Fundamentals of Phishing: A Usability Perspective

Zane Ma

University of Illinois Urbana-Champaign
zanema2@illinois.edu
https://zanema.com
Phishing: Roots

1995 - AOL cracking tool
Phishing: Today

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Victims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phishing/Vishing/Smishing/Pharming</td>
<td>114,702</td>
</tr>
<tr>
<td>Non-Payment/Non-Delivery</td>
<td>61,832</td>
</tr>
<tr>
<td>Extortion</td>
<td>43,101</td>
</tr>
<tr>
<td>Personal Data Breach</td>
<td>38,218</td>
</tr>
<tr>
<td>Spoofing</td>
<td>25,789</td>
</tr>
<tr>
<td>BEC/EAC</td>
<td>23,775</td>
</tr>
<tr>
<td>Confidence Fraud/Romance</td>
<td>19,473</td>
</tr>
<tr>
<td>Identity Theft</td>
<td>16,053</td>
</tr>
<tr>
<td>Harassment/Threats of Violence</td>
<td>15,502</td>
</tr>
<tr>
<td>Overpayment</td>
<td>15,395</td>
</tr>
<tr>
<td>Advanced Fee</td>
<td>14,607</td>
</tr>
<tr>
<td>Employment</td>
<td>14,493</td>
</tr>
<tr>
<td>Credit Card Fraud</td>
<td>14,378</td>
</tr>
<tr>
<td>Government Impersonation</td>
<td>13,873</td>
</tr>
<tr>
<td>Tech Support</td>
<td>13,633</td>
</tr>
<tr>
<td>Real Estate/Rental</td>
<td>11,677</td>
</tr>
<tr>
<td>Other</td>
<td>10,842</td>
</tr>
</tbody>
</table>

Source: 2019 FBI Internet Crime Report

Watch Out for Coronavirus Phishing Scams

Ransomware
Trojans
Other malware
Why haven’t we solved/curtailed phishing, twenty-five years later?
Phishing: Root Causes

Mistaken Identity

accounts-google.com

Misplaced Trust

questionsaboutisp.com
Phishing: Root Causes

Mistaken Identity

URL complexity leads to mistaken identity

Misplaced Trust

Users may (mis)place trust in HTTPS

Measuring Identity Confusion with Uniform Resource Locators

Joshua Reynolds¹ Deepak Kumar¹ Zane Ma¹ Rohan Subramanian¹ Meishan Wu¹ Martin Shelton² Joshua Mason¹ Emily Stark¹ Michael Bailey¹

¹University of Illinois at Urbana-Champaign ²Google, Inc.
joshua3, dkumar11, zanema2, resubra2, meishan2, joshm, mdbailey}@illinois.edu

The Impact of Secure Transport Protocols on Phishing Efficacy

Zane Ma Joshua Reynolds Joseph Dickinson Kaishen Wang Taylor Judd Joseph D. Barnes Joshua Mason Michael Bailey

{zanema2,joshuar3,jddicki2,kwang40,tjudd,jdbarns1,joshm,mdbailey}@illinois.edu

University of Illinois Urbana-Champaign

CHI 2020

CSET 2019
Ubiquitous URLs

Your Santander Bank Account has been blocked.
All services have been withdrawn.
Go to http://santander.onlineupdates.hec.net.pk to reactivate now.

Important: Your Password will expire in 1 day(s)

MyUniversity
to me

Dear network user,

This email is meant to inform you that your MyUniversity network password will expire in 24 hours.
Please follow the link below to update your password
myuniversity.edu/renewal

Google tracks you. We don’t.

DuckDuckGo.com
URLs in Browsers

Everything is trivially spoofable besides the URL
URL Complexity

What is the second-level domain + TLD?

http://example.com
  service=accountsettings&hl=en-US&continue=https%3A%2F%2Fmyaccount.google.com

https://fb.com/login@example.com%2e2e2e2e2e%2emx?
  @bofa.com/login.php#twitter.com
URL Complexity

What is the second-level domain + TLD?

http://example.com

https://paypal.com.accounts.google.com/signin/v2/identifier?
  service=accountsettings&hl=en-US&continue=https%3A%2F%2Fmyaccount.google.com

https://fb.com/login@example.com.2e2e2e2e.mx?
  @bofa.com/login.php#twitter.com
Research Questions

Given that URLs are ubiquitous and complex:

1. How well do users parse identity information from URLs?

2. What URL features or user strategies lead to mistakes?

94 Mechanical Turk participants
User Confidence

“I know how to read a URL”

• 91/94 reported “Very True” or “Mostly True”

“I know how to tell what website I am on”

• 91/94 reported “Very True” or “Mostly True”
Target Identification

Asked users to describe the target of 19-20 URLs, some with one of 13 different URL obfuscations applied

