Drift Mitigating Self-Learning System for Detecting Anomalous Encrypted Traffic

Snehal Mishra (22BIT0325)

Priyanshi Saraf (22BIT0649)

Abstract

With the increasing use of encryption in network traffic, conventional anomaly detection techniques are rendered ineffective, and there is a need for sophisticated solutions to Encrypted Traffic Anomaly Detection (ENTA). Machine learning algorithms, though promising, are plagued by concept drift, where changing data distributions compromise predictive accuracy over time. This paper introduces an adaptive self-learning system that counteracts drift while enhancing anomaly detection in encrypted traffic. The methodology integrates incremental learning, ensemble learning, and hypernetwork-based adaptation for improving model robustness. The system is tested with real encrypted traffic datasets to show its performance in high detection accuracy, even with changing network patterns. Real-time adaptation, labelling expense, and generalization towards new attack patterns are discussed as the main challenges. This work also investigates the consequences of inserting adaptive security frameworks within large network infrastructures. The findings point towards ongoing model development to mitigate the increasing complexity of cyber-attacks and provide robust anomaly detection processes. The research makes a contribution towards emerging adaptive cybersecurity models through suggesting scalable, drift-aware solutions for detecting anomalies in encrypted traffic.

Literature Review

Introduction

The exponential increase in network traffic caused by growing adoption of encryption has rendered encrypted traffic anomaly detection (ENTA) a key research direction in cybersecurity. Conventional machine learning models do not handle concept drift, i.e., changes in the mapping between input attributes and the target variable over time, which compromise model performance. To overcome this issue, self-learning systems have to include adaptive methods that detect and counter drift, maintaining anomaly detection accuracy in real time.

Concept Drift in Machine Learning

Concept drift is variations in the statistical characteristics of data over time that result in poor model predictions. Drift has three categories:

1. Gradual Drift: The relationship between input and output changes progressively over time, such as in spam detection, where spam techniques evolve gradually to bypass filters (Gama et al., 2014).

2. Sudden Drift: A rapid and unexpected change in the data distribution, such as a sudden spike in grocery sales during a pandemic, leading to immediate shifts in consumer behaviour (Widmer & Kubat, 1996).
3. Recurring Drift: Patterns that repeat periodically, such as the fluctuation in movie ticket sales on weekends due to consistent consumer behaviour trends (Žliobaite, 2010).

Concept drift detection and adaptation are necessary for the model to keep up with its accuracy in streaming data settings.

Anomaly Detection in Encrypted Traffic

As encryption is becoming a norm in network communication, conventional payload-based anomaly detection techniques lose their effectiveness. Encrypted Traffic Anomaly Detection (ENTA) then depends on metadata like packet timing, size, and flow behaviour (Bakhshi & Ghita, 2021). Deep learning techniques have been investigated in recent research for ENTA. Bakhshi and Ghita (2021) proposed a hybrid CNN-GRU model that enhanced detection accuracy in encrypted traffic. Static models are ineffective under drift conditions, and adaptive approaches are required.

Adaptive Methods for Concept Drift Mitigation

To effectively address concept drift, various adaptive machine learning techniques have been proposed, each catering to different drift scenarios:

1. Incremental Learning: This technique allows models to evolve over time by updating parameters continuously without requiring full retraining. Li et al. (2023) developed an autoencoder-based anomaly detection system that learns from incoming data streams, enabling real-time adaptation to changing traffic patterns.
2. Ensemble Learning: Adaptive ensemble methods dynamically update classifiers to mitigate the impact of drift. Some approaches introduce weighting mechanisms to prioritize recent models, ensuring improved accuracy in evolving environments.
3. Hypernetwork-based Adaptation: Zhu et al. (2023) proposed METER, a novel framework leveraging hypernetworks that generate model parameters dynamically in response to detected drift, reducing the risk of performance degradation.
4. Drift-aware Feature Engineering: Some approaches focus on continuously monitoring feature importance and adapting feature selection strategies accordingly. This minimizes the impact of drift by ensuring that models prioritize stable, informative features.
5. Self-Supervised Learning: Emerging techniques reduce dependency on labelled data by leveraging pseudo-labelling and contrastive learning, enabling models to adapt autonomously without extensive human intervention.

