[7], however it uses additional model-based features and image processing techniques.

7. Conclusion

This study presents a modified variant of a website classification technique that we have previously proposed that proved to be effective in enhancing the classification accuracy of anti-phishing email filters as well. As evaluated empirically on publicly available phishing and legitimate email data sets, the addition of the LUA technique proved to be effective in enhancing the classification performance of the evaluated email classifier with virtually all evaluated features subsets.

We believe that the positive results are due to the fact that most phishing email messages contain URLs, and enhancing website detection techniques can directly benefit the performance of anti-phishing email classifiers. One of the advantages of the implemented LUA technique is its ability to accuratly classify phishing websites by only analyzing their URLs lexically. Other website classification techniqus also process website contents (e.g. HTML) or send queries over the network.

Acknowledgment

The authors would like to thank Buhooth[5] for funding this work.

References

[1] A. Alnajim and M. Munro. An evaluation of users' anti-phishing knowledge retention. In *Information Management and Engineering, 2009. ICIME '09. International Conference on*, pages 210 –214, April 2009.

[2] Gerry Gaffney. The myth of the stupid user. `http://www.infodesign.com.au/articles/themythofthestupiduser`. Accessed March 2011.

[3] Steve Sheng, Brad Wardman, Gary Warner, Lorrie Faith Cranor, Jason Hong, and Chengshan Zhang. An empirical analysis of phishing blacklists. `http://ceas.cc/2009/papers/ceas2009-paper-32.pdf`, 2009.

[4] Ian Fette, Norman Sadeh, and Anthony Tomasic. Learning to detect phishing emails. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 649–656, New York, NY, USA, 2007. ACM.

[5] Jose Nazario. Phishing corpus. `http://monkey.org/~jose/wiki/doku.php?id=phishingcorpus`. Accessed July 2010.

[6] SpamAssassin. Public corpus. `http://spamassassin.apache.org/publiccorpus/`. Accessed January 2011.

[7] André Bergholz, Jan De Beer, Sebastian Glahn, Marie-Francine Moens, Gerhard Paaß, and Siehyun Strobel. New filtering approaches for phishing email. *J. Comput. Secur.*, 18:7–35, January 2010.

[8] F. Toolan and J. Carthy. Phishing detection using classifier ensembles. In *eCrime Researchers Summit, 2009. eCRIME '09.*, 20 2009.

[9] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Lexical url analysis for discriminating phishing and legitimate websites. In *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, CEAS '11, pages 109–115, New York, NY, USA, 2011. ACM.

[10] Sujata Garera, Niels Provos, Monica Chew, and Aviel D. Rubin. A framework for detection and measurement of phishing attacks. In *Proceedings of the 2007 ACM workshop on Recurring malcode*, WORM '07, pages 1–8, New York, NY, USA, 2007. ACM.

[11] Colin Whittaker, Brian Ryner, and Marria Nazif. Large-scale automatic classification of phishing pages. `http://research.google.com/pubs/pub35580.html`. Accessed July 2010.

[12] Lorrie Cranor, Serge Egelman, Jason Hong, and Yue Zhang. Phinding phish: An evaluation of anti-phishing toolbars. `www.cylab.cmu.edu/files/cmucylab06018.pdf`, 2006.

[13] Mahmoud Khonji, Andrew Jones, and Youssef Iraqi. A study of feature subset evaluators and feature subset searching methods for phishing classification. In *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, CEAS '11, pages 135–144, New York, NY, USA, 2011. ACM.

[14] Saeed Abu-Nimeh, Dario Nappa, Xinlei Wang, and Suku Nair. A comparison of machine learning techniques for phishing detection. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*, eCrime '07, pages 60–69, New York, NY, USA, 2007. ACM.

[15] F. Toolan and J. Carthy. Feature selection for spam and phishing detection. In *eCrime Researchers Summit (eCrime), 2010*, eCrime '10, Dallas, TX, 2010.

[16] Wilfried N. Gansterer and David Pölz. E-mail classification for phishing defense. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ECIR '09, pages 449–460, Berlin, Heidelberg, 2009. Springer-Verlag.

[17] M. Chandrasekaran, K. Narayanan, and S. Upadhyaya. Phishing email detection based on structural properties. In *NYS Cyber Security Conference*, 2006.

[18] SonicWall. Bayesian spam classification applied to phishing e-mail. In *Whitepaper*, 2008.

[19] Mahmoud Khonji, Andrew Jones, and Youssef Iraqi. A novel phishing classification based on url features. In *GCC Conference and Exhibition (GCC), 2011 IEEE*, 2011.

[20] Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2011.

[5] http://www.buhooth.ae/

Copyright © 2013, Infonomics Society