Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification

F Rahmad *, Y Suryanto, K Ramli
Information Network Security Management, Faculty of Engineering, University of Indonesia, Salemba, Jakarta, Indonesia

Email: *fauzi.rahmad81@ui.ac.id

Abstract. This research aims to test the ability to capture malicious e-mail objects, spam e-mails, phishing e-mails, or malware e-mails on Proofpoint Email Protection and Fortimail antispam devices to prevent security threats to PT. XYZ. Email service was one of the primary services for an agency or company in providing electronic mail that could be used to exchange official information. The study was conducted by comparing two antispam devices of Proofpoint and Fortimail. The test was conducted sequentially within three months period on each device. Those devices were configured using the infrastructure of PT. XYZ. Besides comparing the functions or features of each device, device testing was also done on overcoming the problem of email security threats such as spoofing, impostor, and bulk email that often threaten email infrastructure. To measure the accuracy of the results of the two devices, we use the Confusion Matrix method to measure the performance of the classification system. As the result, the slight differences inaccuracy were the indication that the ability of each antispam device was not too far. A large number of spam e-mails which caught through antispam devices were illustrated that PT. XYZ had become the target of spam, phishing, and malware attack. For this reason, anti-spam device technology was needed to maintain the quality of PT. XYZ email service.

1. Introduction
Email continues to be an essential part of daily business and consumer communication. Email is an integral part of internet users because all online activities, including social networks, online shop, instant messaging, and other account activities, require an email for the registration. The total number of business and consumer emails sent and received per day will exceed 293 billion in 2019 and is forecast to grow to over 347 billion by year-end 2023 [1]. Spam emails are sent massively, unauthorized, and without the consent of the receiver. Malware and phishing can be injected using those spam emails. This way, the victim can be flooded by random advertisements, explicit content, and fraud scheme [2].

In Q3 2019, the average percentage of spam in global mail traffic was 56.26% [3]. In the last decades, spam email growth exponentially increases. This condition is not only causing disruptions but also increasing security threats [2]. Therefore, we need a concept and technology to be able to prevent and anticipate security threats. There are several types of email security threats, such as spam email, spoofing email, impostor email, and bulk email. According to (Khandu), Spam e-mails are messages that are sent in large numbers to a random number of unknown addresses, meanwhile According to (Opazo, Whittle, & Shing), Email spoofing is a method of scamming individuals by impersonating a trusted correspondent via email. Impostor email is sent by the attacker using more sophisticated methods, the contents of the email look legitimate, but actually target important information such as
passwords and cryptographic keys [6]. In addition, email impostors also use more sophisticated strategies to bypass or compromise the protectors used [7]. Bulk e-mail is an e-mail that is usually sent from one source of an email address to many people or groups. The purpose of this attack is to send as many emails as possible to the recipient.

In order to maintain the information security of e-mail services from outside and inside attacks, a particular security device is needed to prevent security threats on e-mail services. One of the security devices that can be used to maintain email services security is an anti-spam device. The study was conducted by comparing two antispam devices, Proofpoint and Fortimail. The test was conducted sequentially within 3 months period on each device that is configured using the infrastructure of PT. XYZ. The tests carried out include the Test Function and Test Problem. Then in order to get accurate results from the two devices, the Confusion Matrix is used to measure the performance of a classification system. The confusion matrix is one of the methods that can be used to measure the performance of a classification method. The confusion matrix contains information that compares the results of the classification carried out by the system with the default results of the classification [8].

2. Method
The research Methodology in this study is:
- Comparing the functions or features of each device, device testing was also done on overcoming the problem of email security threats such as spoofing, impostor, and bulk email that often threaten email infrastructure
- Using Confusion Matrix to measure the accuracy of the results of the two devices.

