MAML AE-SAC: A Synergy framework of Meta-Learning and Adversarial Reinforcement Learning for Adaptive Intrusion Detection

Dang Nhat Duy
Hanoi University of Science and Technology
Hanoi, Vietnam
duydang312@gmail.com

Nguyen Linh Giang
Hanoi University of Science and Technology
Hanoi, Vietnam
giangnl@soict.hust.edu.vn

Bui Trong Tung
Hanoi University of Science and Technology
Hanoi, Vietnam
tungbt@soict.hust.edu.vn

Le Van Dong
Hanoi University of Science and Technology
Hanoi, Vietnam
donglevan1@hust.edu.vn

***Abstract*–Modern Intrusion Detection Systems (IDS) suffer enormous challenges dealing with sophisticated adversarial attacks and are challenged by their ability to cope with new attacks, especially under imbalanced conditions where sparse threats are often overlooked. This work introduces MAML AE-SAC, a novel framework that combines adversarial reinforcement learning with model-agnostic meta-learning (MAML)[1] to construct a robust and fast-adapting IDS. Trainining a Soft Actor-Critic (SAC) agent adversarially and adapting with MAML, our solution prepares the system to defensively counter both strong and infrequent threats. Test results with the NSL-KDD[2], AWID[3], and CICIDS2017[4] data sets confirm that MAML AE-SAC sets a new record, achieving significant progress compared to existing benchmarks with regard to detection precision and generalizability.**

***Keywords*–Intrusion Detection, Reinforcement Learning, Meta-Learning, MAML, Adversarial Learning, Cybersecurity.**

I. INTRODUCTION

Conventional Intrusion Detection systems, which work based on known attack signatures, are useless against unknown or smartly manipulated threats. On the other hand, Anomaly detection IDS based on machine learning perform better but are still brittle when attackers manipulate input to deceive them. These vulnerabilities point to the need for a new, robust approach to intrusion detection.

Apart from the successful application of deep learning methods in this field, several challenges still exist that require attention. On the one hand, deep learning models show a high sensitivity to the datasets used during training. On the other hand, the network intrusion detection datasets obtained from real network environments usually include a large amount of normal behavior data along with a small amount of attack behavior data, leading to a highly unbalanced dataset. This imbalance leads to an inadequate detection ability for some intrusions in deep learning models. On the other hand, the development of effective models faces significant challenges. It is well known that deep learning models have a large number of parameters, and the careful tuning of these parameters requires a large amount of both labor and time.

The emergence of deep reinforcement learning (DRL) has brought with it a new solution to solve the intrusion detection problem. Although these reinforcement learning techniques have achieved satisfactory performance in intrusion detection, there are still challenges in dealing with imbalanced data, multi-class cyberattacks and limited samples in certain types of attacks . This paper introduces a novel framework, MAML AE-SAC (Model-Agnostic Meta-Learning with Adversarial Environment Soft Actor-Critic), designed to overcome these multifaceted challenges. Our paper introduces a novel deep reinforcement learning model based on the Soft Actor-Critic (SAC) algorithm, enhanced by the integration of Model-Agnostic Meta-Learning (MAML), adversarial training, and the Synthetic Minority Oversampling Technique (SMOTE).

In this work, we propose a novel architecture that synergizes adversarial RL with meta-learning to enhance rare attack detection, improve adversarial resilience, and enable rapid adaptation to new threats. And we also perform some experiments to evaluate the effectiveness of this MAML integrated framework to the intrusion detection domain.

The remainder of this paper is structured as follows. Section II reviews related literature. Section III details the MAML AE-SAC framework. Section IV describe our experimental setup and present a comprehensive analysis of the results. Finally, Section V concludes the paper and discusses future work.

II. RELATED WORK

Intrusion detection has seen significant developments with a shift from traditional approaches to data-driven ones. The application of Deep Learning (DL) architectures, such as using Autoencoders to learn unsupervised features[5] and applying CNN-LSTM architectures to identify spatio-temporal behaviors[6], has shown to be highly accurate. Nonetheless, such architectures are inherently reactive and suffer from not being able to handle new attack varieties without undergoing rigorous retraining procedures.

