

Web Security

More attacks on Clients:

Privacy



"On the Internet, nobody knows you're a dog."

[Peter Steiner., 1993]

myth



"On the Internet, nobody knows you're a dog."

[Peter Steiner, 1993]

reality





*“Remember when, on the Internet,
nobody knew who you were?”*

Audience poll

- *Is your browser blocking third party cookies?*
- *Do you use an ad blocker in your browser?*
- *Do you use a privacy plugin or a privacy-enhancing browser?*

Privacy threat modelling

Threat modeling = what are we worried about?

1. What are the **privacy/security guarantees** we want ?
Or: what are the bad things we want to prevent?
2. **What can the attacker do?** (aka **attacker modeling**)
 - 2a. What are the **attack vectors**?
What are the **attacker's capabilities**?
 - 2b. Possibly also: What are the **attacker's motivations**?

Privacy threat modelling (1): *What are we worried about?*

Three related but subtly different properties:

- **Privacy**

protecting access to personal information - aka Personal Identifiable Information (PII)

- Many more general characterisations of the notion of privacy are possible

- **Anonymity**

not being identifiable when performing some action

- A way to ensure privacy

- **(Un)linkability**

two or more actions being linkable to the same person

- Weaker property than anonymity: your actions may still be linkable even though you are anonymous

Privacy threat modelling (2): *What is our attacker model?*

1. Network attacker

(passive) **eavesdropper** or (active) **Man-in-the-Middle attacker**

- malicious wifi access point
- compromised router
- your ISP
- lawful interception by government agencies,...

2. Website attackers

- not just website visited, but also (especially) 3rd party content providers

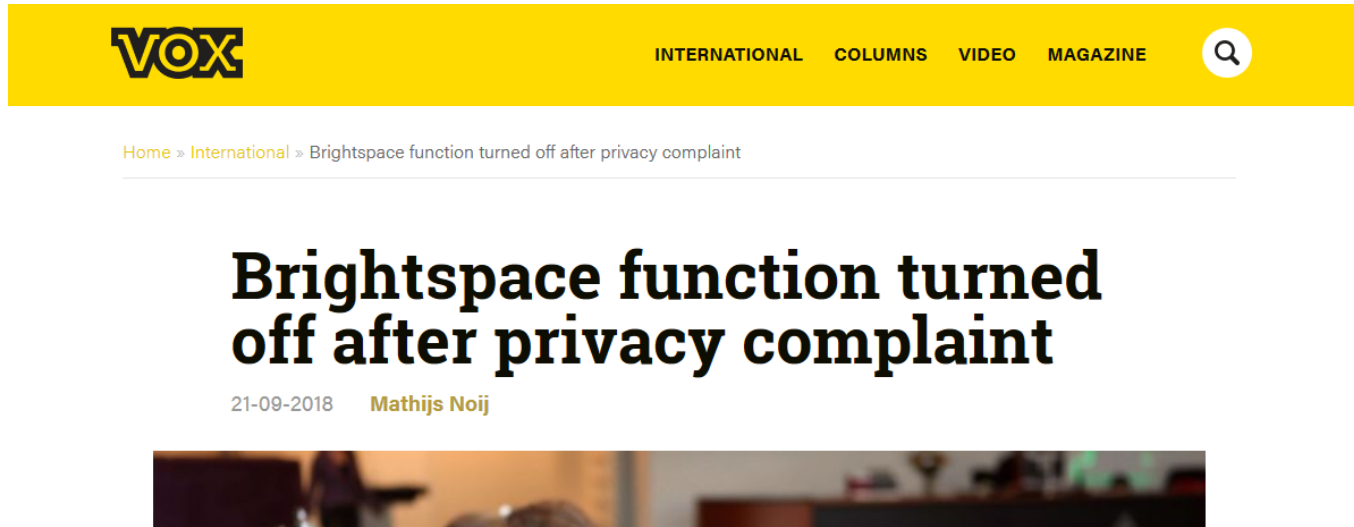
3. Local attackers

someone accessing your device & browser

malware on your device, eg. **browser plugins**

(un)lawful access by government agencies, ...

