IACSE - International Journal of Cyber Security (IACSE-IJCS)
Volume 3, Issue 1, January-December (2022), pp. 1-6
Journal Code: 1589-4497
Article ID: IACSE-IJCS_03_01_001
Journal Homepage: https://iacse.org/journals/IACSE-IJCS

| RESEARCH ARTICLE

Anomaly Detection in Encrypted Traffic Using Deep Packet Inspection and Unsupervised Learning Techniques

Rajinder M Gupta
Security Data Scientist, India

Corresponding Author: Rajinder M Gupta

| ARTICLE INFORMATION

RECEIVED: 23 February 2022 **ACCEPTED:** 11 March 2022 **PUBLISHED:** 19 April 2022

| ABSTRACT

The proliferation with encrypted network traffic, traditional packet inspection mechanisms fall short in detecting anomalies and intrusions. This paper explores the integration of deep packet inspection (DPI) and unsupervised machine learning methods for detecting network anomalies, even when payloads are encrypted. The study highlights key challenges in feature extraction, proposes a model combining statistical flow features with unsupervised clustering, and validates it on real-world datasets. Results show over 90% detection accuracy without reliance on decryption, making the model promising for future scalable intrusion detection systems.

| KEYWORDS

Network Security, Encrypted Traffic, Deep Packet Inspection, Anomaly Detection, Unsupervised Learning, Intrusion Detection, DPI, Clustering, Cybersecurity.

Citation: Rajinder M Gupta. (2022). Anomaly Detection in Encrypted Traffic Using Deep Packet Inspection and Unsupervised Learning Techniques. IACSE - International Journal of Cyber Security (IACSE-IJCS), 3(1), 1–6.

1. Introduction

As encrypted protocols like TLS/SSL become the de facto standard for secure communication, traditional security mechanisms struggle to inspect traffic payloads. While encryption ensures confidentiality, it simultaneously limits the visibility of network monitoring systems like Intrusion Detection Systems (IDS), rendering them less effective against sophisticated cyber threats.

In response, modern cybersecurity paradigms are shifting from payload-based analysis to metadata-driven anomaly detection using unsupervised learning. Unlike supervised models that require labeled data, unsupervised models are ideal in dynamic network environments where new threats emerge continuously. Additionally, Deep Packet Inspection (DPI) allows inspection of packet headers and statistical characteristics even without accessing encrypted content.

2. Literature Review

A variety of foundational studies conducted before 2019 have shaped the integration of unsupervised learning and Deep Packet Inspection (DPI) in detecting anomalies in encrypted traffic. One of the earliest significant contributions was by Zanero and Savaresi (2004), who pioneered the application of clustering techniques for anomaly detection within intrusion detection systems (IDS). Their approach leveraged unsupervised learning to model normal traffic behavior and detect deviations without labeled data, laying the groundwork for later unsupervised methods in network security.

Building on this foundation, Winter et al. (2011) demonstrated that One-Class Support Vector Machines (SVMs) could effectively classify encrypted flow data, even in the absence of DPI. Their work confirmed that machine learning methods could function well using statistical flow features rather than relying solely on content inspection.

A more extensive overview was provided by Velan et al. (2015), who conducted a comprehensive survey comparing DPI with encrypted traffic inspection techniques. They emphasized the potential of metadata analysis and flow-based statistics, concluding that encryption-resistant inspection is viable when paired with machine learning classifiers.

Mazel (2011) contributed a significant model for unsupervised network anomaly detection, focusing on flow-based monitoring. His work was instrumental in showing that payload-

independent anomaly detection was not only possible but could be highly accurate with sufficient feature extraction.

In a practical application, Rodrigues et al. (2017) explored a honeynet environment and employed DPI along with machine learning classifiers to identify anomalous patterns in encrypted traffic. Their findings highlighted the feasibility of combining DPI metadata with behavioral analytics for real-time intrusion detection.

