

# Supporting Trust Decisions

Alessandro Acquisti, Lorrie Cranor, Sven Dietrich, Julie Downs, Jason Hong, Norman Sadeh  
Carnegie Mellon University

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



## SUPPORTING USERS

### Why do users fall for phish?

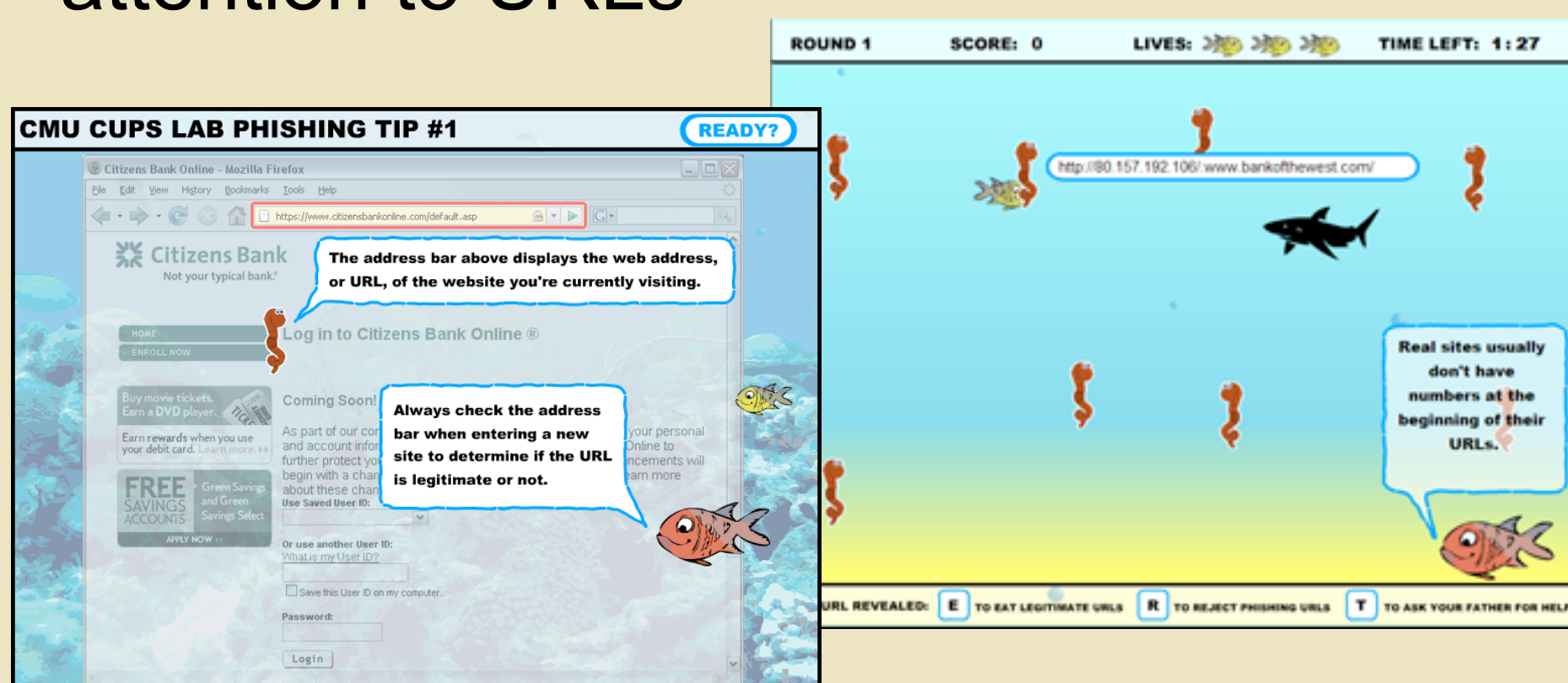
- Conducted 40 mental models interviews, included email role play
- Found limited security knowledge
- Only half knew meaning of phishing: *“Something to do with the band Phish...”*
- Few paid attention to URLs
- Knowledge of one kind of scam didn't help them avoid other kinds of scams

### Can we teach users not to get phished?

- Conducted lab study of 28 users
- After 15 minutes reading training materials, users did significantly better
- People can learn from online training - how do we get them to do it?

