 269 billion messages every day [2]. Email spoofing is crucial in phishing attacks, in which the attacker impersonates someone the target knows or trusts. The attacker has a greater chance of swindling the victim by tricking their email address into being that of an established organization or trusted friend [5].

Since the beginning of 2000, there has been a continuous effort to establish, support and implement protocols for preventing spoofing. Protocols like SPF, DKIM, and DMARC have become Internet standards, enabling email recipients to authenticate the sender's identity [5]. The global deployment of SPF is stated as 53.8%, DKIM at 38.8%, and DMARC at 46.8% [6].

1.2 Simple Mail Transfer Protocol (SMTP)

Simple Mail Transfer Protocol (SMTP) is the Internet's standard for email transfer. SMTP is a protocol for sending and receiving email over the Internet via an application layer. SMTP is created and maintained by the Internet Engineering Task Force (IETF) [7].

SMTP is a protocol for the application layer. The client that wants to send the email establishes a TCP connection with the SMTP server and then sends the email over that connection. The server has the SMTP always-on listening mode configured [8].

The SMTP process will initiate a connection to port 25 as soon as it hears the TCP connections from any client. The client process will send the mail immediately once you have successfully set up a TCP connection. SMTP is usually integrated into an email client application and is composed of four key components [8] [9]:

1. **Mail User Agent (MUA)**

   It is a computer program that assists in the transmission and retrieval of mail. It generates email messages for delivery to the mail transfer agent (MTA).
2. **Mail Submission Agent (MSA)**

   It is a computer software that accepts mail from a Mail User Agent (MUA) and transfers it through the Mail Transfer Agent (MTA).
3. **Mail Transfer Agent (MTA)**

   It is software that allows you to move email from one machine to another via SMTP.
4. **Mail Delivery Agent (MDA)**

   A mail Delivery agent or Local Delivery Agent is a system that helps deliver mail to the local system.

Some of the commonly used SMTP Commands are as follows:

1. **AUTH (Authentication)**

   Indicate an authentication mechanism for the SMTP server. Both PLAIN and LOGIN are supported.
2. **HELO (Hello)**

   Identifies the client to the server, fully qualified domain name, only sent once per session.
3. **EHLO (Extension Hello)**

   Enable SMTP extensions.
4. **MAIL (Mail)**

   Start an e-mail transaction to deliver the e-mail to one or more recipients.
5. **RCPT (Recipient)**

   It follows MAIL and identifies an addressee, typically the fully qualified name of the addressee, and for multiple addressees, use one RCPT for each addressee.

6. **DATA (Data)**

   Consider the lines following the command to be e-mailed from the sender.

7. **STARTTLS (Start Transport Layer Security)**

   Request that the SMTP server initiate a Secure Sockets Layer (SSL) or TLS negotiation with the SMTP client in order to create an SSL or TLS connection.

8. **SEND (Send)**

   Deliver emails to one or more workstations.

As shown in *Figure 1*, SMTP establishes a connection between the user and the server, whilst MTA and MDA offer domain searching and local delivery capabilities [7]. SMTP's fundamental weakness is that it lacks built-in security protections that prohibit anyone (attackers) from faking or spoofing an arbitrary sender address [5].