In pursuit of proactive approaches, Deep Reinforcement Learning (DRL) has been incorporated, often under a competitive setting where there is a defender agent that counteracts an attack agent that chooses particularly challenging samples[7]. Such configuration leads to improved robustness. Follow-up works have further improved this method by utilizing advanced DRL algorithms, for example, Double Deep Q-Network (DDQN)[8], and by dealing with class imbalance with methods like using SMOTE implementation or challenge-aware reward schemes[9], [10]. The groundwork for this current work draws from such advanced approaches, with the previous model being that of the AE-SAC[10]. Such robust DRL systems are specific to particular data distributions, though, and suffer from difficulties with fast, few-shot adaptability to completely new attack behaviors.

Meta-learning, or "learning to learn," is an attractive solution to the problem of rapidly adapting to small datasets. The Model-Agnostic Meta-Learning (MAML) algorithm, as proposed by Finn et al.[1], is an important breakthrough in such areas. Instead of aiming for the optimization of a single optimal set of model parameters, MAML aims to learn a parameter initialization that can be rapidly adapted to new tasks using few examples and gradient updates. This line of work has been successful and spawned further advancements, leading to systems not only that meta-learn like MAML but also which meta-learn further aspects of the optimization process, even down to the updating rule itself[11], [12]. The core insight behind MAML—an versatile initialization well suited for rapid adaptation—serves as an underlying guiding principle behind our very own research work.

Research Gap. The existing body of work shows an important gap. The models based on deep learning lack proactivity, while the current reinforcement learning-based models are not adequately geared for prompt, few-shot adaptation when new threats are encountered. Meta-learning is a plausible remedy for enabling quick adaptation; but no effort is evident which combines this technique in an overarching and adversarial-based reinforcement learning architecture for intrusion detection framework. The present study fills the gap identified as well as establishing an intrusion detection framework which not just proves resilient against known types of attack but is also specifically trained for quick adaptation against new and dynamic threats.

III. METHODOLOGY

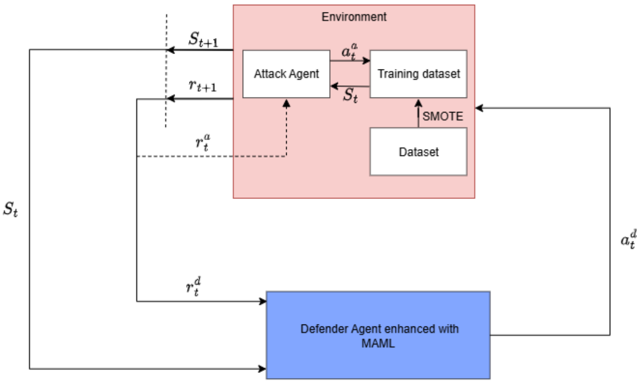

Figure 1: Overview of the MAML AE-SAC Framework.

A. MAML AE-SAC

Our paper proposes a Defender Agent enhanced with the MAML[1] method for network intrusion detection, adapted from the adversarial reinforcement learning model in[10]. As illustrated in Figure 1, the framework consists of two core components: an adversarial environment for data generation and a MAML-enhanced Defender Agent. The environment features an Attack Agent and uses a training dataset pre-processed with SMOTE[13] to oversample minority classes.

The system operates in an interactive loop. The Attack Agent (policy π_{a}) selects challenging data samples from the augmented training set. The MAML-enhanced Defender Agent (policy π_{d}) then classifies these samples. Both agents are implemented using reinforcement learning principles, receiving opposing rewards (r_t^a and r_t^d) to drive the adversarial dynamic. This process forces the Defender to learn a policy that is robust against difficult and rare attack types.

In this RL formulation, the *state* (s_t) represents a single network connection or packet, encapsulated as a feature vector containing attributes like protocol type, duration, and service. The *actions* are defined differently for each agent. For the Defender Agent, an action a_t^d is a classification decision from its action space, which includes all traffic categories (e.g., 'Normal', 'DoS', 'Probe'). For the Attack Agent, an action a_t^a is the selection of a specific attack type from the training set to challenge the Defender, aiming to find its weaknesses.

