2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

Internet traffic anomalies and their detection techniques

Kensuke Fukuda

NII / Sokendai
2-1-2 Hitotsubashi, Chiyoda, 101-8430, Japan
Email: kensuke@nii.ac.jp

The Internet has been a crucial infrastructure in our daily life. Most of traffic (packets) are legitimate one, however, we frequently observe some anomalous traffic hidden in such majority of legitimate traffic. These traffic are due to malicious activities such as virus and DoS (Denial of Service) or misuse of devices (e.g., router configuration).

It is not easy task to find such anomalous behavior in Internet backbone traffic. There are two approach to find anomalies in Internet traffic; signature-based and statistics-based one.

The former assumes on an exact matching of packet field to pre-defined anomalous signature similar to recent virus detection techniques. The accuracy of the detection is high, however the drawback of this approach is a difficulty to scale up to current high speed networks (e.g., 10-100Gbps networks). Furthermore, it naturally requires known packet signatures, thus it is impossible to find recent so-called zero-day attack that use unknown vulnerabilities.

The latter approach relies on statistical methods to differentiate normal and anomalous behaviors. Thus, deviation from the normal behavior corresponds to anomalous one. The accuracy of this approach can be lower than the signature-based approach, however, the main advantage of this approach is a robustness against new types of anomalies. Also, this approach is suitable for high-speed networks because it is robust against packet sampling.

In this talk, we first review past literature in anomaly detection in Internet as described above. Then, we introduce our past and recent activities to tackle this problem. As a basic principle, we do not believe that there is one perfect anomaly detection algorithm for Internet backbone traffic. Therefore, we have been developing several anomaly detectors based on different theoretical backgrounds. We will briefly explain these algorithms (e.g., Multi-scale histogram, PCA, and image processing approach).

However, it is a natural idea to combine several anomaly detection algorithms to increase the accuracy of the anomaly detection, assuming that there is no one perfect algorithm. Based on this principle, we developed a new anomaly detection framework for Internet backbone traffic. The key idea of this is to construct a graph representation of anomalous events; the node is an anomalous event detected by each algorithm, and the link weight between them is the number of packets belonging to them. We apply a community mining algorithm to extract dense parts in the graph, meaning that anomalous behaviors are highlighted as well-connected sub graphs.

We demonstrate the performance improvement of this approach by using publicly available MAWI traffic traces. This longitudinal dataset since 2000 contains a wide variety of network events, so they are suitable to show the effectives of our approach. Indeed, we have been updating our anomaly database that is an annotation to the WIDE traffic traces by our detection algorithms for other researchers (http://www.fukuda-lab.org/mawilab). Thus, if other researchers develop a new anomaly detection algorithm, they can easily compare the performance of their algorithm to our benchmark results.

Finally, we introduce our recent activity on a new type of Internet sensor, DNS backscatter, for anomaly detection. One problem of Internet traffic anomaly detection is a locality of measurement point. Network operators know their network status by observing their networks. However, it is not easy to understand network-wide events (e.g., scans and spams). In this end, we have been developing a new Internet sensor based on DNS. The basic idea of this sensor relies on the fact that a source of a network-wide event triggers a large number of reverse DNS queries at target, though each queries have a small amount of information. Thus, we could detect anomalous events from the DNS queries with the help of collective knowledge. We show the effectiveness of DNS backscatter with DNS logs measured at Root and JP DNS servers.