International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Survey on: Email Security for Targeted Malicious Attacks

Jagdish R. Yadav[1], **A.K. Srivastava**[2]

[1]M.Tech.Student, IET, Alwar, Rajasthan, India

[2]Ph.D (AI), TIFR Mumbai IET, Alwar, Rajasthan, India

Abstract: ***The aim of this survey paper is to detect the Targeted email attacks as well as to find the malicious executables. Beyond spam or phishing designed to trick users into revealing personal information, targeted malicious email (TME) acquire sensitive information from targeted networks. These targeted malicious email attacks are not singular unrelated events; instead they are coordinated and persistent attack campaigns that can span years. This survey categorizes existing email filtering techniques, proposes and implements new methods for detecting malicious email and compares these newly developed techniques to traditional detection methods. Current research and commercial methods for detecting spam and phishing attacks, but not focused on addressing targeted malicious emails. Furthermore, this study first documents the existence of TME and characterizes it as a form of malicious email attack different than spam, phishing and other conventional illegitimate email. unseen malicious executables often arriving as email. Current anti-virus systems attempt to detect these new malicious programs but there is a overhead of updating of anti-virus This approach is costly and oftentimes ineffective. In this survey focus is to design a system that detects new, previously unseen malicious executables accurately.***

Keywords: Targeted malicious email, spam, phishing attack, recipient and persistent threats.

1. Introduction and Problem Statement

All organizations allow email to enter in their network some of the attackers target single user or small group. and extract important information by injecting malicious code in the email as well as in email attachment that creates backdoor in system. If we rely on current conventional detection methods, targeted email attack goes undetected and file attachment have malicious code that is also create trouble in network. A malicious executable is defined to be a program that performs a malicious function, such as compromising a system's security, damaging a system or obtaining sensitive information without the user's permission. Using data mining methods. Every day some malicious programs are created and most cannot be accurately detected until signatures have been generated for them.

2. Literature Review

R.M. Amin, Julie J.C.H. Ryan, and J. René van Dorp (2012) describe the approach to detect targeted email attack for that he focus on some learning as well as random forest algorithm. Based on 83 feature of email they detect either email is targeted or non targeted. Unsolicited email is not only a nuisance but can be potentially dangerous. Methods to filter it out work fairly well with conventional unsolicited commercial email (also known as spam) or email soliciting personal information (also known as phishing), but they don't work as well with targeted malicious email (TME) that facilitates computer network exploitation. Current detection algorithms work well for spam and phishing because it's easy to detect mass- generated email sent to millions of addresses it's possible to gather emails with similar characteristics and message content to prob- abilistically identify them. TME, on the other hand, targets single users or small groups in low volumes. It's tailored specifically to the target recipient and engineered to appear legitimate and trustworthy.

Salvatore J. Stolfo and Shlomo Hershkop (2005) describe the Malicious Email Tracking (MET) system is an online "behavior-based" security system employing anomaly detection techniques to detect deviations from a system's or user's normal email behavior, rather than solely by attempting to identify known attacks against a system via signature-based methods. The Email Mining Toolkit (EMT) is an offline data analysis system designed to assist a security analyst compute, visualize and test models of email behavior for use in MET. In this brief report, he enumerate the features implemented in the EMT system.

Matthew G. Schultz and Eleazar Eskin has write on detecting new malicious executables. To do this they separated their data into two sets: a training set and a test set with standard cross-validation methodology. The training set was used by the data mining algorithms to generate classifiers to classify previously unseen binaries had no examples in it that were seen during the training of an algorithm. This subset was used to test an algorithms' performance over similar, unseen data and its performance over new malicious executables. Both the test and training data were malicious executables gathered from public sources. they implemented a traditional signature-based algorithm to compare with the the data mining algorithms over new examples

2.1 The Email Transferring System

An email message relies on the Simple Mail Transfer Protocol (SMTP, defined in RFC 821 [4]) for transferring from the senders mail client (user agent.UA) to his/her mail server (mail transfer agent. MTA).This MTA in turn uses

SMTP to transfer the email to other intermediate mail servers (relaying MTAs) until the email reaches the recipients mail server (recipients MTA). Each MTA needs to contact its Domain Name Server (DNS) for the IP address of the next MTA before delivering the email. The end-user.s mail user agent (MUA) will normally use POP3 or IMAP4 protocols to retrieve their emails from an ISP.s mail server (recipients MTA).

