Electronic Preprint for Journal of Information Processing Vol.28

Regular Paper

Zero-day Malicious Email Investigation and Detection Using Features with Deep-learning Approach

Sanouphab Phomkeona[1,a)] Koji Okamura[1,b)]

Received: June 17, 2019, Accepted: November 29, 2019

Abstract: Cyber hackers use email as a tool to trick, inject or drop malicious software into the recipient's device. Everyday users have to face off against, phishing or malicious emails and it would be a huge problem for whole organizations even if only one user clicked on a single link from this malicious email. The difficult issue is how to classify and detect those malicious emails from ordinary, especially spear phishing emails, which are designed for a particular target, or zero-day malicious emails that no one has ever found until now. In this paper, we introduce a way to classify and detect zero-day malicious emails by using deep-learning with data investigated from the email header and body itself, combined with dynamic analysis information as a group of features. Four different language email datasets can be used to train and test the system to simulate real-world diversity and zero-day malicious email attack situations. We succeeded in obtaining a satisfactory accuracy rate for detection results for both zero-day malicious email types and normal spam.

Keywords: zero-day malicious spam, spam detection, deep-learning, features extraction

1. Introduction

Nowadays, people are easily accessing the internet compared to the past. Among the thousands of types of internet services, email is still the most popular choice. We are all using emails for work, to contact people, and email is a basic requirement for system registration to apply for services on the internet. On the other hand, international cybersecurity reports such as the Cisco 2018 annual cybersecurity report [1] said that the most common initial hacker attack method is by using malicious emails and APT *AdvancedPersistentThreat* attacks. Moreover, there is confirmation from other cybersecurity organizations, such as the Center for Internet Security report [2], that the number of cyberattack incidents increased in recent years, and doubled in 2017. The report showed that malicious spam (malspam) infection vector continues to stay the primary entry vector, from December 2017, increasing 8% in January 2018. The ISTR report from Symantec [3] also confirmed that the percentage of emails which contain spam is 53%, 53% and 55% from 2015, 2016 and 2017 respectively. This means about 14.5 billion malicious spams are sent every single day in the first quarter of 2018 with spear phishing being the number one infection vector employed by 71% of organized groups in 2017. Spam mail is defined as the type of electronic spam where unsolicited messages are sent by email. Many spam emails are merely advertisements for products or services but some of them also contain disguised links that appear to be phishing websites or sites that host malware as scripts or other executable file attachments. These days, spam emails can also directly contain various malicious files such as Microsoft Office documents, pdf, JavaScript or PE files. Therefore, we call this type of email malicious spam (malspam). Consequently, a zero-day malspam is an email that contains a cyber-attack exploiting a vulnerability that has not been disclosed publicly. Machine learning is popular for automatically detecting both known and unknown threat types of infections. The technology can learn to identify unusual malspam in large numbers and automatically detect new malspam in the future. Above the popular well-known algorithms NB, SVN, or K-Nearest Neighbor (KNN), running the MPL neural network algorithm on test data seems to be the best way to detect spam in terms of efficiency. To detect malspam, email messages, headers, URL links, hash values and other features have been used to train the machine. However, for zero-day malspam, new APT threats or spear phishing, which is particularly designed to fool users and systems, is not easy for AI system to detect by training from those main features above. In this case, zero-day malspam still requires cybersecurity experts to investigate and analyze them manually. The question is, what is the difference between the information that we give to train the AI system and the information that experts use for classifying malspam? To verify this, more than 200 malspams have been manually investigated and analyzed [4], and we developed techniques to detect malware infection on the workstation [5], [6] in our earlier works. We have found the difference between malspam information and legitimate email from both header and body. Malspam usually has suspicious email header information, such as misconfigured time format, the unrelated relationship between domain time zone and language used, an unclear title that might have been taken from machine translation, etc. So, in this research, we

[1] Graduate school of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819–0395, Japan
[a)] sanouphab@fe-nuol.edu.la
[b)] oka@ec.kyushu-u.ac.jp

design approaches to obtain more informational features similar to how cybersecurity experts do when investigating suspicious emails. We then use a deep-learning approach to increase detection accuracy and aim for an automatic system of zero-day malspam detection.

2. Related Research

Much research has already been done on spam and phishing email filtering by using machine learning with basic message information. In addition, Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (K-NN) and Deep-Learning with Artificial Natural Network (NN) are the most common ways to classify email threats and the current accuracy of those classification results are already quite satisfactory. In 2017, Ajaz et al. [12] proposed a sophisticated and robust e-mail abstraction scheme based on Bayesian with a new scheme and efficiently captured the near duplicate of the spams as well as achieved efficient similarity matching and reduced data storage. Al-Jarrah et al. [7] identified potentially useful email header features for email spam filtering and used them as input to several machine learning-based classifiers and compare their performance in filtering email spam. More than 10 features from email header, such as the number of hops, span time, domain address and legality, etc., were used and their results showed that Random Forest classifier has the best performance with an average accuracy, precision, recall, F-Measure, ROC area of 98.5%, 98.4%, 98.5%, and 98.5%, respectively. Wang et al. [16] also proposed the similar spam classification technique by using features of the sender and receiver address (To, CC, BCC), mail user agent and message-ID to train the system. Similar research by Hu et al. [17], Wu et al. [14], Ye et al. [15] and Sheu [13] also classified spam email by using different features information such as the size of an email, time, length of sender-destination field, sender IP or email subject and others with machine learning.

