C02047: Doctor of Philosophy
CRICOS Code: 00099F
Subject Code: 33875
December 2019

Machine Learning based Information Forensics from Smart Sources

Amber Umair

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

Machine Learning based Information Forensics from Smart Sources

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

Doctor of Philosophy
in
Computer Systems

by

Amber Umair

to

School of Electrical and Data Engineering
Dr. Priyadarsi Nanda
University of Technology Sydney
NSW - 2007, Australia

December 2019

Abstract

We live in a world that connects, socializes and interacts using internet. Humans generate tons of information on daily basis, according to Forbes 2.5 quintillion bytes of data is created each day in year 2018. The data creation pace is continuously accelerating with the growth of the Internet of Things (IoT). Extensive social media usage fuels data creation which is primarily generated from mobile phones. In the present scenario, data is the asset and this asset is extremely vulnerable. In our research work we utilized these data sources to aid digital forensics investigation. Due to our technology engulfed lifestyle, we leave a lot of information about ourselves during our routine activities. These traces are used by the wrongdoers for their vicious objectives but also these traces can be used by investigators to understand any incident and to penalize the delinquents. Forensics investigators face challenges with a huge amount of data during investigations. Whether the data source is an online social network, a smart phone or an IOT based environment, huge amount of data adds complexity and delay to forensics investigation. To contribute to the forensics investigation we propose the use of machine learning for forensics data anaylsis. Forensics investigation is a three-phase process including data acquisition, data analysis and presentation. Our research focuses on the first two phases of the forensics investigation cycle i.e. data collection and data analysis. This thesis discusses following research achievements:

1. Data acquisition from smart sources for forensic information especially for IoT
2. Machine learning based data analysis to extract forensic artefacts
3. IoT forensics framework(acquisition and analysis phase) implementation

This thesis is segmented based on the three data sources for data analysis namely online social networks, smart phones and sensor-based networks (IoT). Using IoT based data this thesis proposes a scheme SACIFS(Smart aged care information forensics) for IoT forensic feature extraction and data analysis of elderly patients monitored in a nursing home environment. We developed our machine learning model based on Support Vector Machine to detect an incident and highlight relevant forensic artefacts. We used smart phone data in Digital Forensic Intelligence Analysis Cycle framework to identify strongly connected contacts during triage phase. We classified the contacts of a smart phone user with respect to their closeness, extracted from the data features from Facebook messenger. Moreover, utilizing publically available Online Social Network, we analysed multiple tools to collect, analyse and visualize data from Facebook pages and groups.

AUTHOR'S DECLARATION

I, *Amber Umair* declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Electrical and Data Engineering*, *Faculty of Engineering and Technology* at the University of Technology Sydney, Australia, 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.

[Amber Umair]

DATE: 28th December, 2019

PLACE: Sydney, Australia

Dedication

To my beloved Parents, my loving Husband Umair and my kids Abdullah and Anabia

. . .

ACKNOWLEDGMENTS

Foremost I would like to thank Allah, the sustainer, who kept me going through the highs and lows of this journey and made it achievable for me. He undoubtly is the source of light for me in every aspect of life.

I would like to express my deepest gratitude to my supervisor, Professor Priyadarsi Nanda for accepting me into the Phd. program and giving me the honor of being one of his research students. Prof. Nanda's guidance, patience, and continuous support were essential to the completion of this thesis and to my formation as a researcher. His unmatched knowledge and invaluable feedback immensely helped me go forward in my research. I am greatly indebted to him for helping me get through the difficult times I had during my research. I feel very lucky to have worked with Prof. Nanda without whom this thesis would have only been a dream. I would like to thank my Co-supervisor Professor Xiangjian He, who always provided his valueable and expert feedback on my work. His support, co-operation, and generosity thoughout the research tenure is truely undeniable.

I also wish to thank all of my colleagues and friends from the School of Electrical and Data Engineering at the University of Technology Sydney. Specifically, I would like to thank Upasana, Ashish, Annie, Amjad and Nisha for creating a friendly atmosphere in the group and assisting me in whatever manner possible. Many thanks to all the administrative staff at the School of Electrical and Data Engineering, especially Thomas, Eryani and Aprillia who helped me with administrative issues.I also acknowledge Chandranath Adak (UTS) for providing this thesis template to all HDR students and easing the thesis preparation procedure.

I am eternally grateful to my loving and supportive husband, lovely kids, beloved parents and parents in law for their sacrifices, prayers, encouragements, and endless love. Without their everyday support, I would have not been able to reach this stage. I sincerely dedicate this thesis to them. Words cannot express my gratitude and appreciation for their unwavering support and kindness throughout this journey.

List of Publications

Related to the Thesis :

1. Online Social Network Information Forensics Tools analysis and a survey of how cautious facbook users are.

   A. Umair, P. Nanda and X. He, "Online Social Network Information Forensics: A Survey on Use of Various Tools and Determining How Cautious Facebook Users are?," 2017 IEEE Trustcom/BigDataSE/ICESS, Sydney, NSW, 2017, pp. 1139-1144. doi: 10.1109/Trustcom/BigDataSE/ICESS.2017.364

2. User Relationship Classication of Facebook Messenger Mobile Data using WEKA

   Umair A., Nanda P., He X., Choo KK.R. (2018) User Relationship Classification of Facebook Messenger Mobile Data using WEKA. In: Au M. et al. (eds) Network and System Security. NSS 2018. Lecture Notes in Computer Science, vol 11058. Springer, Cham

3. SACIFS: Smart Aged Care Information Forensics System using Machine-Learning (To be submitted to a suitable journal)

Table of Contents

List of Figures

List of Tables