09. Web Security & Privacy

Blase Ur, April 24th, 2017
CMSC 23210 / 33210
Today’s class

• Trust on the web
  – SSL notifications

• Online tracking
  – Privacy tools
Trust on the web
Overview

- Secure Sockets Layer (SSL) and its successor, Transport Layer Security (TLS) enable secure communication
- Frequently encountered with web browsing (HTTPS) and more behind the scenes in app, VOIP, etc.
What we want to defend against

• People snooping on our communications
  – The contents of what we’re sending
  – Session tokens (see, e.g., Firesheep)

• Man-in-the-middle attacks
  – We want to authenticate that we are talking to the right site, not an imposter
  – Use certificates inside a public-key infrastructure
How we could obtain trust

• Web of trust
  – People you already trust introduce you to people they trust
  – Can get complicated, doesn’t scale well
  – Infrequently seen in practice

• Public-Key Infrastructure (PKI)
  – Certificates are issued by certificate authorities that bind cryptographic keys to identities
Public-Key Infrastructure

• Binding of keys to identities
What does SSL look like to users?

• Compare, e.g., the following:
  – https://www.google.com (normal certificate)
  – Go to Google images and then click on an image and see what happens (mixed content)
  – https://www.thawte.com (EV certificate)
What does SSL look like to users?

<table>
<thead>
<tr>
<th>Browser</th>
<th>HTTPS</th>
<th>HTTPS minor error</th>
<th>HTTPS major error</th>
<th>HTTP</th>
<th>EV</th>
<th>Malware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrome 48 Win</td>
<td><a href="https://www.example.com">https://www.example.com</a></td>
<td></td>
<td></td>
<td><a href="http://www.example.com">www.example.com</a></td>
<td>Symantec Co</td>
<td><a href="https://dow">https://dow</a></td>
</tr>
<tr>
<td>Edge 20 Win</td>
<td><a href="https://mix.example.com">https://mix.example.com</a></td>
<td></td>
<td></td>
<td><a href="http://www.example.com">www.example.com</a></td>
<td>Symantec Co</td>
<td>Unsafe website: demo</td>
</tr>
<tr>
<td>Firefox 44 Win</td>
<td><a href="https://mix.example.com">https://mix.example.com</a></td>
<td></td>
<td></td>
<td><a href="http://www.example.com">www.example.com</a></td>
<td>Symantec Corp</td>
<td><a href="https://space.net">https://space.net</a></td>
</tr>
<tr>
<td>Safari 9 Mac</td>
<td>example.com</td>
<td>mixed.badssl.com</td>
<td></td>
<td></td>
<td></td>
<td>downloadgamer</td>
</tr>
<tr>
<td>Chrome 48 And</td>
<td><a href="https://v.example.com">https://v.example.com</a></td>
<td></td>
<td>mixed.badssl.com</td>
<td></td>
<td>Symantec Corp</td>
<td></td>
</tr>
<tr>
<td>Opera Mini 14 And</td>
<td><a href="http://www.example.com">www.example.com</a></td>
<td>url hidden</td>
<td></td>
<td></td>
<td></td>
<td>Unavailable</td>
</tr>
<tr>
<td>UC Mini 10 And</td>
<td>Example Do.</td>
<td></td>
<td></td>
<td></td>
<td>Example Do.</td>
<td>Blocked</td>
</tr>
<tr>
<td>UC Browser 2 iOS</td>
<td>Example Do.</td>
<td>mixed.badssl</td>
<td></td>
<td></td>
<td>Example Do.</td>
<td>Unavailable</td>
</tr>
<tr>
<td>Safari 9 iOS</td>
<td>example.com</td>
<td></td>
<td>mixed.badssl</td>
<td></td>
<td>Symantec</td>
<td>Unavailable</td>
</tr>
</tbody>
</table>

(From Felt et al. SOUPS 2016)
How does PKI look to browsers?

• Hundreds of trusted certificate authorities
  – Certificate authorities (CAs) sign the certificates binding identities to keys
  – See, e.g., Firefox’s advanced settings
How does PKI look to site admins?

• Apply for a certificate
  – Validation process
  – Certificate authorities (CAs) delegate trust (“chain of trust”)
  – CAs sell you a certificate
Issues with SSL/TLS/PKIs

• Implementation issues
• Communicating to users what is happening
• Compromised Certificate Authorities
• Man-in-the-middle attacks
  – Downgrade/dumbing-down attacks
  – Addition of “rogue” certificates
• Revocation
• Timing attacks and other side channels
One famous implementation issue

- OpenSSL bug
  - Heartbleed (CVE-2014-0160)
  - TLS heartbeat extension misses a bounds check and thus lets an attacker “read” memory
Compromised CAs

• Comodo and Diginotar both suffered breaches in 2011 that let attackers issue rogue certificates

• What about untrustworthy CAs?
  – Compelled certificate creation attacks (see, e.g., Soghoian and Stamm FC ’11)
Man-in-the-middle attacks (MITM)

- Effectively, many corporations perform MITM attacks by adding certificates to users’ computers and presenting “fake” certificates to users.

