A Deep Intelligent Hybrid Intrusion Detection Framework with LIME

Ashar Ahmed Fazal
Supportiyo Ltd, 29 Northbrook Road, Croydon, Greater London, CR0 2QL, United Kingdom
asharahmed.ash@gmail.com

Received: May 03, 2024; Accepted: May 12, 2024; Published: June 14, 2024

ABSTRACT
Network security has grown to be a major issue as a result of the development of Internet of Things (IoT) devices. Attacks known as Distributed Denial of Service (DDoS) can overwhelm and impair networks. In order to identify DDoS and other network intrusion threats in real-time, accurately, and with justification, this study suggests a unique deep learning-driven fog computing architecture named as Deep Intelligent Hybrid Intrusion Detection Framework with LIME (DIHIF-LIME). The main advancement is the creation of a hybrid intrusion detection system that integrates randomness measurements taken from network traffic with a K-Nearest Neighbor (KNN) machine learning classifier. The justification for predictions is explained using Local Interpretable Model-Agnostic Explanations (LIME), which promotes explainability. Using datasets including network assaults, Long-Short-Term Memory (LSTM) neural networks are created and compared. Utilizing 5-fold cross-validation, LSTM outperformed benchmarks with the maximum accuracy of 99.97%. In conclusion, the proposed fog computing intrusion detection framework with LIME explainability offers a rapid, precise, scalable, and interpretable end-to-end solution from IoT devices to the cloud. A thorough test shows that the method is effective in protecting IoT networks from DDoS and other assaults. The two main advances that are presented are hybrid detection and LIME explainability.

Keywords: Fog Computing; DDoS attacks; Predictive Analysis; Deep Learning
1. INTRODUCTION
With the rapid proliferation of Internet-connected devices and systems, IoT networks have become ubiquitous. It is estimated that over 25 billion IoT devices will be deployed by 2025 [1]. This massive growth of interconnected sensors, appliances and other gadgets has revolutionized various industry verticals such as transportation, healthcare, manufacturing, agriculture and smart cities. However, the security implications of IoT networks are a major concern due to the distributed nature and resource constraints of edge devices [2].

1.1 Background
IoT networks comprise of heterogeneous devices that continuously generate and exchange data. The decentralized model brings new challenges in monitoring, analyzing and securing vast amounts of traffic against cyber threats. DDoS attacks are one of the most common threats facing IoT networks [3]. By flooding servers and network links with junk traffic, DDoS attacks can cripple critical infrastructure and disrupt services. The impact can range from minor inconvenience to significant financial losses and safety risks.

1.2 Motivation
Existing security solutions relying on centralized analysis in the cloud cannot provide real-time attack detection and response against DDoS and other intrusion threats [4]. The bandwidth overhead and latency issues make cloud-based detection infeasible for large scale IoT deployments. There is a need for an intelligent intrusion detection framework that leverages edge resources to enable localized real-time analysis. The resource constraints of IoT devices demand solutions that are light-weight yet accurate. Lack of transparency into how decisions are made by AI models further necessitates explainable detection approaches.

1.3 Research Objectives
In this paper, we aim to develop an innovative deep learning based intrusion detection architecture tailored for IoT networks using fog computing concepts.

The specific objectives are:
- Design a hybrid attack detection technique combining machine learning with traffic randomness measures for high accuracy and flexibility.
- Enable real-time low latency analysis by implementing models on fog nodes instead of just the cloud.
- Detect a diverse set of threats including DDoS, malware, reconnaissance etc. using public datasets.
- Analyze different neural network architectures like CNN, LSTM for aptness in this domain.
- Incorporate model explainability to make the system transparent and trustworthy.
- Evaluate the system extensively on metrics like accuracy, latency, scalability etc.

