



# ANALYSIS EMAIL FILTERING IN GOVERNMENT NETWORK

MOHD NASRAN BIN HASAN

MASTER OF COMPUTER SCIENCE  
(INTERNETWORKING TECHNOLOGY)

2017



## **Faculty of Information and Communication Technology**

# ANALYSIS EMAIL FILTERING IN GOVERNMENT NETWORK

**MOHD NASRAN BIN HASAN**

MASTER OF COMPUTER SCIENCE  
(INTERNETWORKING TECHNOLOGY)

**2017**

ANALYSIS EMAIL FILTERING IN GOVERNMENT NETWORK

**MOHD NASRAN BIN HASAN**

**A thesis submitted  
in fulfillment of the requirements for the degree of Master of Computer Science.  
(Internetworking Technology)**

**Faculty of Information and Communication Technology**

**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

**2017**

## DECLARATION

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Computer Science. (Internetworking Technology)

Signature : .....

Supervisor Name : Dr. Mohd.Faizal Bin Abdollah

Date : .....

## APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in term of scope and quality for the award of Master of Computer Science (Internetworking Technology).

Signature : .....

Supervisor Name : .....

Date : .....

## **DEDICATION**

*I lovingly dedicate my thesis to my beloved wife Saleha Mohd Hussin and My*

*Father Hasan Bin Omar who supported me each me step way.*

*To my precious sons, Rabiatul Adawiyah Mohd Nasran, who give me passion and strength with their smile. You are “ Anugerah Allah yang Terindah”*

## ABSTRACT

*Nowadays, email filtering at all still in testing for suitability. The same thing also with Melaka State Development Corporation. Email plays an important role in this department. Thus, anti-spam to be an important component in improving the security of email security. It can check all incoming and outgoing emails to the server. This paper provides an overview of Email, importance, types and studies the effects of implementing email anti-spam email to drive performance.*

## ABSTRAK

*Pada masa kini, penapisan email di semua masih dalam ujian untuk mendapatkan kesesuaian. Perkara ini sama juga dengan Perbadanan Kemajuan Negeri Melaka. Email memainkan peranan yang penting dalam jabatan ini. Jadi, anti-spam menjadi merupakan komponen penting dalam meningkatkan keselamatan keselamatan email. Ia boleh memeriksa semua email yang keluar dan masuk ke dalam server. Kertas kerja ini memberi gambaran keseluruhan tentang Email, kepentingannya, jenis email dan mengkaji kesan melaksanakan anti-spam kepada prestasi perjalanan email.*



## ACKNOWLEDGEMENTS

Immeasurable gratitude to Allah S.W.T. for giving this servant an opportunity to undertake and complete this piece of work.

I would like to express my gratitude to all those who gave me the possibility to complete this independent study. I am deeply indebted to my supervisor Dr Mohd Faizal Bin Abdullah whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this independent study.

My former colleagues from the Master of Science (Internetworking) Faculty of Information Technology And Communication supported me in my research work. I want to thank them for all their help, support, interest and valuable hints. To my parents whose encouragements and wishes for me to succeed in educational field; to all my family members whose warm supports and understanding to motivate me to write and produce this thesis as a token of my strong interest in education.

Last but not least, I would like to give my special thanks to Saleha Mohd Hussin whose patient love enabled me to complete this work.

