Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 7, July 2013, pg.406 – 412

RESEARCH ARTICLE

Identifying Under Attack Hateful Email

M. Ramu[1], B. Triveni[2], L. Srinivasa Rao[3]
[1]Assistant Professor, Department of Computer Science and Engineering,
Teegala Krishna Reddy Engineering College, JNTU Hyderabad, Andhra Pradesh, India
[2]Assistant Professor, Department of Computer Science and Engineering,
Teegala Krishna Reddy Engineering College, JNTU Hyderabad, Andhra Pradesh, India
[3]Scientist, Department of Information Systems, Defence Research and Development Laboratory (DRDL),
Ministry of Defence, Hyderabad, Andhra Pradesh, India

[1] ramumooducse@gmail.com; [2] triveni.banavatu@gmail.com

Abstract— 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 or email soliciting personal information but they don't work as well with under attack hateful email (AHE) 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 nit's possible to gather emails with similar characteristics and message content to probabilistically identify them. AHE, on the other hand, targets single users or small groups in low volumes. It's tailored specifically to the goal recipient and engineered to appear legitimate and trustworthy. If we rely on current conventional detection methods, AHE goes undetected.

Key Terms: - filtering; emails; hateful attacks; records

I. THE UNDER ATTACK HATEFUL EMAIL CHALLENGE

A network defender encounters different classes of threat actors with varying intents and capabilities. Conventional computer network attacks exploit network-based listening services such as Web servers, where asunder attack attacks oft en leverage social engineering through vehicles such as email. Email is especially dangerous because nearly all organizations allow email to enter their networks. In mid-2005, the UK National Infrastructure Security Co-ordination Centre1 and the US Computer Emergency Response Team2 issued technical alert bulletins about targeted, socially engineered emails that drop Trojans to exfiltrate sensitive information. The intrusions occurred over a significant period of time, evaded conventional firewall and antivirus capabilities, and enabled adversaries to harvest sensitive information. In 2007, various government agencies experienced intrusion attempts.3 The US-China Economic and Security Review Commission's 2008 and 2009 reports to Congress summarize open source reporting of under attack attacks against US military, government, and contractor systems to collect sensitive information.4 A report prepared for the US-China Economic and Security Review Commission profiled an advanced cyber intrusion and documented AHE5 In all of these examples, the threat actors weren't necessarily looking for immediate financial gain. For such advanced persistent threats, acquiring valuable information is the real intention. Although many victims of illegitimate email have money, only certain organizations have the type of valuable information that yields long-term strategic advantage. This level of targeting and sophistication suggests a patient threat actor with the resources to reconnoiter a target environment and craft emails relevant to the recipients, using email addresses, subject lines, and content tailored to entice recipients to open the message. The threat actors can then attach hateful files or Web links or repurpose previously sent email appended with hateful content. Clearly, threat actors can't use this sort of advanced targeting on an Internet-wide scale: they're after specific information.

II. Records Set Structure

Given AHE's specific features and the failure of traditional filtering techniques to reliably detect it, developed an alternative filtering procedure. Figure 1 outlines our process. Look at features of the email that other filtering techniques don't typically extract, classifying them as persistent threat and recipient-oriented features. These features on the basis of our analysis of a large Record set of actual AHE from a Fortune 500 company with more than 100,000 email users fully reviewed our intentions with the company's legal counsel and information security personnel. Although we conducted the research using actual Record, we've sanitized all results to anonymize both the company and any users of the company's email system. To the best of our knowledge, reliable, scalable, and automated AHE detection isn't yet possible with commercially available tools. For other researchers looking to create similar Record sets, we recommend partnering with organizations that are already manually identifying AHE Typically; the Record sets used to evaluate email-filtering techniques are incomplete or are an amalgamation of several different Record sets. For example, the PU1 and ling-spam corpora, commonly used for evaluating the performance of spam filters, are made up of known spam and known legitimate emails from different sources.6 Privacy concerns make it difficult to obtain legitimate email for analysis, and to further complicate matters, Record sets sometimes lack email header information or are sanitized to the point where useful information is lost. Our study had to use full and complete emails, because a critical goal was to measure the added value of leveraging features of hateful email that are persistent threat and recipient oriented. We leveraged complete emails from the company and additional recipient context, such as full name, job title, and business division membership. The complete Record set consists of three classes of emails: non-targeted hateful email (referred to as NAHE1),under attack hateful email (referred to as AHE1), and an evaluation set containing both AHEs and NAHEs (referred to as TS1). We used NAHE1 and AHE1 to construct the AHE-filter technique and TS1 for evaluation. Figure 2 provides context for the new features we incorporated for AHE detection.

