A Trusted Knowledge Management System for Multi-layer Threat Analysis

Anonymous Submission

No Institute Given

Abstract. Large scale attacks and the increasing malware sophistication call for a better understanding of the involved threats and resilient-by-design information systems. In this paper, we present *tKMS*, a trusted knowledge management system for multi-layer threat analysis, designed based on the following requirements: trustworthiness, confidentiality, scalability and uniform programmability.

1 Motivation

In recent years, we have seen a surge of cybersecurity incidents ranging from widespread attacks (e.g., large scale attacks against infrastructures or endpoints [1]) to new technological advances (i.e., new generations of malicious code are increasingly stealthy, powerful and pervasive [2]). Facing these incidents, the European Union, Japan, the United States or China have developed national cybersecurity programs, including training of professionals, development of roadmaps for new tools and services and organization of national interest groups on the topic. *There is thus a shared need for better understanding of this kind of large scale threats.* Some of the basic requirements to better understand these large-scale incidents include handling large volumes of data collected from distributed probes and performing efficient cross-layer analysis.

2 System Description

In this paper, we introduce a trusted knowledge management system for multi-layer threat analysis (tKMS). tKMS is capable of supporting a great variety of sensors ranging from honeypots and spam detection systems to real-time intrusion detection systems and online web sources. Moreover, it provides actionable information for cyberdefense systems. Support for a wide array of sources is feasible thanks to the modular architecture and a common lightweight data sharing format – the *n6* API. tKMS is comprised of two basic components: the *Threat Information Sharing component* and the *Cross-layer Analysis module*, as shown in Figure 1. We have designed our system to meet the following requirements: (i) provision of trusted access to multiple sources of data, (ii) confidentiality of the security networks that provide the data, (iii) scalability, (iv) real-time analysis and (v) uniform programmability through support of multiple data types. What follows is a description of tKMS basic components.

Infrastructure and End Point Data. Infrastructure and End Point data are collected from infrastructure (routers, switches, middleboxes, etc.) and endpoint (PC, smartphones, etc.) devices by means of probes and consists of several datasets. Each dataset provides an API which enables other modules and components to access its data.

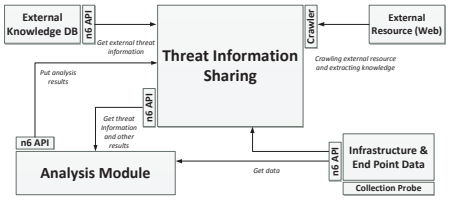

Fig. 1: Architecture of tKMS.

Threat Information Sharing (TIS). This component manages threat information from *Analysis Modules*, *External Knowledge Systems* and *External Resources*. Originally, data posted by these components are unrelated and the system conjectures a relationship among them turning, them into knowledge.

Cross-layer Analysis Module (CAM). The CAM aims at detecting cyber-threats based on the analysis of data coming from the infrastructure and end point layers. The CAM consists of several components. Each component serves the purpose of detecting a certain threat or a number of threats that are somehow related. The analysis results are pushed to the TIS component via the n6 API.

External Knowledge System (EKS). This component designates external sources of cyber threat related information such as software vulnerabilities databases. The information is provided through an API using the common exchange format.

External Resources (ER). This component collects cyber threat information or related information not formatted in any standard scheme. The main difference with the EKS is that the resources do not provide any data sharing interfaces. That kind of information is gathered mainly by web crawlers and other automated data gathering mechanisms which are able to extract knowledge from external sources. Acquired knowledge is managed under the TIS.

The n6 API. n6 is a platform for processing security-related information and its API provides a common and unified way of representing data across the different sources that participate in our knowledge management system. n6 exposes a REST-ful API over HTTPS with mandatory authentication via TLS client certificates, to ensure confidential and trustworthy communications. Moreover, it uses an event-based data model for representation of all types of security information. Each event is represented as a JSON object with a set of mandatory and optional attributes.

Access control and encryption mechanisms are used in order to preserve confidentiality of data across the different components. Moreover authentication mechanisms are used in the components' communication to make our system resilient to hijacking.

References

1. M. Brown. Pakistan hijacks youtube. `http://www.renesys.com/2008/02/pakistan-hijacks-youtube-1/`.
2. D. Kushner. The real story of stuxnet. `http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet`.