Bin LI, Yijie WANG, Li CHENG, 2024. Adaptive and augmented active anomaly detection on dynamic network traffic streams. *Frontiers of Information Technology & Electronic Engineering*, 25(3):446-460. https://doi.org/10.1631/FITEE.2300244

Adaptive and augmented active anomaly detection on dynamic network traffic streams

Key words: Active anomaly detection; Network traffic streams; Pseudo labels; Prior knowledge of network attacks

Corresponding author: Yijie WANG
E-mail: wangyijie@nudt.edu.cn
ORCID: https://orcid.org/0000-0002-2913-4016

Motivation

Active anomaly detection queries labels of sampled instances and uses them to incrementally update the detection model, and has been widely adopted in detecting network attacks. However, existing methods cannot achieve desirable performance on dynamic network traffic streams because:

- their query strategies cannot sample informative instances to make the detection model adapt to the evolving stream, and
- their model updating relies on limited query instances only and fails to leverage the enormous unlabeled instances on streams.

Main idea

- A prior-knowledge forest is constructed using prior knowledge of network attacks to find feature subspaces that better distinguish network anomalies from normal traffic.
- To make the model adapt to the evolving stream, a novel adaptive query strategy is designed to sample informative instances from two aspects: the changes in dynamic data distribution and the uncertainty of anomalies.
- Based on the similarity of instances in the neighborhood, we devise an augmented update method to generate pseudo labels for the unlabeled neighbors of query instances, which enables usage of the enormous unlabeled instances during model updating.

Framework

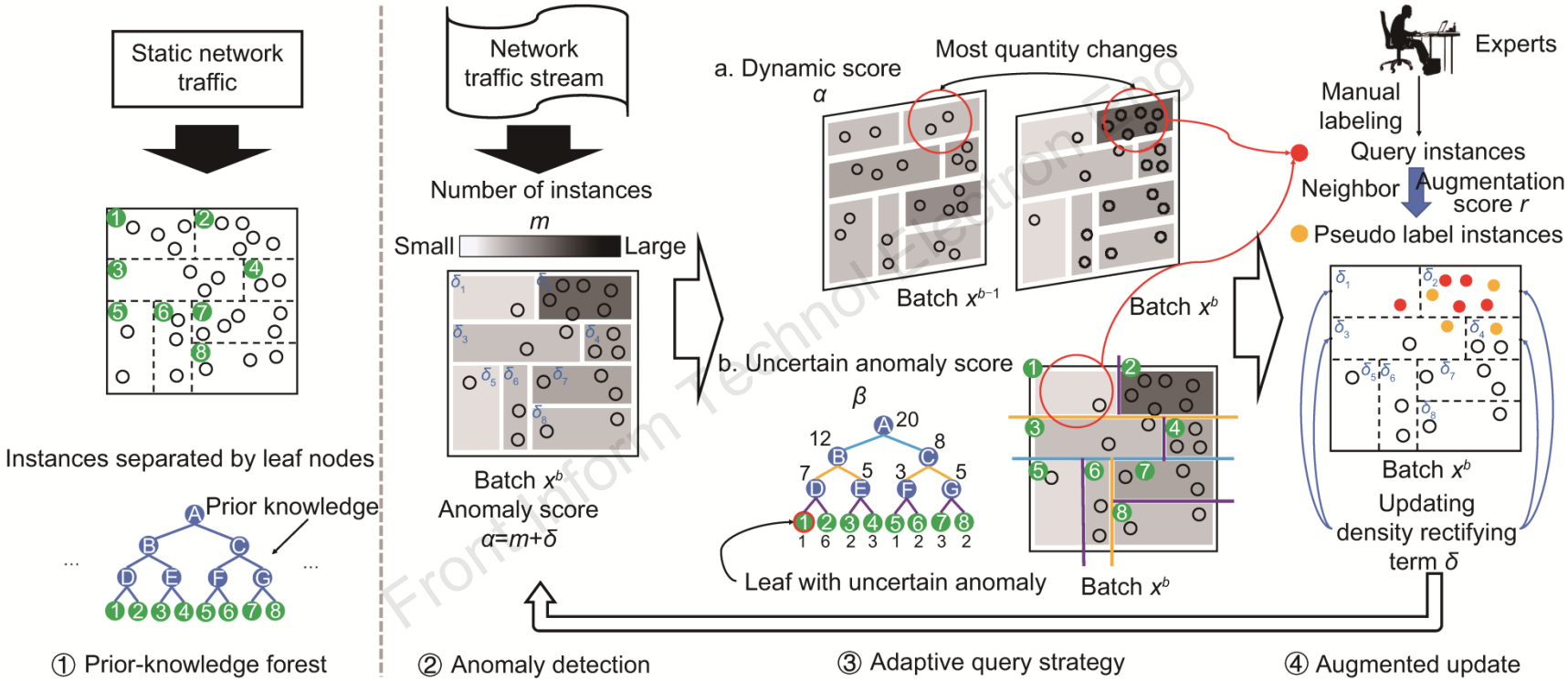


Fig. 1 Overview of A^{3}PF. The initialization stage, constructing prior-knowledge forest based on the static network traffic, is shown on the left of the gray dotted lines, while the testing stage on the network traffic stream is shown on the right. The white arrows show the relationship among anomaly detection, adaptive query strategy, and augmented update. The black arrows show that the network traffic is processed by the model. References to color refer to the online version of this figure