These adaptive methods enhance resilience in machine learning-based ENTA systems, ensuring sustained performance in dynamically evolving network environments.

Integration of Adaptive Learning in ENTA

Chen et al. (2025) introduced a self-evolving encrypted traffic classifier in the direction of drift, accommodating evolving traffic. Their system extended classifier lifespan without requiring labelled data, a critical challenge in network security. Analogously, Soltani et al. (2023)

investigated federated learning-based adaptable models, cutting down on reliance on centralized update while coping with drift.

Challenges and Future Directions

Despite advances in adaptive learning and drift detection, several challenges remain:

1. Real-time Adaptation: Many drift detection mechanisms introduce computational overhead, making real-time anomaly detection difficult. Future research should explore lightweight models and efficient update strategies to minimize latency while maintaining accuracy.
2. Labelling Costs: Many adaptive systems require periodic human labelling to maintain accuracy. Semi-supervised and self-supervised learning approaches should be further developed to reduce reliance on labelled data while preserving detection performance.
3. Generalization to New Attack Patterns: Adaptive models must detect zero-day attacks without prior knowledge. Current approaches struggle with unseen threats, necessitating research into continual learning techniques that can recognize emerging anomalies with minimal retraining.
4. Scalability in Large-Scale Networks: Deploying drift-aware security frameworks in large network infrastructures introduces challenges related to data heterogeneity and computational constraints. Future solutions must balance adaptability with efficiency to support real-world implementation.
5. Hybrid Approaches: Combining signature-based detection with adaptive machine learning models can improve robustness. Further studies should explore how integrating rule-based heuristics with AI-driven anomaly detection can enhance security frameworks.

Future research should aim towards self-supervised learning approaches for decreasing dependency on labelling, enhancing real-time adjustment, and achieving more scalable systems that guarantee reliable anomaly detection within encrypted traffic systems.

The findings from this literature survey underscore the necessity of adaptive anomaly detection in encrypted network traffic. While concept drift is a hindrance for standard models, recent approaches like incremental learning, ensemble models, and hypernetwork-based adaptation hold a lot of potential. These observations form the building blocks for designing an effective, self-learning, and drift-conscious anomaly detection system that can maintain high detection performance over a prolonged period.

References

1. Bakhshi, T., & Ghita, B. (2021). Anomaly detection in encrypted internet traffic using hybrid deep learning. *Security and Communication Networks*, *2021*, Article ID 5363750. https://doi.org/10.1155/2021/5363750
2. Chen, Z., Cheng, G., Li, J., Qin, T., Zhou, Y., & Luan, X. (2025). Drift-oriented self-evolving encrypted traffic application classification. *arXiv Preprint*. https://arxiv.org/abs/2501.04246
3. Li, J., Malialis, K., & Polycarpou, M. M. (2023). Autoencoder-based anomaly detection with incremental learning. *arXiv Preprint*. https://arxiv.org/abs/2305.08977

4. Soltani, M., Khajavi, K., Siavoshani, M. J., & Jahangir, A. H. (2023). A multi-agent adaptive deep learning framework for online intrusion detection. *Cybersecurity*, *6*(1), Article 10. https://doi.org/10.1186/s42400-023-00199-0
5. Widmer, G., & Kubat, M. (1996). Learning in drifting environments. *Machine Learning*, *23*(1), 69-101. https://doi.org/10.1023/A:1018046501280
6. Zhu, J., Cai, S., Deng, F., Ooi, B. C., & Zhang, W. (2023). METER: A dynamic concept adaptation framework for online anomaly detection. *Proceedings of the VLDB Endowment*, *17*(5), 794-807. https://doi.org/10.14778/3636218.3636233
7. Žliobaite, I. (2010). Learning under concept drift: An overview. *arXiv Preprint*. https://arxiv.org/abs/1010.4784