Web Security Origins and Evolution

Mary Ellen Zurko

mez@ll.mit.edu

May 5, 2021

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.© 2018 Massachusetts Institute of Technology. Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

Web Security Evolution Agenda

- The First Web Security Feature
- Protecting Web Pages
- Web Security User Interface
- Open Standards and Web User Security
- Mixing Code with Data
- Open Source and Security Vulnerabilities
- Web Attacks On Humans

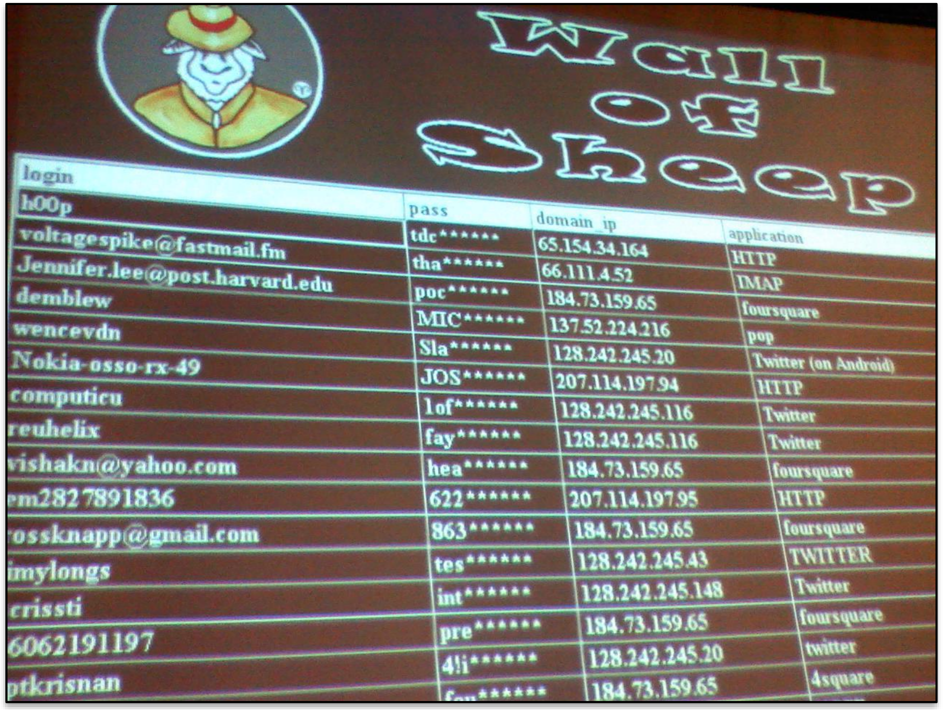
Web Security Evolution Agenda

→

- The First Web Security Feature
- Protecting Web Pages
- Web Security User Interface
- Open Standards and Web User Security
- Mixing Code with Data
- Open Source and Security Vulnerabilities
- Web Attacks On Humans

Web Security In the Beginning (1992)

- TimBL has a vision of the read/write web
  - It begins as a read-only experience for users
  - Web pages are static data
- Only one web security feature is in TimBL's 1992 WWW proposal
- Basic Authentication
  - Password is Base64 encoded
  - Every URL DNS domain (+ realm) does their own authentication
  - Who's asking you for your password?

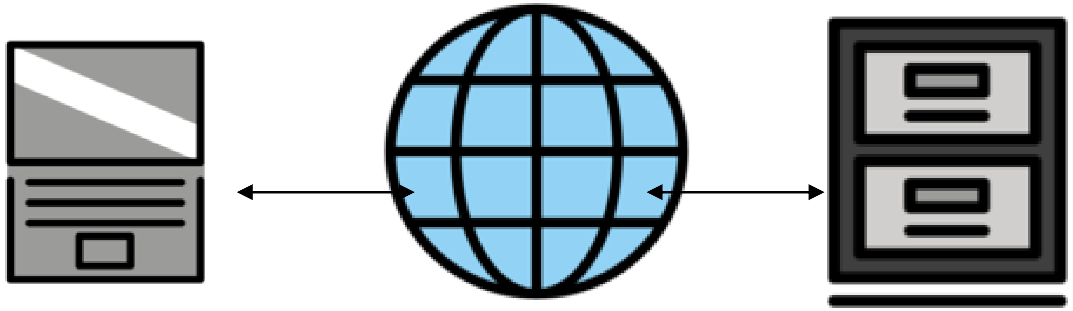


login	pass	domain_ip	application
h00p	tdc******	65.154.34.164	HTTP
voltagespike@fastmail.fm	tha******	66.111.4.52	IMAP
Jennifer.lee@post.harvard.edu	poc******	184.73.159.65	foursquare
demblew	MIC******	137.52.224.216	pop
wencevdn	Sla******	128.242.245.20	Twitter (on Android)
Nokia-osso-rx-49	JOS******	207.114.197.94	HTTP
computicu	1of******	128.242.245.116	Twitter
reuhelix	fay******	128.242.245.116	Twitter
vishakn@yahoo.com	hea******	184.73.159.65	foursquare
em2827891836	622******	207.114.197.95	HTTP
rossknapp@gmail.com	863******	184.73.159.65	foursquare
imylongs	tes******	128.242.245.43	TWITTER
crissti	int******	128.242.245.148	Twitter
6062191197	pre******	184.73.159.65	foursquare
ptkrisnan	4!i******	128.242.245.20	twitter
	fou******	184.73.159.65	4square

Future attacks will be the unanticipated ones, particularly if you're successful

Digest Authentication
Encrypt All The Passwords (1994)

- Digest Authentication Features
  - Cryptographically hash the password
  - Defense against Rainbow Tables
  - Nonces in the server challenge for replay protection
- Deployment Challenges
  - The protocol for negotiating mutual support allows a Man in the Middle to spoof lack of support
  - Three tier architectures need to pass the password
  - No attacks in the wild, no high value web site interactions



Deployment means interoperability and co-existence with systems without the new security feature

Web Security Evolution Agenda

- The First Web Security Feature
- → Protecting Web Pages
- Web Security User Interface
- Open Standards and Web User Security
- Mixing Code with Data
- Open Source and Security Vulnerabilities
- Web Attacks On Humans

How Did We First Encrypt Web Pages?

- Secure HyperText Transfer Protocol - S-HTTP:
- Flexible framework for encryption of the HTML document
  - Page data and submitted data – not the headers
  - The specific URL moved into encrypted portion
- Headers defined to specify type of encryption and algorithm, type of key management
  - Supports pre arranged keys, public/private keys, PGP, etc.
  - Server and client negotiate which enhancements they'll use
- Digital signature option
  - Another form of authentication
- End to end
  - Clients can initiate the encrypted request
  - Resists Man in the Middle

Why Didn’t S-HTTP Take Over The World?

- End to end protection requires client side deployment of secrets
  - Scale of client deployment was much larger than server deployment
- End user had to interact with secrets at the scale of web pages
- Flexible framework meant (too) many choices for deployment
  - Which type of secrets do which users have?
  - Which type of secrets do which web pages require?

Flexibility without use cases leaves questions for someone else to answer

SSL/TLS – HTTPS:

- Encryption, authentication, and security since 1994
- SSL was an open standard with three versions
  - TLS v1.0 superseded it in 1999
- Authentication of the server using public key certificate
- Authentication of the client using public key certificate is an option
- The encryption for network confidentiality part works pretty darn well
  - Except when in the face of attacks and errors…

Certificate Authority Attacks

- 12 CA incidents in 2011
  - Attack on Comodo stole username/password of a Registration Authority
    - 9 fraudulent certificates issued, including login.yahoo.com, mail.google.com, login.skype.com, addons.mozilla.org
    - Certificate revoked upon discovery
  - DigiNotar was attacked and fraudulent certificates issued
  - KPN discovered attack tools on its server during an audit and stopped issuing certificates
    - DDoS tool there for as long as 4 years
- Certificate transparency allows domain owners to see CA issued certificates for their domain

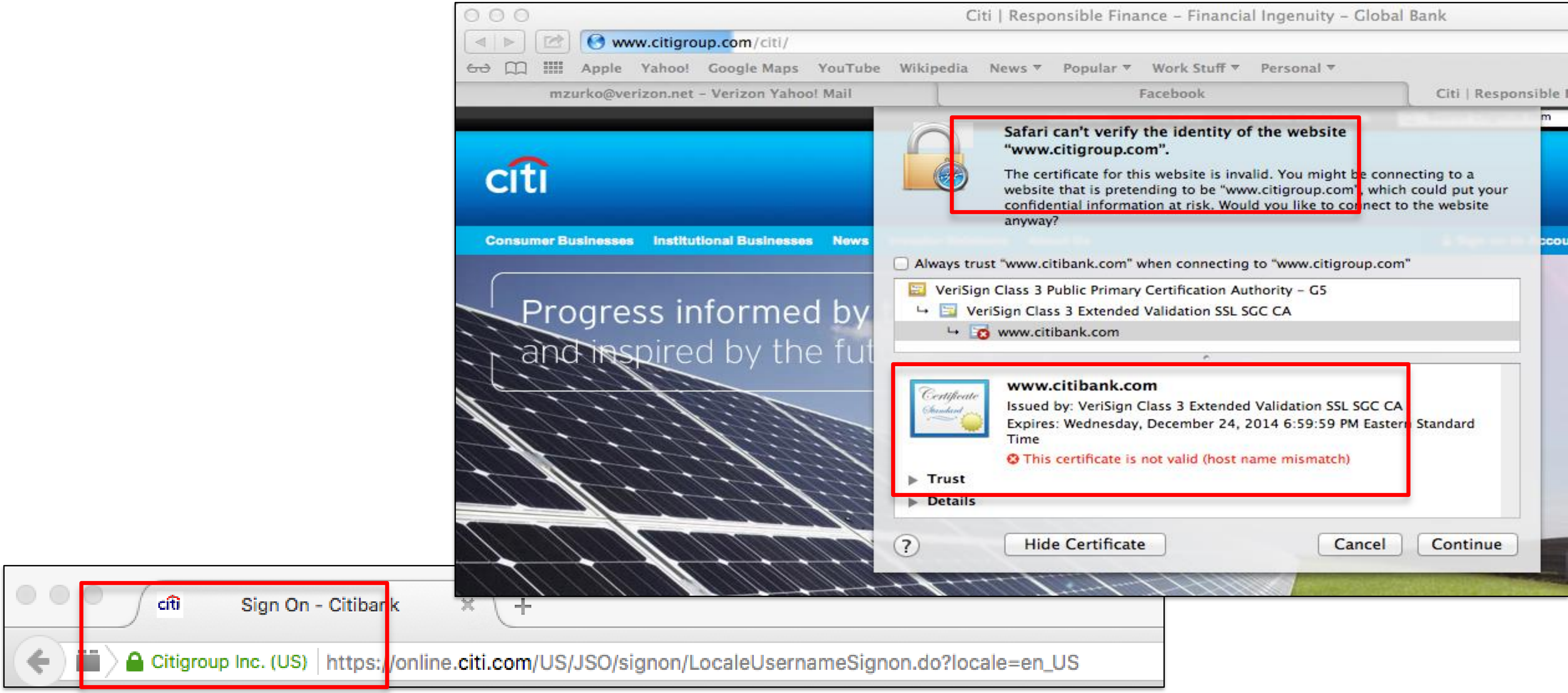
More potential attack targets means more and more-varied attacks

Error Handling in TLS web site authentication

- No one seemed to think asking the users was a problem at protocol design time
- What does it really mean if a server has a self signed certificate?
  - CA issued certificates cost money; economic effects were not considered
  - Users learned to ignore warnings
- Crying Wolf: An Empirical Study of SSL Warning Effectiveness
  - 2009 study using FF2 as a baseline for clickthrough
  - 90% ignore rate in their in-lab user study of a banking scenario
- ImperialViolet documented a 60% rate of bypassing SSL interstitials in 2012
- WWW2013 paper documented a high false positive rate
  - 1.54% false positive warning rate on 3.9 billion TLS connections across 300k academic users

The user is not an exception handling module

Are warnings about domains from HTTPS meaningful?



Citi | Responsible Finance – Financial Ingenuity – Global Bank

www.citigroup.com/citi/

Safari can't verify the identity of the website "www.citigroup.com".

