Enhanced Spam Filter with Hoax Detection -
Project Proposal

Abhishek Aggarwal 2005MCS2942
Prem Piyush Goyal 2006MCS2262
January 21, 2007

“Don’t add Mr. ABC to ur messenger contacts - ITS A VIRUS”, “My friend contracted AIDS froma needle in the theatre seat”, Rapedrug hoax - “The victim can NEVER conceive again”, “Don’t answer a call from 9837429083 its a VIRUS”, so on and so forth..

While hoaxes are recognized as a real and increasing threat with real and increasing costs, the industry has always focused its resources on detecting malicious software, as opposed to malicious text. The torrent of spam that is threatening to make e-mail unattractive enough to discard as a communications medium has had an unexpected and fortunate (for antivirus) side-effect. Recent developments in spam filtering research have led to the development of tools that could be used to detect and filter hoaxes, hence challenging the more traditional methods of addressing hoaxes via awareness and education campaigns. These tools and methods, particularly those based on Bayesian filtering show promise and are worthy of serious consideration. It is conceivable, that using these techniques, implementations could be developed that are specifically tailored to a new approach of dealing with the threat of hoaxes.

The most important difference between spam and hoax is that we receive hoaxes from our own friends and others we communicate with. Hence detecting hoaxes is tougher than detecting usual spam. We therefore propose to enhance the existing STATE-OF-THE-ART SPAM-FILTERS to enable them to handle these issues as well. Our implementation will not only detect the spams as usual, but will also adaptively learn to identify the hoax mails even if the user is unable to do so.

**Resources Required:** 2GB HardDisk Space: For spam-filter installation, storing training and test data sets, and working files storage. 2 good performance Computers (preferably dual core P4 with 1GB RAM) Data Set: Archive of 100000 solicited (including hoaxes) and un-solicited email (with complete mail headers and body).

The data set can be obtained from the iitd mail server itself.