INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, e-ISSN 1841-9844, 14(4), 518-529, August 2019.

Efficient Detection of Attacks in SIP Based VoIP Networks using Linear l_1-SVM Classifier

W. Nazih, Y. Hifny, W.S. Elkilani, T. Abdelkader, H.M. Faheem

Waleed Nazih*
1. College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, KSA
2. Faculty of Computers and Information Sciences, Ain Shams University, Egypt
*Corresponding author: w.nazeeh@psau.edu.sa

Yasser Hifny
Faculty of Computers and Information, Helwan University, Egypt

Wail S. Elkilani
1. Faculty of Computers and Information Sciences, Ain Shams University, Egypt
2. College of Applied Computer Science, King Saud University, KSA

Tamer Abdelkader
Faculty of Computers and Information Sciences, Ain Shams University, Egypt

Hossam M. Faheem
Faculty of Computers and Information Sciences, Ain Shams University, Egypt

Abstract: The Session Initiation Protocol (SIP) is one of the most common protocols that are used for signaling function in Voice over IP (VoIP) networks. The SIP protocol is very popular because of its flexibility, simplicity, and easy implementation, so it is a target of many attacks. In this paper, we propose a new system to detect the Denial of Service (DoS) attacks (i.e. malformed message and invite flooding) and Spam over Internet Telephony (SPIT) attack in the SIP based VoIP networks using a linear Support Vector Machine with l_1 regularization (i.e. l_1-SVM) classifier. In our approach, we project the SIP messages into a very high dimensional space using string based n-gram features. Hence, a linear classifier is trained on the top of these features. Our experimental results show that the proposed system detects malformed message, invite flooding, and SPIT attacks with a high accuracy. In addition, the proposed system outperformed other systems significantly in the detection speed.
Keywords: Machine learning, Support Vector Machines (SVMs), Session Initiation Protocol (SIP), VoIP attacks.

1 Introduction

Voice over Internet Protocol (VoIP) is a technology that enables the user to make voice or telephone calls over the Internet Protocol (IP) networks. Since the internet has been, and continues to be a prominent form of communication, the VoIP services are going to be a promising communication medium because of its low cost and added features. VoIP systems have two main functions: signaling function and media transmission function.

The most popular protocols developed for the signaling function are the Session Initiation Protocol (SIP) and $H.323$ protocol. SIP [19] is an application layer protocol to create, modify, and terminate real-time sessions between participants over an IP based network. Although $H.323$ protocol is more powerful [16], SIP is more popular because it is flexible, simple, easy to implement, and is based on ASCII messages (not binary messages as in $H.323$).

SIP is vulnerable to many attacks [33]. DoS attacks (i.e. malformed message and invite flooding) can disrupt the VoIP service partially or totally. In addition, SPIT attacks are usually

used for products advertisement, harassment of subscribers, or convincing VoIP users to dial specific numbers. Recent statistics showed that spam calls result in huge man labor losses, especially for small business enterprises in the U.S. [32].

In this paper, we developed a machine learning (ML) system that decreases VoIP-SIP attacks using a linear SVM classifier. Our contributions are: (*i*) Introducing a novel approach for VoIP-SIP attacks detection using a fast linear SVM classifier; (*ii*) Using l_1 regularizer in the objective function that leads to sparse solutions[1]; (*iii*) Comparing our results with the published state-of-the art systems.

The rest of the paper is organized as follows. The next section introduces the related work, with a focus on the ML approaches. In section 3, we describe the proposed approach to detect VoIP-SIP attacks. Data is explained in section 4. Experimental setup is illustrated in section 5. Finally, section 6 concludes the paper and discusses future work.

2 Related work

Decreasing VoIP attacks is a hot topic of research in the last few years. Hosseinpour et al. [9] used a Finite State Machine (FSM) to extract parameters of the SIP traffic in normal conditions. These parameters are used with fuzzy logic to detect DoS attacks. Tsiatsikas et al. [27] detected DoS attacks which exploit the SIP message body. They built a Session Description Protocol (SDP) parser using 100 rules, which achieved a high accuracy.

Machine learning (ML) is an artificial intelligence approach that creates a model to recognize some patterns based on training examples. The ML task usually consists of three phases: choosing a learning algorithm, training the algorithm using the training dataset, and evaluating the algorithm performance by running it on another dataset (test-dataset). The detection of VoIP-SIP attacks using ML methods is introduced in many research work. Nassar et al. [14] extracted a set of 38 features from a slice of SIP messages, a SVM classifier decides if this vector is anomaly or not and issues an event, the event correlator uses a set of rules and conditions to filter the classifier events and generates alarms when necessary.

