Malay SMS Spam Detection Framework using Naïve Bayes Technique

Cik Feresa Mohd Foozy, Shamala Palaniappan, Sofia Najwa Ramli, Chuah Chai Wen, MF Abdollah, Rabiah Ahmad

Abstract: Short Message Service (SMS) Spam is one form of mobile device attack that can affect mobile user's security and privacy. This is because such attack applies social engineering method to trick the user for information gathering. This study proposed an SMS Spam detection framework specifically for Malay language by using Naïve Bayes. There are several solutions to detect SMS Spam, but machine learning is one of the most effective technique to detect spam attack. In addition, the existing detection framework using machine learning technique is not effective for Malay language SMS. This is because the features used are not based on Malay language to detect the SMS content as Spam or not Spam. This framework consists of several processes such as Data Collection, Pre-processing, three types of Features Selection, Classification and Detection. Based on the result, it shows that the classification derives acceptable accuracy which is over 90%.

Index Terms: Attack, Detection, Naïve Bayes, Spam

I. INTRODUCTION

Short Message Services (SMS) is one of the alternatives as communication medium by mobile phone. However, SMS can be used to fraud users[1]. There are anti-Spam available to be installed on mobile devices for protection, but it is still lacking to detect SMS Spam in Malay language. Malay language is the main language in Malaysia and it is used in formal and informal communication throughout Malaysia.

The SMS Spam operates by sending SMS to users randomly. The message contains unwanted content such as business promotion or web link. Usually, each SMS sent is charged to the user although unsolicited, and the user need to prompt a reply to stop the SMS. Even the charge to stop the SMS is borne by the user.

Revised Manuscript Received on May 22, 2019.

Cik Feresa Mohd Foozy, Applied Computing Technology (ACT), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor, Malaysia

Shamala Palaniappan, Applied Computing Technology (ACT), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor, Malaysia

Sofia Najwa Ramli, Applied Computing Technology (ACT), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor, Malaysia

Chuah Chai Wen, Information Security Interest Group (ISIG), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor, Malaysia

MF Abdollah, Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal,Melaka, Malaysia

Rabiah Ahmad, Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal,Melaka, Malaysia

This indicates that the security and the privacy of the user's mobile device has been violated. There are several detection techniques that have been applied in SMS Spam detection studies such as Content-Based Filtering, Whitelist or Blacklist, Machine Learning, Matching Pattern and Artificial Immune System.

This paper developed a detection framework for SMS Spam for Malay language which consists of Data Collection for Malay language SMS Spam, Pre-processing, Features Selection based on Malay language, Classification and Detection. In this paper, several experiments have been done to analyze the proposed technique and framework. Moreover, the results in each process of the framework had been validated through Naive Bayes classification technique. By developing this framework, it will help to provide a Malay language SMS Spam feature for future work in spam detection since the existing studies are only focused in English SMS Spam features and less on the Malay language SMS features.

II. LITERATURE REVIEW

Spam and Phishing threats can attack mobile devices via SMS. SMS is one of the most reasonable communication service in terms of cost in most countries. However, because of this benefit, SMS is widely used for marketing advertisement and abused by some irresponsible parties to gain profit from users. Unfortunately, the effectiveness of SMS also increases information security risk. Several studies in SMS Spam is discussed in this section such as SMS Spam, SMS Spam Detection Framework, Dataset, Pre-processing, Classification and Detection.

A. SMS Spam

Spam message is an attack that sends unwanted marketing message[2]. Example of SMS Spam message is *"FREE camera phones with linerental from 4.49/month with 750 cross ntwk mins. 1/2 price txt bundle deals also avble. Call 08001950382 or call2optout/J MF".*

The number of SMS Spam had increased because of the increase in SMS usage[3]. Shirani-Mehr[4] claimed that in year 2012, 30% of SMS sent in Asia were spam.

Cloudmark[5] reported high volume of SMS from three (3) Washington DC phone numbers to many phone numbers in Malaysia which had sent an average of 3000 SMS per day. This message was sent in Malay language as shown as Figure 1.

