Vivekanand Education Society's Institute of Technology

V.E.S.
Since 1962

Department of Computer Engineering

Group No.: 01
Date: 02/08/2024

Project Synopsis Template (2024-25) - Sem V

AI-Powered Cyber Threat Hunting by Fine-tuning LLM for DDoS Detection

Dr. Mrs.Nupur Giri
Head of Department, Computer Engineering (CMPN)

Ronak Ajwani
V.E.S.I.T
2022.ronak.ajwani@ves.ac.in

Hannan Chougle
V.E.S.I.T
2022.abdul.chougle@ves.ac.in

Manit Khira
V.E.S.I.T
2022.manit.khira@ves.ac.in

Shreya Chhatwani
V.E.S.I.T
2022.shreya.chhatwani@ves.ac.in

Dr. Mrs.Nupur Giri
Project Mentor

Table of Contents

Abstract

Cybersecurity threats, particularly Distributed Denial-of-Service (DDoS) attacks, pose substantial risks to modern network infrastructures. Traditional detection methods often struggle with evolving threats, necessitating advanced AI-driven solutions. This project leverages supervised fine-tuning of state-of-the-art Large Language Models (LLMs), specifically Llama 3.1, Llama 3.2, and Mistral, using Low-Rank Adaptation (LoRA). By fine-tuning these models on the CICIDS 2019 dataset, we aim to enhance the accuracy, scalability, and real-time detection capabilities against network threats. Llama 3.2 emerged as the superior model, providing balanced trade-offs between accuracy, precision, recall, and latency. This solution improves real-time threat detection accuracy and reduces false positives, significantly enhancing cybersecurity defenses.

Introduction

Distributed Denial-of-Service (DDoS) attacks remain a critical challenge in cybersecurity, continuously evolving in complexity and scale. Traditional Intrusion Detection Systems (IDS), primarily rule-based or conventional machine learning-based, increasingly fail to adapt rapidly to sophisticated cyber threats. Recent advancements in Large Language Models (LLMs), originally intended for NLP applications, offer significant potential in cybersecurity through their powerful pattern recognition and anomaly detection capabilities. However, pre-trained general-purpose LLMs lack the necessary domain-specific expertise, necessitating fine-tuning for effective cybersecurity deployment. By applying supervised fine-tuning using Low-Rank Adaptation (LoRA), our project investigates the potential of fine-tuned LLMs to detect and classify network-based threats efficiently.

Problem Statement

- Traditional threat detection systems primarily rely on predefined signatures, thus failing against novel DDoS attack patterns.
- Existing AI-based approaches, while promising, often produce high false-positive rates and face significant computational overhead.
- General-purpose LLMs struggle with domain-specific classification, limiting their practical utility in real-time cybersecurity.

Proposed Solution

Our solution addresses these challenges through:

- Fine-tuning state-of-the-art LLMs (Llama 3.1, Llama 3.2, and Mistral) using the LoRA method to effectively classify DDoS attacks from network logs.
- Optimizing structured prompt engineering techniques to enhance the contextual understanding and accuracy of LLMs.
- Employing supervised training with the CICIDS 2019 dataset, resulting in substantial accuracy improvements and reduced computational resources.

Methodology / Block Diagram

1. **Data Collection and Preparation**
   - Acquisition of CICIDS 2019 dataset containing labeled network traffic logs.
   - Splitting the dataset into training (80%) and testing (20%) subsets.

2. **Data Preprocessing**
   - Feature selection based on cyber-threat relevance.
   - Data cleaning (handling missing values, duplicates, label standardization).
   - Scaling numeric data via StandardScaler.

3. **LLM Fine-Tuning**
   - Selection of LLMs: Llama 3.1, Llama 3.2, and Mistral.
   - Supervised fine-tuning using the CICIDS 2019 dataset on Google Colab with T4 GPU resources.

4. **LoRA Implementation**
   - Parameter-efficient fine-tuning (PEFT) via Low-Rank Adaptation (LoRA).
   - Selective fine-tuning of attention layers to enhance model specialization and computational efficiency.

5. **Structured Prompt Engineering**
   - Developing effective prompts tailored specifically for cybersecurity threat detection to improve model response quality and consistency.

6. **Model Evaluation**
   - Evaluation using metrics such as accuracy, precision, recall, F1-score, and latency.
   - Comparative analysis among the chosen LLMs.

7. **Deployment and Visualization**
   - Real-time network monitoring using the optimal fine-tuned model.
   - Visualization dashboard for real-time alerts and cybersecurity status overview.

Hardware, Software and Tools Requirements

- **Hardware Requirements:**
  - GPU (Google Colab with T4 GPU or equivalent)
- **Software Requirements:**
  - Hugging Face Transformers Library
  - LoRA Framework
  - Python (Pandas, NumPy, Scikit-learn)
  - Google Colab (GPU-enabled environment)
  - Weights & Biases for performance visualization
- **Dataset Requirements:**
  - CICIDS 2019 dataset (network logs)

Proposed Evaluation Measures

- **Accuracy**: Overall correct classification of network threats.
- **Precision**: Minimization of false positives.
- **Recall**: Detection completeness of actual threats.
- **F1 Score**: Balance between precision and recall for robust performance.
- **Inference Latency**: Time efficiency in real-time detection scenarios.

Conclusion

This research highlights the significant potential of supervised fine-tuning of LLMs using LoRA for effective and scalable DDoS detection. The fine-tuned Llama 3.2 model offers substantial accuracy improvements over conventional methods, ensuring efficient real-time threat detection. Future efforts will explore real-time deployment scenarios and integration with multimodal threat intelligence to further enhance practical applicability and cybersecurity resilience.

References

[1] M. Guastalla et al., "Application of Large Language Models to DDoS Attack Detection," USC, Los Angeles, CA, USA, 2023.

[2] Q. Li et al., "DoLLM: How Large Language Models Understand Network Flow Data to Detect Carpet Bombing DDoS," arXiv:2405.07638, 2024.

[3] Y. Chen et al., "Large Language Models for Cyber Security: A Systematic Literature Review," arXiv:2405.04760, 2024.

[4] M. Du et al., "DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning," ACM CCS'17, Dallas, TX, USA, 2017.

[5] A. Hekmati et al., "Correlation-Aware Neural Networks for DDoS Attack Detection in IoT Systems," arXiv:2302.07982, 2023.

[6] M. E. Ahmed et al., "DDoS Attack Detection Using Flow Entropy and Packet Sampling on Huge Networks," IEEE Transactions on Network and Service Management, vol. 15, no. 1, 2018.