Akbar et al. [2] introduced Packet-based SIP Intrusion Protector (PbSIP) to prevent SIP flooding attacks and SPIT. PbSIP contains a packet-based analyzer that uses a set of spatial and temporal features to reduce the required processing and memory, features computation module, and Naive Bayes and J48 classifiers. In addition, Asgharian et al. [3] introduced a set of 18 statistics features calculated from SIP headers. They used a SVM classifier to evaluate the proposed features. Pougajendy et al. [17] used a subset of [3] features plus 2 new features, they evaluated the proposed features using a SVM classifier.

Rieck et al. [18] converted the SIP message to a high-dimensional vector space using n-gram tokens. They measured the Euclidean distance between a new message and a built model to detect anomalous messages. Tang et al. [24] proposed a prevention and detection system of SIP flooding attacks. They integrated a three-dimensional sketch design with the Hellinger Distance (HD) technique.

In [23] Su et al. extracted 23 features to detect SPIT attacks using k-nearest neighbor classifier. They added weight to each feature using a genetic algorithm. Vennila et al. [29] introduced two phases model; a SVM classifier to classify the traffic into VoIP and non-VoIP, and an entropy model to classify the VoIP traffic into flooding and non-flooding. Later, in [30] they proposed another two phases model to detect SPIT callers, which used Markov Chain, and incremental SVM classifier.

[1]By sparse solutions we mean that most of the parameters of the model are zeros.

In [25] Tsiatsikas et al. proposed an offline system to detect Distributed DoS (DDoS) attacks, they calculated the occurrence of 6 mandatory headers of SIP message, and implemented headers anonymization using HMAC. They tried 5 classifiers to find the best false alarm rate. Later, in [26] they proposed a real-time detection system and tried a group of DDoS scenarios.

Akbar et al. [1] used kernel tree analysis instead of features extraction, and a SVM classifier to detect malformed DDos attacks. In [21] Semerci et al. detected DDoS attacks using a change-point method which detects the change of Mahalanobis distance between successive feature vectors. If the change exceeds a threshold, the system labeled this as an attack. Kurt et al. [12] extracted a set of features from SIP messages and server logs. A Hidden Markov Model was used to relate these features to hidden variables, and a Bayesian multiple change model used these variables as change point indicators to detect DDoS flooding attacks. Le et al. [13] built a large data-set using a developed interface over a mobile application, which enables the user to label the malicious calls. They started with 29 features and reduced them to 10 features. Many machine learning models were tried (i.e. SVM and neural networks).

We observed a few issues in the developed systems described above. The feature extraction methods based on hand-crafted features are not generic, and tuned for specific datasets and attacks [3, 14, 23]. Hence, there is a need to develop a generic feature extraction method that is suitable for many attacks. Besides, the classification approaches based on distance measures and static threshold are not immune to noisy datasets [9, 29]. Moreover, the dual SVM methods are known to be slow due to the kernel calculations [1, 17]. Furthermore, low detection accuracy in general [21] or in case of low-rate attack [24] were observed. Lastly, rules-based systems require deep knowledge of SIP and numerous manual work [27].

These drawbacks led to the need for a system to detect VoIP-SIP attacks with high detection accuracy and a little processing time. To achieve this, we used the n-gram technique to extract features from SIP messages, and linear l_1-SVM to classify these messages into normal or attack.

3 Proposed approach

Detection of SIP attacks is formulated based on a ML approach. It consists of two steps. The first step is to project the messages into a high-dimensional space since they are more likely to be linearly separable than low-dimensional space [5], as illustrated in Figure 1. A method based on extracting n-gram tokens from a SIP message is used to generate the high-dimensional space. The second step is to use a linear SVM classifier with l_1 regularization to detect the SIP attacks. This classification algorithm optimizes the primal soft-margin objective function, and it is much faster than optimizing the dual objective functions with kernels that were used in the previous research [3, 14, 17].

3.1 Features extraction

In order to classify the SIP messages into normal or attack, the SIP messages are converted into numerical feature vectors. The features can be based on heuristics and domain knowledge as in [3, 23]. The disadvantage of this approach is that the generated features do not capture the diversity of the SIP messages, and they are highly tuned for specific attacks. Alternatively, they can be generated using a generic mathematical method like n-gram tokens as in [18]. The n-gram methods are widely used in speech and language processing [11] and in the network intrusion detection [31].

