Crossroads
The ACM Student Magazine

Introduction

by ***William Stevenson***

Welcome to the first ever issue of *Crossroads* that is completely dedicated to exploring the plague of our inboxes, "Spam." Spam, or unsolicited bulk email, has over the past few years evolved from an infrequent interruption to a nuisance to what is now considered to be one of the biggest threats to the future of the Internet. Many informal estimates suggest that more than half of all email sent is junk, ranging from (relatively) targeted mailings all the way to scams that are intended to defraud the receiver out of money, or worse, their identity.

Many attempts at solutions, ranging from technical to legislative, are presently underway to combat the problem of junk email. However, the laws of individual countries do not necessarily have reach to other countries in the world where Spam often comes from, and many of these countries are not in a financial or technical position to deploy anti-spam technologies. In addition, many junk emailers are aware of the illegality or at least the ill-repute of their actions, and try to cover their tracks by compromising insecure machines that are connected to the internet. For these reasons, the current paradigm on anti-spam technology has tended to have the present solutions (more often than not content-based filtering solutions) deployed on the receiving email server and on the client side. This issue of *Crossroads* describes several such solutions.

In our first article, "Learning How to Tell Ham from Spam," George Sakkis describes learning-based spam filters and discusses their advantages over the early and traditional techniques such as black/white lists and keyword filtering. One of the resounding successes of learning-based filters over the earlier techniques is that they are substantially easier to use from the end-user's perspective (I would not want to have to give a primer on regular expressions to my grandparents). Sakkis describes how Bayes' theorem is used to determine the likelihood that individual messages, represented using a vector-space model, are junk, and how filters that use Bayes learn to differentiate "spam from ham" for the individual user.

Shlomo Hershkop and Salvatore Stolfo follow up on Sakkis by describing behavior-based spam filtering in their article "Identifying Spam Without Peeking at the Contents." In the article, the authors discuss how non-content features of messages, such as the domain name, number of recipients, size, mime type, and number of "re:s" in the subject header can aid in classifying junk email. Although it may seem that their model only increases the accuracy of the spam detection system by a few percent, I am confident that all of us would gladly enjoy any prospect of getting less junk mail, and the work described in the article suggests more places of focus to push accuracy of spam classification even higher.

Our third article, "Peer-to-Peer Collaborative Spam Detection" by Nathan Dimmock and Ian Maddison describe a student project that has evolved a centralized collaborative filtration system into a fully distributed peer-to-peer spam detection system. The authors set out with the objectives that their system be efficient, scalable, secure, private, and able to interface with numerous mail clients. An interesting feature of the system, a built-in trust metric used to make sure that nodes that deliberately feed bad information gain a bad reputation and are ignored, is described in great detail, and seems to be a technology that will be of increasing importance in future systems.

Finally, we turn out attention from spam to the affect that the Internet has had on our rights in "Security, Privacy, and Anonymity" by Thomas Wright. Wright describes these rights, that we often don't consider, and discusses how certain technologies conspire to identify and track Internet users for financial gain. In addition, the author provides suggestions on how Internet users can protect their rights, through secure communications and proxies that anonymize browser activities. It is worth thinking about our basic rights and how to preserve them, as many people tend to naively trust that their communications and activities are private due to society's general faith in technology and science.

Please stop by our web site, **http://www.acm.org/crossroads**, to see an additional online-only feature article "Spam Detection using an Artificial Immune System" by Terri Oda and Tony White. Their article describes the unique application of an artificial immune system model for effectively protecting users from unwanted messages. This system goes head to head with filters like SpamAssassin to classify messages with similar accuracy, but does so with fewer detectors, and is certainly an interesting multidisciplinary application. It's free, act now! :)

Biography

William Stevenson (**billstevenson@acm.org**) is a Ph.D. student in the School of Information Sciences and Technology at the Pennsylvania State University, from which he holds a Master's degree in Computer Science. His main research interests are in Cognitive Science and High Performance Scientific Computing. In his spare time, Bill enjoys cooking, the outdoors, and working on his Mac. He has served as Editor in Chief of *ACM Crossroads* since July 2001.