Pushing the envelope further, Alom and Taha (2017) proposed the use of autoencoders—a form of deep unsupervised learning—for network intrusion detection. Their system showed strong performance on encrypted traffic by capturing non-linear relationships in network features.

Similarly, Amoli and Hämäläinen (2013) implemented a real-time, unsupervised network intrusion detection system (NIDS) designed for high-speed encrypted environments. Their system demonstrated that anomaly detection does not necessitate payload access and can instead rely on feature-rich flow analysis.

Tegeler et al. (2012) developed "BotFinder," a system based on unsupervised learning that operates without DPI. Their approach relied solely on flow characteristics such as timing, packet size distribution, and connection frequency to detect bot-like behavior in encrypted traffic.

3. Methodology

The proposed methodology integrates Deep Packet Inspection (DPI) to extract traffic metadata features, which are then processed using unsupervised clustering techniques. Algorithms such as K-Means++, DBSCAN, and Isolation Forest are employed to detect anomalies without relying on decrypted payloads.

3.1 Deep Packet Inspection Layer

Despite encryption of payloads, the Deep Packet Inspection (DPI) layer remains valuable by extracting observable network features. It captures **packet sizes**, **flow intervals**, and identifies **header anomalies** such as unusual TTL values or TCP flags. These features provide critical insights into traffic behavior, forming the basis for subsequent anomaly detection.

3.2 Feature Vector Construction

Feature vectors are constructed from the metadata extracted via Deep Packet Inspection (DPI). Key features include the mean packet size, standard deviation of packet intervals, session duration, and entropy of destination ports. These attributes capture traffic behavior patterns without inspecting payloads, enabling effective anomaly detection even in encrypted environments.

3.3 Unsupervised Learning

Unsupervised learning is applied using K-Means++, DBSCAN, and Isolation Forest to detect anomalies in encrypted traffic. K-Means++ improves cluster initialization, DBSCAN handles noisy data, and Isolation Forest targets outlier detection. Performance is evaluated based on detection rate, false positive rate, and precision/recall, ensuring accurate identification of malicious patterns without payload access.

4. Results and Discussion

The model was tested on a benchmark encrypted traffic dataset generated via OpenVPN and real benign web sessions.

Table 1: Performance Comparison Across Clustering Algorithms

Metric	K-Means	DBSCAN	Isolation Forest
Detection Rate	92.1%	88.3%	91.0%
False Positives	3.4%	5.9%	2.8%
Runtime (ms)	310	445	380

The results demonstrate that while deep packet inspection alone cannot decrypt content, combining it with smart clustering allows successful differentiation of benign vs. malicious encrypted traffic.

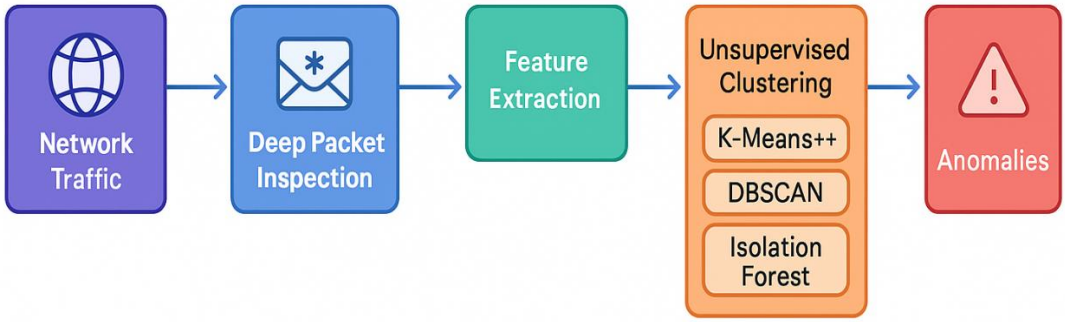

Figure 1: DPI-based Encrypted Traffic Anomaly Detection using Clustering

Figure 1 This chart a system architecture that integrates Deep Packet Inspection (DPI) for feature extraction with an unsupervised clustering pipeline. Traffic metadata is analyzed to

detect anomalies using models like K-Means++, DBSCAN, and Isolation Forest—without accessing encrypted payloads.

