SIP-DDoS: SIP Framework for DDoS Intrusion Detection based on Recurrent Neural Networks

Oussama Sbai, Benjamin Allaert, Patrick Sondi, Ahmed Meddahi

SIP-DDoS: SIP Framework for DDoS Intrusion Detection based on Recurrent Neural Networks

Oussama SBAI[0000−0003−4826−4446], Benjamin ALLAERT[0000−0002−4291−9803], Patrick SONDI[0000−0001−9484−7357] and Ahmed MEDDAHI[0000−0002−9255−2114]

Centre for Digital Systems, IMT Nord Europe, Institut Mines-Télécom, 59000 Lille, France
{oussama.sbai, benjamin.allaert, patrick.sondi, ahmed.meddahi}@imt-nord-europe.fr

Abstract. The rapid evolution of beyond fifth-generation (B5G) and sixth-generation (6G) networks has significantly driven the growth of Internet of Things (IoT) applications. These applications are characterised by: a massive connectivity, high security level, trust, wireless coverage, also ultra-low latency, high throughput, and ultra-reliability, especially for real-time oriented sessions or sensor like cameras. While traditional protocols like MQTT and CoAP are inadequate for such types of applications, under certain conditions, the 3GPP standard Session Initiation Protocol (SIP) emerges as a promising solution. However, SIP faces various Distributed Denial of Service (DDoS) threats, as INVITE flooding attacks presenting a significant challenge. This work presents a GRU-based Intrusion Detection System (IDS) to detect SIP-INVITE flooding attacks. Leveraging recurrent neural networks, the IDS efficiently process sequential SIP traffic data in real time, identifying attack patterns effectively. The GRU's ability to capture temporal dependencies enhances accuracy in classifying and detecting attack behaviors. The results demonstrate that the framework can effectively detect and mitigate INVITE flooding attacks of different intensities, under practical settings. The performance results show that the proposed framework is robust and can be practically deployed, e.g., inference time less than 800 µs and a marginal rate for the misclassified traffic.

Keywords: SIP protocol · DoS/DDoS · SIP-INVITE flooding attack · Recurrent Neural Networks · IoT

1 Introduction

The Internet of Things (IoT) has become a powerful and transformative force, driven by the goal to improve safety, efficiency, and sustainability for a wide range of applications. IoT is one of the most promising technological advancements on the horizon is the integration of sixth-generation (6G) networks and the evolution beyond 5G (B5G) as it has proven capacity for innovation and enhancement. This development has the potential to revolutionize communication

systems by optimizing processes within smart cities, industries, or powering the seamless operation of intelligent transportation systems (ITS). This also by introducing high speeds, ultra-low latency, and a vastly expanded capacity. These attributes are indispensable for accommodating the expected exponential growth in connectivity. The convergence of the Internet of Things (IoT) with 6G and B5G networks presents an unprecedented opportunity to transform the digital landscape. This technologies integration will enable the deployment of innovative IoT applications that require real-time constraints, reliable communication channels, efficient and predictive resource management, and robust security systems.

In this context, the network communications require "lightweight" protocols for session initiation and control, such as Message Queuing Telemetry Transport (MQTT) and Constrained Application Protocol (CoAP). However, MQTT and CoAP are not applicable in advanced IoT applications that require high data rates and low latency [22, 23]. Session Initiation Protocol (SIP) is considered as a good candidate, for supporting a wide variety of IoT application scenarios [7, 11, 22, 23]. Besides instant messaging, it is suitable for handling short or long session semantics for streaming data and publish-subscribe semantics for event notifications, e.g., subscribe, publish, notify or method, thereby enabling more sophisticated IoT network operations. The Session Initiation Protocol (SIP) is well-suited for this purpose and functions as a signaling protocol designed for multimedia communication [9, 22, 23]. Nevertheless, one of the major SIP Distributed Denial of Service (DDoS) attacks is the INVITE flooding attack [5]. In this attack, malicious users flood the proxy or SIP server with a high volume of malicious INVITE messages with the aim of disrupting the service.

In this paper, we propose a comprehensive framework specifically designed to detect SIP INVITE flooding attacks. Recent approaches to detecting SIP-DDoS attacks focus mainly on analyzing the content of SIP messages or the entire SIP dialog [14, 16]. However, these works have not been studied to analyze SIP traffic in time intervals. This temporal segmentation makes it possible to systematically examine and evaluate SIP communication patterns, while facilitating timely anomaly detection and improving the overall effectiveness of the IDS. The proposed framework is designed to address the unique challenges posed by such attacks, and to provide an effective defense mechanism against DoS/DDoS threats targeting SIP infrastructures. The core of the proposed framework is a Gated Recurrent Unit (GRU) based Intrusion Detection System (IDS). By harnessing the power of recurrent neural networks, the IDS can effectively process sequential SIP traffic data and identify patterns indicative of INVITE flooding attacks. The GRU's ability to capture temporal dependencies makes it a valuable asset in recognizing attack behaviors with high accuracy. To evaluate the proposed IDS, the Train/Test/Validation protocol was employed, a widely-used technique for assessing the model's performance. Furthermore, an independent test dataset was used to validate the IDS's effectiveness. The results demonstrate the framework's ability to accurately detect and mitigate INVITE flooding attacks, even under different attack intensities, showcasing its robustness and

practicality in real-world settings. The proposed framework is available online [1], containing the code, dataset and learning models. The main contributions are the followings:

- Generating a realistic dataset: to ensure the effectiveness of the framework, a diverse and realistic dataset was generated using the SIPp [19] and Mr.SIP [21] tools. This dataset contains a legitimate and malicious INVITE messages with different scenarios and message rates, providing a solid basis for training and evaluating our proposed deep learning based IDS.
- Proposing a reliable and reproducible evaluation protocol: a strategy for encoding SIP message frames is proposed to facilitate learning model convergence. Several recurrent neural network models are confronted in order to study their ability to detect anomalies in different attack scenarios. A study of detection accuracy and inference time is carried out.

