Spam Challenge 2008: IBM ISS Spam Filtering Technology

C. Hagemann, D. Harz, R. Iffert, M. Usher
IBM Kassel Lab, Germany
{hagemann, dirk.harz, ralf.iffert, mark.usher}@de.ibm.com

ABSTRACT
The IBM Kassel Lab took part in the CEAS Spam Challenge 2008 using two configurations of the IBM Proventia Network Mail Security System [1]. In this paper we describe in general terms the spam filtering technology used in this product, and discuss its performance in the spam challenge.

1. INTRODUCTION
The IBM Proventia Network Mail Security System [1] is a complete email security appliance providing preemptive protection and spam control for the enterprise messaging infrastructure. It is easily adaptable to the needs of specific enterprises, containing customizable analysis modules to enforce inbound and outbound content filtering. The spam filter can be used "out of the box", as it requires no user training of the classifiers. All classifier training is performed back-end in the IBM Kassel Lab, and made available via content updates.

2. CLASSIFICATION METHODS
The classification methods employed can be placed into three broad categories. These are database methods, non-database analysis and IP-based blocking.

2.1 Database Methods
The database classifiers utilize a large database of spam fingerprints and spam URLs. The database is fed by thousands of spam traps created by the IBM Kassel Lab, as well as customer feedback. Text, email structure and image fingerprints are extracted from the spam emails received by the IBM Kassel Lab and added to the database. All fingerprints employ fuzzy extraction methods to generalize the email content, including a patent-pending image fingerprinting method. URLs are also extracted from the emails and verified for spam content using sophisticated back-end procedures. The database used for the analysis is hosted on the client machine, and updated frequently from the IBM Kassel Lab servers. During live spam analysis, fingerprints are extracted from the emails and checked against the local database for spam content.

2.2 Non-Database Analysis
These methods analyze the email content without reference to the fingerprint / URL database. These procedures include Bayesian analysis using a pre-trained classifier, text and pattern filtering and heuristic analysis. These filters are regularly updated based on the emails received by the IBM Kassel Lab. A similarity flow analysis classifier is also utilized, which is automatically trained on the live email stream.

2.3 IP-Based Blocking
The system contains methods for IP-Reputation and DNSBL. The IP-Reputation module is trained locally on the live email stream. The DNSBL module can be configured for various black lists, including our DNSBL server, trained using the IBM Kassel Lab spam traps.

3. FILTER CONFIGURATION
For the competition we set up two different configurations of the system. The first system was configured to produce as few false positives as possible. The second system used less restrictive settings. Due to technical reasons relating to the competition environment, we were not able to utilize the IP-based blocking methods.

4. RESULTS
The first system (IBM Kassel Lab 1) achieved second place in the LAM competition with a LAM score of 0.00314, corresponding to a false negative rate of 0.05650 and a near perfect false positive rate of 0.00015. This was by far the best false positive rate out of all systems which took part in the competition.

The second system achieved a high place in the LAM competition rankings with a LAM score of 0.00454, corresponding to a false negative rate of 0.02248 and a false positive rate of 0.00088.

4.1 IBM Kassel Lab 1
This system produced only four false positives on the entire ham set. The four emails were identical, containing a quarantine report of another spam filter. They were blocked on the basis of the large number of spam subjects contained within.

4.2 IBM Kassel Lab 2
This system produced 24 false positives. These included eight newsletters, and four mailing list emails discussing spam.

5. CONCLUSION
Our filtering policy is based on real world customer experience, where false positives rank far higher than false negatives. In real scenarios, only a false positive rate of around 0.0001 would be regarded as acceptable. In this respect we are satisfied with the performance of our filter. The false negative rate would be improved by the utilization of our IP-based procedures, and customer-specific rule based filtering allowed by the policy system.

6. REFERENCES
[1] http://www.ibm.com/services/us/index.wss/offering/iss/a1027071

CEAS 2008 – Fifth Conference on Email and Anti-Spam, August 21-22, 2008, Mountain View, California USA