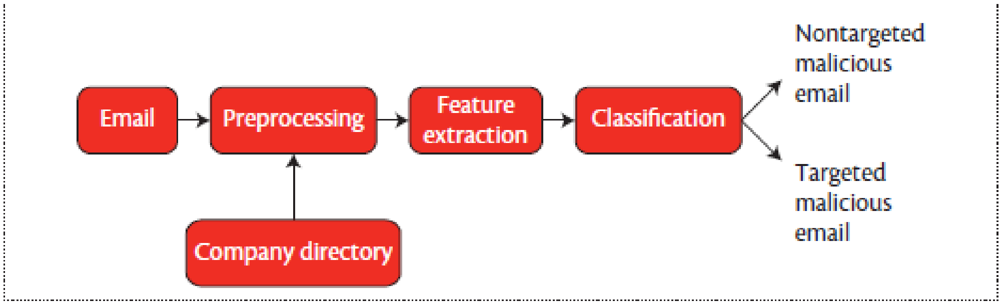


Figure 1. Classification process. A simplified view of our classification process first involves preprocessing email, leveraging company-specific information. Persistent threat and recipient-oriented features are extracted and the associated emails are classified using a random forest classifier

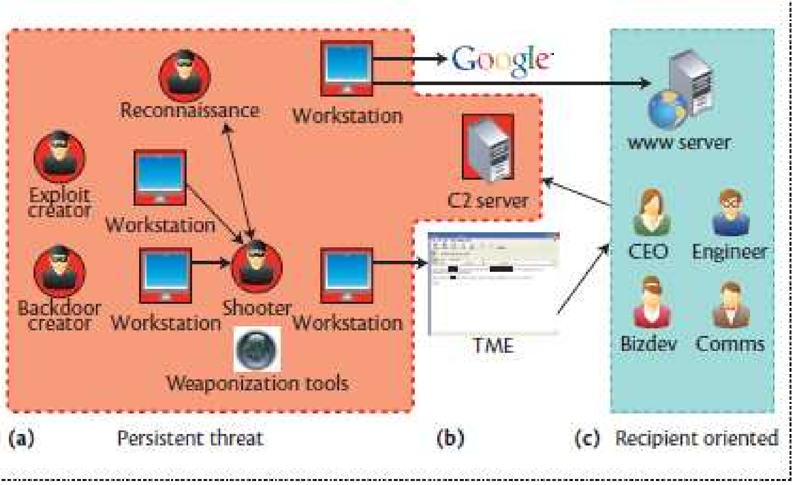


Figure 2. New features used for detecting under attack hateful email (AHE). (a) The supply chain of components and threat actors necessary for creating a AHE (b) The Internet, which stands between the threat actors and email recipients. (c) AHE recipients

III. Trends with AHE and NAHE

Separate analysis of AHE informed the feature selection for our extraction purposes. Figure 3 shows the number of AHEs received by accounts at the target company: the vast majority of accounts received no AHEs, although a select set of accounts received more than one. Figure 4a shows the cumulative distribution function of Google search hits for the company's employees' email addresses, and Figure 4b shows the average amount of AHEs received by those email addresses sharing the same hit count in Google. The *x*-axis uses just the first 24 Google search hit bins, which accounts for 99.93 percent of the total population. Later bins are sparsely populated. We can see a positive correlation between these two variables. Some AHEs appear to target employees with specific job titles. Table 1 shows the top 15 job titles in the target company by number of employees, NAHEs received, and AHEs received. It's interesting that the most common job title, systems engineering, isn't one of the top 15 AHE recipients. Furthermore, the international business development job title, which consists of only 44 people in a company of more than 100,000 employees, is the third most under attack group.

3.1 Persistent Threat and Recipient-Oriented Features

Figure 2 provides the context for the persistent threat and recipient-oriented features that extracted to support the classification objective. Extracted a total of 83 features) and used them as input to a random forest classifier to construct the AHE-filtering technique. Table 2 lists the top 10 of these 83 features.

IV. Persistent Threat

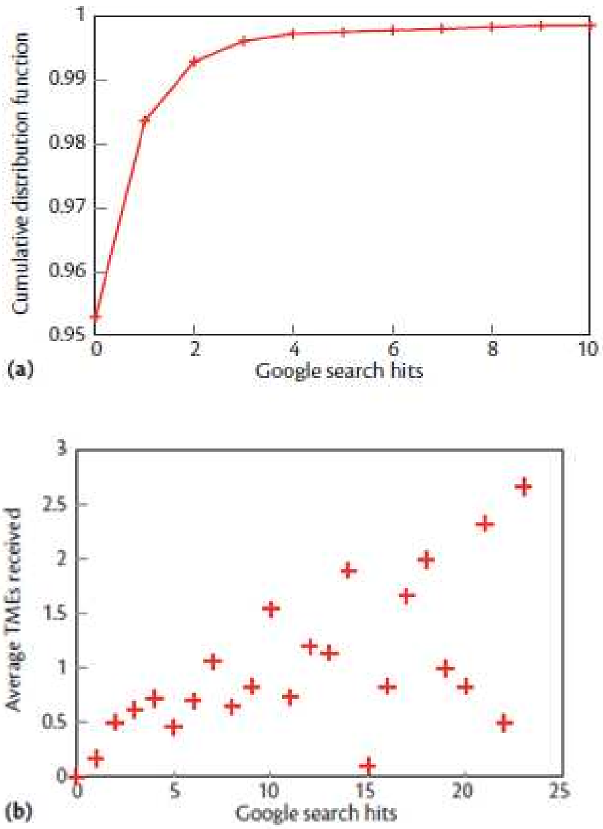
When threat actors weaponize (or embed hateful content in) an email, they can leave fingerprints useful for detection. Inevitably, threat actors, being human, resort to automation or other procedural techniques that can enhance detectability across several repeated intrusion attempts. To save on cost, threat actors might reuse weapons with different delivery vehicles. The combination of tools, techniques, and procedures measures capability. Both the tools used and locale are persistent threat features that incorporate in AHE-filtering techniques.