- A man in the middle can also tell you a site doesn’t support SSL/TLS (downgrade) or any strong ciphers (dumbing down)
  - Why does this create a huge problem?
  - Why is this hard to deal with?
Important question 1

• How do you know if a site supports HTTPS?
  – EFF’s HTTPS Everywhere
  – HTTP Strict Transport Security (HSTS)
  – In both cases, how do you bootstrap/maintain?
Important question 2

• How do you know you have the right certificate for a site?
  – Certificate transparency
  – Public key pinning
  – Perspectives (originally a CMU project)
How do you know a cert is valid?

• Certificates can be revoked in case of a compromise

• Certificate Revocation Lists (CRLs) were used, but they got really large
  – Incremental updates were better

• Online Certificate Status Protocol (OCSP)
  – How does this impact privacy?

• OCSP Stapling
Self-signed certificates

• What happens if someone signs their own certificate and chooses not to use the PKI infrastructure?
  – You get a warning!
Warnings

This applet was signed by "Unlimi–Tech Software Inc.," and authenticated by "Thawte Consulting cc". Do you trust this certificate?

Click Trust to run this applet and allow it unrestricted access to your computer. Click Don't Trust to run this applet with standard Java restrictions.
The server's certificate chain is incomplete, and the signer(s) are not registered. Accept?

Server name:

grey-dev.ece.cmu.edu
The certificate for "grey-dev.ece.cmu.edu" is signed by the unknown Certificate Authority "grey-dev.ece.cmu.edu". It is not possible to verify that this is a valid certificate.

Certificate summary

Holder: grey-dev.ece.cmu.edu

Issuer: grey-dev.ece.cmu.edu

Expires: 02/25/2019 02:38:00 PM GMT

Encryption protocol

256 bit AES (DHE_RSA/SHA)
Opera
The site's security certificate is not trusted!

You attempted to reach grey-dev.ece.cmu.edu, but the server presented a certificate issued by an entity that is not trusted by your computer's operating system. This may mean that the server has generated its own security credentials, which Chromium cannot rely on for identity information, or an attacker may be trying to intercept your communications.

You should not proceed, especially if you have never seen this warning before for this site.

- Proceed anyway
- Back to safety

Help me understand
**Chromium**

You attempted to reach `grey-dev.ece.cmu.edu`, but the server presented a certificate issued by an entity that is not trusted by your computer's operating system. This may mean that the server has generated its own security credentials, which Chromium cannot rely on for identity information, or an attacker may be trying to intercept your communications.

You should not proceed, **especially** if you have never seen this warning before for this site.

[Button] Proceed anyway  [Button] Back to safety

**Help me understand**

When you connect to a secure website, the server hosting that site presents your browser with something called a "certificate" to verify its identity. This certificate contains identity information, such as the address of the website, which is verified by a third party that your computer trusts. By checking that the address in the certificate matches the address of the website, it is possible to verify that you are securely communicating with the website you intended, and not a third party (such as an attacker on your network).

In this case, the certificate has not been verified by a third party that your computer trusts. Anyone can create a certificate claiming to be whatever website they choose, which is why it must be verified by a trusted third party. Without that verification, the identity information in the certificate is meaningless. It is therefore not possible to verify that you are communicating with `grey-dev.ece.cmu.edu` instead of an attacker who generated his own certificate claiming to be `grey-dev.ece.cmu.edu`. You should not proceed past this point.

If, however, you work in an organization that generates its own certificates, and you are trying to connect to an internal website of that organization using such a certificate, you may be able to solve this problem securely. You can import your organization's root certificate as a "root certificate", and then certificates issued or verified by your organization will be trusted and you will not see this error next time you try to connect to an internal website. Contact your organization's help staff for assistance in adding a new root certificate to your computer.
Mozilla Firefox

This Connection is Untrusted

You have asked Firefox to connect securely to grey-dev.ece.cmu.edu, but we can't confirm that your connection is secure.

Normally, when you try to connect securely, sites will present trusted identification to prove that you are going to the right place. However, this site's identity can't be verified.

What Should I Do?

If you usually connect to this site without problems, this error could mean that someone is trying to impersonate the site, and you shouldn't continue.