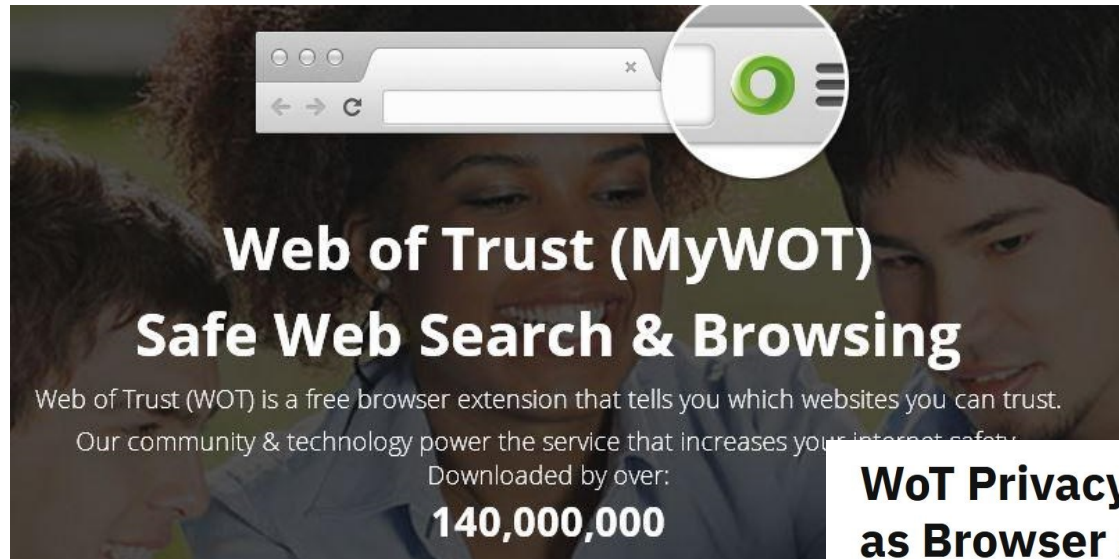
Example server-side threat: Brightspace



“... the university turned off the Class Progress function in Brightspace, at least temporarily. This functionality enabled professors to view all sorts of information at the individual level about the students who are taking their classes: for example, if certain documents have been downloaded, if students have completed certain assignments and, indeed, at what time of day (or night) students logged in.”

<https://www.voxweb.nl/international/brightspace-function-turned-off-after-privacy-complaint>

Example client-side threat: browser plugin



WoT Privacy Breach: Trust Tanks as Browser Add-On Caught Selling User Data

November 10, 2016 @ 10:30 AM

WoT plugin harvested and sold (and even gave away) complete browsing histories.

- **Such histories are typically easy to de-anonymise; e.g. one member of parliament could be identified from the data**

<https://www.ndr.de/nachrichten/netzwelt/Nackt-im-Netz-Millionen-Nutzer-ausgespaehrt,nacktimnetz100.html>

<https://www.kuketz-blog.de/wot-addon-wie-ein-browser-addon-seine-nutzer-ausspaehrt>

<https://youtu.be/1nvYGj7-Lxo>

<https://www.pcworld.com/article/3139814/software/web-of-trust-browser-extensions-yanked-after-proving-untrustworthy.html>

Example local threat: government access

New Zealand can now fine you \$3,200 if you don't hand over your phone password at the border