The certificate for this website is invalid. You might be connecting to a website that is pretending to be "www.citigroup.com", which could put your confidential information at risk. Would you like to connect to the website anyway?

☐ Always trust "www.citibank.com" when connecting to "www.citigroup.com"

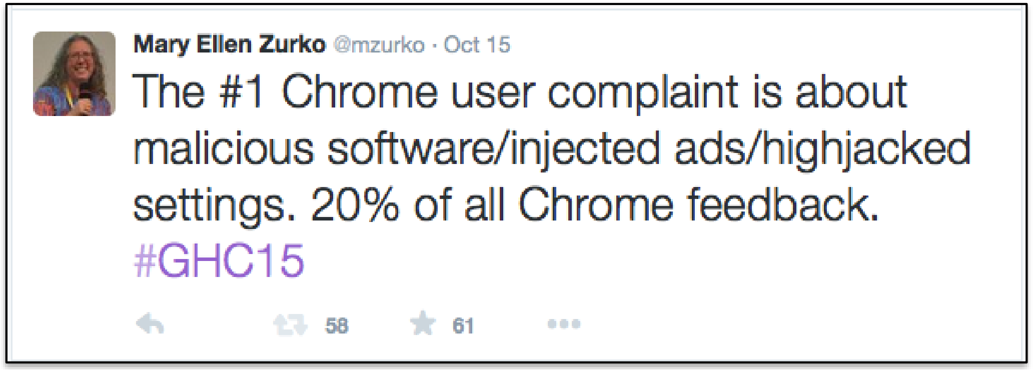
```
VeriSign Class 3 Public Primary Certification Authority – G5
  ↳ VeriSign Class 3 Extended Validation SSL SGC CA
      ↳ www.citibank.com
```

www.citibank.com
Issued by: VeriSign Class 3 Extended Validation SSL SGC CA
Expires: Wednesday, December 24, 2014 6:59:59 PM Eastern Standard Time
This certificate is not valid (host name mismatch)

Citigroup Inc. (US) https://online.citi.com/US/JSO/signon/LocaleUsernameSignon.do?locale=en_US

User Experience and Malware Warnings

- Firefox Click Through Rate (CTR) for malware warnings is 33% (2014)
  - Google Chrome's 70%
- Mock Firefox styling closed that difference by 12 to 20 points in a 10 day at scale controlled experiment
  - Change to text, layout, default button
- Users heed warnings to sites they have not visited
  - Users unpredictable for warnings on sites they have visited
  - Survey said users trust high reputation sites more than malware warnings
- Further change promoted the safe choice and demoted the unsafe choice (2015)
  - Chrome CTR 38%


Mary Ellen Zurko @mzurko · Oct 15

The #1 Chrome user complaint is about malicious software/injected ads/highjacked settings. 20% of all Chrome feedback. #GHC15

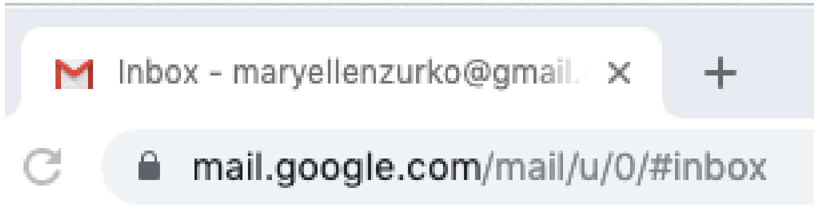
In theory, there is no difference between theory and practice.
In practice, there is. - Yogi Berra

Web Security Evolution Agenda

- The First Web Security Feature
- Protecting Web Pages
- Web Security User Interface
- Open Standards and Web User Security
- Mixing Code with Data
- Open Source and Security Vulnerabilities
- Web Attacks On Humans

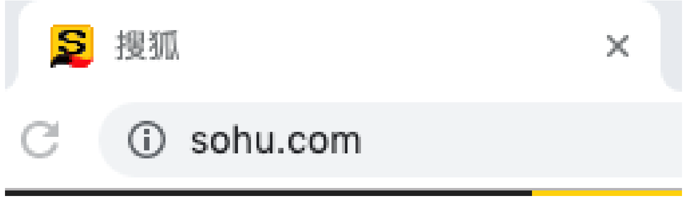
Authenticating the Server to the Human

- TLS provides authentication of the server using its public key certificate
- Can you explain each of these four different types of web server authentication from Chrome in 2019?



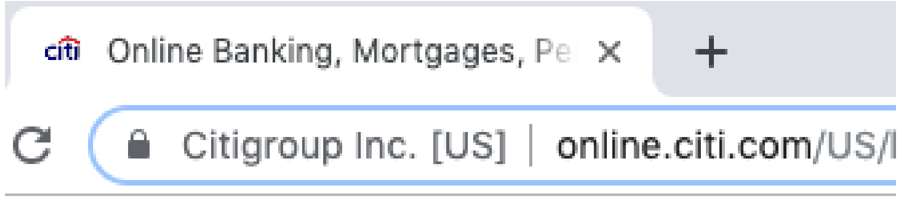
Inbox - maryellenzurko@gmail.

mail.google.com/mail/u/0/#inbox



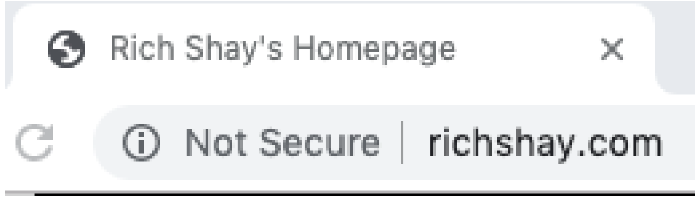
搜狐

sohu.com



Online Banking, Mortgages, Pe

Citigroup Inc. [US] | online.citi.com/US/l

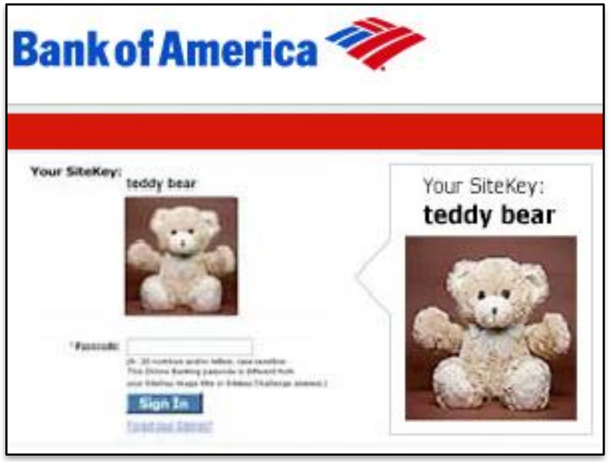


Rich Shay's Homepage

Not Secure | richshay.com

What do users do when web site authentication fails?

- The Emperor’s New Security Indicators (2007)
- Lab study of bank customers (67)
  - 3 groups; as self, role playing + not primed, role playing + security primed
- Removed HTTPS indicators
  - “https” in address bar and lock icon in bottom right
  - 0 withheld password
- Removed the customer selected site-authentication image
  - Replaced it with a bank upgrade maintenance notice
  - 23 of 25 using their own accounts entered their password
  - All 36 role playing entered their password
- Role playing participants behaved statistically significantly less securely
  - Even the group that was security primed



Humans won’t do what technologists assume they will

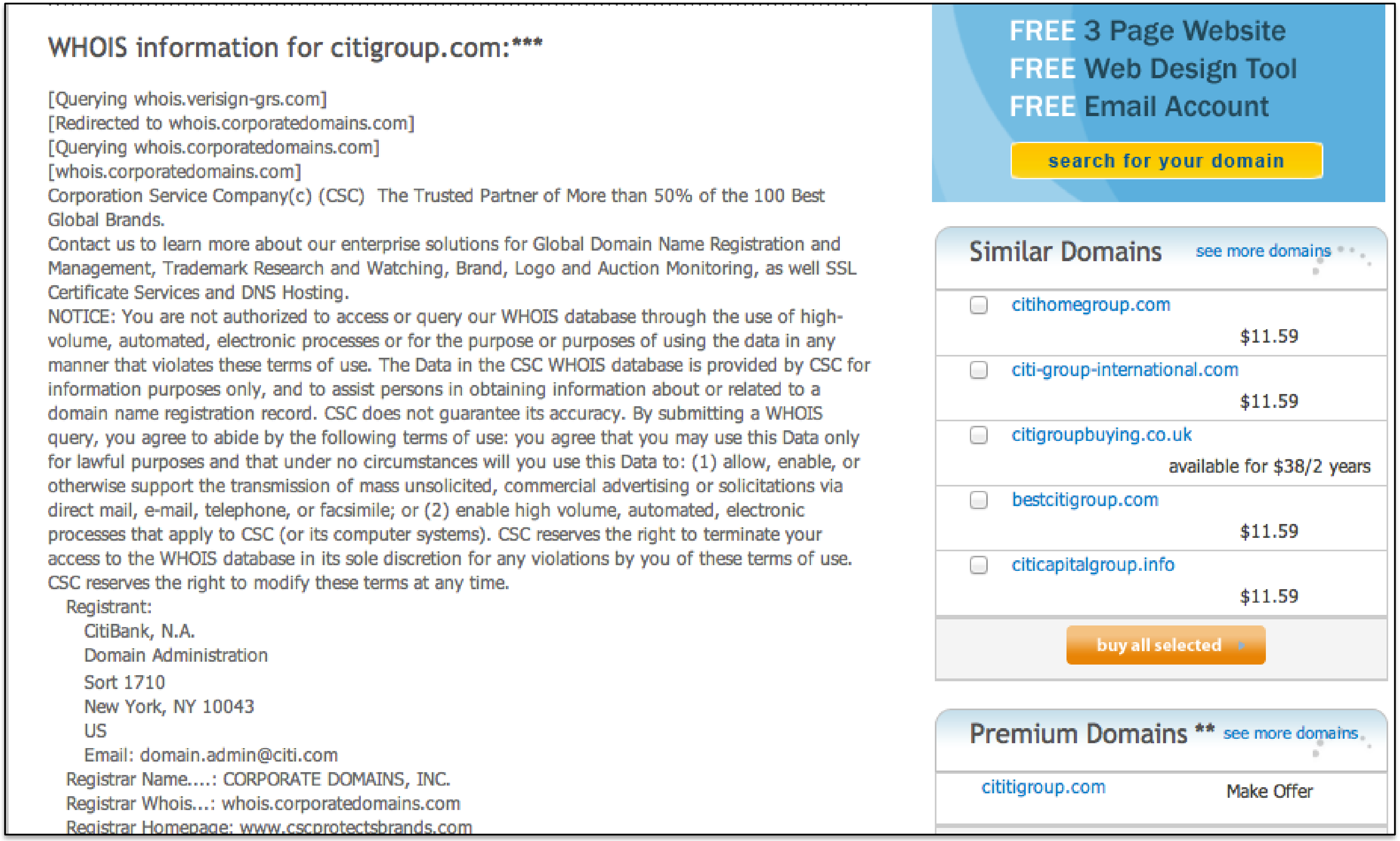
It’s 2012: Which of these domains are not owned by Citibank?

- Citigroup.com
- Citibank.com
- Cititigroup.com
- Citigroup.de
- Citibank.co.uk
- Citigroup.org
- Thisiscitigroup.org

- Citibank.info
- Citicards.com
- Citicreditcards.com
- Citibank-cards.us
- Citimoney.com
- Citigold.net
- Citigrøup.org

It’s 2012: Which of these domains are not owned by Citibank?

