

Reflections on Trusting Identifiers: The Foundations of Social Engineering

Zane Ma

Georgia Institute of Technology

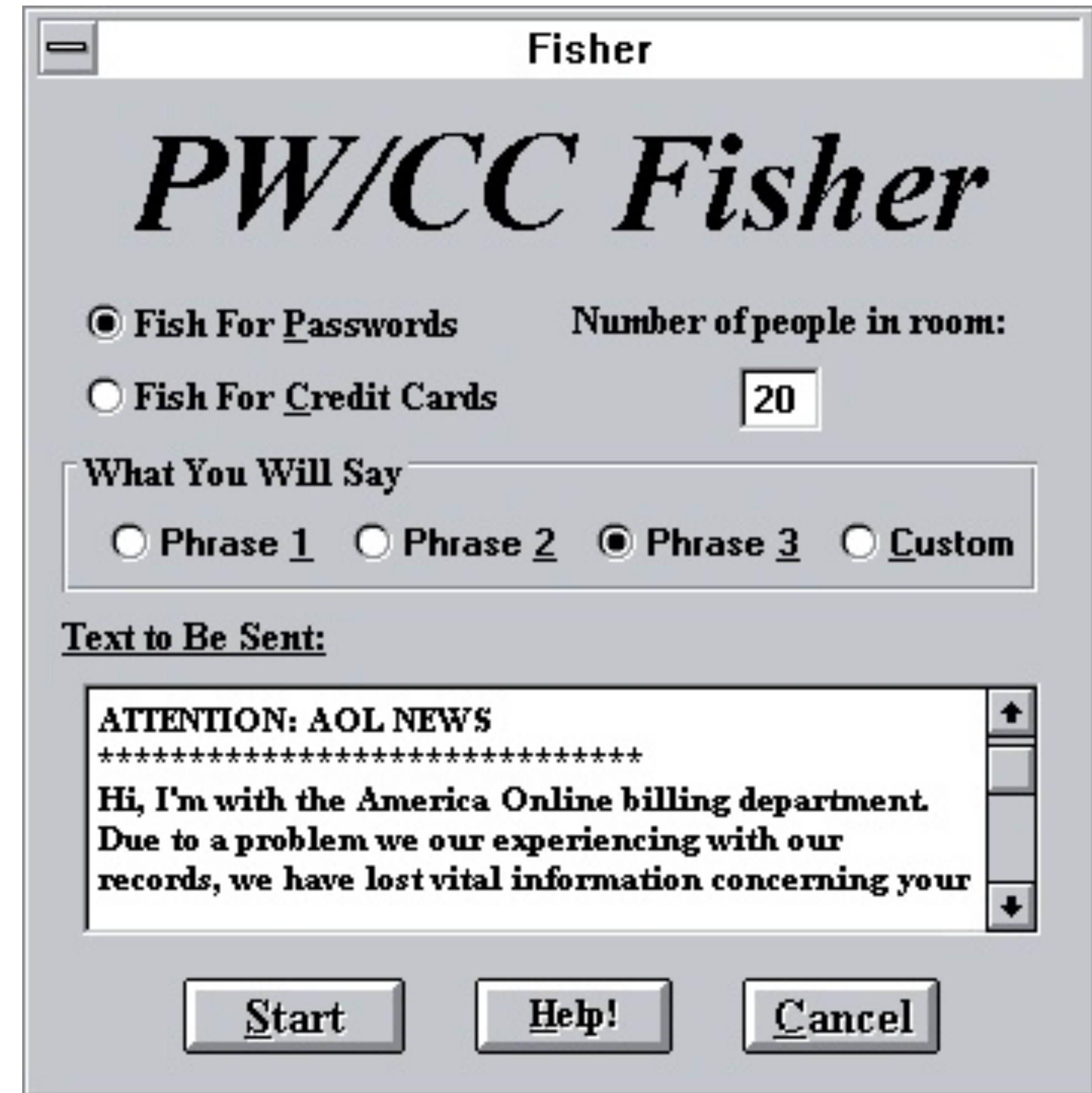
zanema@gatech.edu

<https://zanema.com>

Social Engineering Roots



1995 - AOL cracking tool



Social Engineering: Today

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

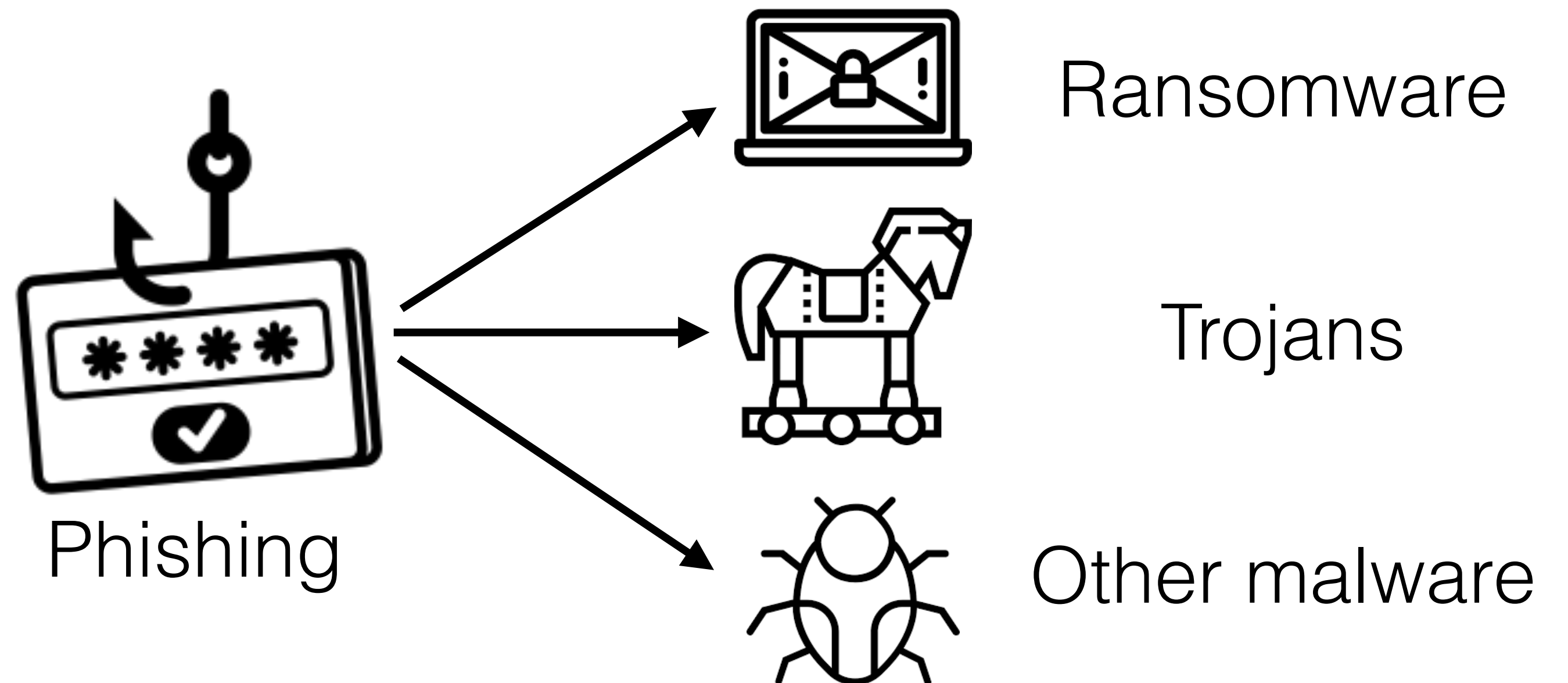
WIRED

LILY HAY NEWMAN

SECURITY 01.31.2020 05:08 PM

Watch Out for Coronavirus Phishing Scams

2020
DEMOCRATIC
NATIONAL
CONVENTION



Source: 2019 FBI Internet Crime Report

Phishing: Today

Crime Type	Victims
Phishing/Vishing/Smishing/Pharming	114,702
Non-Payment/Non-Delivery	61,832
Extortion	43,101
Personal Data Breach	38,218
Spoofing	25,759
BEC/EA	23,775
Confidence Fraud/Romance	19,473
Identity Theft	16,953
Harassment/Threats of Violence	15,512
Overpayment	15,395
Advanced Fee	14,607
Employment	14,493
Credit Card Fraud	14,378
Government Impersonation	13,873
Tech Support	13,633
Real Estate/Rental	11,677
Other	10,842