B. Reward Function

The adversarial nature of the framework is encoded in its reward functions, adopted from[10]. The functions are designed to incentivize the detection of difficult, low-frequency attacks.

$$r_{t+1}^d = \begin{cases} 0, & a_t^d \neq a_t, \\ 1, & a_t^d = a_t, a_t \in A_L, \\ 2, & a_t^d = a_t, a_t \in A_M, \end{cases} \tag{1}$$

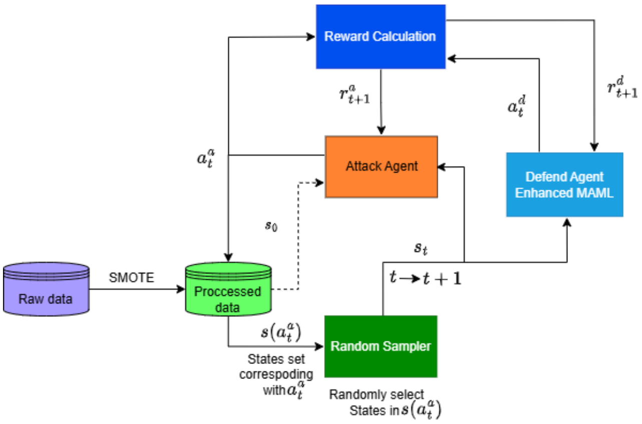

Figure 2: The adversarial sampling process.

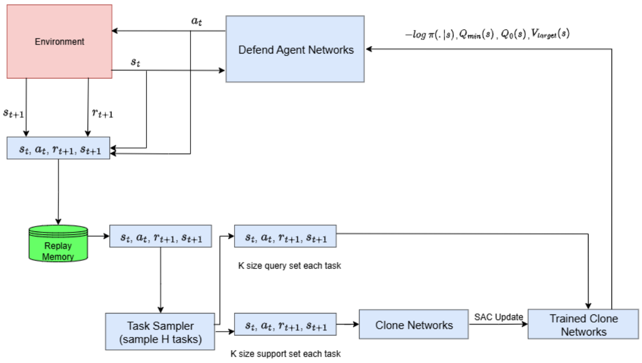
Environment
a_t
s_t
Defend Agent Networks
$-log\,\pi(.|s), Q_{min}(s), Q_0(s), V_{target}(s)$
s_{t+1}
r_{t+1}
$s_t, a_t, r_{t+1}, s_{t+1}$
Replay Memory
$s_t, a_t, r_{t+1}, s_{t+1}$
$s_t, a_t, r_{t+1}, s_{t+1}$
K size query set each task
Task Sampler (sample H tasks)
$s_t, a_t, r_{t+1}, s_{t+1}$
K size support set each task
Clone Networks
SAC Update
Trained Clone Networks

Figure 3: The MAML-SAC meta-update procedure for the Defender Agent.

$$r^a_{t+1} = \begin{cases} 0, & a^d_t = a_t, \\ 1, & a^d_t \neq a_t, a_t \in A_L, \\ 2, & a^d_t \neq a_t, a_t \in A_M, \end{cases} \tag{2}$$

Here, a^d_t is the Defender's action, a_t is the true label, A_L represents high-frequency (common) attack classes, and A_M represents low-frequency (minority) classes. The Defender is rewarded more for correctly identifying rare attacks, while the Attacker is rewarded more for finding rare attacks that the Defender misclassifies.

C. *Training process*

This research improved the training process in[10], which is illustrated in Fig 2. The environment agent selects an action label a^e_t from its policy π_e given the current state s_t. Based on this action label a^e_t, the next state s_{t+1} is selected from the training data, where the actual action label a^c_t is also received. Subsequently, the classifier agent's policy π_d determines the action label a^d_{t+1} given the state s_{t+1}. Following this, the environment agent receives a reward r^e_{t+1}, and the classifier agent receives a reward r^d_{t+1}.

Two experience tuples, $(s_t, a^\text{d}_t, r^c_{t+1}, s_{t+1})$ for the classifier agent and $(s_t, a^\text{a}_t, r^e_{t+1}, s_{t+1})$ for the environment agent, are accumulated into their respective experience replay memories . After collecting sufficient empirical data, both agents are updated using the Soft Actor-Critic (SAC)[14][15][16] update procedure to obtain new policies π_a and π_a. Moreover, with Defend Agent, it's training alternates between a standard SAC update and a specialized MAML meta-update, triggered periodically (controlled by the P frequency). This Meta-Learning Update step is demonstrated in Fig 3. The key difference between the two update mechanism starts with the Task Sampler which treats H most popular attack types from recent experiences as learning tasks, and then samples data from memory corresponding to these tasks. The sample set for each task includes a support set (D_support) and query data sets (D_query). Following this, a temporary copy of our Defender—the Clone Network—is trained on the support set using the standard SAC update.