Shi and Xie [20] propose a reputation-based collaborative anti-spam approach to adopt fingerprinting techniques and for evaluating if a reporter is genuine by using MIME features of emails. By dividing each incoming email into five subparts: email header, text/plain content, text/html content, embedded resources and attachments, then generate a weighing fingerprint for different MIME subparts, they can achieve better performance and robustness than other methods.

Similarly, our research is also using email features information with deep-learning to classify and detect malicious email but we extract and collect more features information from both the email header and body and a relationship between them by using static and dynamic analysis. Moreover, these days zero-day malspam usually first release in one language and then be translated into other languages to spread into world-wide. So, we extract a new method to collect email subjects and titles in different languages to detect another version of malspam by using machine translation which was used by recent year zero-day ransomware attacks. While the related research above do not guarantee to detect zero-day malspam, we do aim to achieve this goal.

Not only bad email classification, but phishing webpage classification is also a popular research applied with machine learning. Mao et al. [8] proposed a learning-based aggregation analysis mechanism to determine the similarity of page layouts and detect phishing pages. Their approach automatically trains classifiers to determine web page similarity from CSS layout features, without requiring a human expert. By using SVM and Decision Tree (DT), the method achieved 93% and more than 95% accuracy results. Moghimi et al. [9] proposed two feature sets to improve a detecting phishing attack performance and preventing data loss in internet banking webpages. By using relevant features plus page resource identity feature set and page resource access protocol feature set with SVM, they got 99.14% true positive with 0.86 false positive alerts. Sonowal et al. [10] proposed a phishing detection model with a multi-filter approach. The result from their experiment shows that the model is capable to detect phishing sites with an accuracy of 92.72%. Basnet et al. [11] evaluated two common feature selection techniques, correlation based and wrapper-based techniques, for phishing detection. They compared the features selection techniques by using two feature space searching techniques, then conducted the experiments and evaluated results on a real world dataset with more than 16,000 phishing webpages and more than 32,000 non-phishing webpages. In our research, because the phishing spams usually contained a link or internal file attachment, we use API to upload and check suspicious links and files with free online dynamic analysis service and collect the results. Feature information such as Link status, File scan result, SHA-256, and others are extracted from the dynamic analysis results and given for the AI system process.

3. Method

Recently, neural networks and deep learning provides the best solutions to many problems in image recognition, speech recognition, and natural language processing. In this paper also we design to use neural network and deep learning by aiming to archive the best result of zero-day malspam detection. However, the result will also depend on which features we give to the system.

Features in machine learning is a piece of information included in the representation of the data they are given. It is well-known that a performance of machine learning algorithms depends on the representation of the data they are given. Most of the popular research on spam classification uses a set of features that come from the email header and body themselves directly, such as domain name, IP address, email title and body, or other features from text words analysis. Similar to what the security experts do, we collect and extract features from the email header analyzed information and email body dynamic analysis information that could be a clue to indicate suspicious email and judge them whether or not to be malspam or even unknown (zero-day) malspam. From analyzing more than 500,000 emails, we have discovered that a relationship between those headers and body features information is very important. For example, more than 99 percent of work emails are sent only within the working time period (8AM–8PM) which corresponds with the sender's domain time-zone and language used (in case the language is not English). In case of Japanese email datasets, the result shows that Japanese people commonly use Japanese languages to communicate with the receiver during working time by using local email

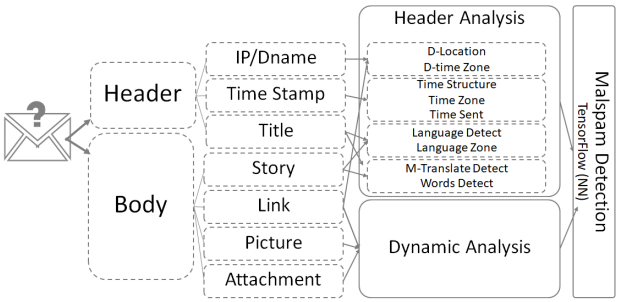

Fig. 1 Email features extraction.

