Swarm-Driven Deep Intelligence: A Multi-Source Hybrid Framework for IoT Forensics and Threat Attribution

Mashıya Afroze F, Dr. V. Poornima
Research Scholar, Department of Computer Science, VEL'S Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India
Associate Professor , Department of Computer Science, VEL'S Institute of Science Technology and Advanced Studies, Chennai, India, Tamil Nadu.

***Abstract*—The exponential rise of the Internet of Things (IoT) has led to the interchange of vast amounts of data between network-connected devices. There are many security lapses and breaches as a result of the wide-ranging connectivity between IoT devices. Numerous benefits come with the quick expansion of IoT devices, but new security and forensics difficulties also arise. Digital investigators and practitioners face significant obstacles when interacting with IoT devices to probe cybercrimes in a timely and forensically sound manner, as a result of the vast amount of evidences generated by the billions of devices in the IoT system. The study's primary goal is to create a framework that performs forensic investigation on resource-constrained IoT using a combination of forensic technologies and machine learning to detect various kinds of attacks. The feasibility of Deep Neural networks (DNNs) in IoT Forensics (IoTF) is examined in this study to detect the attacks using the operating system logs. This work suggests utilizing an optimum set of parameters to train a Salps Swarm Optimization Algorithm (SSOA) for DNN. The suggested SSOA-DNN method is compared with the ML classifiers including KNN, RF, SVM, DT, LDA and NB Classifiers. The following metrics are used to evaluate how effective ML models are: (1) Accuracy, (2) Precision, (3) Recall, and (4) F-Measure. The results show that the SSOA-DNN outperforms with an accuracy of 96.37% than the other ML classification algorithms in IoTF Analysis.**

***Keywords*—Internet-Of-Things Forensics (IoTF), Machine Learning, Attack Prediction, Deep Neural Networks, Salps Swarm Optimization Algorithm (SSOA), Machine Learning**

I. INTRODUCTION

Internet of Things (IoT) devices are currently utilized in a wide range of contexts, including homes, offices, farms, hospitals, and even transit. They are becoming more and more interwoven into daily life [1]. There are 26.5 billion linked devices globally as of 2022, of which 16.4 billion are active IoT devices. It is anticipated that there will be over 30 billion IoT connections by 2025, translating to an average of nearly four IoT devices per person globally [2]. People's lives are now easier due to the growing usage of smart technology, which includes smart sensors that can read human minds and act accordingly. In the era of the IoT, millions of objects are connected to one another through communication networks, and this trend is expanding exponentially. The majority of regular customers are more focused on the advantages of IoT services than the difficulties and issues that come with privacy and security concerns [3]. As the IoT expands commercial prospects, it also brings with it a host of new difficulties pertaining to cyber-attacks [4]. The extraction of evidence from IoT devices, network communications analysis, the development of specialized tools and methodologies, the resolution of challenges brought on by the diversity of IoT devices, the improvement of the consistency and reliability of forensic evidence, and the standardization of IoT forensics practices are some of the key areas in this field[5].

A. IoT Forensics

IoTF is essential to many applications, such as incident response, criminal investigations, and legal proceedings. In order to investigate crimes, it is crucial for criminal investigators to retrieve digital evidence from IoT devices. Information gathering from different IoT devices and storage media is known as data extraction. The diversity of IoT devices presents a number of issues, including variations in operating systems, storage systems, and data formats. Researchers are addressing these variances and creating ways to handle them [6]. The

IoTF investigations can be performed on any of the three layers, Device layer, Network layer and cloud layer.

IoT and traditional forensics differ in that they focus on different areas and employ devices in different ways. Monitoring and detecting IoT devices is just one of the problems with IoT forensics. Other problems include jurisdiction, hazy boundaries, improper evidence management, protecting the chain of control, and a lack of standardization [7]. Because of this, IoT forensics is difficult because the data is distributed across multiple devices and is often multi-formatted.

B. Machine Learning in IoT Forensics

Research has demonstrated the effectiveness of ML in detecting abnormalities, classifying various malware types, and even projecting future assaults based on historical data, showcasing its adaptability in the context of constantly evolving cyberthreats[9].

The detection of attacks can be performed using various machine learning models, including Random Forest Classifier, Decision Tree Classifier, Naive Bayes Classifier, LDA Classifier , MLP Classifier, K-Nearest Neighbor and Support Vector Machines classifiers are a part of forensic analysis using ML models.

C. Optimization Algorithms in Machine Learning

The process of determining which configuration and parameter combinations work best for ML models to get the desired results is known as ML optimization. This is crucial to reduce misclassification errors, increase model accuracy, or improve other performance metrics.

