C02047: Doctor of Philosophy
CRICOS Code: 058666A
33875 PhD Thesis: Computer Systems
March 2021

A Study on

Feature Analysis and Ensemble-based Intrusion Detection Scheme using CICIDS-2017 dataset

Upasana Nagar

School of Electrical and Data Engineering
Faculty of Engineering & IT
University of Technology Sydney
NSW - 2007, Australia

A Study on

Feature Analysis and Ensemble-based Intrusion Detection Scheme using CICIDS-2017 dataset

A thesis submitted in partial fulfilment of the requirements
for the degree of

Doctor of Philosophy
in
Computer Systems

by

Upasana Nagar

under the supervison of

Dr. Priyadarsi Nanda

to

School of Electrical and Data Engineering
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

March 2021

Abstract

One of the primary security research challenges faced by traditional IDS methods is their inability to handle large volumes of network data and detect modern cyber-attacks with high detection accuracy and low false alarms. Hence, there is a need for efficient and reliable IDS schemes that can tackle this ever-changing cybersecurity paradigm. Machine learning techniques are hence, becoming very popular in designing modern intrusion detection systems. Several supervised and unsupervised machine learning techniques have been used in literature; however, the IDS classification efficiency is affected by noisy data in high dimensional datasets. The role of feature selection is significant as the feature selection process eliminates the redundant and noisy data and further selecting optimal feature subset enables reduction of high dimensional IDS datasets. Machine learning algorithms are extensively being used for intrusion detection. However, research has proved that the performance of multiple classifier-based IDS is far better than an IDS classifier, which has given us the motivation to develop an ensemble-based intrusion detection model. Lastly, the benchmark IDS datasets currently being used for the evaluation of IDS schemes are outdated and do not represent modern-day attacks. The CICIDS - 2017 dataset is offered by the University of New Brunswick. It is the latest publicly available dataset for intrusion detection. However, there are a significantly low number of research studies conducted using this dataset which also focus on optimal feature selection. This dataset has a good potential to be used as a future benchmark intrusion detection dataset as it covers the modern-day system setup and threat profile and the dependency on outdated IDS datasets can be removed. There is a need to benchmark the performance of modern IDS datasets using machine learning ensemble-based classifiers. This thesis aims to address the issues by proposing a new intrusion detection framework using ensemble-based feature selection method for generating a low dimensionality feature subset and ensemble-based intrusion detection framework to benchmark the performance of the CICIDS - 2017 dataset. The proposed scheme is beneficial for research community as it combines the use of the latest available IDS dataset with ensemble technique for feature selection and ensemble-based intrusion detection model.

AUTHOR'S DECLARATION

I, *Upasana Nagar* declare that this thesis, is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *School of Electrical and Data Engineering*, *FEIT* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

SIGNATURE: Production Note:
Signature removed prior to publication.

[Upasana Nagar]

DATE: 30th March, 2021

Dedication

I dedicate this achievement to the Supreme Almighty ...
For bestowing the blessing called "Life" ...
&
Immensely grateful for His guidance to "Turn Inwards" !!!

***

I also dedicate this achievement to my parents ...
Papa & Mumma, you are my friend, philosopher and guide !!!!

***

ACKNOWLEDGMENTS

I am immensely grateful to the Almighty God for His kind "Grace". He has helped me conquer my fears with His guiding light, helping me achieve this Doctor of Philosophy Degree.

At this moment of accomplishment, I would like to express my deepest gratitude and appreciation to my supervisor Dr. Priyadarsi Nanda. This dissertation would not be possible without his able guidance and unwavering support. I am grateful for your expert feedback and valuable suggestions that helped me progress in my research. I am indebted to your kindness in offering your time and availability whenever I came to you for research discussion. Thank you for always being the extremely patient, positive, understanding, flexible and motivational guide, through out my research tenure. You gave me hope whenever I felt hopeless. Thank you for carving the researcher in me and showing me how to focus on seeking solutions to problems and always keeping a positive approach, while working on my research. I am blessed to have you as my mentor and my "Guru" for you have never given up on me even in times when I did give up on myself. I also wish to thank my co-supervisor Dr. Xiangjian He, for his valuable feedback and guidance during my research. Thank you for always offering kind and positive words of encouragement to keep me motivated in pursuing my research.