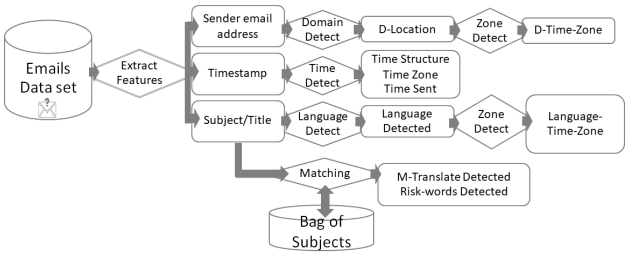

Fig. 2 Features extraction from email header's flows.

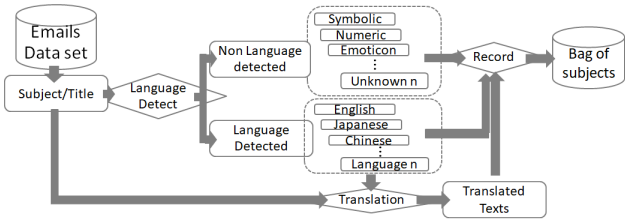

Fig. 3 Bag of subject database flows chart.

domain service. On the other hand, malspam can be any language, sent in random time and might use domain service from a different geolocation. This means, the common Japanese email characteristic should be Japanese language title, correctly configured Japanese time zone, sent during the working time period from domain located in Japan time zone, while malspam can be any language title, misconfigured time zone, sent in any time from any domain outside Japan. Of course, there is a chance that malspam also has all common characteristics in the email header, that is why features from the email body is also important. Generally, the URL-based approach, content-based approach and the combination of those two are popularly used to generate features to detect phishing websites or phishing emails. So, from well-known features such as email title, body message, URL link, IP address, time stamp, etc., in this research, we analyze more information such as the domain location, domain time zone, language detected, machine translated detected, uncommon time format, sent time after normal work hours as well as relationship between each feature are also included. Consequently, the dynamic analysis for links, pictures, and attachment files from the email body includes link status, file scan result, file type, SHA-256, other file names, file size, last scan time, etc. **Figure 1** shows the overall features of extraction flow from both the email header and body.

3.1 Email Header Extraction

In the email header extraction part, **Fig. 2** shows the flow chart of this method. From the email dataset we first extract 3 main features: source address, timestamp and subject/title from email header directly. Then from the source address, we discover and extract domain location and domain time zone by using Whois API. At the same time, from timestamp we detect the time structure, time zone, time sent and create time related features. From the subject/title, we detect language and extract a language time zone feature. We also match subject/title with a bag of subject database to detect machine translated and risk words detected features.

Define an email as the symbol e extracted into 2 main features header and body: $e = \{e_h, e_b\}$. Then from the email header and body part e_h and e_b extract into subfeatures $e_h = \{{e_h}^1, {e_h}^2, ...{e_h}^i\}$ and $e_b = \{{e_b}^1, {e_b}^2, ...{e_b}^i\}$.

3.2 Bag of Subject Database

Bag of subject is a concept similar to bag of words or bag of features which is popularly used in image processing. Because these days the attackers usually translate malspam into many different languages to expand an attack to many targeted countries world-wide, inside the database, we collect all types of email subject/title including both normal and abnormal language and aim to collect several language based email datasets other than English. By using language detection API, we can identify the email subject language and record the original subject/title into the database. Moreover, we use Yandex-translate API to translate the subject/title from the original form into other languages and record them into the database. In this research, we collect 4 languages (English, Chinese, Japanese and Lao) for experiments. However, as we randomly check the translation result from Yandex-API, the translation accuracy is still low compared to a well-known commercial Google translation API. We suggest adding more translation languages and a better accuracy translation result in the future for more efficiency. Finally, in case that subject/title is an abnormal language such as symbolic, digit numbers, emoticon or blank, these none-language subjects are directly sent to the database without translation. **Figure 3** shows a Bag of subject database flows chart.

3.3 Email Body Extraction

From the email body, **Fig. 4** shows how we extract features from the email body. First, we detect and extract 4 features included a story, URL link, picture and attachment file. For the email story, it has the same process as a subject/title feature, which is a detected language, classify words and store in the database. URL link, picture and attachment file will be uploaded into free online dynamic analysis service via Virustotal API. We then extract features from dynamic analysis result obtained from Virustotal API report include link status, link scan result, URL link, link last analyze date, web category, file type, file size, file name, file scan result, file SHA-256, file type detected, file last analyze date and other file names.

Table 1 shows all 27 features with their descriptions that we extracted from emails which used in deep-learning neural network model to classify and detect zero-day malspam.

Figure 5 displays the overall of the proposed method which

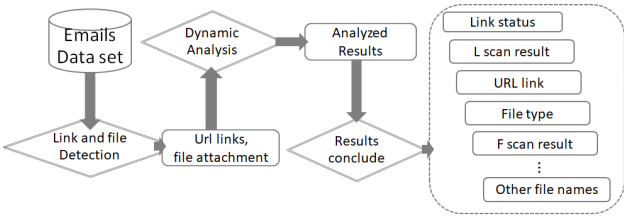

Fig. 4 Features extraction from the email body's flows.