Method

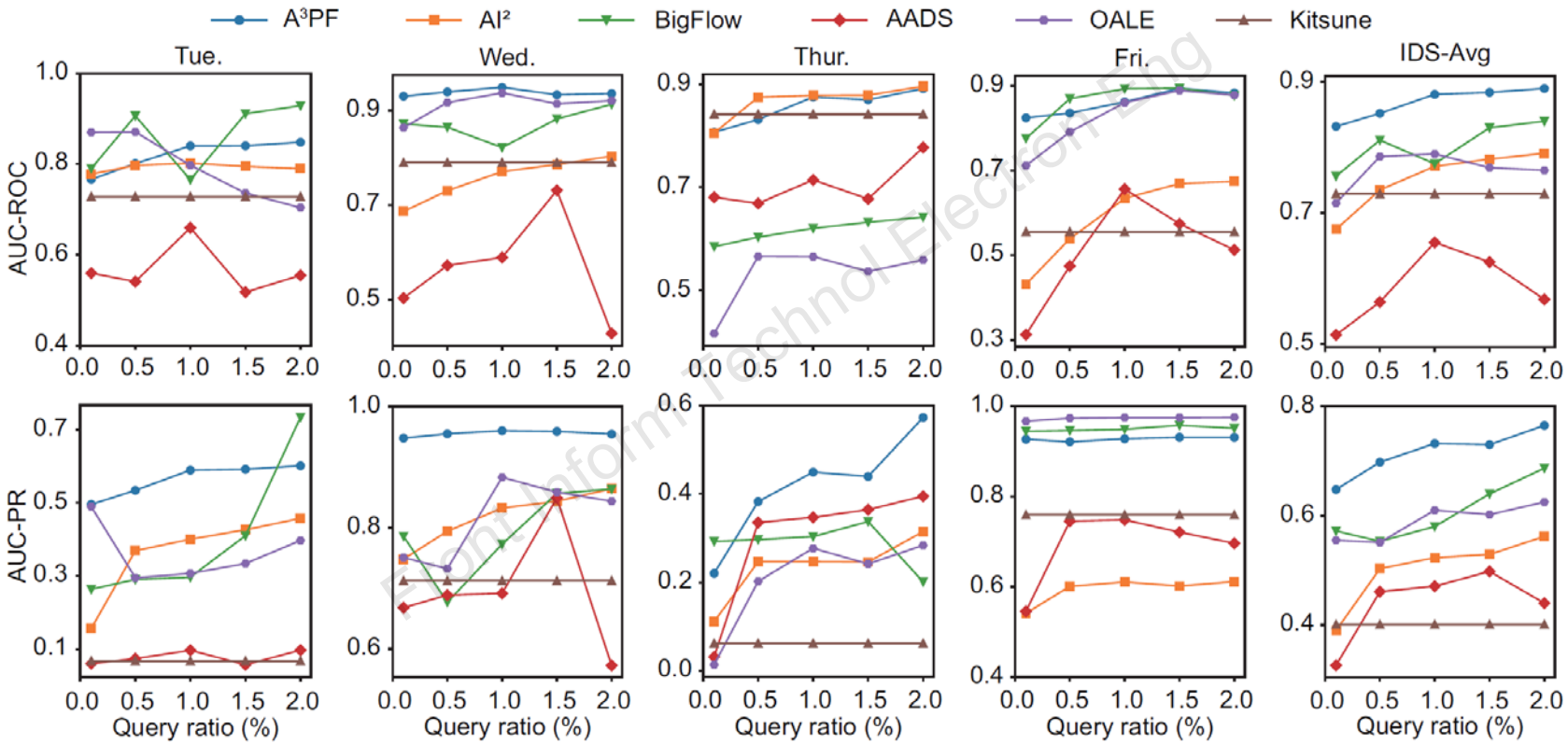
We first construct a prior-knowledge forest using prior knowledge of network attacks to find feature subspaces that better distinguish network anomalies from normal traffic.

Next, a novel adaptive query strategy is proposed to sample the most informative instances for manual labeling and model updating. Informativeness is measured in terms of the dynamic change of data distribution and the degree of uncertainty anomaly. Based on these informative instances, this strategy can make the model better adapt to network traffic changes