- Cititigroup.com
- Thisiscitigroup.org
- Citimoney.com
- Citigrøup.org



WHOIS information for citigroup.com:***

[Querying whois.verisign-grs.com]
[Redirected to whois.corporatedomains.com]
[Querying whois.corporatedomains.com]
[whois.corporatedomains.com]
Corporation Service Company(c) (CSC) The Trusted Partner of More than 50% of the 100 Best Global Brands.
Contact us to learn more about our enterprise solutions for Global Domain Name Registration and Management, Trademark Research and Watching, Brand, Logo and Auction Monitoring, as well SSL Certificate Services and DNS Hosting.
NOTICE: You are not authorized to access or query our WHOIS database through the use of high-volume, automated, electronic processes or for the purpose or purposes of using the data in any manner that violates these terms of use. The Data in the CSC WHOIS database is provided by CSC for information purposes only, and to assist persons in obtaining information about or related to a domain name registration record. CSC does not guarantee its accuracy. By submitting a WHOIS query, you agree to abide by the following terms of use: you agree that you may use this Data only for lawful purposes and that under no circumstances will you use this Data to: (1) allow, enable, or otherwise support the transmission of mass unsolicited, commercial advertising or solicitations via direct mail, e-mail, telephone, or facsimile; or (2) enable high volume, automated, electronic processes that apply to CSC (or its computer systems). CSC reserves the right to terminate your access to the WHOIS database in its sole discretion for any violations by you of these terms of use. CSC reserves the right to modify these terms at any time.

```
  Registrant:
    CitiBank, N.A.
    Domain Administration
    Sort 1710
    New York, NY 10043
    US
    Email: domain.admin@citi.com
  Registrar Name....: CORPORATE DOMAINS, INC.
  Registrar Whois...: whois.corporatedomains.com
  Registrar Homepage: www.cscprotectsbrands.com
```

FREE 3 Page Website
FREE Web Design Tool
FREE Email Account

search for your domain

Similar Domains see more domains

□ citihomegroup.com
$11.59

□ citi-group-international.com
$11.59

□ citigroupbuying.co.uk
available for $38/2 years

□ bestcitigroup.com
$11.59

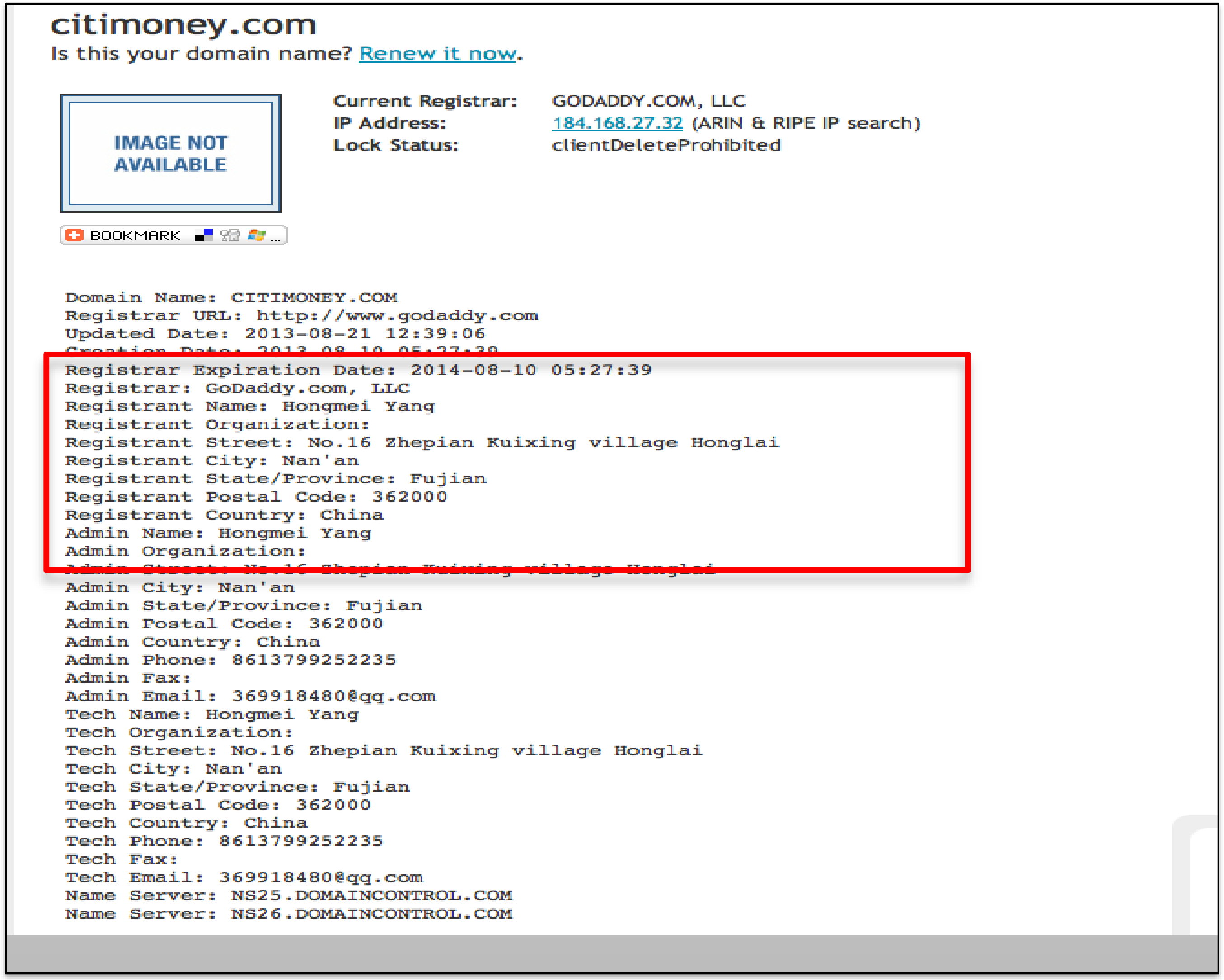
□ citicapitalgroup.info
$11.59

buy all selected

Premium Domains ** see more domains

cititigroup.com Make Offer

Who else thought citimoney.com was an excellent domain name in 2013?



citimoney.com

Is this your domain name? Renew it now.

Current Registrar:	GODADDY.COM, LLC
IP Address:	184.168.27.32 (ARIN & RIPE IP search)
Lock Status:	clientDeleteProhibited

```
Domain Name: CITIMONEY.COM
Registrar URL: http://www.godaddy.com
Updated Date: 2013-08-21 12:39:06
Creation Date: 2013-08-10 05:27:39
Registrar Expiration Date: 2014-08-10 05:27:39
Registrar: GoDaddy.com, LLC
Registrant Name: Hongmei Yang
Registrant Organization:
Registrant Street: No.16 Zhepian Kuixing village Honglai
Registrant City: Nan'an
Registrant State/Province: Fujian
Registrant Postal Code: 362000
Registrant Country: China
Admin Name: Hongmei Yang
Admin Organization:
Admin Street: No.16 Zhepian Kuixing village Honglai
Admin City: Nan'an
Admin State/Province: Fujian
Admin Postal Code: 362000
Admin Country: China
Admin Phone: 8613799252235
Admin Fax:
Admin Email: 369918480@qq.com
Tech Name: Hongmei Yang
Tech Organization:
Tech Street: No.16 Zhepian Kuixing village Honglai
Tech City: Nan'an
Tech State/Province: Fujian
Tech Postal Code: 362000
Tech Country: China
Tech Phone: 8613799252235
Tech Fax:
Tech Email: 369918480@qq.com
Name Server: NS25.DOMAINCONTROL.COM
Name Server: NS26.DOMAINCONTROL.COM
```

Web Security Evolution Agenda

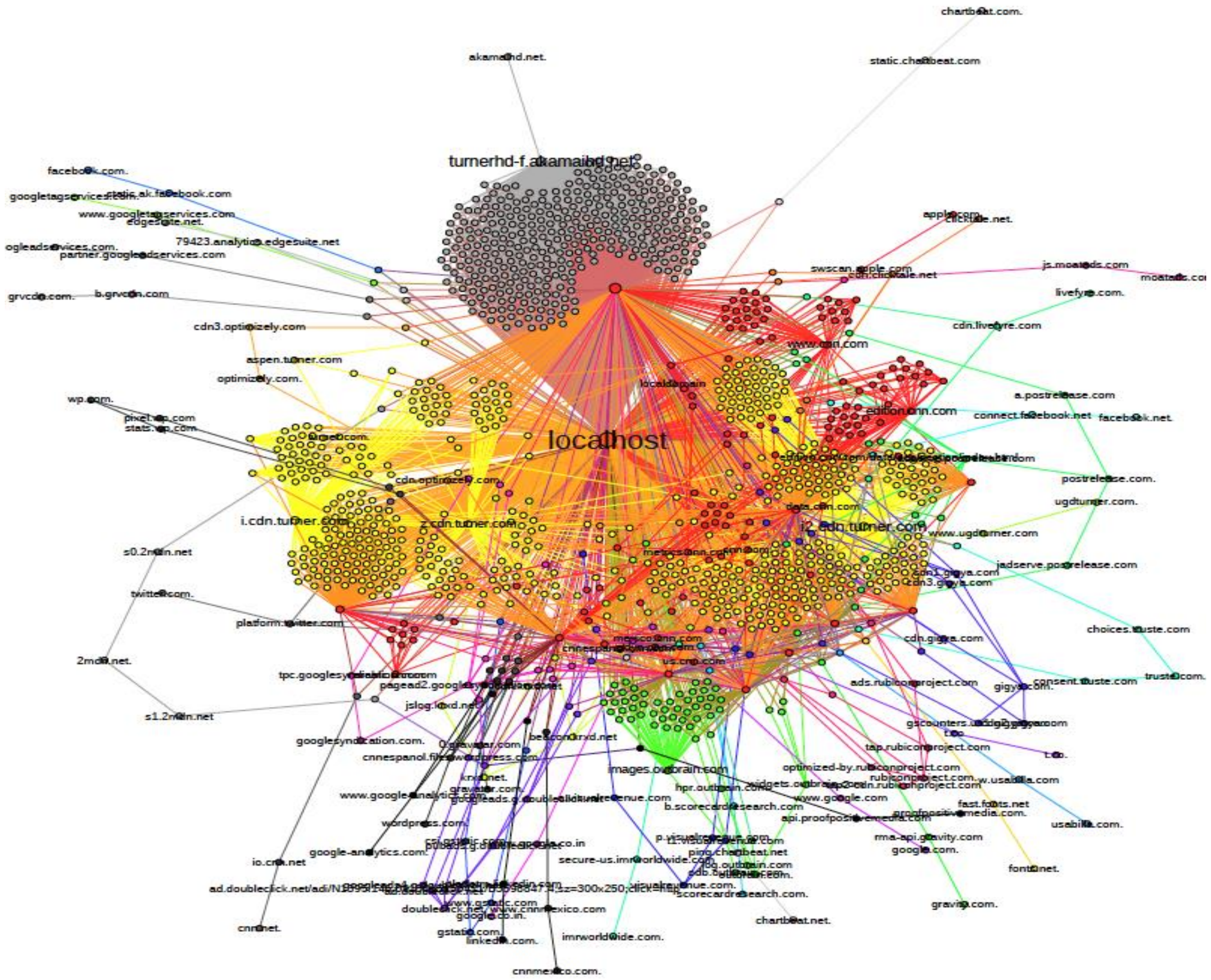
- The First Web Security Feature
- Protecting Web Pages
- Web Security User Interface
- ➡ Open Standards and Web User Security
- Mixing Code with Data
- Open Source and Security Vulnerabilities
- Web Attacks On Humans

W3C Web Security Context (WSC)

- First usable security standard
- Charter: To enable users to come to a better understanding of the context that they are operating in when making trust decisions on the Web
  - Specify a baseline set of security context information and practices for the secure and usable presentation of this information
- Functional areas: TLS encryption, Domain name (authenticated or claimed), Certificate information, Browsing history, Errors
- Principles: Visibility, assurance, attention

Would a standard security user experience make web security more usable?

WSC Tried to Make This Understandable



WSC Recommendations

- Certificate Trust validation
  - Extended Validation, self-signed, and untrusted, and user interactions around validation
- Existence of encryption
- Strong cipher suites
- User interactions for error handling based on error severity
  - Attempting to combat habituation
- Consistent visual presentation of authenticated DNS identity
- MUST NOTs – mixed content, obscuring security info, techno jargon, unsupervised installation, automatic bookmarks

WSC Challenges

- “Successful standards enable”
  - We had a lot of “Don’t do this thing” and constraints
- UI standards are process, not presentation
- Some of the reasons browser vendors participated in standards
  - Interoperability (as required by/for the market)
  - Customer requirements (compliance and laws and features)
- Some of the reasons browser vendors didn’t participate in standards
  - IP/patents
  - Dilution of their brand
  - Market advantage in the area
- And then came mobile apps - technology marched forward

Open standards haven’t worked for security user experience

Web Security Evolution Agenda

- The First Web Security Feature
- Protecting Web Pages
- Web Security User Interface
- Open Standards and Web User Security
- Mixing Code with Data
- Open Source and Security Vulnerabilities
- Web Attacks On Humans

Code Comes to Web Pages

Is it safe?