Table 1 Features list extracted from the email header and body.

Feature Name	From	Description
Sender address	Header	Email address of the sender
D location	Header	Location of the domain that sender use
D time zone	Header	Time zone of the domain location
Time structure	Header	Email time format
Email time zone	Header	Time zone set in email
Email time sent	Header	Time that email was sent
L detected	Header	Language detected from email subject
L time zone	Header	Time zone of the language used in subject
Subject/title	Header	Email subject/title
M detected	Header	Machine translated detected result
W detected	Header	Risk words matching result
N detected	Header	None language detected from subject
Link status	Body	Current status of targeted website
Link scan result	Body	Result of link scan from dynamic analysis
L last analyze	Body	Link's last analyzed date
Url	Body	Link url
Web category	Body	Targeted website category
F detected	Body	Detected a file attach in email
P detected	Body	Detected a picture of logo in email
F SHA-256	Body	SHA-256 of the attach file
F name	Body	File's name
F type	Body	File's type
F type detected	Body	File's type detected from scan result
F size	Body	File's size
F last analyze	Body	File's last analyzed date
F scan result	Body	Result of file scan from dynamic analysis
O file name	Body	Other file's name information

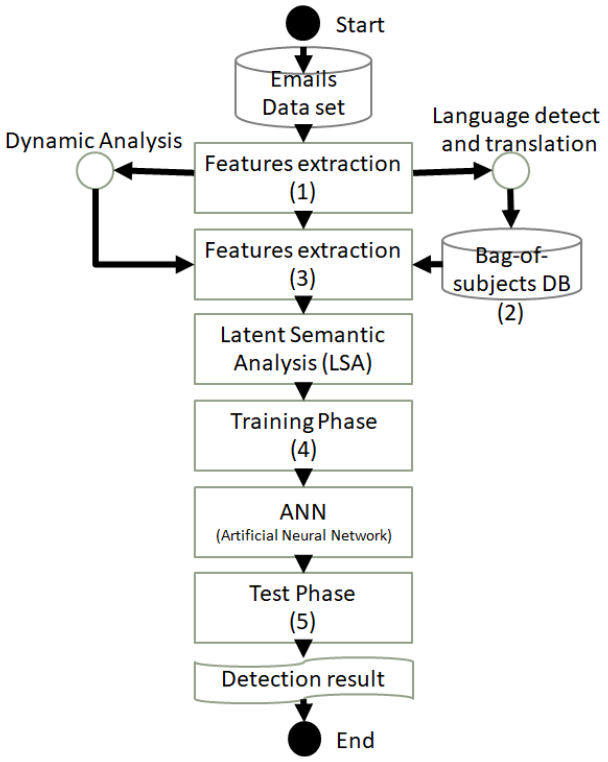

Fig. 5 Procedure of the proposed method.

consists of 5 phases.

4. Experiments

This section our experimentation grouped into 3 subsections: Environment and hardware implementation, Experiments and Evaluation methods.

4.1 Environment and Hardware Implementation

For training on deep-learning, we use Kyushu university's supercomputer ITO-subsystem B [19] which satisfied the resource requirements for the experiment. TensorFlow, CUDA, and Python based APIs are used in software training. **Table 2** shows email dataset that we currently own and use to extract features. Most of the datasets are downloaded from untroubled.org (spam), Enron email dataset (legitimate) and from the Kyushu university zero-day malicious email investigation and analysis lab which contain spam, legitimate, and malicious/phishing spam emails from 2012 to a recent time. We also use our private email dataset (TM) which contain in English, Japanese, Chinese, Lao and other language based subject/title emails. Thus, We keep searching for more email datasets of legitimate email and malicious spam in other language than English to increase capacity and diversity of the database.

4.2 Experiments

After extracting all 27 features from the email dataset, we train and test the system with different email datasets. We have done 4 separate experiments in different number of features in order to check the efficiency of the feature group. First, we used only plain-text words feature which is very popularly used for spam email classification. Second, we used 9 features from header parts including Sender email address, Domain location, Domain time zone, time structure, time zone, time sent, subject/title, language detect and language time zone. Next, we increase the number of features to be 12 which are all features obtained from email header part, Machine translated detect features, risk words detect features, normal and abnormal language detected features are included from the previous experiment case. Finally, we use all 27 features extracted from both the email header and body part which are also included in the features extracted from dynamic analysis results.

In this research, a Multi Layer Perceptron (MLP) artificial neuron network (ANN) algorithm [24] was applied to the extract features. We also use the concept of mutual information [25] to determine the root features in the detection process. The features vector can be defined as a group of features that are used in classification. Here, the mutual information score of each feature F and the candidate features in the training dataset are identified as follows:

$$MI(F) = \sum_{F\in\{0,1\},C\in\{0,1\}} P(F=f, C=c) * \log \frac{P(F=f, C=c)}{P(F=f)(C=c)} \quad (1)$$

where C denotes the class (malspam or normal), $P(F=f, C=c)$ is the probability that the features F occurs ($F=1$) or not ($F=0$) in malspam (C = malspam) or normal (C = normal) emails,

Table 2 Email dataset collection and sources.

Dataset	Amount	Type	Description&Reference
Spam	4,567,714	spam	Spam emails (2012-June 2018) http://untroubled.org/spam/
Zmal	281	Malspam	Malicious & phishing spam (2017-2018) Okamura Lab, Kyushu university. [19]
Enron	517,401	ligitimate	May 7, 2015 version of dataset https://www.cs.cmu.edu/~/enron/
TM	4,251	legitimate	Email dataset (2012-2018)

$P(F = f)$ is the probability that the feature F occurs ($F = 1$) or not ($F = 0$) in all emails, and $P(C = c)$ is the probability that an email is malspam (C = malspam) or normal (C = normal).

A number of features with the highest mutual information scores were selected and we refer to this number as the feature vector size. The highest valued features are most probably the features occurring frequently in one class of normal emails and not so much in the malspam. The algorithms were executed with different feature vector sizes. After the feature vector size and the root features that form the feature vector are identified, it is necessary to determine the range of values that each element of the vector can take. Each vector element corresponds to an input node in the ANN and the value of that element corresponds to the value of the node.

4.3 Evaluation Results

This subsection explains how we calculate Precision, Recall and Accuracy in each part as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + TN} \tag{3}$$

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN} \tag{4}$$

While FP and TN are important values for detection results, FP has a strong relation with Precision and TN with Recall. Next section will show the experiment result with Precision, Recall and Accuracy rate using ANN models.

5. Result and Discussion

In our experiment, to verify an accuracy rate of our method of zero-day malspam detection, we use the extracted 27 features and use the testing datasets different from the training datasets with TensorFlow. **Table 3** shows that the results of both spam and normal email type detection are very similar with 61.71%, 67.06%, 77.11% and 78.12% accuracy in order from 1, 9, 12 and 27 features experiments even using the dataset that have never been trained to the system before.

Currently our experiment cannot show which one is the most effective features to detect zero-day malspam because all features are related to each other and we need to give a set of features to the system. For example, we cannot adjudge that all emails which sent from African's IP address are malspam. What we should do is to check the relationship of the set of features include domain location, domain time zone, email time sent, language detect, language time zone, M-W-N detected features before adjudge. So we believe that the most effective way to detect zero-day malspam is to use a set of features which contain some information from several features that related to each other.

Table 3 Zero-day malspam detection results.

Features Used	Data type	Precision	Recall	TP	FP	Accuracy
1	Spam	0.6152	0.6254	0.6254	0.3746	0.6171
	Normal	0.6191	0.6089	0.6089	0.3911	
9	Spam	0.6648	0.6881	0.6881	0.3118	0.6706
	Normal	0.6768	0.6768	0.6530	0.6530	
12	Spam	0.7785	0.7758	0.7758	0.2242	0.7711
	Normal	0.7736	0.7664	0.7664	0.2336	
27	Spam	0.7782	0.7866	0.7866	0.2133	0.7812
	Normal	0.7843	0.7758	0.7758	0.2241	

From the result, the accuracy rate is significantly increased by increasing a number of features. However, in case of the 27 features, the result only increases 1% from the 12 features experiment. The reason is that the spam email dataset we used contains only normal spam, not malicious or phishing spam. Because normal spam email does not contain malicious file or link, so the system cannot receive any information from the dynamic analysis from email body, and this explains why 14 features from email body are not so efficient. We believe that if the trained dataset also contains enough malspam, the result from 27 features experiment will be improved. We also found that, to extract some features, such as machine translated detection and risk words detection features, the email title database is required to have as much translated text words in as many languages as possible. In addition, the effectiveness of the translation tool is also important. Currently, we are using free Yandex-translation API because of budget issues. The database we offer only supports 4 languages included English, Japanese, Chinese and Lao using Yandex-Translate API. By randomly checking the translation results, we found that some subject/titles are still not translated correctly, compared to the popular commercial translation service such as Google translator, so it is also possible that this issue could be affecting the overall result.

