Security Data Analysis using Web Server Log Files

Tanushri Bhaduri

September 2016

Dissertation submitted in partial fulfilment for the degree of
Master of Science in Big Data

Computing Science and Mathematics
University of Stirling

Abstract

Nowadays everyone is using different website for different purposes like banking, trading, shopping, playing game, reading books etc. But websites are never safe. Hence companies use different security prevention tools like firewalls, secure socket layer, data encryption. But these applications also cannot protect the website completely. Hence, security detection is required to increase the level of security. Security detection is a process of monitoring a website to find out if someone tried to hack or already hacked the website. There are various security detection tools available like scan my server, web inspector. In this project, a security detecting tool has been created by performing analysis of web server log files. Server log files capture the requests of the user, which is checked and may or may not be fulfilled by the web server in the form of web page. Log files are text files generated by the server. This file is in a human and machine readable format. So this model reads the log file, once the path of the log file will be provided. It performs outlier detection, time series analysis and rule-based classification to identify if any kind of suspicious activity has happened in the website. To perform this analysis, first data has been generated by building a web server and creating a website. In order to test this model, simple attacks like brute force attack, cross-site scripting, SQL injection and cross-site request forgery are implanted into the website. This analysis has been performed using the HTTP status code. The status code is a three-digit code send to the user in the response of the request he made. The output of this analysis has been presented in a web page as a real-time report. This result can be generated for a day, week or month as per requirement.