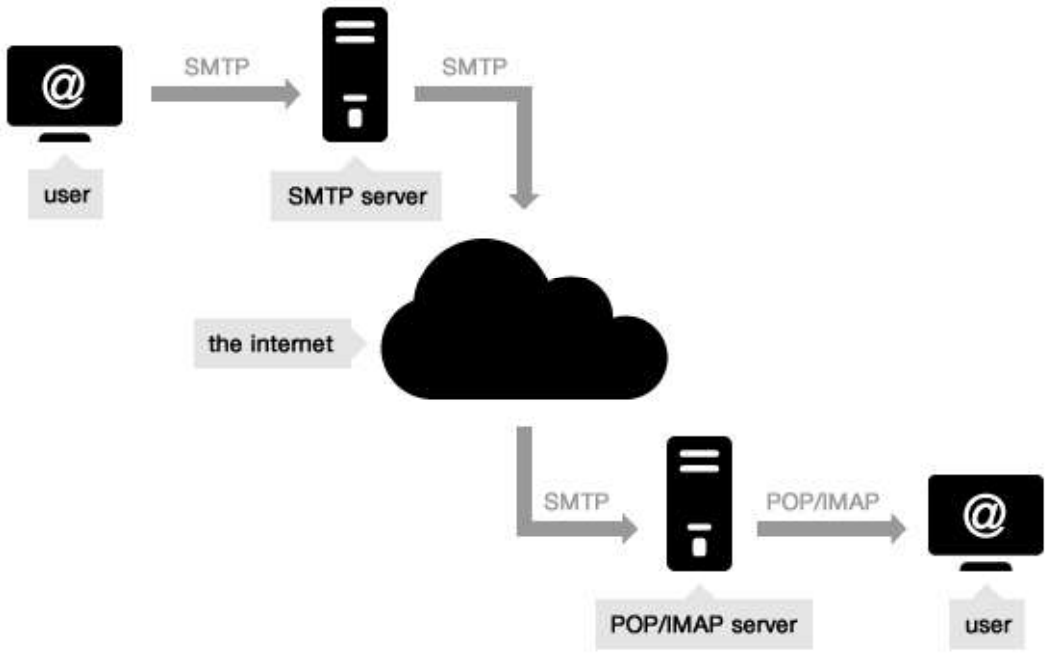

Figure 1: Email Delivery Process

SMTP allows any computer to send an email claiming to be from a certain source address. This is abused by spammers, who frequently use phony email addresses to make it harder to track the message back to its source, allowing them to disguise their identity and avoid accountability. Phishing, in which a user is fooled to disclose personal information after receiving an email purporting to come from the organisation such as a bank, also uses it.

1.3 Email Authentication

Malware frequently searches for email addresses within the machine they have infected and utilizes them both as targets for email and to construct believable faked "From" fields in the emails they send, making these emails more likely to be opened [10]. Email spoofing occurs when messages are created with a false sender address. Because the basic email protocols lack authentication methods, spam and phishing emails frequently utilize spoofing to deceive recipients about the origin of the message.

Attackers can manipulate two key fields to send emails to perform a spoofing attack. First, the attacker can use the "MailFrom" command and set the sender's address to anyone they want to impersonate after establishing a SMTP connection with the target mail server. After that, the “MailFrom” address is inserted into the header as the “Return-Path” [5] Furthermore, the email header can be modified by attackers by adding another field called “From”. The "From" parameter defines the address that appears on the email interface. When the user receives an email, they will see the "From" address. When a user answers to an email, the message is delivered to the "ReturnPath" specified in "MailFrom" [11]. It is to be noted that neither address is guaranteed to be identical [5].

1.4 Problem Statement

Email spoofing is continuously increasing with the increase in use of email. Attackers attempt to exploit the inadequate security of email authentication in SMTP by spoofing emails and attacking victims in some way. This research aims to check whether email spoofing is still possible and to find viable solutions to countermeasure this problem of email spoofing.

1.5 Motivation for the Research

Since the early times of my email usage, I have wondered why we receive so many spam emails daily. There was once a time when the number of spam emails received in my inbox was much greater than the number of genuine emails. The problem of spam in email was always in my mind. And when I got in a position where I could do some serious research, email was my priority.

Going into detail, much research is going on in the area of preventing spam emails. But to my surprise, more research work needed to be done in the area of email authentication to prevent email spoofing. Since my college days, I have been well aware of online tools that one can be easily used to send emails to anyone in the name of any desired email ID. However, with time at present, it is more challenging than it used to be in earlier days but spoofing of email is still possible by doing some tweaks. I always wanted to do my research in an area where I could bring some change to society through research work, and I confirmed that I would be doing my research for my PhD in the field of spoofing of email.

1.6 Objectives of Research

To brief about the main objectives of our research, the following points could be considered:

1. Systematic and detailed analysis of the complete email delivery process
2. Study and comparison of algorithms currently in use for the authentication of email
3. Exploration of new ideas to be applicable in the email delivery system to enhance the security in the present system
4. Development of an adaptive algorithm to countermeasure the everlasting problem of authentication of email
5. Modeling and implementing the suggested method, as well as analyzing the results obtained using the proposed technique
6. Finding the pros and cons of the proposed system and defining the relative future work

1.7 Research Contribution

1.7.1 Research Contribution to Society

Our research will help society by providing lost confidence in email security due to millions of spoofed emails received daily by millions of people. Upon successful implementation of our proposed system, the spoofing of emails will become negligible, resulting in better faith of people in the email and saving billions of dollars in addition to saving people from mental stress due to frauds by spoofers.