- In 1997, Dynamic HTML introduced HTML tags that contain code
  - Postscript format for printing had previously crossed this boundary
- Who vouches for the code on this web site?
  - Javascript used the sandbox + same origin policy
- Web mail was the earliest web application serving data in pages not created by web site developers
  - It broke domain name authentication assumptions and gave rise to cross site scripting (XSS)
- Response - HTML escaping of everything
  - Where are my bold text and dancing pigs?
- Next steps: Whitelist vs Blacklist of HTML tags
  - What are the tradeoffs?

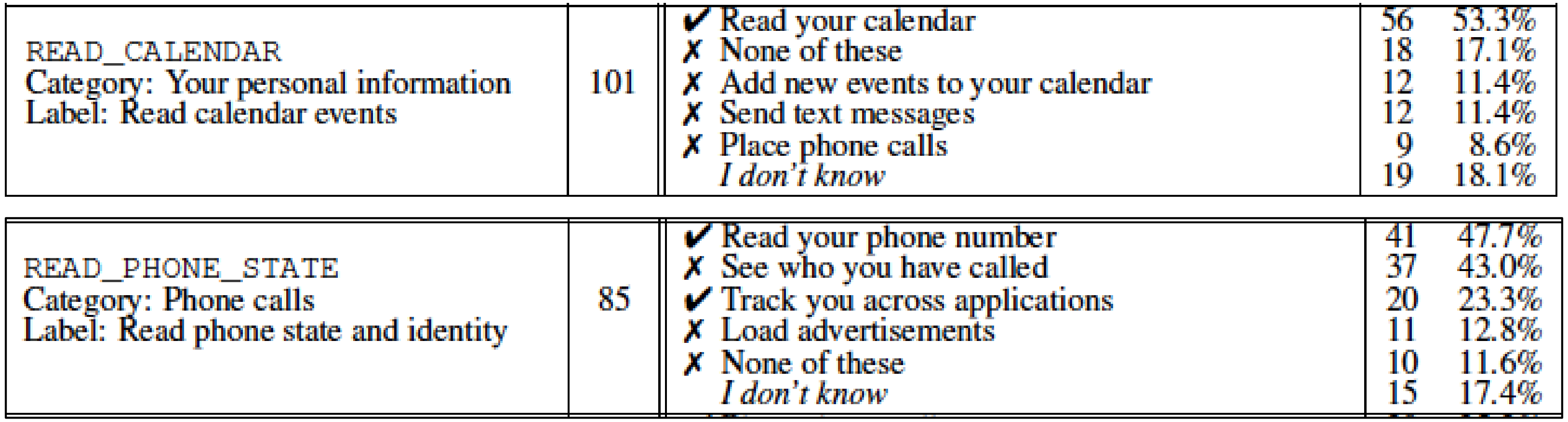
In security, there is a large difference between data and code

Active Content Security Challenge

- Browsers enabled many ways for “code” to execute on your device
- With web applications, GET stopped being safe and idempotent
  - Which gave us CSRF
  - JSON and XML enable CSRF with POST
- Browsers could be used to directly download code
- Browser extensions were a new type of code
- Web based updates/patches were not automatic, because they were code
- Mobile applications allowed anyone to write code for you to download
  - Introduced in 2007 on Apple iPhone iOS
  - Controls included a permissions model

User Experience Installing Code with Android Permissions (2012)

- 308 participants in the Internet study, 25 in the lab
- 17% of participants paid attention to permissions during installation (self reported and lab experiment)
  - 42% aware permissions exist but do not always consider them
- 3% of survey respondents could answer correctly and exactly all three randomly chosen permission comprehension questions
  - 53% of the answers contain at least one correct choice
- READ_CALENDAR
  - 46% correct
- READ_PHONE_STATE
  - 4.7% correct



READ_CALENDAR Category: Your personal information Label: Read calendar events	101	✔ Read your calendar	56	53.3%
		✗ None of these	18	17.1%
		✗ Add new events to your calendar	12	11.4%
		✗ Send text messages	12	11.4%
		✗ Place phone calls	9	8.6%
		I don't know	19	18.1%

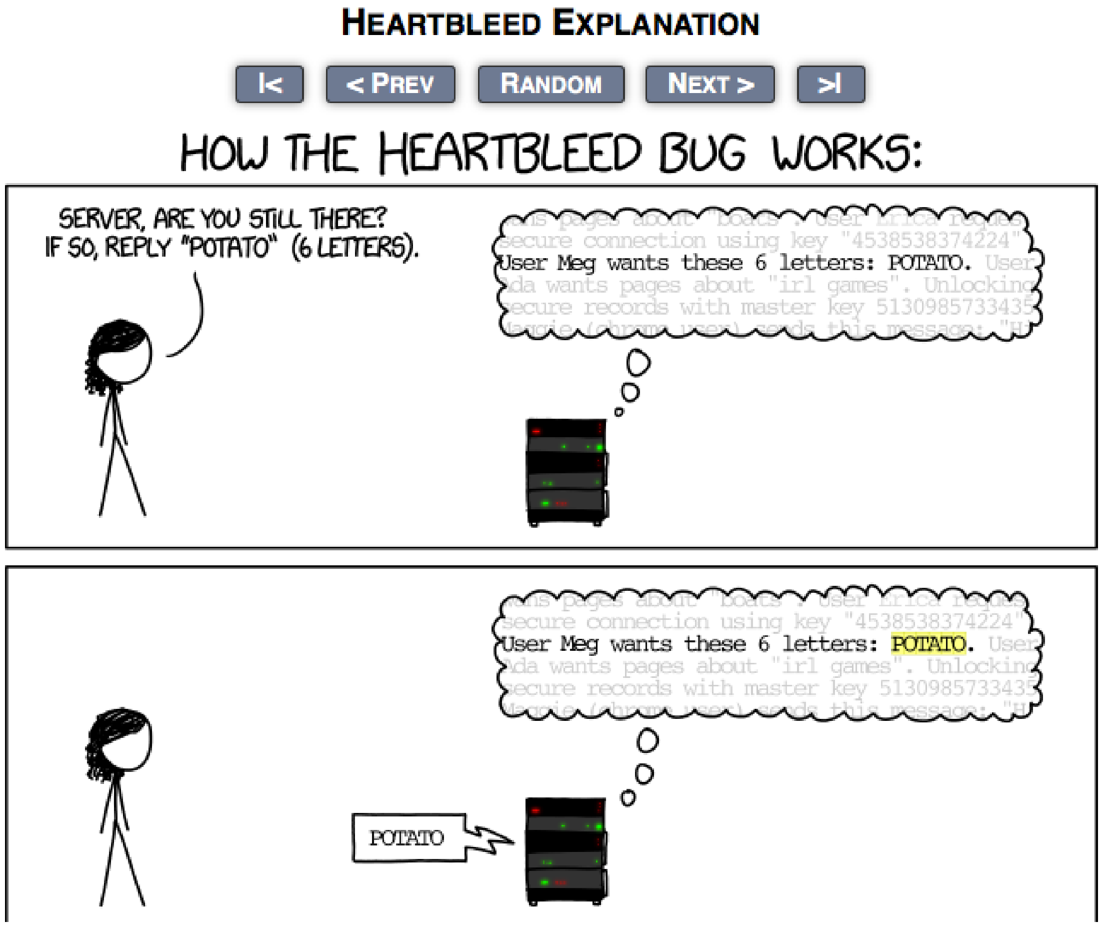
READ_PHONE_STATE Category: Phone calls Label: Read phone state and identity	85	✔ Read your phone number	41	47.7%
		✗ See who you have called	37	43.0%
		✔ Track you across applications	20	23.3%
		✗ Load advertisements	11	12.8%
		✗ None of these	10	11.6%
		I don't know	15	17.4%

Web Security Evolution Agenda

- The First Web Security Feature
- Protecting Web Pages
- Web Security User Interface
- Open Standards and Web User Security
- Mixing Code with Data
- ➡ Open Source and Security Vulnerabilities
- Web Attacks On Humans

Heartbleed Vulnerability

- ”Given enough eyeballs, all bugs are shallow”
- Heartbeat standard is an extension to TLS standard
  - Keep Alive performance enhancement
  - TCP has its own keep alive
- Heartbleed vulnerability was discovered in 2014
  - The code was committed in 2011
- Improper input validation due to a missing bounds check
  - C language – specify string sizes
  - Network protocols
  - Common source of error for programmers (aka humans)


HEARTBLEED EXPLANATION
|<
< PREV
RANDOM
NEXT >
>|
HOW THE HEARTBLEED BUG WORKS:
SERVER, ARE YOU STILL THERE? IF SO, REPLY "POTATO" (6 LETTERS).
ants pages about "boats". User Erica reques
secure connection using key "4538538374224"
User Meg wants these 6 letters: POTATO. User
Ada wants pages about "irl games". Unlocking
secure records with master key 5130985733435
Maggie (chrome user) sends this message: "H
ants pages about "boats". User Erica reques
secure connection using key "4538538374224"
User Meg wants these 6 letters: POTATO. User
Ada wants pages about "irl games". Unlocking
secure records with master key 5130985733435
Maggie (chrome user) sends this message: "H
POTATO

Developers are human, and any mistake might be a vulnerability

What does Heartbleed tell us about Open Source Security?

- OpenSSL was a popular cryptographic library
  - SSL/TLS widely used to secure a variety of communications
  - Over 66% of the Internet deployed OpenSSL
  - 17% of secured web servers (.5 million) were believed to be vulnerable
  - Full recovery would mean changing anything secret that could have been in memory while the vulnerable version was deployed
- Open source security largely relied on the many eyes involved in development, deployment, and use
  - Process for commits – was reviewed by one of the four core developers
  - Security testing did not seem to be part of the development process
    - One of the teams that found this was Codenomicon, developing fuzz tests for the Heartbeat protocol
  - A code audit by a deployer was the other way it was found

Response to Heartbleed: Core Infrastructure Initiative

- Member companies provide money and advice
- Risk score of Open Source projects to focus funding
- Planned and potential activities included some closed source best practices
  - Compensating full time developers
  - Deploying test infrastructure
    - Fuzzing, positive/negative test suites, static checking
  - Developer education on security best practices
  - Reproducible builds
  - Security audits
  - Badging program for best practices in open source security
- What did research have to say about these at the time?

