ANTI-SPAMMING USING HONEYPOT TECHNIQUE AND JAVA APACHE MAIL ENTERPRISE SERVER (JAMES)

NURULAZLINA BT HJ MAZLAN

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

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
KOLEJ UNIVERSITI TEKNIKAL KEBANGSAAN MALAYSIA
2006

BORANG PENGESAHAN STATUS TESIS

JUDUL: ANTI-SPAMMING USING HONEYPOT TECHNIQUE AND JAVA APACHE MAIL ENTERPRISE SERVER (JAMES)

SESI PENGAJIAN: 2005/2006

Saya NURULAZLINA BINTI HJ MAZLAN
(HURUF BESAR)

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

1. Tesis adalah hakmilik Kolej Universiti Teknikal Kebangsaan Malaysia.
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 : No. 14, Prsn. Tmn. Meru 10
Tmn. Meru 2C, 30020 Ipoh Perak.

Tarikh : 4 APRIL 2006

(TANDATANGAN PENYELIA)

Pn. Aslinda Hassan
Nama Penyelia

Tarikh : 4 APRIL 2006

CATATAN: ** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.
^ Tesis dimaksudkan sebagai Laporan Projek Sarjana Muda (PSM)

DECLARATION

I hereby declare that this project report entitled

ANTI-SPAMMING USING HONEYPOT TECHNIQUE AND JAVA APACHE MAIL ENTERPRISE SERVER (JAMES)

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

STUDENT : _____ Date : 5/5/2006

(NURULAZLINA BINTI HJ MAZLAN)

SUPERVISOR : _____ Date : 5/5/2006

(PN. ASLINDA BINTI HASSAN)

DEDICATION

Especially dedicated to my beloved parents, Mr. Hj Mazlan Bin Mat Duni and Mrs. Hjh Khalijah Binti Husin for invoke and giving impulse everyday without bored.

For my supervisor, Mrs Aslinda Binti Hassan at Kolej Universiti Teknikal Kebangsaan Malaysia (KUTKM) for her criticizes advice and replying my endless queries.

And lastly to my entire friend who have encouraged, guided and inspired me throughout my journey of education.

ACKNOWLEDGEMENTS

This project would not have been possible without the help of so many people. I would like to take this opportunity to thank them for their effort in completing the PSM. Firstly is a God who never fails me.

Firstly, my deepest appreciation goes to my supervisor, Puan Aslinda Hassan for her extraordinary level of support and patience. I will always remember those countless hours her spent with me in meeting and presentations. Her expertise and advice have been a great motivation for me and have had a key role in making this a unique and rewarding Degree Project.

Second, I really want to thanks the KUTKM's involves professor, lecturer and staff. Thank you to all the information that given and whatever valuable tips and hints that helping me to develop my project application.

Then, I would like to thank to all my friends because concern about me and giving support for everything which is related to my project structure.

Finally, I'm very thankful to the entire involving person that was achieving my project instead. The knowledge was very important and very expedient in the future. I also never forget everything that I've learned and practice. Thanks to the entire involving person, I love to develop my project indeed.

ABSTRACT

There are different tools and technologies available to prevent spam attack on email application are used. This project uses honeypot concept with Bayesian filtering technique to prevent spam email from being downloaded into mail client. A honeypot is a computer system that appears to be an interesting target to hacker, but actually gathers information about them. It designed to be probed, attacked and compromised and at the same time, monitors the actions taken to complete this task. This project describes about the development and configuration of spam filtering technique using the Java Apache Mail Enterprise Server (JAMES) developed by The Apache Software Foundation. The development process includes the creation of a spam filtering by using Bayesian technique (honeypot concept). The configuration process include the setup of an Apache James mail server to process incoming mail for spam properties before delivering the mail to an internal server.