The paper is structured as follows. Section 2 reviews the related work, discussing existing research in the field. Section 3 presents the proposed framework, detailing the experimental platform, the dataset generation process, and the approach for detecting SIP DDoS attacks. Section 4 describes the detection methodology, and Section 5 presents the results obtained from deploying the proposed approach. Finally, we conclude by highlighting the key findings and suggesting directions for future research.

2 Related Works

The related literature shows that DDoS attacks can be considered as one of the most predominant threats in IoT networks [6, 8, 12, 15]. DDoS attacks pose significant challenges to the security and stability of IoT infrastructures due to their ability to disrupt services by overwhelming target systems with a massive flow of malicious traffic. This section provides an overview of the proposed dataset for SIP-DoS/DDoS transactions, as well as a review of various studies that address approaches to detecting SIP anomalies in terms of data encoding and technologies used.

2.1 SIP dataset

Few datasets have been proposed in the literature for analyzing DDoS attacks. Alvares et al. [1] introduce a dataset dedicated to Voice over Internet Protocol (VoIP) networks, while considering various DoS/DDoS attack types, including the INVITE flooding attack. Nassar et al. [13] present a dataset with multiple SIP-DoS/DDoS attacks, including INVITE flooding. They used "Inviteflood" tool with two intensity scenarios: 100 and 1000 requests per second.

However, these two existing datasets suffer from same limitation or lack of details as they do not provide a comprehensive details regarding the considered

[1] Repository: https://gvipers.imt-nord-europe.fr/benjamin.allaert/sip-ddos

specific attack, especially they do not consider the various types of SIP-INVITE flooding attacks and scenarios. Additionally, they do not address the critical issue of invalid IP addresses, in case of IP spoofing attacks. This information can provide a more holistic and realistic representation of practical intrusion scenarios. These shortcomings tend to reduce the ability to properly evaluate IDSs for SIP-INVITE attacks.

2.2 SIP anomalies detection

Recent approaches to detecting SIP-DDoS attacks focus mainly on analyzing the content of SIP messages or the entire SIP dialog. Pereira et al. [17] present an IDS model for recognizing SIP signaling patterns to identify abnormal SIP dialogs within observed SIP sequences. Their core approach is based on a LSTM architecture. They compared their methods to a probabilistic-based solution and found that their methods achieved higher detection scores in a shorter time. Extension[16] have been proposed for signaling SIP attacks based on Convolutional Neural Networks (CNN) architecture. The performance evaluation demonstrates that both the CNN and LSTM models achieve similar effectiveness in detecting the most probable SIP dialog identifier. Nazih et al. [14] introduce an IDS based on the RNN architecture, designed to analyze message content and identify SIP-INVITE flooding attacks. To evaluate the effectiveness of their proposed solution, the authors conduct experiments using a dataset that includes real legitimate SIP messages (normal usage), while the malicious messages (INVITE flooding attacks) are generated using the SIPp-DD tool [20]. However, these works have not been studied to analyze SIP traffic in time intervals. This temporal segmentation makes it possible to systematically examine and evaluate SIP communication patterns, while facilitating timely anomaly detection and improving the overall effectiveness of the IDS.

Several SIP message encoding and artificial data generation techniques have been proposed to improve the ability of neural networks to detect DDoS attacks. Nazih et al. [14] provide a character- or token-based feature extraction process. Raw data is transformed into sequences, following by padding techniques to ensure consistent sequence lengths. Embedding techniques are then employed to represent the data into a format that can be exploited for a meaningful analysis and pattern recognition, related to the SIP-INVITE message. The results clearly demonstrate that the token-based approach combined with the GRU and LSTM architectures give the best performance. Meddahi et al. [10] introduce Generative Adversarial Networks (GAN) to augment SIP messages. The Authors convert the SIP traffic data into images, enabling image-based techniques to be applied to SIP traffic data. The final phase involves the deployment of a GAN model utilized to generate new SIP messages. These newly generated SIP messages are then integrated into the dataset, expanding its size and diversity, effectively. The authors introduce a parameter γ to measure the gap between the synthetic and real SIP-data. Nevertheless, data augmentation using techniques like GAN introduce information loss in the SIP message fields. This limitation is critical for message and traffic analysis in case of intrusion detection. While these techniques

have proved effective for detecting SIP-DDoS attacks, focusing on the analysis of SIP message content or the SIP dialog itself, they have not been proven effective for analyzing an attack in real-time sequential SIP traffic data.

3 Framework

This section, is dedicated to describe the proposed framework. We provide a comprehensive overview of the SIP Protocol, its core principles, and functionalities. We also introduce the tools used in our experimentation, including the simulation environment and data generation techniques. Finally, we outline our SIP dataset, with a focus on the different SIP-INVITE flooding attack scenarios.

