International Journal of Innovative
Computing, Information and Control
Volume **5**, Number **11(A)**, November **2009**

ICIC International ©2009 ISSN 1349-4198
pp. 3717–3731

AN EFFICIENT SPAM FILTERING METHOD BY ANALYZING E-MAIL'S HEADER SESSION ONLY

Jyh-Jian Sheu[1] and Ko-Tsung Chu[2]

[1]Department of Information Management
National Dong Hwa University
1, Sec 2, Dahsueh Road, Shoufeng, Hualien 97401, Taiwan
jjsheu@mail.ndhu.edu.tw

[2]Department of Finance
Minghsin University of Science and Technology
1, Hsin Hsin Road, Hsin Feng, Hsinchu 30401, Taiwan
ktc1009@must.edu.tw

Received April 2008; revised September 2008

Abstract. *This research manages in-depth analysis on Chinese spam and expects to find more efficient spam judgment rules in the condition of reducing system burden. Different from many spam filtering methods at present which have to investigate the complete content of e-mails, this research doesn't check the content of e-mails to avoid the complexity and executive efficiency. We focus on header's basic attributes such as e-mails titles, senders' names, senders' e-mail address, sending date and apply decision tree data mining technique to analyze the association rules of Chinese spam and propose a systematic method with reversing mechanism to accurately identify spam and legitimate mails. According to the experiment, the accuracy of our spam filtering method proposed in this paper was 96.17% and the precision of our method was up to 98% which was not lower than that of other present filtering methods of checking e-mail content. Thus, the method of this research could efficiently identify the spam e-mails by only checking the header sections. This advantage of this method could reduce the cost for calculation.*
Keywords: Data mining, Decision tree, Spam filtering, Security

1. **Introduction.** As Internet advances, e-mail has become the critical channel of modern people's communication and turned into the principal tool of mass marketing. Because of its convenience and speediness, many firm or website managers use it as marketing tool. For instance, e-mail marketing is used to replace mail-order catalogues or magazines. It can even replace part of customer services and increase customer satisfaction. However, these excess advertising mails caused general e-mail users' trouble. On the other hand, from the views of firms and governments, plenty of spams resulted in the serious burden of e-mail server. Moreover, time wasted when the employees screened and deleted the spam was also the significant cost loss.

Excessive application of e-mail has brought the problem of spams. In 1978, there were about 400 people receiving the first spam on ARPANET; until now, 80% of the mails in the mailboxes of 1/3 e-mails users were spams. In 2003, spam became 50% of e-mails in the world. According to the latest report of spam reorganization proposed by Symantec, the international firm of information security [20], by the end of May, 2007, spam has been 75% of e-mails in the world. Mail boxes are filled with plenty of spam which has become the annoying problem of many people. Spam filtering method thus becomes an important research subject. Therefore, spam prevention becomes the critical issue of Internet application.