Median: 54.1%
Research Questions

Given that URLs are ubiquitous and complex:

1. How well do users parse identity information from URLs?
   - Poorly (54% median accuracy), despite user confidence

2. What URL features or user strategies lead to mistakes?
# URL Obfuscation

Unobfuscated URLs 93% accuracy; obfuscated URLs 40% accuracy

<table>
<thead>
<tr>
<th>Obfuscation</th>
<th>Example</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (Control)</td>
<td><a href="https://example.com/login">https://example.com/login</a></td>
<td>93%</td>
</tr>
<tr>
<td>Typosquatting</td>
<td><a href="https://exemple.com/login">https://exemple.com/login</a></td>
<td>70%</td>
</tr>
<tr>
<td>IDN Homograph</td>
<td><a href="https://%D0%B5%D0%B6%D0%B0%CC%81%D0%BCple.com/login">https://ежа́мple.com/login</a></td>
<td>53%</td>
</tr>
<tr>
<td>Self-Declared Secure</td>
<td><a href="https://secure-example.com/login">https://secure-example.com/login</a></td>
<td>36%</td>
</tr>
<tr>
<td>Fake ID in Credentials</td>
<td><a href="https://example.com@a4930.nz/login">https://example.com@a4930.nz/login</a></td>
<td>32%</td>
</tr>
<tr>
<td>URL Encoding Hides Subdomain as Domain</td>
<td><a href="https://example.com%2e2x-log.in">https://example.com%2e2x-log.in</a></td>
<td>29%</td>
</tr>
<tr>
<td>Long Subdomain Chain</td>
<td><a href="https://example.com.0jg094.05930.3590902sdg9f0.249905930.3590902sdg.mx/login">https://example.com.0jg094.05930.3590902sdg9f0.249905930.3590902sdg.mx/login</a></td>
<td>26%</td>
</tr>
</tbody>
</table>
Observed Parsing Strategies

“... highlight each group of characters that helps you learn the identity of the website it points to”
Observed Parsing Strategies

“... highlight each group of characters that helps you learn the identity of the website it points to”

https://secure-twitter.com@google.com@cnn.com%2ebay.com%46buy-and-sell-online.com?@facebook.com@paypal.com****SECURE-BANK-OF-AMERICA-SITE****
Evaluation Strategies

“When you see a link/URL, how do you decide if it is safe to go there?”

Check for HTTPS

“I know it is safe when it reads https, the s stands for secure for me.”

“I first think about if it is a place I know is a legit website. Then I’m looking for HTTPS cert and if the URL just look sensible.”
Evaluation Strategies

“When you see a link/URL, how do you decide if it is safe to go there?”

Check for HTTPS

Familiarity

“I check the url for familiarity. It’s quite frankly easy to tell if it’s an official link to an authentic website.”

“…Like if I’m opening company A and the URL is companyA.com/... I would click it.”
Evaluation Strategies

“When you see a link/URL, how do you decide if it is safe to go there?”

Check for HTTPS

Familiarity

URL fields

“Check to see if it’s mispelled [sic] or weird”

“If it looks like crazy letters then I don’t click it”

“...Also check the prefix of the site and the domain of it. .com .org .ru things of that nature"
Evaluation Strategies

“When you see a link/URL, how do you decide if it is safe to go there?”

Check for HTTPS

Familiarity

URL fields

External tools/context

“I have a antivirus scanner, so it will check whether the site is safe or unsafe.”

“I consider the context of how it was presented to me. Sketchy email? No thanks. Someone spams a shortened link on a forum advertising something that’s too good to be true? No thanks.”
Target Identification

Asked users to describe the target of 19-20 URLs, some with one of 13 different URL obfuscations applied

Median: 54.1%
Making URLs More Usable

Solutions that work without changing ubiquitous URLs?

Automated familiarity tracking

This is a new, unfamiliar website
Making URLs More Usable

Solutions that work without changing ubiquitous URLs?