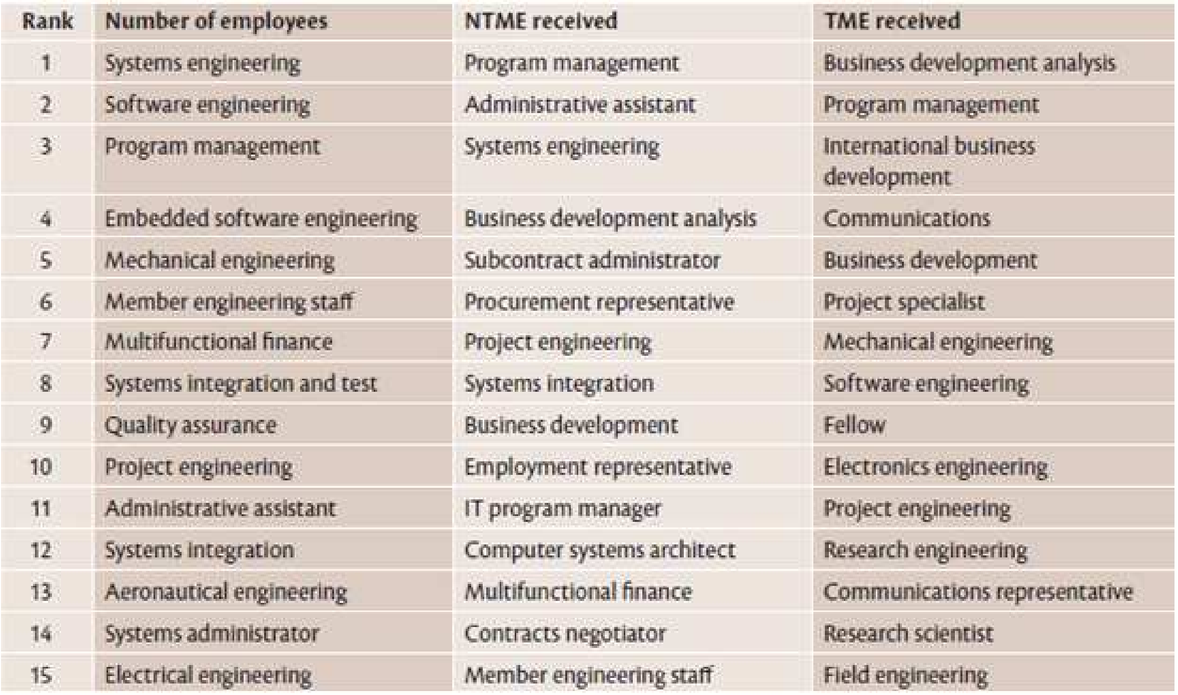
Some automation tools actually leave names in an email, whereas other tools leave more subtle clues. For example, feature 10 in Table 2 is an artifact of certain email tools. When a threat actor is preparing and launching an email weapon, certain elements of his or her locale might be left in attachments or in the email itself. Infer locale through language settings, character encodings, time-zone settings, Internet Protocol addresses, and system host names.

V. Recipient Oriented

Threat actors might send emails to a particular individual because of his or her role in an organization. They might target the company's CEO, thinking his or her system might have sensitive information. Employees in business development might be prone to AHE simply because their email addresses are more readily available as a function of their job; a senior-level employee might be more likely to be under attack than an entry-level employee.

Just as some sender-focused reputation techniques maintain lists of known bad senders, recipient reputation involves maintaining a list of recipients known to receive AHE It's conceivable that threat actors maintain a Record base of email addresses for a specific target organization and that these email addresses might receive a higher volume of AHE over time. Feature 1 in Table 2 was only feasible as a recipient-oriented feature because the company keeps a log of those who have previously been targeted. Another dimension of reputation includes email visibility. Presumably, those email addresses that are more publicly available are more likely to be targeted. Email address visibility can be as straightforward as the number of times an email address appears in Internet search engine results. Furthermore, employees who have left a company might continue to receive AHE to their no-longer-valid





Rank	Number of employees	NTME received	TME received
1	Systems engineering	Program management	Business development analysis
2	Software engineering	Administrative assistant	Program management
3	Program management	Systems engineering	International business development
4	Embedded software engineering	Business development analysis	Communications
5	Mechanical engineering	Subcontract administrator	Business development
6	Member engineering staff	Procurement representative	Project specialist
7	Multifunctional finance	Project engineering	Mechanical engineering
8	Systems integration and test	Systems integration	Software engineering
9	Quality assurance	Business development	Fellow
10	Project engineering	Employment representative	Electronics engineering
11	Administrative assistant	IT program manager	Project engineering
12	Systems integration	Computer systems architect	Research engineering
13	Aeronautical engineering	Multifunctional finance	Communications representative
14	Systems administrator	Contracts negotiator	Research scientist
15	Electrical engineering	Member engineering staff	Field engineering

Table 1. Not ready to attack hateful emails received and attacked hateful emails received

VI. EXPERIMENTAL SETUP

We used the random forest classifier8 to separate NAHE from AHE Several characteristics of this classifier made it ideal for the Record sets in this study:

it can handle a large number of features;
it can handle a large number of emails;
it can handle a mixture of binary, numeric, and categorical features;
it generally doesn't over it;
it can handle missing features;
it trivially parallelizes the algorithm to scale up for huge Record sets;
it can estimate which features are more important than others; and
it can handle unbalanced Record sets (for example, a much greater number of NAHEs than AHEs).