Bill Bostock ⌚ 03 Oct 2018 💧 39

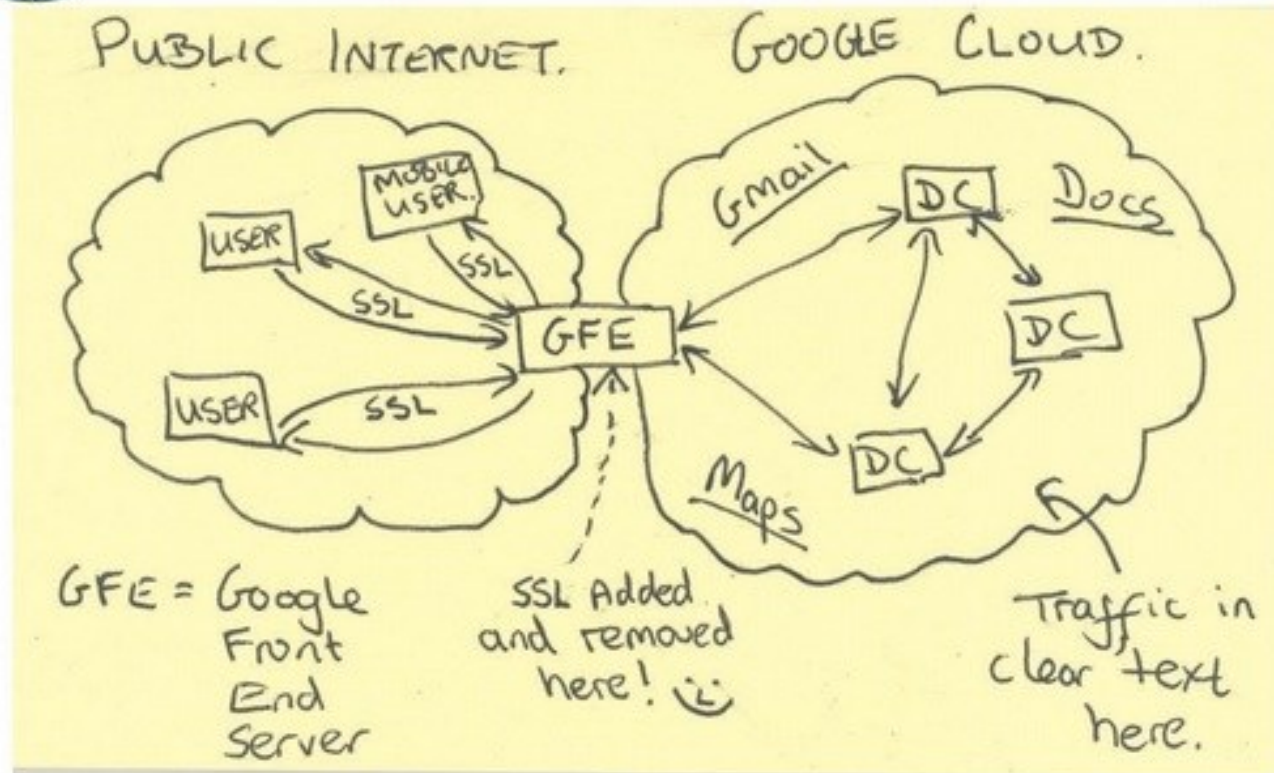


<https://www.businessinsider.nl/new-zealand-border-customs-get-your-phone-password-or-face-3200-fine-2018-10>