Minimizing the loss function is one of the main goals of machine learning optimizations. Minimizing the loss function is one of the main goals of machine learning optimizations. In order to minimize the loss function, it is usual practice to update the model's parameters iteratively using optimization procedures like gradient descent and Meta heuristics algorithms. Machine learning models can benefit from the use of meta-heuristic optimization, which optimizes functions without making major predictions about the function's structure [8].

The goal of this study is to use IoT devices' logs to identify attacks on IoT devices using optimization algorithms and ML. The present study employs a Salps Swarm Optimization Algorithm (SSOA) and Deep Neural Networks (DNN) for IoTF Analysis in identifying the attacks. . The major contributions of the study is to

- Examine and assess the IoT device traces using ML methods
- Suggest a hybrid optimized ML based IoTF system using DNN and SSOA to detect the attacks on IoT devices.

II. METHODOLOGY

A third-party logging server handles the data acquisition and the traffic from IoT devices is routed to a logging server, which creates and stores logs and warnings of malicious attack traffic for use in forensic study. The information regarding attacks and attackers is gathered by regenerating and analyzing these stored logs in a forensic server. With the aid of a dataset produced from these logs, an automated machine learning technique is used to detect attacks.

A. Salps Swarm Optimization Algorithm (SSOA)

Swarm optimization algorithms, like the Salp Swarm Algorithm, are bio-inspired techniques based on swarm intelligence that solve optimization problems successfully by effectively exploring and exploiting search spaces. These algorithms imitate the collective behavior of creatures like salps. Salps have a translucent barrel-shaped body and are members of the Salpidae family. The algorithm simulates how salps move in a coordinated manner when they forage in a chain. While the followers take use of the space surrounding the leader, the leader searches the search space for the best areas. The salps travel across N-dimensional search space, and the following formula is used to update the leader's salp location according to the distance between the salp and the food source:

$$X_i^1 = F_i + r_1\big((ub_i - lb_i) * r_2 + lb_i\big) \qquad r_3 \geq 0 \quad (1)$$

$$X_i^1 = F_i + r_1\big((ub_i - lb_i) * r_2 + lb_i\big) \qquad r_3 < 0 \quad (2)$$

In equation (1) and (2) X_i^1 is the position of the leader (i^{th} position), F_i is the food position, $ub_i \; and \; lb_i$ denotes the upper and lower bound of the i^{th} position,

r_1, r_2, r_3 are the random numbers. Here r_1 establishes a balance between search space exploration and exploitation.

$$r_{1=2e^{(-\frac{4l}{L})^2)}} \quad (3)$$

In equation (3), l is the current iteration, while L is the maximum number of iterations. Newton's law of motion is used to update the position of followers.

$$x_i^j = \frac{1}{2} at^2 + V_0^t \quad (4)$$

$$x_i^j = \frac{1}{2} (x_i^j + x_i^{j-1} \quad (5)$$

In equation (4) x_i^j indicates the position of j^{th} follower in the ith position. V_0 is the initial speed and t are the iterations.

The direction and speed of an object can be controlled with the use of inertia. Nonetheless, this approach is effective for minor-dimensional unimodal and multimodal situations. However, it produces unsatisfactory results that show poor convergence for higher dimension issues. In order to arrive at the best answer, the position updates equation must be changed in each iteration [10].

B. Deep Neural Networks (DNN)

Deep neural networks are a further development of traditional artificial neural networks. Artificial neural networks with numerous hidden layers, known as DNNs, are more complicated and resource-intensive than traditional neural networks. For quicker training times, they function better with GPU-based architectures and are utilized in a variety of applications. Each node layer in a DNN trains on a distinct set of characteristics depending on the output of the layer before it. Therefore, the more you penetrate the net, the deeper you may move forward within it, and the more intricate features the nodes are able to identify. It does this by recombining and learning from the prior layer's features.

C. Proposed Method -DNN with SSOA

Due to the DNN model's black box nature, which prevents both its operation and the traits it learns from being inspected, a number of issues arise. In order to improve the accuracy of attack prediction, the PSO algorithm is employed to address the issue that the DNN is prone to falling into the local minimum value and that the number of hidden layer nodes is not set. The following is a summary of the steps in the proposed SSOA-DNN:

1. The SSA's potential individuals are initially selected at random, which represents the DNN's potential solutions.
2. Assessing the solutions' fitness. The bits of solutions that indicate the potential values of the connection weights and biases are given randomly generated values in this stage. The DNN then evaluates these solutions. Finding the optimal network structure for the DNN that, when applied to a training portion of the dataset, produces the increase in accuracy.
3. Modify each solution by shifting the search agents' locations within the SSA algorithm.
4. Until the halting condition is met, steps two and three are repeated. Creating the ideal DNN structure with the best weight settings and biases is the final step in the suggested model. The dataset's testing and validation sections are used to test this DNN structure. As a result, the MSE-derived error rate is the lowest.

