# Smart Email Security Assistant

**Cristian PASCARIU[1], Ioan BACIVAROV[2]**
[1] Faculty of Electronics, Telecommunications and Information Technology,
University POLITEHNICA of Bucharest, Romania
crpascariu@gmail.com
[2] EUROQUALROM, Faculty of Electronics, Telecommunications and Information Technology,
University POLITEHNICA of Bucharest, Romania
ioan.bacivarov@upb.ro

**Abstract**

*With security incidents and breaches growing each year, email is still used as the major entry point to server malicious content that results in credential theft or malware infections enabling malicious threat actors to mount complex attacks. This paper is intended to document a new approach for detecting suspicious and malicious emails leveraging techniques such as security analytics, natural language processing to discover the intent of the email, as well as artificial neural networks to support more complex rules for classification. This solution can be used in a basic mode to flag which emails are safe and which are not, at the same time it can also be used by security analysts to gain a better understanding of the attack vectors and speed up the investigation process.*

**Index terms:** artificial neural networks, email security, indicators of compromise, natural language processing, phishing

## 1. Introduction

Along with the digital revolution, more and more companies and institutions decide to store and manage their sensitive information in a digital format based on cloud or on-premise software solutions. In the private sector, the new business model leverages hardware and software IT solutions to manage their services offered to their clients. In the public sector, more and more institutions now offer online services as an alternative to people to reduce the amount of paperwork, people involved and time.

From an information security perspective, when the process involved a lot of physical paper documents, the security controls meant to protect the information were mainly around physical security. The documents were classified and then stored in vaults. Some of the major security risks involve either theft or disclosure by unauthorized personnel or natural disasters like floods that can destroy documents if inadequate measures are in place.

In the digital world sensitive information is stored in databases that are deployed on servers either based on-premise or in the cloud. Although this offers a lot of benefits in terms of availability and redundancy, this creates new avenues of attack, as anybody with access to the Internet can gain access to these systems. To address these types of risks, these databases are segregated at a network level.

In order to gain access to sensitive information and to bypass user access controls, malicious threat actors need to first compromise and steal credentials from an authorized user and use those credentials to access sensitive information.

Phishing emails are maliciously crafted messages send to many employees within a company or to individuals in an attempt to trick them to either download computer viruses disguised as legitimate computer programs or click on links to malicious websites hosted by the attackers.

These phishing websites are replicated and made to look like legitimate services and login pages. When the victim enters his or her valid credentials, these will be captured by the attacker for future use. At this point in time, the credentials of the victim are considered compromised, although the user might still be unaware that he has become a victim of phishing.

According to reports on security breaches, email phishing accounts for 96% of security breaches. This is a very high percentage as the malicious threat actors are targeting the human element as this is the most susceptible to such attacks. The motivation for this research paper was influenced by these high numbers [1]. Throughout this paper, apart from the proposed solution, existing solutions and techniques will be documented and analyzed based on their features and where they fall short.

## 2. State of the art

The email system, at its core, is a simple solution for sending and receiving digital messages. When it was invented, the current security risks were not an issue at that time. Email also relies on network protocols to ensure the identity of the sender and the recipient as well as secure the messages while they are transmitted over the Internet.

The Internet Message Access Protocol (IMAP) is an application layer protocol that enables users to retrieve messages from an email server. Most email client applications use IMAP to retrieve the messages and the Simple Mail Transfer Protocol (SMTP) so send the composed messages from the sender to the email server or relay.

Although these protocols ensure controls for authentication of the owner of the mailbox, the first versions of these protocols sent data in clear text over the Internet. From a security perspective this has a high impact on the confidentiality aspect of the data sent between the sender and recipients. With the rapid growth of end-to-end encryption, newer revisions of these protocols support the transfer of data over a secure communication channel.

Both SMTP and IMAP can be used with TLS or SSL which has been recently deprecated due to some high severity vulnerabilities. To make a distinction between the secure version of the protocol and the normal one the letter “s” is appended to the name. SMTPS and IMAPS are the more secure versions [2].