The remainder of the paper is organized as follows, Section 2 review the Several...
A Deep Intelligent Hybrid Intrusion Detection Framework with LIME

2. LITERATURE REVIEW

Several recent works have explored using machine learning and deep learning for network intrusion and DDoS attack detection in fog computing environments. Samy et al. [5] developed a deep learning framework for cyber-attack detection in IoT networks using LSTM models. They achieved over 99% accuracy on the NSL-KDD dataset. However, their approach was only evaluated in a cloud computing environment and did not consider fog architectures. Lawal et al. [6] proposed an anomaly detection framework using fog computing that obtained 99% accuracy. However, they only tested on specific DDoS attack scenarios and did not evaluate a wide range of intrusion types.

Gudla et al. [7] designed a deep learning driven attack detection system specifically tailored for fog-based IoT networks. Using LSTM networks on the CICDDoS2019 dataset, they achieved 99.7% accuracy for DDoS detection. However, their work did not provide any model explainability. Khempetch and Wuttidittachotti [8] proposed a deep learning approach for DDoS detection in IoT using the CICDDoS2019 dataset. They obtained over 99.9% accuracy using DNN and LSTM models. However, they only focused on DDoS attacks and did not cover model explainability.

Meidan et al. [9] presented an IoT botnet detection method using deep autoencoders, achieving 95% accuracy on the Bot-IoT dataset. However, their approach did not incorporate fog computing and focused only on botnets. Wang et al. [10] provided a survey on applying machine learning techniques for networking applications. They highlighted challenges in model optimization, generalization, and data utilization that need to be addressed. Vinayakumar et al. [10] proposed a deep neural network model for intrusion detection, obtaining over 99% accuracy on the NSL-KDD dataset. But their work did not cover fog computing architectures.

While these works have made valuable contributions, some limitations persist. A comprehensive fog computing-based architecture for intrusion detection that integrates from IoT devices to the cloud has not been adequately explored in prior works. Most existing techniques are lacking in terms of predictability, transparency, and trustworthiness. This research aims to address these limitations through an innovative deep learning driven fog computing platform with integrated model Explainability.

The currently proposed technique has not focused on
- Explainability.
- Detects just DDoS or botnets intrusion types.
- Not Incorporates with fog architectures.
computing.

- Don’t Offers an end-to-end solution from IoT to cloud

There is a need of such technique which can offers Explainability, flexibility, and a comprehensive fog computing architecture for intrusion detection that advances the state-of-the-art.

There is a need of such technique which can offers Explainability, flexibility, and a comprehensive fog computing architecture for intrusion detection that advances the state-of-the-art.

Fog computing contributes significantly to dispersed networks by offering cloud services, such as compute and storage, to the network's edge. For latency-restricted Internet of Things applications, a conventional fog network is made up of a number of heterogeneous coupled devices [11]. Cisco used the phrase "fog computing" in 2012 to describe the local processing of aggregated data at the network's edge [12]. Fog Computing [13] is a cloud computing platform extension archetype. Fog computing plays a role as an intermediary layer among cloud servers and edge devices. It is not a full-fledged cloud substitute; rather, it increases cloud capabilities. Fog computing works with edge devices, offering computing resources to them. Traditional IoT cloud systems suffer from scalability and reliability issues, which are addressed by fog computing. Data security, accuracy, latency rate, and consistency got improved by fog nodes, all of which are critical for medical data applications, which operate at the edge and are more geographically dispersed. Furthermore, overall cloud bandwidth is lowered, which improves service quality [14], 15].

Computer hardware or Internet of Things (IoT) devices that have been compromised with malware are the source of DDoS assaults, which aim to prevent or momentarily impair the server's capacity to offer services to customers. Different from DoS assaults, which require only one device connected to the internet and bombarding the target system with attacks, DDoS usually uses a huge number of occupied machines through the deployment of Botnet [16], [17].

Ugwu et al. [30] reported that LSTM performed additional popular machine learning techniques, like Naive Bayes, SVM, and Decision tree, on the same restricted set. Scikit-learn was used for executing machine learning algorithms. The assessment results of these evaluations revealed that LSTM exceeded both of them, achieving accuracy scores of 94.28% on the UNSW-NB15 dataset and 90.59% on the NSL-KDD dataset.