D. Swarm Intelligence-Based Optimization for Anomaly Detection

A Swarm Intelligence Optimization Algorithm (e.g., SSOA) will be employed for hyper parameter tuning, optimization of weight distributions, and identification of relevant features to further improve the predictive capabilities of the DL model. Swarm intelligence is the model used to reproduce natural systems' collective behavior, enabling the model to converge quickly and efficiently to an optimal solution. Swarm intelligence improves the accuracy of anomaly detection while minimizing false positives by dynamically selecting relevant features and tuning model parameters. Consequently, the outcomes of the deep learning model's predictions and the forensic analysis will be performed with greater accuracy and reliability. Let the DL model have hyper parameters $\Theta = \{ \theta_{CNN}, \theta_{LSTM}, W, b\}$. the optimization objective is to minimize a loss function $L(\Theta)$

$$\Theta^* = \arg\min_{\Theta} L(\Theta) \quad (9)$$

Using a swarm intelligence optimization Algorithm (SSOA), particles P_k represent candidate solutions:

$$P_k(t+1) = P_k(t) + v_k(t+1) \quad (5)$$

$$v_k(t+1) = wv_k(t) + c_1 r_1 \left(p_k^{best} - P_k(t)\right) + c_2 r_2 (g^{best} - P_k(t)) \quad (11)$$

where v_k is velocity , p_k^{best} is particle's best position, g^{best} is global best , w inertia weight , c_1, c_2 acceleration coefficients and r_1, r_2 random numbers[18].

III. RESULTS AND DISCUSSIONS

The Python was used to conduct the research and the packages of NumPy, SciPy , Pandas and Scikit-learn were among the extension packages used to accomplish the data processing and machine learning assessment.

Experiments were performed on network flows, host logs and IoT/IIoT telemetry for the purpose of anomaly detection and cyber-attack classification performance. Quantitative metrics such as accuracy, precision, recall, F1-score, MCC, AUC-ROC, FPR, FNR and F2-score were used to assess the detection reliability, robustness and discrimination capability. Experiments with the baseline models of Random Forest, Isolation Forest, Autoencoder, SVM, and LSTM show the better performance of the proposed framework in detecting normal and malicious behaviors, and illustrate the effectiveness of the proposed framework on multi-source IoT forensic analysis.

Table 1 Performance Comparison of the Proposed IoT Forensic Framework

Method	Accuracy	Precision	F1-Score
Proposed Framework	99.62%	99.41%	99.62%
Random Forest (RF)	98.5%	98.7%	98.5%
Isolation Forest (iForest)	97.8%	97.5%	97.7%
Autoencoder (AE)	96.5%	96.2%	96.5%
Support Vector Machine (SVM)	95.0%	94.8%	94.9%
Long Short-Term Memory (LSTM)	94.0%	93.7%	94.0%

A. Findings

The hybrid optimized DNN- SSOA technique is used in IoTF Analysis to detect the attacks in the IoT devices. In ML, particularly advanced DL optimizations are crucial because they allow us to manage high-dimensional data, reduce overfitting, speed up training, handle non-convex optimization issues, and increase model performance. Models in ML/DL frequently contain millions or even billions of parameters. Optimizations facilitate the process of determining the best values for these parameters by effectively traversing the high-dimensional parameter space. Training ML/DL models on such massive datasets would be computationally impossible without optimization strategies. It also allows us to expedite the training procedure. ML models' precision, effectiveness, and potential for generalization can be improved by utilizing optimization approaches. Since algorithms are trained to carry out tasks in the most efficient manner, optimization is at the core of ML models. Optimizing hyper-parameters is essential to producing a precise model. The accuracy of the model and its capacity to accomplish particular tasks are directly impacted by the choice of model settings. Further, both over-optimized and under-optimized models are susceptible to failure, therefore it is crucial to get it right.

This study uses SSOA algorithm, which is a meta-heuristic algorithm that mimics the predatory behavior of salps by simulating their group joining end-to-end in the shape of a chain and moving in a sequential manner. It uses intelligence produced by elaborate processes like competition and cooperation among members of the biologic colony to efficiently solve optimization problems. It still uses a population-based global search approach as compared to evolutionary computation, but its velocity-displacement search model is straightforward and simple to use. Some drawbacks of the SSOA include its poor optimization capabilities and slow convergence time.