Antispam devices could help protect and control inbound and outbound emails. By using the email protection feature, antispam devices could protect data and information security from attacks via email. Before determining the need for an appropriate antispam device, firstly, we had to identify the existing email traffic, which includes inbound and outbound traffic to be used for estimating the amount of antispam device capacity. The identification matrix was describes on table 1:

| Table 1. Identification Matric of Traffic Email |
|----------------------------------------------|
|                                            |
| **Peak Hourly Message Volume**              |
| Inbound | Outbound | Overall |
| 2000    | 2000     | 4000    |
| **Average Message Size (KB)**               |
| Inbound | Outbound | Overall |
| 215     | 215      | 430     |

Both incoming and outgoing emails in the PT. XYZ has a total number of 4000 emails with an average email size of 430 KB. We configured Proofpoint and Fortimail antispam devices on the PT. XYZ to become the main edge/mail gateway for all incoming e-mails with the logic flow topology and the e-mail system and antispam topology as figure 1.
The device is configured at the DMZ layer and is directly connected to the existing mailbox and has been protected by a firewall (See Figure 2).

Figure 1. Topology of Email Communication Flow.

Figure 2. Topology of Email System with Antispam Flow

Besides configuring the network, the configuration was also done on the DNS Server of PT. XYZ by changing the MX Record towards the IP of the antispam device. The antispam device was deployed using the transparent mode to reduce changes in the existing email infrastructure, as shown above. Logically, all emails would enter through the gateway, and then all incoming emails will be sent to the antispam device to be processed and identified.

A comparison between the two antispam devices was made by testing the functions of each device, as well as the ability to overcome problems related to spam, spoofing, impostor, and bulk email.

3. Results and Discussion

3.1. Function Test of Antispam Device: Proofpoint and Fortimail

Function tests were conducted by comparing the essential functions of anti-spam devices. The table 2 was the Table of Function Comparison.
Table 2. Function Comparison of Antispam Devices

| No | Function                                                                 | Fortimail | Proofpoint |
|----|--------------------------------------------------------------------------|-----------|------------|
| 1  | Mail Quarantine Management (search filtering, manual release)            | OK        | OK         |
| 2  | Integration with third-party spam URL and real-time blacklists (SURBL/RBL) | OK        | OK         |
| 3  | Global and local sender reputation                                       | OK        | OK         |
| 4  | Deep email header Inspection                                              | OK        | OK         |
| 5  | Realtime email activity tracking                                          | OK        | OK         |
| 6  | Logging Admin Activity                                                    | OK        | OK         |
| 7  | Logging Email History (Subject, Sender, Attachment)                       | OK        | OK         |
| 8  | Mail Queue Management                                                     | OK        | OK         |
| 9  | Behavioral/Content Analysis                                               | OK        | OK         |
| 10 | Threat Prevention                                                         | OK        | OK         |
| 11 | Dashboard Aktivitas Email                                                 | OK        | OK         |

Based on the table 2, both Proofpoint and Fortimail have all the essential functions of a functioning anti-spam device. These functions are obtained from the datasheet of each device.

3.2. Problem test spam email
The spam categories were mostly based on the spam email identification algorithm, email phishing, and malware email, which were the proprietary of each product owner's principal and the identification of free DNS Blacklist lists on the internet. The results of the e-mail spam problem test described on the Figure 3 and Figure 4.

![Figure 3. Result of Problem Test on Proofpoint Email Spam Device.](image-url)
Within three months periods of the anti-spam device testing, data reporting samples were taken within one month and obtained spam email identification on the Proofpoint anti-spam device by 17.56%. Whereas on the Fortimail antispam device, 29.8% of spam emails came to the XYZ.com domain.

![Statistics Summary (This Week)](image)

**Figure 4.** Result of Problem Test on Fortimail Email Spam Device

The difference in the test results percentage of the two devices was due to differences in testing time, so the results of the two could not be compared. However, from the test results of the problem, it could be stated that the two devices succeeded in identifying incoming e-mail spam, so it could be concluded that the e-mail spam test problem on both devices was successfully conducted.

3.3. Problem Test Spoofing Email

Email spoofing problem testing was conducted by activating the email spoof detection feature on both devices. The following were the results of the test problem (See Figure 5 and Figure 6).
In the figure 5, it could be seen that the spoofing feature on the Proofpoint antispam device has been active, but throughout the trial period, no emails were detected as spoofing.

**Figure 5. Problem Test Result of Email Spoofing Result on Proofpoint Device**

![Figure 5. Problem Test Result of Email Spoofing Result on Proofpoint Device](image)

In the figure 5, it could be seen that the spoofing feature on the Proofpoint antispam device has been active, but throughout the trial period, no emails were detected as spoofing.