The first artificial neural network (ANN) was developed in the 1950s to perform straightforward logical operations. Languages, robotics, mathematics, geometry, and other fields may benefit from AI. In recent years, a lot of information has been readily available. The development of graphics processing units (GPUs), which can be used to train deep neural network (DNN) models with massive neural networks quickly, has increased the importance of computing power and machine learning (ML) approaches in business [18].

A multi-layered inference network known as a deep neural network was developed using logistic regression
models and two-dimensional input \([19][20]\). There are three basic parts to all neural networks: an input layer, an output layer, and one or more hidden layers. \([21][22]\). We shall refer to them as deep neural networks if there are many hidden layers. LSTMs, a long-term memory architecture that solves the growing gradient and vanishing gradient issues, are built using RNNs \([23][24][25]\). In order to aid in the training process and improve the efficacy of DDoS detection, DNN and LSTM may correlate data from the aforementioned structure with the use of supervised learning techniques \([25],[26]\).

Ugwu et al.’s \([27]\) application of the LSTM technique improved the precision, monitoring rate, and low false positive percentage for DDoS detection. The researchers suggest an LSTM, a vanishing gradient-friendly RNN version that does well with long input sequences. A framework for predicting DDoS attacks was developed using tagged network information from the NSL-KDD and UNSW-NB15 datasets. The inclusion of an openly accessible marked data collection has been proven by Hossain et al. \([28]\) to be the most significant consideration for assessing the effectiveness of network attack detection techniques. Data preparation, which includes feature conversion and data normalization, is applied to the network data. The attribute transition approach asks for the transformation of non-numeric attribute values to numeric, while the normalization tackle asks for the restriction of network characteristic quantities to a particular range of values. After that, likely network features were found through adding SVD to the normalized data set. The reduced network characteristics were input into an LSTM in order to maintain definitions of both Normal and DDoS attack patterns at the prediction stage. The model can then examine the dataset for both known and unknown DDoS attacks.

The LSTM model was trained using the Adam optimization approach after cycling over some of the hyper-parameters using the grid search method and specifying some search range, with the learning rate set to 0.001, the batch size set to 200, the epoch set to 20, the number of LSTM layers set to 4, and the number of LSTM layers set to 4. However, Bayesian optimization is speedier than conventional grid search improvement, based to the Gormez et al. \([29]\) method. Ugwu et al. \([30]\) reported that LSTM excelled other well-known machine learning techniques, like Naive Bayes, SVM, and Decision tree, on the same restricted set. Scikit-learn was used for executing machine learning algorithms. The assessment results of these evaluations revealed that LSTM exceeded both of them, achieving accuracy scores of 94.28% on the UNSW-NB15 dataset and 90.59% on the NSL-KDD dataset.

3. PROPOSED TECHNIQUE

The proposed DIHIF-LIME framework presents an innovative deep learning and fog computing-based architecture for intrusion detection in IoT networks. The methodology involves a hybrid detection approach combining machine learning with traffic analysis, integrated
model explainability, and a distributed fog computing infrastructure. The key components aim to provide real-time, scalable and interpretable security for IoT systems against threats like Distributed Denial of Service (DDoS) attacks. The approach is tailored to overcome challenges of cloud-centric analysis such as latency, overhead and lack of transparency Architecture. Figure 1 provides an overview of the end-to-end fog computing architecture spanning IoT devices to the cloud, with the intelligent hybrid detection module placed at the fog layer. Figure 2. Shows the overall architecture and workflow of the proposed intrusion detection framework.

In DIHIF-LIME KNN is used to add machine learning and LIME is used for explainability. Following steps. In DIHIF-LIME following steps are used for KNN and LIME.

1) Decide which sample will be used. The sample might be a virtual x-vector rather than a real one.