There are some miss-detection cases of normal email that adjudge to be a malspam. Those email are the email that have emoticon or some special symbol in title that our system cannot extract feature information correctly. Those group of email that contain an archived a file locked with a password are also get the result depends on only features from the email header part. Our proposed method still cannot detect some zero-day malspam especially in spear phishing email. Normally a zero-day malspam is generated from virus infected victim's devices or created by updating from its previous version, or both. This means that combinations of these versions might be changed but at least it is still using some part from previous version such as it uses a same story but change a sender address and malicious file type. In some patterns, zero-day malspam change email story, sender address, file name/type but the hash value of the attach file still be the same and already discovered or can be detected by using dynamic analysis. However, the system cannot detect zero-day malspam in case of spear phishing technique which is good design for both the information from email header part and also new composition (a new malicious file or new phishing URL link) that cannot detect by using dynamic analysis.

In **Table 4** we compare our result with some related research.

Table 4 Results comparison.

	Spam filter Trained dataset			Zero-day malspam filter (Untrained)
Approach	Precision	Recall	Accuracy	Accuracy
One feature	0.8961	0.8977	0.8977	0.6171
Sheu	0.9667	0.9630	0.9650	-
He et al.	0.9350	0.9230	0.9670	0.6300
Al-Jarrah et al.	0.9890	0.9920	0.9850	-
Our approach	0.9127	0.9286	0.9286	0.7812

The spam filter results by using a same source of training and testing dataset. Our approach resulted in 92.86% accuracy while others such as Al-Jarrah et al., He et al. and Sheu resulted in a higher accuracy rate. However, those approaches are not confirmed in cases of a testing dataset that are different from training. While our approach got the best accuracy rate at 78.12%, we can say that we have successfully improved a method to detect a zero-day malspam. Thus, we believe that our method still has a lot of things to improve and evaluate to get better results in the future.

6. Conclusion

In this paper we propose a method by using new features extracted from email and deep-learning approach to detect zero-day malspam. We have successfully extracted 27 features from email's header and body part, included machine translation detected, risk words detected and other features by using several APIs. We also use 4 different languages email dataset for more diversity and realistic purpose to build a words database and create features. Our experiment results show the accuracy rate of a zero-day malspam detection is about 78% and 92.8% for normal spam. Thus, we believe that the system still can be improved by adding more malicious spam datasets to train the system, as well as using a translation API having better accuracy.

Acknowledgments This research was supported by Management Expenses Grants of Cybersecurity Center, Kyushu University and Strategic International Research Cooperative Program, Japan Science and Technology Agency (JST) and JSPS KAKENHI Grant Number JP16K00480.

