Misinformation, Technology, and Usability
Lessons from Usable Security

Mary Ellen Zurko

mez@ll.mit.edu

December 11, 2019

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

How do people react to misinformation?

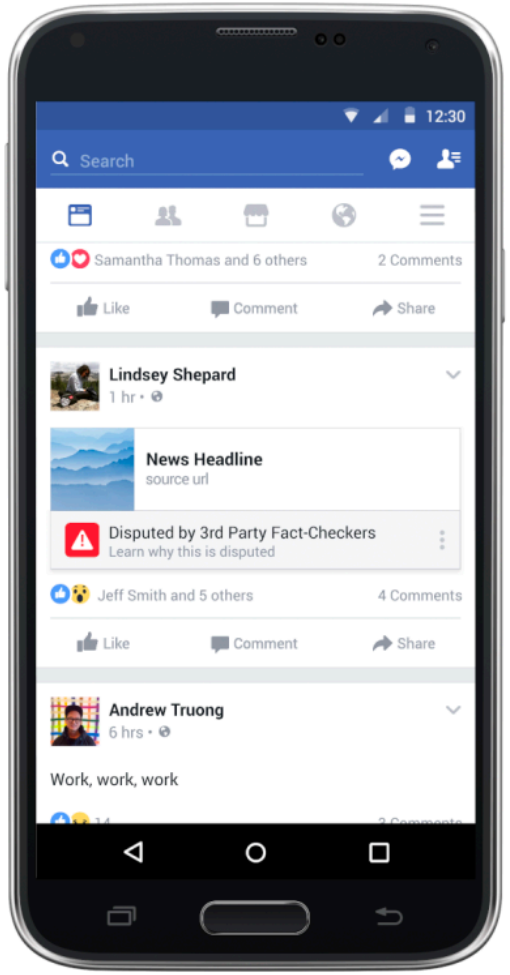
Weekly World News

The World's Only Reliable News

People will pay to read misinformation

How about warnings about misinformation?

Facebook's Handling of Misinformation



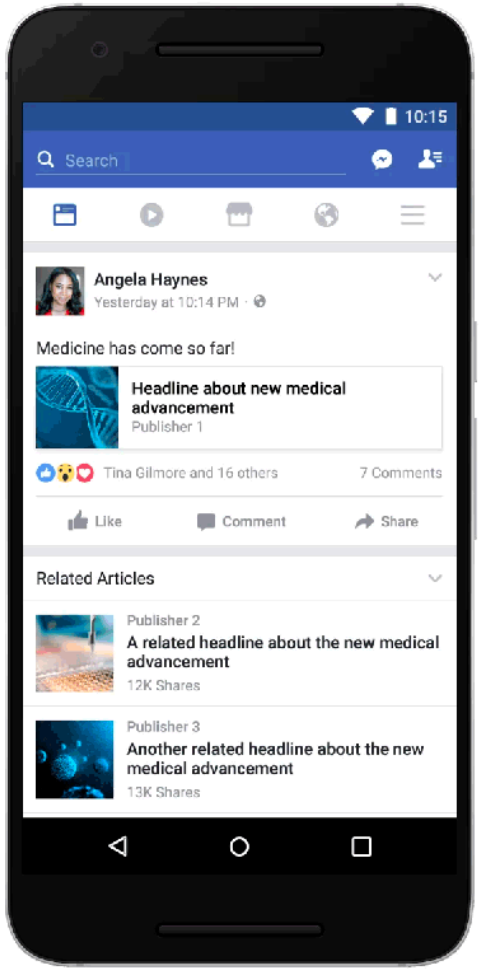
aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science in the Public Interest
13(3) 106–131

DOI: 10.1177/1529100612451018
http://pspi.sagepub.com
SAGE

Misinformation and Its Correction: Continued Influence and Successful Debiasing

Stephan Lewandowsky[1], Ullrich K. H. Ecker[1], Colleen M. Seifert[2], Norbert Schwarz[2], and John Cook[1,3]
[1]University of Western Australia, [2]University of Michigan, and [3]University of Queensland