3.1 SIP protocol overview

The Session Initiation Protocol (SIP) stands as a foundational signaling protocol within the domain of real-time communication [RFC3261-3265], primarily designed to initiate, maintain, modify, and terminate communication sessions between Internet Protocol (IP) devices. SIP finds application in a diverse array of scenarios, encompassing Voice over IP (VoIP), video conferencing, instant messaging, and presence services.

SIP adopts a text-based protocol architecture, following a request-response paradigm. Typically transmitted over User Datagram Protocol (UDP), although Transmission Control Protocol (TCP) is also viable, SIP messages are structured with a header-body format. The header comprises essential information concerning message attributes, source and destination addresses, and various parameters, while the body carries supplementary data, such as session descriptions or multimedia streams.

SIP session establishment is initiated through the transmission of an INVITE message from the caller to the callee. The INVITE message includes a session description, detailing the caller's preferred media types and supported codecs. The callee, upon accepting the session, responds with a 200 OK message, marking the establishment of the session. Subsequently, participants can exchange media streams using the designated codecs.

SIP allows for session modifications and terminations via corresponding SIP messages. To adjust a session, participants may employ a REINVITE message, while session termination is accomplished through a BYE message. The detailed process of the SIP call is illustrated in Fig. 1.

Distinguished by its adaptability and extensibility, SIP incorporates various noteworthy features:

- Scalability: SIP is architected to scale efficiently, accommodating a substantial user base and numerous concurrent sessions;
- Reliability: SIP ensures message reliability through mechanisms designed to guarantee successful delivery;
- Security: SIP boasts a range of security mechanisms, encompassing authentication and encryption;

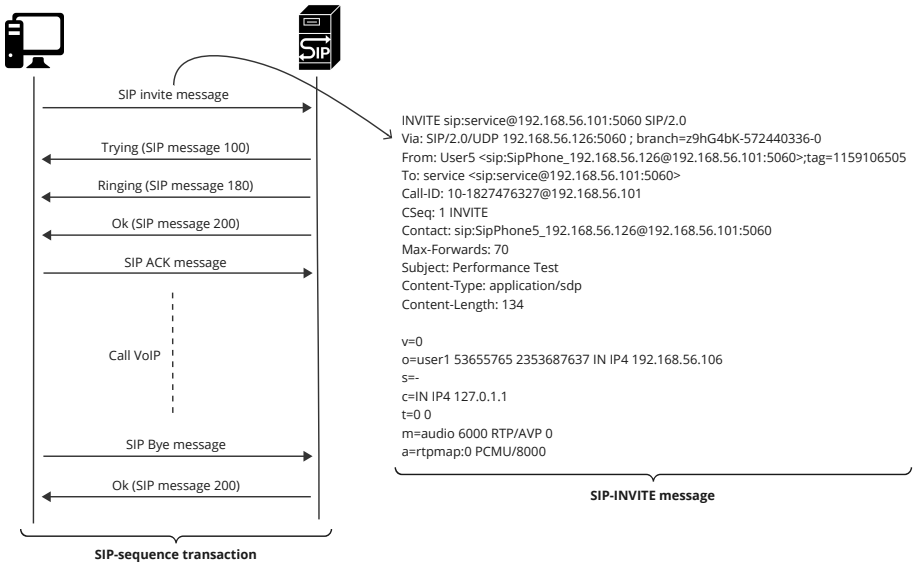

Fig. 1. An overview of SIP session transaction and an example of SIP-INVITE message

– Mobility: SIP supports user mobility, allowing seamless session transition as users move between locations.

Widely adopted in the telecommunications sector, the SIP protocol is the backbone of VoIP and real-time communication services offered by many providers. It also finds extensive application in enterprise solutions, including video conferencing and instant messaging.

Several real-world instances underscore SIP's ubiquity:

– VoIP Calls: SIP underpins VoIP calls via mobile applications, orchestrating call initiation and management;
– Video Conferencing: SIP facilitates the establishment and control of video conferences, whether through web browsers or dedicated conferencing clients;
– Instant Messaging: In messaging applications, SIP may be leveraged for message delivery;
– Presence Services: SIP plays a role in delivering presence information, informing users about the online status of their contacts.

SIP stands as a versatile and influential protocol underpinning a broad spectrum of real-time communication applications, exemplifying its indispensable role within contemporary telecommunications infrastructure.

3.2 SIP communication network simulation

The proposed dataset was generated using the SIPp simulator, which served as the User Agent Client (UAC) for generating legitimate INVITE messages, and

as the User Agent Server (UAS) for receiving and manipulating SIP call sessions. To simulate the DDoS attack scenarios, the Mr.SIP simulator was utilized as the malicious UAC, generating the INVITE flooding attacks. The network architecture adopted for simulating SIP communication is depicted in Fig. 2, providing an overview of the setup used for the experiments.

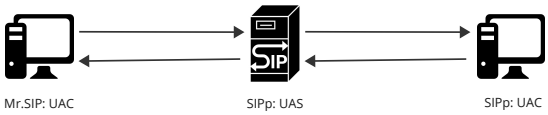

Fig. 2. Architecture of simulated network, where Mr.SIP is used for DDoS attacks traffic generation and SIPp for SIP server and legitimate users

To ensure the quality and sophistication of the generated data, some modifications have made to the source code of the SIPp simulator. These modifications allowed for the generation of more elaborate SIP INVITE messages, including the main SIP fields such as the source IP address, branch number, URI, and call number. Fig. 3 shows that SIPp is not full compliant with SIP specifications as the fields of the original SIP INVITE message are not randomly generated. This fields are based on a specific format that does not comply with the SIP specifications RFC 3261 [18]. As a result, the generated SIPp messages deviated from the conventional format. The same figure shows an example of the new SIP message generated by SIPp, after the source code modifications were made to enhance and modify the SIP message structure. Furthermore, in addition to generating legitimate INVITE messages, the Mr.SIP tool plays a crucial role in simulating multiple attackers and generating DoS/DDoS attacks through IP-spoofing techniques. The ability to mimic multiple attackers allows us to create realistic and challenging scenarios, simulating the presence of malicious actors attempting to overwhelm the SIP infrastructure with a barrage of malicious INVITE messages.