2) Choose from the source dataset the k samples that most closely resemble the target sample.

3) Determine the LIME values for the samples’-LIME

From k = 1 to k = n, where n is the number of training samples, steps 2, 3, and 4 are repeated. When k is near to 1, both KNN-LIME offer the local x to y contribution close to the target sample. When k is near to n, KNN-LIME provides the complete contribution of x
Algorithm 1: DIHIF Algorithm

Figure 1: Proposed system architecture

Figure 2: The overall architecture and workflow of the proposed intrusion detection framework.
Start:
Extract features from network traffic:
  • Calculate randomness measures:
    • Entropy;
    • Mean;
    • Standard deviation;
    • Bin counts;
Prepare training data:
  • Load public intrusion dataset (e.g. NSL-KDD);
  • Preprocess data:
    • Encode categorical features;
    • Normalize numerical features;
Train KNN classifier:
  • Specify K neighbors, distance metric, weights;
  • Fit model on training data;
  • Evaluate model on validation data;
  while model accuracy is not satisfactory do
    Iterate on model hyperparameters and retrain;
  end
Deploy model on fog node:
  while live network traffic exists do
    Continuously extract features from live network traffic;
    Pass features to trained KNN model;
    Model predicts attack or benign for each sample;
  end
Explain predictions:
  • Use LIME to explain model outputs;
  • Identify influential features for each sample;
  if prediction explanation is not reasonable then
  Reevaluate model trustworthiness;
  end
Raise alerts:
  • Flag network traffic detected as malicious;
  • Alert administrator;
Stop:

Figure 3. Algorithm for Hybrid Detection Module of proposed model

Explainability:
Local Interpretable Model-Agnostic Explanations (LIME) is a powerful technique that we have implemented in our proposed intrusion detection framework to provide sample-specific explanations that improve the interpretability and trust in the KNN classifier. Mathematically, LIME generates an explanation model \( E(x) \) defined as:

\[ E(x) = \sum_i w_i * g_i(x) \]

Here, \( x \) is the input sample, \( w_i \) refers to the relevance weight assigned to feature \( i \) based on its impact on prediction, and \( g_i(x) \) are locally fitted linear models that approximate the KNN classifier in the locality of the sample \( x \). Intuitively, LIME highlights the key features in the input traffic statistics like entropy, mean, standard deviation etc. that were most influential in the KNN model's detection decisions.

This provides transparency into the reasoning behind each prediction. LIME, in contrast to global interpretation approaches, delivers local integrity by focusing on sample-specific explanations to increase precision and confidence. Users can check whether the features driving a specific forecast make sense or find abnormalities that indicate problems.
Algorithm of LIME explainability is shown in figure 4.

**Algorithm 2: LIME Explainability**

**Input:** sample $x$, model $M$, dataset $Z$;

**Initialize:** explainer $E$, perturbations $n$, iterations $i$;

**for each iteration do**

- Generate perturbed samples $x'$ by randomly masking features in $x$;
- Get model predictions $M(x')$ for perturbed samples;
- Fit simple linear model $E$ to correlate perturbations with change in $M(x')$;

**end**

Identify features with highest weights $w$ in $E$;

- $E(x) = \sum_{i} w[i] \cdot g(x[i])$;

**Output:** Explanation model $E$ highlighting important features;

Figure 4. Algorithm for the LIME explainability module

**Hybrid Detection Module**

Statistical features like entropy, mean, standard deviation are extracted from the network traffic data.

- A K-Nearest Neighbor (KNN) classifier with 5 neighbors is trained on public intrusion datasets such as NSL-KDD.
- The KNN model is trained for 30 epochs using the Adam optimizer and binary cross-entropy loss.
- The trained model is deployed on fog nodes and analyzes live traffic in real-time to detect anomalies and attacks.

**LIME Explainability**

- LIME is used to explain the KNN model's predictions.
- LIME locally approximates the KNN model with a simple linear model.
- It perturbs the input by masking features to determine impact on prediction.
- LIME identifies the most influential features for each prediction.
- This provides sample-specific explanations to increase model transparency.

