Department of Computer Science & Engineering

A Project Report on

Windows Based Malware Prediction Using Deep Learning Techniques

Submitted in partial fulfilment of the requirements for the award of the degree of

Bachelor of Engineering in Computer Science & Engineering

By

Devika Anil	1MS15CS040
Aravind P Anil	1MS15CS024
Aatish Kayyath	1MS15CS002
Abishek Padaki	1MS15CS005

Under the guidance of

Dr. S. Rajarajeswari
Associate Professor

M S RAMAIAH INSTITUTE OF TECHNOLOGY
(Autonomous Institute, Affiliated to VTU)
BANGALORE-560054
www.msrit.edu
May 2019

RAMAIAH Institute of Technology

Department of Computer Science & Engineering

CERTIFICATE

Certified that the project work entitled "Windows Based Malware Prediction Using Deep Learning Techniques" carried out by Devika Anil – 1MS15CS040, Aravind P Anil – 1MS15CS024, Aatish Kayyath – 1MS15CS002 and Abishek Padaki – 1MS15CS005, a bonafide student of M.S.Ramaiah Institute of Technology Bengaluru in partial fulfilment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belgavi during the year 2018-19. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the department library.

The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

Project Guide **Head of the Department**

Dr. S. RAJARAJESWARI **Dr. ANITA KANAVALLI**

External Examiners

Name of the Examiners: **Signature with Date**

1.

2.

RAMAIAH Institute of Technology

Department of Computer Science & Engineering

DECLARATION

We, hereby, declare that the entire work embodied in this project report has been carried out by us at M.S.Ramaiah Institute of Technology, Bengaluru, under the supervision of **Dr. S. RAJARAJESWARI, Associate Professor,** Dept of CSE. This report has not been submitted in part or full for the award of any diploma or degree of this or to any other university.

Signature
Aravind P Anil
1MS15CS024

Signature
Aatish Kayyath
1MS15CS002

Signature
Devika Anil
1MS15CS040

Signature
Abishek Padaki
1MS15CS002

ACKNOWLEDGEMENT

We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We would like to express our profound gratitude to the Management and **Dr. N.V.R Naidu** Principal, M.S.R.I.T, Bengaluru for providing us with the opportunity to explore our potential.

We extend our heartful gratitude to our beloved **Dr. Anita Kanavalli**, HOD, Computer Science and Engineering, for constant support and guidance.

We whole heartedly thank our project guide **Dr. S. Rajarajeswari,** for providing us with the confidence and strength to overcome every obstacle at each step of the project and inspiring us to the best of our potential. We also thank her for her constant guidance, direction and insight during the project.

This work would not have been possible without the guidance and help of several individuals who in one way or another contributed their valuable assistance in preparation and completion of this study.

Finally, we would like to express sincere gratitude to all the teaching and non-teaching faculty of CSE Department, our beloved parents, seniors and my dear friends for their constant support during the course of work.

Aravind P Anil

Devika Anil

Aatish Kayyath

Abishek Padaki

ABSTRACT

Malware attack is a very large domain of cybersecurity attacks. Cryptanalysts across the world has been in a long drawn out battle which has intensified in the past decade with malware. With increase in the technological capabilities of computers, there is also a distinct sharp increase in the capabilities of what malware can do and more importantly, how a malware can prevent from being detected. This project aims at developing a model which accurately predicts the probability that Windows operating system will be hit by a malware. It works on an operating systems dataset that has over 7 million recorded operating systems and their various features, generated by combining heartbeat and threat reports.

In order to understand the working of a wide variety of models on such a problem, three different models will be developed and assessed for its accuracy at predicting malware. Three different models under consideration are recurrent neural network, LightGBM technique and lastly a factorization method called XDeepFM.

This approach is assessing vulnerability of the system rather than the attacker. If the attacker is constantly evolving and learning new techniques against the system's defence, then efforts to defend against certain types of attacks are futile. Hence, predicting an attack in a more generic sense before it has even happened by assessing the system itself is the better alternative. Even though there are variants, a malware always targets a vulnerability or an exploit of the system to attack. If these weak points on the system are found and patched up before an attack happens, we can develop a very secure and malware proof security configuration.

Table of Contents

List of Tables

List of Equations

List of Figures