Detection of Signaling Vulnerabilities in Session Initiation Protocol

Diogo Pereira, Rodolfo Oliveira

Detection of Signaling Vulnerabilities in Session Initiation Protocol

Diogo Pereira[1], Rodolfo Oliveira[1,2]

[1]Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[2]Instituto de Telecomunicações, 1049-001 Lisboa, Portugal

dfca.pereira@campus.fct.unl.pt, rado@fct.unl.pt

Abstract. This paper investigates the detection of abnormal sequences of signaling packets purposely generated to perpetuate signaling-based attacks in computer networks. The problem is studied for the Session Initiation Protocol (SIP) using a dataset of signaling packets exchanged by multiple end-users. The paper starts to briefly characterize the adopted dataset and introduces a few definitions to propose a deep learning-based approach to detect possible attacks. The solution is based on the definition of an orthogonal space capable of representing the sampling space for each time step, which is then used to train a recurrent neural network to classify the type of SIP dialog for the sequence of packets observed so far. When a sequence of observed SIP messages is unknown, this represents possible exploitation of a vulnerability and in that case, it should be classified accordingly. The proposed classifier is based on supervised learning of two different sets of anomalous and non-anomalous sequences, which is then tested to identify the detection performance of unknown SIP sequences. Experimental results are presented to assess the proposed solution, which validates the proposed approach to rapidly detect signaling-based attacks.

Keywords: Deep Learning, Multimedia Networks, SIP Protocol.

1 Introduction

Nowadays, the Session Initiation Protocol (SIP) [1] plays an important role in the telecommunications arena. SIP supports a plethora of communication services, including voice calls and legacy Public Switched Telephone Network systems through Voice over Internet Protocol (VoIP) [2]. Not less important is the SIP role on cellular networks, where it is used to support all IP Multimedia Subsystem (IMS) services' signaling [3, 4], including multimedia and non-multimedia services. The SIP protocol allows the establishment of sessions through adequate authentication mechanisms and signaling control flows that are dynamic enough to accommodate several purposes, e.g., session initiating, maintaining, and terminating between two peers.

The security of the SIP protocol is an important aspect given its importance in the telecommunication operators and in supporting non-commercial VoIP services. It is

well-known that SIP is exposed to a significant number of vulnerabilities [5, 6]. In this paper, we are interested in exploring the vulnerabilities caused by the combination of different signaling patterns, which can cause denial-of-service, unauthorized access to a call, billing errors, and other types of attacks [5]. Consequently, it is important to identify potential malicious SIP signaling sequences received by the SIP servers, including new signaling sequences never observed before. While the already known potential malicious sequences can be detected in an automated way, the SIP sequences never observed before need to be analyzed by domain experts who can then assess their level of vulnerability. However, the detection of anomalous SIP signaling sequences is challenging due to the high number of different signaling sequences, the order of the messages in the dialog, and the dialogs' variable length.

Motivated by the advantages of adopting deep learning techniques and admitting that a SIP server can access all SIP messages as they occur over time, in this paper we propose a deep learning scheme to detect anomalous and/or unknown SIP signaling sequences as they transverse the SIP servers. Apart from describing the proposed solution, this paper also reports its performance evaluation tested with practical data.

1.1 Research Question and Contribution

The research question tackled in this paper has to do with the design of innovative learning and classification methodologies capable of identifying unknown or abnormal SIP sequences in the shortest amount of time. The research question is stated as follows:

Considering that a SIP server, or a SIP agent, has access to all SIP messages as they occur over time, is it possible to classify SIP signaling patterns, so that the detection of abnormal or unknown SIP dialogs can be quickly done in an efficient manner? Additionally, is it possible to predict abnormal or unknown SIP dialogs when only a few elements of the SIP signaling sequence are known?

The main contributions of this work are: (1) an innovating modeling approach, capable of representing the knowledge related with the SIP sequences; (2) the knowledge model is then explored in a deep learning scheme, based on Long Short-term Memory (LSTM) Recurrent Neural Networks (RNN); (3) a statistical analysis is proposed to automatically detect unknown SIP sequences; (4) the proposed methodology is evaluated using practical data, showing that the detection of unknown SIP dialogs can reach 99.84%, which demonstrates the practical advantage of the proposed scheme.