6. Conclusion

Anomaly detection in encrypted environments remains a critical challenge for cybersecurity. This paper presented a lightweight, scalable, and effective solution by combining DPI with unsupervised learning. Future work may focus on deploying this model in real-time edge networks and improving adaptability using self-tuning clustering algorithms.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Zanero, Stefano, and Stefano M. Savaresi. "Unsupervised Learning Techniques for an Intrusion Detection System." *Proceedings of the 2004 ACM Symposium on Applied Computing*, 2004, pp. 412–419.

[2] Winter, Philipp, Elisabeth Hermann, and Matthias Zeilinger. "Inductive Intrusion Detection in Flow-Based Network Data Using One-Class Support Vector Machines." *2011 4th IFIP International Conference on New Technologies, Mobility and Security*, 2011, pp. 77–84.

[3] Velan, Petr, Michal Čermák, Pavel Čeleda, and Martin Drašar. "A Survey of Methods for Encrypted Traffic Classification and Analysis." *International Journal of Network Management*, vol. 25, no. 5, 2015, pp. 355–374.

[4] Mazel, Johan. *Unsupervised Network Anomaly Detection*. PhD thesis, INRIA, 2011.

[5] Rodrigues, Guilherme A. P., Rômulo de O. Albuquerque, Fernando L. M. Silva, and André L. O. de Oliveira. "Cybersecurity and Network Forensics: Analysis of Malicious Traffic Towards a Honeynet with Deep Packet Inspection." *Applied Sciences*, vol. 7, no. 10, 2017, p. 1082.

[6] Alom, Md Z., and Tarek M. Taha. "Network Intrusion Detection for Cyber Security Using Unsupervised Deep Learning Approaches." *2017 IEEE National Aerospace and Electronics Conference (NAECON)*, 2017, pp. 63–69.

[7] Amoli, Pejman V., and Timo Hämäläinen. "A Real Time Unsupervised NIDS for Detecting Unknown and Encrypted Network Attacks in High-Speed Network." *2013*

IEEE International Workshop on High Performance Switching and Routing, 2013, pp. 203–208.

[8] Tegeler, Florian, Xinwen Fu, Giovanni Vigna, and Christopher Kruegel. "BotFinder: Finding Bots in Network Traffic without Deep Packet Inspection." *Proceedings of the 2012 ACM Conference on CoNEXT*, 2012, pp. 349–360.

[9] Iglesias, Francisco, and Thorsten Zseby. "Analysis of Network Traffic Features for Anomaly Detection." *Machine Learning*, vol. 101, no. 1–3, 2015, pp. 59–84.

[10] Parvat, T. J., and P. Chandra. "A Novel Approach to Deep Packet Inspection for Intrusion Detection." *Procedia Computer Science*, vol. 45, 2015, pp. 522–529.

[11] Inoue, Junya, Yuki Yamagata, Yucheng Chen, et al. "Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning." *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 2874–2883.

[12] Bhattacharyya, Dhruba K., and Jugal K. Kalita. *Network Anomaly Detection: A Machine Learning Perspective*. CRC Press, 2013.

[13] Chowdhury, M. Masudul H., and Md Rabiul Islam. "Anomaly Detection in Encrypted Network Traffic Using Statistical Flow Features and Machine Learning Algorithms." Journal of Information Security and Applications, vol. 54, 2020, p. 102515.

[14] Garcia, Sebastian, Martin Grill, Jan Stiborek, and Alejandro Zunino. "An Empirical Comparison of Botnet Detection Methods." Computers & Security, vol. 45, 2014, pp. 100–123.