Get me out of here!

▸ Technical Details

▸ I Understand the Risks
You have asked Firefox to connect securely to grey-dev.ece.cmu.edu, but we can't confirm that your connection is secure.

Normally, when you try to connect securely, sites will present trusted identification to prove that you are going to the right place. However, this site's identity can't be verified.

What Should I Do?

If you usually connect to this site without problems, this error could mean that someone is trying to impersonate the site, and you shouldn't continue.

Get me out of here!

Technical Details

grey-dev.ece.cmu.edu uses an invalid security certificate.

The certificate is not trusted because it is self-signed.

(Error code: sec_error_untrusted_issuer)

I Understand the Risks

If you understand what's going on, you can tell Firefox to start trusting this site's identification. Even if you trust the site, this error could mean that someone is tampering with your connection.

Don't add an exception unless you know there's a good reason why this site doesn't use trusted identification.

Add Exception...
Discuss Felt et al. 2016

- Coding process
- Scale
  - Not at all to Extremely
- Recruitment
Deploying certs more widely

• EFF’s Let’s Encrypt
  – https://letsencrypt.org/
Online tracking
Online Tracking

• First party = the site you are visiting (whose address is in the URL bar)

• Third party = other sites contacted as a result of your visit to that site

• First-party tracking (e.g., for search)
  – Consider DuckDuckGo and alternatives
Online Behavioral Advertising (OBA)
Let your computer read for you

- Platform for Privacy Preferences (P3P)
- W3C specification for XML privacy policies
  - Proposed 1996
  - Adopted 2002
- Optional P3P compact policy HTTP headers to accompany cookies
- Lacks incentives for adoption
Merrell Primo Chill Slide - Men's Tan: Merrell Shoes
Buy Merrell Primo Chill Slide - Men's Tan and find Spring trends at Onlineshoes. Free Shipping and Exchanges on all Merrell!!...
http://yhs.trafficdashboard.com/track.htm?id=1031...
Privacy Policy - Similar Pages

Merrell "Primo Chill Slide Shoes, Chocolate, Women's
Italian styled winter slide for convenience and warmth. Easy-on and water resistant, the Primo Chill gives your feet after-sport comfort in casual style. Water-resistant pigskin leather upper with sheepskin lining. Removable wool fleece footbed. Injection-molded nylon shank for increased arch support. Air Cushion EVA midsole for softer flex and increased comfort. Merrell Pilot sole with sticky rubber sports a weight-saving design that is siped and barred for traction....
http://clickserve.cc-ct.com/link/ddiprod?lid=41000...
Privacy Policy - Similar Pages

Merrell Primo Chill Slide
http://shopping.yahoo.com/p:Merrell%20Primo%20Chill...
Privacy Policy - Similar Pages

Merrell Primo Chill Slide (Men's)
http://www.shoebuy.com/cgi-bin/abref.cgi?link=yps&...
Privacy Policy - Similar Pages

Merrell Shoes Primo Chill Slide - Men's
Why limit your casual winter footwear wardrobe to unimaginative, straight-laced shoes? Merrell's Primo Chill Slides offer the slip-in convenience of traditional post-sport footwear...
Impact of privacy information on decision making

- Online shopping study conducted at CMU lab
- Paid participants to make online purchases with their own credit cards, exposing their own personal information
- Participants paid fixed amount and told to keep the change – real tradeoff between money and privacy
- Studies demonstrate that when readily accessible and comparable privacy information is presented in search results, many people will pay more for better privacy


http://privacyfinder.org/
P3P in Internet Explorer

• P3P implemented in IE 6, 7, 8, 9, 10 …

• Default privacy setting
  – Rejects third-party cookies without a CP
  – Rejects unsatisfactory third-party cookies
No P3P syntax checking in IE

- IE accepts P3P policies containing bogus tokens or missing required tokens
- Example of valid compact policy:
  - CAO DSP COR CURa ADMa DEVe OUR IND PHY ONL UNI COM NAV INT DEM PRE
- Examples of invalid policies accepted by IE:
  - AMZN
  - Facebook does not have a P3P policy. Learn why here: http://fb.me/p3p

Do not track

- Proposed W3C standard
- User checks a box
- Browser sends “do not track” header to website
- Website stops “tracking”
- W3C working group trying to define what that means
Tools to stop tracking, effective?

- Browser privacy settings
  - Cookie blocking
  - P3P
  - Tracking Protection Lists
  - Do Not Track

- Browser add-ons

- Opt-out cookies

- Digital Advertising Alliance (DAA) AdChoices icon and associated opt-out pages
Existing Privacy Tools
Existing Privacy Tools

Privacy Badger detected 45 potential trackers on this page. These sliders let you control how Privacy Badger handles each one. You shouldn't need to adjust them unless something is broken.