1.2 Related Work

The SIP [1] is an application-layer protocol designed to initiate, maintain, and terminate multimedia sessions through the exchange of SIP messages between each user agent. Each SIP message can be either a request or a response. Initially, a SIP message must be sent with a request that can be identified by a specific method. In response to one of those methods, a response SIP message is sent with a specific code. Every SIP request

exchanged between agents initiate a SIP transaction, and multiple SIP transactions exchanged between two peers form a SIP dialog, which represents the peer-to-peer relationship over time. A user agent can identify the different dialogs through the SIP Call-ID, i.e., a unique identifier for every dialog's message.

The vulnerabilities of the SIP protocol have been identified in several works such as [5, 6]. These include service interruption, service destruction, or unauthorized access to previously reserved computing resources. SIP service interruption can be caused by flooding attacks, and different solutions include threshold-based classifiers that compare the traffic patterns with the prior statistics [7]. Malformed SIP messages are another way of compromising SIP. Malicious SIP messages are usually detected through intrusion detection systems or identification of deviations from a priori statistics [8]. Another class of SIP vulnerability, aka SIP signaling vulnerability, take advantage of defective implementations of the protocol, where protocol implementation issues can be explored by sending SIP messages to allow improper authentication mechanisms [5]. A mitigation approach for this type of vulnerability was proposed in [9], where a rule-based methodology is used according to the contextual information of the SIP traffic. More recently, the work in [10] has proposed a methodology based on the SIP sequences and their timings that are then used to detect deviations that may represent vulnerabilities. Although different SIP signaling vulnerabilities have already been proposed in [9, 10], this work is not assuming a fixed probabilistic model of the SIP operation. Contrarily, we propose a methodology capable of learning from past SIP sequences, which is used to detect unknown SIP dialogs that can be further categorized by domain experts. Moreover, the vulnerability of the abnormal dialogs can also be evaluated based on prior trustworthy SIP data.

2 Applied Artificial Intelligence Systems

A wide range of services has been supported by innovative applied Artificial Intelligence (AI) systems that have been applied in several areas, such as e-commerce, banking, and social media, just to mention a few. While in several application scenarios the amount of data and time constraints are not a big concern, in other applications the AI systems need to cope with very large amounts of data and a quick response might be required. Traditionally, the security of individuals and computational systems has been an area of massive adoption of AI systems. In this work, we address the security of the SIP protocol, an important tool for network operators and society in general. This work is mainly centered on the adoption of AI tools to learn from prior SIP data and to classify SIP dialogs in the shortest amount of time. Recent advances in AI, particularly on deep learning techniques that deal with sequential data are natural candidates to be used in the classification of SIP dialogs. However, the high volume of signaling traffic on the network operator servers needs well-tailored solutions capable of making an accurate decision using the minimum amount of data so that they achieve a quick response. Our work aims at fulfilling this challenge, by adopting state-of-art deep learning techniques based on RNNs that are feed with real-time SIP packets traversing the SIP servers. The focus is on the efficient classification of SIP dialogs observed so far, as well as on the detection of unknown SIP dialogs never observed before.

3 Modeling, Learning, and Classification

3.1 System Model

A **SIP message** m_k, $k \in \mathcal{M} = \{1,2,\dots,M\}$, represents a specific type of SIP request or SIP response. The total number of types of SIP requests plus responses is denoted by M, and $\mathcal{M}$ represents the set of all types of SIP messages. A SIP dialog is composed of SIP messages. More specifically, a **SIP dialog** is a sequence of consecutive SIP messages represented by $\boldsymbol{d}_k =< m^{(1)}, m^{(2)}, \dots, m^{(L_d)} >$, where $m^{(j)}$ represents the $j-th$ message of the SIP dialog. L_d represents the SIP dialog length. All SIP messages contained in a SIP dialog share the same SIP Call-ID, and sender and receiver URIs. Given the number of possible SIP methods in a request and possible reply codes in a response, the number of different dialogs that can be created between the user agents is high. Besides that, the dialogs can be legitimate or anomalous.

An **observation** k taken by a user agent or a SIP server is a sequence of consecutive SIP messages represented by $\boldsymbol{o}_k =< m^{(1)}, m^{(2)}, \dots, m^{(L_o)} >$. Each SIP message is represented by $m^{(h)} = m_i,\ i \in \mathcal{M}, h \in \{1,2,\dots,L_o\}$. The symbol L_o represents the length of the observation.

