ABSTRACT

The back-end database is pivotal to the storage of the massive size of big data Internet exchanges stemming from cloud-hosted web applications to Internet of Things (IoT) smart devices. Structured Query Language Injection Attack remains an intruder's exploit of choice on vulnerable web applications to pilfer confidential data from the database with potentially damaging consequences. The existing solutions of mostly signature approaches were all before the recent challenges of big data mining and at such lacks the functionality and ability to cope with new signatures concealed in web requests. An alternative Machine Learning predictive analytics provides a functional and scalable mining to big data in detection and prevention of SQLIA. Unfortunately, lack of availability of readymade robust corpus or data set with patterns and historical data items to train a classifier are issues well known in SQLIA research. Here we explore the generation of data set containing extraction from known attack patterns including SQL tokens and symbols present at injection points. The trained classifier to be deployed as a web service that is consumed in a application implementing a web proxy Application Programming Interface (API) to intercept and accurately predict SQLIA in web requests thereby preventing malicious web requests from reaching the protected back-end database. This will demonstrate a full proof of concept implementation of an ML predictive analytics and deployment of resultant webservice that accurately predicts and prevents SQLIA.