3.3 Training SIP data generation

The INVITE flooding attack is executed by an attacker who generates malicious INVITE messages, by spoofing the IP address and/or URI of legitimate users. The primary objective of this attack is to overwhelm the SIP server by not responding to the SIP 200 message with an ACK SIP message. Consequently, the server enters an ACK wait phase (timeout) while attempting to reestablish communication with the client through a SIP 200 message. The attack's effectiveness lies in the large volume of INVITE messages sent without corresponding ACK messages, leading to a denial of service for the SIP server.

The Mr.SIP tool uses IP spoofing attack to hide the true source of its malicious traffic, making it difficult for the target system to identify the attacker and detect the attack. This enables to evaluate the proposed framework's robustness

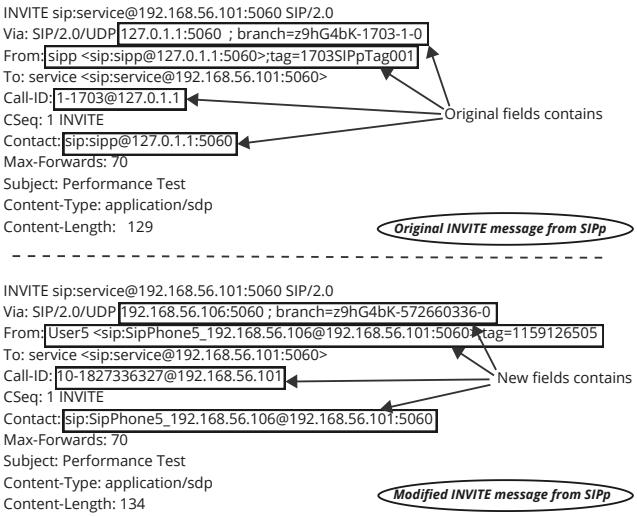

Fig. 3. Changes made to the content of INVITE message fields generated by SIPp to ensure the consistency of the generated data.

against sophisticated DDoS attacks that try to obfuscate their origins, mirroring real-world threat scenarios.

In the proposed dataset, the focus is on capturing the root cause of the attacks, which is the excessive volume of INVITE messages without ACK messages. To highlight this aspect, only the SIP transaction of the INVITE message is saved, omitting other related messages. This dataset structure allows for a more specific analysis of the attack pattern. From the literature [5], on the SIP-INVITE flooding rate, the attack can be classified flowing four categories:

Abrupt flooding behavior: the most common type of flooding attack in SIP networks. They occur when the SIP proxy or server receives a large number of requests in a short period of time, in an unusual way. This suggests that the attack does not have a specific rate or pattern.

Very high flooding behavior: the goal of this attack is to overload the system with a large number of INVITE requests in a short space of time. The system may then become unavailable or crash.

Stealthy flooding behavior: designed to evade detection by IDS. Attackers launch stealthy flooding attacks at a slow rate in order to avoid triggering any alarms or alerts. The low rate of this form of attack makes it extremely difficult to distinguish it from regular behavior, and it keeps the flooding rate slightly below the threshold.

Low rate flooding behavior: attacks involving floods can occasionally be launched at a slow rate. The low rate attack has roughly constant flooding rate, unlike stealthy behavior.

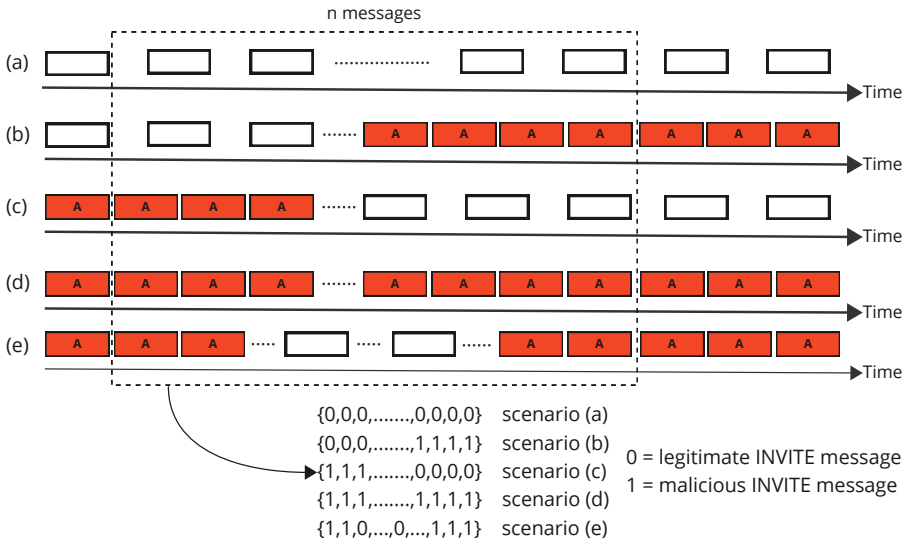

Fig. 4. The different generated sequence scenarios of INVITE traffic. (a) legitimate sequence, (b) a sequence that ends with malicious traffic, (c) a sequence that starts with malicious traffic, (d) a sequence containing nothing but malicious traffic, (e) a sequence containing legitimate traffic in the middle.