References

[1] Ciscio 2018 Annual Cybersecurity Report, available from ⟨https://www.cisco.com/c/m/en_au/products/security/offers/cybersecurity-reports.html⟩ (accessed 2019-06-14).
[2] The Center for Internet Security (CIS): Top 10 malware October 2017, available from ⟨https://www.cisecurity.org/blog/top-10-malware-of-october-2017/⟩ (accessed 2019-06-14).
[3] Symantec: The Internet Security Threats Report, Volume 23 (ISTR23) (2018), available from ⟨https://www.symantec.com/content/dam/symantec/docs/reports/istr-23-2018-en.pdf/⟩ (accessed 2019-06-14).
[4] Phomkeona, S., Edwards, K., Ban, Y. and Okamura, K.: Zero-day Malicious Email Behavior Investigation and Analysis, *Asia Pasific Advance Network Workshop* (*APAN44 Dalian*) (Aug. 2017).
[5] Phomkeona, S., Kono, K. and Okamura, K.: An Unknown Malware Detection Using Execution Registry Access, *The 42nd IEEE International Conference on Computers, Software & Appications* (*COMPSAC2018*) (July 2018).
[6] Phomkeona, S. and Okamura, K.: The design of an active method for spyware detection, *The 12th International Conference on Future Internet Technologies 2017*, Kyushu University (2017).
[7] Al-Jarrah, O., Khaterz, I. and Al-Duwairi, B.: Identifying Potentially Useful Email Header Features for Email Spam Filtering, *ICDS 2012: The 6th International Conference on Digital Society* (2012).
[8] Mao, J. et al.: Detecting Phishing Websites via Aggregation Analysis of Page Layouts, *2018 International Conference on Identification, Information and Knowledge in the Internet of Things* (2018).
[9] Moghimi, M. and Varjani, A.Y.: New rule-based phishing detection method, *Expert Systems with Applications*, Vol.53, pp.231–242, Elsevier Journal (2016).
[10] Sonowal, G. and Kuppusamy, K.S.: PhiDMA – A phishing detection model with multi-filter approach, *2017 Journal of King Saud University – Computer and Information Sciences* (2017).
[11] Basnet, R.B. et al.: Feature selection for improved phishing detection, *Advanced Re-search in Applied Articial Intelligence*, pp.252–261, Springer (2012).
[12] Ajaz, S., Nafis, M.T. and Sharma, V.: Spam Mail Detection Using Hybrid Secure Hash Based Naive Classifier, *The International Journal of Advanced Research in Computer Science*, Research Paper, Vol.8, No.5, ISSN No.0976-5697 (2017).
[13] Sheu, J.-J.: An Efficient Two-phase Spam Filtering Method Based on E-mails Categorization, *International Journal of Network Security*, Vol.9, pp.34–43 (July 2009).
[14] Wu, C.-H.: Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks, *Expert Systems with Applications*, Vol.36, pp.4321–4330 (Apr. 2009).
[15] Ye, M. et al.: A Spam Discrimination Based on Mail Header Feature and SVM, Proc. Wireless Communications, Networking and Mobile Computing, *WiCOM'08, 4th International Conference on Dalian* (Oct. 2008).
[16] Wang, C.-C. and Chena, S.-Y.: Using header session messages to anti-spamming, *Computers & Security*, Vol.26, pp.381–390 (Jan. 2007).
[17] Hu, Y. et al.: A scalable intelligent non-content-based spam-filtering framework, *Expert Syst. Appl.*, Vol.37, pp.8557–8565 (2010).
[18] Aski, A.S. and Sourati, N.K.: Proposed efficient algorithm to filter spam using machine learning techniques, *Pacific Science Review A: Natural Science and Engineering*, Vol.18, No.2, pp.145–149 (July 2016).
[19] R.I.I.T. of Kyushu University Supercomputer system ITO, available from ⟨https://www.cc.kyushu-u.ac.jp/scp/system/ITO/01_intro.html⟩ (accessed 2019-06-14)
[20] Shi, W. and Xie, M.: A Reputation-based Collaborative Approach for Spam Filtering, *2013 AASRI Confernece on Parallel and Distributed Computing Systems* (2013).
[21] Sorour, S.E., Goda, K. and Mine, T.: Evaluation of Effectiveness of Time-Series Comments Using Machine Learning Techniques, *Journal of Information Processing*, Vol.23, No.6, pp.784–794 (Nov. 2015).
[22] Phomkeona, S. and Okamura, K.: A Design Method for Zero-day Malicious Email Detection Using Email Header Information Analysis (EHIA) and Deep-Learning Approach, *Computer Security Symposium 2018* (*CSS2018*).
[23] Phomkeona, S. and Okamura, K.: Collecting useful features for zero-day malicious emails detection, *The 81st National Convention of IPSJ*.
[24] Bishop, C.: *Neural Networks for Pattern Recognition*, Oxford University (1995).
[25] Androutsopoulos, I. and Koutsias, J.: An Evaluation of Naive Bayesian Networks, Potamias, G., Moustakis, V., van Someren, M. (Eds.), *Machine Learning in the New Information Age*, pp.9–17 (2000).
[26] Malicious spam and phishing email dataset, available from ⟨https://drive.google.com/file/d/1_iJVTRbS64F3qDgO4Sg5APaP6593_kST/⟩ (accessed 2019-09-30).
[27] Ozgur, L., Gungor, T. and Gurgen, F.: Spam Mail Detection Using Artificial Neural Network and Bayesian Filter, Lecture Notes in Computer Science (Aug. 2004).

Appendix

A.1 Risk Words List

A.1.1 List of Risk Words in English

Here is the risk word list (English version) contains 455 key words. For more 3 other languages (Japanese, Chinese and Lao version please download from our dataset [26].

The risk words are: {100% #1 $$$ 100% free 100% Satisfied 4U 50% off Accept credit cards Acceptance Access Accordingly Act Now Action Ad Additional income Addresses on CD Affordable All natural All new Amazed Amazing Amazing stuff Apply now Apply Online As seen on Auto email removal Avoid Avoid bankruptcy Bargain Be amazed Be your own boss Being a member Beneficiary Best price Beverage Big bucks Bill