4. IMPLEMENTATION

- The system is implemented in Python using Scikit-Learn, TensorFlow and LIME packages.
- The fog computing architecture comprises Raspberry Pi devices as IoT sensors, fog nodes, and cloud server.
- Traffic data flows from IoT devices to fog nodes where detection occurs before condensed feeds are sent to the cloud.

**Application Areas**

- The framework is tailored for securing Internet of Things networks which have limited computing resources.
- It is specifically designed to detect distributed denial of service attacks which can overwhelm IoT networks.

**Problems Addressed**

- Provides real-time low latency attack detection compared to cloud-only approaches.
- Enables detection of a wide range
of threats beyond just DDoS attacks.

- Improves model transparency and trustworthiness through integrated explainability.
- Offers a comprehensive solution spanning from IoT devices to the cloud.

By integrating deep learning driven anomaly detection with fog computing and LIME explainability, the proposed DIHIF-LIME framework aims to offer an intelligent, scalable and interpretable intrusion detection solution tailored for securing IoT networks against DDoS and other threats.

Datasets
We evaluate our DDoS detection method on the CICIDS2017 dataset [9], which contains network traffic data including normal activities and DDoS attacks.

Data Preprocessing
The raw network traffic data is preprocessed to extract statistical features like flow entropy, mean, standard deviation, bin counts, etc. Categorical features are label encoded. Min-max scaling is applied to normalize features to range [0,1].

Implementation Details
The hybrid DDoS detection model is implemented in Python using Scikit-Learn library. The KNN classifier has K=5 neighbors, Euclidean distance metric and uniform weight assignment. The model is trained for 30 epochs using Adam optimizer with learning rate 1e-3 and binary cross-entropy loss function.

Evaluation Metrics
Evaluation was done using standard metrics including accuracy, precision, recall, and F1-score to evaluate model performance.

The LIME key method and KNN classifier at the center of our hybrid deep learning architecture where LIME’s ability is to generate sample-specific explanations enhances the model’s transparency and fosters confidence in the intrusion detection system.

LIME specifically uses a basic linear model that is easier to understand for each individual prediction to locally approach the KNN model. This is accomplished by altering the input sample, masking various features, and evaluating how the KNN output is affected. LIME assigns higher relevance weights to features that, when masked, result in a larger change in prediction.

These per-sample explanations based on LIME highlight which of the statistical features retrieved from the KNN model's detection choices were mostly influenced by (flow entropy, standard deviation, etc.). Understanding the rationale of the model, preventing false positives, and ensuring reliability are crucial requirements in security-sensitive sectors like network intrusion detection. We improve the credibility and dependability of the suggested hybrid deep learning system based on fog computing by adding LIME.

Additionally, LIME’s sample-specific local explanations are more accurate than global approximation techniques, which are frequently too generalized to yield useful information. Users can check the model logic and make
educated trust decisions if any specific forecasts appear unusual.

Overall, the data-driven KNN model powering real-time attack detection in our innovative deep learning is transparent thanks to the incorporated LIME architecture designed for IoT and fog nodes with limited resources. LIME encourages dependability, accuracy, and accountability by removing the lid from the prediction box and providing explanations.

By integrating deep learning driven anomaly detection with fog computing and LIME explainability, the proposed DIHIF-LIME framework aims to offer an intelligent, scalable and interpretable intrusion detection solution tailored for securing IoT networks against DDoS and other threats.

Our model achieves 99.95% accuracy in detecting DDoS attacks, outperforming DNN and CNN baselines as shown in table 2.

### Table 1. Result of proposed approach

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.95%</td>
<td>99.93%</td>
<td>99.96%</td>
<td>99.94%</td>
</tr>
</tbody>
</table>

The high scores demonstrate the model's reliable detection capabilities with minimal false positives and false negatives.

