International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

March 2023 | Issue 54 | Volume 15 | Number 1 | Pages 116-120

DETECT MALICIOUS EMAILS USING DART LANGUAGE

A.Z. Ablahd

Technical College of Kirkuk, Northern Technical University, Kirkuk, Iraq, drann@ntu.edu.iq

Abstract- The communication of emails, nowadays become an official and important. The email was a vector of conducting attack in cyber-crimes such as phishing, spoofing and any malicious email. Such attack used as a malicious transferring mode into the victim's device. The wide growing of malicious Email leads to online criminal activities. By such activity the confidential and secure information will stalled. To avoid such criminal activity, an On-line application (classifier) where prepared for inspecting each suspicious Email address. This classifier acts as interface in receiving Email address and early detects malicious, phishing and spoofing emails to avoid user from being victim in such attack. The proposed system built by using Dart-SDK language Windows supported with flutter-windows-3.0.1 platform. With this classifier the address of each email has been analyzed used one of machine learning algorithms called "Naive Bayesian algorithm" to distinguish between benign and malicious Email address. This system provides real time detection for any Email malicious crimes. Accuracy rate of this classifier is about 99.2%, 8000 Email address where tested, 4700 are malicious, 3300 are benign.

Keywords: Malicious, Email, Naive Bayesian, Machine Learning, Classifier.

1. INTRODUCTION

Due to the wide and rapidly growing of the internet, Email became a source of intruder and hackers [1]. The term malicious called to all things contribute in disturbing the victim computer systems for accessing to all personal and secure information. The malicious Email is a trick technique applied as fake Email in deceiving addresses seems as benign Emails in attempt for stealing personal specifics of receiver [2][10]. The malware Email attached with a small code that downloaded in a victim computer through opened it. This type of emails in last years can encrypt, steal bank account or passwords, and backup all the information of the victim computer and take full control of it, even if stored in server or his/her cloud [3][4].

In this proposed system it prepared a real time classifier, for detect any Email malicious crimes by testing all the suspicious Email addresses to distinguish between malicious and benign Emails. Dart-SDK language Windows supported with flutter-windows-3.0.1 platform used in preparing this classifier.

With this classifier the address of each email has been analyzed used one of machine learning algorithms called Naive Bayesian algorithm. Accuracy rate of this classifier is about 99.2%, 8000 Email address where tested, 4700 are malicious, 3300 are benign.

2. RELATED WORK

In this section an overview of some works in detecting malicious Email. There are a few research efforts focused on tracking malicious Emails. The researcher Zhan, et al. used a Stochastic Learning-Based Weak Estimators (SLWE) in filtering and detecting phishing Emails for a real environment life [1]. The proposed system of Zhan suffers by choosing enormous number of email features and unlimited training that reduced the performance of the system and consuming a huge storage. But Chandrasekaran system based on special Email structural features (used SVM algorithm) in detecting Email phishing and avoids the suspicious email from reaching users [2]. Lueg offered a survey for gaps explorer for retrieving and filtering the information are applied with hypothesis malicious Email detection. This survey did not present tools that used in simulation, the algorithm of machine learning and the environment architecture of email [3]. Wang used various techniques in detecting unsolicited Email. There was some limitation in this article that caused most of Email featured was not covered [18]. M.N. Marsono, et al. introduce a motor for examining and detecting unsolicited Email. This article orders up to 117 (million features) in every second that leads to slow down of its work [4]. Sahami, et al suggest of using some features for filters scrap Email and build a classifier using Bayesian. The result of this classifier is not accurate [5]. That is why it prepared a real dart language with flutter Email classifier to distinguish between benign or malicious Email address.

3. NAIVE BAYES ALGORITHM

This is one of a multi and binary machine learning classification algorithm used in proposed system. The classier built by using this algorithm [11]. Mostly is powerful in task of classification, because the decision in textual data analysis is precisely good [22]. In the fields of probability and statistics this algorithm is very important [12]. The Equation (1) is stated in Bayes algorithm.

$$p(A/B) = \frac{P(B \mid A)P(A)}{P(B)} \quad (1)$$

whereas, $P(B)\neq 0$, A, B: events, $P(A)$, $P(B)$: probability of event [19].