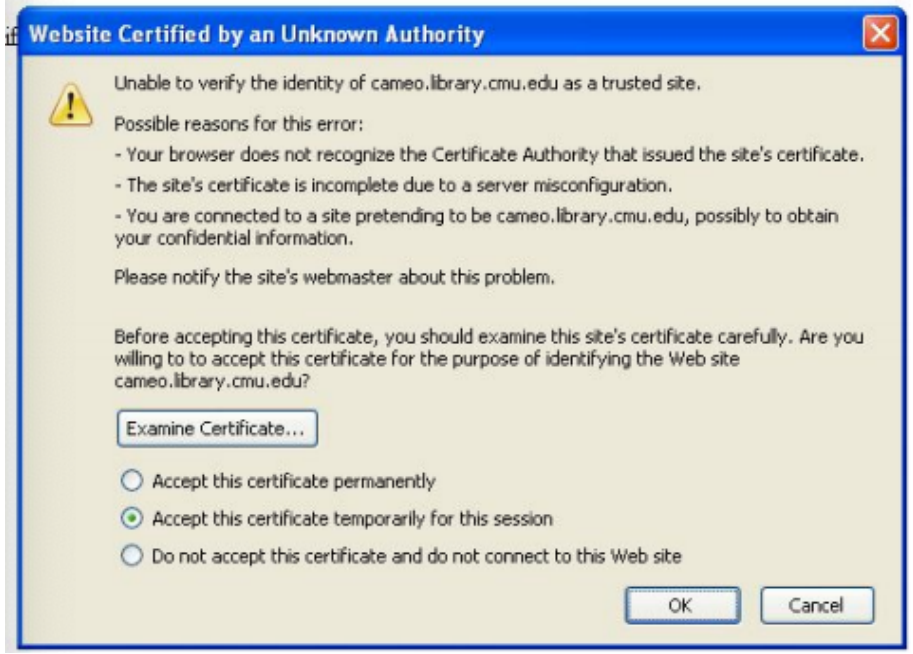


Facebook went from warnings to more information

How do people react to warnings for cybersecurity?

Error Handling in TLS web site authentication

- Servers got a self signed certificates
  - CA issued certificates cost money
  - 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 1.54% false positive warning rate on 3.9 billion TLS connections across 300k academic users