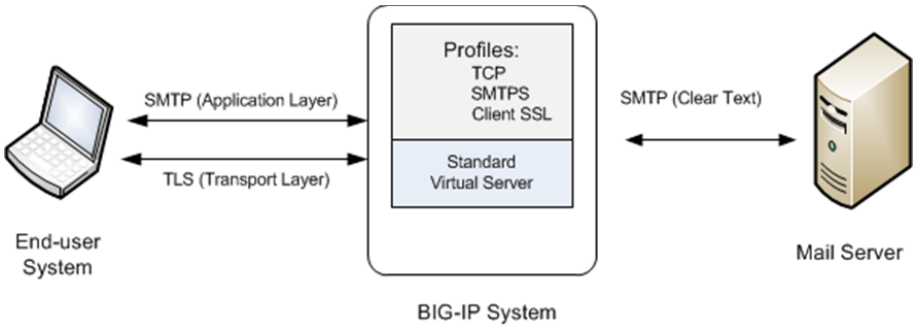


**Fig. 1.** Securing SMTP traffic [3]

Even though these protocols provide security controls to ensure authenticated access to the mailbox as well as the confidentiality, availability and integrity of the messages as they are sent between the sender and the receiver, these do not provide any control to prevent malicious threat actors from masquerading as a trusted sender from a legitimate company or public institution.

The Domain-based Message Authentication, Reporting and Conformance (DMARC) is a solution designed to combat malicious impersonation also known as spoofing. It is designed to combat techniques used by phishing emails that forge the sender address to match the ones of legitimate organizations. DMARC relies on two other protocols: the Sender Policy Framework known as SPF and DomainKeys Identified Mail known as DKIM. The end goal of a successful implementation is to ensure that there is a tight correlation between the email server that sends the email and the specific "From:" field in the email that is visible to the user [4].

Although SPF and DKIM can provide additional email security, it still relies on a secure configuration of the policies to prevent phishing and spam emails for reaching the employees of an organization.

All of these controls reduce significantly the amount of phishing and spam emails; however, they are not full proof as malicious threat actors can temporarily register domains, acquire expired ones, or leverage trusted mail services to bypass these controls.

The next set of email protections are targeted at the email itself rather than the entire infrastructure. The Naive Bayes spam filtering technique has been used as a baseline security control to prevent spam emails [5]. This is available as for the free email services as well being available to a wide variety of people. This solution is based on a statistical technique that classifies emails as being spam depending on common words that are used in other spam emails. This can be fine-tuned over time, if a message that is suspicious is not marked as spam by the mail service, it can be classified later as spam by the user and the system will gather specific words from the new mail message that are not present in the legitimate messages and if a new message is received with the new keywords, this will be marked as SPAM [6].

## 3. Proposed solution

The proposed solution (Figure 2) is not meant to be a replacement for the existing solution, but rather a set of complementary controls that can be used by basic users to reduce the amount of phishing emails and spam emails that they receive. At the same time this can also be used by security analysts to gain better insight into suspicious emails and increase the speed and the efficiency of their analysis process.

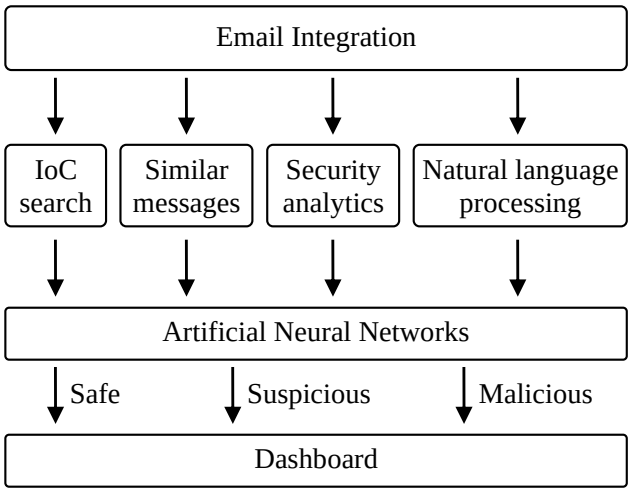


**Fig. 2.** Design diagram

### A. Email integration

The email integration module is an interface module, the scope of this is to facilitate the connection to one or multiple email accounts to gather data and provide it to the next modules for

future analysis. From a technical perspective, this module implements a simple mail client, the most common programming languages have dedicated modules for interfacing with a mail server and account.

