

# Filtering spam: Researchers propose new method to rid inboxes of unwanted email

November 20 2012

---

Spammers have recently turned high-tech, using layers of images to fool automatic filters. Thanks to some sophisticated new cyber-sleuthing, researchers at Concordia University's Institute of Information Systems Engineering are working toward a cure.

Once upon a time, Spam came in a can and could be easily avoided. Nowadays, spam plagues email inboxes around the world, hawking miracle pills and enticing the gullible with tales of offshore bank accounts containing untold fortunes.

These once text-based email infiltrators have recently turned high-tech, using layers of images to fool automatic filters. Thanks to some sophisticated new cyber-sleuthing, researchers at Concordia University's Institute of Information Systems Engineering are working toward a cure.

PhD candidate Ola Amayri and thesis supervisor, Nizar Bouguila, have conducted a comprehensive study of several [spam filters](#) in the process of developing a new and efficient one. They have now proposed a new statistical framework for spam filtering that quickly and efficiently blocks unwanted messages.

"The majority of previous research has focused on the textual content of spam emails, ignoring visual content found in [multimedia content](#), such as images. By considering patterns from text and images simultaneously, we've been able to propose a new method for filtering out spam," says Amayri, who recently published her findings online in a series of

international conferences and peer-reviewed journals.

Amayri explains that new [spam messages](#) often employ sophisticated tricks, such as deliberately obscuring text, obfuscating words with symbols, and using batches of the same images with different backgrounds and colours that might contain random text from the web. However, until now, the majority of research in the domain of email spam filtering has focused on the automatic extraction and analysis of the textual content of spam emails and has ignored the rich nature of image-based content. When these tricks are used in combination, traditional spam filters are powerless to stop the messages, because they normally focus on either text or images but rarely both.

So how do we stop spam before it sullies our inboxes? "Our new method for spam filtering is able to adapt to the dynamic nature of spam emails and accurately handle spammers' tricks by carefully identifying informative patterns, which are automatically extracted from both text and images content of spam emails," says Amayri.

By conducting extensive experiments on traditional spam filtering methods that were general and limited to patterns found in texts or images, she has developed a much stronger way, based on techniques used in pattern recognition and data mining, to filter out unwanted emails. Although the new method has been tested on English spam emails, Amayri says it can be easily extended to other languages.

While this new spam-detecting approach is still in the development stage, Amayri and Bouguila are currently working on a plug-in for SpamAssassin, the world's most widely used open-source spam filter. Amayri hopes that this plug-in will allow other researchers to perform further tests and make more progress in the field of spam detection.

"Spammers keep adapting their methods so that they can trick the spam

filters, says Amayri. "Researchers in this field need to band together to keep adapting our methods too, so that we can keep [spam](#) out and focus on those messages that are really important."

Provided by Concordia University

Citation: Filtering spam: Researchers propose new method to rid inboxes of unwanted email (2012, November 20) retrieved 19 April 2024 from <https://phys.org/news/2012-11-filtering-spam-method-inboxes-unwanted.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--