Method

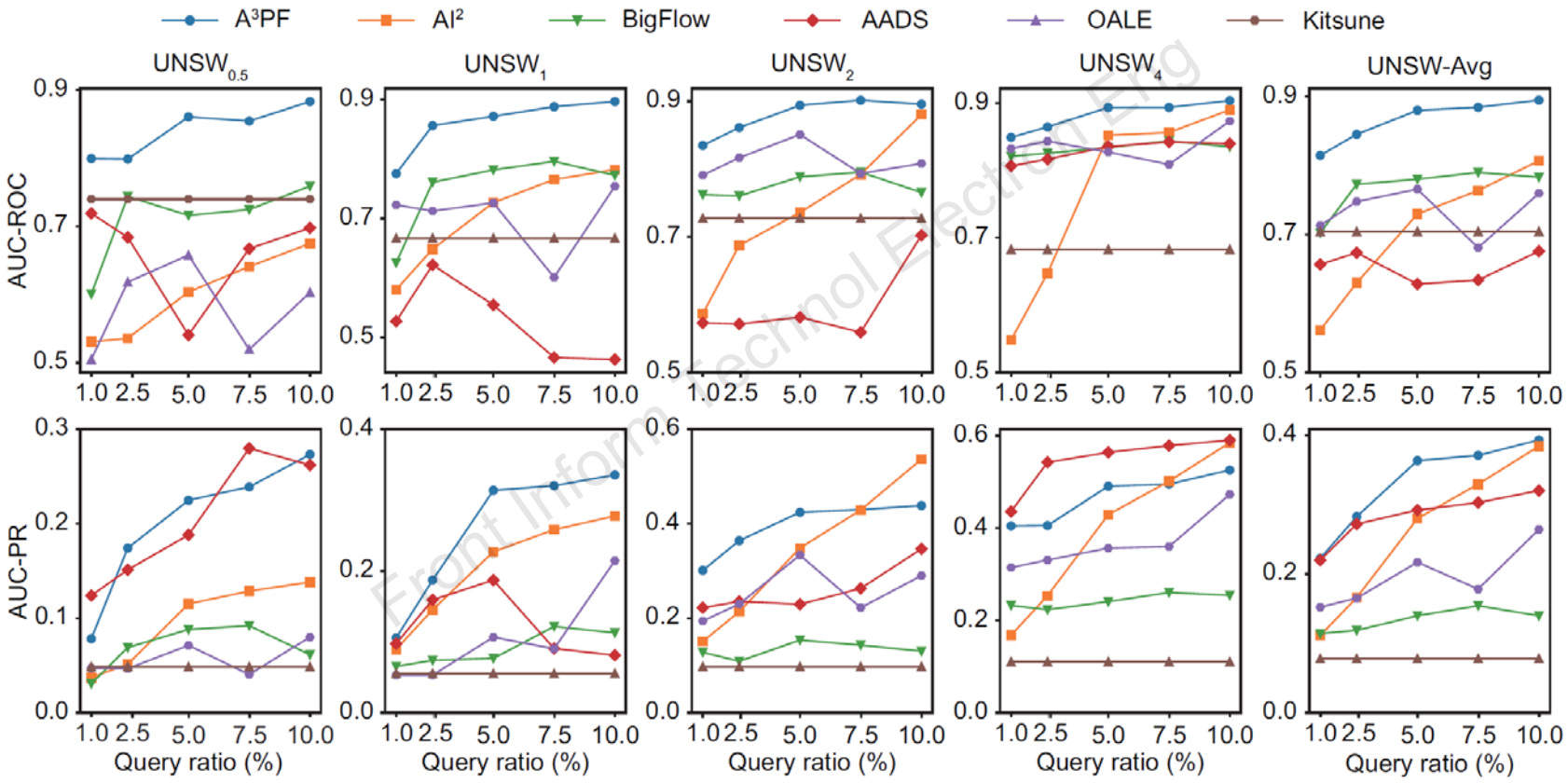
Finally, we further devise an augmented update mechanism to leverage the enormous unlabeled instances in model updating. Specifically, A^{3}PF generates pseudo labels for the unlabeled neighbors of query instances, assuming that the neighboring instances tend to belong to the same class. The instances with pseudo labels are used to incrementally update the model together with the query instances, which makes the detection model more generalizable.

Major results



AUC-ROC and AUC-PR w.r.t. query ratios and methods on CIC-IDS2017

Major results



AUC-ROC and AUC-PR w.r.t. query ratios and methods on UNSW-NB15

Conclusions

- A^{3}PF builds the forest model with prior knowledge of network attacks to detect anomalies based on local density.
- To sample the most informative instances of network traffic streams for manual labels, an adaptive query strategy is proposed to annotate limited instances, which makes the detection model adapt to the dynamic streams.
- To leverage the unlabeled instances in model updating, pseudo labels are generated for unlabeled instances and together with these query instances, incrementally updating the detection model.

Bin LI received the M.S. degree in computer science and technology from the National University of Defense Technology (NUDT), China, in 2018. He is currently a Ph.D. candidate in the College of Computer of NUDT. His current research interests lie in network intrusion detection and data stream classification.

Yijie WANG received the Ph.D. degree in computer science and technology from NUDT, China, in 1998. Currently, she is a professor of the National Key Laboratory of Parallel and Distributed Computing, NUDT. Her research interests include cloud computing, distributed storage, and big data analysis.

Li CHENG received the Ph.D. degree in computer science and technology from the College of Computer of NUDT, China in 2020. His current research interests lie in clustering analysis and outlier detection.