These events interpreted in Naive Bayes classifier as probability classes. The probability theory divided the event frequency instance by instance total number. $P(A|B)$ represent the conditional probabilities of events [13, 14]. The Naive classifier strategy says that the last probability for "A" event done when "B" event is true. This algorithm used in analyzed Email address to separate between malicious and benign Email address using Equation (1).

4. GENERALN CONCEPTS OF EMAIL

Email is mnemonic of (Electronic Mail) that means a communication way for sending and receiving messages by using Internet [15]. There are many protocols used by email within the suite of TCP/IP protocol [19]. The Email is very important way for communication because it is easy, cheap, quickly send and receive the messages in different distances for one person or at once to people group [16][20].

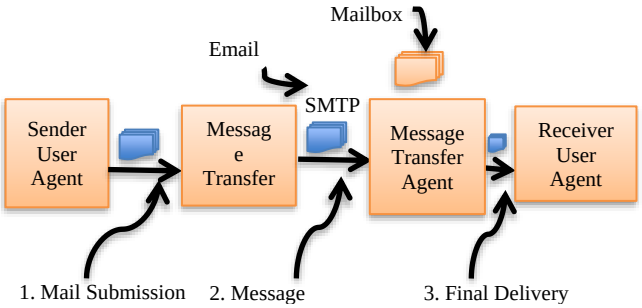

Figure 1. The architecture of email [1]

The architecture of the system of email see Figure 1. It consists of 2 sub-systems [17][21]. The first one is agent's user, that specify compose, sent, display replies to incoming message. For Example, of common user agents are Micro-Outlook, Google-Gmail, Apple-Mail, Mozilla Explorer [23]. The second sub system is message transfer agents, which called a server mail, used in sending messages with helping of (SMTP) protocol from source to destination [24].

5. EMAIL STRUCTURE COMPONENTS

There are different components of email like format of message, [18] Email addresses, agents and protocol. Figure 2 represents the structure of email.

5.1. Message Format

There is a standard format of the Email messages. Email consists of a header and body for each message. The message header contains: subject, attachment, From, CC, Bcc and to commands. Each command has its job as follows:

- From: used to determine the sender address.
- To: used to determine the receiver address.
- Subject: to include the subject of sender message.
- CC Bcc: CC mnemonic of "Carbon Copy" to declare second receiver.
- Attachment: represents the attached file like text, video, picture, etc.

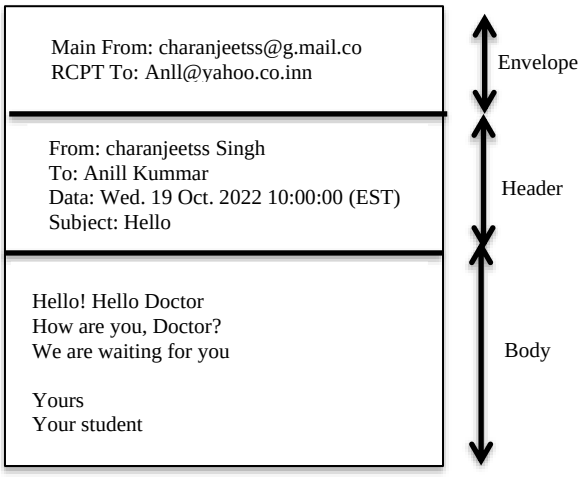

Figure 2. Basic structure of email [4]

Figure 3 represents analysis of email body. But the body of the message will be in text format that generated by email sender system automatically [20]. These contents are different according to varied systems of Emails.