WIRED

LILY HAY NEWMAN

SECURITY 01.31.2020 05:08 PM

Watch Out for Coronavirus Phishing Scams

2016
DEMOCRATIC
NATIONAL
CONVENTION

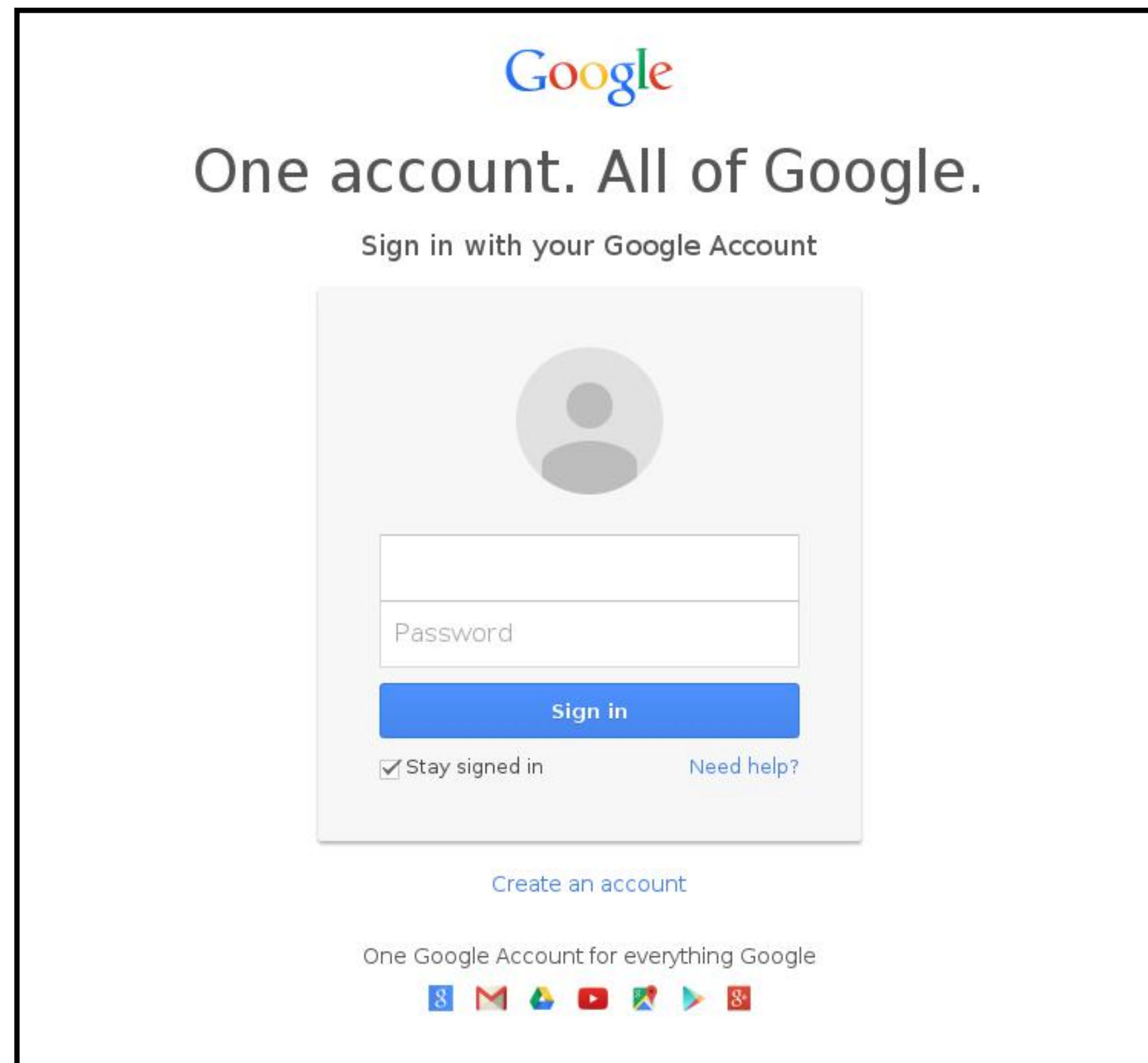
Why haven't we solved/curtailed social engineering, twenty-five years later?



Source: 2019 FBI Internet Crime Report

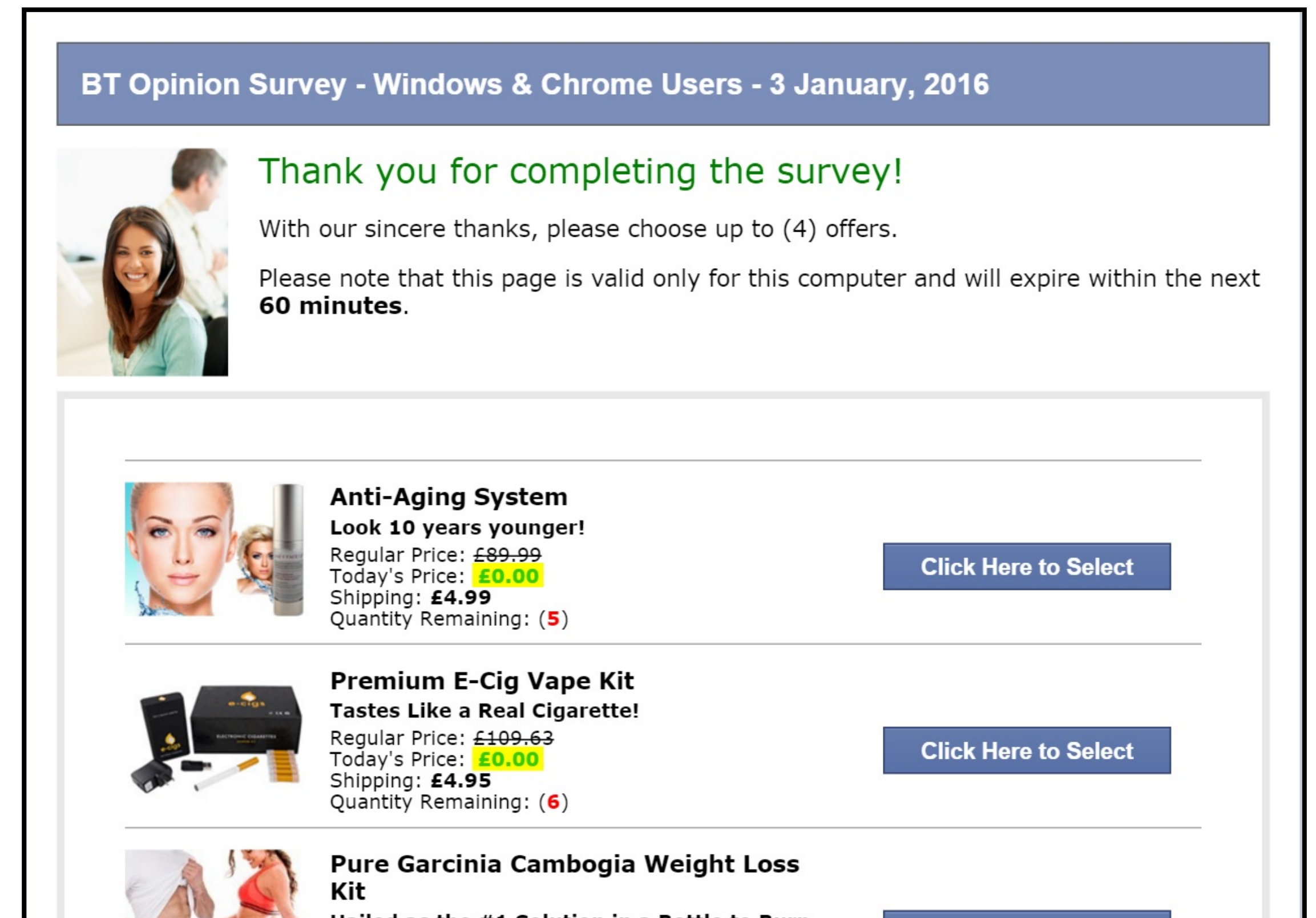
Social Engineering: Root Causes

Mistaken Identity



accounts-google.com

Misplaced Trust



questionsaboutisps.com

Social Engineering: Root Causes

Mistaken Identity

Misplaced Trust

Measuring Identity Confusion with Uniform Resource Locators

Joshua Reynolds[†] Deepak Kumar[†] Zane Ma[†] Rohan Subramanian[†] Meishan Wu[†]
Martin Shelton[‡] Joshua Mason[†] Emily Stark[‡] Michael Bailey[†]
[†]University of Illinois at Urbana-Champaign [‡]Google, Inc.
{joshuar3, dkumar11, zanema2, rcsupra2, meishan2, joshm, mdbailey}@illinois.edu

CHI 2020

The Impact of Secure Transport Protocols on Phishing Efficacy

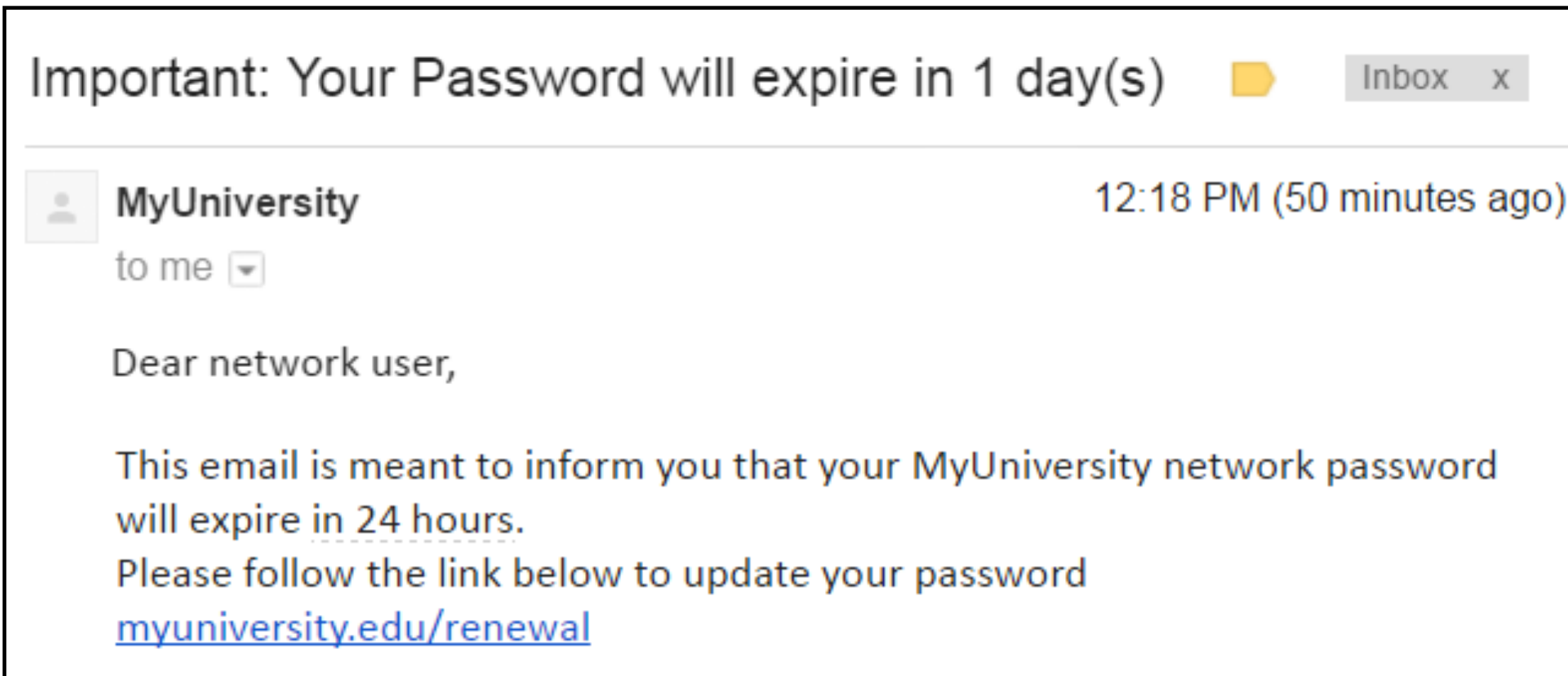
Zane Ma Joshua Reynolds Joseph Dickinson Kaishen Wang
Taylor Judd Joseph D. Barnes Joshua Mason Michael Bailey
{zanema2,joshuar3,jddicki2,kwang40,tjudd,jdbarns1,joshm,mdbailey}@illinois.edu
University of Illinois Urbana-Champaign

