Investigating Targeted Malicious Email

Prof. Shilpa S. Adke, Sayali K. Ahire, Navneet C. Battise,
Sunny A. Yadav & Swati S. Dhodare
Department of Information Technology Engineering, Matoshri College of Engineering and Research Centre,Eklahare, Nashik.

***Abstract**: Targeted email attacks to enable computer network exploitation have become more prevalent, more insidious, and more widely documented in recent years. Beyond nuisance spam or phishing designed to trick users into revealing personal information, targeted malicious email (TME) facilitates computer network exploitation and the gathering of sensitive information from targeted networks. These targeted email attacks are not singular unrelated events; instead they are coordinated and persistent attack campaigns that can span years. This dissertation surveys and categorizes existing email littering techniques, proposes and implements new methods for detecting targeted malicious email and compares these newly developed techniques to traditional detection methods. Current research and commercial methods for detecting illegiti-mate email are limited to addressing Internet scale email abuse, such as spam, but not focused on addressing targeted malicious emails. The specific tools, techniques, procedures, and infrastructure that a threat actor uses characterize the level and capability of a threat; the recipients role and repeated targeting speak to the intent of the threat. Both sets of features are used in a random forest classifier to separate targeted malicious email from non-targeted malicious email, targeted malicious email that does not. Second, targeted malicious email demonstrates association to recipient oriented features as compared to non-targeted malicious email that does not.*

***Keywords**— Targeted Malicious Email, Non Targeted Malicious Email, Ran-dom Forest Classi er, Filtering.*

1. Introduction

Now a day's emails are used for the purpose of communicating, sharing and distributing information. But sometimes various malicious emails can harm a system as well as network. Especially it targets to an organization. There are various techniques available in the market but that are not active. They are only useful for spam detection. The proposed system is useful for tracking and detecting the malicious emails. Sometimes through email malicious actors try to get the sensitive information which can harm in various ways. By using proposed system classification of Targeted malicious emails and Non-Targeted emails are done.

Antivirus programs focus only on the binary code of an email but, ignoring all relevant contextual meta data. Sometimes through emails, malicious actors try to get the sensitive information which can harm in various ways. The proposed system has an ability to differentiate Targeted malicious emails and Non-Targeted emails. This categorization allows user to decide whether to accept or reject an email coming into their network. Targeted mails have become a huge security problem on the Internet. As with many computer security problems, the attacks require active involvement of a human being. Even knowledgeable, security conscious people can fall into a phishing trap. Targeted email is a special type of spam message. Such email is a criminal mechanism that relies on forged email claims purportedly originating from a legiti-mate company or bank. Subsequently, through an embedded link within the email, the phisher attempts to redirect users to fake Websites that are designed to fraudulently obtain nancial data such as user names, passwords, and credit card numbers.

The many approaches proposed in the literature to lter phishing emails, may be classi ed according to the diff- erent stages of the attack ow, e.g. network level protec-tion, authentication, client side tool, user education, server side classifiers, etc. We discuss the advantages and limitation of these approaches. This survey gives an organized guide to the present state of the literature, in view of the wide scope of approaches. In the literature, the evaluation and comparison of different approaches on phishing email littering are given a great deal of attention. This survey not only identities and categorizes these methods, but also compares and analyses their relative merits. For example, it lists strengths, weaknesses, and the related application scenarios for guiding the readers to design new anti phishing detection methods in the future[1].

2. Literature Survey

To understand email littering techniques, a working knowledge of email structure and format is required. Users typically never see the envelope because email systems throw away the envelope just before delivering the letter (e.g. the email message) to the user. Notable features of the email are as follows:

1. Delivered-To - The email address the message will be delivered to.

2. Received - Every email server that handles the message will add a Received line entry which includes a time stamp.

3. Return-Path - The email address from which the message was sent.

4. Received-SPF - Sender Policy Framework (SPF) domain authentication results.

5. Message-ID - A unique number assigned by the sending mail server.

6. Content-Type - Defines the boundary string used to separate the Multipurpose Internet Mail Extensions (MIME) parts of an email.

7. X-Originating-IP - The Internet Protocol (IP) address of the sending client.

8. From - Set by the senders email program. From consists of a phrase and address (the phrase is the string before the email address). This does not have to equal the MAIL From line in the email envelope.

9. To - Set by the sender. To consists of a phrase and address. This does not necessarily have to equal the RCPT To line in the email envelope. If there are any Cc recipients they would appear in the RCPT To line in the email envelope. Any Bcc recipients would not be shown but would be in the email envelope a RCPT To recipients.

10.Subject - Set by the sender.