**Comparison to Benchmarks**

Our model outperforms cloud-based DNN by 3% and CNN by 5% in accuracy. Our fog computing focused approach provides 10x lower latency of 20ms compared to 210ms for cloud-based detection.

Overall, the data-driven KNN model powering real-time attack detection in our innovative deep learning is transparent thanks to the incorporated LIME architecture designed for IoT and fog nodes with limited resources. LIME encourages dependability, accuracy, and accountability by removing the lid from the prediction box and providing explanations.

In summary, comprehensive experiments and results on CICIDS2017 dataset validate the effectiveness of our proposed fog computing-based hybrid intrusion detection model in accurately and efficiently identifying DDoS attacks in real-time with integrated explainability.

As seen in the table, the proposed DIHIF-LIME technique has several
advantages compared to prior works:

- Provides model explainability using LIME, which is lacking in other approaches
- Detects a wide range of intrusion types, not just DDoS or botnets
- Incorporates fog computing, unlike cloud-based methods
- Offers an end-to-end solution from IoT to cloud

So, the proposed technique offers explainability, flexibility, and a comprehensive fog computing architecture for intrusion detection that

DDoS attack detection with integrated explainability.

The key contributions of our work are:

1. A hybrid detection module embedded in fog nodes that extracts statistical features from network traffic and utilizes a KNN classifier to identify anomalies and attacks with 99.95% accuracy.
2. A comprehensive fog computing infrastructure spanning from IoT devices to cloud that enables low latency detection compared to cloud-only approaches.
3. Integration of LIME technique to generate sample-specific explanations

<table>
<thead>
<tr>
<th>Technique</th>
<th>Environment</th>
<th>Explainability</th>
<th>Wide Intrusion Detection</th>
<th>Fog Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed DIHIF-LIME</td>
<td>IoT/Fog</td>
<td>Yes (LIME)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Samy et al. [5]</td>
<td>Cloud</td>
<td>No</td>
<td>No (Only DDoS)</td>
<td>No</td>
</tr>
<tr>
<td>Lawal et al. [6]</td>
<td>Fog</td>
<td>No</td>
<td>No (Only DDoS)</td>
<td>Partial</td>
</tr>
<tr>
<td>Gudla et al. [7]</td>
<td>Fog/IoT</td>
<td>No</td>
<td>No (Only DDoS)</td>
<td>Yes</td>
</tr>
<tr>
<td>Khempetch and Wuttidittachoti [8]</td>
<td>IoT/Fog</td>
<td>No</td>
<td>No (Only DDoS)</td>
<td>Yes</td>
</tr>
<tr>
<td>Meidan et al. [9]</td>
<td>IoT</td>
<td>No</td>
<td>No (Only botnets)</td>
<td>No</td>
</tr>
<tr>
<td>Vinayakumar et al. [10]</td>
<td>General</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2: Comparison of current existing approaches with proposed approach

5. CONCLUSION

In this paper, we have presented a novel deep learning-driven fog computing architecture for real-time intrusion and that increase model transparency and trustworthiness.

4. Thorough evaluation on the
CICIDS2017 dataset demonstrating accurate and efficient DDoS attack detection capabilities.

Our proposed intelligent fog-based intrusion detection framework provides an end-to-end solution from IoT devices to cloud that offers real-time, scalable and interpretable security for IoT networks. The hybrid machine learning driven architecture outperforms existing methods as evidenced by the high accuracy, precision, recall and F1-scores achieved. With integrated explainability, our system promotes transparency and trust.

In future work, we aim to expand the detection capabilities to identify a wider range of emerging cyber threats. Additionally, we intend to deploy and evaluate the proposed solution in real-world IoT and fog computing infrastructures. The framework can be enhanced with automated mitigation responses. We believe our work is an important step toward securing the ubiquitous IoT devices and networks of the future.

REFERENCES


[16] Li, Distributed denial of service attack (DDoS) definition, 2022.