ABSTRAK

Terdapat pelbagai fungsi dan teknologi yang tersedia untuk menghentikan serangan *spam* ke atas aplikasi email yang digunakan. Projek ini menggunakan konsep honeypot dengan teknik penapis *Bayesian*. *Honeypot* adalah sistem komputer yang mengeluarkan satu lokasi untuk penggodam tetapi sebenarnya adalah untuk mengambil maklumat tentang penggodam. Ia direka untuk mencari, menyerang dan berkompromi dan pada masa yang sama mengawal aktiviti tugasan supaya berjalan lancar. Projek ini menerangkan tentang pembangunan dan pembinaan teknik penapis *spam* menggunakan *Java Apache Mail Enterprise Server (JAMES)* yang di bina oleh *The Apache Software Foundation*. Pembangunan menerangkan tentang pembuatan penapis *spam* dengan menggunakan teknik *Bayesian* (konsep *honeypot*). Pembinaan pula menerangkan tentang pembentukan Apache James server mail bagi memproses email yang datang dalam bentuk spam sebelum email tersebut sampai ke destinasi sebenar (inbox).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

JAMES	Java Apache Mail Enterprise Server
SMTP	Simple Mail Transfer Protocol
POP3	Post Office Protocol 3
API	Application Program Interface
DNS	Domain Name Server
IP	Internet Protocol
HTTP	Hypertext Transfer Protocol
URL	Uniform Resource Locator
ICT	Information and Communication Technology
SMI	Small Medium Industry
NNTP	Network News Transfer Protocol
UDP	User Datagram Protocol
TCP	Transmission Control Protocol
OS	Operating System
SQL	Structured Query Language
PSM	Projek Sarjana Muda
PCs	Personal Computers
PII	Personally Identifiable Information
DFD	Data Flow Diagram
JDK	Java Development Kit
UTP	Unshielded Twisted Pair
DMZ	Demilitarize Zone
ERD	Entity Relationship Diagram
DFD	Data Flow Diagram
JVM	Java Virtual Machine

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

1.1 Project Background

This project application defines an email application as a way for user to send or receive e-mail from anywhere where there is an Internet connection. In other word, it's become an enterprise solution that eliminates spam threats at the Internet gateway or mail server. The activities of a small number of people such as sending spam are becoming a bigger problem for the Internet. Many developers have been actively engaged in fighting spam for years and help fight spam to keep the Internet useful for everyone. Also take advantage of the information they gathered to make experience on the Internet better.

This project application runs in the Java Apache Mail Enterprise Server (Apache JAMES) which is a 100% pure Java SMTP and POP3 Mail server. In addition, the Apache JAMES is also a mail application platform and it's developed using a Java API which is a Java code to process emails that call the mailet API. A mailet can generate an automatic reply, update a database, prevent spam and build a message archive. Apache JAMES accept mail messages meeting one of two criteria:

1. Messages sent by a user of the mail-domain the server is responsible for
2. Messages addressed to a user in that domain.

Domain Name System (DNS) was common for mail servers to accept mail from anyone, for anyone. The server would then make best efforts to relay the mail to the mail-server of the appropriate mail-domain. [1]

Honeypot concept is a system left open and unprotected to entice hackers to break into it. Usually this is done so that system administrators can monitor the methods used to break in, the frequency of attack, or just to throw off attackers from the real goodies [2].

Indeed, this application defining spammers routinely abuse open relays that find, by pumping huge amounts of spam through this application, often crashing this application in the process. By addressing each message to a large number of users, this application can send a relatively small number of messages to the relay, and reach a large number of in-boxes.

There are three parts in this application will figure it out:

1. Selective relaying – this application attempts to identify relay test messages, and relays only those messages to the destination on the envelope. Other messages are considered to be spam, and are not relayed. Instead, they are filed for reference.
2. Mail server – this application saves full details of all spam mail submitted to it as a collection.
3. Configurable – A lot of the behavior of this application is configurable, user can:
    a. control which ports and IP addresses that mail application serves on.
    b. switch on and off the SMTP and POP3 services independently.

