COLORADO SCHOOL OF MINES
engineering the way

COMPUTER SCIENCE DIVISION
COLORADO SCHOOL OF MINES

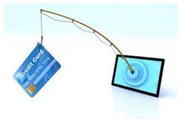
The Highly Insidious Extreme Phishing Attacks

In Proceedings of IEEE International Conference on Computer Communication and Networks (ICCCN), 2016.

Rui Zhao*, Samantha John†, Stacy Karas†, Cara Bussell†, Jennifer Roberts†, Daniel Six†, Brandon Gavett†, and Chuan Yue*
* Colorado School of Mines, † University of Colorado Colorado Springs

Introduction

- Phishing: uses spoofed websites to steal users' passwords and online identities.
- Defense:
  - Blacklist-based
  - Heuristics-based
  - Whitelist-based



- Phishing reporting and verification services:
  - APWG & PhishTank
- Phishing attacks have also been quickly evolving to evade the detection and defense.

1

Introduction – cont.

- First-layer context: a spoofed email or instant message
  - To lure users to the phishing websites

- The success is limited by two constraints
  - If phishing emails or instant messages are suspicious
    - Users would not click on phishing URLs

  - If phishing emails are captured by spam filters
    - Cannot even reach users in the first place

2

Introduction – cont.

- Second-layer context: look and feel similar to a targeted legitimate website
  - To lure users to submit their login credentials
- The success is limited by two constraints
  - If phishing websites trigger warnings if they are detected by browsers

  - If the look and feel of the undetected phishing websites are suspicious

3

Our Goal

- We explore the feasibility of the extreme of phishing attacks!
  - that have the almost identical look and feel as those of the targeted legitimate websites
- We evaluate the effectiveness of such phishing attacks by performing a user study

4

Metrics for Look and Feel

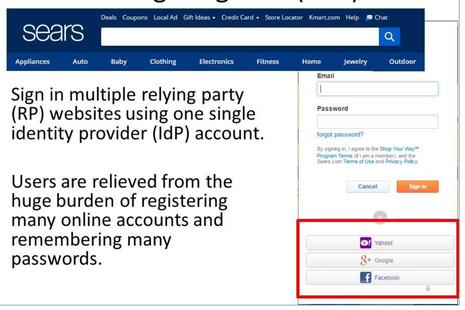
- We focus on the second-layer context

	Metrics			
	Appearance	Page Depth	Support to Dynamic User Interaction	Phishing Types
Extreme Phishing	Similar in every way	Unlimited levels of pages with completely modified links	Yes	Traditional & High-quality SSO
Advanced Phishing	Mostly similar	Limited levels of pages with partially modified links	No	Traditional & Low-quality SSO
Simple Phishing	Somewhat similar	One page with partially modified links	No	Traditional

- The *appearance*: page layouts, text contents, images, styles
- The *page depth*: the levels of webpages that are organized and linked together
- The *support to dynamic user interaction*: user interactions such as clicking, searching, and form submission as well as the triggered JavaScript executions
- The *phishing types*: traditional phishing and Web Single Sign-On (SSO) phishing

5

Web Single Sign-On (SSO) ?



- Sign in multiple relying party (RP) websites using one single identity provider (IdP) account.
- Users are relieved from the huge burden of registering many online accounts and remembering many passwords.

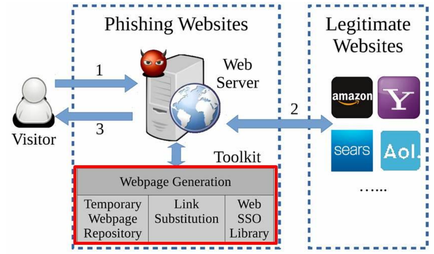
6

A Measurement of Existing Phishing Websites

- In 2015, measured and inspected 471 live phishing websites reported on PhishTank
  - 30% do not contain any link
  - 22% contain invalid links
  - 17.6% contain links to the targeted legitimate websites
  - 26.4% contain links to other websites
- The majority of them, 449 (95%) of 471 – Simple phishing
- Only a handful of them – Advanced Phishing
  - 2 Yahoo, 7 Paypal, and 11 Gmail are mostly similar
  - 2 Paypal contain over two levels of webpages
  - 10 phishing websites support low-quality Web SSO phishing
- None of them – Extreme Phishing

