duAE-IDS: A dual Autoencoder-based Intrusion Detection System

Student: Hsiao-Te Hu
Advisor: Yu-Lun Huang
Institute: Graduate Degree Program of Cyber Security,
National Yang Ming Chiao Tung University

Abstract

Recently, the deep learning (DL) technique is being used as a potential solution to build an anomaly-based IDS. Despite the success with an anomaly-based IDS built with DL technology, the high false positive rate made by the anomaly-based IDS causes poor adoption in the industry. In this thesis, we propose a dual Autoencoder-based Intrusion Detection System, named duAE-IDS, as a way to cope with the ever-changing network attacks. duAE-IDS is a protocol-based IDS, which divides network traffic by the application layer protocols. Then, duAE-IDS determines the network traffic's abnormality by considering both the criteria and the application layer protocol. The criteria are obtained by training a neural network model, named duAE model, with traffic containing only one type of application layer protocol. Each network traffic is represented using 67 features extracted by the Feature Extractor of duAE-IDS named TSFlowMeter. Among 67 features, 8 TCP analysis features are adopted to improve the accuracy of the intrusion detection. The duAE model trained using benchmark datasets can reach an 87% f1-score for network traffic collected under different network environments. The results show that our duAE-IDS can be used in any network without pre-collecting the traffic of the target network.

duAE-IDS

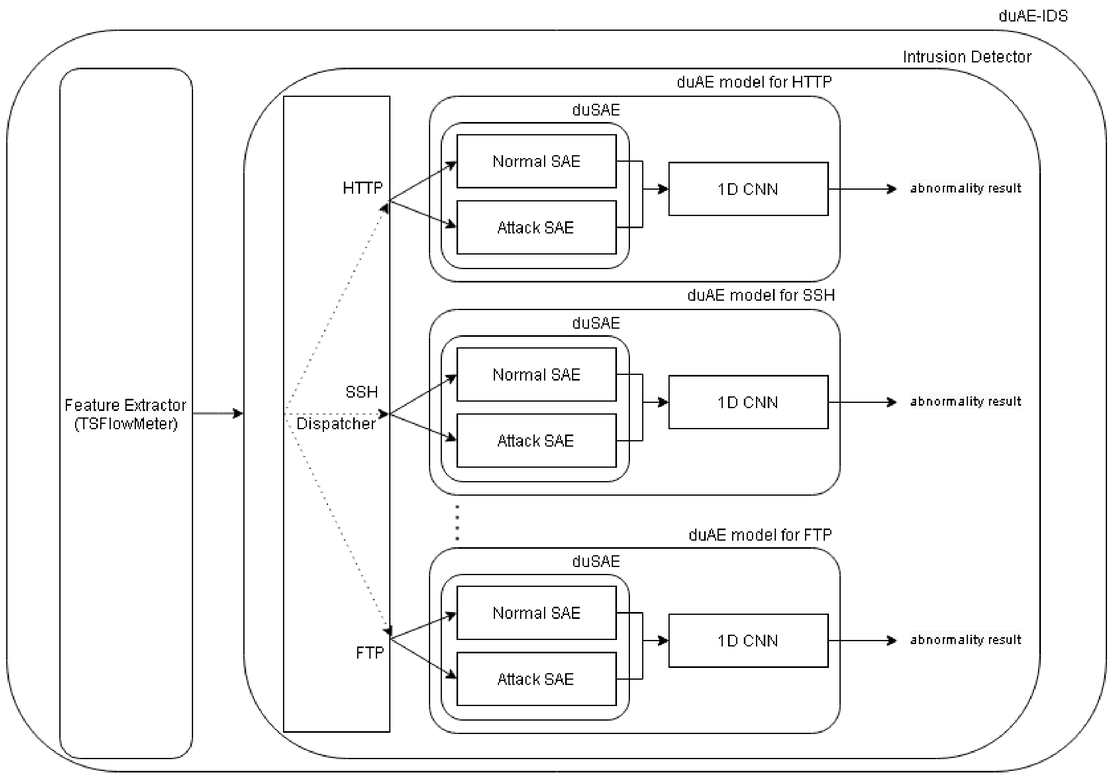
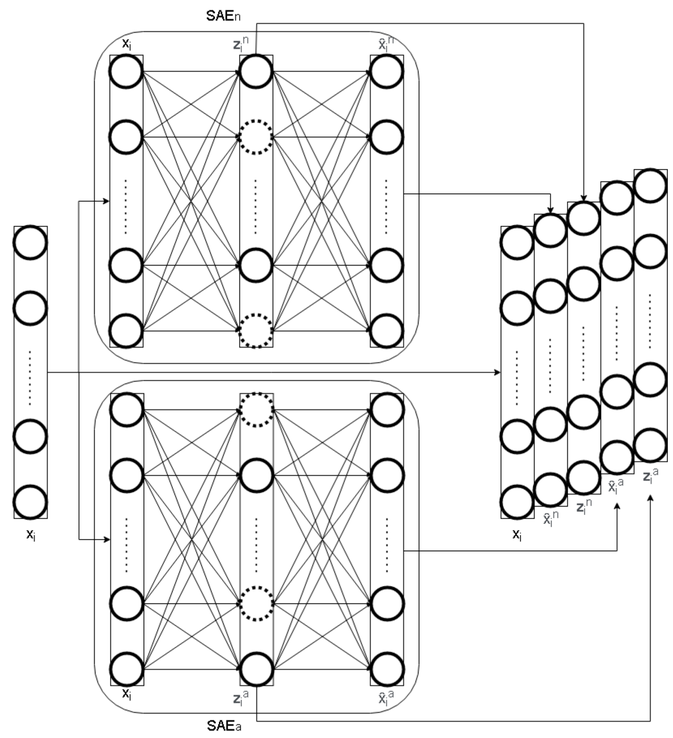