Clubbing Seals: Exploring the Ecosystem of Third-party Security Seals

- Do sites with seals have better security than sites without?
  - Statistically significant difference for 3 of 9 passively discoverable security mechanisms, 2 to 1 in favor of web sites without seals
- Are sites with seals clean from basic and well known vulnerabilities?
  - Stood up a website with 12 vulnerabilities with 8 security seal providers
  - Seal providers found from 0 to 5 of the vulnerabilities
  - 3 automated scanning tools found from 5 to 6 of the vulnerabilities
    - Automated scanners can tolerate more false positives, leading to more true positives
- At least security seals do not decrease the security of websites?
  - Transition from visible to invisible, plus site's status on the seal provider, form an indicator of a known vulnerability on a web site
  - 2 months of monitoring 8k websites showed 333 seal transitions
  - Attacker who can purchase a seal and craft their website can capture likely seal scanning information for replay or analysis to identify potential vulnerabilities
- Seals can be visually spoofed or directly included with a simple ruse

Web Security Evolution Agenda

- The First Web Security Feature
- Protecting Web Pages
- Web Security User Interface
- Open Standards and Web User Security
- Mixing Code with Data
- Open Source and Security Vulnerabilities
- Web Attacks On Humans

Attacks on Humans that Use the Web

- Fraudulent e-commerce sites joined the real ones (~1998)
  - How about that TLS server authentication?
- Phishing for credit cards, then credentials (~2004)
  - First research paper on the potential efficacy of targeted phishing (2005)
- Fraudulent tech support scams
- Misinformation, Disinformation, and Influence Operations

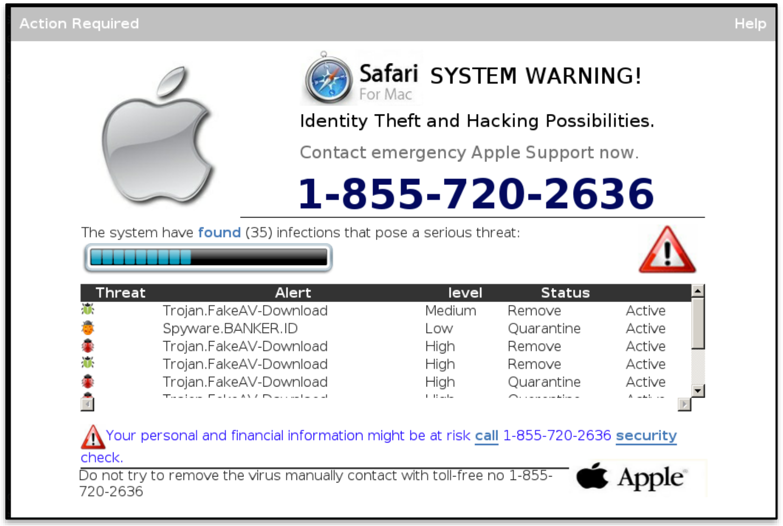
Technology turns old attacks into new attacks

Anatomy of a Tech Support Scam

- Fraudulent tech support scams
  - Charge for the “service” of removing (nonexistent) malware
  - Sometimes also spread malware
  - $1.5 billion industry in first 10 months of 2015
- Contact starts with cold calls, or with pop ups or web sites claiming the user has malware and should call the fake tech support
- Talos security researchers called one to understand their methods and infrastructure
  - Set up a virtual machine
  - Recorded the interactions
  - Identified individuals on LinkedIn associated with the web sites and finances of the tech support scam company

Step 1: Get connected over the web

- Called the phone number, and talked to “Kelly Thompson”
- “Are you using a phone?” as the device that needs cleansing
  - Confirmed their computer was a Toshiba, not a Macbook
  - Kelly asserted she could still take care of the issue
- Instructed to follow a (shortened) URL
  - The URL loaded TeamViewer which provides remote control of a computer
    - Which has a built in warning about exactly this sort of thing
  - Promptly instructed by Kelly to ignore the warning
    - “Tap on Trustworthy”



SYSTEM WARNING!

Identity Theft and Hacking Possibilities.

Contact emergency Apple Support now.

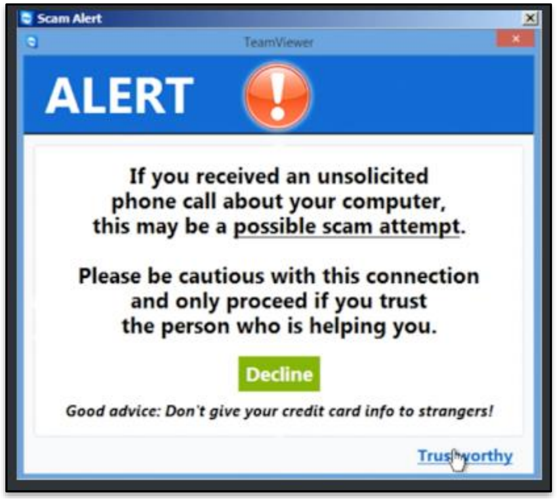
1-855-720-2636

The system have found (35) infections that pose a serious threat:

Threat	Alert	level	Status	
	Trojan.FakeAV-Download	Medium	Remove	Active
	Spyware.BANKER.ID	Low	Quarantine	Active
	Trojan.FakeAV-Download	High	Remove	Active
	Trojan.FakeAV-Download	High	Remove	Active
	Trojan.FakeAV-Download	High	Quarantine	Active

Your personal and financial information might be at risk call 1-855-720-2636 security check.

Do not try to remove the virus manually contact with toll-free no 1-855-720-2636



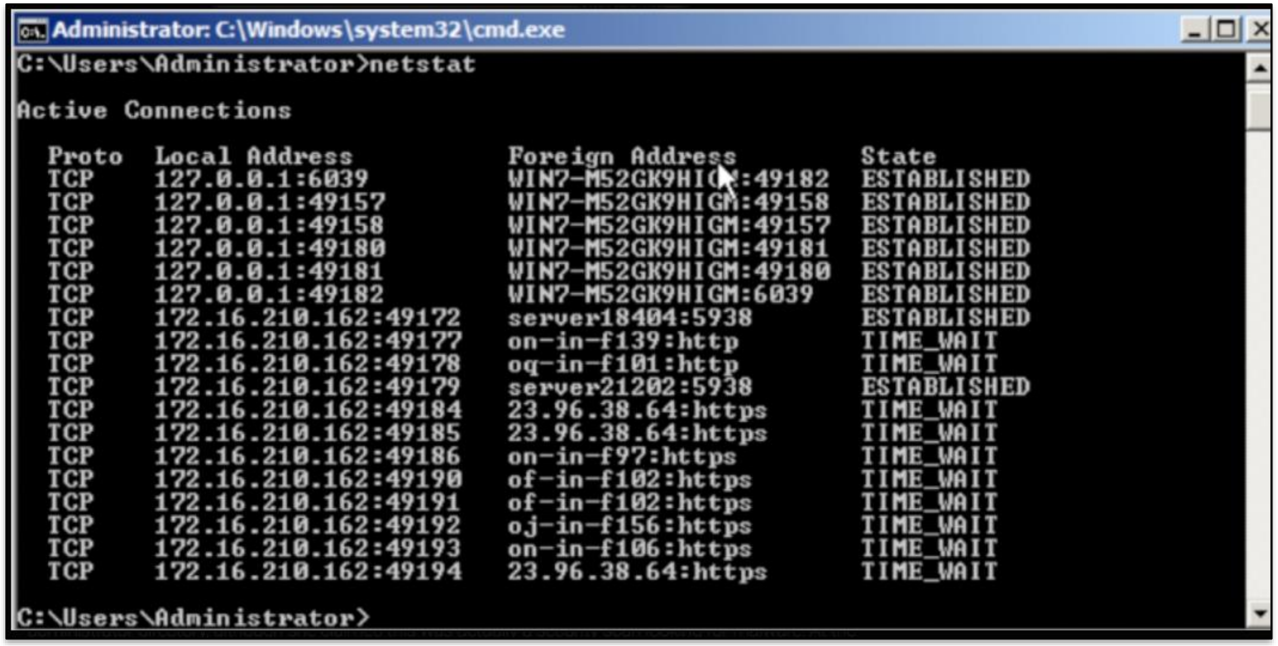
If you received an unsolicited phone call about your computer, this may be a possible scam attempt.

Please be cautious with this connection and only proceed if you trust the person who is helping you.

Good advice: Don't give your credit card info to strangers!

Step 2: Hackers are infiltrating your computer

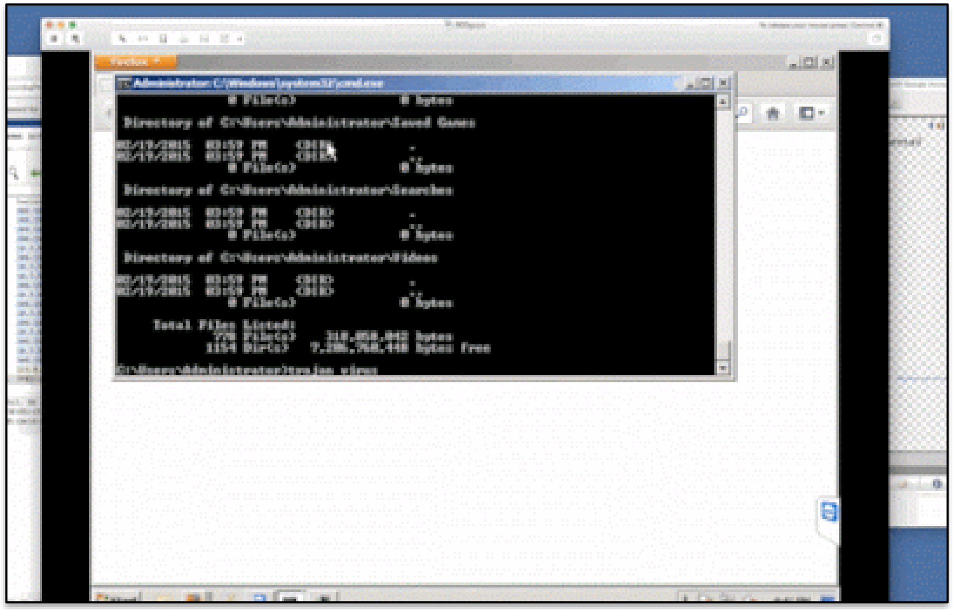
- **Kelly now has remote access**
- **Displayed a variety of harmless processes as evidence of malicious activities**
  - **Netstat shows network connections with “foreign addresses”**
  - **These are hackers infiltrating your computer from another country!**



```
C:\Users\Administrator>netstat

Active Connections

  Proto  Local Address          Foreign Address        State
  TCP    127.0.0.1:6039         WIN7-M52GK9HIGM:49182  ESTABLISHED
  TCP    127.0.0.1:49157        WIN7-M52GK9HIGM:49158  ESTABLISHED
  TCP    127.0.0.1:49158        WIN7-M52GK9HIGM:49157  ESTABLISHED
  TCP    127.0.0.1:49180        WIN7-M52GK9HIGM:49181  ESTABLISHED
  TCP    127.0.0.1:49181        WIN7-M52GK9HIGM:49180  ESTABLISHED
  TCP    127.0.0.1:49182        WIN7-M52GK9HIGM:6039   ESTABLISHED
  TCP    172.16.210.162:49172   server18404:5938       ESTABLISHED
  TCP    172.16.210.162:49177   on-in-f139:http        TIME_WAIT
  TCP    172.16.210.162:49178   oq-in-f101:http        TIME_WAIT
  TCP    172.16.210.162:49179   server21202:5938       ESTABLISHED
  TCP    172.16.210.162:49184   23.96.38.64:https      TIME_WAIT
  TCP    172.16.210.162:49185   23.96.38.64:https      TIME_WAIT
  TCP    172.16.210.162:49186   on-in-f97:https        TIME_WAIT
  TCP    172.16.210.162:49190   of-in-f102:https       TIME_WAIT
  TCP    172.16.210.162:49191   of-in-f102:https       TIME_WAIT
  TCP    172.16.210.162:49192   oj-in-f156:https       TIME_WAIT
  TCP    172.16.210.162:49193   on-in-f106:https       TIME_WAIT
  TCP    172.16.210.162:49194   23.96.38.64:https      TIME_WAIT

C:\Users\Administrator>
```

Step 3: Discovery of a trojan on the computer

- Kelley typed in a command that showed a long recursive directory listing
- Kelly typed “trojan virus” at the end of it
  - Look, that shows you have a trojan virus!
- Kelly showed the wikipedia page on Trojans to explain the problem
  - Which had a link to an article on “social engineering”
  - Which the researcher clicked on
  - Kelly was undeterred

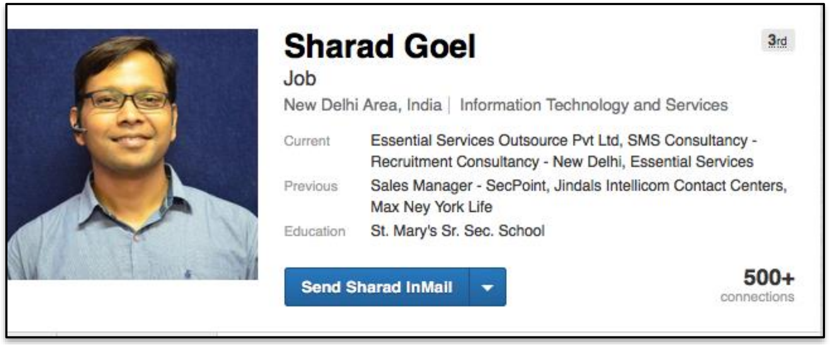


Administrator: C:\Windows\system32\cmd.exe

```
               0 File(s)              0 bytes

 Directory of C:\Users\Administrator\Saved Games

02/19/2015  03:59 PM    <DIR>          .
02/19/2015  03:59 PM    <DIR>          ..
               0 File(s)              0 bytes

 Directory of C:\Users\Administrator\Searches

02/19/2015  03:59 PM    <DIR>          .
02/19/2015  03:59 PM    <DIR>          ..
               0 File(s)              0 bytes

 Directory of C:\Users\Administrator\Videos

02/19/2015  03:59 PM    <DIR>          .
02/19/2015  03:59 PM    <DIR>          ..
               0 File(s)              0 bytes

     Total Files Listed:
             778 File(s)    318,058,042 bytes
            1154 Dir(s)   7,286,768,448 bytes free

C:\Users\Administrator>trojan virus
```