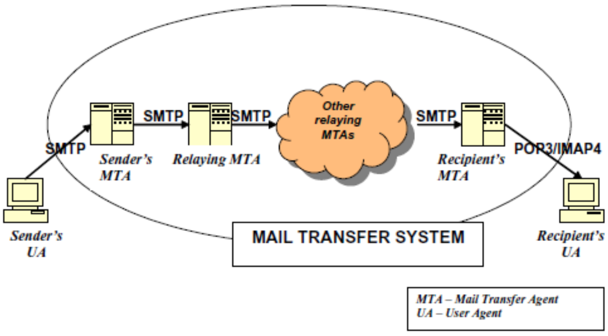

Figure 1. Basic model of Email Transferring System

MT Proxy works by acting as a proxy for the recipients MTA. Incoming SMTP connections (over TCP/IP) terminate on MT Proxy rather than the recipients MTA directly. MT Proxy establishes a new SMTP connection to the recipient. MTA for each incoming SMTP from an intermediate MTA. However, as each email is transferred through MT Proxy it is evaluated for evidence of being spam. If an SMTP connection appears to be carrying spam, MT Proxy slows down TCP/IP connection over which that particular SMTP connection is established.

2.2 Current email filtering techniques

2.2.1 Authentication

Wong and Schlitt (2006) and Lyon and Wong (2006) describe two methods of domain authentication called Sender Policy Framework (SPF) and Sender-ID. Both rely on the sending system publishing valid email server records in the Domain Name System (DNS). The receiving system is then able to verify that an email advertised as coming from a particular domain actually came from email servers authorized to send email on behalf of that domain. SPF and Sender-ID are very similar in approach and differ in the fields they use on the receiving end for the lookup.

Crocker et al. (2005), Otis et al. (2005), Leslie et al. (2005) discuss the components of Certified Server Validation (CSV) which is another authentication scheme leveraging DNS for domain validation. However CSV differs in that it uses the domain name in the Simple Mail Transfer Protocol (SMTP) HELO transaction. CSV first checks to ensure the server sending IP address matches the IP address in DNS for the domain used in the HELO transaction. Second, CSV verifies the reputation of that domain vs. the domain name advertised in the email headers. This difference is important when considering situations where individuals are sending email through a mail server where the From address and the mail server may not match. This difference between CSV and SPF/Sender-ID results in a different approach for handling spoofing. With the former, spoofing controls need to be handled on the sending server side to ensure the server is only sending email that it is supposed to send. With the latter, spoofing controls need to be handled on the receiving server side to ensure that the sending IP address is valid for the advertised From domain in the email headers. Another authentication mechanism, Domain Keys, is described by Delaney (2007), Allman et al. (2007) and Leiba and Fenton (2007). Domain Keys is different in that it leverages a public/private key cryptographic solution where the sending email server signs the email with a private key and the receiving email server validates the signature by retrieving the public key for the From domain in the email headers via DNS. This approach is similar to SPF/Sender-ID in that Domain Keys validates that a particular email server is authorized to send email for a domain advertised in the From email headers. Domain Keys differs from SPF/Sender-ID, however, because it does not require the sending domain to maintain lists of authorized email servers for the domain. Taylor (2006) describes Google Mail's (Gmail) approach to establishing sender reputation and it is heavily based on both SPF and Domain Keys as complementary approaches because each has its strengths and weaknesses. Other authentication like approaches include the Occam protocol described by Fleizach et al. (2007), The Occam protocol works in real-time on a per-email basis where the receiving email server asks the sending email server to validate that it sent a particular email based on the email Message-ID field. SPA uses a cryptographic based email address which encapsulates the policy in the email address itself. Not designed for person-to-person interaction, someone looking to receive communication from a party at a later date (e.g. online retailer) would generate a SPA and give that to the party for their explicit use. The policy in the SPA defines an expiration date and authorized senders who are allowed to use the SPA. Enforcement of this policy is done by the receiver. TEA is a challenge-response authentication scheme that uses hashes of previously exchanged email between two email addresses to authenticate that a new received email is being sent from the correct email server and not being spoofed.