Automated familiarity tracking

Alternate URL presentations

https://paypal.com.accounts.ggle.com

https://com.ggle.accounts.com.paypal
Phishing: Root Causes

Mistaken Identity

URL complexity leads to mistaken identity

Misplaced Trust

Users may (mis)place trust in HTTPS

Measuring Identity Confusion with Uniform Resource Locators

Joshua Reynolds¹, Deepak Kumar², Zane Ma³, Rohan Subramanian¹, Meishan Wu¹
Martin Shelton¹, Joshua Mason¹, Emily Stark¹, Michael Bailey¹

¹University of Illinois at Urbana-Champaign ²Google, Inc.
³(joshuar3, dkrmar11, zanema2, rsprod2a, meishan2, joshm, mdlbailey)@illinois.edu

The Impact of Secure Transport Protocols on Phishing Efficacy

Zane Ma, Joshua Reynolds, Joseph Dickinson, Kaishen Wang,
Taylor Judd, Joseph D. Barnes, Joshua Mason, Michael Bailey
{zanema2, joshuar3, jdicki2, kwang40, tjudd, jdbarns1, joshm, mdlbailey}@illinois.edu

University of Illinois Urbana-Champaign

CHI 2020

CSET 2019
Existing Security Protocols Lack Trustworthiness

Not designed to protect against phishing

TLS = Confidentiality + Integrity + Identity/Authenticity

TLS secures connections, not content

Prior work:

1. Some users look at connection security indicators when exposed to phishing

2. Users confuse “connection security” and “site security”
Experimental Goals

1. Does the presence of secure transport protocols make phishing more effective?

   Methodology: A/B test HTTP/HTTPS and SMTP/SMTP+TLS

2. Does browser URL bar UI (e.g. security indicators) influence phishing susceptibility?

   Methodology: Generate and feature code browser screenshots, correlate URL bar features with phishing outcomes
Phishing Experiment

1. Open Email
2. Access Site
3. Submit Credentials
4. Opt-In To Survey

**University of Illinois Phishing Survey**

Demographics

1. Are you male or female?
   - [ ] Male
   - [ ] Female
   - [ ] Other
   - [ ] Prefer not to answer

2. What is your age?
   - [ ] 18-24
   - [ ] 25-34
   - [ ] 35-44
   - [ ] 45-54
   - [ ] 55-64
   - [ ] 65 or older

---

**Technology Services**

University of Illinois Technology Services - Phishing Awareness Drill

1. Purpose of the Study
2. Experiment
3. Follow-Up Survey & Compensation
4. Participation
5. Education
6. Contact Information

---

**Fraudulent Email**

From: zane@emailillinois.edu

To: John Doe

Subject: Network Abuse Warning

Dear John,

This notice is being served as a warning that the computer registered to you has been discovered attempting to make repeated connections to prohibited/illegal sites. Technology Services takes the misuse of the UNIVERSITY campus network seriously and will blacklist and report this device according to the terms of the Technology Access Policy of Computers and Network Systems at the University. For more information of if you believe you have received this notification in error, please follow the link below.

Follow this link or paste the following into your browser:

http://technology.illinois.edu/abuse-warning/hdb/Ch9huG48wC2uAK7pQ5bM39CwA5-xe42UJ17r1

Kevin Randolph,
Office of Technology Services
Legal Compliance Officer

"You are now doing your job well"

---

**Phishing Experiment**

1. Open Email
2. Access Site
3. Submit Credentials
4. Opt-In To Survey
Phishing Campaign

Target population: 266 employees of a university IT organization

0. Send Email
266 Users 100%

1. Open Email
140 Users 53%

2. Access Site
92 Users 35%

3. Submit Credentials
57 Users 21%
Q1: Phishing Effectiveness

2. Access Site

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Success Rate</th>
<th>p-value</th>
<th>3. Enter Credentials</th>
<th>Success Rate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HTTP</strong></td>
<td>45/75 = 60.0%</td>
<td>0.17</td>
<td>25/45 = 55.6%</td>
<td></td>
<td>0.31</td>
</tr>
<tr>
<td><strong>HTTPS</strong></td>
<td>47/65 = 72.3%</td>
<td></td>
<td>32/47 = 68.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TLS Email</strong></td>
<td>45/71 = 63.3%</td>
<td>0.96</td>
<td>30/47 = 63.8%</td>
<td></td>
<td>0.87</td>
</tr>
<tr>
<td><strong>No TLS Email</strong></td>
<td>45/69 = 65.2%</td>
<td></td>
<td>27/45 = 60.0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q2: Browser UI Correlation