Step 4: Payment

- $100 for the virus removal,
  $50 to fix security drivers
  - “I do not have credit or debit cards”
    “Can I pay by check?”
- Pay to Essential Services Worldwide,
  4630 Border Village Road Suite N1497,
  San Ysidro, CA, 92173
- What do the researchers find out from this?
- Used Yellow pages, corporatedir.com, WHOIS, and LinkedIn to identify a company director and a DNS domain administrative contact



Sharad Goel 3rd

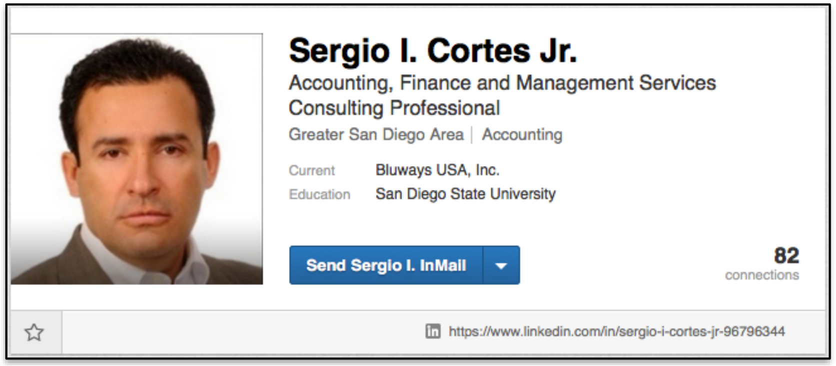
Job

New Delhi Area, India | Information Technology and Services

Current: Essential Services Outsource Pvt Ltd, SMS Consultancy - Recruitment Consultancy - New Delhi, Essential Services

Previous: Sales Manager - SecPoint, Jindals Intellicom Contact Centers, Max Ney York Life

Education: St. Mary's Sr. Sec. School



Sergio I. Cortes Jr.

Accounting, Finance and Management Services Consulting Professional

Greater San Diego Area | Accounting

Current: Bluways USA, Inc.

Education: San Diego State University

https://www.linkedin.com/in/sergio-i-cortes-jr-96796344

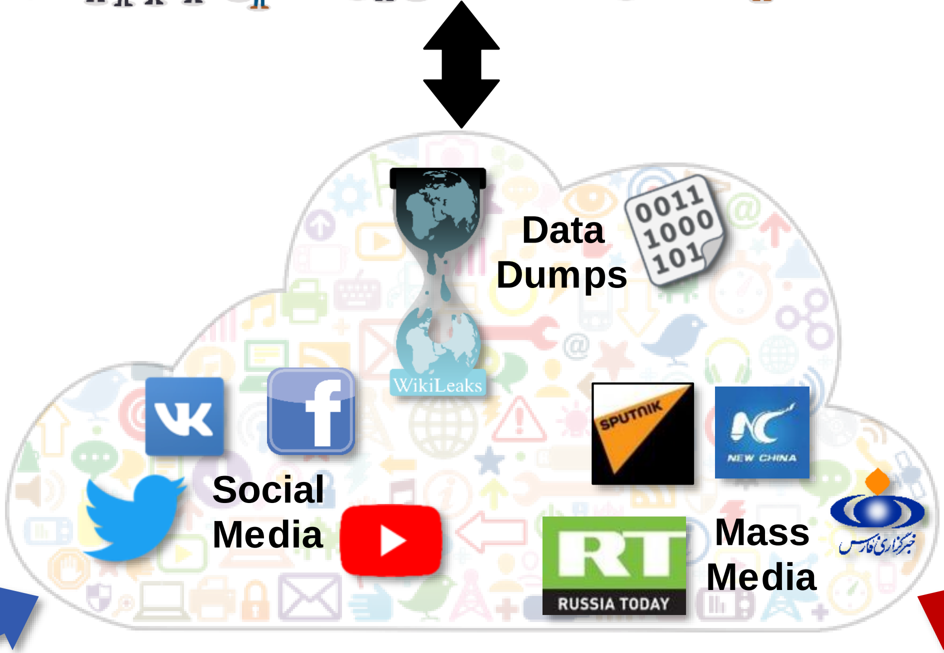
Overview of Influence Operations (IO)

Objective: Influence attitudes, behaviors, and decisions of target audience

U.S. & Allies

- Promote U.S. positions
- Strengthen relationship with allies
- Defend U.S. and western democracy
- Maintain peace and stability

- Positive narratives of U.S. positions
- Counter with new information



Data Dumps

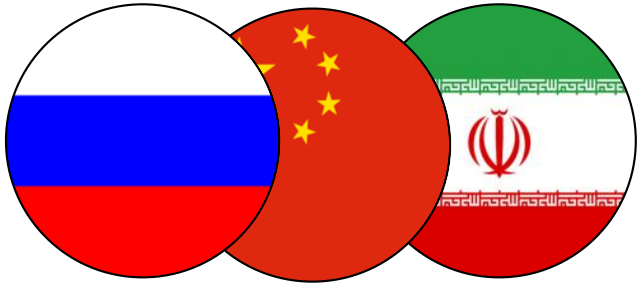
Social Media

Mass Media

Battlespace: Information Environment

with magnified scale, speed, and reach

Adversaries

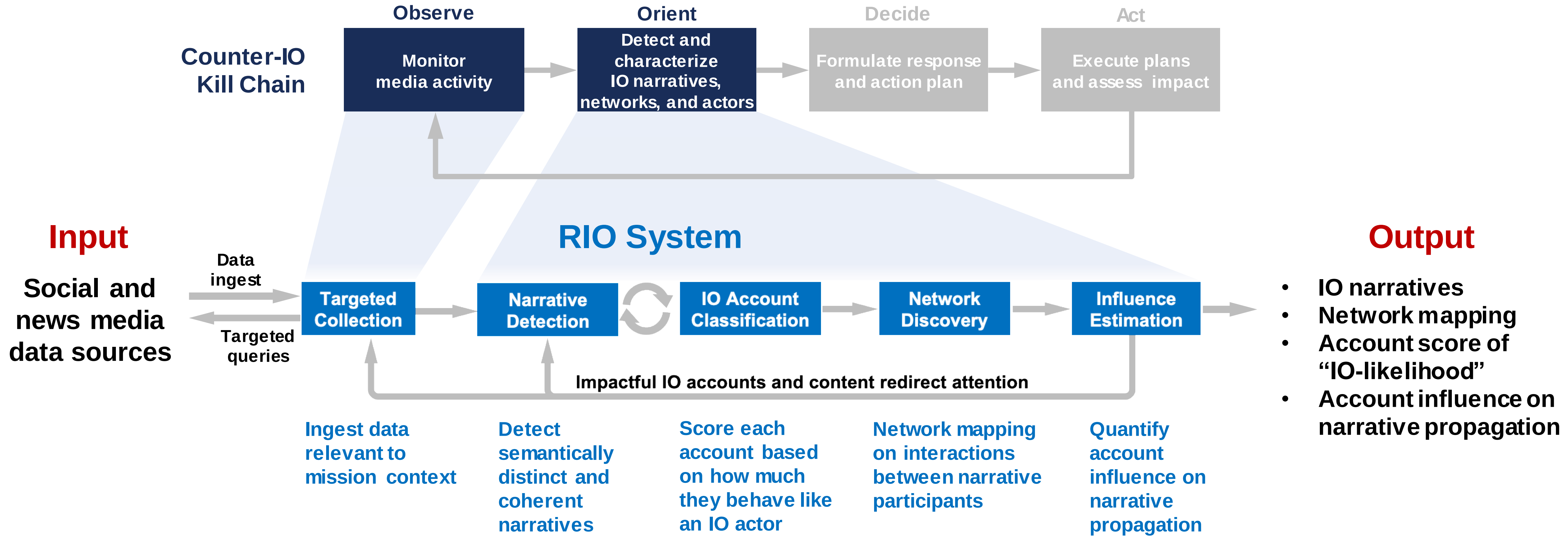


- Undermine U.S. influence
- Weaken NATO and EU alliances
- Attack U.S. and western democracy
- Incite local unrest

- Propaganda & disinformation
- Dismiss, Distort, Distract, Dismay (4Ds)

Reconnaissance of Influence Operations (RIO) Technical Approach

Objective: Automate detection of IO narratives, networks, and influential actors to provide actionable intelligence for countering the threat at its source



Counter-IO Kill Chain

Observe: Monitor media activity

Orient: Detect and characterize IO narratives, networks, and actors

Decide: Formulate response and action plan

Act: Execute plans and assess impact

Input

Social and news media data sources

Data ingest

Targeted queries

RIO System

Targeted Collection

Narrative Detection

IO Account Classification

Network Discovery

Influence Estimation

Impactful IO accounts and content redirect attention

Ingest data relevant to mission context

Detect semantically distinct and coherent narratives

Score each account based on how much they behave like an IO actor

Network mapping on interactions between narrative participants

Quantify account influence on narrative propagation

Output

- IO narratives
- Network mapping
- Account score of "IO-likelihood"
- Account influence on narrative propagation

* Smith at al. (2021), Automatic detection of influential actors in disinformation networks, *Proc. Natl. Acad. Sci. U.S.A.* 115(4) e2011216118

Security Lessons Recap

- Future attacks will be the unanticipated ones, particularly if you're successful
- Deployment means interoperability and co-existence with systems without the new security feature
- Flexibility without use cases leaves questions for someone else to answer
- The user is not an exception handling module
- More potential attack targets means more and more-varied attacks
- In theory, there is no difference between theory and practice. In practice, there is
- Humans won't do what technologists assume they will
- Open standards haven't worked for security user experience
- Developers are human, and any mistake might be a vulnerability
- In security, there is a large difference between data and code
- Technology turns old attacks into new attacks

Cyber Operations and Analysis Technology Group

Mary Ellen Zurko

mez@ll.mit.edu

MISSION: Design, prototype, and transition cyber technology to enable effective missions, operations, and assessments

Backup

Targeted Collection and Narrative Detection

Targeted Collection → Narrative Detection ⟳ IO Account Classification → Network Discovery → Influence Estimation

Targeted Data Collection

Challenge:

IO signal is buried in a massive amount of social and news media data

Approach:

- Targeted collection of PAI* within the boundary of policies
- Analyst provides cues on key topics, accounts, and spatiotemporal regions

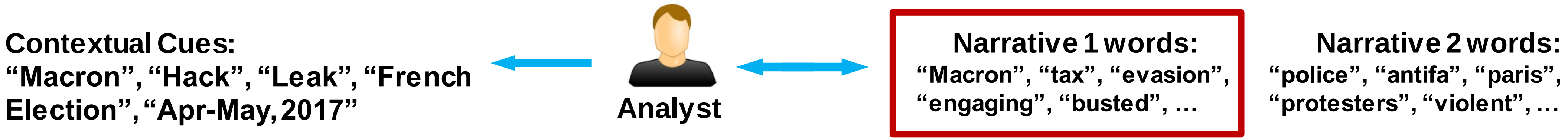
Narrative Detection

Challenge:

IO narratives are often complex and not narrowly defined by hashtags and keywords

Approach:

- Narrative detection using natural language processing algorithms in original language
- Topic modeling to identify distinct and coherent narratives
- Analyst selects from detected narratives



Contextual Cues:
“Macron”, “Hack”, “Leak”, “French Election”, “Apr-May, 2017”