CSET 2019

URL complexity leads
to mistaken identity

Users may (mis)place
trust in HTTPS

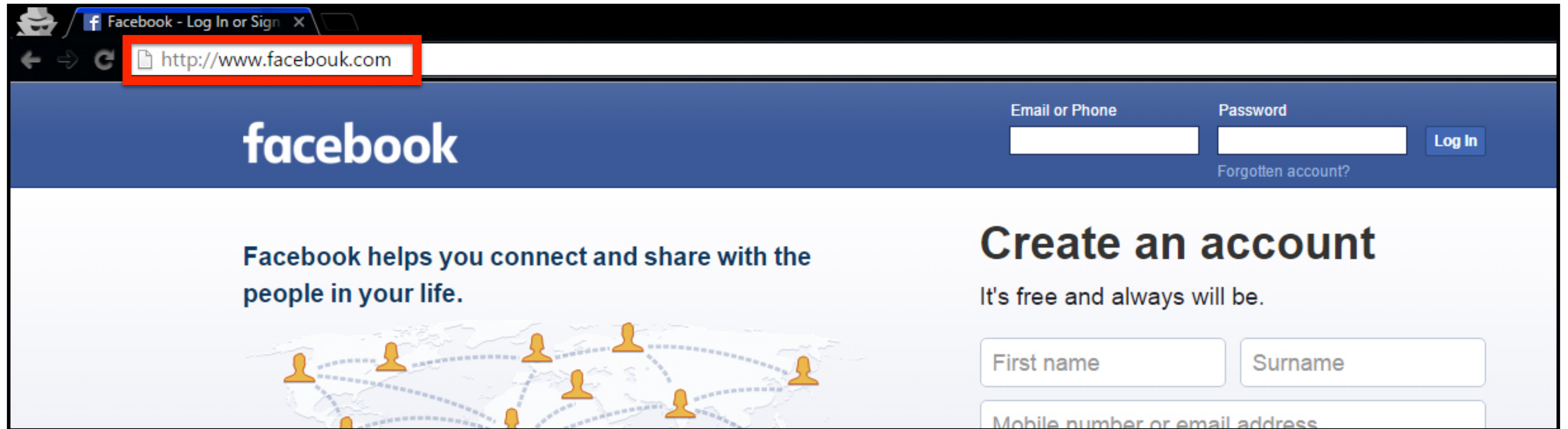
Ubiquitous URLs



Your Santander Bank Account has been blocked.
All services have been withdrawn.
Go to <http://santander.onlineupdatesecures.he.net.pk>
to reactivate now.



URLs in Browsers



Everything is trivially spoofable besides the URL

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>

[@bofa.com/login.php#twitter.com](https://fb.com/login@example.com%2e2e2e2e2e2e%2emx)

URL Complexity

What is the second-level domain + TLD?

<http://example.com>

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

[@bofa.com/login.php#twitter.com](https://fb.com/login@example.com%2e2e2e2e2e2e%2emx)

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”

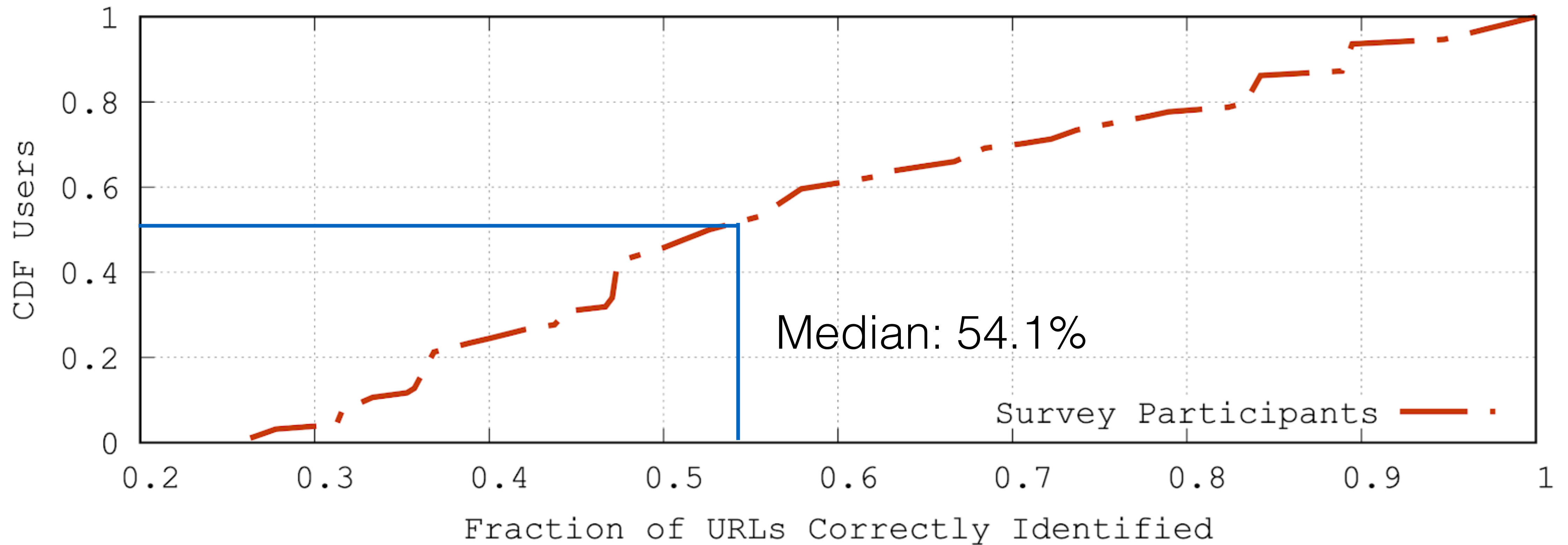
- 91/94 reported “Very True” or “Mostly True”

“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



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?

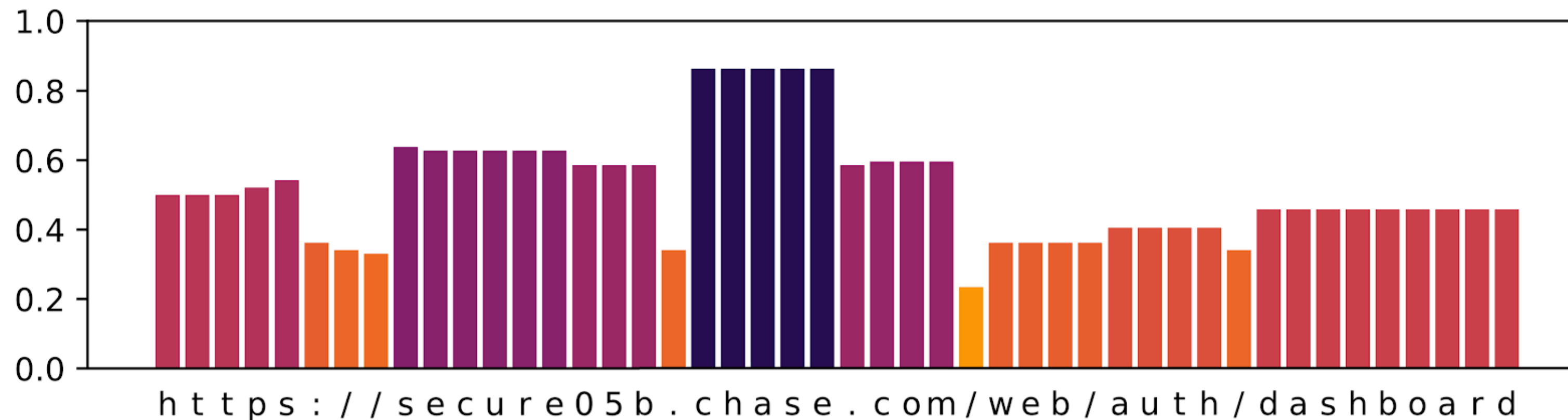
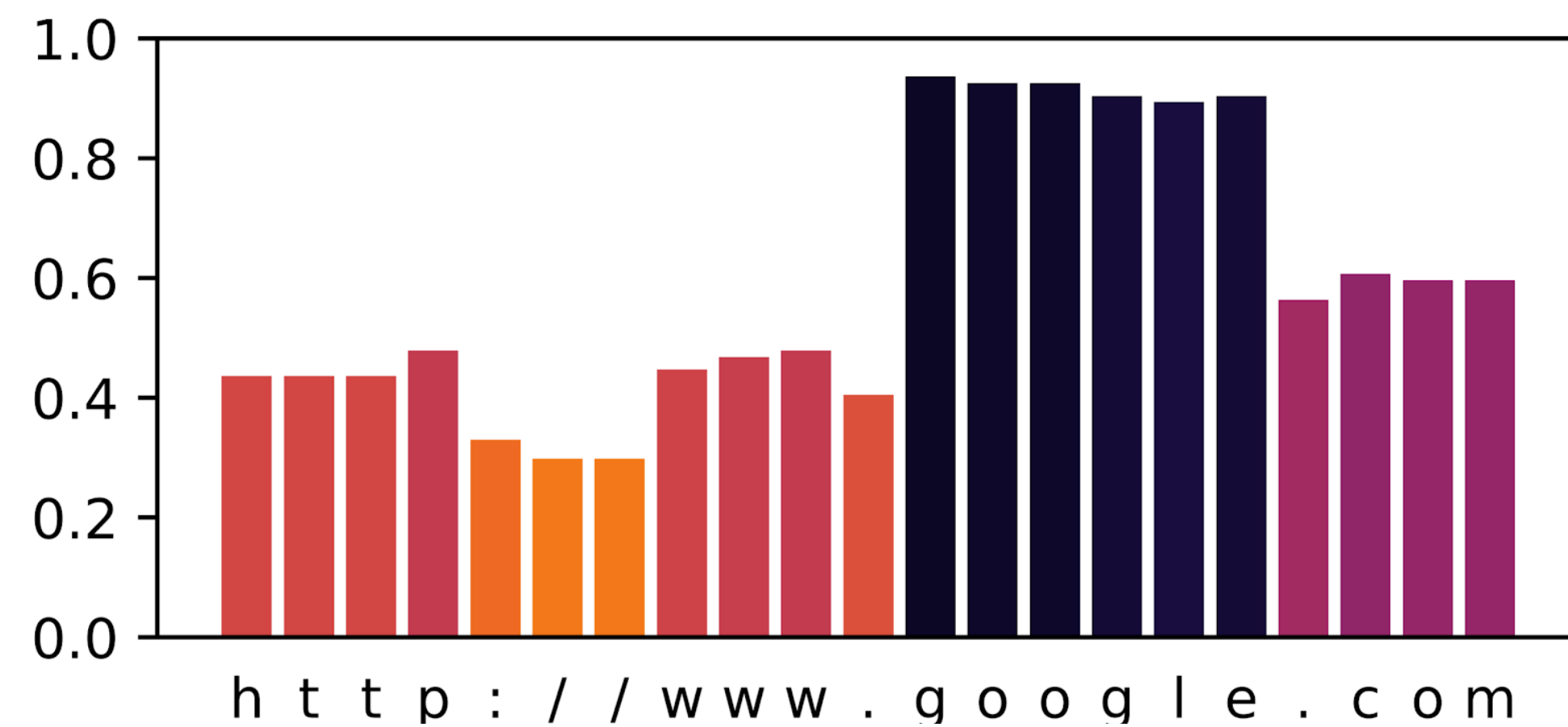
URL Obfuscation