Example network threat: NSA

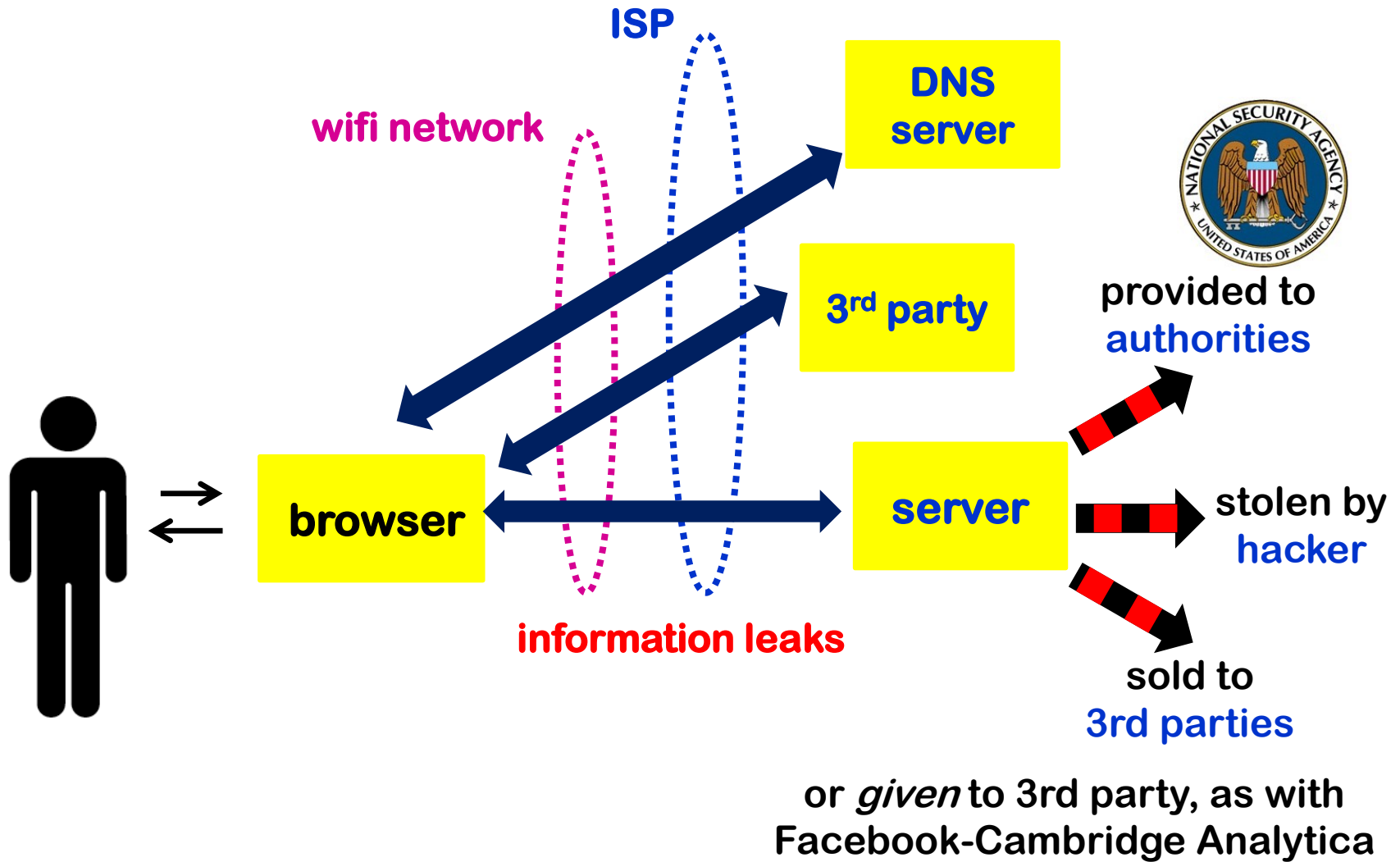


Current Efforts - Google



TOP SECRET//SI//NOFORN

Privacy



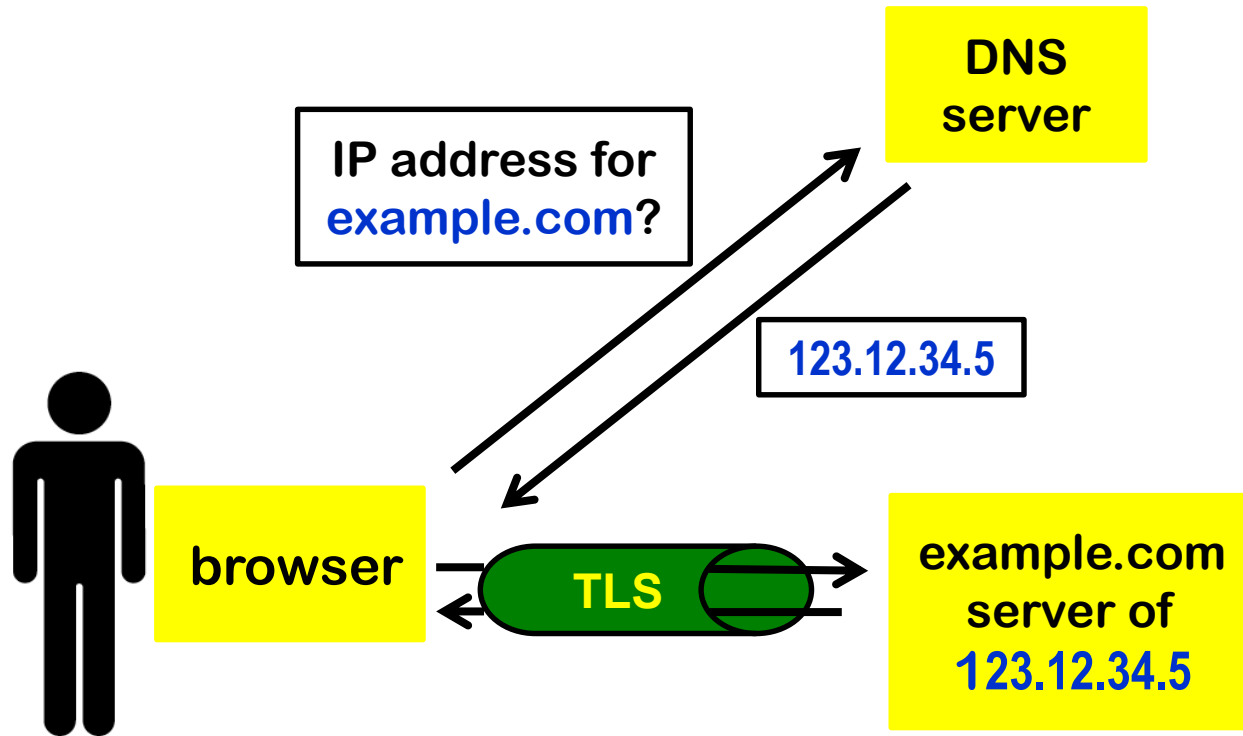
What information?