Analyst

Narrative 1 words:
“Macron”, “tax”, “evasion”, “engaging”, “busted”, …

Narrative 2 words:
“police”, “antifa”, “paris”, “protesters”, “violent”, …

* PAI: Publicly available information

IO Account Classification

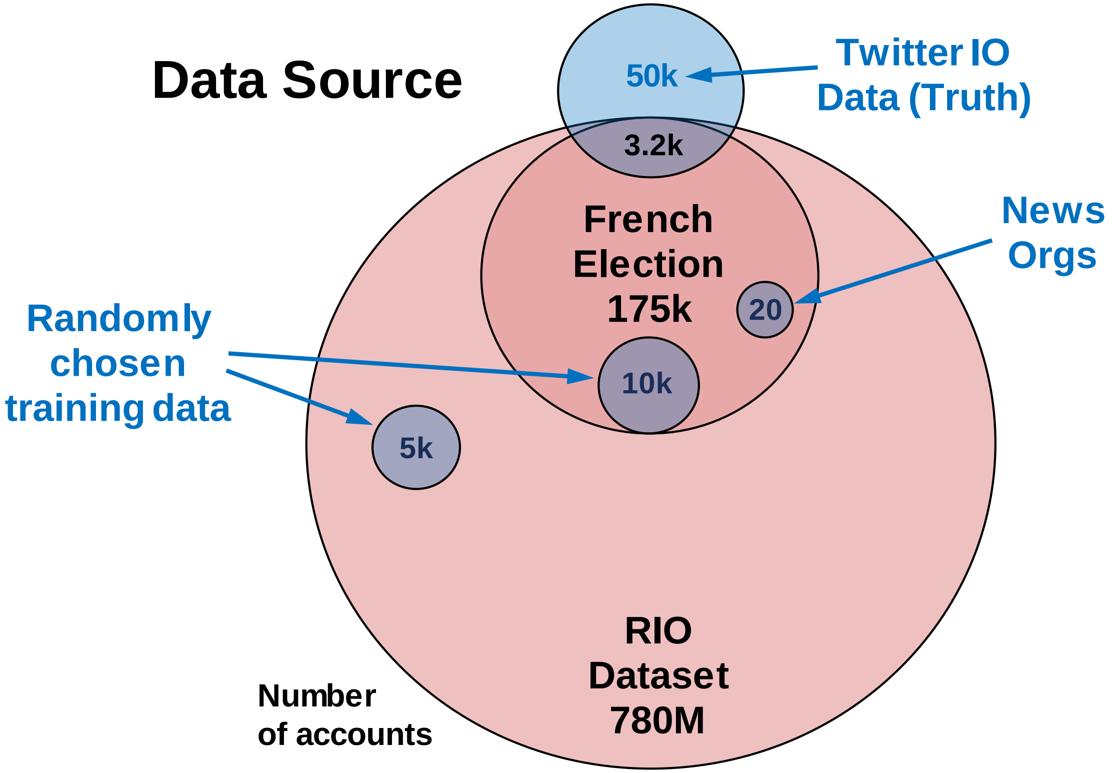
Targeted Collection → Narrative Detection ⟳ **IO Account Classification** → Network Discovery → Influence Estimation

Challenge: Need to automate detection of IO accounts operated by both bots and humans
Approach: Principled feature engineering and machine learning with ensemble tree classifier

Construct Training Set

- **Identify known IO and known non-IO accounts**
- **Select random accounts from collection**
- **Classify unknown accounts with semi-supervised learning heuristics**

Data Source



IO Account Classification

Targeted Collection → Narrative Detection ⟳ **IO Account Classification** → Network Discovery → Influence Estimation

Challenge: Need to train classifier when known IO accounts are limited in number and may not have engaged in target narrative

Solution: Use semi-supervised learning to label accounts with strong IO behavior* for training data

Snorkel †

- **Uses heuristic labeling functions to label accounts**
- **Can label large training sets with minimal effort**
- **Allows for training the classifier in narratives with limited labeled Twitter data**

Our Snorkel Labeling Functions

- **Each gives a label of IO, REAL, or ABSTAIN**
- **Functions are:**
  - **Independent of narrative**
  - **Based on profile and behavioral characteristics only**
  - **Learned from observations, IO account vs general pop.**
  - **Validated on small set of hand labeled accounts**

* Ratner, et al. Snorkel: Rapid training data creation with weak supervision, *Proc. VLDB Endowment* (2017)

† Luceri, et al. Don't feed the troll: Detecting troll behavior via inverse reinforcement learning, *Proc. Intl. Conf. Web and Social Media* (2020)

IO Account Classification

Targeted Collection
Narrative Detection
IO Account Classification
Network Discovery
Influence Estimation

Label as IO if:

- No account profile
- Frequent interactions with suspect news accounts
- Following excessive number of accounts
- Most tweets include links
- Tweets in too many languages
- Has many tweets in an undetermined language
- Has almost no or far too many favorites

Label as REAL if:

- Has a follower-following ratio consistent with real people
- Had few or no interactions with suspect news accounts
- Tweeted very few links
- Has a reasonable number of likes
- Profile length normal
- Has a very large number of followers (typical of organizations)

If criterion not met, label ABSTAIN

- Accounts receive mix of labels, may conflict
- Resolve label set into single probability *p* in [0,1]
- Accounts with *p*>=0.7 labeled as IO in training set

Influence Estimation Using Network Causal Inference*

Targeted Collection → Narrative Detection ⟳ IO Account Classification → Network Discovery → **Influence Estimation**

Potential outcomes of account i :
(number of narrative tweets)

$$Y_i(\boldsymbol{z}, \boldsymbol{A})$$

Source vector   Influence network

Causal influence of account *k*:

$$\zeta_k = \text{Average}[Y_i(\boldsymbol{z}_{k+}, \boldsymbol{A}) - Y_i(\boldsymbol{z}_{k-}, \boldsymbol{A})]$$

k present (observed)

k absent (counterfactual, imputed using outcome model)

Network potential outcome model:

$$Y_i \sim Poisson(\lambda_i)$$

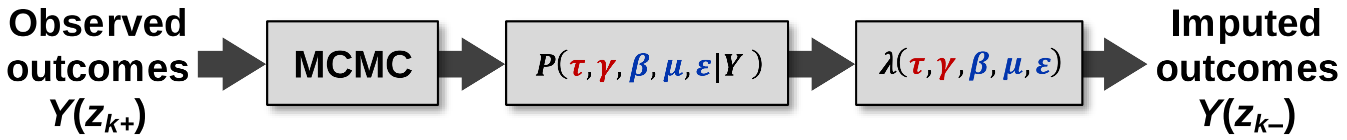
$$\log\lambda_i = \tau Z_i + \sum_{n=1}^{N_{hop}} \prod_{k=1}^{n} \tau\gamma_k s_i^{(n)} + \beta^{\mathrm{T}} x_i + \mu + \varepsilon_i$$

Exposure to source

Individual baseline

Adjusts for confounders (e.g. node degrees and community membership)

Bayesian imputation:



Observed outcomes $Y(z_{k+})$
MCMC
$P(\tau, \gamma, \beta, \mu, \varepsilon | Y)$
$\lambda(\tau, \gamma, \beta, \mu, \varepsilon)$
Imputed outcomes $Y(z_{k-})$

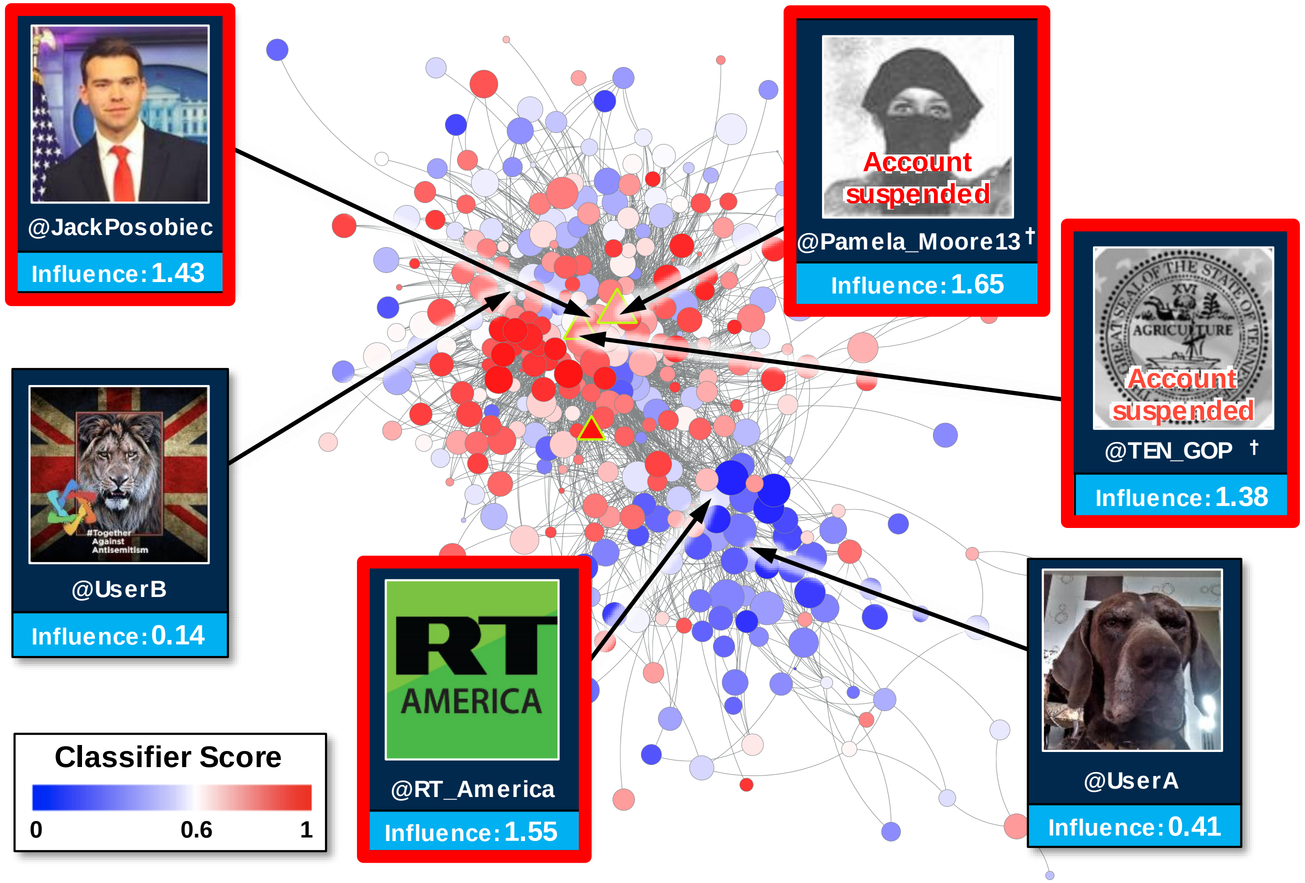
- Causal influence captures each account’s contribution to the overall narrative tweets
- Outcome model expresses narrative propagation on the network
- Causal framework disentangles social confounders (e.g. homophily) from actual influence

* Smith et al., System and technique for influence estimation on social media networks using causal inference, U.S. Patent Application No. 62/654,782

Causal Influence Estimation Performance Evaluation

Targeted Collection
Narrative Detection
IO Account Classification
Network Discovery
Influence Estimation

Anti-Macron Narrative Network



Nodes are colored by IO classifier score, and sized by causal influence

Screen name	T	RT	F	Earliest time	Pagerank Centrality	RIO
@RT_America*	39	8	386k	12:00	2706	1.55
@JackPosobiec	28	123	23k	01:54	4690	1.43
@UserA	8	0	1.4k	22:53	44	0.14
@UserB	12	15	19k	12:27	151	0.41
@Pamela_Moore13†	10	31	56k	18:46	97	1.65
@TEN_GOP†	12	42	112k	23:15	191	1.38