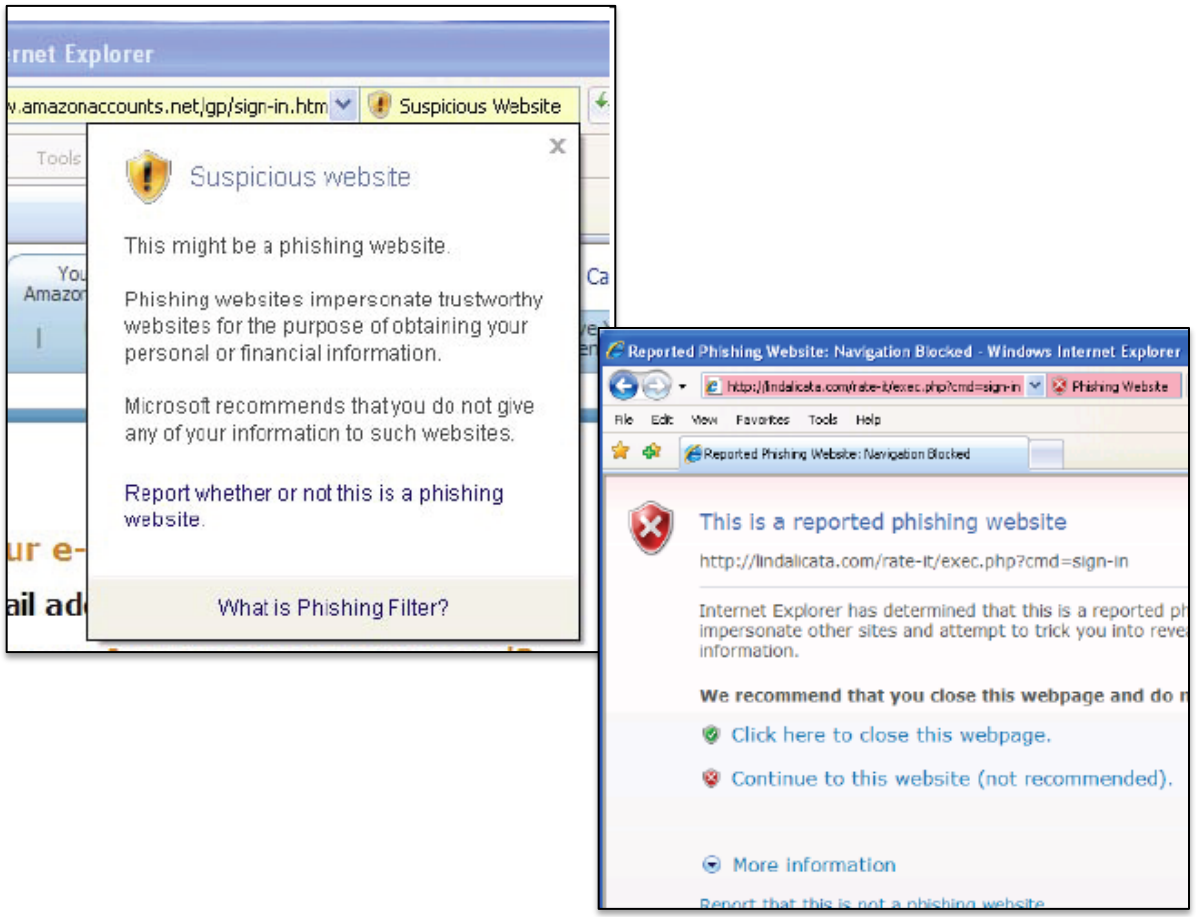


False positive warnings cause habituation

You've Been Warned

An Empirical Study of the Effectiveness of Web Browser Phishing Warnings

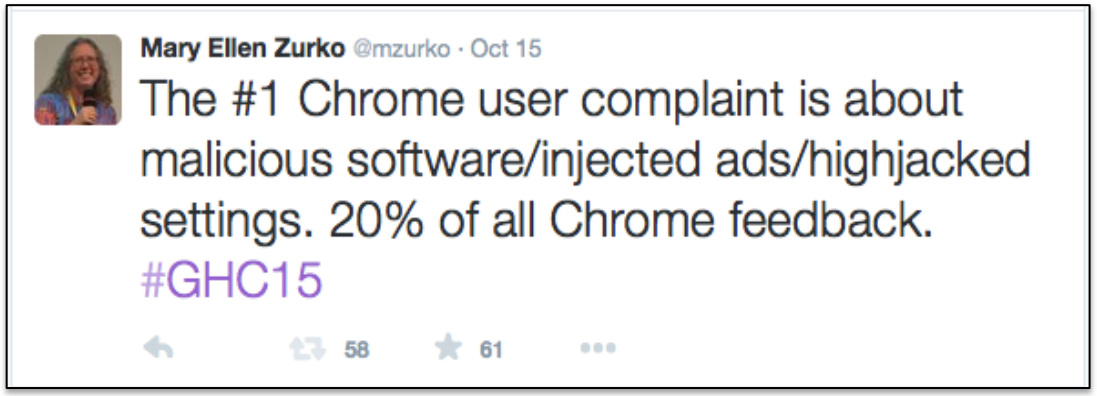
- **Simulated spear phishing**
  - **97% fell for at least one**
  - **79% heeded active warnings when presented**
- **Active warnings directly interrupt the task, give the user choices, and make recommendations**
  - **Fail safely**
- **Correlations between understanding a warning and heeding it**