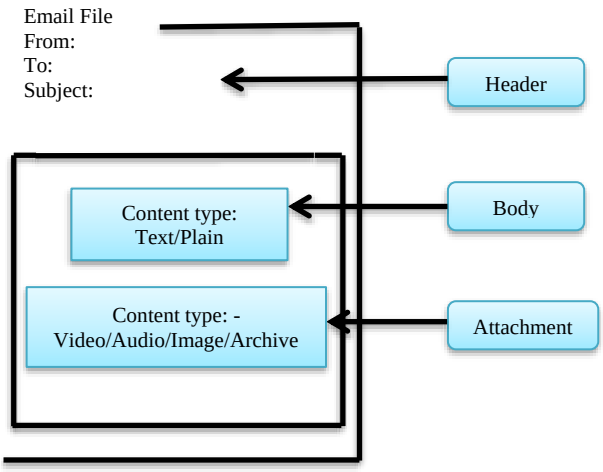

Figure 3. Email body analysis [5]

5.2. Email Address

To identify between different users, there is an Email address that use in sending and receiving messages [3]. The Email address contents 2 parts, local and domain parts (local-part@domain) this is the general form of email address [5]. The domain may be (a domain name) or (IP address in closed brackets), e.g. "Hario.Nash@gmail.com".

5.3. Email Protocols

There are different three protocols responsible for delivered the Email by Internet. The protocols are SMTP

(Simple Mail Transfer Protocol), POP (Post Office Protocol), IMAP (Internet Message Access Protocol [15]. All these protocols use TCP. The SMTP protocol lets the users on separate or same computers to interchange Emails. This protocol supported by the ability of sending one message to different recipients [15]. The messages may be delivered via Internet as text, video, graphic, audio, …, etc. The POP protocol is a simple that allows messages to be downloaded to computer from Inbox. But the IMAP is more advanced by allowing users to view all folders mail servers.

6. THE THREADS OF EMAIL

The cybercriminals every year was developed by introducing new different scenarios of tricks to be attached by email messages to attack the victims. But the aim of the malicious threats is same like stealing password account, losing data, disturbing the computer work, etc. The malicious threats consist of malware software like spyware, viruses, Trojan, bots, and type of attacks that infect PDF or any Office files.

The idea of launch different malware, through opened the attached Email messages. According to report in 2019 says that 94% of computer malwares delivered by email messages [3]. The malicious URL (Uniform Resource Locator) is a link can be clickable included within body or may be attacked by email messages [4]. Most of time, the malicious URL is disguised in text, image, and buttons. By clicking on this URL, a malware can be downloaded, installed and executed in computer victims.

7. DART LANGUAGE

The Dart language is an open source, objects-oriented programming, is an optimized client language used for developing or preparation a fast application in most platforms [22]. Developed by Google company. It aims is offering the more productive programming in multiple developed platforms paired with runtime execution [19]. Dart SDK were used in proposed system, that has command line tools and libraries like Dart - Web Develop, server applications, etc.

Other part of programming language used through preparation proposed system the open-source code that build by Google Company too is a Flutter windows 3.0.1 [12]. Flutter is a natively compiled framework (in a single codebase) used for building or developing multi beautiful platform (mobile, desktop, web) or embedded apps [16]. The flexibility of flutter appears through building an apps (same code) with Dart portability and Flutter framework to be work in Android, iOS, browser, and desktop. Flutter has the ability to support multi web application with rich interactive and graphic contents that is clear and attractive to different users.

8. THE PROPOSED CLASSIFIER SYSTEM

In this proposed system, built a classifier using Dart language with Flutter by depending in Naïve Base algorithm for detect malicious Email. In this classifier the set of features wear extracted as in Figure 4. There are 4 parts of email that been extracted to set of features like URL, Header, Script, and body of email.

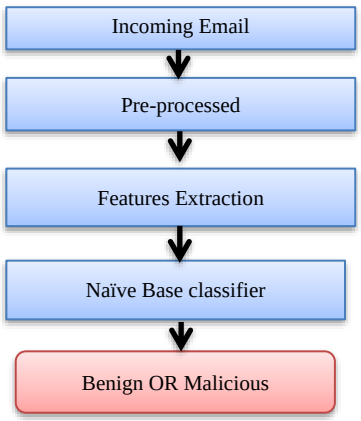

Figure 4. Represents structure of proposed system

The Pre-Processed step deals with preparing the raw data to be suitable for features extraction to build learning and training models as in Figure 5.

