Position Paper
A Cloud Computing Based Architecture for Cyber Security Situation Awareness

Wei Yu* and Chao Lu
Cyber-Physical Networking System and Security (CPNSS) Research Laboratory
Department of Computer and Information Sciences
Towson University, Towson, MD 21252
http://www.towson.edu/~wyu, http://www.towson.edu/~lu/
http://pages.towson.edu/wyu/CPNSSwebsite/CPNSS.html
*Corresponding Author

Motivation: The computer and communication techniques and the innovations in social web and mobile technologies have led to an incredible growth of the cyber space, which has been reshaping the way people communicate, learn, and work. Nonetheless, the fast growth of the cyber space leads to opportunities for the proliferation of cyber-threats. In the past, cyber-threats have increasingly evolved from simple and benign types to sophisticated types, featured by multi-stage phased attacks targeting business operations, critical infrastructures, etc. Hence, there is an urgent need for effective threat monitoring and detection systems, which can prevent and combat these threats quickly and accurately.

Cyber security situation awareness is recognized as a promising technology, which can enhance the security of cyber space in characterizing, monitoring, detecting, and mitigating cyber threats in a timely and accurate manner. Nonetheless, it has been an increasing number of applications and devices distributed in cyber space and a large amount of data stream collected from different applications, operating systems, and network devices for the cyber security analysis. Efficiently storing and processing such large scale stream data is challenging. Cyber security situation awareness (e.g., threat monitoring and analysis, network fraud and intrusion detection are characterized by very high volume data streams (big data) and real-time processing requirements. Hence, the increasing volume of data stored in central databases and the high computation power requirement can seriously hinder the effectiveness of cyber security situation awareness.

Our Approach: To address the aforementioned challenges, we propose a cloud computing based architecture to assist cyber security situation awareness, which offers vast storage, flexible deployment, more computation resources, less expensive infrastructure investment, and ubiquitous sharing of information across all members of the cloud. It is worth noting that our proposed architecture can be used to combat threats in large enterprise networks and cyber-physical systems (CPS) such as energy CPS and transportation CPS). Our proposed cloud computing based architecture, which consists of components: data sources, cloud infrastructure, and operation center. Data sources consist of various end user devices (e.g., mobile devices, computers, network devices, etc.) connected to the cloud computing infrastructure. Due to the limited resources, end-user devices can use services such as Infrastructure as a service (IaaS), Platform as a service (PaaS), and Software as a service (SaaS). To conduct the cyber security situation awareness, the threat information needs to be effectively collected and transmitted to the cloud for further security analysis. Here, a cloud infrastructure is a distributed system deployed with a number of servers in the data center, which provides the storage and computation for the cloud. In our proposed system, there are two types of servers inside the cloud: storage servers and application servers. The monitored stream data will be pushed and stored in storage servers in real-time while application servers will provide threat analysis and detection capabilities. The operation center is one special user of cloud computing infrastructure, which monitors the security status of end users' devices. Additionally, the operation center can dynamically push the cyber operation policies and configuration to the monitored end-users' devices and cloud servers. The operation center plays the intelligence role in the cloud computing infrastructure.

The operation center remotely sends queries to cloud storage servers and retrieves useful information for threat analysis. In addition, the visualization schemes will be developed in the operation center to assist the security administrator in monitoring and detecting threats effectively.

Based on the proposed architecture, we can collect the data streams from end-user devices for threat monitoring and analysis. Nonetheless, to meet real-time requirements posed by cyber security situation awareness, there are several issues. On one hand, a large number of threat monitoring data streams, including both host-based and network-based detection data, are collected from end user devices distributed over the network and streamed to the central database for detection and analysis purposes. The mounting volume of data stored in the central database and the continuously increasing storage capacity incur a high cost and can significantly slow data extraction and the overall performance of the system. On the other hand, the real-time data processing to generate security alerts is time critical and the latency should be kept at a minimum in order to assure the effectiveness of cyber security detection and analysis.

To overcome these issues, we will leverage the large storage and computational resources in the cloud to conduct threat detection and analysis. To do so, a streaming-based storage model should be developed in order to reduce the storage processing time. To efficiently collect and process the large data stream from end-user devices, we will leverage the stream processing algorithm such as MapReduce-based data processing. To ensure that the threat detection methods are efficient, techniques such as MapReduce based machine learning (MML) schemes can efficiently deal with threat monitoring over big data. The core idea of the MML system is to speed up the machine learning (ML) process using cloud computing. The first step is to collect the characteristics of traffic flows (e.g., flow duration and average bytes per packet of the flow, average bytes per seconds of the flow, and etc.). To accurately and rapidly detect traffic anomalies, MapRduce based ML schemes should be developed to profile the dynamic characteristics of traffic flows and then to detect anomalies. In the proposed MML schemes, the computational burden of the learning process is spread across multiple machines. The learned computational results from multiple machines are then integrated into one single learned classifier. Lastly, the learned classifier will then be used to recognize whether a new traffic flow is normal or abnormal (benign or malicious). Besides the MML schemes, techniques such as digital signal processing and image processing based approaches should be integrated to the cloud environment to efficiently filter unrelated information and conduct the attack scene investigation. For example, given a large amount of data, deriving the spatial and temporal graph requires intensive computation time. By using MapReduced-based data processing, the time taken for obtaining the spatial and temporal graph related to attacks can be significantly reduced.

As a critical computing infrastructure, the cloud computing system with a large number of computing resources can also be a target for a cyber adversary. Because cloud computing systems have been connected through networks for easy use, they are exposed to cyber-threats. After cloud system components are compromised, the adversary can manipulate computing resources, resulting in disruptions to computing operations. Hence, addressing the security of cloud computing is important. To provide secured cloud computing service, we will develop an effective cyber security monitoring system to deal with cyber-threats on cloud computing systems. To accomplish this, we will develop a framework to systematically explore the space of threats, model and analyze the impact of these threats. Based on the understanding of threats, we will develop an integrated defense system against diversified cyber-threats on cloud computing systems, which span techniques for attack protection, detection, and attribution. The challenge is that the developed defense techniques need to balance the detection performance (accuracy, efficiency) and overhead to the cloud systems (e.g., time, code, memory, compute, I/O, storage, architecture heterogeneity, and etc.).