Ensemble Optimization for the Classification of Encoded Network Traffic Communication

Love Allen Chijioke Ahakonye, JaeHyun Lee, Cosmas Ifeanyi Nwakanma †, Jae Min Lee, Dong-Seong Kim
IT Convergence Engineering, † ICT Convergence Research Center,
Kumoh National Institute of Technology Gumi, South Korea
(loveahakonye, leejaehyun, cosmas.ifeanyi, ljmpaul, dskim)@kumoh.ac.kr

Abstract

The accelerating heterogeneity of encoded network traffic has increased complexity and difficulty in network ad- ministration. An efficient approach is vital to find the maximum objective function to detect and classify network traffic effectively. This study proposed an ensemble optimization (EO) approach to classify diverse encrypted network traffic. The proposed approach achieved better performance in classifying the diverse complex encrypted network traffic. The experimental results on the ISCXVPN2016 dataset demonstrate the usefulness and via- bility of the suggested optimization approach, which outperforms state-of-the-art optimization techniques used for secure network traffic analysis

Ⅰ. Introduction

Due to the recent development within the industrial internet of things (IIoT), intuitive data transmission methodology is swiftly rising. Furthermore, data transmission in such systems is made by various sensors [1] and transferred via broadband communications. High-speed internet with broad encryption modules is for a secure circulation of encoded data over end-to-end networks. Effective detection and classification approaches are required to increase network traffic monitoring for secure protection [2].

Given the rapid rise of encoded network data, traditional network data classification and single learners are becoming ineffective for reliably identifying communications. These improvements are critical for effectively analyzing complex encoded network packets. Broad bandwidth-based [4 - 7] and port-based [8, 9] conventional inspection methods are standard. Due to the deployment of encoding techniques such as secure sockets layer, transport security layer, and non-standard ports, these early detection approaches failed to classify networks precisely traffic[3]-[5]. As a result, machine learning (ML) techniques resolve the shortcomings of these classification methods.

Network traffic detection and classification are critical to minimizing intrusions and monitoring network traffic[2],[6]. There is a need to effectively manage the massive encoded traffic created by various network applications utilizing different encoding and ensemble methods. Complex encrypted network traffic can be noisy[3], necessitating a practical ensemble approach for precise detection and classification of the application types.

The ensemble parameter optimization *(EO)* as a meta optimization function is where each iteration of a specified hyper-parameter structure necessitates the training of a model[7], thus offering this benefit. The outcome is usually the best-optimized setting, with model training resulting in the optimal model parameter setup. Various strategies can combine the output of numerous weak learners into a single high-quality ensemble prediction.

Current network traffic classification methods employ individual learners to identify encoded network data. Nevertheless, classifying complicated composite network traffic that consists of more forms of encrypted files is challenging. In light of this, we suggested an ensemble optimization strategy to classify complex encoded network traffic into several reference network applications.

Specifically, this study focuses on the following:

1) Selection of the optimum ensemble optimization model based on detection accuracy, prediction speed, misclassification error (MCE), and model training time.
2) Classification of the encoded network traffic (ISCX TOR-nonTOR dataset) into various types of network application traffic, such as chats, streaming, peer-to-peer (P2P), email, voice over internet protocol (VoIP), and file transfer.
3) Compared the proposed approach's performance with some state-of-art ML optimization approaches for reliability tests.

Section I is followed by section II, which is the overall system model. Section III is the evaluation of the AI techniques. Section IV concludes the paper.

Ⅱ. System Methodology

Ensemble Optimization (EO): The prediction accuracy of a trained model is greatly influenced by optimizing the parameter. Since *(EO)* varies for different datasets and is not dependent on the training process, a meta-process for parameter optimization is required. It is a meta-optimization problem in which each attempt of a specific optimization framework necessitates training a model. Furthermore, the outcome is the effective parameter, and model performance yields the optimum solution.