1.2 Problem Statement

Spam emails, which also known a unsolicited bulk emails are large collection of messages where the recipient was not granted verifiable permission for the messages to be sent. It is a big problem, if every user has to spend between 20-40 minutes a day 'peeking' into spam or likely spam emails. The worse fear is huge, email filling advertises of full blown incompatible scenes and should be contain a lot of viruses that makes user computer very slow and uncomfortable.

1.3 Objectives

Achievement on the objective is:

1. To send, compose and receive email while using James Mail Server as the default mail server.
2. To identify relay test messages and relays only those messages to the destination on the envelope.
3. To saves full details of all spam mail to be submitted to this application. Therefore, the application can control all incoming and outgoing spam mail in an every time when mail arrived.
4. To filter email arrives into James Mail Server using Bayesian filtering technique based on honeypot concepts.

1.4 Scopes

Scope of this project module is:

1. This application will relay mail to any email-addresses that it considers it must to be in inbox.

2. This application will control spam in every second when mail arrived using Bayesian filtering technique based on honeypot concepts.
3. Running up the apache JAMES as a mail server and locate the application that will be integrate by honeypot concept inside the server.
4. The filter is based on the domain name server and email addresses.

1.5 Project Significance

For all above, this application is for users in all over the world that use an email for their needs. In addition, this application can be use in ICT industry as well as SMI Companies. This application is a useful for system administration and education also for the end users where most of them use email to reach or send important messages. With this application they able to minimize spam threat.

1.6 Conclusion

This project application will be implemented an organization that needs anti-spamming abilities. Using Bayesian filtering technique based on honeypot concepts as guidance and running in Apache JAMES can be considered as a creative solution for this project although some mail servers have been implemented and working properly before.

Next chapter which is on literature review defines of a collection of published researches that are relevant to the problem statement. All good research and writing is guided by a review of the relevant literature. Regardless of the research methodology used, the purpose of the literature review remains the same. It is an essential test of the research question against that which is already known about the subject.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

Literature review is the review of a collection of published research relevant to the research question. All good research and writing is guided by a review of the relevant literature. An integral component of the scientific process, a literature review is the mechanism by which research is viewed as a cumulative process. The literature review has two components which is the actual literature search and the writing of the review. Regardless of the research methodology used, the purpose of the literature review remains the same. It is an essential test of the research question against that which is already known about the subject. [3]

Project methodology is a guideline to help agencies establish a project management competency within their organizations and roadmap to implementing a successful project. In addition, process that includes phases, activities, tasks and key templates or deliverables. [4]

2.2 Fact and Finding (based on topic)

For fact and finding, there are few researches that are related to this project. The first fact describes the ways of spam explosion can be implemented and the second fact is on the Java Apache Mail Enterprise Server (a.k.a. Apache James) as a mail server. Both facts and finding will be used as reference in this project. In addition, this project will look into the definitions and the difference between Bayesian filters technique and Honeypot concept.

2.2.1 Spam Explosion

Everyone complains about spam, but the data also shows that spam is now up getting into of all email messages on the Internet. Actually spam means flooding the Internet with many copies of the same message. It attempts the force of messages that the people that use Internet did not wish to get. Most spam is commercial and it cost the recipient to pay for receiving instead of the sender.

There are two main types of spam, that which are cancelable usenet spam and usenet spam. Cancelable usenet spam means the spam messages are sent over up to 20 or more usenet newsgroup. While the usenet spam is aimed to the person that read the newsgroup but never reply or post in the newsgroup. Plus, usenet spam subverts the ability of system administrator and owner to manage the topics they accept on their systems.

Email spam targets individual users with direct mail messages. Email spam typically cost users money to receive. Many people, with measured phone service, read or receive their mail while the meter is running. Spam costs them additional money. On top of that, it costs money for ISPs and online services to transmit spam, and these costs are transmitted directly to subscribers.