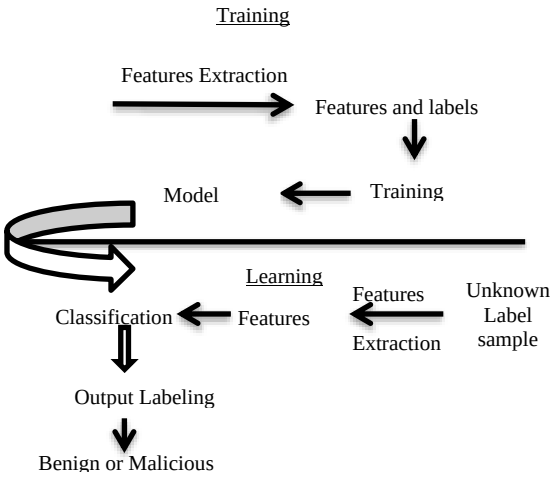

Figure 5. Malicious Email Detection

The Pre-Processed step increases the efficiency and accuracy of the proposed system in interpreted the features by proper data. The cleaning text used in this step by converting the data into a word list and removing the blanks, tab, punctuation, newlines and stop words Figure 6 represents stop words.

i, me , my , myself , we , our , ours , ourselves , you , you re, you 've, you'll, you'd, your , yours , yourself , yourselves , he , him , his , himself , she , she 's, , are , was , were , be , been , being , have , has , had , having , do , does , did , doing , a , an , the , and , but , if , or , because , as , until , while , of , at , by , for , with , about , against , between , , to , from , her , hers , herself , it , it' s, its , itself , they , them , their , theirs , themselves , what , which , who , whom , this , that ,etc,

Figure 6. Represents stop words

Through Pre-Processed step there is a tokenization that refers to split the large text or essays into small segments (like small document or lines or sentences) to facilities splitting the sentences into words which is called tokens. The algorithm of the proposed system as in Algorithm 1.

Algorithm 1. The proposed system

1: Email Address Selecting.
2: Extracting features by tokenization with word count algorithm.
3: Dataset Training by Naive Bayesian Classifier
4: Find the probability of malicious and benign Emails Address.
5: Dataset Testing.
6: Classify mailing and benign mails.
7: Calculate error rate and Email incorrectly categorized.
8: pretend the text data rate of errors, then compute the wrongly fraction in categorized word.

The features extraction in this proposed system, responsible for extracting all features of email to use by Naive Base algorithm for distinguishing between malicious or benign Email address. The most important features that be extracted are: URL containing of Hexadecimal and IP Address, the Dots Number in Domain Name, HTML Email, Links Number, appearance of Java Script, Number of Linked to Domain, From Body Match Domain Check, Malware Based Malicious Email. The combinations of all these features form a set of effectively classified Emails into malicious and benign.

The URL for different normal websites contains website names like (http://www.ethiotelcom.com) that means connect to ethiotelcom website. Most hackers used IP address or hexadecimal format in URL like:

- http://172.24.12.1/signin.ethiotlecom.com
- http://0xd4:0xeA:0x27:0x92/signin.ethiotlecom.com

The existence of IP or the hexadecimal in URLs are indicator of malicious Email. The dots number in domain name should be limited to three or less in benign address [5], if the Email address contains dots greater than 3 that are considered a malicious email. The format of email is defined in a standard form called "MIME" that defines the contents type attribute like plain text which is indicated as "text/plain". While HTML indicated as "text/html" [6]. The number of embedded links in email is important classification features. The malicious Email usually contains multiple links numbers.

The URLs of each Email are extracted to count the number of different domain names. The counted number used as a feature [7]. Each domain name is counted only once and any subsequent happened will be discarded. There is other feature like "From Body Match Domain Check" all the Email domain names are extracted to match the domain names with the domain of sender (domain name indicate to field "From" for the oneself Email). If there is a contrast between both of them that suggest this Email is a malicious [8].

Other features were extracted is "Malware Based Malicious Email", this feature involves installing different malware software within victim's device to gather private information of it. [9] In this status the malware directs the attached malicious link to great machinations in victim machine.