Finally is the meta-optimization step, the query set is fed into the trained clone networks and then it's output is used to update the parameters of the original Defender Agent, training it to generalize and learn effectively from small quantities of new task-specific data. The environment agent then continues data sampling using the updated policy π_a.

The initial state s_0 is randomly selected from the training data. The policies π_a and π_d assign probability values to all possible actions, and actions a^e_t and a^c_t are selected based on these probabilities. According to[7], the action sets A_a and A_d in the NSL-KDD dataset[2] are not necessarily equivalent. The classifier agent's action set is defined as $A_d \in \{0, 1, 2, 3, 4\}$, corresponding to the number of classifier classes. In contrast, the environment agent's action set is defined as $A_a \in \{0, 1, \ldots, 22\}$, representing all possible attacks in the training dataset.

IV. EXPERIMENTAL RESULTS

A. *Experiment scenarios*

Datasets: Our methodology's effectiveness is evaluated by applying our approach to three well-known benchmark datasets: NSL-KDD[2], a classical dataset with severe class imbalance; AWID[3], a modern and large-scale dataset capturing 802.11 network traffic; and CICIDS2017[4], another popular large-scale benchmark.

Preprocessing: The preprocessing methods utilized include data cleaning, one-hot encoding for categorical data, and Min-Max normalization for numerical data. Notably, with respect to our NSL-KDD training data, we use SMOTE[13] to over-sample minority attack classes with less than 60 instances, thus increasing their representation to 3,000 instances. The purpose of this is to provide a more balanced starting view for the learning agents.

Evaluation Metrics: The metrics used include Accuracy, Precision, Recall, and F1-Score. Due to the common presence of class imbalance in the datasets, the F1-Score is used as our primary metric of overall performance.

Baselines: We compare MAML AE-SAC against a comprehensive set of baselines, focusing on the top-performing models from traditional ML, DL, and state-of-the-art DRL.

B. Performance Evaluation

We assessed our MAML AE-SAC architecture on three commonly accepted benchmark datasets. Across all experimental tests, our algorithm systematically achieved a new state-of-the-art accuracy, thus emphasizing the valuable benefits derived from combining meta-learning within an adversarial reinforcement learning framework.

In the established **NSL-KDD** data, where a severe imbalance exists amongst classes, MAML AE-SAC achieves a remarkable F1-Score of 83.23% (Table 1). This result, by virtue of exceeding not just the immediate predecessor's performance, AE-SAC (80.39%), but even that of the specially developed AESMOTE technique's optimised variant, 82.43%, points to a gain derived from the versatility cultivated by MAML, rather than by data augmentation methods alone. With regards to the complex and real-world **AWID** wireless network dataset, the proposed model proves to be better-performing, with an obtained F1-Score of 98.33% (see Table 2). Notably, it shows a highest precision across all models tested, with a reading of 98.55%. This higher precision has significant real-world relevance, since it corresponds to a lower rate of false alarm incidence, a major operational challenge faced by security operatives.

Finally, the comprehensive **CICIDS2017[4]** benchmark shows that MAML AE-SAC exhibits considerable scalability and robustness with an F1-Score of 99.11% (see Table 3). This very high score further supports its role as an excellent methodology, consistently outperforming powerful deep learning baselines and other reinforcement learning-based benchmarks, in a benchmark dataset covering a broad set of modern attacks. In conclusion, experimental results on various benchmarks provide disparate yet strong evidence that highlights the effectiveness and improved generalizability of the MAML AE-SAC algorithm.

Table 1: Performance on the NSL-KDD Dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Traditional Machine Learning				
RBF-SVM[10]	80.65	81.30	80.65	80.56
Linear-SVM[10]	75.60	76.22	75.60	72.95
Deep Learning				
DNN[10]	78.50	81.60	78.50	76.50
1D-CNN[10]	78.75	80.94	78.75	76.33
Deep Reinforcement Learning				
AESMOTE[10]	82.09	84.11	82.09	82.43
AE-SAC	81.01	81.23	81.01	80.39
MAML AE-SAC (Ours)	**83.98**	**83.50**	**83.98**	**83.23**

Table 2: Performance on the AWID Dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Traditional Machine Learning				
J48 (Decision Tree)[10]	96.26	96.20	96.30	96.30
Random Tree (RT)[10]	96.23	95.90	96.20	94.80
Deep Learning				
GRU[10]	95.27	93.65	95.27	93.67
1D-CNN[10]	95.37	94.24	95.37	93.57
Deep Reinforcement Learning				
SSDDQN[10]	98.19	98.40	98.19	98.22
AE-SAC	95.84	97.40	95.84	96.31
MAML AE-SAC (Ours)	**98.29**	**98.55**	**98.29**	**98.33**

Table 3: Performance on the CICIDS2017 Dataset.