All of these authentication approaches are focus on that an email being received is actually being sent by a system or person authorized to send an email from the advertised email address. A threat actor could break this authentication by registering in new domain, equip it with the appropriate authentication capabilities and then send spam from that domain. However, Internet-wide real-time block-lists would be quickly updated and tag email from this domain as illegitimate. Trying to scale this sort of approach would introduce a non-trivial cost to the actor. These authentication approaches can also be used to prevent a more advanced threat actor from sending a targeted spoofed email. However, these techniques being able to prevent targeted social engineering malicious email attacks. Furthermore, approaches like SPF, Sender-ID and Domain Keys, which are the predominant email authentication approaches in use today.

2.2.2 Contextual

The bulk of email filtering comes under the contextual analysis category. These are techniques which leverage the actual content of the email while making filtering decisions. The result of contextual analysis is based on a probabilistic answer with regards to the legitimacy or illegitimacy of an email instead of a binary answer typically associated with the authentication approaches described above. Basic approaches to contextual analysis include processing a set of rules, or heuristics, that assign a score to the presence of certain words or phrases in an email. Rules can be established using words or phrases commonly found in the types of email that are being sought. Stone (2007) uses a rules-based approach based on Natural Language Processing (NLP) and is able to achieve a 75 percent detection rate using four rules for detecting phishing emails. Evading these types of filtering techniques is rather trivial since a threat actor only needs to craft emails to change words that avoid any of the rules in the defined rule set.

Sahami et al. (1998) and Pantel and Lin (1998) describe machine learning Bayesian based approach for filtering spam. Interestingly, Sahami et al. incorporate additional properties in the classification vector for each email such as whether an attachment is present. They note that most junk email does not have an attachment and is sent at night. But this approach may have some exceptions each and every time this approach is not applicable.

2.2.3 Characterization

Some email filtering techniques are based on behavioral characterization techniques designed to focus on the behaviors of the actors sending the email. Similar to contextual approaches, probabilistic answers are the result. Bhattacharyya et al. (2002) created a tool called “MET” (Malicious Email Tracker) that mention a client/server architecture to track statistics of email sent and received to determine if there are viral propagations occurring. Any identified viral emails can be filtered out once identified, and new viral propagations can be discovered early. Alternative behavioral characterization approaches focus on identifying a fingerprint of the author of emails. the characterization based approaches to filtering email only suitable for filtering spam. Since email attacks are generally low volume and mimic other normal email characteristics such as rate and message content, filtering using these mechanisms is problematic.

2.2.4 Reputation

Reputation based approaches to filtering email are based on maintaining White lists and blacklists or calculating a level of trust through relationship linkages. White lists, blacklists and DNS-based Real-time Block Lists (DNS RBLs) are examples of list based reputation filtering. In these approaches, the reputation of an email is calculated based on known bad senders can decrease an email’s overall reputation whereas known good sending IP addresses can increase an email’s overall reputation. Erickson et al. (2008) use a combination of challenge-response and a persistent white list per user for filtering legitimate email. this approach manage the white list, to detect spam mails. However, they do make a fairly significant assumption that sender-based authentication services, described above, are a prerequisite to prevent simple spoofing

Jung and Sit (2004) analyze DNS based black lists and find that across spam analyzed over a roughly three year period, approximately 80 percent of spam sources are listed in at least one of seven popular DNS based black lists. However, they show that relying on only one or two lists is not sufficient since some lists are more conservative than others when determining which sources get listed.

A fundamental assumption in most of the reputation based approaches for filtering email is that senders are authenticated, if not, email from address will significantly minimize the effectively of these approaches.

3. Proposed system

This Survey propose a system that detect malicious email attacks as well as the malicious executables and doing so we are using a large data set of malicious links, IP addresses and Email ID. Between sender and receiver of Email there is a administrator who will verify the incoming Email by the request of receiver. if receiver has doubt about receiving E-mail ,he send verifying request to administrator who take the decision about it, either the email contains the malicious code or not. if the E-mail contains the malicious code then it will not place the E-mail in receivers Inbox rather than that mail will be placed in malicious box. The goal of this survey is to automatically design and build a scanner that accurately detects malicious code and URL in the Email before they have been given a chance to run.

4. Conclusion and Future Scope

Current detection algorithms mainly focus for detecting spam and phishing email sent to millions of addresses, science this survey approaches a technique to develop a system to detect the malicious email attack which acquire sensitive information of email receiver. To protect the sensitive information from unauthorized thread actor the system administrator will verify the E-mail and take decision about the email.