Unobfuscated URLs 93% accuracy; obfuscated URLs 40% accuracy

Obfuscation	Example	Accuracy
None (Control)	https://example.com/login	93%
Typosquatting	https://exemple.com/login	70%
IDN Homograph	https://ежаmple.com/login	53%
Self-Declared Secure	https://secure-example.com/login	36%
Fake ID in Credentials	https://example.com@a4930.nz/login	32%
URL Encoding Hides Subdomain as Domain	https://example.com%2e2x-log.in	29%
Long Subdomain Chain	https://example.com.0jg094.05930.3590902sdg9f0.249905930.3590902sdg.mx/login	26%

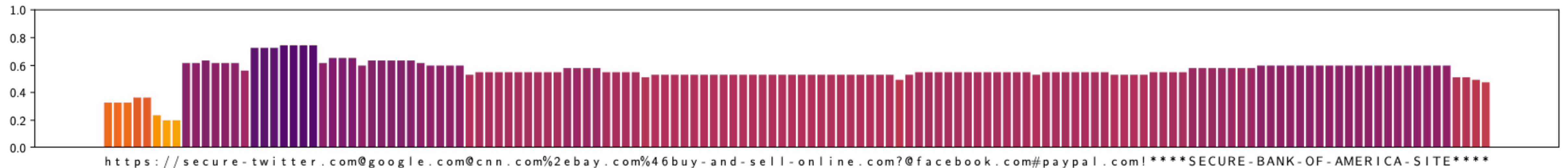
Observed Parsing Strategies

“... 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 misspelled [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

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

Familiarity

URL fields

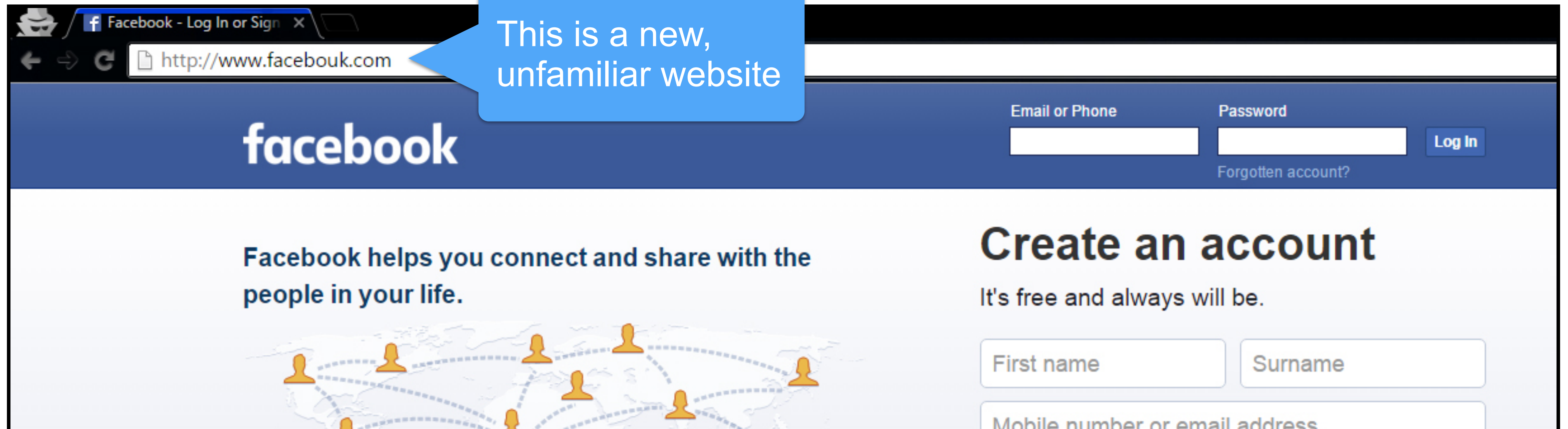
External tools/context

“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.”

Making URLs More Usable

Solutions that work without changing ubiquitous URLs?

Automated familiarity tracking



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>

Social Engineering: Root Causes

Mistaken Identity

Misplaced Trust

Measuring Identity Confusion with Uniform Resource Locators

Joshua Reynolds[†] Deepak Kumar[†] Zane Ma[†] Rohan Subramanian[†] Meishan Wu[†]
Martin Shelton[‡] Joshua Mason[†] Emily Stark[‡] Michael Bailey[†]
[†]University of Illinois at Urbana-Champaign [‡]Google, Inc.
{joshuar3, dkumar11, zanema2, rsubra2, meishan2, joshm, mdb Bailey}@illinois.edu

CHI 2020

URL complexity leads to mistaken identity

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, mdb Bailey}@illinois.edu
University of Illinois Urbana-Champaign

CSET 2019

Users may (mis)place trust in HTTPS

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

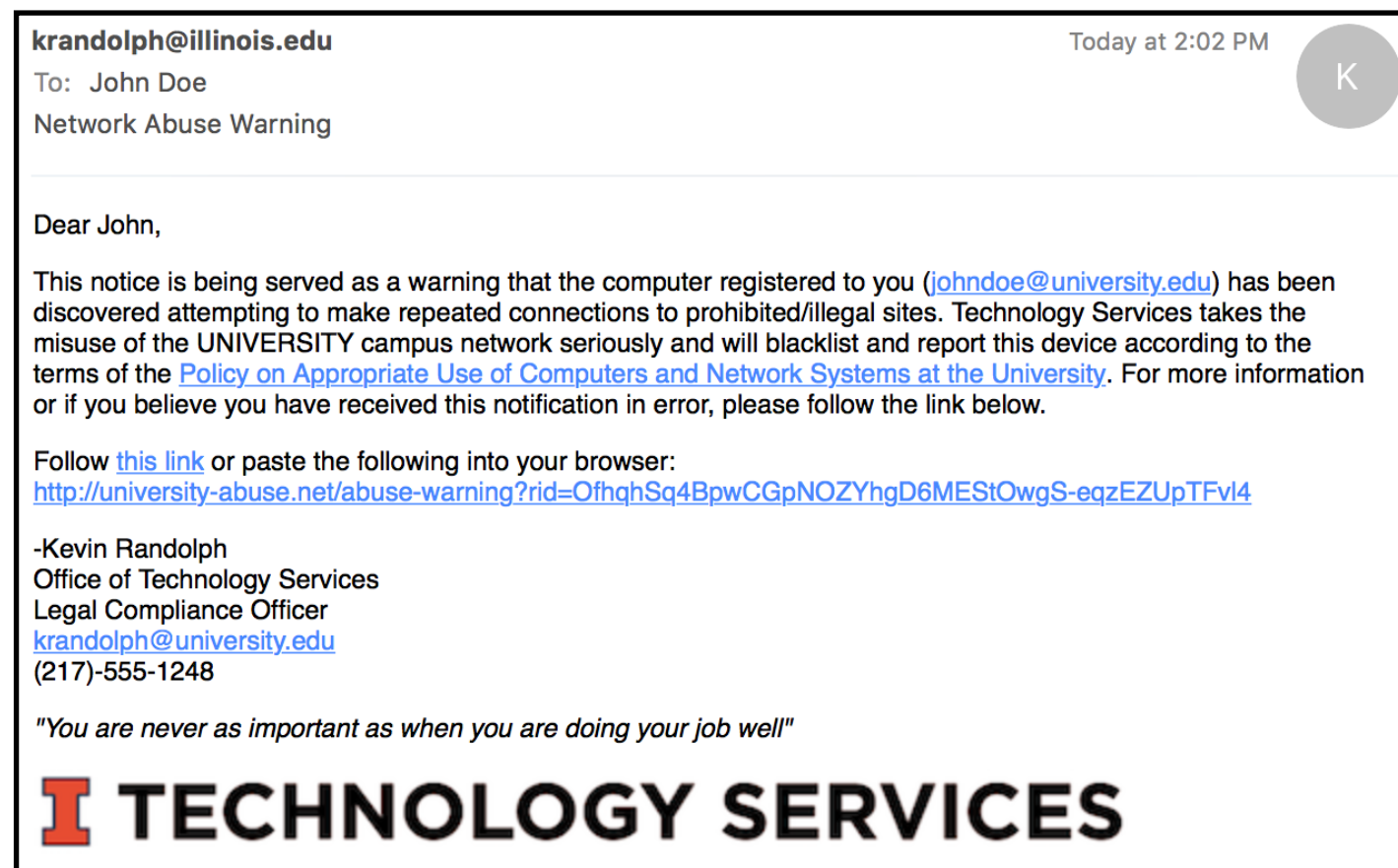
2. Does browser URL bar UI (e.g. security indicators) influence phishing susceptibility?