9. EXPERIMENTAL OF THE PROPOSED SYSTEM AND EVALUATION

This part represents the performance metrics for the optimal solution of the malicious Email detection. The accuracy and precision were evaluated. Through running the proposed system Email Tester, a platform will appear in waiting to input the Email for testing. After analysis of input Emails, it detects the Email as malicious or benign.

Evaluation Results
Accuracy |Precession | Recall | F1-measure

Malicious Email:
0.90625 | 0.985174 | 0.825 | 0.8979591836

Benign Email:
0.90625 |0.849572 | 0.9875 | 0.9132947976

Confusion Matrix (Malicious Email)
| Malicious | Benign |
Malicious | TP = 330 | FN = 70 |
Benign | FP = 5 | TN = 395 |

Confusion Matrix (Benign Class)
| Malicious | Benign |
Malicious | TP = 395 | FN = 5 |
Benign | FP = 70 | TN = 330 |

Figure 7. Result of using Malicious Email Detector

There are 4 situation was used to distinguish between (TP, FP, TN, FN) (True Positive, False Positive, True Negative, False Negative) respectively. The (TP, TN) refers to presence or correct prediction. Where:

- TP: Represents Email malicious no. correctly identified.
- FN: Represents the number of malicious detected as benign Emails.
- FP: Refers to benign emails detected as malicious Emails.
- TN: Represents the number of benign Emails detected as benign Emails.

While FP, FN refer to absence or incorrect prediction. Figure 7 represents TP-FP-TN-FN. The accuracy rate is calculated by using the Equation (2) [19].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

10. CONCLUSION

The wide growing of malicious Email leads to online criminal activities. By such activity the confidential and secure information will stalled. To avoid such criminal activity, an On-line application (classifier) where prepared for inspecting each suspicious Email address. This classifier acts as interface in receiving Email address and early detects malicious, phishing, and spoofing emails to avoid user from being victim in such attack.

In this proposed system it identified a various kinds of malicious Email detection. With this classifier the address of each email has been analyzed used one of machine learning algorithms called Naive Bayes algorithm to distinguish between benign and malicious. In detection (malicious links or contents) with accuracy 99.2%, 8000 Email address where tested, 4700 are malicious, 3300 are benign. The proposed system built by using Dart-SDK language Windows supported with flutter-windows-3.0.1 platform. This system provides real time detection for any Email malicious crimes.

REFERENCES

[1] J. Zhan, L. Thomas," A Survey of Learning Based Techniques of Phishing Email Filtering", International Journal of Digital Content Technology and its Applications (JDCTA), Issue 18, No. 14, Vol. 6, pp. 55-59, Malaysia, April 2021.

[2] M. Chandrasekaran, M.S. Padhyaya, "Phishing Email Detection Based on Structural Properties", NYS Cyber Security Conf., pp. 20-30, Barracuda, Thailand, 2006.

[3] Y. Haddi, A. Kharchaf, A. Moumen, "Study of a Mobile Robot Obstacle Avoidance Behavior in a Radioactive Environment with A High Level of Autonomy", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 50, Vol. 14, No. 1, pp. 34-41, March 2022.

[4] M. Atify, M. Bennani, A. Abouabdellah, "Structure Optimization of a Hexapod Robot", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 50, Vol. 14, No. 1, pp. 42-49, March 2022.

[5] A.S. Tanenbaum, D.J. Wetherall, "The Application Layer in Computer Networks", Pearson Education Journal, Vol. 5, pp. 90-95, USA, 2011.

[6] I. Fette, N. Sadeh, A. Tomasic, "Learning to Detect Phishing Emails", The 16th International Conference on World Wide Web, pp. 3-7, 2007.

[7] H. Bhuiyan, A. Ashiquzzaman, T.I. Juthi, S. Biswas, J. Ara, "A Survey of Existing E-Mail Spam Filtering Methods Considering", Global Journal of Computer Science and Technology: C, Software and Data Engineering, Vol. 18, Issue 2, pp. 20-23, USA, 2021.