The credentials for accessing an account has to be specified within the configuration files of the solution, this file will be restricted using granular access permission controls.

**B. IoC search**

Some malicious phishing mails deliver malicious payloads either as attachments or as malicious links for websites that host malicious payloads. The Indicator of Compromise (IoC) search module is meant as a front line of defense against such attacks.

This component takes advantage of two lists, a whitelist of safe and secure domains, URLs and IP addresses and a blacklist that is composed of malicious and suspicious domains, URLs, and IP addresses for detecting malicious network activity to and from the infrastructure controlled by the attacker. Any network indicator that is included within the email as well as the originating sender IP and domain are scanned against the whitelist and the blacklist. These lists of indicators are updated regularly from open-source intelligence services that make these available to the wider information security community.

The same whitelist and blacklist concepts are applied to scanning attachments. These can be extracted from the email for further analysis, with basic hash functions, the hash value can be calculated for the attachment, this hash value is later then validated against a list of known bad files included in the blacklist. There might also be false positives, and to reduce the numbers the whitelist is used for these files is not subject to further analysis.

A more efficient solution for scanning malicious files is the use of YARA, this is a pattern matching solution that can scan a file for specific signatures using a set of rules [7].

Open-source intelligence as well as public security breach reports made available by governmental institutions or security vendors can be used to keep an updated database on malicious indicators of compromise.

Another widely popular web security service within the security community, VirusTotal.com, provides integration libraries and modules for the most known programming languages such as python. With these APIs, the IoC Search module can validate directly all indicators found within a email message including both network based indicators as well as file attachments.

**C. Similar messages**

A large amount of phishing emails embedded addresses for phishing websites that are crafted to look like the legitimate login pages of trusted services such as online banking. From an analyst perspective, this can be trivial to uncover as the malicious website will be hosted on a different service using a different domain, sometimes one that is similar to the original one.

The goal of this module is to identify phishing mails that resemble the notifications received from the authentic services. Because the messages will be similar this is one of the query criteria, the second indicator is to check for links that point to a different domain rather than the authentic one.

Using a predefined list of words and expressions contained in the legitimate notification emails from the authentic web services, both suspicious and safe emails will be flagged, the second search criteria is to validate the originating address and domain as well as links within the message itself, to check if these are corresponding to the trusted domain, if this is false that means that the analyzed message is likely a phishing mail.

From an extensibility perspective, the list of words and expressions used can be increased with new works and expressions corresponding to other legitimate notifications such as online banking, government tax services or online shops.

**D. Security analytics**

One of the requirements of smart systems is to make use of an existing knowledge base when taking decisions. Based on this requirement, the Security Analytics module is meant to analyze the current mail message in relation to past messages based on common criteria such as the same sender, same or partly the same content.

Although this module might not provide significant value for a basic user, from a security analyst perspective this provides great insight whether the current message is part of a new campaign or an existing phishing campaign but with some minor changes.

Because of the low difficulty of writing phishing emails as opposed to the high difficulty of writing exploits, attackers often make small changes to their attack strategy starting with the phishing mail itself. Other times, in order to avoid blacklists, attackers will change their mail infrastructure but use the same message.

The goal of this module is to detect the reuse of indicators such as same sender address, same malicious links within the message or same attachments based on the previous messages received with the same criteria.

From a security analysis perspective, this module can also help reduce the amount of generic spam messages in order for the analyst to focus on higher priority tasks during the incident response process.