Possible information leaks include both **content** and **meta-data**

- visits to certain web site
- geographical location
- browser history
- “content”, entered certain data at web site
 - search queries
 - look at certain subpages, topics,...
 - email addresses, email content, telephone number
- all information exchanged
- video & sound via camera and microphone
- ...

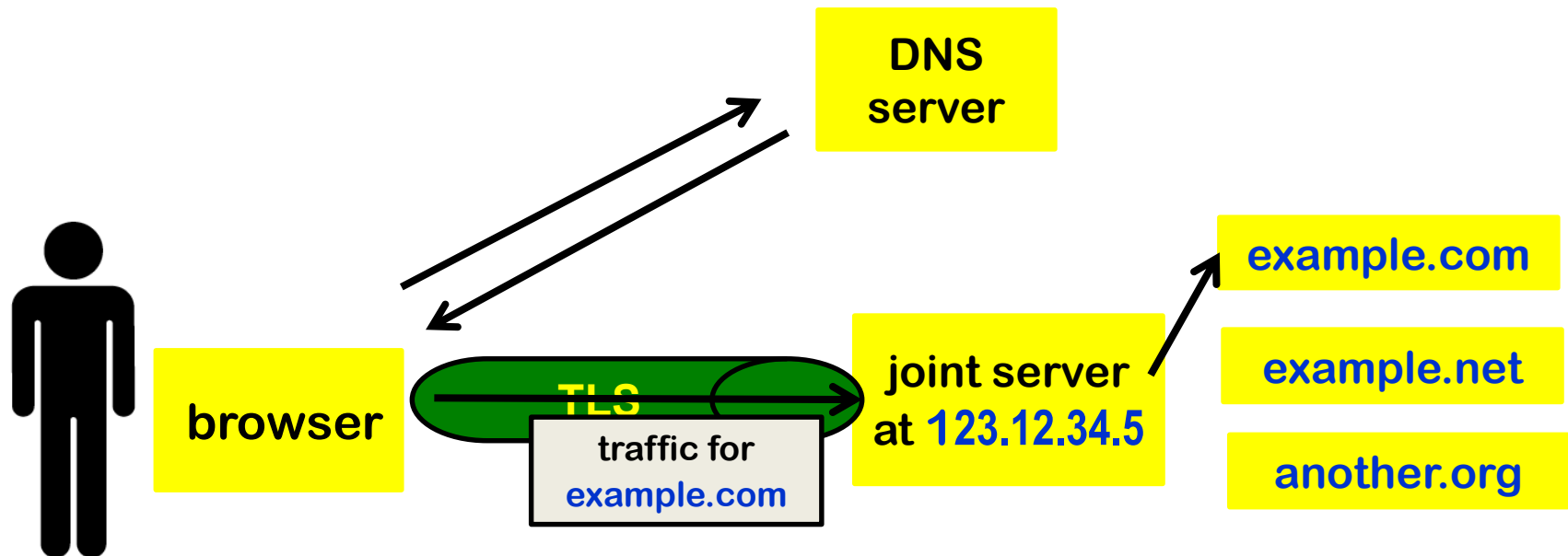
HTTPS information leaks

HTTPS information leaks (2) : DNS



If DNS traffic is in the clear, **network attacker can observe host name.**
Solutions: **DNSSEC, DoH, DoT**