The SIP message is converted to a feature vector by moving a window of length n over the message and extracting all sub-strings (n-gram tokens). The length of the feature vector equals the number of unique n-grams in the training set. For each n-gram, we compute the number

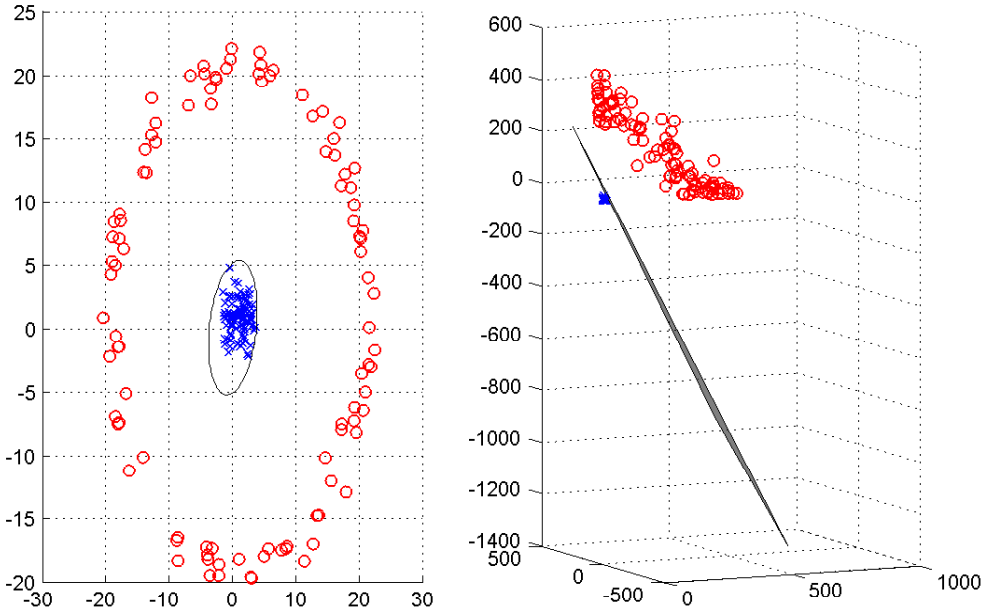

Figure 1: Moving to three-dimensional space, a nonlinear decision boundary for a two-dimensional classification problem becomes linear

of its occurrences in the message and use it to set its value in the feature vector. Figure 2 summarizes the features extraction process (n=4).

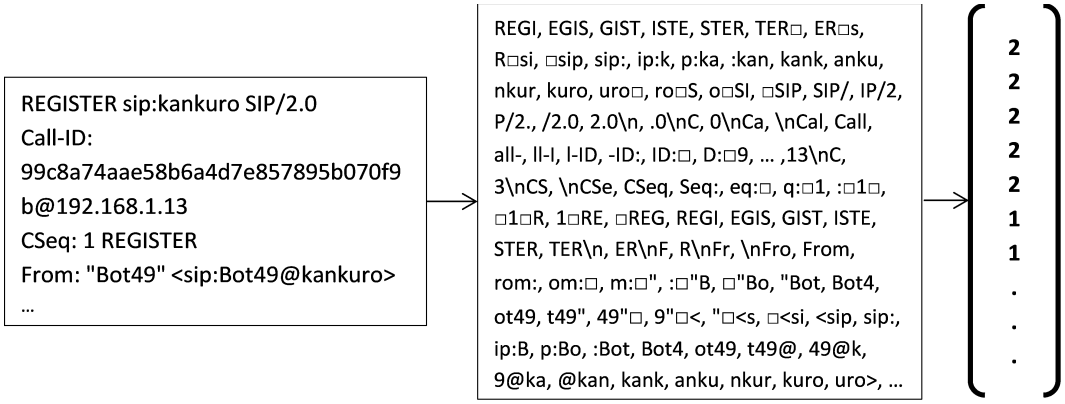

Figure 2: A part of a SIP message, its n-grams, and the occurrences of n-grams in the message

Although the extraction of features based on n-gram tokens provides a generic and an effective way for representing the SIP message, the length of the resulting feature vector is very large, which slows down the detection speed in the classification process. To overcome this problem, we set a cutoff hyper-parameter, and add to the feature vector the n-grams that exceed the cutoff.

3.2 Linear l_1-SVM classifier

Given a training set D that has m examples:

$$D = [(x_1, y_1), \ldots, (x_m, y_m)], \tag{1}$$

where y_i are either 1 or -1, each indicating the class to which the point x_i belongs (i.e. normal or attack). Each x_i is a d-dimensional real vector (i.e. the unique numbers of $n-$gram tokens in the training set). The soft-margin SVM classifier is computed by minimizing the *primal* objective function given by:

$$J = \min_w \left[\frac{1}{m} \sum_{i=1}^{m} \max\left(0, 1 - y_i(w \cdot x_i - b)\right) \right] + \lambda \|w\|^2, \tag{2}$$

where w is the vector of the parameters and $\max\left(0, 1 - y_i(w \cdot x_i - b)\right)$ is the *hinge* loss function. The term $\|w\|^2$ is the l_2 regularization penalty. The hyper-parameter λ is used to determine the trade-off between increasing the margin-size and ensuring that the x_i lie on the correct side of the margin.