<table>
<thead>
<tr>
<th>Tracker</th>
<th>Block Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>weather.api.cnn.io</td>
<td></td>
</tr>
<tr>
<td>rtax.criteo.com</td>
<td></td>
</tr>
<tr>
<td>ad.doubleclick.net</td>
<td></td>
</tr>
<tr>
<td>googleads.g.doubleclick.net</td>
<td></td>
</tr>
<tr>
<td>securepubads.g.doubleclick.net</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.m.facebook.net">www.m.facebook.net</a></td>
<td></td>
</tr>
</tbody>
</table>

Disable Privacy Badger for This Site

Did Privacy Badger break this site? Let us know!

Donate to EFF

15 Trackers found on www.cnn.com

14 Blocked

<table>
<thead>
<tr>
<th>Tracker</th>
<th>Block Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertising</td>
<td></td>
</tr>
<tr>
<td>Amazon-Associates</td>
<td>✅</td>
</tr>
<tr>
<td>ChartBeat</td>
<td>✅</td>
</tr>
<tr>
<td>Criteo</td>
<td>✅</td>
</tr>
<tr>
<td>DoubleClick</td>
<td>✅</td>
</tr>
<tr>
<td>Google Publisher Tags</td>
<td>✅</td>
</tr>
<tr>
<td>Krux-Digital</td>
<td>✅</td>
</tr>
<tr>
<td>NetRatings SiteCensus</td>
<td>✅</td>
</tr>
<tr>
<td>Outbrain</td>
<td>✅</td>
</tr>
<tr>
<td>Rubicon</td>
<td>✅</td>
</tr>
<tr>
<td>Rubicon</td>
<td>✅</td>
</tr>
<tr>
<td>ShareThrough</td>
<td></td>
</tr>
</tbody>
</table>

Trust Site

Restrict Site

Pause Ghostery

Map These Trackers

Site Analytics

2 Trackers  2 Blocked
Existing Tools’ Connection Graphs
User study results

• Problematic defaults
• Poorly designed interfaces and jargon
• Feedback
• Misconceptions about opt-out tools
• Users unable to make meaningful decisions on a per-company basis

Do people understand OBA + tools?

• Opinions about OBA mixed – both useful and creepy

• Participants did not understand OBA technologies

• Some of the worst fears based on misconceptions

• Participants did not know how to effectively exercise choice

What Do Online Behavioral Advertising Disclosures Communicate to Users?

Pedro Giovanni Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. WPES 2012
The industry claims total success

“The DAA has revolutionized consumer education and choice by delivering a real-time, in-ad notice more than 10 billion times every day through the increasingly ubiquitous DAA Advertising Option Icon (also known as the ‘Ad Choices’ Icon)”

Objectives

• Evaluate the effectiveness of different OBA disclosures at communicating notice and choice about OBA

• Find ways to improve effectiveness of OBA disclosures
Methodology

• Large scale between-subjects online study
  – 1,505 participants
  – Over 100 participants per treatment

• Participants recruited through Amazon Mechanical Turk

• Guided browsing scenario

• Online survey
First exposure to OBA disclosures
Second exposure to OBA disclosures

- Why did I get this ad?
- Interest based ads
- AdChoices
- Sponsor ads
- Learn about your ad choices
- Configure ad preferences
- ‘No tagline’
Do icons and taglines suggest tailored ads?

• To what extent, if any, does this combination of the symbol and phrase, placed on the top right corner of the above ad suggest the following?
  – This ad has been tailored based on websites you have visited in the past. [true]
This ad has been tailored based on websites you have visited in the past.

- Why did I get this ad?: 80% definitely or probably
- Interest based ads: 68% definitely or probably
- Learn about your ad choices: 66% definitely or probably
- Configure ad preferences: 58% definitely or probably
- AdChoices: 58% definitely or probably
- Blank: 34% definitely or probably
- Sponsor Ads: 26% definitely or probably

Colors represent:
- Definitely not
- Probably not
- Not sure
- Probably
- Definitely
Takeaways

• OBA icons and taglines are not noticed
• “AdChoices” was outperformed by other tagline treatments at communicating notice and choice about OBA
• Users are afraid to click on icon
Browser fingerprinting

• Use features of the browser that are relatively unique to your machine
  – Fonts
  – GPU model anti-aliasing (Canvas fingerprinting)
  – User-agent string
  – *(Often not)* IP address *(Why not?)*
Browser fingerprinting

- https://panopticlick.eff.org/