Accurate warnings alone are not enough

User Experience and Warnings Continued To Be a Challenge

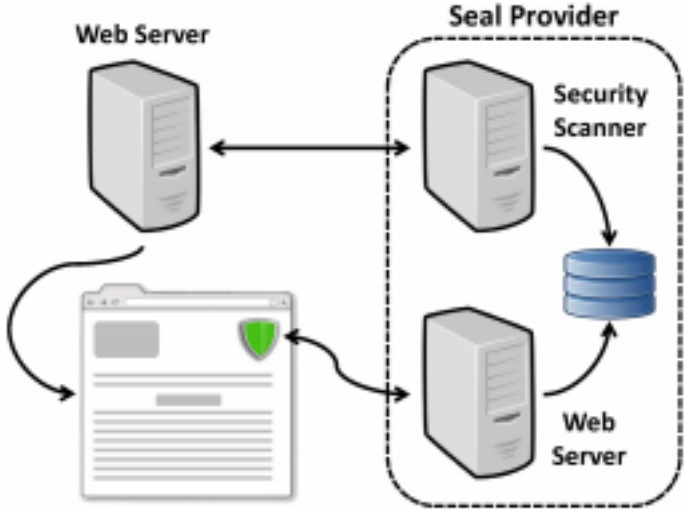
- Firefox Click Through Rate (CTR) for malware warnings is 33% (2014)
  - Google Chrome's 70%
- Mock Firefox styling closed that difference by 12 to 20 points in a 10 day at scale controlled experiment
  - Change to text, layout, default button
  - Users heed warnings to sites they have not visited
  - 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%



Accurate warnings with opinionated design had impact

What about more information to influence cybersecurity?

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 well known vulnerabilities?
  - 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 status on seal provider, an indicator of 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 also 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

Lessons from Usable Security

- Technology that identifies misinformation is unlikely to be insufficient to influence
- False warnings will decrease warning impact through habituation
- Warnings can easily not have the impact their designers intended
- Humans may trust familiar or confirming information more than warnings about something that it is harmful
- Influencing through additional positive information can be ineffective or have surprising consequences