The l_2 regularization penalty in Equation 2 does not lead to sparse solutions. The $l_1 = \|w\|$ regularizer or Lasso penalty is often used to increase the model sparseness since it can lead to solutions that have some elements with zero values [8]. In the proposed system, regularization is implemented by adding the l_1 norm penalty term to the hinge loss criterion (i.e. linear l_1-SVM classifier):

$$J = \min_w \left[\frac{1}{m} \sum_{i=1}^{m} \max\left(0, 1 - y_i(w \cdot x_i - b)\right) \right] + \lambda \|w\|, \tag{3}$$

In real implementation of Equation 3, the hinge loss is weighted by C:

$$J = \min_w \left[C \sum_{i=1}^{m} \max\left(0, 1 - y_i(w \cdot x_i - b)\right) \right] + \|w\|, \tag{4}$$

where the hyper-parameter $C = \frac{1}{m\lambda}$.

The solution of this objective function can be used to classify new points z:

$$c = \operatorname{sgn}(w \cdot z - b) \tag{5}$$

where c is the class identifier and b is a bias term.

The primal objective function in Equation 2 is commonly solved using a dual form with the Lagrangian [4, 28]. The dual form is given by:

$$J = \min_\alpha \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i \alpha_i k(x_i \cdot x_j) y_j \alpha_j, \tag{6}$$

$$\text{subject to} \sum_{i=1}^{m} \alpha_i y_i = 0, \text{ and } 0 \leq \alpha_i \leq \frac{1}{2m\lambda}; \text{for all i.} \tag{7}$$

where α are the parameters to optimize and $k(x_i \cdot x_j)$ is a kernel function. Some common kernels functions are: Polynomial (homogeneous): $k(x_i, x_j) = (x_i \cdot x_j)^d$, Gaussian radial basis function: $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, for $\gamma > 0$, and Hyperbolic tangent: $k(x_i, x_j) = \tanh(\kappa x_i \cdot x_j + c)$, for some (not every) $\kappa > 0$ and $c < 0$. The new points z can be classified by computing:

$$c = \operatorname{sgn}\left(\left[\sum_{i=1}^{m} \alpha_i y_i k(x_i, z) \right] - b \right), \tag{8}$$

The main advantage of the dual form solution is producing a nonlinear classifier. However, it is slow in training (i.e. $O(m^2)$) and classification phases due to the usage of kernels [3,14,17]. On

the other hand, the linear l_1-SVM classifier optimizes the primal soft-margin objective function, and is much faster than optimizing the dual objective functions with kernels.

The detection time is an important factor in the detection of attacks, and the dual form classifiers (i.e. Equation 8) may not be suitable for this task. The linear l_1-SVM classifier detection in Equation 5 is relatively fast. Hence, our main classifier is based on the primal form objective function solutions.

4 Data

To evaluate the proposed system, we need VoIP datasets. Unfortunately, real VoIP datasets are not available because of privacy concerns, so we used two generated datasets.

The first dataset was produced by INRIA [15]. They used two SIP proxy servers (i.e. Opensips and Asterisk), and VoIP attack tools to generate different scenarios of the invite flooding and SPIT attacks. The test-bed consists of a PC that acts as a server, two PCs generate the normal traffic using VoIP bots, and a PC that generates the attack messages. This dataset contains about 266,450 SIP messages.

In addition, the SIP-Msg-Gen tool [7] was used to generate the second dataset. It is a synthetic SIP message generator that generates normal SIP messages according to the SIP grammar defined in the RFC 3261 [19], and malformed SIP messages according to the SIP test messages defined in RFC 4475 [22]. The SIP-Msg-Gen tool can generate 14 different scenarios of the malformed SIP messages. All of these scenarios were used in the dataset generation. This dataset contains about 246,750 SIP messages.

For all experiments, we divided each dataset into three parts, 60% for training, 20% for cross validation, and 20% for testing. The training dataset was used to build the classification model. The cross validation dataset was used to tune the model hyper-parameters, and the test dataset was used to evaluate the final detection accuracy of the model.

5 Experiments

In this section, we evaluated the proposed approach using INRIA and SIP-Msg-Gen datasets. We projected the SIP messages into a high-dimensional space, and a linear l_1-SVM classifier was used for detection. In our proposed system we aim to achieve fast and high detection accuracy. Hence, we turned our attention to the primal form SVMs with l_1 regularization (i.e. linear l_1-SVM) to produce a sparse solution that will accelerate the detection process and decrease the number of active features. We compared the primal form SVMs with the dual form SVMs classifier. LibLinear [6] was used for the primal form l_1-SVM experiments. It is an open source library that solves large scale linear classification problems, and supports l_1 and l_2 regularizations. In addition, the LibSVM toolkit [10] was sued for dual form SVM experiments. All experiments are done in a machine with Intel Core i5 CPU, 3.2 GHz Quad-core and 8 GB RAM memory.

