A COMPARISON AND EFFECTIVE PENETRATION TESTING APPROACHES WITH NMPREDICTOR BASED ON MACHINE LEARNING

By

MUHAMMAD NOMAN KHALID

A thesis
Presented to the Bahria University, Karachi Campus
In partial fulfillment of the requirement
For the degree of
MS (Computer Science)

SPRING, 2018

DEPARTMENT OF COMPUTER SCIENCES
BAHRIA UNIVERSITY

A COMPARISON AND EFFECTIVE PENETRATION TESTING APPROACHES WITH NMPREDICTOR BASED ON MACHINE LEARNING

By

Muhammad Noman Khalid

A thesis report submitted in partial fulfillment of the requirements for the degree of MS (Computer Science)

Supervisor: Dr. Humera Farooq

Nationality: Pakistani

Bahria University Karachi Campus

13, National Stadium Road

Pakistan

Bahria University
Discovering Knowledge

MS-13

Thesis Completion Certificate

Student's Name: M. NOMAN KHALID Registration No. 24120

Programme of Study: MS (CS)

Thesis Title: A COMPARISON ON PENETRATION TESTING APPROACHES AND A PROPOSED FRAMEWORK FOR WEB VULNERABILITIES BASED ON MACHINE LEARNING

It is to certify that the above student's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for Evaluation. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index at 9% that is within the permissible limit set by the HEC for the MS/MPhil degree thesis. I have also found the thesis in a format recognized by the BU for the MS/MPhil thesis.

Principal Supervisor's Signature: _____

Date: 19/3/18 **Name:** Dr. Humera Farooq

Acknowledgement

First of all I would like to thank Almighty Allah who give me strength to complete my thesis. Then this study could not have been carried out without the cooperation, constructive advice and inspiration of my supervisor **Dr Humera Farooq.** I am equally grateful to **Sir Iqbal** who not only supported my idea but also gave me wise advises regarding my research. I would also like to thanks my colleague and friends who helped a lot to achieve goal of this research. Last but not the least I am thankful to my parents because their prayers and support are always with me.

Table of Contents

ABSTRACT

Vulnerabilities are known to be difficult to detect and prevent, especially in the context of web application. Although a significant research on web application security has been ongoing for a while, these applications have been a major source of problems and their security continues to be challenged. An important part of the problem derives from vulnerable source code of web applications. In order to overcome web vulnerabilities, different penetration tester used variety of techniques such as secure programming, static analysis, dynamic analysis, hybrid analysis and machine learning. Machine learning is consider an approach to prevent web vulnerabilities with a wide range of web applications because it is more preferable and does not have problems of false positive rate.

There are numerous method proposed for detecting web vulnerabilities based on machine learning. It is very difficult to measure, which method is efficient to secure web application. Furthermore, there is lack of study found that targets the comparison of machine learning method to find out optimal method. However, comparative study is required to understand the path that could be followed by different penetration tester. In this thesis we use six different methods based on machine learning. In order to find optimal method for existing studies, decision were taken on Drupal metrics file with J48 and random forest. We have implemented NMPREDICTOR method with the feature extraction, performance parameters, classifiers with default parameters and 10k cross validation. Training data is passed through J48 and random forest to form a training model on which testing data is predicted and analyzed. Our results state that, to prevent web vulnerabilities VULPREDICTOR shows better results as compared to all others methods. We have found much higher accuracy of NMPREDICTOR method with respect to those reported by existing studies.