	Actually AHE	**Actually NAHE**
Predicted as AHE	True positive (TP)	False positive (FP)
Predicted as NAHE	False negative (FN)	True negative (TN)

Table 2. Possible outcomes from the classifier

Traditional decision-tree classification algorithms split each node using the *best split* from all available features. The best split is that which provides the most separation in the Record. With random forests, each node splits (using the best split) from a randomly selected set of features at that node. In addition, they create multiple decision trees using bootstrap samples (random selections with replacements) from the Record set. These trees are created independently of each other and are classified according to a simple majority vote from the trees in the forest. The algorithm8,9 is as follows:

1. In this study, trees grow to maximum size: k = number of trees to create; m = number of random features to select for node splitting; and d = maximum depth of the trees.
2. Select k vectors from the training Record such that vector θk is chosen independent of $\theta 1, \ldots, \theta k - 1$.
3. For each of the bootstrap samples, grow a tree Tk, where each node splits using the best split from m randomly selected features. The result is multiple tree classifiers $Tk : h(\mathbf{x}, \theta k)$, where **x** is an input vector of un
4. To classify **x**, process that feature vector down each tree in the forest. Each tree will output a classification, also known as a *vote*. If $Ck(\mathbf{x})$ represents the classification of the kth tree in the forest, then the aggregate classification of the forest, $Cforest(\mathbf{x})$ = majority vote $C\mathbf{x}\{()\}kk1$.

The 83 features extracted from email are represented as a vector of features. The output of the random forest classifier for a particular email is binary, classified as either AHE or NAHE using the email's specific vector of persistent threat and recipient-oriented features as input.

When the classifier correctly predicts a AHE, it's a *true positive* (TP). When the classifier correctly predicts an NAHE, it's a *true negative* (TN). When the classifier predicts an NAHE as AHE, it's a *false positive* (FP) or Type I error. When the classifier predicts a AHE as NAHE, it's a *false negative* (FN) or Type II error. Table 3 shows the possible outcomes from the classifier.

The false positive rate (FPR) is the proportion of NAHE that was incorrectly classified as AHE The specificity is equal to 1 − FPR, where the FPR is FP/FP+TN

The false negative rate (FNR) is the proportion of AHE that was incorrectly classified as NAHE The sensitivity is equal to 1 − FNR, where the FNR is FN/FN+TP.

VII. OUTCOME

Recall the construction of the NAHE1, AHE1, and TS1 Record sets. At first, a 10-fold cross validation as our evaluation method for the joint NAHE1–AHE1 Record set. Later, we used the joint NAHE1–AHE1 Record set for training, but instead of doing cross validation, we used the independent TS1 Record set to evaluate the AHE filter constructed using the joint NAHE1–AHE1 Record set.

7.1 NAHE1–AHE1 Record set

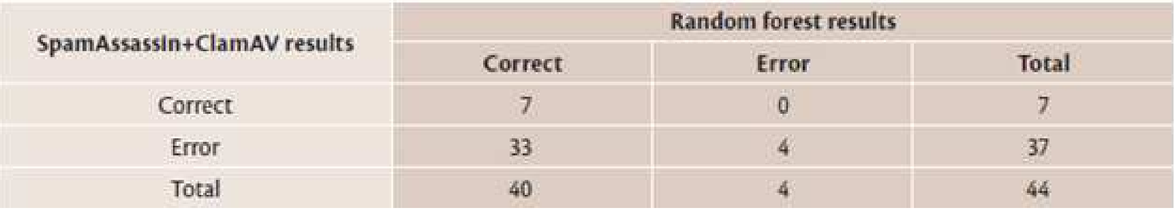
A. Analysis of the NAHE1–AHE1 Record set by using the random forest classifier's measure of feature importance. A random forest using all 83 features with k = 50 and m = 30 produced the best result—an FNR of 0.6 percent.