**E. Natural Language Processing**

With the rise in computing power, making computers understand normal human speech has become a requirement to simplify the interaction between people and computers using natural language rather than a development language.

This module is used to parse sentences from the message to discover the intent and scope of the message. In regard to generic phishing and spam messages, most of them contain the same message just phased differently, sometimes other words or sentences are used in order not to trigger the traditional spam filter.

Using detection techniques based on Natural Language Processing (NLP), this module is able to detect suspicious messages that are phrased differently but hide the same intent. Libraries such as the Natural Language Toolkit for the Python language implement methods to parse and tokenize sentences directly.

The goal of this module is not to duplicate existing functionality from the other module but complement it.

**F. Artificial Neural Network**

The analysis modules will provide very narrow insight into the suspicious messages, the artificial neural network is used as a classifier.

The main advantage of using a feed forward neural network is its ability to classify patterns with various degrees of truth.

Writing correlation rules for a very large amount of email messages can be a tedious process and prawn to human error and even if this is completed it might still miss out on some specifically crafter messages.

Taking advantage of the artificial neural networks, in the learning phase a large amount of maliciously crafted messages as well as legitimate emails can be fed to the neural networks to build all of these correlation rules automatically.

The inputs to the neural network are the outputs of the analysis modules. Some might flag some generic phishing emails, other modules might flag specifically crafter messages. The goal of the module is to classify these into three major categories: safe, suspicious, or malicious.

Extensibility is another feature of this module. If new analysis modules are added, using the correlation rules technique, this will require re-writing all rules to include the new analysis module. This limitation does not exist with the neural network as this can be retrained using the same training set.

### G. Dashboard

The dashboard is the main method of how the users of this solution will interact with it.

For the basic users the dashboard will display the results of the artificial neural network classification module.

For security analysts apart from the basic results, the results of the analysis modules will also be displayed, this is meant to give security analysts gather insight into malicious emails. This can also be used to fine tune the solution and its modules.

## 4. Conclusion

The aim of this paper is to leverage security analysis, natural language processing, and artificial neural networks to identify hidden threats in emails. The smart email security assistant can be used by security conscious computer users to add an extra layer of defense against phishing emails that attempt to deliver malware or steal credentials.

Security analysis can gain a lot more benefits by using this solution, as the techniques aggregated into this solution have been successfully applied individually within the incident response process with a great rate of success. The value added consists of having a standardized workflow for analyzing suspicious emails.

## References

[1]. "2018 Data Breach Investigations Report," Verizon. [Online]. Available: https://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_execsummary_en_xg.pdf

[2]. "SMTPS: Securing SMTP and the Differences Between SSL, TLS, and the Ports They Use," Agari. [Online]. Available: https://www.agari.com/blog/smtps-how-to-secure-smtp-with-ssl-tls-which-port-to-use

[3]. "Overview: Securing client-side SMTP traffic," F5. [Online]. Available: https://techdocs.f5.com/kb/en-us/products/big-ip_ltm/manuals/product/bigip-ssl-administration-13-1-0/12.html

[4]. M. Kucherawy, E. Zwicky (Eds.), "Domain-based Message Authentication, Reporting, and Conformance (DMARC)," IETF. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc7489

[5]. T. Lv, P. Yan, H. Yuan and W. He, "Spam Filter Based on Naive Bayesian Classifier," J. Phys.: Conf. Ser. 1575 012054. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1575/1/012054/pdf

[6]. T. Subramaniam, H.A. Jalab, and A.Y. Taqa, "Overview of textual anti-spam filtering techniques," International Journal of the Physical Sciences, Vol. 5(12), pp. 1869-1882, Oct. 4, 2010. [Online]. Available: https://www.cs.rug.nl/~tanguyen/pubs/article-Subramaniam.pdf

[7]. "Yara. The pattern matching swiss knife for malware researchers," Yara. [Online]. Available: https://virustotal.github.io/yara/