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

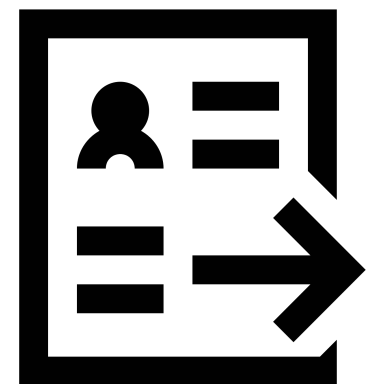
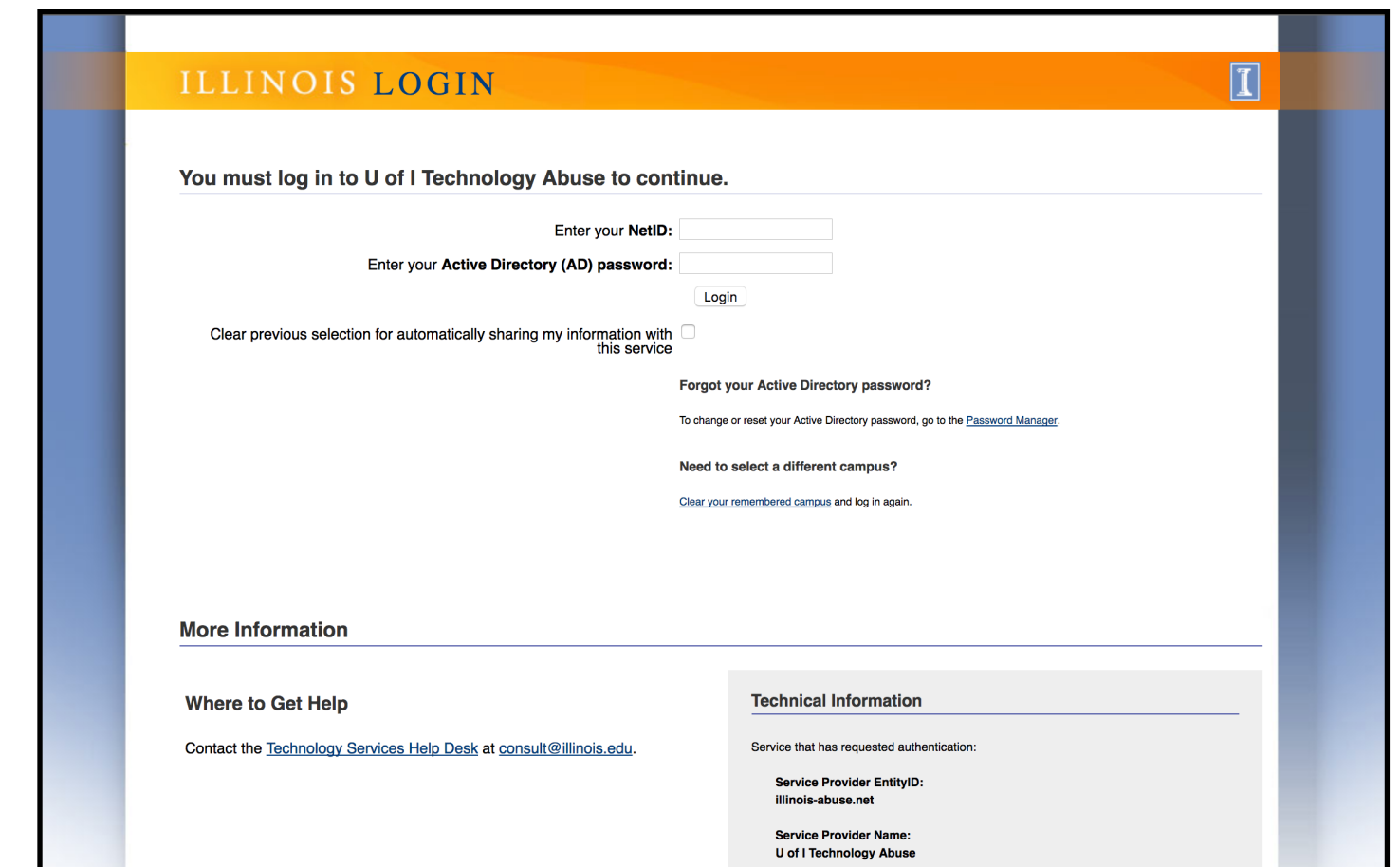
Phishing Experiment



