Contribution ID: **32** Type: **Oral Presentation**

A weakly-supervised method for encrypted malicious traffic detection

Monday, March 21, 2022 4:00 PM (30 minutes)

In order to defend against complex threats and attacks in cyberspace, IHEPSOC has been developed and deployed at the Institute of High Energy Physics Chinese Academy of Sciences (IHEP), which is considered as an integrated security operation platform to ensure a secure state of the network and scientific researches in IHEP. Nowadays the integration of the state-of-the-art cyber-attack detection algorithms for IHEPSOC has become an important task for enhancing the attack discovery capability of IHEPSOC. Meanwhile, malicious traffic detection comes to be increasingly challenging recently. With the extensive application of data encryption techniques to protect communication security and privacy, malicious software could also hide their attack information, making most of malicious traffic identification methods such as port-based methods and DPI-based methods ineffective.

Machine learning based detection methods have been proposed to address the encrypted malicious traffic detection problem, which usually construct statistical features of internet traffic flow and train classification models for detecting encrypted malicious traffic. There have been some drawbacks with above mentioned detection methods. On the one hand, feature selection is a time-consuming procedure that depends on expert experience. On the other hand, most traffic classification schemes employ supervised learning methods, while the acquisition of the large fine-grained labeled datasets is a tedious task. In this paper, we propose a weakly-supervised method for encrypted malicious traffic detection, which combines the generative adversarial network (GAN) and the multiple instance learning detector to achieve the fine-grained classification of encrypted traffic with a small number of coarse-grained labeled samples and a large number of unlabeled samples. Thus, we could focus on the accuracy of the malware detector instead of spending efforts on dataset annotation with the weakly-supervised learning approach. First of all, we convert the traffic data into single-channel grayscale images in the proposed method, and then input them into the GAN, so that the original traffic features can be learned without manual effort. Secondly, the improved semi-supervised learning generative adversarial network, which based upon convolution neural network (CNN) architecture, generates more synthetic samples for data augmentation, and addresses the insufficiency of labeled samples. In addition, a multi-instance detector with an attention mechanism is used to identify encrypted malicious traffic from coarse-grained labeled data. We validate the proposed approach on two datasets, a real-world dataset and a public dataset. Compared with other malicious traffic detection methods, the experimental results show that our proposed framework can effectively perform fine-grained detection of encrypted malicious traffic.

Primary authors: LIU, Junyi; LI, Zhenyu; WANG, Jiarong (Institute of High Energy Physics); YAN, Tian (IHEP); QI, Fazhi (Institute of High Energy Physics,CAS); CHEN, Gang (Institute Of High Energy Physics)

Presenter: LIU, Junyi

Session Classification: Artificial Intelligence

Track Classification: Track 10: Artificial Intelligence (AI)