AntiPhish – Lessons Learnt

André Bergholz
Fraunhofer Institute Intelligent Analysis
and Information Systems (IAIS)
Schloss Birlinghoven
St. Augustin, Germany
andre.bergholz@iais.fraunhofer.de

ABSTRACT

Phishing emails usually contain a message from a credible looking source requesting a user to click a link to a website where she/he is asked to enter a password or other confidential information. Most phishing emails aim at withdrawing money from financial institutions or getting access to private information. Phishing has increased enormously over the last years and is a serious threat to global security and economy. There are a number of possible countermeasures to phishing. These range from communication-oriented approaches like authentication protocols over blacklisting to content-based filtering approaches [3].

We argue that the first two approaches are currently not broadly implemented or exhibit deficits. Therefore content-based phishing filters are necessary and widely used to increase communication security. A number of features are extracted capturing the content and structural properties of the email. Subsequently a statistical classifier is trained using these features on a training set of emails labeled as ham (legitimate), spam or phishing. This classifier may then be applied to an email stream to estimate the classes of new incoming emails.

AntiPhish is a specific targeted research project funded under Framework Program 6 by the European Union. It is aims at developing improved anti-phishing technologies that help to protect and secure the global email communication infrastructure. The project on the one hand developed the filter methodology in a test laboratory setting, but on the other hand implemented this technology in real world settings, to be used to filter all email traffic online in real time. In this talk we summarize our experience with phishing filtering with benchmark data and in addition with different real-life email streams.

First we describe a number of novel features that are particularly well-suited to identify phishing emails [1]. These include statistical models for the low-dimensional descriptions of email topics, sequential analysis of email text and external links, the detection of embedded logos as well as indicators for hidden salting [2]. Hidden salting is the intentional addition or distortion of content not perceivable by the reader. For empirical evaluation we have obtained a large realistic corpus of emails pre-labeled as spam, phishing, and ham (legitimate). In experiments with benchmark data our methods outperform other published approaches for classifying phishing emails.

The second part of the talk describes the application of these approaches to real-life email streams. On the one hand we investigate how we can identify new phishing emails arriving from a honeypot system. This allows to spot new types of phishing mails. Subsequently the characteristics of these new phishing emails can be used to update client-based phishing filters. A second experiment investigates the capabilities of the AntiPhish system when monitoring emails in an ISP framework. It turns out that active learning approaches are very efficient to maintain and improve filtering accuracy.

We discuss the implications of these results for the practical application of this approach in the workflow of an email provider. Finally we describe a strategy how the filters may be updated and adapted to new types of phishing.

CSI-KDD'09, June 28, 2009, Paris, France.

ACKNOWLEDGMENTS

This talk is based upon work performed within the FP6-027600 project AntiPhish (http://www.antiphishresearch.org/). The authors would like to thank the European Commission for partially funding the AntiPhish project as well as all the AntiPhish project partners for their interest, support, and collaboration in this initiative.

REFERENCES

[1] Andre Bergholz, Jan De Beer, Sebastian Glahn, Marie-Francine Moens, Gerhard Paass, Siehyun Strobel 2009. New Filtering Approaches for Phishing Email. Accepted for publication for Journal of Computer Security (JCS)

[2] Andre Bergholz, Gerhard Paass, Frank Reichartz, Siehyun Strobel, Marie-Francine Moens and Brian Witten 2008. Detecting Known and New Salting Tricks in Unwanted Emails. Fifth Conference on Email and Anti-Spam, CEAS 2008, Aug 21-22, 2008.

[3] Markus Jakobson and Steven Myers 2007. Phishing and Countermeasures - Understanding the Increasing Problem of Electronic Identity Theft. Wiley, Hoboken, New Jersey.

BIOGRAPHY

André Bergholz is a senior research engineer in the text mining group at Fraunhofer IAIS. He is interested in text and data analysis and management. Prior to joining Fraunhofer André Bergholz worked as a research engineer at Xerox Research Centre Europe in the domain of document management. André Bergholz holds a PhD and a German diploma degree, both from Humboldt-University Berlin. At the time he specialized in management of semistructured data, where he also did a postdoc at Stanford University.