Contribution ID: **33** Type: **Oral Presentation**

Imbalanced Malicious Traffic Detection Based on Coarse-grained Data Labels

Wednesday, March 23, 2022 2:00 PM (30 minutes)

In order to resist complex cyber-attacks, IHEPSOC is developed and deployed in the Institute of High Energy Physics of the Chinese Academy of Sciences (IHEP), which provides a reliable network and scientific research environment for IHEP. It has become a major task to integrate cutting-edge cyber-attacks detection methods for IHEPSOC to improve the ability of threat detection. Malicious traffic detection based on machine learning is an emerging security paradigm, which can effectively detect known and unknown cyber-attacks. However, the existing studies usually adopt traditional supervised learning, which often encounters some problems contrary to the implicit hypothesis in the real-world service. For example, most studies are often based on data sets that already have accurate data labels, but these labeled data sets take a lot of manual effort to carry out such accurate data labels according to the requirements. In addition, in the real-world service, the benign traffic data is much more than the malicious traffic data, and the imbalance between benign and malicious categories also makes many machine learning detection models difficult to apply in the production environment. Based on these problems, we propose an imbalanced malicious traffic detection method based on coarse-grained data labels. First of all, malicious traffic detection is modeled as a weakly supervised learning problem of multi-instance and multi-classification learning, which only needs to use coarse-grained data labels for traffic detection. Specifically, experts only need to confirm whether there is malicious traffic in the original data stream for a period of time when labeling data manually. There is no need to find out the accurate data of malicious traffic, which greatly reduces the difficulty of data labeling. In addition, in weakly supervised learning with only coarse-grained labels, the above class imbalance of benign and malicious is more serious, and the solutions to class imbalance in traditional machine learning, such as sampling, cost-sensitive functions, etc., will destroy the premise of coarse-grained labeling in weakly supervised learning. In view of this, we design a corresponding scheme to deal with the imbalance under weak supervision. The possibility of malicious traffic is pre-estimated by integrating the results of multiple clustering models and updated during the training process to shield the negative impact of the majority benign traffic in a group of coarse-grained labeled data. We change the fine-grained data labels in Android Malware dataset to coarse-grained data labels, and show that the proposed method with coarse-grained data labels outperforms traditional supervised learning method. In addition, we carried out an ablation study to verify the effectiveness of each module in our method.

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

Presenter: LI, Zhenyu

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations