Email Classification Tool to Detect Phishing Using Hybrid Features

H.V. Mahesh

169321L

Faculty of Information Technology

University of Moratuwa

March 2019

Email Classification Tool to Detect Phishing Using Hybrid Features

H.V. Mahesh
169321L

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of Degree of Master of Science in Information Technology.

February 2019

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student (s)	Signature of Student (s)
H. V. Mahesh	……..
	Date ……..

Supervised by

Name of Supervisor	Signature of Supervisor
S.C. Premarathne	…….
	Date: ……...

Acknowledgements

First and foremost, I would like to express my sincere gratitude towards my supervisor, Mr. Saminda Premarathne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his guidance, supervision, advices and sparing valuable time thorough the research project.

A special thanks goes to Mr. Indika Gunawardhana, CIO, LAUGFS Holdings Ltd for the valuable support and guidance given for the research. I would also like to thank all batch mates of my MSc IT program who gave their valuable feedbacks to improve the results of the research. Further I express my warm thanks to my family & wife for the support and encouragement they have given during this time.

Abstract

Phishing is a fraudulent attempt of trying to gather personal sensitive information such as user ID and passwords, credit card and bank account details through network. Social messaging and websites are used as medium to trigger attacks in addition to the use of emails, which is the most common and leading method currently used to perform phishing attacks. In an attack, the attacker is sending an email with a URL of the phishing website camouflaged as a legitimate source.

Nowadays phishing has become more complicated and critical problem to many organizations. The phishers can bypass the filters and rules set by anti-phishing procedures and techniques. This research build a web based phishing email detection tool using data mining classification model.

To build an efficient classification model, varieties of extracted email features have been used. These selected features can be categorized according to email header, email body, URL and Web Page Content of URL. In this model, classification accuracy will be enhanced by using these hybrid features.

This model will be used to implement the web-based tool to detect phishing emails with more accuracy even without opening the emails. This can be used as preventive and proactive technique for phishing detection.

Contents