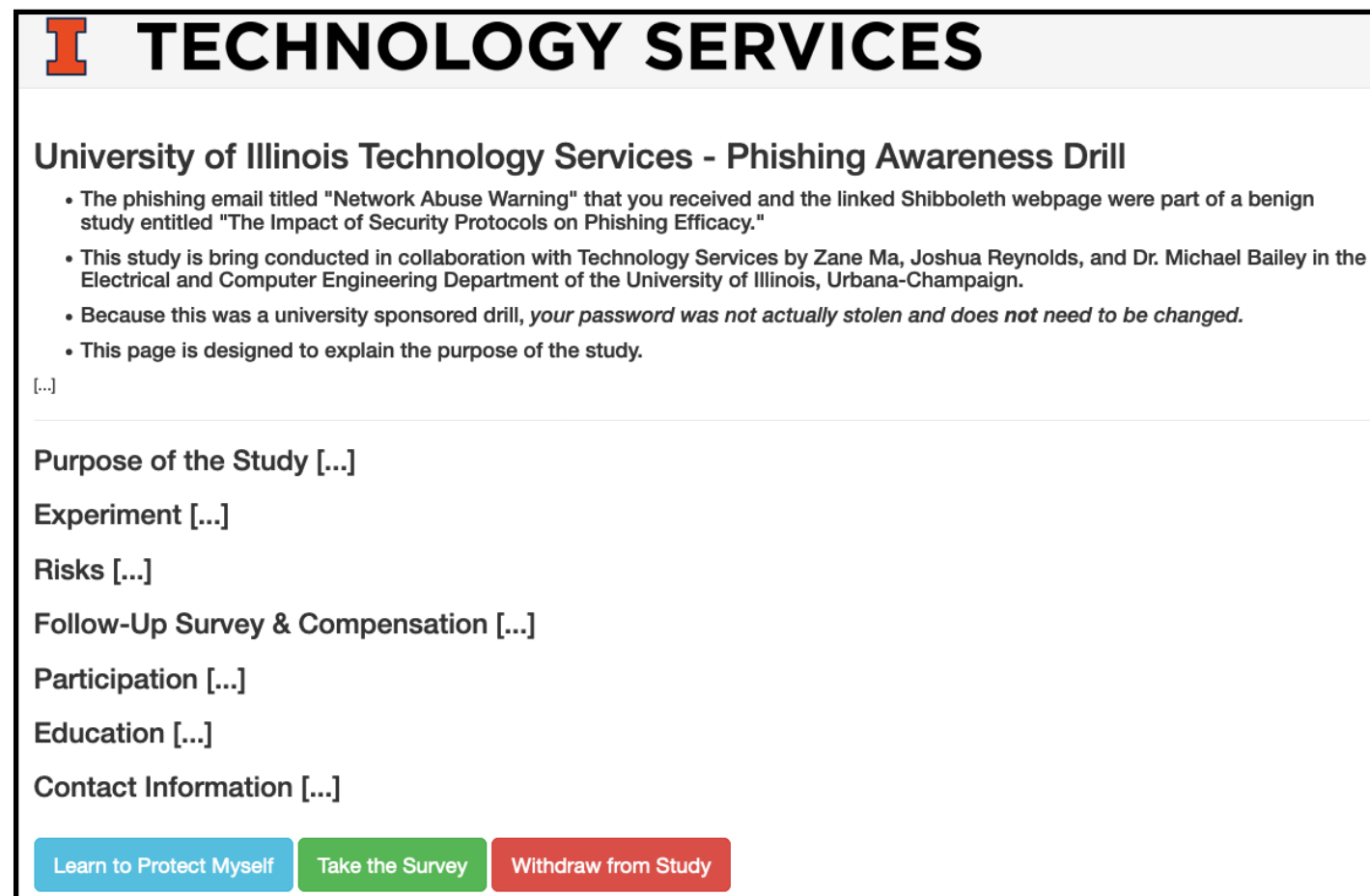
1. Open Email



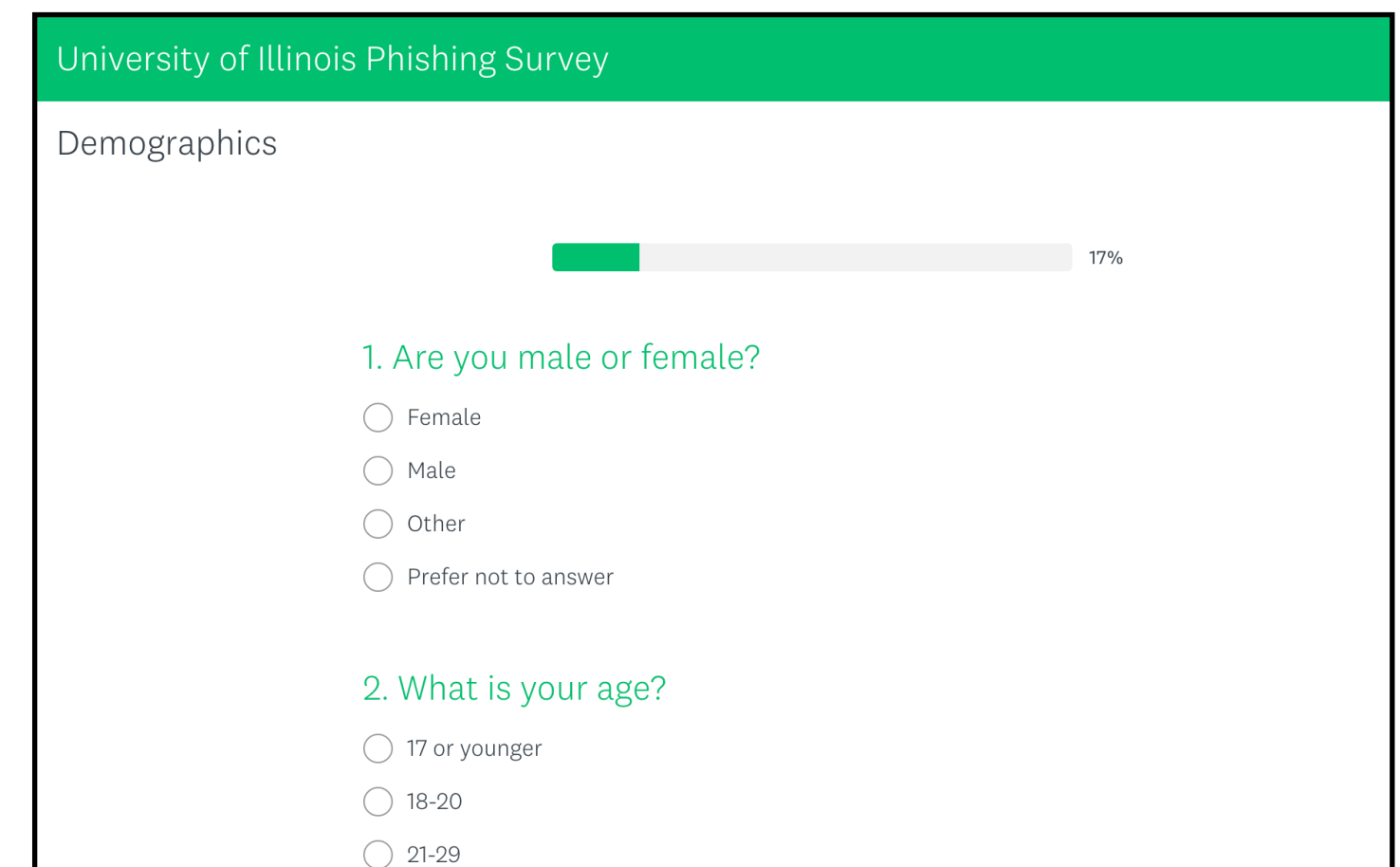
2. Access Site



3. Submit Credentials

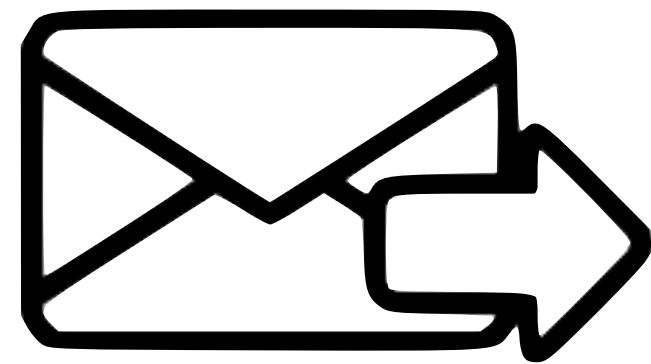


4. Opt-In To Survey



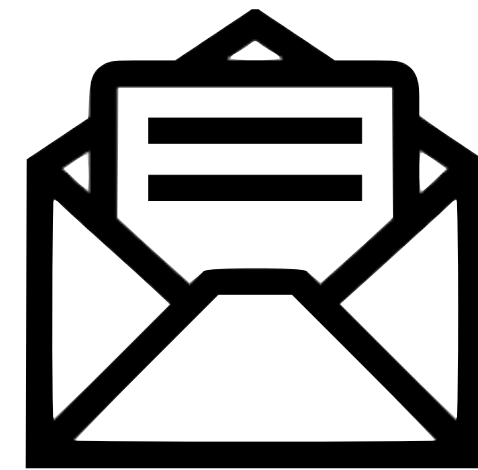
Phishing Campaign

Target population: 266 employees of a university IT organization



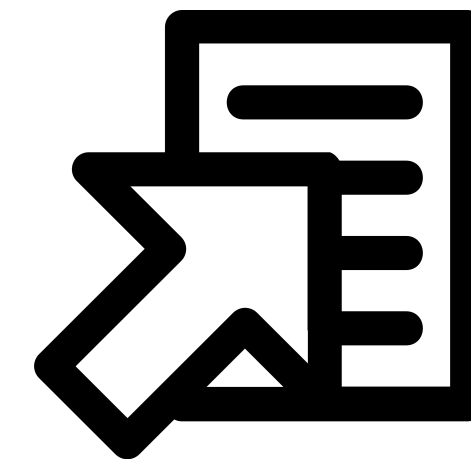
0. Send Email

266 Users
100%



1. Open Email

140 Users
53%



2. Access Site

92 Users
35%
(66%)



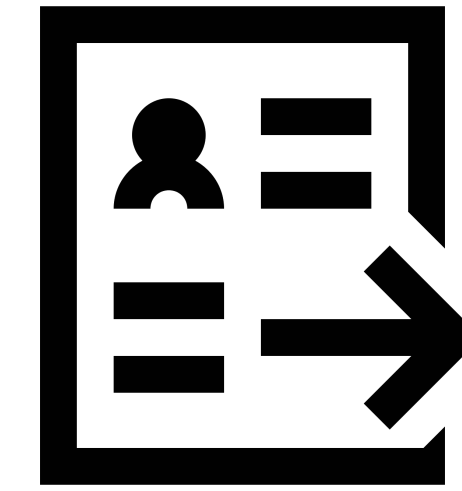
3. Submit Credentials

57 Users
21%
(62%)

Q1: Phishing Effectiveness



2. Access Site



3. Enter Credentials

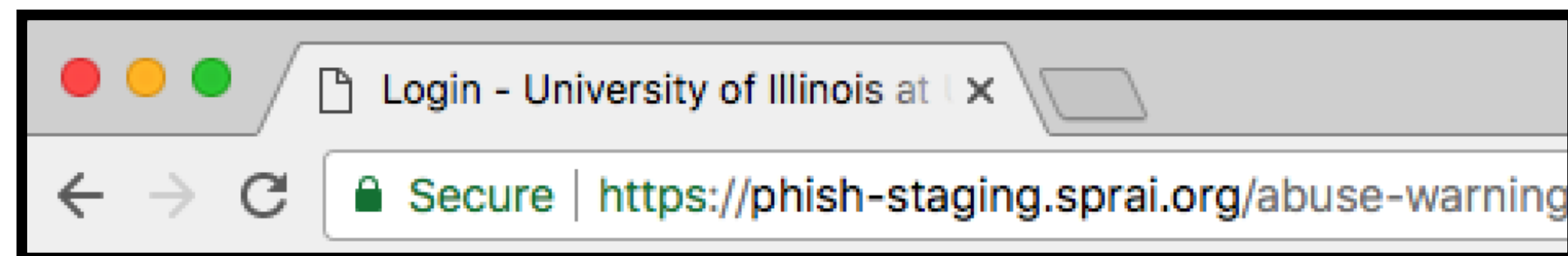
No TLS Email	45/69 = 65.2%	p = 0.96	27/45 = 60.0%	p = 0.87
TLS Email	45/71 = 63.3%		30/47 = 63.8%	
HTTP	45/75 = 60.0%	p = 0.17	25/45 = 55.6%	p = 0.31
HTTPS	47/65 = 72.3%		32/47 = 68.0%	

Q2: Browser UI Correlation

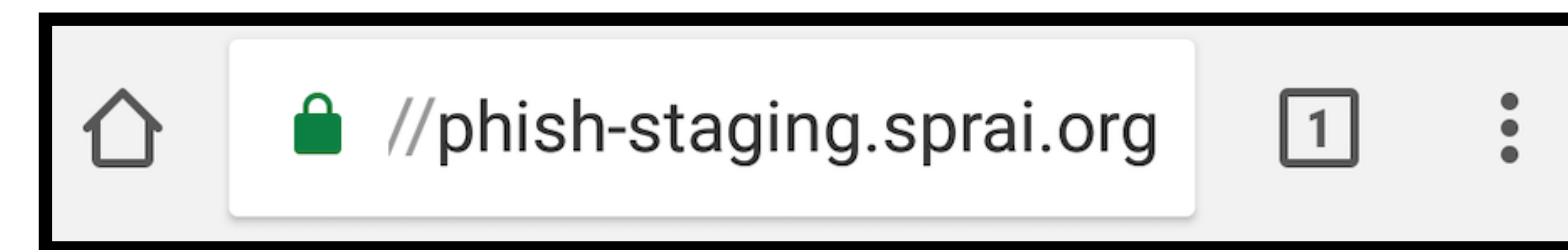
Feature coded 2,882 screenshots across different browsers / platforms / OS

Correlate features with HTTP User-Agent for susceptible users

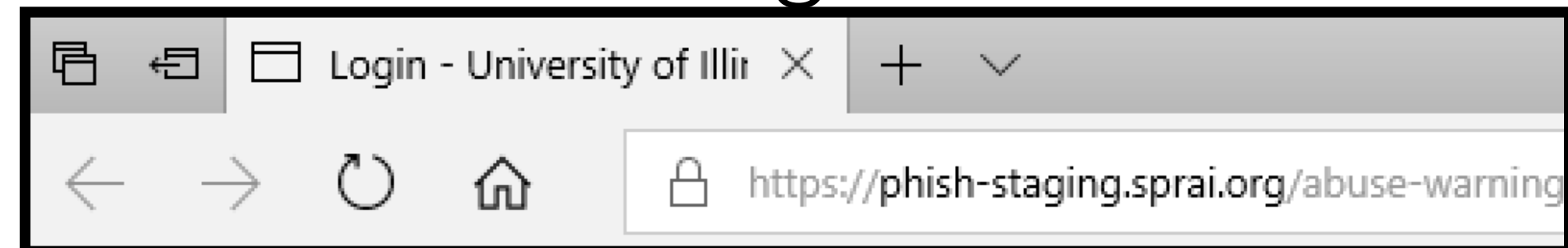
Mac 10.13 Chrome 63



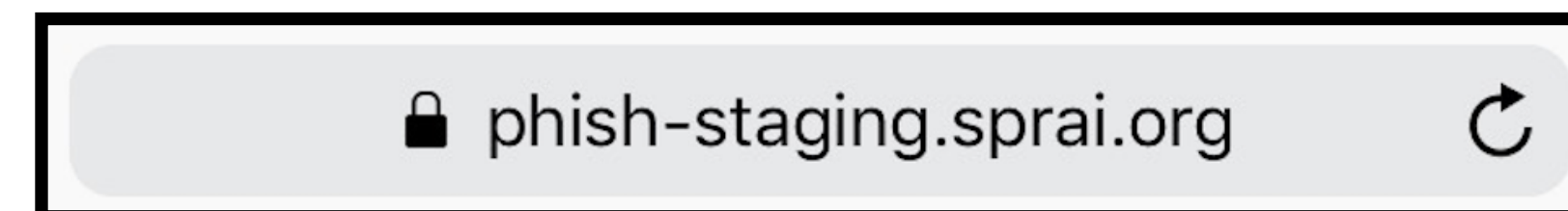
Galaxy S7 Android 7.0 Mbl. Chrome 63



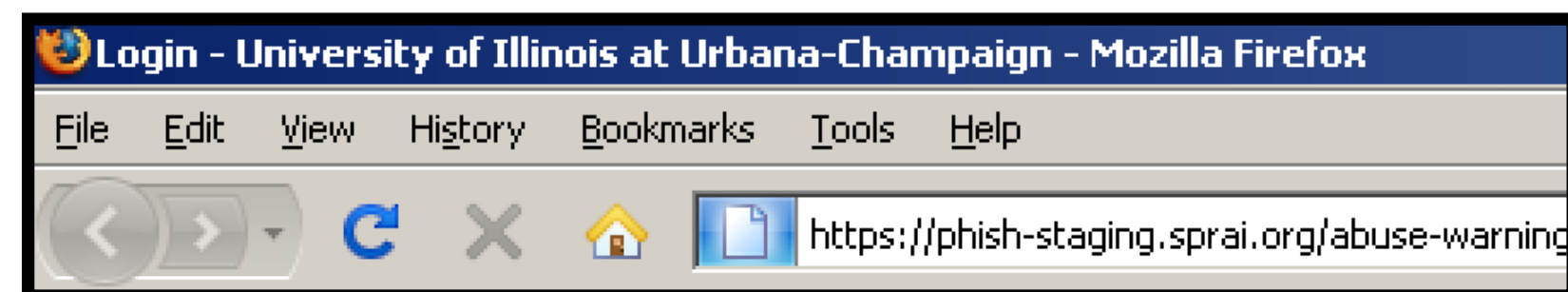
Windows 10 Edge 16



iPhone 8 iOS 11 Mbl Safari 11.0



Windows XP SP2 Firefox 3.0

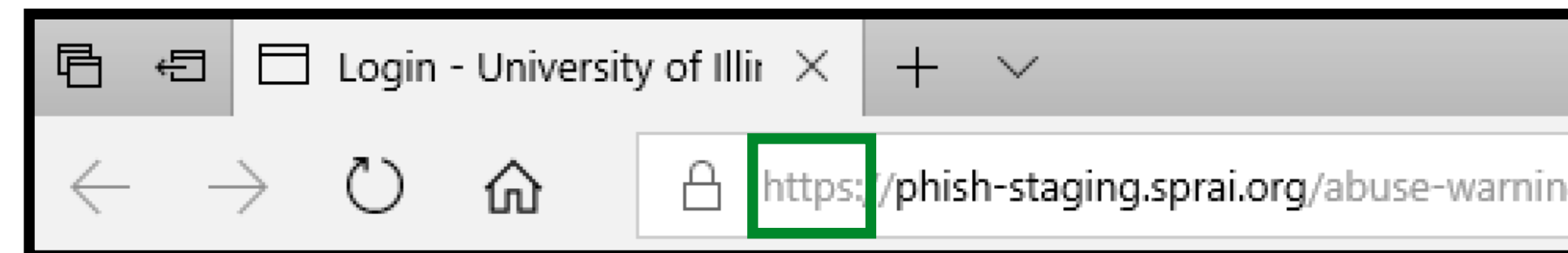


<https://github.com/teamnsrg/url-bar-coding>

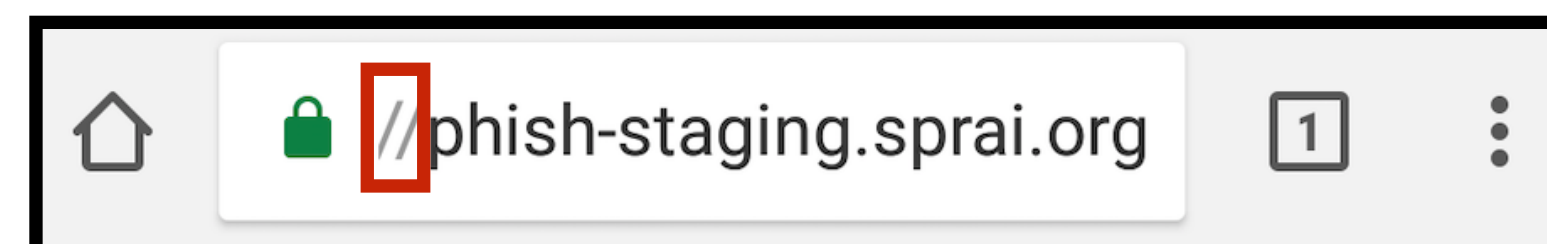
Q2: Browser UI Correlation

Feature	p_{exp}
Any Icon?	0.25
Lock Icon?	0.32
Lock Position	0.98
Lock Color	0.55
Detailed Lock?	0.54
Lock Additions	0.27
Favicon?	0.56
Favicon Position	0.32
Default Favicon	0.06
Protocol Visible?	0.07
Protocol Emphasis	0.63
Additional Text?	0.62
Add. Text Emphasis	0.62
Add. Text Background	0.97
Icon/URL Separator?	0.42

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



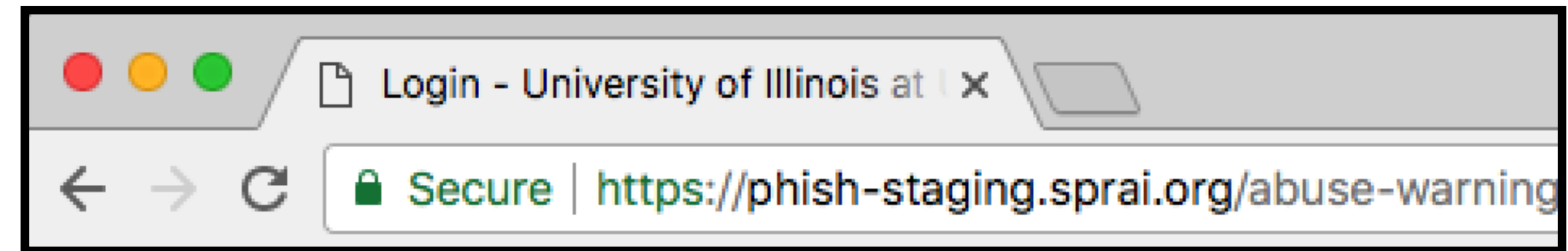
27/46 = 58.7% of users who did not see protocol submitted credentials



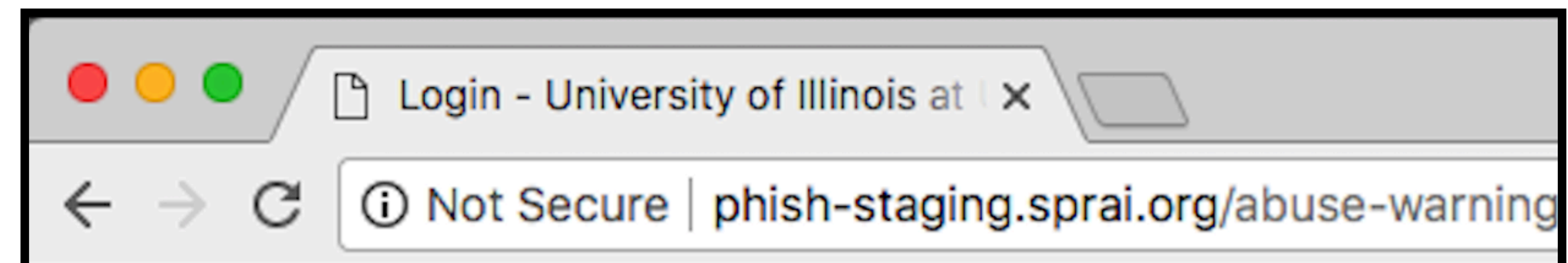
Q2: Browser UI Correlation

Feature	p_{exp}
Any Icon?	0.25
Lock Icon?	0.32
Lock Position	0.98
Lock Color	0.55
Detailed Lock?	0.54
Lock Additions	0.27
Favicon?	0.56
Favicon Position	0.32
Default Favicon	0.06
Protocol Visible?	0.07
Protocol Emphasis	0.63
Additional Text?	0.62
Add. Text Emphasis	0.62
Add. Text Background	0.97
Icon/URL Separator?	0.42