5.1 Evaluation

To evaluate the performance of our proposed model, we used F1 score [20]. It is the harmonic average of the precision and recall that takes into account the false positives and false negatives. The precision is the number of positive predictions divided by the total number of positive class predicted:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}, \tag{9}$$

and the recall is the number of positive predictions divided by number of positive class values.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}. \tag{10}$$

The F1 score is given by:

$$\text{F1} = 2 * \frac{\text{Precision * Recall}}{\text{Precision + Recall}} \tag{11}$$

where the precision and recall equally contributed into F1 score. The best F1 score is 1 and the worst is 0. The F1 score is usually more useful than accuracy, especially in case of different classes distribution.

The model sparseness was measured using the compression ratio (CR) criterion, given by

$$\text{CR}(C) = \frac{\text{\#Param - \#Pram(C)}}{\text{\#Param}} \tag{12}$$

where #Param is the number of features before the training process and #Pram(C) is the number of features after the training process, and is a function of the C parameter.

The target of the proposed system is to maximize the detection accuracy and minimize the required time for message classification. The average detection time $T_{\text{detection}}$ is computed as follows:

$$T_{\text{detection}} = \frac{\text{Features Extraction Time + Detection Time}}{\text{\# Messages in Test Dataset}} \tag{13}$$

where the features extraction time is the total time required to compute the features for all messages in the test dataset and the detection time is the total time required to run the l_1-SVM classifier on the test dataset. The average train time T_{train} was computed using the same equation but over the *training* dataset.

5.2 Results

The proposed classifier was trained using the training dataset and different values of the hyper-parameter C were tried to achieve the best detection accuracy. The F1 score, $T_{\text{detection}}$, CR (C), and T_{train} were reported for these experiments.

Our first experiment was performed on INRIA dataset, the feature vectors were created using n-gram with n=4 and cutoff=5. All n-grams that existed in the training dataset more than 5 times are stored in the dictionary, the dictionary of INRIA dataset have 120,209 4-grams. This dictionary was used to create feature vectors for the cross validation and the test datasets. Then the primal form l_1-SVM classifier was tried with different C values.

In the second experiment, we used SIP-Msg-Gen dataset with the same hyper-parameters (n=4 and cutoff=5), and the dictionary has 290,166 4-grams. The size of this dictionary is bigger than the INRIA dictionary because the malformed messages usually contain random content.

The F1 Score for INRIA and SIP-Msg-Gen datasets with different C values is shown in figure 3. For INRIA dataset, the $l1$-primal SVM classifier achieved 100% detection accuracy at C= 0.0039063 using only 37 features out of the 120,209 in 0.737 milliseconds average detection time per message. For the SIP-Msg-Gen dataset, the 100% detection accuracy is achieved at C= 0.5 using 9,800 features out of the 290,166 features in 0.570 milliseconds.

Figure 4 shows the results of CR for INRIA and SIP-Msg-Gen datasets. Because a high detection accuracy was achieved with a few number of features, the compression ratio for both datasets is high. INRIA achived 99% compression ratio while SIP-Msg-Gen achieved 96% compression ratio.