A requirement to meet when working with sequential neural networks is that the length of the input data must be described over consecutive discrete-time events. The set of events is represented by the observation. However, because the length of the observations can be variable, we transform each observation into a fixed-length stuffed sequence, denoted as a pad sequence. A **pad sequence** $\boldsymbol{s}_k$ associated to the observation $\boldsymbol{o}_k$, is a sequence of length $L_N = L_o + n$, where n represents the number of zeros added to the observation as follows, $\boldsymbol{s}_k =< \boldsymbol{o}_k, \underbrace{0,0,\dots,0}_{(n)} >$. L_N represents a fixed length in all pad sequences. Another aspect that must be evaluated is the type of data handled by the neural network, i.e., numerical, or categorical data, since it influences how the input data is normalized. The SIP methods and responses are categorical data, and its encoding is done through an **encoded SIP message** $\boldsymbol{m}_i'$, which represents a Boolean vector describing the SIP message m_i. This vector has a length of $L_M = M + 1$, where 1 represents the zero-pad symbol and is obtained using a One Hot Encoder [12].

3.2 Deep Learning and Classification

The learning of the SIP dialogs observed in a server is tackled in this subsection through the adoption of a recurrent neural network, more specifically a LSTM.

Since the goal of this work is to classify the input data in multiple SIP dialogs contained in trustworthy or non-trustworthy datasets, it was designed the LSTM architecture to detect an observed input sequence of L_o SIP messages. This architecture is presented in Figure 1, consisting of one LSTM RNN model and an Unknown SIP Dialogs' Classifier. The LSTM RNN model receives a $1 \times L_N \times M$ input sequence $\boldsymbol{s}_k$, produced by the Pad Sequence and the One Hot Encoder. Then the LSTM layer processes one step at each L_N discrete time units of $\boldsymbol{s}_k$ and returns a $1 \times N$ sequence, $\boldsymbol{h}_0$, with real values in the interval $[-1,1]$. The Dense layer receives the LSTM output and generates the output vector, $\boldsymbol{y_k}$, a $1 \times N$ sequence with real values in $[0,1]$. The

variable N stands for the total number of unique SIP dialogs contained in the trustworthy dataset. Through this output vector, it is possible to identify the most probable dialog by finding the position with maximum value.

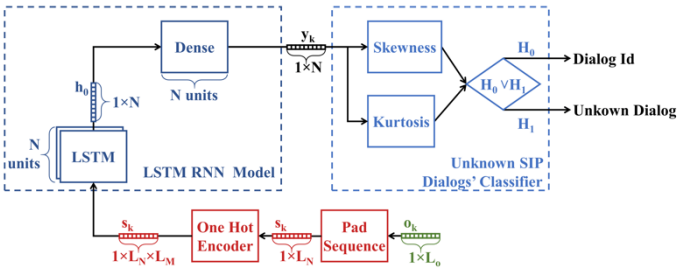

Fig. 1 LSTM architecture, including the unknown SIP dialog's classifier.

3.3 Unknown SIP Dialogs' Detection

The unknown SIP dialogs are the ones not included in the training datasets of trustworthy and non-trustworthy dialogs. The goal behind the detection of unknown SIP dialogs is to identify SIP dialogs not included in the training datasets and thus detect SIP dialogs never observed before. The output of the LSTM RNN model is used as input of the unknown SIP dialogs' classifier. From this input, it is computed the central moments to evaluate if they can be used to distinguish unknown SIP dialogs. Next, we describe a semi-supervised threshold-based classifier used to distinguish known and unknown SIP dialogs.

The classifier starts to compute the normalized kurtosis and skewness of each observed SIP dialog. The kurtosis and skewness thresholds are given by $\lambda_k = \mu_k - \sigma_k^2$ and $\lambda_s = \mu_s - \sigma_s^2$, respectively, where μ_k and μ_s represent the mean of the kurtosis and skewness values computed from the LSTM RNN model outputs for all SIP dialogs in the training set, and σ_k^2 and σ_s^2 represent the variance of the kurtosis and skewness values, respectively.

A SIP dialog is classified as a known dialog, hypothesis H_0, or unknown dialog, hypothesis H_1, according to the following conditions:

$$H_0: \mu^3 \geq \lambda_s, \mu^4 \geq \lambda_k,$$
$$H_1: \mu^3 < \lambda_s, \mu^4 < \lambda_k,$$

where μ^3 and μ^4 represent the skewness and kurtosis of the LSTM output obtained with the SIP dialog to classify.