Giving this, the dataset contains various sequences and scenarios of the INVITE flooding attack, as depicted in Fig. 4. Additionally, it includes different sub-classes of attacks, providing further granularity and insights:

- Very low (VL): 10 messages/sec
- Low (L): 25 messages/sec
- Medium (M): 50 messages/sec
- High (H): 100 messages/sec
- Very High (VH): 500 messages/sec
- Very High ++ (VH++): 1,000 messages/sec

4 SIP-INVITE flooding detection

First, the methodology is described, including the sliding window technique and its specific parameters, the datasets distribution and the validation process designed to rigorously evaluate the proposed IDS. In addition, we discuss the selected features for our training model and describe the deep learning algorithms

used in our IDS. The effectiveness of our approach is measured, thus the performance metrics we used to evaluate the model's classification of malicious and legitimate SIP-INVITE traffic are described.

4.1 Analyze SIP traffic in time intervals

The proposed approach to detecting malicious sequences employs a sliding window technique with specific window size and forward step values. The sliding window size is fixed to 10 messages, corresponds to the "very Low" scenario rate flooding attack, e.g., 10 messages per second. This window size represents the worst case scenario. The assumption is to train the learning model to detect the "worst case" scenario, so that the model is able to detect other, less discrete attack scenarios. The forward step value p between two successive sub-sequences of transmission flows to be analyzed is set at 2, representing a shift of 2 messages per model inference, thus ensuring overlap between two successive analyses. This sliding window value is justified as the attacker may send only one malicious INVITE message to discover and identify the SIP server or proxy. However, once the attacker sends more than one malicious INVITE message, it indicates a potential flooding attack. By using a forward step of 2 messages, the proposed approach can effectively capture and identify the transition from potential to an effective flooding behavior. The Fig. 5 describes the used sliding window technique. Furthermore, regarding the encoding of SIP-INVITE sequences, a sequence is labeled as malicious if it contains a least 2 malicious messages. Otherwise, it is a legitimate sequence. This rule is explicitly outlined in Equation (1).

$$\forall\ x_i, x_j \in S;\ i \neq j;\ \text{ if } \exists\ x_i = 1 \text{ and } x_j = 1 \Rightarrow S = 1 \tag{1}$$

Where "x" represents an INVITE message and "S" is a set of 10 INVITE messages representing SIP-INVITE sequence (S = $\{x_0, \ldots, x_9\}$). Each "x" within the set "S" can take one of two values: "0" signifies a legitimate message, while "1" is a malicious message ($x \in \{0, 1\}$).

By configuring the sliding window with these specific parameters, e.g., 10 messages/sec and forward step 2, we enhance the detection capability of the proposed approach to accurately identify and classify malicious sequences in SIP communication. Indeed, the choice of a step of 2 messages can be seen as a pessimistic approach, as it places us in the worst-case scenario. However, this approach has several advantages. First, it allows us to detect attacks more quickly, the second, it is more robust to scenarios with fewer malicious messages in the sliding window.

4.2 Dataset distribution and validation protocol

To evaluate the performance of the proposed IDS system, the Train, Validation and Test protocol was adopted with distinct Train/Validation datasets and two

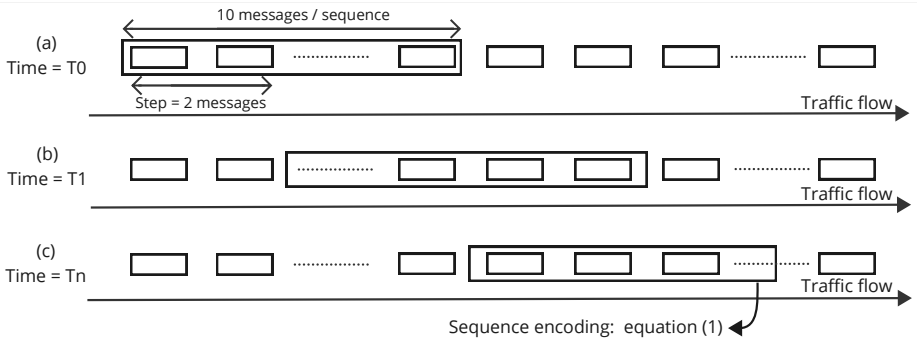

Fig. 5. Description of INVITE sequence detection approach: the sliding window size is fixed to 10 messages and the step forward to 2 messages per time unit. If the sequence contains at least 2 malicious messages, it is labeled as malicious. Otherwise, it is a legitimate sequence.

separate Test datasets. This protocol allows to assess the effectiveness of the system across various datasets and ensure its generalisation. The first Test dataset is generated on the same local network used for training. This enables to evaluate the performance of the IDS on typical data, guaranteeing its effectiveness in detecting DDoS attacks in known scenarios. The second Test dataset is generated in a other local network, i.e., the IP address of the SIP client and the SIP server are different from those used in the training phase. This scenario simulates new data (previously unseen), representing real-world type scenarios. Evaluating the IDS on this new data help to asses its generic aspect while detecting DDoS attacks in diverse SIP environments.