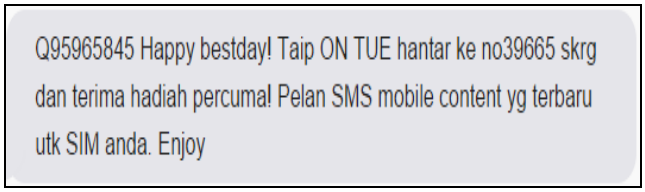

Fig. 1 Example SMS Spam in Malay Language[5]

According to Figure 1, when the user gets a message from an unknown contact, it is hard to identify if the message is a real advertisement SMS or a fraud SMS. Thus, it is significant to study the text behavior of SMS Spam. Since no study was done in Malay language, SMS Spam corpus in Malay language had been developed to evaluate the result accuracy after classification and detection of SMS Spam using Naïve Bayes technique.

B. SMS Spam Characteristics

To recognize attack detection using machine learning technique, it needs to identify the attack characteristics. The characteristic of SMS Spam will contain marketing campaign[6],[7],[8],[9],[10],[11] and [12]. In addition to that, another feature of SMS Spam is to spread news[10].

C. Existing study on SMS Spam Framework

11 studies of SMS Spam frameworks were done for review. A number of techniques were introduced such as content based techniques, blacklist, whitelist, machine learning, text classification, artificial intelligence tools and application. These techniques had also been applied in the various studies as stated in Table 2 from 2013 to 2018.

Table. 2 Studies in SMS Spam Detection from 2013-2018

No.	Author	SMS Filtering/Detection Process
1.	Tiago Almeida et al.[14]	1. Tokenizers 2. Classifiers and Baseline 3. Protocol 4. Results.
2.	Androulidakis et al. [15]	1. Incoming SMS 2. Analyze Message 3. Identified SMSC (Short Message Service Centre) fulfil with originator's number for each received SMS, Identified illegal character 4. Identified time zone contact 5. Identified blacklist words 6. SMS Detection.
3.	Charninda et al. [16]	1. Sender Identification Module 2. SMS content Extractor 3. Tokenizer 4. Filter 5. Categorization 6. Training 7. Detection
4.	Najadat et al. [7]	1. Download SMS 2. Separate SMS into Ham and Spam manually 3. Pre-processing 4. Classification 5. Testing 6. Evaluation.
5.	Shirani-Mehr [4]	1. Preprocessing 2. Feature Extraction 3. Machine Learning Technique 4. Result.
6.	Karami and Zhou [17]	1. Feature Extraction 2. Classification.
7.	Ahmed et al.[18]	1. SMS Collection 2. Preprocessing 3. Feature Selection 4. Vector Creation 5. Filtering Process 6. Updating the System.
8.	Pham et al.[19]	1. Preprocessing 2. Features Selection 3. Classification
9.	Ezpeleta et al.[20]	1. Spam filtering 2. Sentiment Analysis
10.	S. Ali et al.[21]	1. Pre-processing 2. Classification
11.	P. Navaney et al.[22]	1. Classification using various supervised machine learning algorithms

According to Delany et al.[13], no best technique exists to fix the detection of SMS Spam attack. However, in this paper, the framework applies several processes from the existing SMS Spam detection framework such as Data Collection,Pre-processing, Features Selection, Classification and Detection.

III. METHODOLOGY

The detection framework consists of SMS Data Collection, Pre-processing such as Tokenization and Lemmatization, Feature Selection, Classification and detection. The majority spam detection caters for SMS Spam in English language. As a matter of fact, SMS Spam in Malay language had increased by 100% in the past 3 years. Consequently, the fact helped motivate the study to detect SMS Spam in Malay language.

A. Malay language SMS Spam Data Collection

According to Almeida et al.[23], public and real datasets for this research area was still lacking. Therefore, there was a need to prepare data SMS Spam for this study.

A review on text dataset (corpus) was done as a guidance on how the SMS datasets are collected. From the finding, this research is the first study to collect new SMS in Malay language.

SMS Spam in Malay language, was collected from respondents and the Internet. There were 210 SMS collected and after the redundancy process, there were 178 SMS that would apply for this study. The collection consisted of 49 SMS Ham and 129 SMS Spam.

Two steps of validation process for Malay language SMS corpus was done. The first was using the Naive Bayes classification technique and the second was validated by seven (7) expert persons in Information Security.

There are two types of features applied in this study. The features are Generic Features and Payload Features.