1618 Billing Billing address Billion Billion dollars Bonus Boss Brand new pager Bulk email Buy Buy direct Buying judgments Cable converter Call Call free Call now Calling creditors Can't live without Cancel Cancel at any time Cannot be combined with any other offer Cards accepted Cash Cash bonus Cashcashcash Casino Celebrity Cell phone cancer scam Cents on the dollar Certified Chance Cheap Check Check or money order Claims Claims not to be selling anything Claims to be in accordance with some spam law Claims to be legal Clearance Click Click below Click here Click to remove Collect Collect child support Compare Compare rates Compete for your business Confidentially on all orders Congratulations Consolidate debt and credit Consolidate your debt Copy accurately Copy DVDs Costs Credit Credit bureaus Credit card offers Cures Cures baldness Deal Dear [email/friend/somebody] Debt Diagnostics Dig up dirt on friends Direct email Direct marketing Discount Do it today Don't delete Don't hesitate Dormant Double your Double your cash Double your income Drastically reduced Earn Earn $ Earn extra cash Earn per week Easy terms Eliminate bad credit Eliminate debt Email harvest Email marketing Exclusive deal Expect to earn Expire Explode your business Extra Extra cash Extra income F r e e Fantastic Fantastic deal Fast cash Fast Viagra delivery Financial freedom Financially independent For free For instant access For just $ (some amount) For just $xxx For Only For you Form Free Free access Free cell phone Free consultation Free DVD Free gift Free grant money Free hosting Free info Free installation Free Instant Free investment Free leads Free membership Free money Free offer Free preview Free priority mail Free quote Free sample Free trial Free website Freedom Friend Full refund Get Get it now Get out of debt Get paid Get started now Gift certificate Give it away Giving away Great Great offer Guarantee Guaranteed Have you been turned down? Hello Here Hidden Hidden assets Hidden charges Home Home based Home employment Home based business Human growth hormone If only it were that easy Important information regarding In accordance with laws Income Income from home Increase sales Increase traffic Increase your sales Incredible deal Info you requested Information you requested Instant Insurance Insurance Internet market Internet marketing Investment Investment decision It's effective Join millions Join millions of Americans Junk Laser printer Leave Legal Life Life Insurance Lifetime Limited limited time Limited time offer Limited time only Loan Long distance phone offer Lose Lose weight Lose weight spam Lower interest rates Lower monthly payment Lower your mortgage rate Lowest insurance rates Lowest Price Luxury Luxury car Mail in order form Maintained Make $ Make money Marketing Marketing solutions Mass email Medicine Medium Meet singles Member Member stuff Message contains Message contains disclaimer Million Million dollars Miracle MLM Money Money back Money making Month trial offer More Internet Traffic Mortgage Mortgage rates Multi-level marketing Name brand Never New customers only New domain extensions Nigerian No age restrictions No catch No claim forms No cost No credit check No disappointment No experience No fees No gimmick No hidden No hidden Costs No interests No inventory No investment No medical exams No middleman No obligation No purchase necessary No questions asked No selling No strings attached No-obligation Not intended Not junk Not spam Now Now only Obligation Offshore Offer Offer expires Once in lifetime One hundred percent free One hundred percent guaranteed One time One time mailing Online biz opportunity Online degree Online marketing Online pharmacy Only Only $ Open Opportunity Opt in Order Order now Order shipped by Order status Order today Outstanding values Passwords Pennies a day Per day Per week Performance Phone Please read Potential earnings Pre-approved Presently Price Print form signature Print out and fax Priority mail Prize Problem Produced and sent out Profits Promise Promise you Purchase Pure Profits Quote Rates Real thing Refinance Refinance home Refund Removal Removal instructions Remove Removes wrinkles Request Requires initial investment Reserves the right Reverses Reverses aging Risk free Rolex Round the world S 1618 Safeguard notice Sale Sample Satisfaction Satisfaction guaranteed Save $ Save big money Save up to Score Score with babes Search engine listings Search engines Section 301 See for yourself Sent in compliance Serious Serious cash Serious only Shopper Shopping spree Sign up free today Social security number Solution Spam Special promotion Stainless steel Stock alert Stock disclaimer statement Stock pick Stop Stop snoring Strong buy Stuff on sale Subject to cash Subject to credit Subscribe Success Supplies Supplies are limited Take action Take action now Talks about hidden charges Talks about prizes Teen Tells you it's an ad Terms Terms and conditions The best rates The following form They keep your money—no refund! They're just giving it away This isn't a scam This isn't junk This isn't spam This won't last Thousands Time limited Traffic Trial Undisclosed recipient University diplomas Unlimited Unsecured credit Unsecured debt Unsolicited Unsubscribe Urgent US dollars Vacation Vacation offers Valium Viagra Vicodin Visit our website Wants credit card Warranty We hate spam We honor all Web traffic Weekend getaway Weight Weight loss What are you waiting for? What's keeping you? While supplies last While you sleep Who really wins? Why pay more? Wife Will not believe your eyes Win Winner Winning Won Work from home Xanax You are a winner! You have been selected Your income}

Sanouphab Phomkeona was born in 1986. He received his master degree from Toyohashi University of Technology, Japan in 2011. Since 2012, he has been working as a lecturer at the Department of Computer Engineering and Information Technology, Faculty of Engineering, National University of Laos. Currently, he is a Ph.D. student of Kyushu University. His research interest is a information security, malware analysis and developing national cybersecurity status.

Koji Okamura received his B.S., M.S. and Ph.D. from Kyushu University in 1988, 1990 and 1998, respectively. He became an associate professor of the Computer Center and Graduate School of Information Science and Electrical Engineering in 1998 and a professor at Kyushu University in 2011. He serves as the director of the Cybersecurity Center at Kyushu University and vice director of the Research Institute for Information Technology, and vice CISO of Kyushu University. He is a member of IPSJ, IEICE, IEEE-CS.