4 Performance Evaluation

To validate the performance of the architecture proposed in Section 3, it was adopted the SIP dataset created in [11]. Regarding the characteristics of this dataset, it contains 1492 types of SIP dialogs, which correspond to a total of 18782 SIP dialogs. The length

of these SIP dialogs is between 2 and 56, and the number of unique SIP messages (M) is 17. Through this information, we can identify some of the variables previously defined in Subsection 3.1 and 3.2, particularly, L_N=56, L_M=18, and N=1492.

Three different datasets were used to train, validate, and test the proposed neural network architecture. The datasets were randomly divided containing 50%, 20%, and 30% of the transactions exchanged on the SIP dataset, representing the training, validating, and testing datasets, respectively. Finally, the results here reported were computed in Python running over a 64bit Ubuntu 18.04.5 LTS with 128 GB of RAM and running over an Intel(R) Core(TM) i7-9800X CPU @ 3.80GHz and GeForce RTX 2080 Ti 11GB. Regarding the execution in a RTX 2080 Ti, its advantage is mainly observed for training the neural network. Equivalent or even lower computational times can be achieved by implementing the neural network in dedicated hardware (e.g. adopting FPGAs).

To perform the detection of the most likely SIP dialog for each given sequence of SIP dialogs the LSTM RNN model was trained for 500 epochs and the detection probability is reported in Table 1. The detection probability shows that the LSTM RNN

Table 1. Detection probability of the LSTM RNN model in each dataset.

Dataset	Train	Validation	Test	Joint Datasets
Detection Probability	100.00%	93.54%	92.80%	96.60%

model can classify all the dialogs it learned, since it detects all the SIP dialogs belonging to the training set. However, the same cannot be stated for the validation and test datasets, because the detection probability is lower than 100%. The lower performance is due to the existence of SIP dialogs that were never learned before. This is a consequence of their lower occurrence in the SIP dataset meaning that when they were distributed over each dataset they were placed in the test and validation datasets.

The detection performance of the proposed model is also depicted in Figure 2, where the prediction performance is studied as a function of the relative amount of information that constitutes the SIP dialog, i.e., L_o/L_d. In this test, the model uses as inputs an increasing sequence of SIP messages until reaching the complete SIP dialog. For each

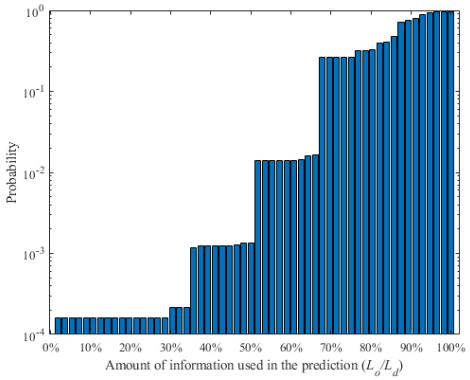

Fig. 2 SIP dialogs prediction probability over the amount of available information.

added new message it is checked if the model is capable of successfully classifying that specific dialog. The results show that the proposed model correctly predicts some dialogs with the minimum amount of information available, close to 1.79%. However, to correctly identify 50% of the dialogs it is required to know approximately 85.71% of the information of the SIP dialog. Finally, when the complete SIP dialog is known ($L_o/L_d = 100\%$) we enter the detection stage where its probability is close to 0.9660. Additionally, the average detection computing time is 1.969 *ms*.

As referred above, there are SIP dialogs that are mis-detected by the proposed LSTM RNN model. Thus, it is important that these dialogs are classified as unknown, to prevent the occurrence of attacks. The classification of unknown SIP dialogs was achieved through the addition of a classifier that was parametrized to distinguish between the dialogs that were already trained and the unknown ones. The classifier is based on the skewness and kurtosis of the LSTM RNN model's output, as explained in Subsection 3.3, and the decision is based on two thresholds, $\lambda_k = 0.9816$, and $\lambda_k = 0.96826$ that were computed using the training dataset. Concerning the performance of the classifier, Figure 3 illustrates the distribution of the dialogs classified into four different labels. These labels present the dialogs that were correctly classified as