The suggested strategy is to detect and classify encoded network traffic communication. Fig. 1 depicts the proposed scheme's workflow and system model. The optimizable ensemble learning on the MATLAB R2021b platform and the simulation parameters are in Table 1. The suggested architecture is in three major stages: model training, testing, and validation. In the train and test stage, the encoded traffic dataset, which contains a variety of network traffic

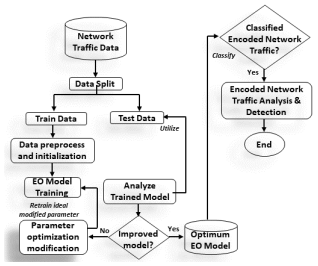

Fig. 1. Pipeline of the Proposed Ensemble Optimization for Detection and Classification of Encoded Network Traffic

generated by network applications using various types of encryption topology, is divided into the train and test batches and introduced to diverse ML techniques employing 5-K fold cross-validation. The test batch validates the performance of the model training. The performance indicators govern the best model selection based on the specified performance metrics. The list of the performance metrics includes the model's accuracy, training duration, and mis-classification error (MCE).

Table 1
Ensemble RUSBoost Network Parameters

Parameters	Settings
Ensemble Method	Optimizable Ensemble
Learning Rate	0.0764
Macimum number of splits	3065
Optimizer	Bayesian Optimization
Iterations	30
Number of Learners	34
K-fold Cross-validation	k=5
No of Observations	8044
Predictors	28
Response	8

Ⅲ. Performance Evaluation

Leveraging the ISCXVPN2016 dataset[8] and MATLAB R2021b, the proposed ensemble parameter optimization approach is for detecting and classifying encoded network traffic, and 30% of the dataset is for validation. The support vector machine optimization (SVMO), Naive Bayes optimization (NO), is compared for a reliability test of the proposed approach's performance. The comparison focuses on the parameter metrics regarding model training time, detection accuracy, prediction speed, and MCE. From the experimentation, the results demonstrate that the proposed EO had significant performance in a combined advantage of the parameter evaluation metric, as shown in Table 2.

Table 2
Performance Comparison of the Proposed Aproch with other State-of-the-art Techniques

parameter Metrics	EO	SVMO	NO
Accuracy(%)	99.4	94.5	70.9
Training Time (s)	185.8	10171	1243.5
MCE(#)	49	444	2344
Prediction Speed (obs/sec)	29000	28000	390

Ⅳ. Conclusion

This study proposed a parameter optimization approach for the efficient detection and classification of encoded network communication traffic. The experimental analysis considered the significant performance in accuracy, MCE, train time, and prediction speed. Nevertheless, considering the complexity of the encoded network traffic, the EO candidate was significantly outstanding based on the specified evaluation metrics.

ACKNOWLEDGMENT

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP by MSIT, Korea.Put sponsor acknowledgments.Put sponsor acknowledgments.

참 고 문 헌

[1] D.-S. Kim and H. Tran-Dang, "Industrial Sensors and Controls in Communication Networks," Computer Communications and Networks. Springer International Publishing, Cham, 2019.

[2] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Efficient Classification of Enciphered SCADA Network Traffic in Smart Factory Using Decision Tree Algorithm," IEEE Access, vol. 9, pp. 154 892 - 154 901, 2021.

[3] G. Li, M. Dong, K. Ota, J. Wu, J. Li, and T. Ye, "Deep Packet Inspection Based Application-Aware Traffic Control for Software Defined Networks," in 2016 IEEE Global Communications Conference (GLOBE- COM). IEEE, 2016, pp. 1 - 6.

[4] K.-S. Shim, J.-H. Ham, B. D. Sija, and M.-S. Kim, "Application Traffic Classification using Payload Size Sequence Signature," International Journal of Network Management, vol. 27, no. 5, p. e1981, 2017.

[5] D. Sanvito, D. Moro, and A. Capone, "Towards Traffic Classification Offloading to Stateful SDN Data Planes," in 2017 IEEE Conference on Network Softwarization (NetSoft). IEEE, 2017, pp. 1 - 4.

[6] H. Tran-Dang, S. Bhardwaj, T. Rahim, A. Musaddiq, and D.-S. Kim, "Re- inforcement Learning Based Resource Management for Fog Computing Environment: Literature Review, Challenges, and Open Issues," Journal of Communications and Networks, 2022

[7] L. A. C. Ahakonye, G. C. Amaizu, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Enhanced Vulnerability Detection in SCADA Systems using Hyper-Parameter-Tuned Ensemble Learning," in 2021 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2021, pp. 458 - 461.

[8] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of Encrypted and VPN Traffic using Time-Related," in Proceedings of the 2nd international conference on information systems security and privacy (ICISSP), 2016, pp. 407 - 414.