Thank you for your attention and questions

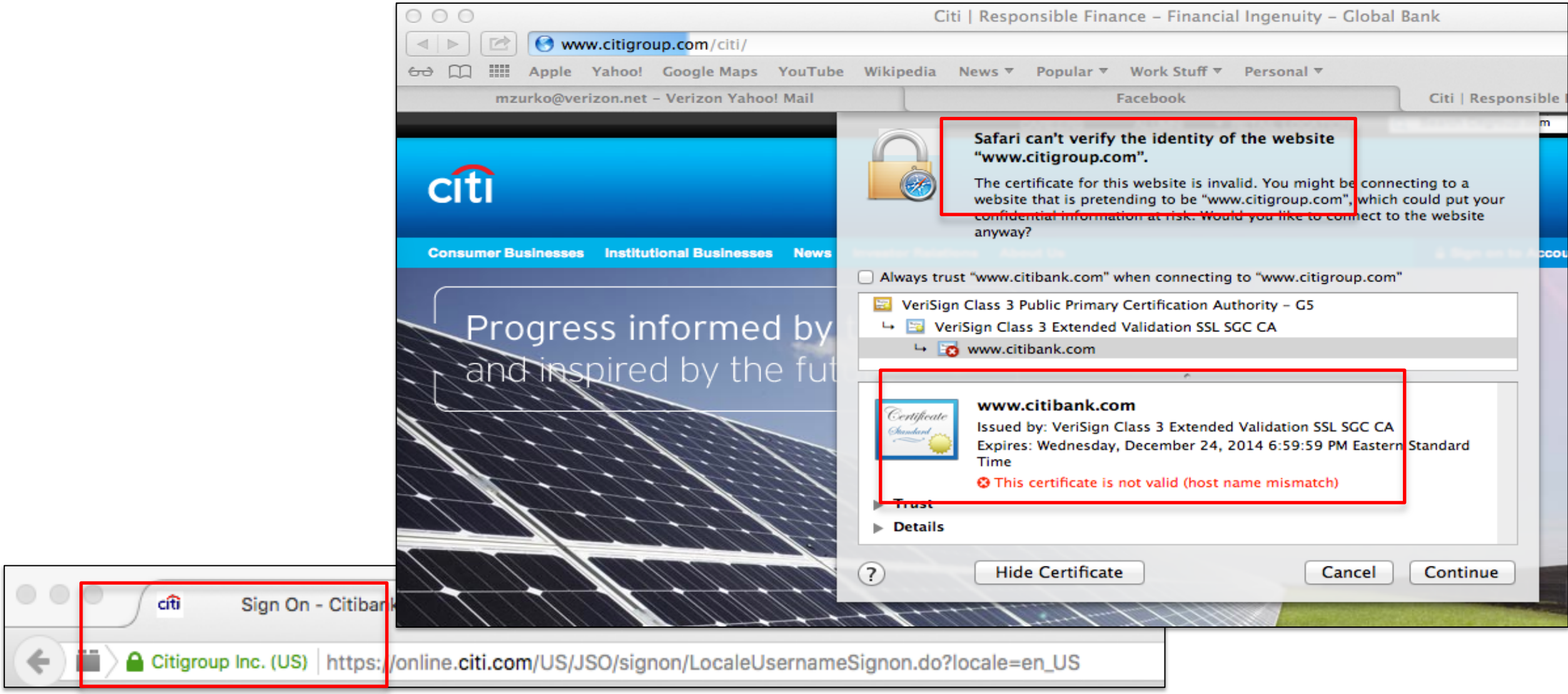
mez@ll.mit.edu

Backup

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

Are warnings about domains from HTTPS meaningful?

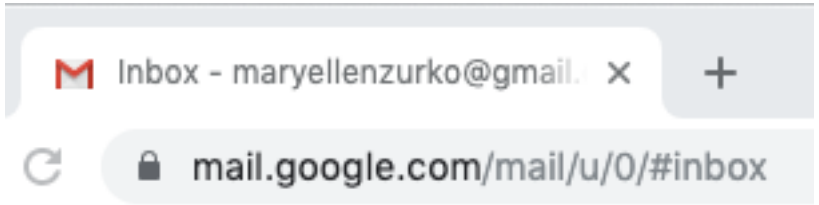


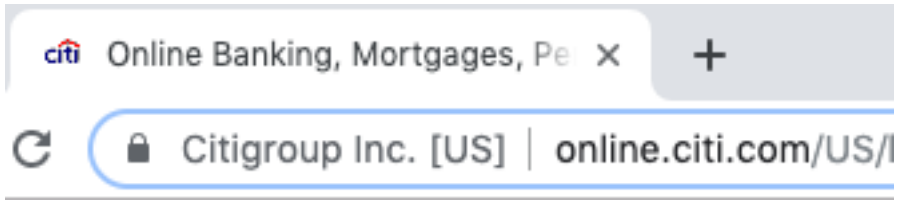
W3C Web Security Context (WSC)

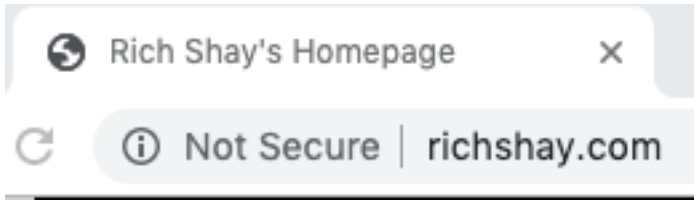
- First usable security standard
- Charter: To enable users to come to a better understanding of the context that they are operating in when making trust decisions on the Web
  - Specify a baseline set of security context information and practices for the secure and usable presentation of this information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

SSL/TLS – HTTPS:

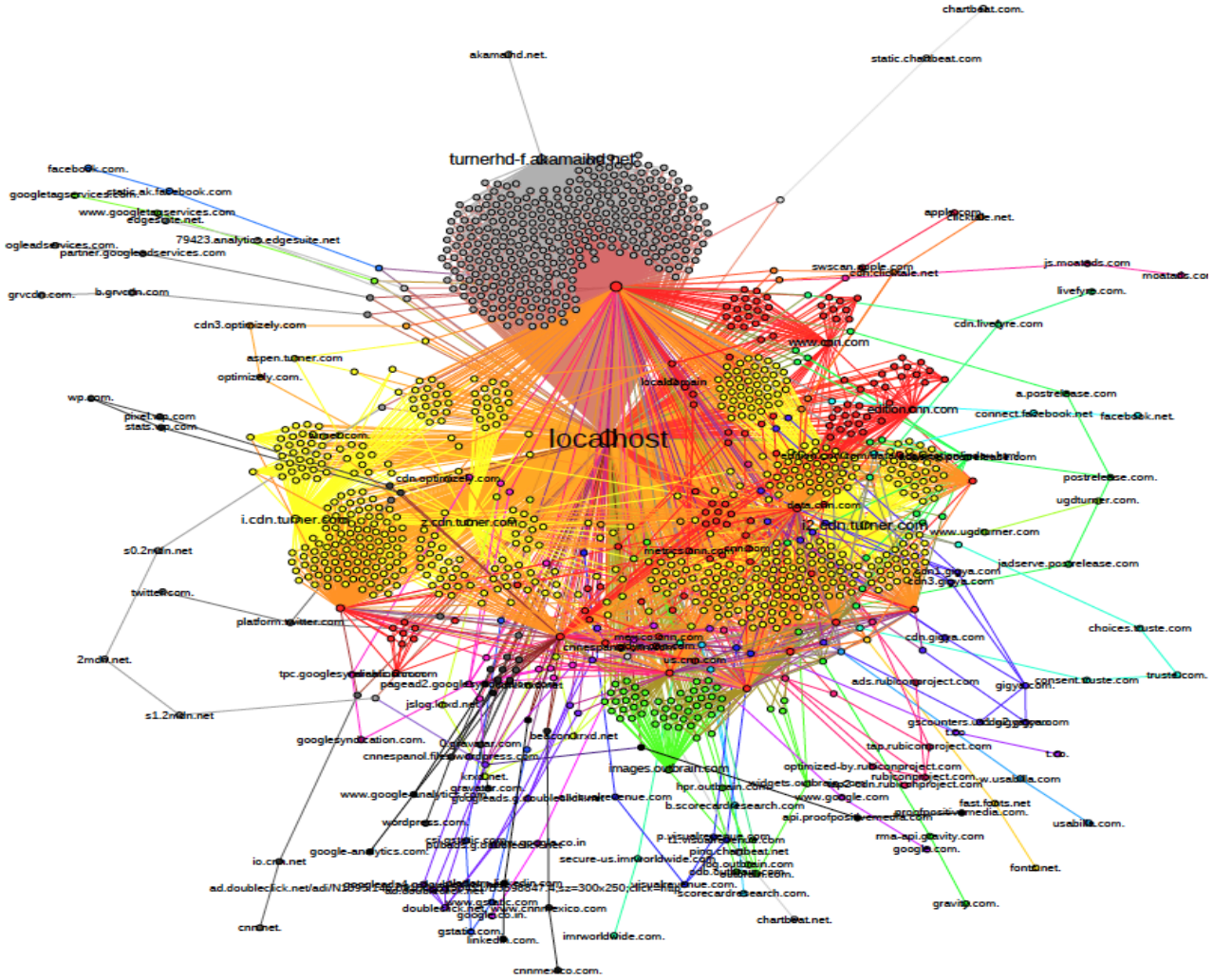
- Encryption! Authentication! Security!
- Open standard
- Authentication of the server using public key certificate
  - Trust, Trustworthy, and Trust for What?
- Authentication of the client using public key certificate is an option
- The encryption part works pretty darn well
- The authentication part…







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

- Standards Challenges
  - "Successful standards enable"
    - We had a lot of "Don't do this thing" and constraints
  - UI standards are process, not presentation
- Context Challenges
  - Browser vendor participation
    - Some of the reasons vendors participate:
      Interoperability (as required by/for the market)
      Customer requirements (compliance and laws and features)
    - Some of the reasons vendors don't participate:
      IP/patents
      Dilution of their brand
      Market advantage in the area
  - And then mobile
    - Technology marches forward