Feature coded 2,882 screenshots across different browsers / platforms / OS

Correlate features with HTTP User-Agent for susceptible users

Mac 10.13 Chrome 63

Windows 10 Edge 16

Windows XP SP2 Firefox 3.0

Galaxy S7 Android 70 Mbl. Chrome 63

iPhone 8 iOS 11 Mbl Safari 11.0

https://github.com/teamnsrg/url-bar-coding
Q2: Browser UI Correlation

<table>
<thead>
<tr>
<th>Feature</th>
<th>$P_{exp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any Icon?</td>
<td>0.25</td>
</tr>
<tr>
<td>Lock Icon?</td>
<td>0.32</td>
</tr>
<tr>
<td>Lock Position</td>
<td>0.98</td>
</tr>
<tr>
<td>Lock Color</td>
<td>0.55</td>
</tr>
<tr>
<td>Detailed Lock?</td>
<td>0.54</td>
</tr>
<tr>
<td>Lock Additions</td>
<td>0.27</td>
</tr>
<tr>
<td>Favicon?</td>
<td>0.56</td>
</tr>
<tr>
<td>Favicon Position</td>
<td>0.32</td>
</tr>
<tr>
<td>Default Favicon</td>
<td>0.06</td>
</tr>
<tr>
<td>Protocol Visible?</td>
<td>0.07</td>
</tr>
<tr>
<td>Protocol Emphasis</td>
<td>0.63</td>
</tr>
<tr>
<td>Additional Text?</td>
<td>0.62</td>
</tr>
<tr>
<td>Add. Text Emphasis</td>
<td>0.62</td>
</tr>
<tr>
<td>Add. Text Background</td>
<td>0.97</td>
</tr>
<tr>
<td>Icon/URL Separator?</td>
<td>0.42</td>
</tr>
</tbody>
</table>

14/16 = 87.5% of users who saw protocol submitted credentials

![Phishing Example](https://phish-staging.sprai.org/abuse-warning)

27/46 = 58.7% of users who did not see protocol submitted credentials

![Phishing Example](https://phish-staging.sprai.org/abuse-warning)
Q2: Browser UI Correlation

<table>
<thead>
<tr>
<th>Feature</th>
<th>( p_{exp} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any Icon?</td>
<td>0.25</td>
</tr>
<tr>
<td>Lock Icon?</td>
<td>0.32</td>
</tr>
<tr>
<td>Lock Position</td>
<td>0.98</td>
</tr>
<tr>
<td>Lock Color</td>
<td>0.55</td>
</tr>
<tr>
<td>Detailed Lock?</td>
<td>0.54</td>
</tr>
<tr>
<td>Lock Additions</td>
<td>0.27</td>
</tr>
<tr>
<td>Favicon?</td>
<td>0.56</td>
</tr>
<tr>
<td>Favicon Position</td>
<td>0.32</td>
</tr>
<tr>
<td>Default Favicon</td>
<td>0.06</td>
</tr>
<tr>
<td>Protocol Visible?</td>
<td>0.07</td>
</tr>
<tr>
<td>Protocol Emphasis</td>
<td>0.63</td>
</tr>
<tr>
<td>Additional Text?</td>
<td>0.62</td>
</tr>
<tr>
<td>Add. Text Emphasis</td>
<td>0.62</td>
</tr>
<tr>
<td>Add. Text Background</td>
<td>0.97</td>
</tr>
<tr>
<td>Icon/URL Separator?</td>
<td>0.42</td>
</tr>
</tbody>
</table>

9/10 “Secure” submitted credentials

8/10 “Not Secure” submitted credentials
Takeaways

• The presence of HTTPS in phishing tended to increase effectiveness, but...need more data, more diverse target population

• Protocol presence may increase phishing susceptibility, while “Secure/Not Secure” had minimal distinction

• Another hint that users conflate credibility/trustworthiness with connection security
Collaborators

Michael Bailey  Josh Mason  Deepak Kumar  Joshua Reynolds

Not pictured: Martin Shelton, Emily Stark, Kaishen Wang, Joseph Dickinson, Rohan Subramanian, Meishan Wu, Illinois Tech Services
Phishing: Root Causes

Mistaken Identity

- accoounts-google.com

Misplaced Trust

- questionsaboutisps.com
Phishing: Root Causes

Mistaken Identity

URL complexity leads to mistaken identity

Misplaced Trust

Users may (mis)place trust in HTTPS

Measuring Identity Confusion with Uniform Resource Locators

Joshua Reynolds1, Deepak Kumar1, Zane Ma1, Rohan Subramanian1, Meishan Wu1, Martin Shelton1, Joshua Mason1, Emily Stark1, Michael Bailey1

1University of Illinois at Urbana-Champaign 2Google, Inc.

{zanema2,joshuar3,jddicki2,kwang40,tjudd,jdbarns1,joshm,mdbailey}@illinois.edu

CHI 2020

The Impact of Secure Transport Protocols on Phishing Efficacy

Zane Ma, Joshua Reynolds, Joseph Dickinson, Kaishen Wang, Taylor Judd, Joseph D. Barnes, Joshua Mason, Michael Bailey

{zanema2,joshuar3,jddicki2,kwang40,tjudd,jdbarns1,joshm,mdbailey}@illinois.edu

University of Illinois Urbana-Champaign

CSET 2019
Fundamentals of Phishing: A Usability Perspective

Zane Ma

University of Illinois Urbana-Champaign
zanema2@illinois.edu
https://zanema.com