1.7.2 Research Contribution to Computer Science

In the field of computer science, our research will contribute in the following ways:

1. Testing the possibility of email spoofing on the top 10 most used email servers considering all possible configurations of the anti-spoofing protocols. Our experiment concluded that even after using anti-spoofing protocols by various email servers, it is still possible to spoof an email.
2. Finding various ways attackers or spoofers bypass the existing mechanism and successfully send a spoofed email. This helped us and will also help other future researchers find ways to countermeasure the problem of email spoofing. We determine under what conditions each email server will successfully transmit a faked email to the recipient's mailbox, designate it as spam, or reject it completely.
3. Proposed Adaptive Authentication Framework for Email (AAFE) for Prevention and Detection of Spoofing of Email. The proposed framework works in addition to the existing anti-spoofing protocols like DMARC, DKIM, & SPF and. Thus, it takes advantage of the current system along with an extra authentication layer of its own.
4. If the existing email servers, like Gmail, Outlook, etc., adopt our proposed framework, the number of spoofed emails is expected to become negligible. There is no proper measure of the amount lost by individuals and organisations due to spoofed emails, which are expected to be in the billions. Using our framework, the billions lost because of spoofed emails can be saved along with other non-monetary losses.

1.8 Scope of Problem Statement

The scope of our research is to find possible solutions to detect and prevent the spoofing of email, which will be carried out by authenticating the genuine sender. Also, in our research, we are limited to email authentication, skipping confidentiality, and other security issues. Assuming that the communication channel and the email system are already secure enough to prevent an attacker or spoofer from impersonating any communication in between, we research to find out possible ways to detect and prevent email spoofing by using some sort of authentication mechanism.

We created our own email server for all our experiments at our location. To check the spoofing of emails by various email service providers like Gmail, Yahoo, etc., we created accounts on these platforms and checked these servers from delivery of spoofed emails. After successfully implementing our proposed framework, we could test the delivery of spoofed emails using our own email server, as it is under our control. However, we could not try to deliver spoofed email on these popularly used email servers while using our proposed algorithm. All the tests we conducted were based only on our own email server as the sender of spoofed email.

1.9 Organization of the Thesis

The thesis is organized into six chapters, each with a brief introduction at the beginning and a summary at the end.

Chapter 1, as already discussed above, briefly introduces the email and clearly states the research's motivation, the study's objective, the research contributions and the scope of the problem statement of our research work.

Chapter 2 comprises the literature survey conducted before and during the actual research work, including the basics of email security, email spoofing and various existing protocols currently in use to countermeasure the problem of email spoofing. This chapter also briefly states the core research gaps in the current system that lay the path for our research work.

Chapter 3 covers the architectural setup for performing various experiments of our research work. It includes the steps taken to set up our own email server, the testing mechanism for email delivery using our own email server and possible ways to bypass various anti-spoofing protocols like DMARC, DKIM and SPF.

Chapter 4 describes our proposed system architecture and methodology to countermeasure the existing authentication problem in email security. It briefly explains the composition and step-by-step working of our proposed framework and how it provides enhanced security in addition to the current anti-spoofing protocols.

Chapter 5 contains a full performance study of our proposed system, including several test scenarios and their findings. This chapter also briefly compares the existing framework and our proposed framework, which clearly shows our proposed framework to be far better and almost flawless for preventing email spoofing.

Chapter 6 summarizes our entire study effort, including the constraints of our suggested system, the findings, and the future scope of work that we and other researchers may do in email authentication.

The References Section lists the references we have used to learn more about the problem of email spoofing and to understand the issues faced in the current situation while the existing anti-spoofing protocols are already in existence. These references also helped us find out what other researchers are doing to countermeasure the problem of email spoofing so we can focus on our proposed solution.

The Research articles Section covers the articles that resulted from this research study. A few papers are not listed here because they are still being reviewed and have not yet been published as of the time this thesis was written.

1.10 Conclusion and Summary

A brief introduction to email security has been discussed in this chapter, along with the basics of email security, email authentication, and the essential workings of the SMTP protocol, which is the core of the email system. This chapter also covers the motivation for the research, the research objective, research contributions, the scope of the problem statement of our research work and the brief organization of this thesis.