For this purpose, 160 SMS were collected from respondents. However, after the redundancy process, only 178 SMS were used for the experiment which were 49 SMS Ham and 129 SMS spam.

Respondents were asked to provide any SMS that was suspected to contain fraudulent content or marketing SMS from ‘unknown’ senders. Table 3 shows examples of the Malay SMS Spam that had been collected.

Table. 3 Example of the SMS Spam that has been collected

No.	SMS Spam
1.	RM0.00 Mumia berkepala 3 mumia? kaum makan manusia masih di dunia? PERCUMA 7 hari Majalah Misteri di sini! Htr ON MISTERI ke 27272.2 SMS/mgg
2.	RM0.00 Pelanggan sekalian: Dptkn Servis SMS content terbaru yg tlh kami kemaskini pd bulan julai. Hantar ON CK ke 377744 segera Dftr free EMSB 80625114
3.	RM0.00 Bonus: Balas ON HT ke 33311 utk enjoy Free mobile SMS content!
4.	RM0.00 Peringatan Akhir Kepada 019XXXXXXX diminta terima servis SMS mobile content Baru 2012. Taip ON PA ke-36333 untuk daftar Servis Tanpa Caj.0380600842
5.	RM0.0 Promosi Tahniah sim kad Anda tlh meraih HADIAH PERCUMA RM 19,000.00 dr …. No reg 552299 Sila call di Talian 0178320403 TQ
6.	“RM0.00: 01XXXXXXXXX Happy bestday! Taip ON THU dan hantar ke no 3XXXX skrg dan terima hadiah percuma! Pelan SMS mobile content yg terbaru utk phone anda. Enjoy
7.	RM0, Dptkan komik yg lucu, mereng & kerek terus ke telefon anda! Boleh dibaca & dikongsi bila2 saja! Pasti kelakar habis! Hanya RM0.50/MMS. Htr ON KOMIK ke 23238
8.	RM0. Pelanggan Dihormati, Sila Aktifkan GPRS anda utk muat turun koleksi MP3/MPG/MOV/MMS Melayu & English terkini. pdftaran FREE dgn sms ON MMS ke 32121 sekarang!
9.	RM0. Ingin mengetahui segalanya tentang artis Hollywood kegemaraan anda? RM0.30/SMS. Hantar ON HOLLY ke 23288 sekarang
10.	Peringatan pengguna: Servis wireless&mms anda telah dibatalkan. Sila htr ON WWQ dan ON WWP ke 32770 utk dptkan SMS content skrg!

B. Pre-processing

In machine learning, data Pre-processing is important to ensure the data is free from unimportant features but in this experiments, all words were included to identify the highest frequency spam word that is used in Malay SMS Spam. For this framework, the pre-processing consists of tokenization and lemmatization. Tokenization splits the sentence into individual words and the process of lemmatization is to group similar words. The next process is for the clean data to be applied for features selection process.

C. Features Selection

There are two types of features that had been introduced in this experiment. Firstly, the Generic Features which was proposed based on the literature study. The Generic Features are Total words, Number of Character bi-grams, Number of Character tri-grams, Average number of length, and Average number of word. For this study, additional features based on the Malay language Spam characteristics was also included such as Advertisement or announcement, Contest, Malicious URL, Telephone

Number, Winning, SMS asking for help to get money and SMS asking to respond or subscribing to services.

After features selection is applied, the number of features is ranked according to the significance of the features.

Secondly, the Payload Features is based on the SMS content. The Payload Features or Bag of Words is a list of frequency of words that had been applied in the document classification study. For this study, the list of words from the Malay language SMS content were extracted using JAVA Programming and were applied to the Malay language SMS class classification. Since this is the first study for detecting SMS Spam in Malay language, a set of bag of words for Malay SMS Spam and Phishing had been introduced for this study.

D. Classification

According to previous studies, the result shows that Naive Bayes is commonly used for SMS classification technique. The reason Naive Bayes was used is because this technique is one of the most simple probability technique which can predict the classification for better results[24]. Thus in this experiment, Naïve Bayes was applied to classify the dataset into Ham and Spam classes.

E. Detection

After classification using the training data, the model detection would test other data to judge the accuracy. The high detection accuracy result for testing data would evaluate the SMS Spam attack detection framework as acceptable or not for the datasets used.