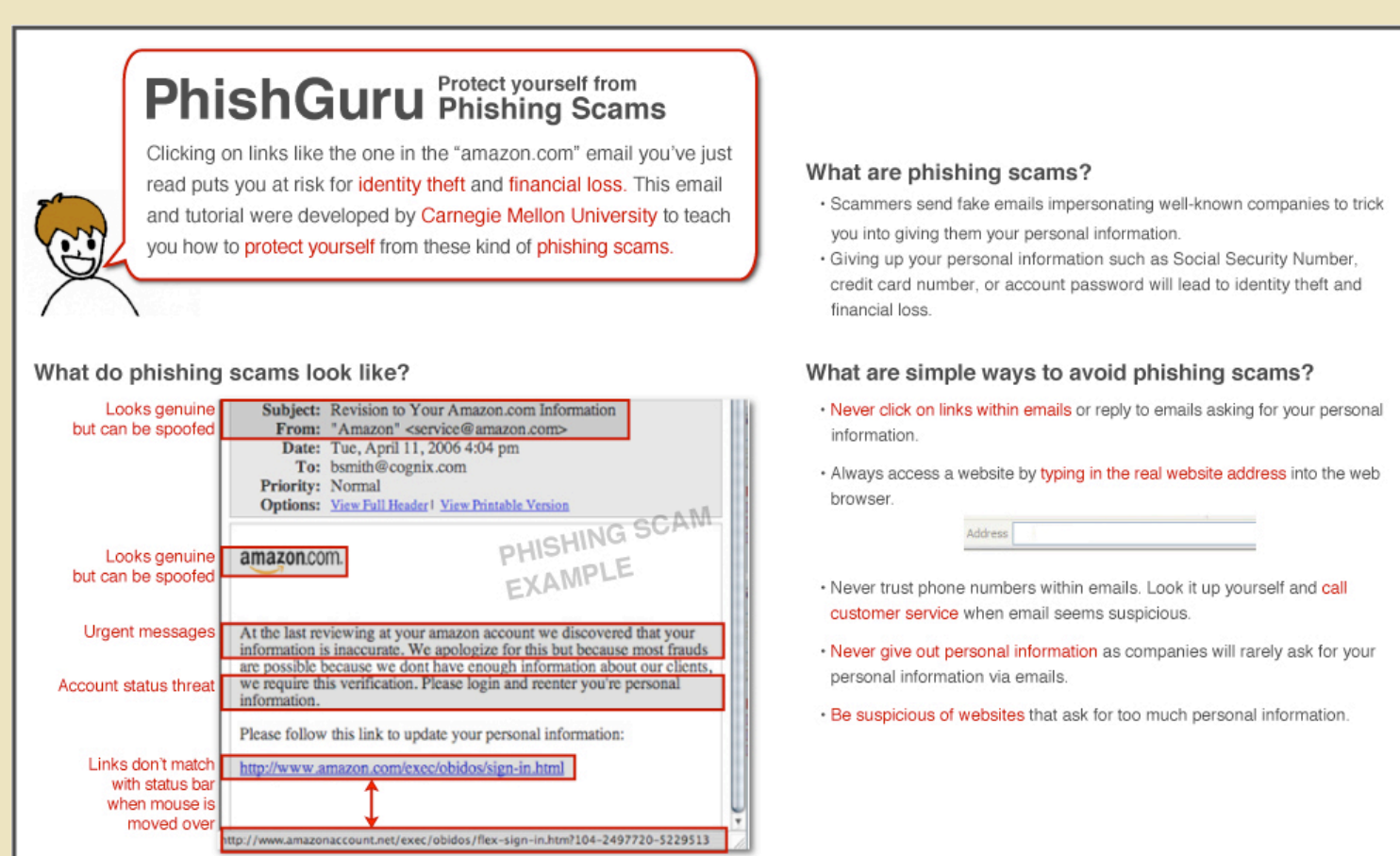
### Anti-phishing Phil

- A web-based interactive game to teach people how to avoid phish by paying attention to URLs



### PhishGuru “Embedded training”

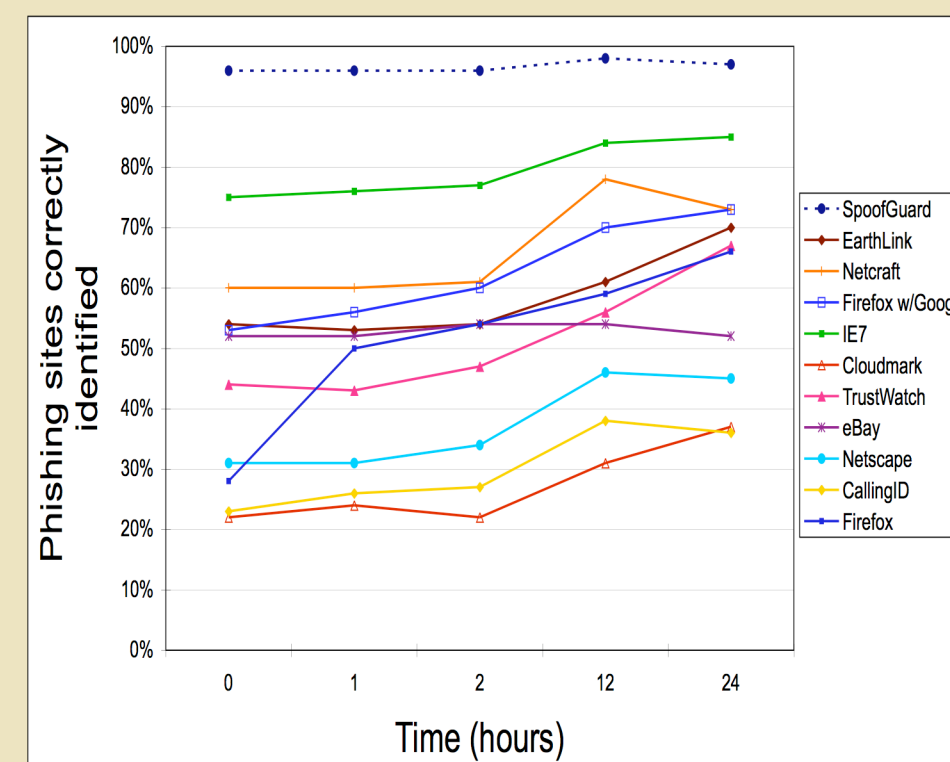
- Users get sent periodic training emails that look like phishing attacks
- If a user clicks a “phishing” link, the training intervenes and highlights cues in a succinct and engaging format
- Lab study showed this approach is better than security notice emails