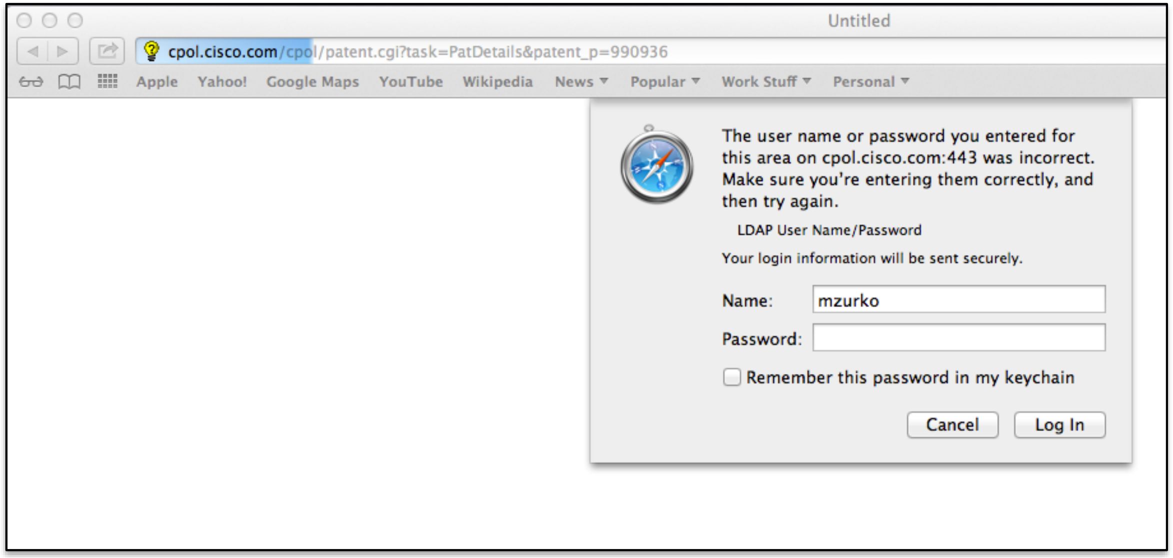
Tweets (T), Retweets (RT), Followers (F), Causal influence estimate (RIO)
**RT_America = "Russia Today" America*

- Causal influence score measures contribution to narrative flow on the network, beyond activity-based and topological statistics
- Results are corroborated by evidence from Twitter† and journalist reports
- RIO finds key actors that do not stand out based on traditional statistics for measuring influence

†U.S. HPSCI. Exhibit of user accounts that Twitter has identified as being tied to Russia's "Internet Research Agency." (Nov. 2017)

(Basic) Authentication

- **Security the way Tim intended**
- **Server says: WWW-Authenticate: Basic realm="*insert realm*”**
- **User prompted for their password**
- **Client says: Authorization: Basic QWxhZGluOnNlc2FtIG9wZW4=**
  - **User agent remembers and sends for that URI domain/realm**



The user name or password you entered for this area on cpol.cisco.com:443 was incorrect. Make sure you're entering them correctly, and then try again.

LDAP User Name/Password

Your login information will be sent securely.

Name: mzurko

Password:

□ Remember this password in my keychain

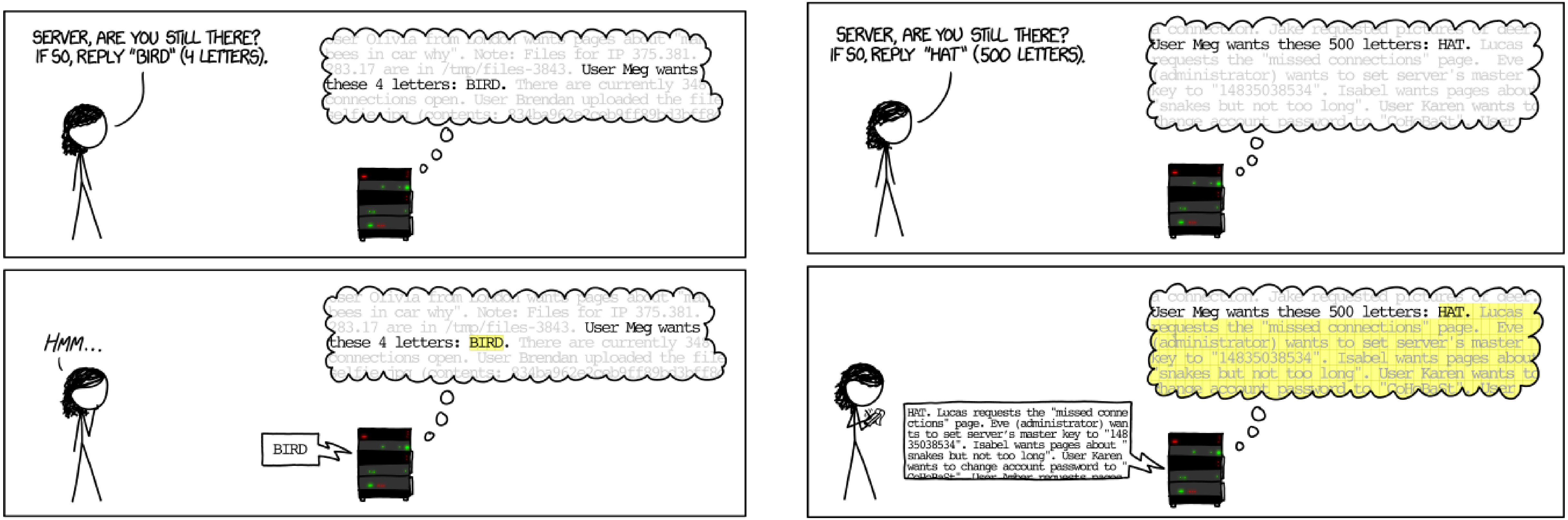
Cancel | Log In

Digest Authentication

- Cryptographically hash the password
- With the username and realm
  - Defense against Rainbow Tables
- Nonces in the server challenge for replay protection
- Started in 1994; RFC in 1997
- Resists passive attacker on the network
- Minimizes handling of password plaintext
  - No passing the password itself in the protocol
  - No need to store the password in the clear

xkcd

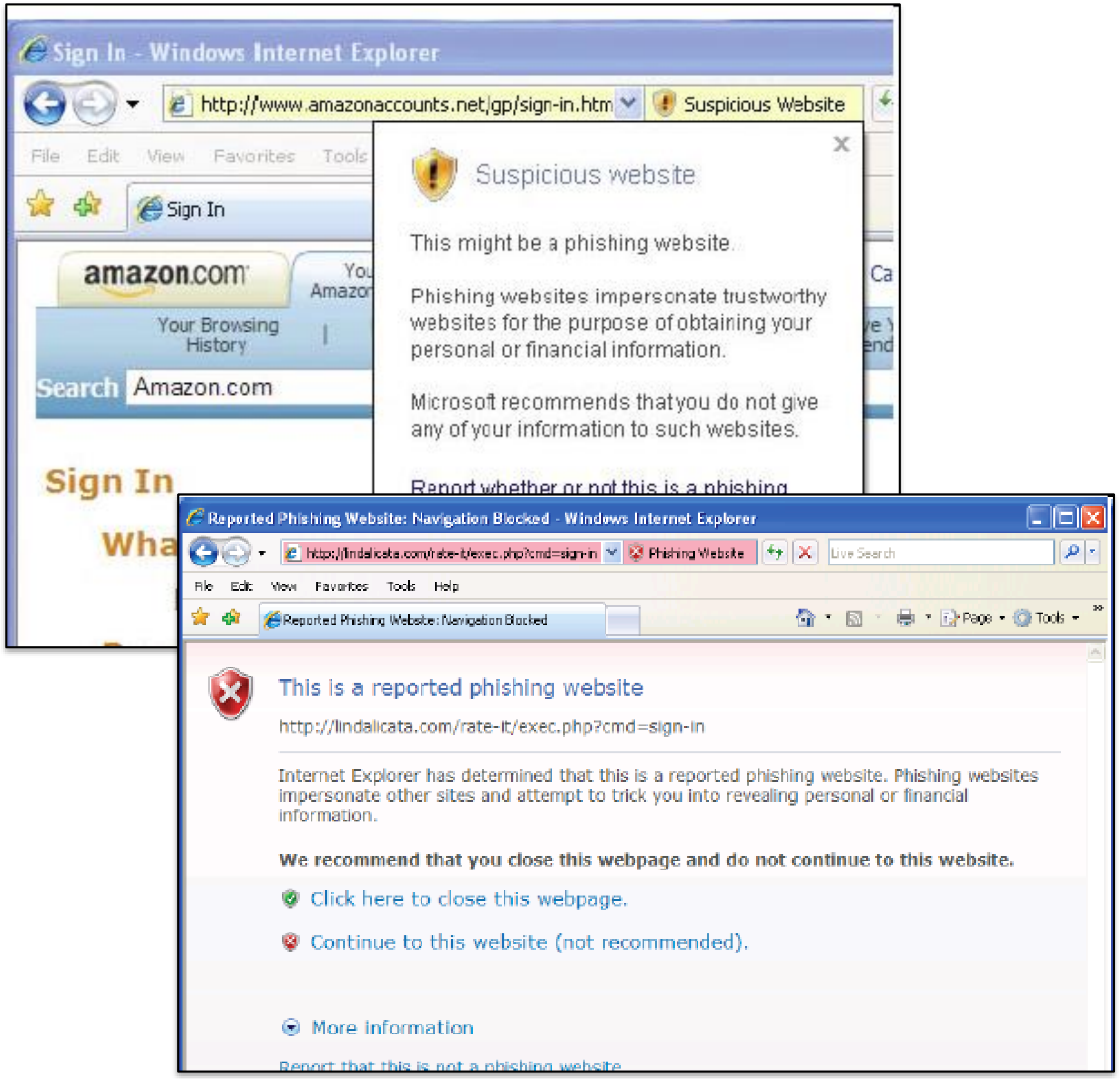
A WEBCOMIC OF ROMANCE, SARCASM, MATH, AND LANGUAGE.


SERVER, ARE YOU STILL THERE? IF SO, REPLY "BIRD" (4 LETTERS).
User Olivia from London wants pages about "ma bees in car why". Note: Files for IP 375.381. 283.17 are in /tmp/files-3843. User Meg wants these 4 letters: BIRD. There are currently 346 connections open. User Brendan uploaded the file selfie.jpg (contents: 834ba962e2ceb9ff89bd3bff8
HMM...
User Olivia from London wants pages about "ma bees in car why". Note: Files for IP 375.381. 283.17 are in /tmp/files-3843. User Meg wants these 4 letters: BIRD. There are currently 346 connections open. User Brendan uploaded the file selfie.jpg (contents: 834ba962e2ceb9ff89bd3bff8
BIRD
SERVER, ARE YOU STILL THERE? IF SO, REPLY "HAT" (500 LETTERS).
a connection. Jake requested pictures of deer. User Meg wants these 500 letters: HAT. Lucas requests the "missed connections" page. Eve (administrator) wants to set server's master key to "14835038534". Isabel wants pages about "snakes but not too long". User Karen wants to change account password to "CoHoBaSt". User
a connection. Jake requested pictures of deer. User Meg wants these 500 letters: HAT. Lucas requests the "missed connections" page. Eve (administrator) wants to set server's master key to "14835038534". Isabel wants pages about "snakes but not too long". User Karen wants to change account password to "CoHoBaSt". User
HAT. Lucas requests the "missed conne ctions" page. Eve (administrator) wan ts to set server's master key to "148 35038534". Isabel wants pages about " snakes but not too long". User Karen wants to change account password to " CoHoBaSt". User Amber requests pages

You’ve Been Warned

An Empirical Study of the Effectiveness of Web Browser Phishing Warnings

- Simulated spear phishing
  - 97% fell for at least one
  - 79% heeded active warnings when presented
- Active warnings directly interrupt the task, give the user choices, and make recommendations
  - Fail safely
- Correlations between understanding a warning and heeding it


Sign In - Windows Internet Explorer
http://www.amazonaccounts.net/gp/sign-in.htm
Suspicious Website
File
Edit
View
Favorites
Tools
Sign In
amazon.com
Your Browsing History
Search
Amazon.com
Sign In
Wha

Suspicious website

This might be a phishing website.

Phishing websites impersonate trustworthy websites for the purpose of obtaining your personal or financial information.

Microsoft recommends that you do not give any of your information to such websites.

Report whether or not this is a phishing

Reported Phishing Website: Navigation Blocked - Windows Internet Explorer
http://lindalicata.com/rate-it/exec.php?cmd=sign-in
Phishing Website
Live Search
File
Edit
View
Favorites
Tools
Help
Reported Phishing Website: Navigation Blocked
Page
Tools

This is a reported phishing website

http://lindalicata.com/rate-it/exec.php?cmd=sign-in

Internet Explorer has determined that this is a reported phishing website. Phishing websites impersonate other sites and attempt to trick you into revealing personal or financial information.

We recommend that you close this webpage and do not continue to this website.

Click here to close this webpage.

Continue to this website (not recommended).

More information

Report that this is not a phishing website