The dataset used for training contain 1,135,855 sequences, providing a substantial amount of data for the model. To ensure the effectiveness of the model, a separate validation set was created, consisting of 77,671 sequences. This validation set is used for the evaluation and fine-tuning of the model's performance during the training process. Additionally, a test set comprising 215,455 sequences was generated to assess the final performance and generalisation of the trained model. The test set used in this study is entirely independent of the train and validation datasets to ensure the reliability and generic aspect of the proposed framework.

Table 1 provides the classes distribution in each dataset. This table gives an overview of the distribution of the two different sequences classes (labeled as legitimate and malicious), present in the Training, Validation, and Test datasets. This information is critical for assessing the balance of the two classes within the datasets, which can impact the model's performance and lead to any required adjustments during the training and evaluation phases.

Table 1. Data distribution of the datasets used to evaluate the models. Dataset 1 is generated on a unique local network. Dataset 2 corresponds to the same scenarios, but where the SIP client and SIP server IP addresses are different from those used in the training phase.

	Dataset 1 (Training)			**Dataset 2 (Inference)**
	Train	**Validation**	**Test**	**Test**
Legitimate sequence	678,343	44,844	124,019	124,019
Malicious sequence	457,512	32,827	91,436	91,436
Total	1,135,855	77,671	215,455	215,455

4.3 Features selection

Table 2 presents the different main SIP fields, their corresponding data types in the dataset and the selected SIP fields as a features to be used by the model.

- Time (Temps): The reception time of the INVITE message is an essential feature for identifying the reception time of the sequence associated with potential DDoS attacks. By analyzing the reception times of INVITE messages, security analysts can identify patterns that may indicate a DDoS attack is in progress. For example, if a large number of INVITE messages are received within a short period of time, this may be indicative of a flooding attack.
- Source (IP address): This feature indicates the IP address of the client from the sent INVITE message. The source IP address is essential in pinpointing the origin of the message and help to identify suspicious or malicious sources of the DDoS attacks.
- From (SIP URI of the client): The "From" SIP-field includes the SIP Uniform Resource Identifier (URI) of the client, to identify the sender of a SIP request. This information helps in characterizing the clients involved in the communication and allow to distinguish legitimate clients from malicious attackers.
- Call-ID (URI of the call): The Call-ID field contains the URI of the session, providing a unique identifier for each session. This data is used for detecting patterns and correlations between different INVITE messages associated with the same session.
- Contact: This field identifies a SIP URI to which a SIP response can be sent. The SIP contact information can help us to know the client identity.

By selecting these fields as features for the training data, the models can learn and recognize patterns that reflect SIP-DDoS attacks. The combination of timing reception and the selected features (selected SIP fields) enhances the models' ability and effectiveness to differentiate between legitimate and malicious INVITE SIP traffic. The rationale behind excluding certain features lies in their lack of relevance for the considered scenarios. We carefully selected features to streamline the training phase, mitigates the risk of over-fitting and allow the models to focus on the most informative attributes.

Table 2. Dataset's column and selected features.

SIP field	Data type	Selected features
Seq	object	✗
Temps	object	✓
Source	object	✓
MSG	object	✗
Via	object	✗
From	object	✓
To	object	✗
Call-ID	object	✓
CSeq	object	✗
Contact	object	✓
Max-Forwards	int64	✗
Subject	object	✗
Content-Type	object	✗
Content-Length	int64	✗

4.4 Deep learning models

Within the IDS system, the implemented algorithms are: Recurrent Neural Network (RNN) [3], Long Short-Term Memory (LSTM) [4], and Gated Recurrent Unit (GRU) [2]. These algorithms are good candidates for sequence analysis and have proven to be effective in capturing temporal dependencies of data. By leveraging these algorithms, we can effectively analyze and detect malicious sequences in the SIP communication. The RNN architecture is designed to handle sequential data by processing information in a recurrent manner. LSTM, a variant of RNN, incorporates memory cells to capture long-term dependencies and prevent the vanishing gradient problem. Similarly, GRU also addresses the vanishing gradient problem while maintaining a simpler architecture compared to LSTM. By using RNNs, LSTM and GRU in the IDS-Sequence system, we can leverage their temporal modeling capabilities to capture the temporal properties of SIP data traffic. The pre-training span 10 epochs and used a batch size of 250. We implemented our approach using the PyTorch libraries and as our working environment we used Google Colab with "T4 GPU".

4.5 Evaluation metrics

To evaluate the proposed IDS, the following metrics are used: accuracy, precision, recall, and F1-score. These metrics are used for assessing the performance of our detection model for classifying malicious and legitimate SIP-INVITE traffic. In addition to the above metrics and for the sequences based approach, the Intersection over Union metric (IoU) and detection time are also used for the sequences classification. IoU is a typical metric used for tasks related to object detection and segmentation. It measures the overlap between the predicted

region and the ground truth region of an object. In the context of sequence classification, IoU can be utilized to determine the similarity between predicted and expected classifications.

Detection time is a critical metric used for measuring the time cost and efficiency of the proposed model for detecting malicious SIP-INVITE traffic. It assesses the delay between the time of the received sequence and the time the sequence is correctly classified as malicious or legitimate.

By exploiting these metrics, we provide an extensive and comprehensive performance evaluation for detecting and classifying the malicious and legitimate SIP traffic, with effectiveness.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{4}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{6}$$

Where True Positives (TP) correspond to the number of instances detected as true malicious traffic. True Negatives (TN) is the number of instances detected as false malicious traffic. False Positives (FP) is the number of instances the legitimate traffic is detected as a malicious traffic. And, False Negatives (FN) correspond to the number of instances the malicious traffic is detected as a legitimate traffic.