Processing the NAHE1–AHE1 Record set using Spam Assassin (http://spamassassin.apache.org) results in an FNR of 73 percent, indicating that a large volume of AHE evades Spam Assassin. Clam AV (www.clamav.net) has even poorer performance with a 90 percent FNR. Serializing both Spam Assassin and Clam AV (SpamAssassin+ClamAV) resulted in a 67 percent FNR.

Figure 5a shows the top 20 features (in order of feature importance) that have to be removed from the random forest classifier before its FNR approaches that of SpamAssassin+ClamAV. Figure 5b presents analysis results following a similar process, but it removes the least important features first. Even with only one feature remaining (removed in order of increasing feature importance), the random forest classifier outperforms Spam Assassin and ClamAV.

A Mc Nemar test, summarized in Table 4, comparing the random forest classifier against SpamAssassin+ClamAV, yields a $\chi 2$ test statistic of 1,541.1, which is greater than the critical value of 6.635 at the $\alpha = 0.01$ level of significance.

We must reject the null hypothesis that the two detection methods are the same in their abilities to detect AHE Table 4 shows that the random forest–based AHE filter technique developed herein identifies 2,300 out of 2,315 AHEs correctly (99 percent),



SpamAssassin+ClamAV results	Random forest results		
	Correct	Error	Total
Correct	7	0	7
Error	33	4	37
Total	40	4	44

Table 3.Evaluation contingency table for AHE1 detection

VIII. CONCLUSION

Threat actors might inadvertently leave remnants of information such as file paths, time zones, or even author names.10 all these features might associate multiple intrusion attempts into a related campaign. In addition, organizations can track features that characterize the types and amounts of email received by a particular email address. For example, for each recipient, the number of emails and attachments received over a fixed time period might help uncover email that falls outside of his or her normal receiving patterns.

REFERENCES

[1] under attackTrojan Email Attacks, briefing 08/2005, Nat’l Infrastructure Security Co-ordination Centre, 2005; www.egovmonitor.com/reports/rep11599.pdf.

[2] under attackTrojan Email Attacks, tech. cybersecurity alert TA05-189A, US-CERT, 2005; www.us-cert.gov/cas/techalerts/TA05-189A.html.

[3] J.A. Lewis, “Holistic Approaches to Cybersecurity to Enable Network Centric Operations,” statement before Armed Services Committee, Subcommittee on Terror¬ism, Unconventional Threats and Capabilities, 110th Cong., 2nd sess., 1 April 2008.

[4] 2009 Report to Congress of the U.S.-China Economic and Security Review Commission, report, Nov. 2009; www.uscc.gov/annual_report/2009/annual_report_full_09.pdf.

[5] B. Krekel, Capability of the People’s Republic of China to Conduct Cyber Warfare and Computer Network Exploita¬tion, Oct. 2009; www.uscc.gov/researchpapers/2009/NorthropGrumman_PRC_Cyber_Paper_FINAL_Approved%20Report_16Oct2009.pdf.

[6] I. Androutsopoulos et al., “An Experimental Compari¬son of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-mail Messages,” Proc. 23rd Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval, ACM, 2000, pp. 160–167.

[7] R.M. Amin, “Detectingunder attackHateful Email through Supervised Classification of Persistent

Authors Bibliography

M Ramu, working as Assistant Professor in the Department of Computer Science and Engineering and having 5 years of teaching experience. He has done M.TECH in Software Engineering.

B Triveni, working as Assistant Professor in the Department of Computer Science and Engineering and having 7 years of teaching experience. She has done M.TECH in Software Engineering.

L Srinivas Rao, working as Scientist in the Department of Information Systems and having 12 years of Industrial experience in DRDL in Hyderabad. He has done M.TECH in Computer Science and Engineering.