IV. RESULT

This section discusses the result of the Pre-processing phase of this research. In this phase, the Pre-processing consisted of collecting the Malay language SMS Spam, Data Pre-processing, Features Selection, Classification using Naive Bayes technique.

A. Malay language SMS Spam Data Collection

Table 4 is information on the Malay language SMS that have been collected. The SMS had also been validated by several experts in spam message detection in Malaysia.

Table. 4 Malay language SMS Corpus Specification

Malay SMS Corpus Specification	Malay SMS
No of messages	210
No of words	26122
Average no. of words per message	124.4
Collection methods and procedures	Website and respondents

B. Malay language SMS Spam Features Selection

The Malay language SMS Spam that was collected had gone through the Pre-processing step such as tokenization. After that, the features selection algorithm was applied to the dataset. This is to identify the highest occurrence of word in the Malay language SMS dataset. The highest occurrence will be selected as the payload features. The result of the algorithm is shown in Table 5.

Table 5 shows several words that have been sorted based on the word frequency in SMS Spam. Based on these words, the highest and lowest frequency of words in Malay language SMS Spam is shown.

For Generic Features are Total words, Number of Character bi-grams, Number of Character tri-grams, Average number of length, and Average number of word.

Table. 5 Example of Features Frequency in Malay language SMS

Features	Occurrences	Features	Occurrences
anda	59	caj	14
sila	45	talian	14
on	37	taip	13
ke	36	kasih	13
sms	32	memenangi	13
utk	24	hubungi	12
hadiah	22	yg	12
tahniah	21	tunai	11
terima	20	sim	11
no	19	mobile	11
shell	19	telah	11

Based on the Payload Features SMS Spam in Table 5, there are some words with meaning. However, all features of SMS content was applied for features selection process because each word in the SMS contributes to the detection of SMS Spam. This result was then applied in the Classification phase.

A. Malay language SMS Spam Classification

There are three types of features examined in the Classification phase. The features are Generic Features, Payload features and both features. The purpose of these experiments is to gauge the detection accuracy. The result of the experiments are shown at Table 6.

Table. 6 Malay language SMS Spam Classification Technique using Naïve Bayes

No.	Experiments Naive Bayes	Accuracy %
1	Generic Features	98.3146 %
2.	Payload Features	93.8202 %
3.	Generic & Payload Features	99.4382 %

According to Table 6, the experiments have been done for class Ham and Spam. The experiments show results for both features is higher than other types of feature. As a preliminary study for the Malay language SMS Spam, it can be further studied in other machine learning techniques and also applied as an enhancement feature for the Malay language SMS Spam Detection.

V. CONCLUSION

The existing SMS attack detection framework can only detect specific features attack. By detecting SMS Spam in the Malay language, spam features in Malay language has been introduced which contributes in detecting SMS Spam in Malaysia.

There are five (5) text mining techniques that can be applied to detect these attacks using the proposed framework.

The experiments from data mining tool showed the acceptable result by using Naive Bayes. The basis to the selection of this technique is because it is commonly used by other researchers in SMS Spam attack detection and its availability in existing machine learning tools.

As a conclusion, according to the study that has been done, it shows that it is significant to detect SMS Spam in Malay language using machine learning technique because most studies focus on detecting SMS Spam in English and that creates a limitation in detecting Malay language SMS Spam. The increasing number of SMS Spam on mobile device violates mobile device user's security and privacy. Although there are detection and filtering mechanisms to prevent SMS Spam, it is still lacking for SMS Spam in Malay language and needs more suitable features to detect Malay language SMS Spam attack.

ACKNOWLEDGMENT

This research is supported by the Ministry of Education Malaysia under the Fundamental Research Grant Scheme (FRGS) Vot K075 and Universiti Tun Hussein Onn Malaysia under TIER 1 Vot H237.