The primary advantage of the DNNs model used in this study is capable to handle unstructured and unlabeled data, which comprises the vast bulk of

data. The majority of data, particularly in the medical industry, is unlabeled and unstructured. DNNs are therefore suitable for processing this volume of data. Deep learning models use this to categorize photos, convert audio to text, forecast stock prices, and carry out a variety of other challenging tasks.

In contrast to conventional computer and network forensics, the field of IoT forensics is still relatively new. Due to the wide range of IoT devices, the quick advancement of technology, the absence of standards, and the restricted forensics capabilities included in many IoT devices, it faces certain particular difficulties. IoT devices range in size, shape, and configuration from massive industrial control systems to tiny sensors. They have different storage capacities, hardware architectures, operating systems, and communication protocols. The collection and evaluation of evidence is complicated by this heterogeneity.

IV. CONCLUSION

This study proposes an optimized ML based approach to improve the DNN architecture with an SSOA, swarm intelligence algorithm to detect the attacks in the IoT environment. In order to achieve good results in the experimental dataset, the DNNs architecture is optimized by integrating the SSOA method to determine the optimal hyper-parameters.

The accuracy of the DNN-SSOA methodology is 96.37% , which is higher than that of the DNN, RF, DT, KNN, SVM, MLP, NB, and LDA approaches.
The proposed novel multi-source IoT forensics framework which is able to efficiently integrate deep learning with swarm intelligence techniques to detect, characterize and report the anomalies in heterogeneous IoT environments[11]. The CNN-LSTM hybrid model structure is adopted in the framework to capture the spatiotemporal patterns, and the swarm intelligence is utilized to optimize the model parameters and the feature selection in order to enhance the prediction accuracy and the reliability. Future work will focus on addressing the limitations of the framework, such as deployment of the lightweight models on resource-constrained devices, integration of other data sources such as cloud logs, and cross-domain interaction of IoT devices. Transfer learning and incremental learning strategies can be explored in an order to deal with evolving attack patterns. Furthermore, combining the results with interpretability and AI techniques will increase the overall interpretability and reliability of forensic results.

REFERENCES

[1] Grispos G, Studiawan H, Alrabaee S. Internet of things (IoT) forensics and incident response: The good, the bad, and the unaddressed. Forensic Science International: Digital Investigation. 2024 Mar 1;48:301671.

[2] Lueth KL. State of the IoT 2020: 12 billion IoT connections, surpassing non-IoT for the first time. IoT Analytics. 2020 Nov;19(11).

[3] Stoyanova M, Nikoloudakis Y, Panagiotakis S, Pallis E, Markakis EK. A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. IEEE Communications Surveys & Tutorials. 2020 Jan 6;22(2):1191-221.

[4] Friedl S, Pernul G. IoT Forensics Readiness-influencing factors. Forensic Science International: Digital Investigation. 2024 Jun 1;49:301768.

[5] Ahmed, Abdulghani Ali, Khalid Farhan, Waheb A. Jabbar, Abdulaleem Al-Othmani, and Abdullahi Gara Abdulrahman. 2024. "IoT Forensics: Current Perspectives and Future Directions" Sensors 24, no. 16: 5210. https://doi.org/10.3390/s24165210

[6] Kouahla Z, Benrazek AE, Ferrag MA, Farou B, Seridi H, Kurulay M, Anjum A, Asheralieva A. A survey on big IoT data indexing: Potential solutions, recent advancements, and open issues. Future Internet. 2021 Dec 31;14(1):19.

[7] Alazab, Ammar, Ansam Khraisat, and Sarabjot Singh. "A Review on the Internet of Things (IoT) Forensics: Challenges, Techniques, and Evaluation of Digital Forensic Tools." (2023).

[8] Qureshi SU, He J, Tunio S, Zhu N, Nazir A, Wajahat A, Ullah F, Wadud A. Systematic review of deep learning solutions for malware detection and forensic analysis in IoT. Journal of King Saud University-Computer and Information Sciences. 2024 Aug 27:102164.

[9] Usman N, Usman S, Khan F, Jan MA, Sajid A, Alazab M, Watters P. Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics. Future Generation Computer Systems. 2021 May 1;118:124-41

[10] Alsalman, Dheyaaldin. "A comparative study of anomaly detection techniques for IoT security using adaptive machine learning for IoT threats." IEEE Access 12 (2024): 14719-14730

[11] https:// research.unsw.edu.au/projects/toniot-datasets