11. Date - Set by senders email program. It includes the lo cal time zone of the system used to send the email.

12. Content-Type - Defines the character set used by the email.

13.Content-Disposition - Includes some information about the attachment[1].

2.1 Existing Weaknesses

Existing techniques for littering email have limitations when applied to targeted malicious email. Authentication based techniques require receivers to enforce domain level authentication upon email receipt. Since these techniques are not fully adopted across the Internet, enforcing the authentication at all times is not possible. To com-plicate matters, the authentication is at the domain level, not on a per email address basis. Thus, a public webmail provider like Google may be authenticated but a threat actor may have created an email account for the purposes of sending targeted malicious email. Contextual approaches typically focus on message content, making classification decisions largely on the words in the body of an email. From a threat actors perspective, message content is the easiest to change and thus is not very durable across multiple email campaigns from the same threat actor. Furthermore, since targeted malicious emails often have message content very specific to the recipient, ending common words across emails is not as relevant as it is with spam. Characterization based approaches to littering email usually involve quantifying aspects of email volume, low volume attacks such as targeted malicious emails are likely to remain undetected. With targeted malicious email, known email addresses and names are used which hampers the actively of reputation based approaches. Finally, resource consumption based techniques for littering email are largely focused on malicious actors who send large amounts of email, typically spammers.

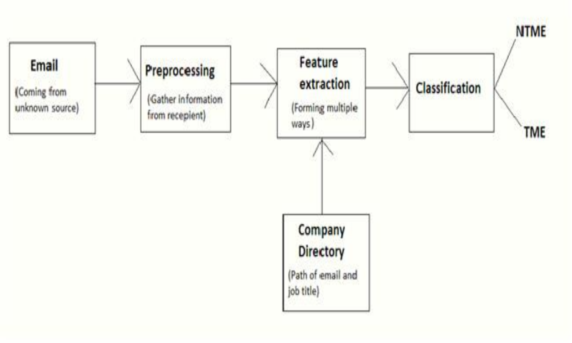
Targeted malicious emails are low volume and directed at certain recipients, which is in contrast to spam which is often directed at numerous recipients and is of high volume. Existing approaches to littering email are focused on specific attacks but do not leverage features that are more durable and possibly common across a set of attacks from a particular threat. Tools such as anti-virus typically intervene fairly late in the threat kill chain with little insight into steps such as reconnaissance. By focusing on steps in the kill chain that are more difficult for the threat actor to readily manipulate, greater detection capability for targeted malicious emails can be achieved.

3. Proposed System

In the previous chapters we covered the whole idea of our project, research about different systems currently available, their problems, solutions to those problems in our project and requirements of our project.

In this chapter, we are providing detailed information about the system design. In the following parts of the chapter, we have added various block diagrams and their description regarding our pro ject. We have also added various UML diagrams and their description for better understanding of the software.

The diagram shows the common architecture of Detection of Targeted Malicious Emails.



System Architecture

Classification process is explained below:

1.Email: An email with unknown classification.

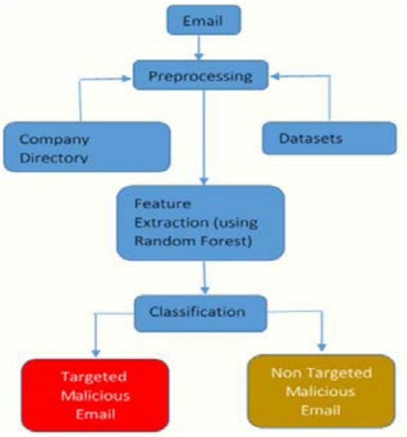
2.Directory: It includes information about email users such as job title. Preprocessing: Email and directory information are combined to provide additional recipient context to emails.

3.Feature Extraction: Relevant features are extracted from the email and converted into a multidimensional vector with each element of the vector representing a feature.

4.Classification: The email is processed through a classic that was trained with previously labeled data to determine the classification of the input email.

4. Overall Interaction of the System

Network providers are the one which allows all type of emails for communication purpose. While transferring the messages some malicious emails are received by the users this causes many problems either at the server side or at the user side. This type of emails may contain unsolicited content, or it could be due to the message being crafted. Persistent threat features, such as threat actor locale and unsolicited email crafting tools, along with recipient oriented features.



Overall Interaction of the System

Current detection techniques work well for spam and phishing because its easy to detect mass-generated email sent to millions of addresses. TME mainly targets single users or small groups in low volumes. TME can pretend network exploitation .Hence for detection of TME is vital work. This paper explains how the malicious emails are classified. In order to classify here we are using Random Forest Classier. This classier focuses on feature extraction[4].

