NetFlow-Based DDoS Detection: Interim Report for Project B

Evan Ji
Mentor: Pending

1 Weekly Meeting Notes URL

(Provide a link to your weekly notes stored online as per requirements.)

2 Introduction

2.1 Problem Statement

DDoS attacks remain a persistent threat to network security. This research focuses on improving DDoS detection using NetFlow analysis on real-world datasets. The goal is to identify attack patterns and distinguish normal vs. malicious traffic to enhance cybersecurity defenses.

2.2 Research Questions

- How effectively can NetFlow analysis detect and characterize DDoS attacks?
- What are the most critical NetFlow features for distinguishing attacks?
- Can machine learning improve the accuracy of DDoS detection compared to traditional statistical methods?

2.3 Novelty and Significance

Our approach differs from existing work by focusing on NetFlow-based detection without requiring packet payload analysis. This enhances scalability, making it feasible for real-time deployment in security systems like firewalls and intrusion detection systems (IDS).

2.4 Personal Interest

I am particularly interested in cybersecurity and machine learning applications in network security. This research provides hands-on experience with real-world data and security challenges.

3 Related Work

Prior studies on NetFlow-based DDoS detection include:

3.1 Statistical Anomaly Detection

Carl et al. (2006), Feinstein et al. (2003) used statistical methods like entropy-based anomaly detection and CUSUM to identify network attacks. **Limitation:** These methods require manually set thresholds, which may lead to false positives.

3.2 Machine Learning Approaches

Doshi et al. (2018), Kallitsis et al. (2016) explored ML-based attack classification using supervised learning on NetFlow data. **Limitation:** Models trained on specific datasets may not generalize well.

3.3 ISP-Level Detection (AMON-SENSS - Tandon et al.)

Used binning-based anomaly detection for large-scale networks. **Limitation:** Lacks real-time response capabilities.

Comparison to Our Work: Unlike prior research, our approach integrates both statistical methods and ML models, testing their comparative effectiveness on a real-world FRGP dataset.

4 Research Progress Findings (So Far)

4.1 Research Goals and Methodology

Our approach consists of:

4.1.1 NetFlow Data Preprocessing

- Extract features like flow duration, packet rate, byte count, entropy of source/destination IPs.
- Normalize and clean data for ML training.

4.1.2 Statistical Anomaly Detection

- CUSUM (Cumulative Sum Control Chart) to detect traffic spikes.
- Entropy-based analysis to identify deviations in traffic behavior.
- Flow correlation analysis to detect unusual patterns.

4.1.3 Machine Learning Models

- **Supervised:** Random Forest, Gradient Boosting Machines (GBM).
- **Unsupervised:** K-Means Clustering, Isolation Forest.
- Compare model effectiveness using precision, recall, and F1-score.

4.2 Preliminary Results

4.2.1 Model Performance Comparison

The results indicate that Random Forest outperformed Logistic Regression in almost every metric:

- **Logistic Regression Accuracy:** 99.06%
- **Random Forest Accuracy:** 99.84%

Random Forest demonstrated a significantly lower misclassification rate, with only 4 total errors compared to 24 errors in Logistic Regression.

4.2.2 Confusion Matrix Insights

Model	False Positives	False Negatives	Total Errors
Logistic Regression	19	5	24
Random Forest	2	2	4

Table 1: Comparison of Confusion Matrices

4.2.3 Precision, Recall, and F1-Score

Model	Precision (DDoS)	Recall (DDoS)	F1-Score (DDoS)
Logistic Regression	0.99	1.00	0.99
Random Forest	1.00	1.00	1.00

Table 2: Performance Metrics

5 Next Steps for Research Project C

To refine and extend this work, we will:

- Enhance dataset labeling by cross-referencing NetFlow logs with NetScout alerts.

- Optimize machine learning models to reduce false positives and improve real-time detection speed.
- Implement additional machine learning algorithms, including XGBoost and GBM.
- Develop and implement feature extraction techniques to identify the most influential NetFlow attributes.
- Expand the number of data points used for training and testing.
- Test hybrid approaches that combine statistical anomaly detection with ML models.
- Deploy and evaluate detection models on a simulated live network.

6 Bibliography

References

[1] G. Carl, G. Kesidis, R. R. Brooks, and S. Rai, ”Denial-of-service attack-detection techniques,” *IEEE Internet Computing*, 2006.

[2] L. Feinstein, D. Schnackenberg, C. Bowers, and D. Kindred, ”Statistical approaches to DDoS attack detection and response,” *Proceedings of the DARPA Information Survivability Conference*, 2003.

[3] R. Doshi, N. Apthorpe, and N. Feamster, ”Machine learning DDoS detection for consumer IoT devices,” *IEEE Security and Privacy Workshops*, 2018.

[4] M. Kallitsis, M. Antonakakis, R. Perdisci, and D. Dagon, ”A study on the effectiveness of machine learning-based botnet detection,” *IEEE Symposium on Security and Privacy*, 2016.

[5] A. Tandon et al., ”AMON-SENSS: Scalable ISP-Level DDoS Detection,” *IEEE Transactions on Network Security*, 2021.