Contribution ID: **21** Type: **Oral Presentation**

Cyber security monitoring and data analysis at IHEP

Wednesday, 3 April 2019 15:00 (20 minutes)

In recent years, along with the rapid development of large scientific facilities and e-science worldwide, various cyber security threats has becoming a noticeable challenge in many data centers for scientific research, such as DDoS attack, ransomware, crypto currency mining, data leak, etc.

Intrusion and abnormality detection by collecting and analyzing security data is an important measure for enhancing the sensitivity of security status perception, level of security protection, and agility of security incident response. However, as the scale of data center growing, it's difficult to use a single security box to process the large volume of various data generated by network traffic, device and host logs, threat intelligence, and so on.

In high energy physics (HEP) community, people are trying to establish a security operation center (SOC) for handle this problem. There is a SOC working group for the Worldwide LHC Computing Grid (WLCG). With help of this working group, we are building a cyber security monitoring and analysis framework at Institute of High Energy Physics (IHEP), Chinese Academy of Sciences. At IHEP, we have 4x10Gbps IPv4 and IPv6 dual-stacked internet connection, and 4x40Gbps inner data center network. There are also hundreds of web information servers, thousands of PC clients and thousands of computing nodes. It's really a challenge for us to handle the security related data generated by such a set of information assets.

In this framework, Malware Information Sharing Platform (MISP) is deployed for threat intelligence exchanging with collaborated HEP institutes and universities. Network traffic is collected from switches and firewalls by a 10Gbps network shunt, and then flows to a Bro instance for traffic analysis. Bro logs and hosts/web logs, security device logs, along with vulnerability scanning results and assets detection results, etc., are defined as cyber security data. All of these data are collected by Flume/Logstash/Syslog to a data pipeline named Kafka cluster. In this cluster, there are some Spark jobs running for stream processing, which are aimed at rapid intrusion and abnormality detection as well as data correlation and enrichment. Then all the processed data are written to Elasticsearch, MySQL and InfluxDB, and then visualized by Kibana and Grafana. At the same time, the processed data can be written to local storage, HDFS, or tap storage for backup and long-term analysis.

With help of this security data collection an analysis framework, it is possible for us to handle the large amount of security data generated at IHEP, and it's also very flexible and scalable for even larger amounts of and different kinds of data in future.

Primary author: Dr YAN, Tian (IHEP)

Co-authors: Mr DEHAI, An (IHEP); Mr QI, Fazhi (Institute of High Energy Physics,CAS); Ms HU, Hao (Institute of High Energy Physics); ZENG, SHAN (IHEP)

Presenter: Dr YAN, Tian (IHEP)

Session Classification: Networking, Security, Infrastructure & Operations

Track Classification: Network, Security, Infrastructure & Operations