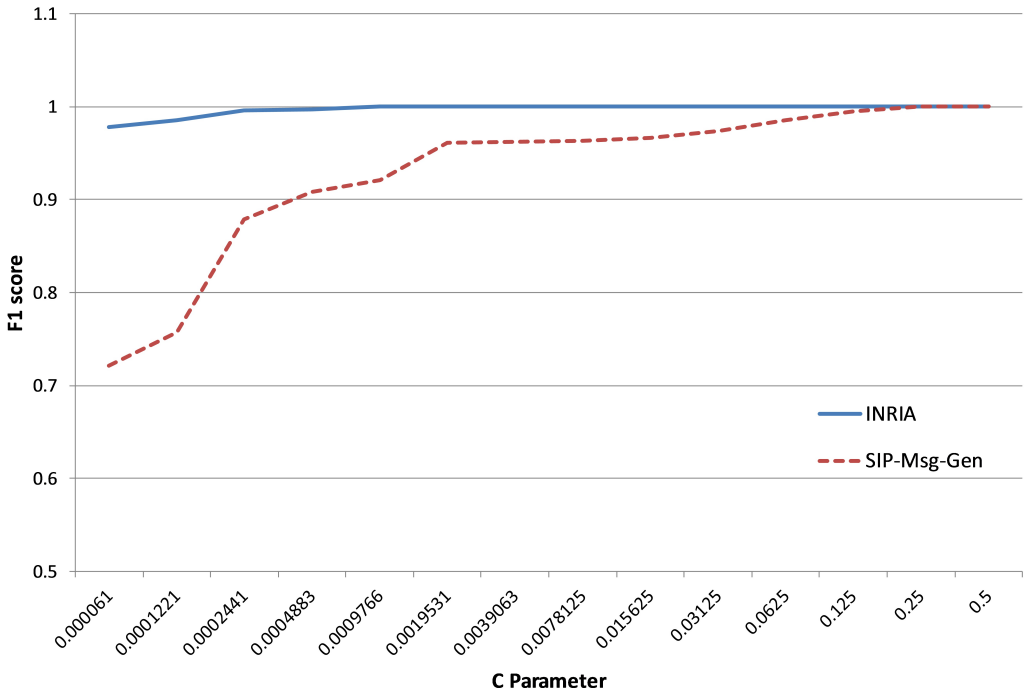

Figure 3: F1 score for INRIA and SIP-Msg-Gen datasets

The SIP-Msg-Gen dataset contains malformed messages, which usually have random content more than the invite flooding and SPIT messages in INRIA dataset. Hence, the l_1-SVM primal classifier needs more features with the SIP-Msg-Gen dataset to achieve high accuracy, and this leads to the low compression ratio with this dataset compared with the INRIA dataset.

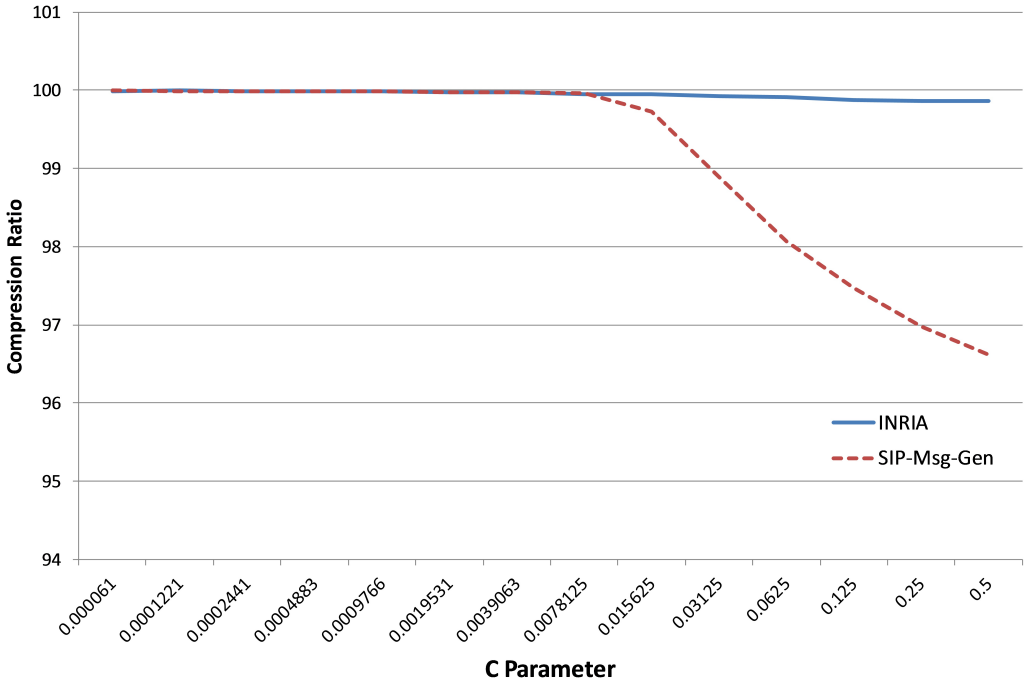

Figure 4: Compression ratio for INRIA and SIP-Msg-Gen datasets

Moreover, we calculated the proposed system throughput in megabits per second (Mbps) to measure the run-time performance. We achieved 2,700 Mbps using INRIA dataset, and 2,500 Mbps using SIP-Msg-Gen dataset which is considered a high throughput compared with previous work such as [18].

One of the important constraints for any learning algorithm is its scalability. When dealing with very large datasets, the dual form SVMs with kernels will be much slower than the primal form SVMs in detection and training. Although a fast detection is important, the fast training is also important especially in the case of the online adaption of the system.

Comparing between the dual form SVM with RBF kernel and the primal form SVM when both of them achieved 100% detection accuracy is given in Table 1. The primal form training time was about 17 times faster than the dual form using the INRIA dataset, and about 400 times faster with the SIP-Msg-Gen dataset. The detection time of the primal form was about 13 times faster than the dual form using the INRIA dataset, and about 100 times faster with the SIP-Msg-Gen dataset. This dual form setup is similar to the one used in [3].