I would also like to thank my senior researchers Dr. Zhiyuan Tan, Dr. Aruna Jamdagni and Dr. Mohammed Ambu Saidi for their time and valuable guidance during the beginning of my research that helped me move ahead with my work. Thank you for all the knowledge sharing sessions that helped me build my research foundations. Many thanks to Eriyani and Aprillia from the School of Electrical and Data Engineering at the University of Technology Sydney for helping me with guidance with administrative tasks. I would also like to acknowledge Dr. Chandranath Adak for providing the PhD thesis template without which writing this thesis would have been an uphill cumbersome task. I would like to thank John Hazelton for his meticulous and prompt proof reading for my dissertation.

I am grateful for my time spent at UTS doing my research work as this gave me an opportunity to foster some great friendships. Big heartfelt thanks to Ashish Nanda and Amber Umair for always being there to inspire and motivate during my research. You both have helped me see the silver lining behind the dark clouds when I needed

it most. Thank you to Annie Baskaran, Nazar Waheed, Chau Nguyen, Nisha Malik, Madhumita Takalkar and Sara Farahmandian for their wonderful camaraderie, friendly banter, advice and support during my research and even outside. A very special thank you to my lovely friends Rattandeep Kaur, Sai Kiran Tadepalli (Sai Garu) and Ramya Bati for their nurturing friendships.

I take this opportunity to thank my father Dr. T.N.Nagar, and my mother Mrs. Shobha Nagar, for their unconditional love, trust, support, encouragement, guidance and motivation in my life, especially as I encountered challenges during my research journey. You have taught me the power of prayer and built in me the resilient, "never give up" attitude towards life. You inculcated values of dedication, perseverance, sincerity, integrity, truthfulness and loyalty. These golden values have made me accomplish my goals and become who I am today. You have always reinforced the belief in me that I can move mountains if I put my mind to it! Thank you from the bottom of my heart for being my safety net and giving me the mental strength and confidence to realize my dreams into reality. I owe you my existence and I can not thank God enough for blessing me with the most awesome and caring parents. This PhD has been made possible only because of your presence and relentless support in the final year of my research. I would like to dedicate my thesis to you both.

Special thanks to my younger sister, Kankana Nagar, for being the playful, happy person and also for being the wiser one between us. You always brighten up my moods and lift my spirits with your presence.

I thank my mother-in-law, Hurshit Nagar, for her love, best wishes and blessings. I would also like to fondly remember my late father-in-law, Hursh Nagar, whose blessings are always with me. He would be so proud of my big achievement.

Last but certainly not least, I would also like to thank my incredible husband, Prasannavadanam Nagar for believing in my dream and showering unconditional love and support during my research journey and in life. My heartfelt gratitude goes to my precious children and my forever cheerleaders, Aum and Anika. Their magical smiles always inspire me to keep giving my best in life.

Words fail to express my gratitude towards my extended family and all my family friends, who have always encouraged me to chase my research goals. I wish to conclude this acknowledgement, with a quote from one of my favourite books, "The Alchemist", by Paulo Coelho :

"And, when you want something, all the universe conspires in helping you to achieve it."

List of Publications

1. Nagar, U, Nanda, P & He, X Feature Analysis and Ensemble-based Intrusion Detection Scheme using CICIDS - 2017 dataset. (Submitted to Wiley, Concurrency and Computation Practice and Experience Journal.)

2. Nanda, P, Arain, A & Nagar, U 2019, 'Network Packet Breach Detection using Cognitive Techniques', Smart Systems and IoT: Innovations in Computing, Second International Conference on Smart IoT Systems - Innovations in Computing (SSIC-2019), Springer, Jaipur, India.

3. Nagar, U, Nanda, P, He, X & Tan, Z 2017, 'A Framework for Data Security in Cloud using Collaborative Intrusion Detection Scheme', SIN'17: Proceedings of the 10th International Conference on Security of Information and Networks, International Conference On Security Of Information And Networks, ACM Digital Library, Jaipur, India, pp. 188-193.

4. Ambusaidi, MA, He, X, Tan, Z, Nanda, P, Lu, L & Nagar, U 2014, 'A novel feature selection approach for intrusion detection data classification', 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE Computer Society, Beijing, pp. 82-89.

5. Tan, Z, Nagar, U, He, X, Nanda, P & Liu, R 2014, 'Enhancing Big Data Security with Collaborative Intrusion Detection', IEEE Cloud Computing Magazine, pp. 34-40.

TABLE OF CONTENTS

List of Figures

Figure **Page**

List of Tables