9/10 “Secure” submitted credentials



8/10 “Not Secure” submitted credentials



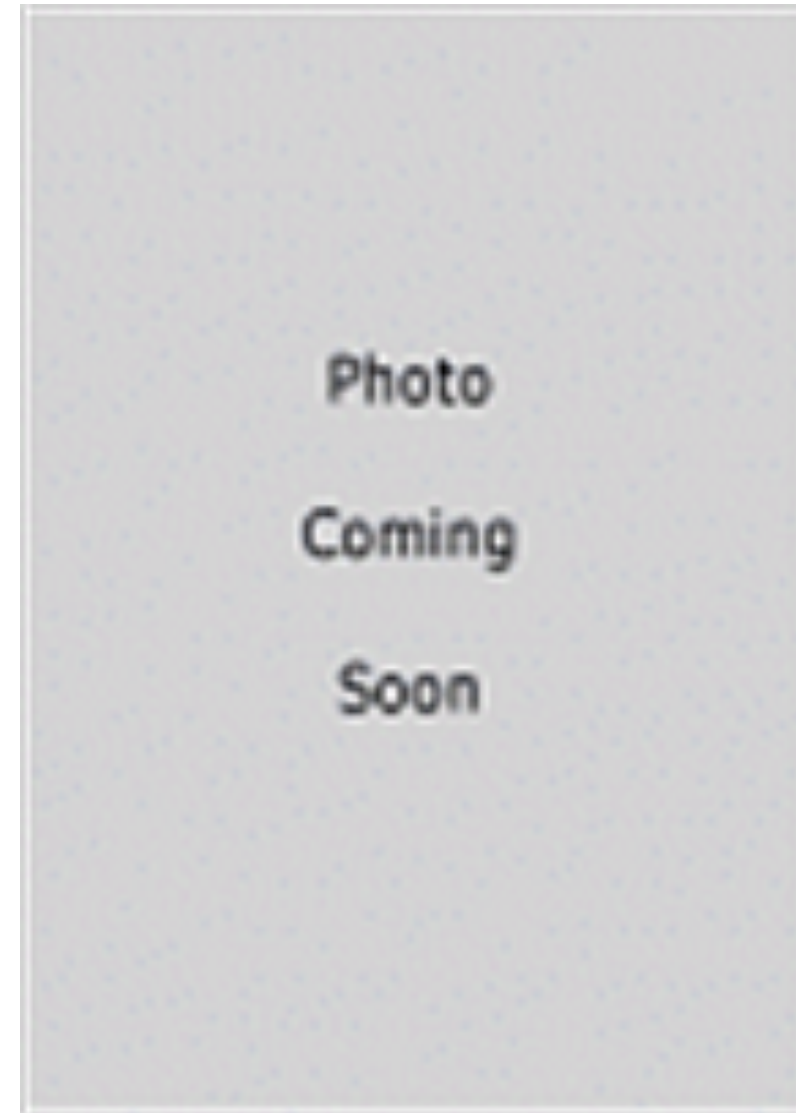
Takeaways

- 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

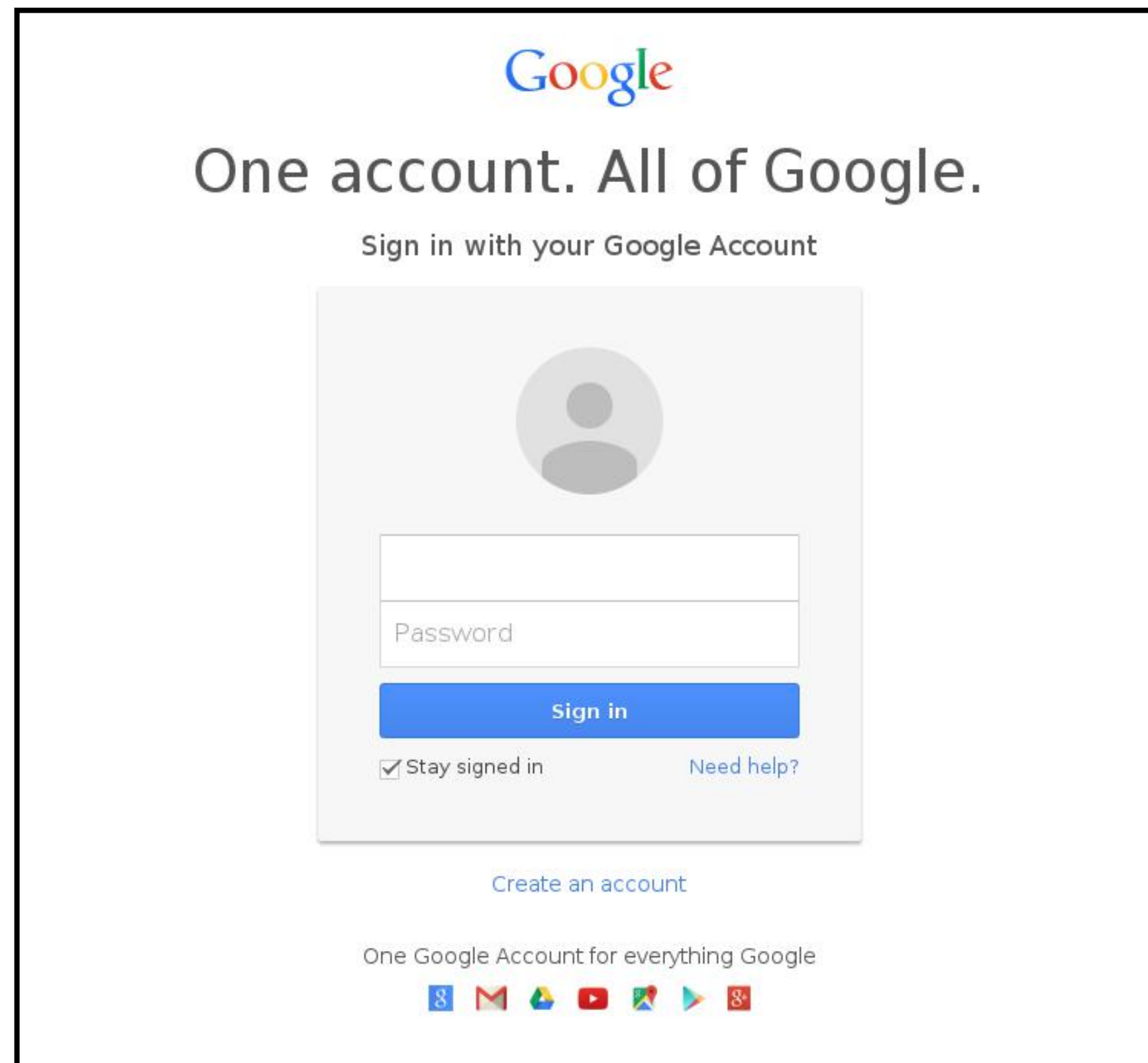


Joshua Reynolds

Not pictured: Martin Shelton, Emily Stark, Kaishen Wang, Joseph Dickinson, Rohan Subramanian, Meishan Wu

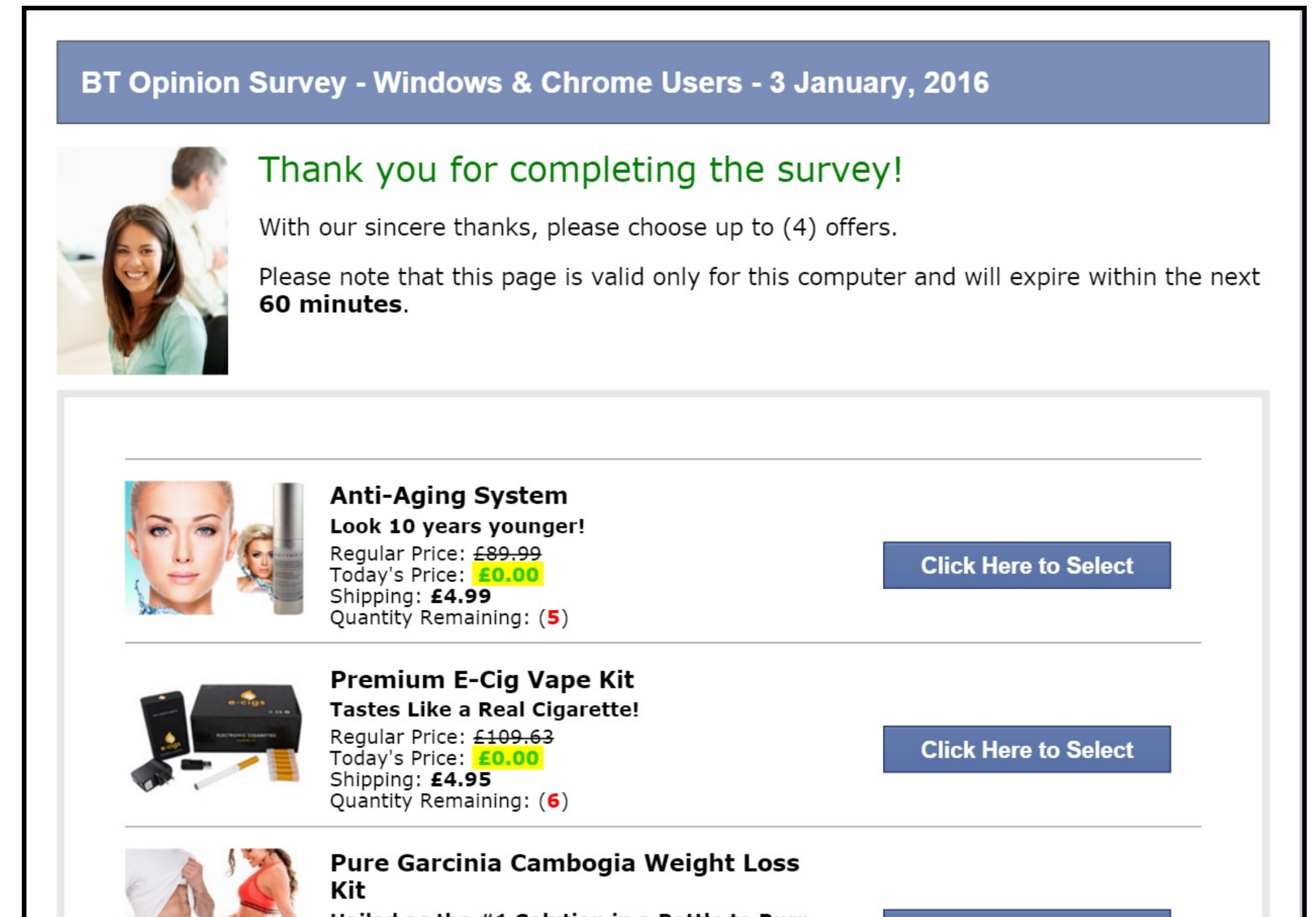
Social Engineering: Root Causes

Mistaken Identity



accounts-google.com

Misplaced Trust



questionsaboutisps.com

Social Engineering: Root Causes

Mistaken Identity

Misplaced Trust

Measuring Identity Confusion with Uniform Resource Locators

Joshua Reynolds[†] Deepak Kumar[†] Zane Ma[†] Rohan Subramanian[†] Meishan Wu[†]
Martin Shelton[‡] Joshua Mason[†] Emily Stark[‡] Michael Bailey[†]
[†]University of Illinois at Urbana-Champaign [‡]Google, Inc.
{joshuar3, dkumar11, zanema2, rsubra2, meishan2, joshm, mdb Bailey}@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, mdb Bailey}@illinois.edu
University of Illinois Urbana-Champaign

CSET 2019

URL complexity leads
to mistaken identity

Users may (mis)place
trust in HTTPS