Table 1: Comparison between the training time and the detection time for the primal and the dual form SVMs

Dataset	**SVM**	$T_{\mathbf{train}}$	$T_{\mathbf{detection}}$
INRIA	Primal	0.7445 ms	0.7370 ms
INRIA	Dual-RBF	13.0130 ms	10.1331 ms
SIP-Msg-Gen	Primal	0.5757 ms	0.5702 ms
SIP-Msg-Gen	Dual-RBF	219.5450 ms	57.0270 ms

Finally, we compared our linear l_1-SVM classifier to the state-of-the-art systems in Table 2. This comparison can only be considered as an indicator that the linear l_1-SVM classifier outperforms the other systems in terms of speed (detection time), because many factors such as the hardware configuration and the test dataset are different.

Table 2: Comparison between linear l_1-SVM and state-of-the-art systems.

Method	**Performance**	$T_{\mathbf{detection}}$	**Attacks**
linear l_1-SVM	F1 100% [1] F1 100%	0.7370 ms 0.5702 ms	Flooding and SPIT Malformed Msgs
Change Point [21]	F1 88%	0.76 ±0.45 sec	DDoS
Bayesian Change Point [12]	F1 95%	–	DDoS
Markov Chain and SVM [30]	Acc. 96.3%	15.145 ms	SPIT
Dual SVM [3]	Acc. 95-97%	–	Flooding
Sketch Design and Hellinger Distance [24]	Acc. 88% [2] Acc. 100%	–	Flooding
Dual SVM [17]	Acc. 99.9%	0.057-0.384 sec	Flooding
SDP Parser [27]	Acc. 100%	17-60 ms	Malformed Msgs
Dual SVM [1]	Acc. 99.9%	–	(D)DoS

6 Conclusions

In this paper, we proposed a machine learning system to detect the attacks of SIP based VoIP networks. The system projects the messages into a high-dimensional feature vector using n-gram

[1] F1 in the case of INRIA and SIP-Msg-Gen datasets respectively.

[2] Accuracy of low-rate and high-rate attack respectively.

tokens. In addition, a linear classifier to detect the SIP attacks (i.e. l_1-SVM classifier) is used for classification. Our work considers the fact that optimizing the primal soft-margin objective function is much faster than optimizing the dual objective function with kernels. Hence, we avoid the main drawback of the traditional dual form SVMs. Using the INRIA and SIP-Msg-Gen datasets, the proposed linear l_1-SVM classifier achieved competitive detection results to the other systems. Moreover, it is much faster than the state-of-the-art systems in the detection speed. Future work may focus on capturing a real VoIP dataset from a site under many VoIP attacks.

Bibliography

[1] Akbar, A.; Basha, S.M.; Sattar, S.A. et al. (2016). An intelligent SIP message parser for detecting and mitigating DDoS attacks, *Int. J. Innov. Eng. Technol*, 7(2), 1-7, 2016.

[2] Akbar, M. A.; Farooq, M. (2014). Securing SIP-based VoIP infrastructure against flooding attacks and Spam Over IP Telephony, *Knowledge and information systems*, 38(2), 491-510, 2014.

[3] Asgharian, H.; Akbari, A.; Raahemi, B. (2015). Feature engineering for detection of Denial of Service attacks in session initiation protocol, *Security and Communication Networks*, 8(8), 1587-1601, 2015.

[4] Cortes, C.; Vapnik, V. (1995). Support-vector networks, *Machine learning, Springer*, 20(3), 273-297, 1995.

[5] Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE transactions on electronic computers*, 3, 326-334, 1965.

[6] Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J. et al. (2008). LIBLINEAR: A library for large linear classification, *Journal of machine learning research*, 1871-1874, 2008.

[7] Ferdous, R. (2012). SIP-Msg-Gen : SIP Message Generator, [Online]. Available: https://github.com/rferdous/SIP-Msg-Gen, Accessed on 8 May 2019.

[8] Friedman, J.; Hastie, T.;Tibshirani, R. (2001). The elements of statistical learning, *Springer series in statistics New York*, 1(10), 2001.

[9] Hosseinpour, M.; Hosseini Seno, S.A.; Yaghmaee Moghaddam, M.H. et al. (2016). An anomaly based VoIP DoS attack detection and prevention method using fuzzy logic, *Telecommunications (IST), 2016 8th International Symposium on. IEEE*, 713-718, 2010.

[10] Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. et al. (2003). *A practical guide to support vector classification*, National Taiwan University, Taipei, 2003 (last updated 2016).

[11] Jurafsky, D.; Martin, J. H. (2014). *Speech and language processing*, Pearson London, 3-ed, 2019.

[12] Kurt, B. et al. (2018). A Bayesian change point model for detecting SIP-based DDoS attacks, *Digital Signal Processing*, Elsevier, 77, 48-62, 2018.

[13] Li, H.; Yildiz, C.; Ceritli, T.Y. et al. (2018). A Machine Learning Approach To Prevent Malicious Calls Over Telephony Networks, *arXiv preprint arXiv:1804.02566*, 2018.

[14] Nassar, M.; State, R.; Festor, O. (2008). Monitoring SIP traffic using support vector machines, *International Workshop on Recent Advances in Intrusion Detection*, Springer, 311-330, 2008.

[15] Nassar, M.; State, R.; Festor, O. (2010). Labeled VoIP data-set for intrusion detection evaluation, *Meeting of the European Network of Universities and Companies in Information and Communication Engineering*, 97-106, 2010.

[16] Packetizer, I. (2011). H. 323 versus SIP: A Comparison, [Online]. Available: http://www.packetizer.com/ipmc/h323_vs_sip, Accessed on December 2018.

[17] Pougajendy, J. and Parthiban, A. R. K. (2017). Detection of SIP-Based Denial of Service Attack Using Dual Cost Formulation of Support Vector Machine, *The Computer Journal*, Oxford University Press, 60(12), 1770-1784, 2017.

[18] Rieck, K.; Wahl S.; Laskov, P.; Domschitz, P. et al.(2008) A self-learning system for detection of anomalous SIP messages, *Principles, Systems and Applications of IP Telecommunications. Services and Security for Next Generation Networks*, Springer, 90-106, 2008.

[19] Rosenberg, J. (2002). SIP: Session Initiation Protocol, *IETF RFC 3261*, 2002.

[20] Sasaki, Y. (2007). The truth of the F-measure, *Teach Tutor mater*, 1-5, 2007.

[21] Semerci, M.; Cemgil, A. T.; Sankur, B. (2018). An intelligent cyber security system against DDoS attacks in SIP networks, *Computer Networks, Elsevier*, 136, 137-154, 2018.

[22] Sparks, R.; Hawrylyshen, A.; Johnston Avaya, A. et al. (2006). *Session initiation protocol (SIP) torture test messages*, 2006.

[23] Su, M.-Y.: Tsai, C.-H. (2015). Using data mining approaches to identify voice over IP spam, *International Journal of Communication Systems, Wiley Online Library*, 28(1), 187-200, 2015.

[24] Tang, J.; Cheng, Y.; Hao, Y. (2012). Detection and prevention of SIP flooding attacks in voice over IP networks, *INFOCOM, 2012 Proceedings IEEE*, 1161-1169, 2012.

[25] Tsiatsikas, Z.; Fakis, A.; Papamartzivanos, D. et al. (2015). Battling against DDoS in SIP: Is Machine Learning-based detection an effective weapon?, *12th International Joint Conference on e-Business and Telecommunications (ICETE)*, IEEE, 4, 301-308, 2015.

[26] Tsiatsikas, Z., Geneiatakis, D.; Kambourakis, G. et al. (2016). Realtime DDoS Detection in SIP Ecosystems: Machine Learning Tools of the Trade, *International Conference on Network and System Security*, Springer, 126-139, 2016.

[27] Tsiatsikas, Z.; Kambourakis, G.; Geneiatakis, D. et al. (2019). The Devil is in the Detail: SDP-Driven Malformed Message Attacks and Mitigation in SIP Ecosystems, *IEEE Access, IEEE*, 7, 2401-2417, 2019.

[28] Vapnik, V. (2013). The nature of statistical learning theory, *Springer science & business media*, 2013.

[29] Vennila, G.; Manikandan, M.; Aswathi, S. (2015). Detection of SIP signaling attacks using two-tier fine grained model for VoIP, *TENCON 2015-2015 IEEE Region 10 Conference*, IEEE, 1-7, 2015.

[30] Vennila, G.; Manikandan, M.; Suresh, M. (2017). Detection and prevention of spam over Internet telephony in Voice over Internet Protocol networks using Markov chain with incremental SVM, *International Journal of Communication Systems, Wiley Online Library*, 30(11), 2017.

[31] Wang, K.; Parekh, J.J.; Stolfo, S.J. (2006). Anagram: A content anomaly detector resistant to mimicry attack, *International Workshop on Recent Advances in Intrusion Detection*, Springer, 226-248, 2006.

[32] [Online]. Marchex. (2018). Spam Phone Calls Cost U.S. 2018 Small businesses half-billion dollars in lost productivity, Available: http://goo.gl/jTrgp3, Accessed on 10 March 2019.

[33] [Online]. Nettitude. (2015). VoIP Attacks on the Rise, Available: https://www.nettitude.com/uk/, Accessed on December 2018.