Data Collection Framework for real-time streaming of cybersecurity events

Herbert Maosa

Abstract:

Enterprises have threat response systems that are based on analysing data from various internal sources such as intrusion detection systems, firewalls and event logs. The analytics could be signature based or behavioural. At the same time, some enterprises also ingest data from various threat intelligence sources using various cyber threat intelligence platforms. However, the threat intelligence data is merely used as data repository with no integration with the threat response systems.

Alert correlation is common step in cyber analytics, aiming to cluster and aggregate alerts for ease of analysis and to improve efficiency by reducing false positive rates. It is evidenced that the more sources are used for correlation, the better the efficiency. However, most systems ingest from internal sources only, thereby limiting the cyber situational awareness of enterprise systems.

In this research, we propose a model for an intelligence driven threat response system that takes advantage of the plethora of threat intelligence data to model a systems threat state and inform real time response actions. We implement correlation of internal sources with external threat intelligence sources in order to improve the detection performance.

Data Collection is an important first step in the analytics process. In the pre-liminary, we have implemented a framework for data collection, pre-processing and real-time streaming of network events for cyber security analytics. Real malware executables

are introduced into Windows Host machines in a sandbox environment based on virtualisation using VMWare Workstation Pro Hypervisor. The malware network activities are captured by custom built sensors into pcap files. A collector client application is developed and sends the packet captures to central collector. After pre-processing, the malware packets are streamed as JSON records into separate Kafka and MQTT message streams to online servers and brokers in real-time.

The overarching aim of this research is to develop an intelligence-driven cyber analytics model for real time threat detection and response.

We have used several research approaches, being analytical, theory building, systems building and observations.

We find that the proposed data collection framework is successful in streaming network events in real time with reduced network and storage capacity demands due to packet feature selection