5 Evaluation

The results in terms of classification, based on the performance metrics, give a high rate for RNN, LSTM, and GRU for each dataset 1 and 2 (Accuracy = 99.9%, Precision = 99.9%, Recall = 99.9%, F1-score = 99.9% and IoU = 99.9%). Also, the performance gap between RNN, LSTM and GRU models is significantly low, with standard deviation of ± 0.02. Given this, we propose to focus the analysis based on the comparison of the other metrics: FP, FN and "misclassifed sequence" and "inference time".

The Table 3 presents the results obtained from the RNN, LSTM and GRU models for the two distinct datasets. The evaluation is performed based on misclassified sequences, inference metrics, False Negatives (FN), and False Positives (FP). As shown, GRU outperforms LSTM and RNN models in terms of misclassified sequences. In the dataset 1, the GRU model misclassifies only 63 sequences,

demonstrating its superiority over the RNN and the LSTM models, which reach respectively 102 and 109 sequences. For the dataset 2, the GRU model also outperforms LSTM and GRU, with 48 misclassified sequences, while the LSTM misclassifies 85 sequences, and RNN misclassifies 103 sequences.

Table 3. Malicious sequence detection results on the two datasets.

Dataset 1			
	RNN	**LSTM**	**GRU**
Misclassified sequences	102	109	63
Inference Time (10 messages)	700.00 µs	749.00 µs	763.00 µs
FP	0.05 %	0.1 %	0.05 %
FN	0.04 %	0.01 %	0.01 %

Dataset 2			
	RNN	**LSTM**	**GRU**
Misclassified sequences	103	85	48
Inference Time (10 messages)	695.00 µs	753.00 µs	757.00 µs
FP	0.05 %	0.06 %	0.03 %
FN	0.04 %	0.01 %	0.01 %

From the inference metrics, the RNN models showcase better inference times. For dataset 1, RNN processes 10 INVITE messages within 700 µs, whereas the GRU model is slightly higher, with a processing time of 763 µs. For dataset 2, RNN shows a lower inference time of 695 µs, while GRU gives 757 µs for the inference time. Regarding the False Positives (FP) parameter, GRU also outperforms RNN and LSTM, showcasing its effectiveness in minimizing FP. For dataset 1, GRU achieves an height performance results, with a False Positive rate of 0.05%, equivalent to the RNN. For dataset 2, GRU maintains its superiority, achieving a False Positive rate of 0.03%. In contrast, RNN reports a slightly higher False Positive rate of 0.05%, while LSTM shows the highest False Positive rate of 0.06%. Regarding the False Negatives (FN) metric, GRU emerges as the front-runner, showcasing its capability to minimize missed classification. For both datasets 1 and 2, GRU achieves a significant low False Negative rate of 0.01%, outperforming RNN by 0.04% while matching the performance of LSTM.

Table 4 provides a detailed and comprehensive illustration regarding the distribution of misclassified sequences within the dataset. It highlights specific sequence patterns that consistently pose challenges for the classification models. These specifics sequences patterns are the most frequently misclassified patterns compared to the others and need further investigations.

In addition, it shows the frequency of misclassifications for GRU, LSTM and RNN. "T-1" and "T-2" in the table indicate their performance for Test datasets 1 and 2. Particularly the sequences (1, 1, 0, 0, 0, 0, 0, 0, 0, 0) and

(0, 0, 0, 0, 0, 0, 0, 0, 1, 1) are of particular interest. Indeed, these sequences are somehow challenging for GRU, LSTM and RNN, as they are frequently misclassified. Furthermore, the sequence (0, 0, 0, 0, 0, 0, 0, 0, 1, 0) appears to be challenging for both LSTM and GRU, as they are also misclassified.

Table 4. Results for the frequency of pattern classification errors in dataset 1 (T-1) and dataset 2 (T-2).

Sequence pattern	**Train**	**Test**	**LSTM**		**GRU**		**RNN**	
			T-1	**T-2**	**T-1**	**T-2**	**T-1**	**T-2**
(1, 0, 0, 0, 0, 0, 0, 0, 0, 0)	500	95	0	7	0	0	21	29
(1, 1, 0, 0, 0, 0, 0, 0, 0, 0)	529	93	31	26	11	8	11	12
(1, 1, 1, 0, 0, 0, 0, 0, 0, 0)	628	110	9	6	0	2	0	0
(1, 1, 1, 1, 0, 0, 0, 0, 0, 0)	1,080	200	8	3	0	0	0	0
(0, 0, 0, 0, 0, 0, 0, 0, 1, 0)	31	7	5	5	7	6	0	0
(0, 0, 0, 0, 0, 0, 0, 0, 1, 1)	553	100	9	9	18	12	20	19
			62	56	36	28	52	54

In conclusion, the performance results highlight that GRU stands out as the most effective and efficient candidate for our classification and use case. Although the GRU model shows the lowest inference times (763 µs for Test dataset 1, 757 µs for Test dataset 2), the gap is not significant compared to the other models (LSTM and RNN) and does not impact the IoT-SIP sessions (few ms to few sec).

