ARIA
APPLIED RESEARCH IN ACTION

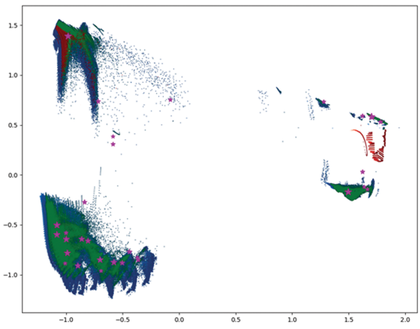
Network Traffic Anomaly Detection by Clustering with Autoencoder Ensembling

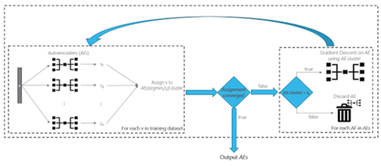
An expectation-maximization approach to detect network traffic anomalies in real-time by clustering network packets with autoencoders.

Xuduo Gu

Dehan Kong
ACADEMIC SUPERVISOR

Azharkhan Pathan
INDUSTRY SUPERVISOR





PROJECT SUMMARY

Cybersecurity has been an important type of anomaly for software-as-a-service (SaaS) providers as cyberattacks could cause serious consequences to both customers and the company. While different types of anomaly detectors exist, each kind may succumb to certain limitations. For instance, traditional rule-based detectors might miss novel attacks, and many machine learning (ML) methods use deep networks that require computation resources. In this work, an ML-based method is proposed to counter these problems. In this method, the statistical and individual features of each intercepted packet will be extracted and made into a feature vector. An ensemble of 2-layer autoencoders is used to cluster the feature vectors of normal packets in the grace period, and each autoencoder describes a distribution of packets by generating a low reconstruction error for the packets in the corresponding cluster. During the detection time, the lowest reconstruction error generated by the autoencoders will be used as the anomaly score of a packet. The proposed method achieves a processing speed of ~2000 packets per second with approximately 30% of CPU resource (i7-1165G7 @ 2.8GHz 2.8GHz) and catches the anomaly in the test datasets and real cloud environment.

REFERENCES

Yisroel Mirsky, Tomer Doitshman, Yuval Elovici, & Asaf Shabtai (2018). Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. CoRR, abs/1802.09089.

SOTI

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing