Anti-Phishing Filtering & Education at Carnegie Mellon

http://cups.cs.cmu.edu/trust

Security Education Challenges

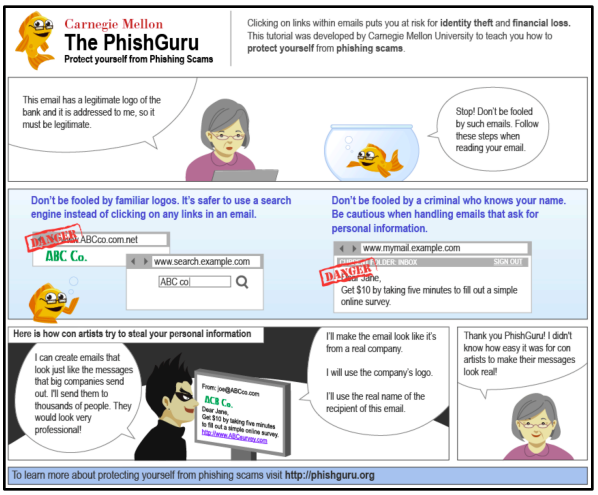
- Security is a secondary task
- Users are not motivated to learn about security
- It is difficult to teach people to identify threats without increasing their tendency to misjudge non-threats

Our Solutions

- Find teachable moments
- Engage users and make learning fun
- Provide actionable educational messages
- Use automated detection to reduce user burden

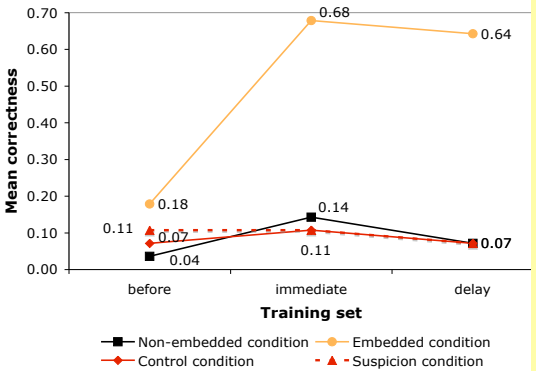
PhishGuru – Embedded Training

- Users get sent periodic training emails that look like phishing attacks
- If a user clicks a "phishing" link, they are shown succinct and engaging information on protecting themselves from phishing



User study results

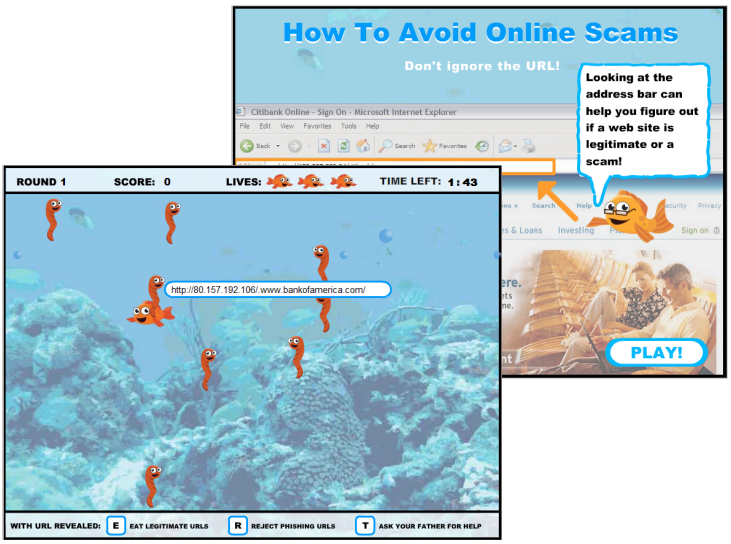
- Sending training materials through normal email is ineffective but users are motivated to learn after falling for fake phishing attack
- Users retain and transfer knowledge better when trained after falling for fake phishing attack versus getting training materials in normal email
- Real world studies confirm our lab study results



Lab study results: mean correctness before, immediate, and after one week delay. **People in the embedded training condition learned the most and retained what they learned.**

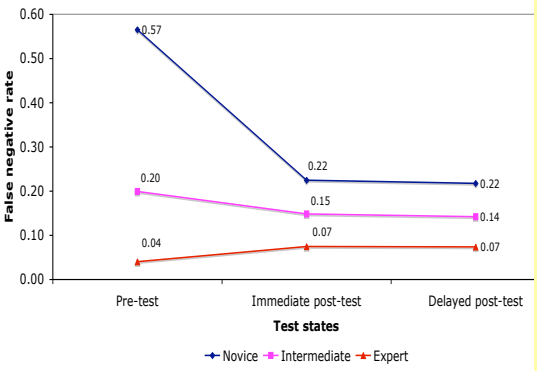
Anti-Phishing Phil – A Training Game

- A web-based interactive game to teach people how to avoid phish by paying attention to URLs
- Players move Phil around screen to examine bait and identify URLs as phishing or legitimate



User study results

- Lab study showed people made better decisions after playing our game versus reading online training materials
- 4,517 people participated in online study
- Online study demonstrated that playing our game helped people learn to make better decisions
- People retained the knowledge after one week



Online study results: false negative before, immediate, and after one week delay. **Novices that were initially poor at identifying phish are much better at identifying phish after playing our game.**

This work is funded by the US National Science Foundation under Grant No. 0524189, Fundação para a Ciência e Tecnologia (FCT) Portugal under a grant from the Information and Communications Technology Institute (ICTI) at CMU, and ARO/CyLab. Principal Investigators: Lorrie Cranor, Alessandro Acquisti, Julie Downs, Jason Hong, and Norman Sadeh

User education is important to prepare users for evolving attacks. However, automated detection can significantly reduce the burden on end users.

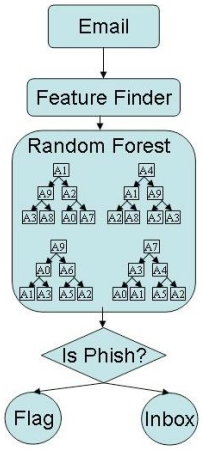
PILFER – Powerful Email Filtering

Rationale: Spam filters let a large number of phishing emails slip through

Solution: Use advanced machine learning in combination with features specifically targeting phishing emails

Implementation: Can work standalone or in combination with spam filter – e.g. available as Spam Assassin (SA) plugin

Evaluation: 90,000 emails



	False positives	False negatives
PILFER	**0.13%**	**4.79%**
SA	3.19%	7.68%

- **Catches more phish than Spam Assassin**
- **25 x fewer false positives**
- **20 x faster**

CANTINA – Powerful Web Filtering

Rationale: Blacklists are slow to update and only contain large-scale phishing attacks

Solution:

- Create fingerprint for a suspicious web page and use search engines and machine learning to identify
- No human intervention needed
- Protects against spear-phishing too

Implementation: Makes use of search engines and locally running machine learning algorithms

Evaluation:

- 90% true positive rate in correctly identifying phish
- 5% false positive rate in mis-identifying legitimate sites as phish

Selected Publications

I. Fette, N. Sadeh, and A. Tomasic. **Learning to Detect Phishing Emails.** In Proceedings of the 16th International conference on World Wide Web, Banff, Alberta, Canada, May 8-12, 2007.

P. Kumaraguru, Y. Rhee, A. Acquisti, L. Cranor, J. Hong, and E. Nunge. **Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System.** In CHI 2007: Conference on Human Factors in Computing Systems, San Jose, California, 28 April - May 3, 2007, 905-914.

P. Kumaraguru, Y. Rhee, S. Sheng, S. Hasan, A. Acquisti, L. Cranor and J. Hong. **Getting Users to Pay Attention to Anti-Phishing Education: Evaluation of Retention and Transfer.** Proceedings of the 2nd Annual eCrime Researchers Summit, October 4-5, 2007, Pittsburgh, PA, p. 70-81.

S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. Cranor, J. Hong, and E. Nunge. **Anti-Phishing Phil: The Design and Evaluation of a Game That Teaches People Not to Fall for Phish.** In Proceedings of the 2007 Symposium On Usable Privacy and Security, Pittsburgh, PA, July 18-20, 2007.

Y. Zhang, J. Hong, and L. Cranor. **CANTINA: A content-based approach to detecting phishing web sites.** In Proceedings of the 16th International conference on World Wide Web, Banff, Alberta, Canada, May 8-12, 2007.

Soon to be commercially available

For more information: **http://cups.cs.cmu.edu/trust**
or contact Lorrie Cranor <lorrie@cmu.edu>