Implementation of Prevention of XPath Injection Attack using PyBRAIN Machine Learning Library - Gajendra Deshpande

About the speaker

Mr. Gajendra Deshpande holds a masters degree in Computer Science and Engineering and working as Assistant Professor at the Department of Computer Science and Engineering, KLS Gogte Institute of Technology, Belgaum. He has a teaching experience of 9.4 years and Linux and Network Administration experience of one year. He is listed as a speaker on Microsoft Technical Community in the subject area “Programming Languages and Visual Studio”. He is independent consultant to many firms and educational institutions for technology implementation. He has worked as system analyst for inventory management software for KLS Gogte Institute of Technology, Belgaum. He has developed a many websites for various organizations, i.e., Shantiniketan Public School, Belgaum Education Society to name a few. He has worked as coordinator for the website of KLS and its institutions and developing software to manage various processes in educational institutions for Belgaum Education Society and its institutions. He is Technical Director for Sestoauto Networks Pvt. Ltd. and Founder of Desh Labs Pvt. Ltd.

Abstract

Injection attacks are considered to impact the most wide spread vulnerabilities in web applications by Open Web Application Security Project (OWASP). XML is used as an alternative technology to Database Systems to store data in XML format, which can be queried to produce the desired results. XPath a query language for XML has injection issues similar to SQL. XPath can be used by the attacker to exploit the vulnerabilities in web applications by injecting malicious XPath query. If the Web Service is injected with malicious XML code, then it affects all the applications which integrate the infected web service.

In this paper we propose a solution which uses count-based validation technique and long-short term memory (LSTM) modular neural networks to identify and classify atypical behavior in user input. Once the atypical user input is identified, the attacker is redirected to sham resources to protect the critical data. . For our simulation we considered the synthetic dataset where the input sample size is 100. We have performed two experiments to classify atypical behavior in user input: one with single neural network and second with modular neural network. The results are recorded for 10 iterations.

Our experiment results in over 90% accuracy in classification of input vectors. Our results also show that use of modular neural network results in improved response time of the web application compared to single neural network. For our experimentation we have used modular machine learning library PyBRAIN which is an acronym for Python-Based Reinforcement Learning, Artificial Intelligence and Neural Network Library on fedora Linux Operating System. PyBrain is a modular Machine Learning Library for Python. For web services we used BottlePy micro web framework for Python. BottlePy framework comes with its own web server known as WSGIRefServer. We used two web servers one for handling normal requests i.e., and the other for handling malicious requests i.e., Apache web server. Various graphs showing the comparison between True Positives, False Positives, False Negatives and True Negatives were drawn using Scipy, Numpy and Matplotlib libraries.