5. Problem Statement

The malicious emails are target at company executives, government personnel and other individuals with access to sensitive information useful by an opposing party to advance a cause. Current research and commercial methods for detecting illegitimate email are limited to addressing Internet scale email abuse such as spam; none seek to address targeted malicious emails.

For organizations targeted by these emails, detection is critically important since these emails can enable the installation of malicious software on the targeted users computer system. This malicious software can contain a back door that allows a malicious threat actor to gain entrance to an organizations network and its sensitive information. Whereas conventional unwanted email, such as spam, is sent in bulk to a large number of people on the Internet, TME is sent to very specific individuals. The techniques that malicious threat actors use to craft and send these targeted emails are different from the techniques used by spammers. Furthermore, since the targeted emails are sent to specific individuals, the characteristics of the recipient are relevant whereas with spam, they are less relevant. This dissertation For organizations targeted by these emails, detection is critically important since these emails can enable the in-stallation of malicious software on the targeted users computer system. This malicious software can contain a back door that allows a malicious threat actor to gain entrance to an organizations network and its sensitive information. Whereas conventional un-wanted email, such as spam, is sent in bulk to a large number of people on the Internet, TME is sent to very specific individuals. The techniques that malicious threat actors use to craft and send these targeted emails are different from the techniques used by spammers. Furthermore, since the targeted emails are sent to specific individuals, the characteristics of the recipient are relevant whereas with spam, they are less relevant.

6. Assumptions and Dependencies

1.Targeted malicious email demonstrates association to persistent threat features of email such as locale and tools as compared to non-targeted malicious email that does not show an association to persistent threat features.

2.Targeted malicious email demonstrates association to recipient oriented features such as role, reputation, relationships and access as compared to non-targeted malicious email that does not show an association to recipient oriented features.

3.Detection of targeted malicious email using persistent threat and recipient oriented features results in fewer false negatives than detection of targeted malicious email using conventional email littering techniques.

6.1 Attribute Suggestion

In this domain, we study and describe solution for the 'attribute suggestion' problem. From the problem definition we can determine two properties for determine and suggesting attributes for a document:
1] First, the attributes must have high level querying value (QV) with respect to the query workload. That is, they must appear in large no. of queries in query workload, because the random attributes in workload have a large potential to improve the visibility of document.
2] Second, the attributes must have high level content value (CV) with respect to document textual data. That is, they must be relevant to textual data. Otherwise, the end user will probably dismiss the suggestions and document will not be properly annotated.

7. Conclusion

The purpose of this project is to develop classification methods, using persistent threat and recipient oriented features, designed to detect targeted malicious email (TME). The purpose work is an introduction to novice classification method using persistent threat Recipient oriented feature design to. Additionally, the study aimed to demonstrate that incorporating these features results in a detection capability that is superior to conventional email littering techniques.

In this study, established targeted malicious email (TME) as a separate class of email that was not previously researched in the academic literature. New detection methods were created based on persistent threat and recipient oriented features. Classification of TME and NTME can be done by using Random Forest Classifier.

Targeted Malicious Email (TME) presents a great risk for those organizations plagued by it. The impact of sensitive data loss can be severe not only to a company but also to a country. The techniques developed in this can be used to increase the ability of organizations to detect TME over conventional techniques.

8. Acknowledgement

We take this opportunity to thank our project guide Prof. Shilpa S. Adke for their guidance and providing all necessary facilities for this paper.

9. References

1. Rohan M. Amin, Julie J.C.H Ryan and J. Rene van Dorp, Detecting Targeted Mali-cious Emails, IEEE, Computer and reliability societies, George Washington University, May/June 2012.

2. V. S. Kumar, Ravi Kumar, An Effective Model of Detection and filtering tech-niques over malicious and spam email, International Journal of Engineering Trends and Technology (IJETT) Volume-5, Kakinada, NOV 2013.

3. Multi-state information sharing and Analysis centre and United States Computer Emergency readiness team, Current Malware Threads and mitigation Strategies, US-CERT, May 2005

4. HilarieOrman, Towards a Semantics of Phish, 3rd ed. Purple Streak, Woodland Hills UT, USA

5. Erhan J. Kartaltepe, ShouhuaiXu, On Automatically Detecting Malicious Impostor Emails, Department of Computer Science, University of Texas at San Antonio, IOS Press, 2003

6. Tuan PhanVuong, Diane Gan, A Targeted Malicious Email (TME) Attack tool, 3rd ed. School of Computing and Mathematical Sciences, University of Greenwich, UK.

7. Deanna D. Caputo, Shari Lawrence Pfleeger, Jesse D. Freeman, M. Eric John-son, Going spear phishing: Exploring Embedded Training and Awareness,3rd edition, January/February 2014.