**Figure 6. Problem Test Result of Email Spoofing on Fortimail Device.**

![Figure 6. Problem Test Result of Email Spoofing on Fortimail Device.](image)
In the problem test result of Email Spoofing on Fortimail Device on figure 6, the SPF Failure was 2.1%, and the DMARC Failure was 1.3%. Domain-based Message Authentication, Reporting & Conformance (DMARC) performs email authentication with Sender Policy Framework (SPF) and Domain Keys Identified Mail (DKIM) checking. SPF compares the client IP address to the IP address of the authorized senders in the DNS record. If the test fails, the email is treated as spam [9].

3.4. Problem Test of Impostor Email
Email impostor problem testing was conducted by activating the email impostor detection feature on both devices. The following were the results of the test problem (See Figure 7 and Figure 8).

Figure 7. Problem Test Result of Impostor Email on Proofpoint Device.

Figure 8. Problem Test Result of Impostor Email on Fortimail Device.
In the figure 7 and figure 8, it can be seen that both antispam devices had a feature to detect email impostors that had been activated. However, during the trial period, no emails were detected as impostors.

3.5. Problem Test of Bulk Email

Bulk Email problem testing was conducted by activating the bulk email detection feature on both devices. The following were the results of the test problem (See Figure 9 and Figure 10).

In the figure 9, it can be seen that the feature to detect bulk email on Proofpoint antispam devices have been active, but throughout the trial period, no emails were detected as Bulk Email.
There are 15.2% of emails with email traffic exceeding the session limit that has been configured, and this indicates there is an email that is identified as Bulk Email.

3.6. Analysis and Comparison Performance of Tested Antispam Device

For the two anti-spam devices tested, a comparison of spam identification results was performed to see the performance of each device. A comparison of spam identification performance is performed using statistical test score calculations in the 200 email test cases. The emails were then assessed for compatibility with the manual assessment by the administrator of PT. XYZ, who better understands the habits of emails in the xyz.com domain.

In the data mining concept, there is a method that can be used to measure the data accuracy so that the data can be used in decision support systems, called the Confusion Matrix. There are 4 (four) terms in the confusion matrix that describe the classification of performance measurement results, namely True Negative (TN), False Positive (FP), True Positive (TP), and False Negative (FN).

![Confusion Matrix](image)

**Figure 11. Confusion Matrix.**

True Negative (TN) value is the amount of harmful data that is detected correctly, while False Positive (FP) is harmful data but detected as positive data. Meanwhile, True Positive (TP) is positive data that is detected correctly. False Negative (FN) is the opposite of True Positive, so the data is positive but detected as harmful data [10].

Precision is data that is taken based on a lack of information. In binary classification, precision can be made equal to positive predictive values. The following formulation is the precision rule:

\[
\text{Precision} = \left( \frac{TP}{TP + FP} \right) \times 100\% \quad \text{Equation 1}
\]

The recall is deletion data that retrieved successfully from data relevant to the query. In binary classification, recall known as sensitivity. The establishment of relevant taken data that is approved by the query can be seen with recall. The following is the recall rule:

\[
\text{Recall} = \left( \frac{TP}{TP + FN} \right) \times 100\% \quad \text{Equation 2}
\]

Accuracy is a percentage of the total data which is identified and assessed. The following is the accuracy rule:

\[
\text{Accuracy} = \left( \frac{TP + TN}{TP + TN + FP + FN} \right) \times 100\% \quad \text{Equation 3}
\]

Each email was then categorized into four truth values. These conditions include:

- **True Positive (TP) Value.** True Positive value is the amount of positive data that is classified correctly by the system. In this report, the true positive value is the number of clean emails (administrator identification), which are also identified as clean emails by the antispam system.

- **True Negative (TN) Value.** True Negative value is the amount of harmful data that are classified correctly by the system. In this report, the true negative value is the number of spam emails (administrator identification), which are also identified as spam emails by the system.
- **False Positive (FP) Value.** False Positive value is the amount of positive data but is classified incorrectly by the system. In this report, true positive is the number of clean emails (administrator identification) but identified as spam emails by the system.
- **False Negative (FN) Value.** False Negative value is the amount of harmful data but is clarified incorrectly by the system. In this report, a true negative value is the number of spam emails (administrator identification) but identified as clean emails by the system.