## TABLE OF CONTENTS

	PAGE
<b>DECLARATION</b>	
<b>APPROVAL</b>	
<b>DEDICATION</b>	
<b>ABSTRACT</b>	I
<b>ABSTRAK</b>	II
<b>ACKNOWLEDGEMENTS</b>	III
<b>TABLE OF CONTENTS</b>	V
<b>LIST OF TABLES</b>	IV
<b>LIST OF FIGURES</b>	IX
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	
1.1 Introduction	1
1.2 Research Background	2
1.3 Research Problem	3
1.4 Research Question	4
1.5 Research Objective	4
1.6 Research Scope	5
1.7 Research Contribution	5
1.8 Thesis Organisation	5
<b>2 LITERATURE REVIEW</b>	
2.1 Introduction	9
2.2 Background of Research Organization	10
2.3 Email Spam Overview	12
2.4 Email Filtering Overview and Techniques	13
2.4.1 Origin-Based Filters	15
2.4.1.1 Blacklists	15
2.4.1.2 Whitelists	18
2.4.1.3 Challenge/Response Systems	18
2.4.2 Content Filters	19
2.4.2.1 Bayesian Filters	20
2.4.2.2 Rule-Based Filters	20
2.4.3 Other Filters	21
2.4.4 Sender Authentication Systems	21
2.4.4.1 Reverse DNS	22
2.4.4.2 The Sender Policy Framework	23
2.4.4.3 DomainKeys	25
2.5 Email Spam Characteristic	26
2.5.1 Email headers	27
2.5.2 Content of Message	29
2.6 False Positives and False Negatives	30
2.7 Summary	31

<b>3</b>	<b>RESEARCH METHODOLOGY</b>	
3.1	Introduction	32
3.2	Chapter Outline	32
3.3	Research Methodology	33
3.3.1	Phase 1: Analysis – Antispam Selection	34
3.3.2	Phase 2: Design and Development	34
3.3.3	Phase 3: Implementation	35
3.3.4	Phase 4: Testing and Evaluation	36
3.4	Research Tools and Project Requirement	36
3.5	Project Schedule and Milestone	37
3.6	Summary	38
<b>4</b>	<b>IMPLEMENTATION</b>	
4.1	Introduction	39
4.2	Technical Implementation	40
4.2.1	SMTP Log Collection	46
4.2.2	SMTP Log Analysis	47
4.3	Anti-Spam Filtering Technique To Choose	47
4.3.1	Bayesian Anti-spam Filtering	47
4.3.2	DNS Blocklists Anti-spam Filtering	49
4.3.3	REVERSE DNS	50
4.3.3.1	Configuration PKNM Public RDNS	52
4.4	Summary	54
<b>5</b>	<b>EVALUATION AND TESTING</b>	
5.1	Introduction	55
5.2	Experiments Result for Real Data	55
5.3	Anti-Spam Filtering Technique Comparison	58
5.3.1	Anti-Spam Filtering Technique Comparison with User Feedback	58
5.3.2	Real Time Black Lists spam Filtering Comparison	59
5.4	Anti-Spam Filtering Using Bayesian Algorithms	63
5.5	Anti-Spam Filtering Using DNS Black Lists	64
5.6	Anti-Spam Filtering Using Reverse DNS	64
5.7	Summary	65
<b>6</b>	<b>CONCLUSION AND DISCUSSION</b>	
6.1	Introduction	66
6.2	Summary of Research	66
6.3	Summary of Contributions	66
6.4	Limitation of research	67
6.5	Incoming work	67
6.6	Conclusion	68
	<b>REFERENCES</b>	69

## LIST OF TABLE

<b>TABLE</b>	<b>TITLE</b>	<b>PAGES</b>
1	Spam Characteristic	27
2	Milestone for master project I and11	38
3	Summary of hardware and software used in the experiment	44
4	Item names and result values in the database	56
5	Spam percentage detected by Anti-Spam Filtering In the experiment	56
6	Anti-Spam Filtering Technique Comparing with User feedback	58
7	Real time black list spam by the percentage detected in the experiment	60
8	Comparison Email Spam Detected By Country	62

## LIST OF FIGURE

FIGURE	TITLE	PAGE
1	The population of spam in e-mail traffic 2015	3
2	Research Structural Process	6
3	Melaka State Development Corporation Network	11
4	The Complaint for regarding Email Spam Issues	12
5	How anti-spam Technique work	14
6	An Overview Blacklisting Technique by Spamhaus.org	16
7	An overview Reverse DNS work	22
8	Sample SPF record	23
9	An overview Sender Policy Work	24
10	An overview Domain Keys Technique	25
11	The Structure Of Chapter Three	32
12	Main Phases of Research Methodology	33
13	Pre-liminary Selection in Analysis Phase	34
14	Design and Development Phase	35
15	Implementation Phase	35
16	Testing and Evaluation Phase	36
17	Network Diagram before project Implementation	40
18	Project Diagram Implementation	41
19	An incoming and rejected e-mail route from SMTP Server	41
20	Posfix in a Zimbra E-mail Server Environment	42

21	E-mail Filtering with spammasian Project	43
22	Spammasian Interface 1	45
23	Spammasian Interface 2	46
24	How Bayesian Filtering Work	48
25	The interface for user enter the filtering by word in Zimbra server	48
26	Zimbra Administration pages	49
27	Interface in User Inbox How User To add Block Message from e-mail receiving	50
28	Conventional DNS Resolution	51
29	Reverse DNS Resolution	51
30	Response from Telekom regarding the request to do PTR record	52
31	Command to configure SMTP Banner in Zimbra server to match with the external RDNS	53
32	Using MX tool website to check configuration for PTR record	53
33	Spam percentage detected by Anti-Spam Filtering Technique in the experiment	57
34	Anti-Spam Filtering Technique Comparison with User mail Inbox	59
35	The message worked as a spam using blacklist	61
36	Spam Detected By Country	61
37	Chart Spam Detected Percentage By Country In Melaka State	61

Development Corporation

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

This study was conducted study to “Analysis Email Filtering Techniques for Melaka State Development Corporation In Government Network”. Malacca State Development Corporation operates in the area of Tower MITC Melaka International Trade Centre where the central administration of the State Government of Melaka in Ayer Keroh, Melaka.

It has 11 sections which have different functions at each other. Main objective of the Melaka State Development Corporation was established Melaka Spur Economic Development in the Field of Industrial Property, Entrepreneurship, Investments and Exploring New Business Opportunities Creative & Innovative Towards the Year 2020.

Malacca State Development Corporation has 11 departments have different tasks to each other. It is divided into two main departments, Department of Technical and Management. In facilitating the management of parts of the communication between each other is very important. Furthermore, the Melaka State Development Corporation of dealing with foreign investors, tenants of the building and the SMI, as well as other government departments.



Effective communication, faster and cheaper is needed in the management of government organizations. There was also a rapid service delivery and regularly is important in the management of government departments. Email is preferred for this communication other than the telephone line. It is because the ability to communicate using email for managing the transmission of information documents, pictures and so more easily. However, with the development of technology seumpana there are also other problems that occur. Among the spam email. I not only can interfere with daily work email users even in the present nie it is able to spread a computer virus and may result in attack by computer hackers to steal information.

## **1.2 Research Background**

E-mail is one of the most popular communications over internet today. However, each day we spent time to delete spam, unsolicited e-mails advertising, offering loans at low interest rates, drugs and others. The spam filters are able to check the majority of the e-mail spam and at the same time, spammers are continuously developing new techniques to send spam messages to more and more people. Using advance of technology mobile devices and other portable electronic devices are now Wi-Fi enabled and internet telephony VoIP (voice over internet protocol) has made communicating across the world easier and inexpensive. Social networks like Facebook are very popular means of connecting with others across the world using the internet.

There are many ways spammers can get to know your e-mail address and send you spam even though you never open any spam mails or click any suspicious links. If you are on any social network and do not set your privacy settings your data is available to anyone which includes your address, e-mail and also your friend lists. And if you have subscribe

with newsgroups your E-mail address can be easily harvested. Dictionary attack is also of one technique to harvest e-mail addresses. So it is easy to find information with less time and effort and spammer have a lots of it. Most of the spammers do the job using bots so even if they get just one user effected by their spam it is worth the effort to send to hundreds of people e-mail. The Population of spam in the world according report from Kapersky 2015 email traffic is refer to figure 1. Population Spam in mail traffic is not the same throughout the year, where in 2015 it was found spam to be identified is high at the beginning of the year. And progressively less until the end of the year.

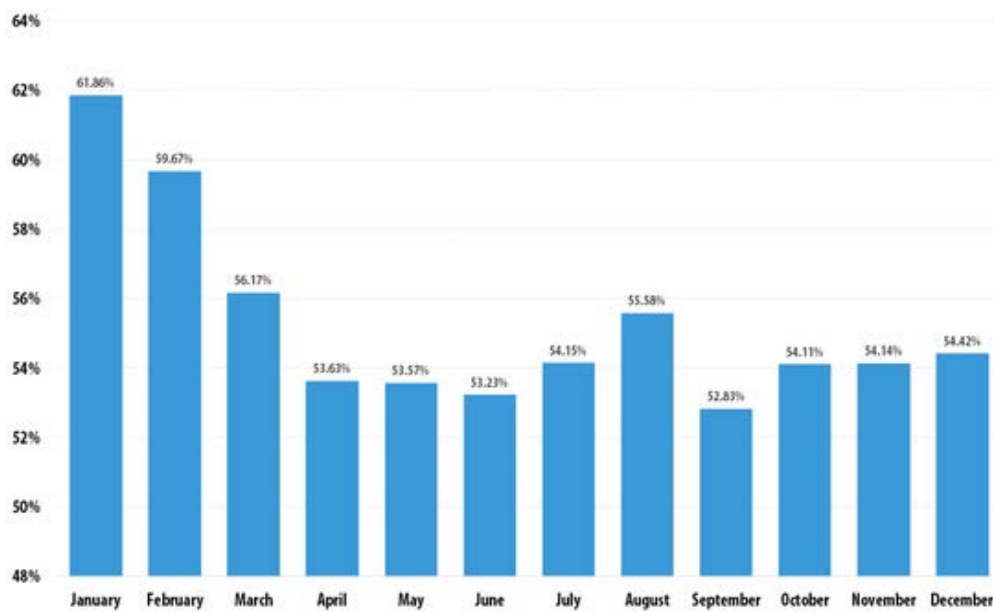


Figure 1: The population of spam in email traffic, 2015

### 1.3 Research Problem

The exponential growth of spam is become more serious and caused a lot of problem to Perbadanan Kemajuan Negeri Melaka staffs. With the development of increasingly

sophisticated technology, e-mail is preferred in order to facilitate communication, accelerate work time and also a government department. Nowadays, spammers are more intelligent because they have try to evade their spam email from being detect by anti-spam solution by perform some word obfuscation on the email keyword. This thesis examines different of technique to reduce the number of spam messages coming to user inbox and proposed a spam solutions to Melaka State Development Corporation in managing and overcome spam problem.

#### **1.4 Research Question**

Referring to the Research Problem in Section 1.1, three research questions are formed to represent the research problems which are:

- RQ1. What is the best possible Anti-spam Filtering Techniques?
- RQ2. How to analyses the significant parameters for Email Filtering Techniques?

#### **1.5 Research Objective**

Based on the research questions formulated in Section 1.2, the research objective has developed as follows:

- RO1. To study anti-spam technique
- RO2. To analysis and compare current anti-spam technique.
- RO3. To propose the best anti-spam technique.

## **1.6 Research Scope**

In order to achieve the Research Objectives, this research will be focus on some issues as stated below:

Simple approach method of measuring accuracy of current anti-spam filtering technique to implement in Melaka State Development Corporation at Government Sector.

## **1.7 Research Contribution**

Based on the research method formulated in Section 1.6, the research contribution has developed as follows:

- RC1. An enhancement prediction the best Anti-Spam Filtering Technique.
- RC2. Approach simple method to prove the outcome on propose the the best anti-spam filtering technique for future use.

## **1.8 Thesis Organization**

This thesis is structured into seven chapters, each focusing on areas related with this research project. A brief outline of each chapter is as follows:

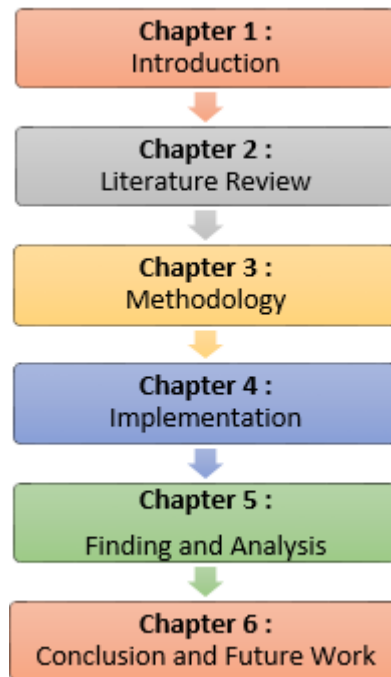


Figure 2: Research Structural Process

### **Chapter One: Introduction**

Introduction describes the problem area of this research in managing huge amount of spam mail alerts which are exhaustive for email administrator due to the false positive alert in anti-spam filtering techniques. Some of the issues related to the anti-spam filtering are briefly discussed. Thereafter, the research question, research aim and research objectives are stated. The operational framework is then designed in line with the research objectives and research significant contributions as stated. This chapter ends with the thesis outline.

### **Chapter Two: Literature Review**

Literature review elaborates the problems described in Chapter One, starting by reviewing the email spam issues. The chapter then reviews the current approach of

antispam filtering in detecting email spam and alert correlation technique for SMTP log and presents analysis and critiques to the current technique in relation to their abilities to detect the email spam. This chapter fulfills the first research objective (RO1) which is to. Some reviews and analysis has also been covered related to alert correlation framework and; evaluation and validation issues.

### **Chapter Three: Research Methodology**

Chapter Three provides discussion on methodology of this research. The research methodology is developed in order to achieve all research objectives. There are four main phases involved in this research methodology which are Phase 1- Analysis, Phase 2-Design and Development, Phase 3-Implementation and Phase 4- Testing and Evaluation.

### **Chapter Four: Implementation**

Chapter Four describes the overview of the general anti-spam filtering techniques. This chapter also describes the preliminary experiment in constructing the spam attack pattern and spam attack model. Consequently, the email spamming patterns are analyzed.

### **Chapter Five: Evaluation And Testing**

This chapter will propose the implementation of anti-spam filtering better based on analysis done in Chapter Two. Anti-spam filtering better for SMTP Log. One scenario would be selected to represent the implementation of the regulations. The

aim of this chapter is to present the capabilities of the regulations in the selection of anti-spam techniques. This exercise should reduce spam e-mail in the inbox of individuals.

### **Chapter Six : Discussion and Conclusion**

This chapter concludes the study of the other chapters in this thesis. This chapter also presents the final conclusions and limitations of this research. Chapter Six ends by identifying some areas for future research in this area.

## CHAPTER 2

### LITERATURE REVIEWS

#### 2.1 Introduction

In this chapter, the study tries to understand the false positive problem generated by current anti-spam filtering to detect the email spamming, to overcome the problem in detail and to see the previous and current approaches of Email Filtering detection technique. The information gathered can be used to identify the prospect and limitations in reducing false positive alarm generated by Spam attack. Furthermore, it acts as a basis to propose a suitable approach to eliminate or at least minimize such problem. To achieve the above intention, some literature reviews on spam issue, email filtering problem, current approach used in detecting email spam and others related issue are reviews and analyze. Findings from the literature reviews will discover the research gap and then the first and second research question (RQ1, RQ2) formulated in Chapter One and the first and second research objective (RO1, RO2) which are to analyze and classify the existing anti-spam detecting technique and to generate an improved generic taxonomy for email filtering technique respectively are accomplished in this chapter.