Model / Study	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Traditional Machine Learning				
SVM[17]	95.50	97.72	99.12	98.40
Stacking-based Ens.[18]	98.15	98.20	98.15	98.16
Deep Learning				
CNN+LSTM[17]	97.16	97.41	99.10	98.24
Attentive Transformer[19]	97.03	97.02	91.10	96.97
Deep Reinforcement Learning				
AE-SAC	97.79	99.42	97.79	98.53
MAML AE-SAC(Ours)	**98.85**	**99.45**	**98.85**	**99.11**

C. Discussion

The consistent improvement in performance compared to the baseline standard AE-SAC across a number of datasets undeniably confirms benefits arising from its combination with MAML. While robustness is improved with adversarial training, using a meta-learning approach combined with a loss based on MAML leads to better generalization capabilities for the agent. Pretraining with MAML assists with "fast-learning." This leads to a better policy that shows robust generalization abilities across diverse scenarios, shows improved adaptivity, and functions with higher efficiency. This reflects higher overall F1-scores.

The obtained results are promising, but research does come with a certain set of limitations. The MAML AE-SAC algorithm is characterized by significant computational requirements, contains two agents and several stages included within its inner loop, leading to increased computational intensity during algorithm operation as well as practical implementation. In our case, concerning our own NSL-KDD dataset[2], our model's predictive performance falls short of predictive accuracy expectations for the U2R minor attack type (as shown in 5), the reason is due to restricted quantity of U2R instances which is only 52 samples included within our training data—with a less-than-ideal implementation of the SMOTE technique[13]. In contrast, for the AWID dataset[3], our model shows improved performance metrics for the two lowest data classes of injection and impersonation when compared to the larger flooding data type, as seen in 4.

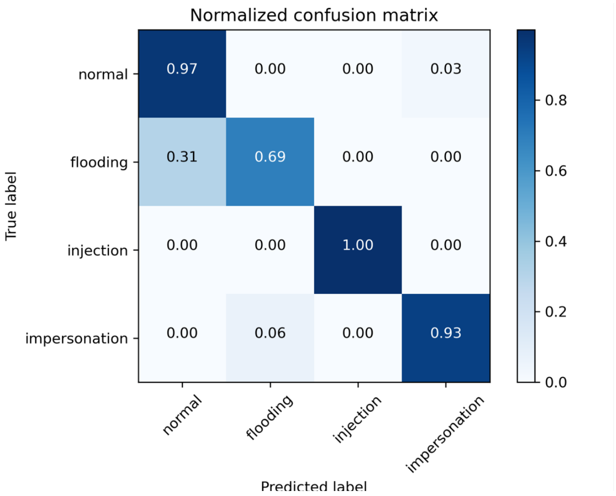

Figure 4: Confusion matrix of MAML AE-SAC on AWID dataset.

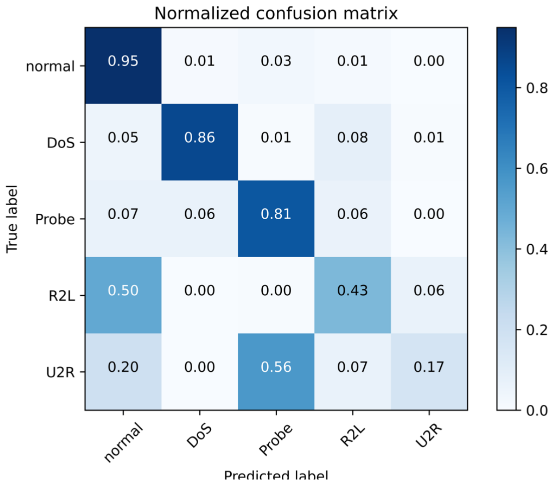

Figure 5: Confusion matrix of MAML AE-SAC on NSL-KDD dataset.