Figure shows the structure of duAE-IDS. The duAE-IDS consists of two components, a Feature Extractor named TSFlowMeter and an Intrusion Detector. The Feature Extractor is responsible for grouping raw packets into network flows and extracting features from each of the network flow. The feature vector for each network flow then passes to the Intrusion Detector for anomaly detection. The Intrusion Detector contains a Dispatcher plus several duAE models for different application layer protocols to be monitored. The Dispatcher is responsible for sending the feature vectors to the corresponding duAE model by the application layer protocol of the network flow. Inside the duAE model, we maintain a dual Sparse Autoencoder (duSAE) and an 1D CNN to extract more features from the feature vector and perform classification on those features.

TSFlowMeter contains three components, General Feature Extractor, Application Layer Protocol Decoder, and TCP dissector, to generate features for the network flow. General Feature Extractor is used to extract features that can be calculated directly from the flow packet headers. The features are the same extracted by CICFlowMeter [5], then remove the features that are duplicate or with value zero across all the network flows. Application Layer Protocol Decoder decode the application layer protocol used by the network flow. The application layer protocol can reduce the false positive significantly. In specific, some attacks can only happen in certain application layer protocols. The TCP dissector tracks the state of each TCP network flow and provides eight additional TCP analysis features. The following is the list of TCP analysis features:

- **Initial Round Trip Time (iRTT)**: iRTT tracks the completeness of TCP 3-way handshake to detect scanning behaviors such as TCP half-open scan.
- **Port Number Reused**: Multiple port number reused suggest the host may be compromised to perform attacks such as port scanning or DoS attack.
- **TCP Duplicate ACK**: Total duplicate ACK packets suggest the host is under a large amount of requests, useful for detection of brute force attacks.
- **TCP Retransmission**: Multiple times of TCP retransmission suggest the host is getting too many requests to handle, indicating an DoS attack in action.
- **TCP Fast Retransmission**: A type of TCP retransmission happens in certain malicious attack.
- **TCP Spurious Retransmission**: A type of TCP retransmission happens in certain malicious attack.
- **TCP Zero Window**: TCP zero window means the TCP buffer on the receiver side is full. A large number of TCP zero window packets may indicate malicious behavior in action.
- **TCP Window Update**: TCP window update means the TCP buffer on the receiver side is ready to receive new data. TCP window update is used with TCP zero window to determine the abnormality of the traffic.

In duAE model, duSAE is used to perform feature learning. One AE for normal traffic analysis and the other for malicious traffic analysis. After training the duSAE, we reconstitute each of the network flow by concatenating the original input vector x_i, the vector $\hat{x}_i^n$ produced by feeding x_i into the SAE_n, the vector z_i^n produced by feeding x_i into the encoder part of SAE_n, the vector $\hat{x}_i^a$ produced by feeding x_i into the SAE_a, the vector z_i^a produced by feeding x_i into the encoder part of the SAE_a. In summary, each network flow can be represented in a 5-tuple format (x_i, $\hat{x}_i^n$, z_i^n, $\hat{x}_i^a$, z_i^a). We expect to use the newly discovered representations to achieve better detection rate than using only the original features.



In duAE model, 1D CNN is used to discover the relationship between the vectors in a 5-tuple network flow from the duSAE. If the network flow is normal, then vector x_i is closer to vector $\hat{x}_i^n$ than vector $\hat{x}_i^a$, and the number of active neurons in vector z_i^a is larger than the number of active neurons in z_i^n. The architecture design for 1D CNN is based on the classic network architecture VGG 16. We apply three building blocks on the 5-tuple format network flow. We intend to build a deep yet less compute-intensive 1D CNN model for real-time unknown traffic intrusion detection.

Reference

[1] Y. Mirsky and T. Doitshman and Y. Elovici and A. Shabtai, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection", arXiv:1802.09089 [cs.CR], 2018
[2] G. Andresini, A. Appice, N. Di Mauro, C. Loglisci and D. Malerba, "Exploiting the Auto-Encoder Residual Error for Intrusion Detection," 2019 IEEE European Symposium on Security and Privacy Workshops, Stockholm, Sweden, 2019, pp. 281-290, doi: 10.1109/EuroSPW.2019.00038.
[3] G. Andresini, A. Appice, N. D. Mauro, C. Loglisci and D. Malerba, "Multi-Channel Deep Feature Learning for Intrusion Detection," in IEEE Access, vol. 8, pp. 53346-53359, 2020, doi: 10.1109/ACCESS.2020.2980937.
[4] D. Gümüşbaş, T. Yıldırım, A. Genovese and F. Scotti, "A Comprehensive Survey of Databases and Deep Learning Methods for Cybersecurity and Intrusion Detection Systems," in IEEE Systems Journal, doi: 10.1109/JSYST.2020.2992966.
[5] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018
[6] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), 2015, pp. 1-6, doi: 10.1109/MilCIS.2015.7348942.