7

High Level Design of A Toolkit For Extreme Phishing



8

Link Substitution

- Our toolkit needs to ensure that all the links on each phishing webpage will be modified to point to the phishing website.
  - To keep holding visitors on a phishing website.
  - To maximize the chances of collecting their login credentials.
- Static Link Substitution:
  - Legitimate domain -> phishing domain & HTTPS -> HTTP & customizable rules for special links (in <head> and <script>)
- Dynamic Link Substitution:
  - Injects JavaScript to intercept the dynamic link generation and modification events & legitimate domain -> phishing domain & HTTPS -> HTTP

9

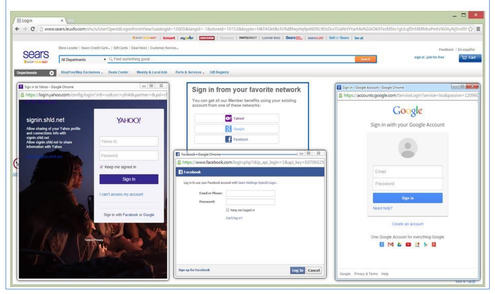
Web SSO Login Window Generation

- More profitable and insidious because
  - The value of IdP accounts is highly concentrated
  - The attack surface area is highly enlarged
  - The difficulty of phishing detection is highly increased

- We achieve
  - The automatic and dynamic construction
  - The automatic inclusion of Web SSO phishing login windows

10

The Single Sign-On login windows on the Sears phishing website (the fake Yahoo, Facebook, and Google login "windows" have the almost identical look and feel as those of legitimate login windows)



User Study

- We provided a computer for all the participants
  - Modified the *hosts* file
  - Installed and configured *five popular browsers*
- This testbed – Realistic!
  - Allows participants to use their real login credentials
  - Perform real browsing activities
- Participants - 94 adults
  - 57 younger (18-38 years), 37 older (50-88 years)
  - 62 female, 32 male
- Each participant performed 4 tasks on 4 websites
  - 2 were extreme phishing websites (traditional, SSO)
  - 2 were legitimate websites (traditional, SSO)
  - Each task – browse the corresponding website as he or she usually does, log into it, and sign out
- Data collection through behavioral observation & questionnaire

12

User Study - Summary

- 87 (92.6%) were not suspicious about extreme phishing websites
- 91 (96.8%) submitted their login credentials to extreme phishing websites
- **No significant difference** in lack of susceptibility to the entire phishing testbed between
  - Those who did and did not report noticing something suspicious about the Web browsing tasks
  - Those who with and without reported awareness of phishing
  - Those who with and without a past history of reported phishing susceptibility
- Note that the success rate of existing phishing attacks in terms of the second-layer context is about 10% as reported in previous measurement studies.

13

Discussion - Defenses

- Heuristics-based solution
  - That only uses features extracted from visual or non-visual elements on webpages may fail
  - That uses features extracted from URLs may become either inaccurate or incorrect on the detection of Web SSO extreme phishing attacks
    - Host the base webpages for Web SSO phishing attacks on legitimate websites
- Blacklist-based solution
  - Indirectly affected if the construction of their blacklists relies on heuristics-based techniques or anti-phishing communities
- Whitelist-based solution - more robust

14

Discussion - Suggestions

- We suggest that researchers should seriously consider extreme phishing in their heuristics-based phishing detection solutions.
- We suggest that Web users should be trained to
  - Be aware of extreme phishing
  - Pay more attention to the domain name of a URL
  - Differentiate the spoofed Web SSO login "windows" from real ones

15

Conclusion

- Explored the extreme phishing attacks and investigated the techniques for constructing them
- Designed and implemented a concrete toolkit
  - Traditional phishing and Web SSO phishing
  - Automatically construct unlimited levels of phishing webpages based on user interactions
- Designed and performed a user study with 94 participants
- Demonstrated that extreme phishing attacks are indeed highly effective
- Discussed the impacts of extreme phishing on existing phishing defense mechanisms
- Provided suggestions

This project is supported in part by the NSF grant CNS-1624149.

16