

SPAM FILTERING IMPLEMENTATION USING OPEN SOURCE SOFTWARE

AHMAD BAKHTIAR BIN IBRAHIM

A PROJECT PAPER SUBMITTED

IN PARTIAL FULFILMENT OF REQUIREMENT

BACHELOR OF SCIENCE (Hons) IN DATA COMMUNICATION AND

NETWORKING

FACULTY OF INFORMATION TECHNOLOGY AND QUANTITATIVE SCIENCE

MARA UNIVERSITY OF TECHNOLOGY

SHAH ALAM

APRIL 2006

**SPAM FILTERING IMPLEMENTATION USING OPEN SOURCE
SOFTWARE**

AHMAD BAKHTIAR BIN IBRAHIM

2004219847

This project submitted to the
Faculty of Information Technology and Quantitative Science
MARA University of Technology
In Partial fulfillment of requirement for the
BACHELOR OF SCIENCE (Hons) In DATA COMMUNICATION AND
NETWORKING

Approved By the Examining Committee:

.....
Mr. Mohammad Yusof bin Darus
Project Supervisor

.....
Mr. Ali bin Isa
Examiner

MARA UNIVERSITY OF TECHNOLOGY
SHAH ALAM
APRIL 2006

ABSTRACT

Spam is a serious problem that has been increasing plaguing users of the Internet. Programs known as spam filters are employ to assist the user in deciding if an email is worth reading or not. This paper is focuses on build the mail server with spam filter and analysis the effectiveness of rule based (Heuristic) and Bayesian filtering method that use in SpamAssassin. The mail server consolidates software which is available in public domain as a powerful and low cost solution for seeking option in fighting spam. The rule based method using set of rules to define the incoming email as spam or not and Bayesian method using probability of incoming mail to define that the mail is belong to either spam or legitimate mail category after its require a training period for SpamAssassin. The combination of both methods show that SpamAssassin is a better spam checking.

TABLES OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION	
1.1 Background	1
1.2 Problem Statement	3
1.3 Objectives	4
1.4 Project Scope	4
1.5 Significances Of Study	5
1.6 Organization of This Report	6
2. LITERITURE REVIEW	
2.1 Spam Overview	7
2.1.2 Definition of Spam	7
2.1.2 Spamming In Different Media	8
2.1.2.1 Email Spam	8
2.1.2.2 Messaging Spam	9
2.1.2.3 Forum Spam	9
2.1.2.4 Online Games Messaging	10
2.1.3 Spam Filtering Methods	10
2.1.3.1 Ruled Based (Heuristic) Filtering	10
2.1.3.2 Bayesian Filtering	11
2.1.4 Mail Filtering – SpamAssassin	12
2.1.5 How Spammers Operate	14
2.2 Email Server Overview	15
2.2.1 How Does Email works	15
2.2.2 Email System	16

2.2.2.1 Simple Mail Transfer Protocol (SMTP)	16
2.2.2.2 POP3 Server	17
2.2.2.3 IMAP Server	17
2.2.3 Email Filtering	18
2.2.4 Understanding of Mail Transfer Agent (MTA), Mail User Agent (MUA) and Mail Delivery Agent (MDA)	19
2.2.4.1 Mail Transfer Agent	19
2.2.4.2 Mail Delivery Agent	19
2.2.4.3 Mail User Agent	20

3. METHODOLOGY

3.1 Introduction	21
3.2 Hardware And Software Requirement	22
3.2.1 Minimum Hardware Requirement	22
3.2.2 Software Requirement	22
3.3 Process Phase	26
3.3.1 Information Gathering	26
3.3.2 Planning	26
3.3.3 Installation and configuration	27
3.3.3.1 SuSE Linux 10.0	27
3.3.3.2 Postfix – Mail Transfer Agent (MTA)	28
3.3.3.3 MailScanner	29
3.3.3.4 MailWatch	30
3.3.3.5 SpamAssassin	31
3.3.3.6 Bayesian Database	31
3.3.4 Testing	32
3.3.5 Analysis	33