CSRL See5 Transformation: Using PHP for Data Transformation Between SNORT® IDS Databases and See5 Flat Files

1. Abstract

When incorporating multiple off the shelf systems into a new system, a method must be developed to transform the output of one system into a format acceptable as input for another. This work centers around the design and development of a PHP based tool for transforming data from SNORT®, an intrusion detection system, and See5, a data-mining application. The Computer Security Research Lab at California State University Dominguez Hills is using this tool to further its research in developing a software system which will be able to automatically create firewall rules based on the classification of malicious network traffic.

2. Introduction

In the Computer Security Research Lab (CSRL) at California State University, Dominguez Hills, student researchers are working to create a system able to distinguish malicious traffic from regular network traffic in a sensor network and be able to automatically update the network’s firewall rules to prevent attacks. To accomplish this, the system, called JACED, incorporates two existing software systems, SNORT® and See5. SNORT® is used as an intrusion detection system [1], and See5 is being used as a data mining tool [2]. SNORT® stores output data in a MySQL database, and See5 requires input in the form of a flat file with a data extension [3]. In order to use the two components in conjunction with each other, there must be a process of data transformation on the alerts in the SNORT® database to allow the alerts to be processed by See5. As there are no currently existing tools to perform this transformation, the CSRL See5 TransformationGUI (CSRL-See5-GUI) provides researchers with a tool to facilitate the transformation of the SNORT® sensor data for research purposes.

4. CSRL See5 Transformation GUI

Given that the size of a See5® identification results database can be thousands of records long and that there are multiple sensors each recording information in different databases, manual transformation of the data from MySQL to a data flat file for processing by See5 is both time consuming and prone to human error. The CSRL See5 Transformation GUI allows researchers to effortlessly aggregate data from various tables in the See5® database and automatically create a data file ready for processing by See5.

4.2 See5 GUI Tool

The CSRL See5 Transformation GUI -
• Written in PHP, and interfaces directly with a PHP script which does the data transformation on the See5® database;
• Is accessible as a web page hosted by the same server which contains the See5® databases.

Provides an easy, graphical method for the user to:
• See which databases are available;
• Perform various functions on the data contained within.

Provides the following functionality:
• Dynamically creates a table listing the databases which can be processed for use with See5;
• Determines if the databases meet CSRL classification standards and provides access to a PHP function to classify them if not;
• Determines whether the databases contain the required tables for data transformation and provides access to a PHP function to create them if not;
• Provides access to a PHP function which performs the data transformation on a selected database;
• Provides access to a PHP function which creates the data flat file from a selected database.

3. Existing Systems

3.1 SNORT®

SNORT®, developed by SourceFire, is an open-source software package designed for packet capturing and real time network traffic analysis [4]. In JACED, SNORT® is being used to identify traffic across a honeypot sensor network as being either malicious or benign, and, if malicious, to provide a signature of the traffic. The results of SNORT®’s identification are then stored in a MySQL database for later analysis. The information stored in the SNORT® database includes the malicious identification signature provided by SNORT®, as well as the IP, and TCP, UDP, or ICMP header information of the sensor network traffic [5].

3.2 See5

See5, developed by Ross Quinnan as the successor to the IDS and C45 systems, is a data mining tool which provides data classification capabilities [6]. In JACED, we are using See5 in order to discover patterns and regularities in the SNORT® database, present them in an intelligible form, and use these patterns and regularities to make predictions about other malicious traffic. See5 does not directly interface with MySQL, but instead accepts data for classification in the form of a flat file with a data extension [7].

5. Results

Student researchers performed black box tests on the See5 GUI, selecting various portions of a number of databases and attempting to create data flat files from the databases. They found that there was no difference in output generated by the See5 GUI and output created by manually entering commands into MySQL.

6. Ongoing Work

As the JACED system progresses, the CSRL See5 Transformation GUI will also grow and expand to meet the additional needs of researchers. Plans for the GUI include automation of the fusion of database information from multiple sensors, automation of the creation of data flat files for all databases, and to ultimately allow for control of all aspects of the JACED project via a user interface.

7. References