[8] I. Vayansky, S. Kumar, "Phishing-Challenges and Solutions", Computer Fraud and Security Journal, Vol. 1, pp. 15-20, 2018.

[9] A. Almomani, B.B. Gupta, S. Atawneh, A. Meulenberg, E. Almomani, "A Survey of Phishing Email Filtering Techniques", IEEE Communications Surveys and Tutorials Jour., Vol. 15, No. 4, pp. 2070-2090, 2013.

[10] A.Z. Ablahd, "A New Cryptography Method Based on Hill and Rail Fence Algorithms", Diyala Journal of Engineering Sciences, Vol. 10, No. 01, pp. 39-47, Kirkuk, Iraq, March 2017.

[11] S. Smadi, N. Aslam, L. Zhang, R. Alasem, M.A. Hossain, "Detection of Phishing Emails Using Data Mining Algorithms", The 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA) IEEE, pp. 1-8, December 2015.

[12] A.Z. Ablahd, S.A. Dawood, "Using Flask for SQLIA Detection and Production", TJES Tikrit Joural Engineering and Science, Vol. 25, No. 2, pp. 1-14, Kirkuk, Iraq, 2020.

[13] A. Tizkar Sadabadi, "Composition of a Game Based Simulation for Software Development Process", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 2, Vol. 2, No. 1, pp. 73-78, March 2010.

[14] N.M. Tabatabaei, S.R. Mortezaeei, "Review of Multi-Agent Systems (MAS), a New Tool for the Control and Management of Modern Power Systems", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 1, Vol. 1, No. 1, pp. 27-31, December 2009.

[15] V. Shahrivari, M.M. Darabi, M. Izadi, "Phishing Detection Using Machine Learning Techniques", arXiv Preprint, v. 1, p. 11116, September 2020.

[16] Symantec, "Internet Security Threat Report (ISRT)", Cyber Criminals Target Payment CARD DATA Conference, p. 48, USA, 2019.

[17] N. Lord, "Social Engineering", Social Engineering Attacks Conference, pp. 3-6, December 2020.

[18] A.A. Akinyelu, O.A. Aderemi, "Classification of Phishing Email Using Random Forest Machine Learning Technique", Jour. of Applied Mathematics, pp. 5-8, 2022.

[19] A.Z. Ablahd, "Detect Malicious Web Pages by Using NAIVE Bayesian Classification Technique", International Journal of Applied Engineering Research Journal, Vol. 06, No. 02, pp. 254-258, 2021.

[20] P.M.V. Divya, U.R. Mouli, "Web Based Optical Character Recognition Application Using Flask and Tesseract", Elsevier, pp. 56-62, 2021.

[21] A. Zainab, C. Hewage, L. Nawaf, I. Khan, "Phishing Attacks", Frontiers Journal in Computer Science, Vol. 3, p. 6, 2021.

[22] S.N.P. Tandale, "Different Types of Phishing Attacks and Detection Techniques", International Conference on Smart Innovations in Design Environment Management Planning and Computing (ICSIDEMPC), pp. 295-299, 2020.

[23] J. Sande, M. Galloway, "Dart Apprentice", Raywenderlich Tutorial Team, pp. 50-70, 2022.

[24] S. Sinha, "Quick Start Guide to Dart Programming: Create High-Performance Applications for the Web and Mobile", Apress, pp. 40-80, 2022.

BIOGRAPHY

First Name: **Ann**
Middle Name: **Zeki**
Surname: **Ablahd**
Birthday: 04.04.1966
Birth Place: Mosul, Iraq
Bachelor: Computer Science Department, Mosul University, Mosul, Iraq, 1988
Master: Computer Science Department, Mosul University, Mosul, Iraq, 2001
Doctorate: Computer Science Department, Mosul University, Mosul, Iraq, 2013
The Last Scientific Position: Assist. Prof., Computer Engineering Department, Kirkuk Technical College, Northern Technical University, Mosul, Iraq, 2017
Research Interests: Cypher Security, Web Application
Scientific Publications: 11 Papers, 1 Book, 3 Theses
Scientific Memberships: Kirkuk Technical Institute Promotions Committee