V. CONCLUSION

We present MAML AE-SAC, a new meta-learning intrusion detection framework, which combines adversarial reinforcement learning with ideas drawn from model-agnostic meta-learning[1]. While prioritizing goal-specific training with a view to supporting rapid adaptation, our proposed solution reveals a capability to effectively meet challenging threats with a ability to produce rapid reactions to novel and unexpected challenges. Extensive experiments on the NSL-KDD, AWID, and CICIDS2017[4] datasets verified our approach's superiority compared to a wide array of state-of-the-art methods.

Despite the promising results, the computational overhead of the dual-agent meta-learning framework presents a limitation. Future work will explore more efficient variants, such as Second-Order MAML for fast convergence but may require more computation, and investigate alternative oversampling techniques to further improve performance on extremely rare attack classes. This research paves the way for a new generation of proactive and intelligent Cybersecurity defenses.

ACKNOWLEDGEMENT

This research project is supported by the Bach Khoa Cybersecurity Center (BKCS) through project KC.01.04/21-30. We would like to thank BKCS and the Hanoi University of Science and Technology for their support.

REFERENCES

[1] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning*, PMLR, 2017, pp. 1126–1135.

[2] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, IEEE, 2009, pp. 1–6.

[3] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis, "Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 184–208, 2015.

[4] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy - ICISSP,*, SCITEPRESS, 2018, pp. 108–116. DOI: 10.5220/0006639801080116.

[5] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018.

[6] J. Zhang, Y. Ling, X. Fu, X. Yang, G. Xiong, and R. Zhang, "Model of the intrusion detection system based on the integration of spatial-temporal features," *Computers & Security*, vol. 89, p. 101 681, 2020.

[7] G. Caminero, M. Lopez-Martin, and B. Carro, "Adversarial environment reinforcement learning algorithm for intrusion detection," *Computer Networks*, vol. 159, pp. 96–109, 2019.

[8] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, "Application of deep reinforcement learning to intrusion detection for supervised problems," *Expert Systems with Applications*, vol. 141, p. 112 963, 2020.

[9] X. Ma and W. Shi, "Aesmote: Adversarial reinforcement learning with smote for anomaly detection," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 943–956, 2020.

[10] Z. Li, C. Huang, S. Deng, W. Qiu, and X. Gao, "A soft actor-critic reinforcement learning algorithm for network intrusion detection," *Computers & Security*, vol. 135, p. 103 486, 2023.

[11] Z. Li, F. Zhou, F. Chen, and H. Li, *Meta-sgd: Learning to learn quickly for few-shot learning*, 2017. [Online]. Available: https://arxiv.org/abs/1707.09835.

[12] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *International Conference on Learning Representations*, ICLR, 2017, pp. 1–11.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[14] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *International Conference on Machine Learning*, PMLR, 2018, pp. 1861–1870.

[15] T. Haarnoja, A. Zhou, K. Hartikainen, *et al.*, *Soft actor-critic algorithms and applications*, 2018. [Online]. Available: `https://arxiv.org/abs/1812.05905`.

[16] P. Christodoulou, *Soft actor-critic for discrete action settings*, 2019. [Online]. Available: `https://arxiv.org/abs/1910.07207`.

[17] M. Roopak, G. Y. Tian, and J. Chambers, “Deep learning models for cyber security in iot networks,” in *Proceedings of the 2019 IEEE Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2019, pp. 0426–0431.

[18] M. Ali, Mansoor-ul-Haque, M. H. Durad, *et al.*, “Effective network intrusion detection using stacking-based ensemble approach,” *International Journal of Information Security*, vol. 22, no. 6, pp. 1781–1798, 2023. DOI: `10.1007/s10207-023-00718-7`. [Online]. Available: `https://doi.org/10.1007/s10207-023-00718-7`.

[19] D. Z. Rodríguez, O. D. Okey, S. S. Maidin, E. U. Udo, and J. H. Kleinschmidt, “Attentive transformer deep learning algorithm for intrusion detection on iot systems using automatic xplainable feature selection,” *PLoS ONE*, vol. 18, no. 10, e0286652, 2023. DOI: `10.1371/journal.pone.0286652`. [Online]. Available: `https://doi.org/10.1371/journal.pone.0286652`.