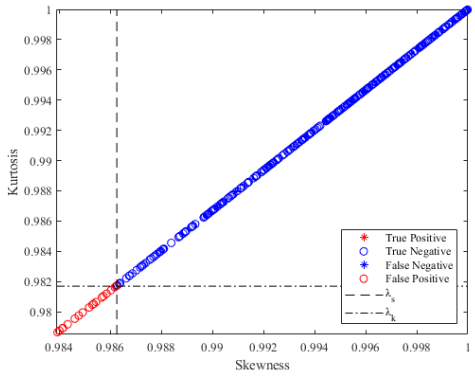

Fig. 3 Performance of the unknown SIP dialog's classifier.

unknown and already trained, respectively, true positive and true negative, but also the dialogs that were misclassified (false negative and false positive). Through the labels, we have computed the probability of correctly classifying the dialogs as unknown (P_D) and the probability of misclassifying a trustworthy dialog (P_{FA}). The results are presented in Table 2, showing that the proposed classifier can detect 99.84% of the unknown SIP dialogs. This is an important result since it can be used to detect possible vulnerabilities through the appropriate evaluation of the SIP dialogs by domain experts capable of assessing its vulnerability level.

Table 2. Unknown SIP dialogs classifier metrics.

$P_D = \frac{TP}{TP + FN}$	$P_{FA} = \frac{FP}{TN + FP}$
99.84%	7.24%

5 Conclusions

Giving the high importance of SIP security for telecommunication companies and non-commercial VoIP services, this paper proposed a dynamic methodology capable of identifying important protocol vulnerabilities related to SIP signaling. More concretely, we have proposed an innovative methodology to detect SIP sequences according to prior data, which can be labeled as trustworthy or non-trustworthy data. Additionally, a statistical methodology is also described to detect unknown SIP dialogs that can constitute an important source of attacks. The experimental assessment described in the paper and its results show the effectiveness of the proposed methodology to improve the security of SIP-based services.

References

1. J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. Sip: Session initiation protocol. RFC 3261, RFC Editor, June 2002.
2. Uzelac and Y. Lee. Voice over ip (voip) sip peering use cases. RFC 6405, RFC Editor, November 2011.
3. H. Khlifi and J. Grégoire, "IMS Application Servers: Roles, Requirements, and Implementation Technologies," in IEEE Internet Comput., vol. 12, no. 3, pp. 40-51, May-June 2008.
4. A. Achour, K. Haddadou, B. Kervella and G. Pujolle, "A SIP-SHIM6-based solution providing interdomain service continuity in IMS-based networks," in IEEE Commun. Mag., vol. 50, no. 7, pp. 109-119, July 2012.
5. D. Sisalem, J. Kuthan, and S. Ehlert. Denial of service attacks targeting a sip voip infrastructure: attack scenarios and prevention mechanisms. IEEE Netw., 20(5):26–31, 2006.
6. D. Geneiatakis, T. Dagiuklas, G. Kambourakis, C. Lambrinoudakis, S. Gritzalis, K. S. Ehlert, and D. Sisalem. Survey of security vulnerabilities in session initiation protocol. IEEE Commun. Surveys Tuts., 8(3):68–81, 2006.
7. I. M. Tas, B. G. Unsalver and S. Baktir, "A Novel SIP Based Distributed Reflection Denial-of-Service Attack and an Effective Defense Mechanism," in IEEE Access, vol. 8, pp. 112574-112584, 2020.
8. N. Hentehzadeh, A. Mehta, V. K. Gurbani, L. Gupta, T. K. Ho, and G. Wilathgamuwa. Statistical analysis of self-similar session initiation protocol (sip) messages for anomaly detection. In 2011 4th IFIP International Conference on New Technologies, Mobility and Security, pages 1–5, 2011.
9. A. Lahmadi and O. Festor. A framework for automated exploit prevention from known vulnerabilities in voice over ip services. IEEE Trans. Netw. Service Manag., 9(2):114–127, 2012.
10. D. Golait and N. Hubballi. Detecting anomalous behavior in voip systems: A discrete event system modeling. IEEE Trans. Inf. Forensics Security, 12(3):730–745, 2017.
11. Mohamed Nassar, Olivier Festor, et al. Labeled voip data-set for intrusion detection evaluation. In Meeting of the European Network of Universities and Companies in Information and Communication Engineering, pages 97–106. Springer, 2010.
12. David Harris and Sarah Harris. Digital Design and Computer Architecture, Second Edition. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2012.