

**Masters Project Final Report December 2015**

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Title** |  | | |
| **Student Name** | J R O Sirimanne | | |
| **Registration No. & Index No.** | 2013/MIS/021  13 | | |
| **Supervisor’s Name** | Dr. Chamath Keppitiyagama | | |
| **Please Circle the appropriate** | **Masters Program** | **Type** | |
| **MIS** | **Research** | **Implementation** |
| **For Office Use Only** | | | |
|  | | | |

## De-Anonymization of Anonymous

## E-mails in digital forensic

**W.A.V.M.G.Wickramasinghe**

**2014**

## 

## De-Anonymization of Anonymous

## E-mails in digital forensic

**A dissertation submitted for the Degree of Master of  
Science in Information Security**

**W.A.V.M.G.Wickramasinghe**

**University of Colombo School of Computing  
2014**



**Declaration**

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Students Name: W.A.V.M.G.Wickramasinghe

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature: Date:

This is to certify that this thesis is based on the work of Mr. W.A.V.M.G.Wickramasinghe under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Dr.H.A.Caldera

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature: Date:

**Abstract**

Rapid growth of e-mail users and popularity of internet also increase the number of cybercrime related with e-mail. Anonymity provides sophisticated mechanism to preserve privacy of user’s online life. Internet has become an ideal place for criminals over the past decades due some of its characteristics. Misuse of Anonymity by Cyber criminals for their wrong doings is a threat to regular online users. E-mail the modern solution to the snail mail is the most frequent method of cybercrime communication.

Individual writing style may contain hidden patterns that can use to uniquely identify an individual using his writing samples. Traditional method of digital forensic on e-mail does not address the use of authorship analysis techniques. E-mail header only based e-mail forensics may mislead the entire investigation towards a wrong direction and may prosecute the innocent at the end. Mining of information from the e-mail text will provides the knowledge about the true authorship of the text. Mining of information on e-mail can use statistical or machine learning based methods. Use of content based authorship analysis in e-mail forensic will increase the probability of proving guiltiness of a criminal in the court of law.

This project focus on use of computational stylometry and related text mining techniques in natural language processing to analyze and identification of authors writing discriminators for recognize the best matching author of an anonymous e-mail.

# **Acknowledgement**

I wish my gratitude

**Table of Contents**

This should list all the chapters, sections, and subsections of the report giving the page number on which each starts. Word processors normally generate this automatically.

**List of Figures**

A list giving the number, title, and the page of each figure used in your report should be provided.

**List of Tables**

A list giving the number, title, and the page of each table used in your report should be provided.

**List of Abbreviations**

|  |  |
| --- | --- |
| ARFF | Attribute Relation File Format |
| BJA | Bureau of Justice Assistance |
| Email | Electronic mail |
| FBI | Federal Bureau of Investigation |
| FERC | Federal Energy Regulatory Commission |
| GUI | Graphical User Interface |
| IC3 | Internet Crime Complaint Center |
| IETF | Internet engineering task force |
| NLP | Natural Language Processing |
| NW3C | National White Collar Crime Center |
| PHI | Personal Health information |
| PII | Personally identifiable information |
| PIN | Personal identification number |
| RFC | Request for comments |
| SVM | Support Vector Machine |
| TOR | The Onion Routing |
| WEKA | Waikato Environment for Knowledge Analysis |
|  |  |
|  |  |
|  |  |

# **INTRODUCTION**