6 Conclusion

We have proposed a framework for detecting and mitigating SIP-INVITE flooding attacks in SIP based IoT infrastructure. We also demonstrate the effectiveness and robustness of our proposed approach through extensive and comprehensive experiments and evaluations under various simulated scenarios. The framework comprises a dataset generated exploiting SIPp and Mr.SIP tools, containing legitimate and malicious INVITE message simulations under various scenarios with different message rates. Scenarios range from 10 messages/s to 1000 messages/s, allowing comprehensive evaluation of the proposed framework. The core component of the proposed framework is a contribution for a GRU-based IDS designed to analyze incoming SIP traffic in real time. GRU constitutes a good candidate for capturing temporal patterns and characteristics in sequential data, while identifying INVITE flooding patterns. By leveraging our datasets for training, the GRU-based IDS is able to distinguish in real time and with high accuracy legitimate and malicious SIP-INVITE messages. The results confirm that the proposed approach and framework can be effectively deployed in SIP based IDS for IoT environments.

For short-term perspectives, we envision two directions: the first one is to investigate the impact of different step values on detection metrics, e.g., performance comparison for step value equal 5 and 10. The second direction is to deploy our proposed solution in real IoT based IDS platform and test-bed.

Acknowledgment

This work has been carried in the context of the project Beyond5G, funded by the French government as part of the economic recovery plan, namely "France Relance" and the investments for the future program.

References

1. Alvares, C., Dinesh, D., Alvi, S., Gautam, T., Hasib, M., Raza, A.: Dataset of attacks on a live enterprise voip network for machine learning based intrusion detection and prevention systems. Computer Networks **197**, 108283 (2021)
2. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
3. Elman, J.L.: Finding structure in time. Cognitive science **14**(2), 179–211 (1990)
4. Graves, A., Graves, A.: Long short-term memory. Supervised sequence labelling with recurrent neural networks pp. 37–45 (2012)
5. Hussain, I., Djahel, S., Zhang, Z., Naït-Abdesselam, F.: A comprehensive study of flooding attack consequences and countermeasures in session initiation protocol (sip). Security and Communication Networks **8**(18), 4436–4451 (2015)
6. Inayat, U., Zia, M.F., Mahmood, S., Khalid, H.M., Benbouzid, M.: Learning-based methods for cyber attacks detection in iot systems: A survey on methods, analysis, and future prospects. Electronics **11**(9), 1502 (2022)
7. Khalil, H., Elgazzar, K.: Leveraging blockchain for device registration and authentication in tsip-based phone-of-things (pot) systems. In: 2023 International Wireless Communications and Mobile Computing (IWCMC). pp. 1605–1612. IEEE (2023)
8. Kumari, P., Jain, A.K.: A comprehensive study of ddos attacks over iot network and their countermeasures. Computers & Security p. 103096 (2023)
9. Mahajan, N., Chauhan, A., Kumar, H., Kaushal, S., Sangaiah, A.K.: A deep learning approach to detection and mitigation of distributed denial of service attacks in high availability intelligent transport systems. Mobile Networks and Applications **27**(4), 1423–1443 (2022)
10. Meddahi, A., Drira, H., Meddahi, A.: Sip-gan: Generative adversarial networks for sip traffic generation. In: 2021 International Symposium on Networks, Computers and Communications (ISNCC). pp. 1–6. IEEE (2021)
11. Meshram, C., Lee, C.C., Bahkali, I., Imoize, A.L.: An efficient fractional chebyshev chaotic map-based three-factor session initiation protocol for the human-centered iot architecture. Mathematics **11**(9), 2085 (2023)
12. Mittal, M., Kumar, K., Behal, S.: Deep learning approaches for detecting ddos attacks: A systematic review. Soft computing pp. 1–37 (2022)
13. Nassar, M., State, R., Festor, O.: Labeled voip data-set for intrusion detection evaluation. In: Meeting of the european network of universities and companies in information and communication engineering. pp. 97–106. Springer (2010)

14. Nazih, W., Hifny, Y., Elkilani, W.S., Dhahri, H., Abdelkader, T.: Countering ddos attacks in sip based voip networks using recurrent neural networks. Sensors **20**(20), 5875 (2020)
15. Omolara, A.E., Alabdulatif, A., Abiodun, O.I., Alawida, M., Alabdulatif, A., Arshad, H., et al.: The internet of things security: A survey encompassing unexplored areas and new insights. Computers & Security **112**, 102494 (2022)
16. Pereira, D., Oliveira, R.: Detection of abnormal sip signaling patterns: A deep learning comparison. Computers **11**(2), 27 (2022)
17. Pereira, D., Oliveira, R., Kim, H.S.: Classification of abnormal signaling sip dialogs through deep learning. IEEE Access **9**, 165557–165567 (2021)
18. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.: Sip: session initiation protocol. Tech. rep. (2002)
19. SIPp: Sipp, https://sipp.sourceforge.net/
20. Stanek, J., Kencl, L.: Sipp-dd: Sip ddos flood-attack simulation tool. In: 2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN). pp. 1–7. IEEE (2011)
21. Tas, I.M., Unsalver, B.G., Baktir, S.: A novel sip based distributed reflection denial-of-service attack and an effective defense mechanism. IEEE access **8**, 112574–112584 (2020)
22. Yang, I.F., Lin, Y.C., Yang, S.R., Lin, P.: The implementation of a sip-based service platform for 5g iot applications. In: 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring). pp. 1–6. IEEE (2021)
23. Yang, S.R., Lin, Y.C., Lin, P., Fang, Y.: Aiottalk: A sip-based service platform for heterogeneous artificial intelligence of things applications. IEEE Internet of Things Journal (2023)