REFERENCES

1. S. Chhabra, "Fighting spam, phishing and email fraud," UNIVERSITY OF CALIFORNIA RIVERSIDE, 2005.
2. K. Yadav, P. Kumaraguru, A. Goyal, A. Gupta, and V. Naik, "SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering," Proceedings of the 12th Workshop on Mobile Computing Systems and Applications. ACM, Phoenix, Arizona, pp. 1–6, 2011.
3. H. Shirani-Mehr, "SMS Spam Detection using Machine Learning Approach," 2014.
4. Cloudmark, "Annual Security Threat Report 2014," Cloudmark, San Francisco, USA, 2014.
5. E. Vall, #233, and P. Rosso, "Detection of near-duplicate user generated contents: the SMS spam collection," Proceedings of the 3rd international workshop on Search and mining user-generated contents. ACM, Glasgow, Scotland, UK, pp. 27–34, 2011.
6. H. Najadat, N. Abdulla, R. Abooraig, and S. Nawasrah, "Mobile SMS Spam Filtering based on Mixing Classifiers," 2014.
7. K. Yadav, S. K. Saha, P. Kumaraguru, and R. Kumra, "Take Control of Your SMSes: Designing an Usable Spam SMS Filtering System," in Mobile Data Management (MDM), 2012 IEEE 13th International Conference on, 2012, pp. 352–355.
8. T. M. M. and A. M. Mahfouz, "SMS Spam Filtering Technique Based on Artificial Immune System," IJCSI Int. J. Comput. Sci. Issues, vol. 9, no. 2, 2012.
9. Q. Xu, E. Xiang, J. Du, J. Zhong, and Q. Yang, "SMS Spam Detection using Content-less Features," Intell. Syst. IEEE, vol. PP, no. 99, p. 1, 2012.
10. M. Taufiq Nuruzzaman, C. Lee, M. F. A. bin Abdullah, and D. Choi, "Simple SMS spam filtering on independent mobile phone," Secur. Commun. Networks, vol. 5, no. 10, pp. 1209–1220, 2012.
11. M. Z. R. que and M. Farooq, "SMS Spam Detection By Operating On Byte-Level Distributions Using Hidden Markov Models (HMMS)," Virus Bulletin Conference September 2010. 2010.
12. S. J. Delany, M. Buckley, and D. Greene, "SMS spam filtering: Methods and data," Expert Syst. Appl., vol. 39, no. 10, pp. 9899–9908, 2012.
13. Tiago A. Almeida and J. M. G. Hidalgo, "SMS Spam Collection Data Set," 2012. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection.
14. I. Androulidakis, V. Vlachos, and A. Papanikolaou, "FIMESS: filtering mobile external SMS spam," in BCI, 2013, pp. 221–227.
15. T. Charninda, T. T. Dayaratne, H. K. N. Amarasinghe, and J. Jayakody, "Content based hybrid sms spam filtering system," 2014.
16. A. Karami and L. Zhou, "Improving static SMS spam detection by using new content-based features," in 20th Americas Conference on Information Systems, AMCIS 2014, 2014.
17. I. Ahmed, D. Guan, and T. C. Chung, "SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset," Int. J. Mach. Learn. Comput., vol. 4, no. 2, 2014.
18. T. H. Pham and P. Le-Hong, "Content-based approach for Vietnamese spam SMS filtering," in Proceedings of the 2016 International Conference on Asian Language Processing, IALP 2016, 2017.
19. E. Ezpeleta, I. Garitano, U. Zurutuza, and J. M. G. Hidalgo, "Short Messages Spam Filtering Combining Personality Recognition and Sentiment Analysis," Int. J. Uncertainty, Fuzziness Knowledge-Based Syst., 2017.
20. S. S. Ali and J. Maqsood, ."Net library for SMS spam detection using machine learning: A cross platform solution," in Proceedings of 2018 15th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2018, 2018.
21. P. Navaney, G. Dubey, and A. Rana, "SMS Spam Filtering Using Supervised Machine Learning Algorithms," in Proceedings of the 8th International Conference Confluence 2018 on Cloud Computing, Data Science and Engineering, Confluence 2018, 2018.
22. T. A. Almeida et al., "Contributions to the study of SMS spam filtering: new collection and results," Proceedings of the 11th ACM symposium on Document engineering. ACM, Mountain View, California, USA, pp. 259–262, 2011.
23. M. T. Nuruzzaman, L. Changmoo, and C. Deokjai, "Independent and Personal SMS Spam Filtering," in Computer and Information Technology (CIT), 2011 IEEE 11th International Conference on, 2011, pp. 429–435.