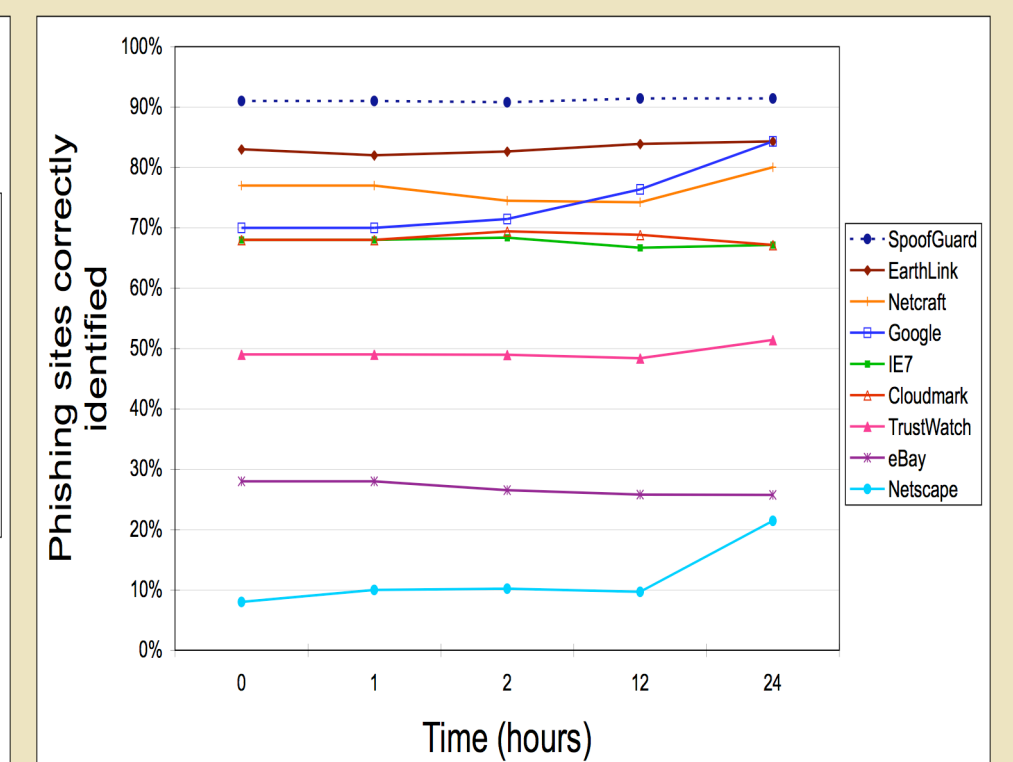
## AUTOMATING DETECTION

### Evaluating anti-phishing tools

- Developed semi-automated testbed
- Tested 10 tools on phish from 2 feeds



APWG phish feed



Phishtank.com

### CANTINA - Carnegie Mellon ANTI-phishing and Network Analysis Tool

- Apply term frequency - inverse document frequency (TF-IDF) algorithm to web page to find group of terms most important and unique to that web page
- Feed top 5 terms into Google
- Not phish if domain is in top 30 results
- 97% catch rate, 6% false positives
- Combined with other heuristics: 89% catch rate, 1% false positives



### PILFER - Phishing Identification by Learning on Features of Email Received

- Phishing classifier using advanced machine learning techniques
- Uses features indicative of phish
  - HTML present
  - JavaScript tricks to mask URLs
  - Use of non-matching HTML links
  - Links to recently registered domains
  - Number of dots in a link
  - Number of domains linked to
  - Result of a spam filter
- Random forest trained on a corpus of phishing and non-phishing emails
- 98.5% catch rate, .1% false positives