

Do Web Interfaces to Email Services Support Users in Coping with Spam Filter Brittleness?

Christopher Lueg

School of Information Technology,
Charles Darwin University
lueg@it.uts.edu.au

Michael Twidale

Grad. School of Library and Information Science,
University of Illinois at Urbana-Champaign,
twidale@uiuc.edu

ABSTRACT

For a number of reasons, spam filters cannot be expected to be totally reliable. Regardless, most research on spam filtering focuses on finding "better" filtering criteria to increase filter effectiveness and reduce the number of false-positives. A more user-oriented approach to addressing spam filter problems would focus on helping users understand how their spam filters actually process email and why they may classify genuine email as spam (and what could be done about it). In this work we present results from an investigation of how popular web-based email services support users in dealing with spam filters.

Author Keywords

email, spam, classification, information filtering

ACM Classification Keywords

H.5 INFORMATION INTERFACES AND PRESENTATION (e.g. HCI) H.5.2 User Interfaces H.5.3 Group and Organization

SPAM FILTER BRITTLINESS

According to technical criteria typically employed in the spam filtering community, anti-spam technologies are performing extremely well in terms of effectiveness and the number of false-positives. The way spam filters are presented to users as service features typically suggests that spam filtering is a straight-forward business. Lueg (2004) suggests however that there is a significant conceptual gap between what has been defined as "spam" in discourse and (technical) definitions of spam as operationalized by spam filtering software. Email users also seem to perceive the situation differently. Fallows (2003) found that 30% of email users surveyed were concerned their email filters might filter genuine incoming email and 23% of users were concerned email they send to others may be filtered.

RESEARCH QUESTIONS

A user-oriented approach to addressing spam filter problems would focus on helping users understand how their spam filters actually process email and why they may fail when determining "spamminess" (and what can be done about it). In this research, we looked at information provided by four popular web-based email services in order to help users understand what the built-in spam filters are doing and why. Issues investigated include, for example, the question whether users are informed about how a

particular service determines spamminess (important because of the above mentioned conceptual gap). We were also curious about the means provided to assist users in dealing with spam filtering outcomes, in particular if genuine email had falsely been classified as spam. Another area of investigation was whether users are actually able and also encouraged to make informed decisions regarding the filtering of their email. The webmail services we looked at were the bigpond.com.au, hotmail.com, yahoo.com and gmx.net. We accessed all services using computers connected to Australian ISPs.

FINDINGS AND DISCUSSION

All four webmail services acknowledge the problem of spam filter brittleness in some way. However, none of them provides a technical definition of what they consider spam in the sense that the definition could be used to verify filtering outcomes. It is of concern that users with two of the email providers (bigpond.com.au, hotmail.com) may be excluded from receiving emails suspected to be spam as respective services state they may discard these emails instead of delivering them eg marked as spam. None of the services explicitly addresses conceptual issues such as the fact that bulk email may be solicited or that unsolicited email may in fact be appreciated; the only solution offered requires users to add senders of bulk email to "white lists".

We conclude that interfaces provided by web email services are far from being perfect. Even *if* configuration options are offered, sound technical knowledge may be required to understand the impact of selecting or de-selecting options. There is little support offered to improve understanding of how spam filters actually work, why they can fail when determining spamminess, and how this could be improved.

ACKNOWLEDGMENTS

CL thanks Computer Sciences Corp. (CSC) for supporting his position at CDU and Nicola Bidwell for feedback.

REFERENCES

1. Fallows, D. (2003). Spam. How it is Hurting Email and Degrading Life on the Internet. Pew Internet & American Life project. <http://www.pewinternet.org>.
2. Lueg, C. (2004). The hidden impacts of anti-spam measures and their contributions to the digital divide: an exploratory study. *Proc. ASIS&T AM*, pp. 176-183.