HTTPS information leaks (2) : SNI

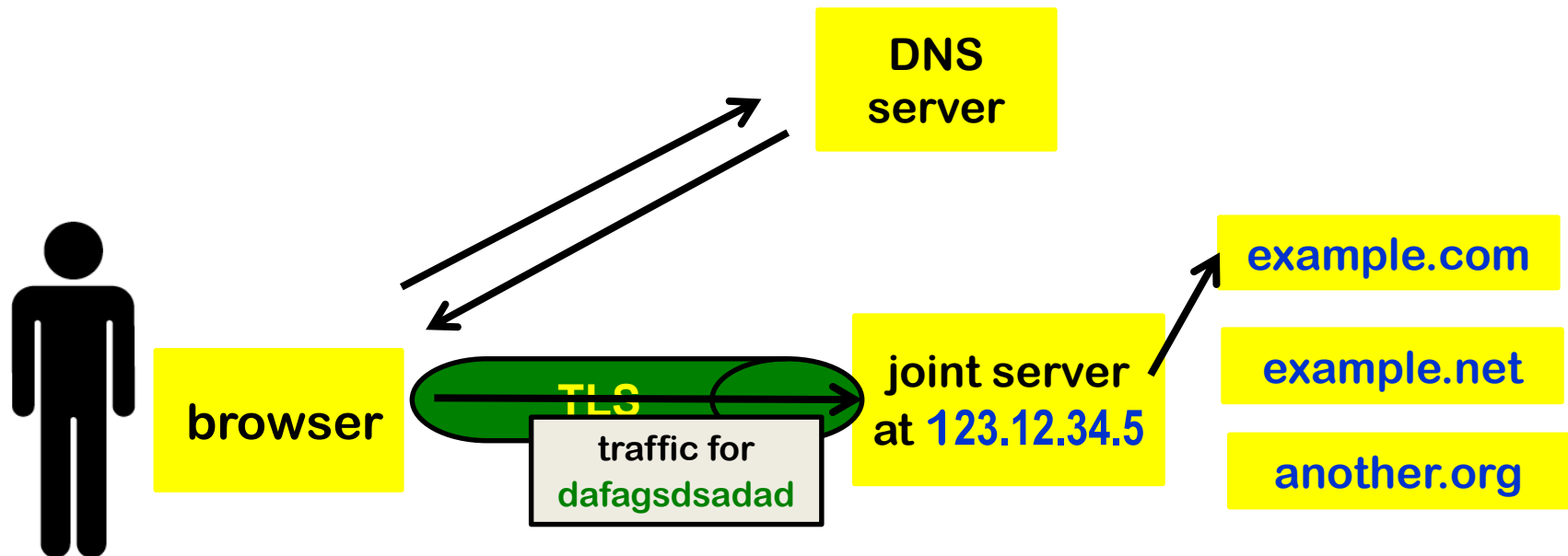


If server hosts several sites at same IP address, traffic has to include **plaintext Server Name Indication (SNI)** so server knows which certificate to use.

Again: **network attacker can observe domain name**

Solution: **ENSI (Encrypted SNI)**

HTTPS information leaks (2) : SNI



With **Encrypted SNI (ENSI)**

a server hosting several sites now has a privacy advantage:

Network attacker cannot observe which site you connect to

Preventing Information leaks in DNS

- **DNS:** protocol to obtain the **IP address** of some **domain name** from a **domain resolver**
 - eg 131.174.123.82 for ru.nl
- DNS traffic is **in the clear & unauthenticated**
 - i.e. no confidentiality & no integrity
 - Your wifi router, ISP and your domain resolver *can see it*
 - These parties can also *fake it*
- Old largely unsuccessful effort to secure it: **DNSSEC**
- More recent efforts to secure it
 - **DoT = DNS over TLS**
 - **DoH = DNS over HTTPS**

Mozilla announced DoH as default in Sept 2019;

Google announced support for DoH for Chrome 78 in Sept 2019

DoH (DNS over HTTPS)

- The good news: for privacy
 - DNS traffic no longer visible to router or ISP
- The (potential) bad news: for privacy, security & functionality
 - DNS traffic not visible to systems administrators, ISP, and parental control providers
 - Change in control

Important difference between DoH and DNS:

Normally the choice of DNS resolver is configured in the OS

With DoH, application (e.g. browser) chooses its DNS resolver

- *Do we want Cloudflare (DNS resolver chosen by Firefox) or Google to see all your DNS queries?*
- *Who do you trust more: your ISP (and national government) or DNS resolver chosen by an app?*

Answers of political activists will vary by country

Some governments don't like privacy measures

China is now blocking all encrypted HTTPS traffic that uses TLS 1.3 and ESNI

The block was put in place at the end of July and is enforced via China's Great Firewall.