In future this system can be extended to detect malicious attachment.

References

[1] R.M. Amin, Julie J.C.H. Ryan, and J. René van Dorp “Detecting Targeted Malicious Email ” P George Washington Univ. June 2012. www.computer.org/security

[2] Salvatore J. Stolfo, Shlomo Hershkop, Ke Wang, Olivier Nimeskern,” EMT/MET: Systems for Modeling and Detecting Errant Email” Columbia University {sal, shlomo, kewang, on2005}@cs.columbia.edu

[3] Matthew G. Schultz, Eleazar Eskin and Erez Zadok “Data Mining Methods for Detection of New Malicious Executables” Department of Computer Science Columbia University{mgs,eeskin}@cs.columbia.edu

[4] J. B. Postel, "RFC 821 Simple Mail Transfer Protocol", August 1982, http://www.cis.ohio-state.edu/cgi-bin/rfc/rfc0821.html

[5] Sebastiani, F., 2002. Machine learning in automated text categorization. ACM Computing Surveys, 34(1):1-47.

[6] Dwork, C., Goldberg A., Naor M.. On memory-bound functions for fighting spam. In Proceedings of the 23rd Annual International Cryptology Conference (CRYPTO 2003), August 2003.

[7] R.J. Hall. How to avoid unwanted email. Communications of the ACM, March 1998.

[8] Golbeck, J., Hendler, J. Reputation network analysis for email filtering. In Proceedings of the First Conference on Email and Anti-Spam (CEAS), 2004.'

[9] J. Lyon and M. Wong. RFC4406 - Sender ID: Authenticating E-Mail, April 2006. URL http://www.ietf.org/rfc/rfc4406.txt.

[10] J. Leslie, D. Crocker, and D. Otis. Domain Name Accreditation (DNA), February 2005. URL http://tools.ietf.org/html/draft-ietf-marid-csv-dna-02.

[11] E. Allman, J. Callas, M. Delaney, M. Libbey, J. Fenton, and M. Thomas. RFC4871 Domain Keys Identified Mail (DKIM) Signatures, May 2007. URL http://www.ietf.org/rfc/rfc4871.txt

[12] Bradley Taylor. Sender Reputation in a Large Webmail Service. In proceedings of CEAS - Conference on Email and Anti-Spam 2006, 2006. URL http://www.ceas.cc/2006/19.pdf.

[13] Chris Fleizach, Geoffrey M. Voelker, and Stefan Savage. Slicing Spam with Occams Razor. In proceedings of CEAS 2007 - Fourth conference on email and anti-spam, August 2007.

[14] Allen Brian Stone. EBIDS-SENLP: A System to Detect Social Engineering Email Using Natural Language Processing. Master's thesis, University of Maryland, 2007.

[15] Mehran Sahami, Susan Dumais, David Heckerman, and Eric Horvitz. A Bayesian approach to filtering junk email. In Learning for Text Categorization: Papers from the 1998 Workshop, Madison, Wisconsin, 1998.

[16] Manasi Bhattacharyya, Shlomo Hershkop, and Eleazar Eskin. MET: An experimental system for Malicious Email Tracking. In Proceedings of the 2002 workshop on New security paradigms, pages 3– 10. ACM, 2002. Doi: http://doi.acm.org/10.1145/844102.844104.

[17] David Erickson, Martin Casado, and Nick McKeown. The Effectiveness of White listing: a User-Study. In Conference on Email and Anti-Spam, 2008.

[18] Jaeyeon Jung and Emil Sit. An empirical study of spam traffic and the use of DNS black lists. In IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pages 370–375, New York, NY, USA, 2004. ACM. doi:http://doi.acm.org/10.1145/1028788.1028838.

Author Profile

Jagdish R.Yadav has done B.E.(Comp.Engg.) BDCOE,Sevagram and presently pursuing M. Tech (CS), IET, Alwar (Rajasthan) , India

Dr. Anoop Kumar Srivastava has done B.Tech. (Elect.). KNIT Sultanpur, , M.M.S. (Operations) CRKIMR Mumbai, and Ph.D. (Computer Science) TIFR Mumbai, FIE, LMISTE, SMLICEIT.