The accuracy and precision values can be obtained from 4 (four) classification of measurement results as above, TN, FP, TP, and FN.

The accuracy value is obtained by comparing correctly classified data and the aggregate data. The accuracy value indicates how close the measurement results are to the actual value.

The equation can obtain the accuracy value:

\[
\text{Accuracy} = \left( \frac{TP+TN}{TP+TN+FP+FN} \right) \times 100\% \quad \text{Equation 4}
\]

Because in the case of measurement of antispam devices using different e-mails, the measurement of the precision value is deemed inappropriate. The dimensions of email objects that are continually changing make precision values irrelevant.

In other statistical terms, this accuracy is also known as Accuracy / Validity, which is usually used as a value of the ability to measure devices to measure what we want to be measured. From the test case identification results obtained a summary of the statistical set values as describes on table 3:

| Table 3. Comparison of Antispam Devices |
|-----------------------------------------|
|                                    | Fortimail | Proofpoint |
| TP                               | 126       | 144        |
| TN                               | 42        | 33         |
| FP                               | 32        | 14         |
| FN                               | 0         | 9          |
| **Accuracy**                      | 84%       | 89%        |

Based on the accuracy data in the analysis, the Antispam Proofpoint device has a higher accuracy value than Antispam Fortimail. Based on these values, it can be concluded that the Antispam Proofpoint device has little ability to capture spam emails better with a test case of 200 emails used as samples. Differences in accuracy values that are not too large can also indicate that the ability of each antispam is not too far.

4. Conclusion

To ward off security threats to the email service, PT. XYZ tested the installation of anti-spam devices, namely Proofpoint Email Protection and Fortimail, to test the ability to capture dangerous e-mail objects, whether spam e-mails, phishing e-mails, or malware e-mails. With a large number of spam emails caught through antispam devices, it is illustrated that PT. XYZ has become one of the objects of spam email attacks, email phishing, or email malware. For this reason, anti-spam device technology is needed to maintain the quality of PT. XYZ. Besides, to maintain the availability of excellent email services and avoid a single point of failure, antispam devices need to have a High Availability configuration in both DC and DRC PT. XYZ.

Acknowledgement

Author would like to thank for the support from Mr. Yohan Suryanto and Mr. Kalamullah Ramli as the lecture of Information Network Security Management, Faculty of Engineering, University of Indonesia.
who give an insights and advice for this research, my family and all parties that involved in this study either directly or indirectly.

References
[1] The Radicati Group, Inc. A Technology Market Research Firm. 2019. Email Statistics Report, 2019-2023 Executive Summary. The Radicati Group.
[2] Bhowmick, A., & Hazarika, S. M. 2018. Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends. ResearchGate.
[3] Kaspersky. 2019, November 30. Kaspersky. Retrieved from Kaspersky Lab Report, Spam and phishing in Q3 2019: https://securelist.com/spam-report-q3-2019/95177/
[4] Om, K. 2017. Secure Email Gateway. IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 49-53.
[5] Opazo, B., Whittleker, D., & Shing, C.-C. 2017. Email trouble: Secrets of spoofing, the dangers of social engineering, and how we can help. 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017), pp.2812-2817.
[6] Kartaltepe, E. J., & Xu, S. 2006. Towards Blocking Outgoing Malicious Impostor Emails. Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM’06).
[7] Kartaltepe, E. J., & Xu, S. 2005. On Automatically Detecting Malicious Impostor Emails. Proceedings of the 2005 International Workshop for Applied PKI (IWAP).
[8] Prasetyo, E. 2012. Data Mining: Konsep dan Aplikasi menggunakan Matlab. Yogyakarta: Andi Offset.
[9] Fortinet. 2016, May 30. The Fortinet Cookbook. Retrieved from The Fortinet Cookbook: https://cookbook.fortinet.com/fortimail-email-authentication-spf-dkim-dmarc/
[10] Rahmandani, M., Nugroho, H. A., & Setiawan, N. A. 2018. Cardiac Sound Classification Using Mel-Frequency Cepstral Coefficients (MFCC) and Artificial Neural Network (ANN). 3rd International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE).