COMPARISON OF SPAM FILTERING MECHANISM

LESTA ANAK ENGKAMAT

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS *

JUDUL: COMPARISON OF SPAM FILTERING MECHANISM

SESI PENGAJIAN: 2010/2011

Saya LESTA ANAK ENGKAMAT
(HURUF BESAR)

mengaku membenarkan tesis (PSM/ Sarjana/ Doktor Falsafah) ini disimpan di Perpustakaan Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____ SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____ TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/ badan di mana penyelidikan dijalankan)
___/___ TIDAK TERHAD	

(TANDATANGAN PENULIS)
Alamat Tetap:Rumah Usek,
Sungai Kriok, 98200 Niah,
Sarawak.
Tarikh: 12 Julai 2011

(TANDATANGAN PENYELIA)
En. Mohammad Radzi bin Motsidi
Tarikh: 12 Julai 2011

CATATAN: *Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)

**Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

COMPARISON OF SPAM FILTERING MECHANISM

LESTA ANAK ENGKAMAT

This report is submitted in partial fulfillment of the requirements for the Bachelor of Science Computer (Computer Networking)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2011

DECLARATION

I hereby declare that this project report entitled

COMPARISON OF SPAM FILTERING MECHANISM

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ Date: 12 JULAI 2011
(LESTA ANAK ENGKAMAT)

SUPERVISOR : _____ Date: 12 JULAI 2011
(EN. MOHAMMAD RADZI BIN MOTSIDI)

DEDICATION

To my beloved parents, your care, your love and your support give me strength.

To my friends, it is for your support and motivation.

To my lecturer, for the guide until the completion of this project, it gives me the challenge to be a better person.

ACKNOWLEDGEMENTS

During the completion of this project, I gained a lot of new knowledge especially in email system. I also learned a new things in life, be strong whenever face a problem and do not give up because of failed many times before. I would like to thank all people that involved with this project.

To my supervisor Mr. Mohammad Radzi bin Motsidi, I would like to thank him for assisting me in completion of this project. He helped me a lot by giving me ideas and suggestions on how to run this whole project. I really appreciate it.

Apart from that, I also want to express my sincere thanks to my all my friends who gave me the support and motivation during the completion of this project.

Lastly, I would like to express my deepest thanks to my family especially my mum who has give me support in completion of this project.

ABSTRACT

Email spam or unsolicited bulk email (UBE) that usually sent to many people either through email. Spam usually wasting our time by deleting it one by one in our mail box, and also eats a lot of network bandwidth. Nowadays, there are lots of organizations and individuals take a step to fight spam with variety of techniques. This research paper is focus on two different types of spam filtering software which are SpamAssassin and ThunderBayes. A testing done to make comparison on both of this spam filtering software based on how this program filter and classified email as spam. The results of test show which program is effective to be used as spam filter software on mail server.

ABSTRAK

Email spam ataupun *unsolicited bulk email (UBE)* biasanya dihantar ke ramai orang sekaligus melalui emel. Spam selalunya membazir masa kita kerasa kita perlu membuang emel tersebut satu per satu di dalam Inbox emel kita. Selain itu, spam juga memakan banyak *bandwidth* rangkaian. Pada masa kini, kebanyakan organisasi dan juga individu mengambil langkah untuk melawan spam dengan pelbagai teknik. Kajian ini tertumpu pada dua jenis program yang digunakan untuk menapis emel spam, iaitu *SpamAssassin* dan *ThunderBayes.* Ujian dilakukan untuk membandingkan bagaimana kedua-dua program ini menapis dan mengklasifikasikan emel tersebut adalah spam. Dari keputusan ujian tersebut, dapatlah diketahui mana satu program yang lebih efektif untuk menapis emel spam.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SMTP	Simple Mail Transfer Protocol
MTA	Mail Transfer Agent
MUA	Mail User Agent
POP3	Post Office Protocol 3
IMAP	Internet Message Access Protocol
IP	Internet Protocol
DNS	Domain Name System
FQDN	Fully Qualified Domain Name
UBE	Unsolicited Bulk Email
UCE	Unsolicited Commercial Email
TCP	Transmission Control Protocol
LAN	Local Area Network
TCP/IP	Transmission Control Protocol/Internet Protocol
SOM	Self-Organizing Map
ADG	Automatically Defined Groups
ASSP	Anti-Spam SMTP Proxy
PSFL	Python Software Foundation License
GNU	General Public License
MIME	Multipurpose Internet Mail Extensions
HTML	Hyper Text Markup Language
OS	Operating System

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

This chapter is the early reviews for this project, which will explain about the project that will be developed, that is spam filtering mechanism in email system. Sub-chapters that will be discussed in this chapter including project background, problem that related to this project, objectives of this project, scopes of the project, project significance, expected output of this project and the conclusion of this chapter.

1.1 Project Background

Most spam emails on the Internet today are advertisements from individuals and the occasional small business that looking for a way to make a fast money. E-mail spam, which also known as junk e-mail or unsolicited bulk e-mail (UBE) is a subset of spam that involves nearly identical messages sent to numerous recipients by e-mail. Definitions of spam usually include the aspects that e-mail is unsolicited and sent in bulk. Usually when user log in to their email, they can see few spam e-mail go through to their inbox. This spam sometime annoying because when it comes with large amount of spam email, it will take time to remove all the junk mail and also it use a lot of storage space on the email. Some spam emails also can contain malware. For regular users which don't have knowledge about malware, they just simply open the spam email that may contain malware. The malware can affect the user's computer and worst case which can happen is the malware will spread itself to

the entire network of the company or organization if the computer is connected to a company networks.

To overcome this problem, the spam filtering mechanism is important because it can recognize the spam and block the spam e-mail or move the spam e-mail to the Junk folder in client email.

For this project, spam filtering implementation will use two types of spam filtering tools, which are SpamAssassin and ThunderBayes. These two applications will be installed and configured alongside with email system so that it can used to filter spam emails. Testing will be done to compare the usage of the software and how the software work to identify email that classified as spam or ham.

1.2 Problem Statement

Spam emails wasting the Internet's two most precious resources, which are the bandwidth of long-distance communications links and the time of network administrator who keep the Internet working from day to day. Spam emails also wasting the time of computer users around the world, which they need to delete those spam emails. Furthermore, in order to deliver the emails, persons who send spam emails are increase which can lead to fraud and computer abuse.

1.3 Objective

There are few main objectives to be established during completion of this project. The objectives have been summarized as below:

i. To setup and configure mail server.

ii. To implement the spam filtering tool into mail server that will detect and filter spam i[illegible]

iii. To compare the implemented spam filtering mechanism software. The spam filtering tools that will be used is SpamAssassin and ThunderBayes.

1.4 Scope

The scope of this project is to implement and compare spam filtering tool which can be used to detect spam in email. The software that will be used to filter spam emails are SpamAssassin and ThunderBayes. The spam filtering tool will be implemented in Linux operating system environment. This spam filtering mechanism is recommended for any organization or small companies that provide email service to their users or workers.

1.5 Project Significance

This project will implement email system with spam filtering tool. This spam filtering tool will detect spam emails that sent to user's email box. For a large organization or perhaps a small company, an email system is an essential way to deliver latest information to employees. When there's update or new task that need to be done in the organization, administrator or whoever in charged with it can inform the workers inside the organization using emails.

For a large company which have a lot of branch, Internet connection needed in order to connect the email system. When it connected to Internet, the risk of being attack with spam emails is high. So, spam email mechanism needed to handle those spam emails. Because some spam email can contain malware which can lead to fraud, it can affect the security of the large organization. Malware such as spyware can stole information which stored inside the organization's server.

When the spam filtering tool implemented inside an email server, it can help large organization company filters spam emails.

1.6 Expected Output

This project will implement email system with spam filtering tool. This spam filtering tool will able to detect and filter spam or ham emails that incoming to user's mail box. Two types of spam filtering tool will be used and comparison should be made based of how the software filters the incoming emails. For SpamAssassin, it will implement on server side while ThunderBayes will implement in client side. Testing will be done in order to compare these two types of spam filtering tool.

1.7 Conclusion

In conclusion, this chapter describes the background of this project that will be developed. This project will used two types of open source spam filtering tool to detect and filter incoming spam or ham. The spam filtering tool will be implemented alongside with email server inside a Linux based operating system.

Next chapter will be discussing about the literature review and project methodology that will be used to implement this project.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

The aim of this project is to review previous research that related to this project. The most relevant methodology that will be used to develop this project also will be identified in this chapter. Several researches are being conducted to gain more understanding and ideas in order to complete this project.

The methodology that will be used through completion of this project is Top-Down approach.

2.2 Literature Review

According to the objectives and scopes of this project, this literature review will give a summary about previous research about problem with spam emails and how to handle the problem. This will give reader what knowledge and ideas have been done that related to this project and also state the strengths and weaknesses of the project.