By [Catalin Cimpanu](#) for [Zero Day](#) | August 8, 2020 -- 18:04 GMT (19:04 BST) |

Topic: [Security](#)

Russia wants to ban the use of secure protocols such as TLS 1.3, DoH, DoT, ESNI

Amendment to IT law would make it illegal to use encryption protocols that fully hide the traffic's destination.



By [Catalin Cimpanu](#) for [Zero Day](#) | September 22, 2020 -- 12:33 GMT (13:33 BST) |

Topic: [Security](#)

Privacy threat: IP addresses

IP addresses

- Any **eavesdropper on the network** will also see source and destination IP addresses of internet communication
- Any **server** you connect to will see your IP address
 - and probably log it
- IP address usually gives **accurate country & town information**
- In Dutch law, IP address counts as **personal information (persoonsgegeven)**, so processing it is subject to **GDPR** (EU General Data Protection Regulation),
 - in Dutch: **AVG** (Algemene Verordening Gegevensbescherming)

Potential problems
of leaking
your IP address...

NIEUWS

TECH

Friso en Mabel veranderden zelf informatie op Wikipedia

door Bas Benneker 29 aug 2007



'Beeldvorming verdringt de inhoud,' verweet Mabel vorig jaar de media

Prins Friso en prinses Mabel hebben zelf de informatie over Mabel op de Engelstalige Wikipedia-encyclopedie gewijzigd. Dat heeft de Rijksvoorlichtingsdienst (RVD) woensdagavond namens het paar gezegd.

1 0 37

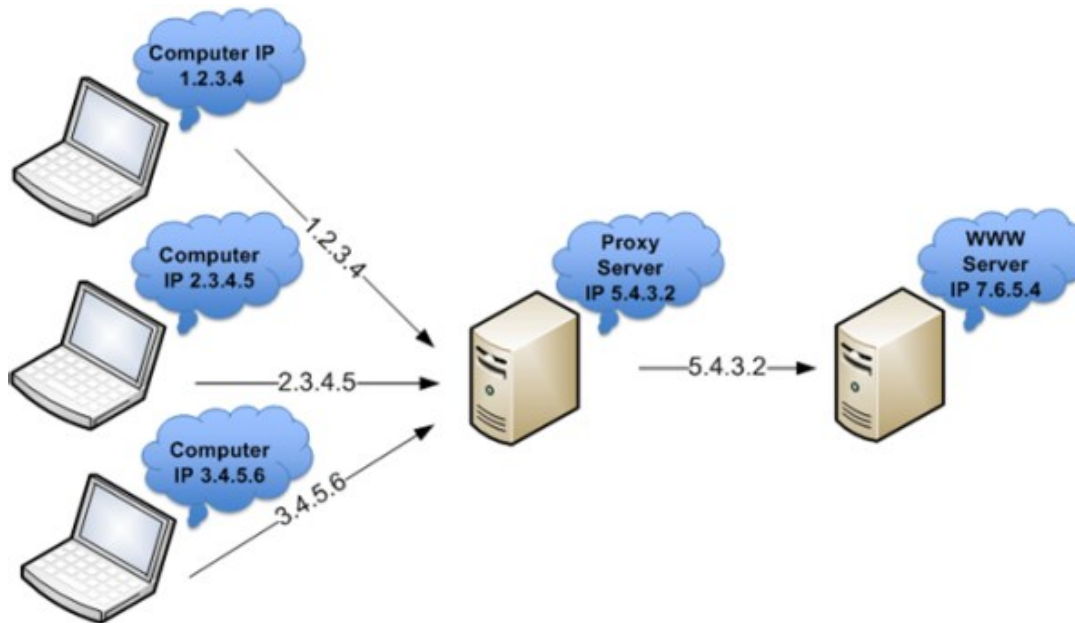


Vanaf het IP-adres van Huis Ten Bosch is informatie in het Wikipedia-artikel over Mabel Wisse Smit gewijzigd om te verdoezelen dat ze heeft gelogen over haar verleden als gangsterliefje.

Datschrift **NRC Handelsblad** woensdag, op basis van gegevens

Using a proxy

- Countermeasure to revealing IP address (and location): set up VPN to a proxy as **intermediary** for internet traffic

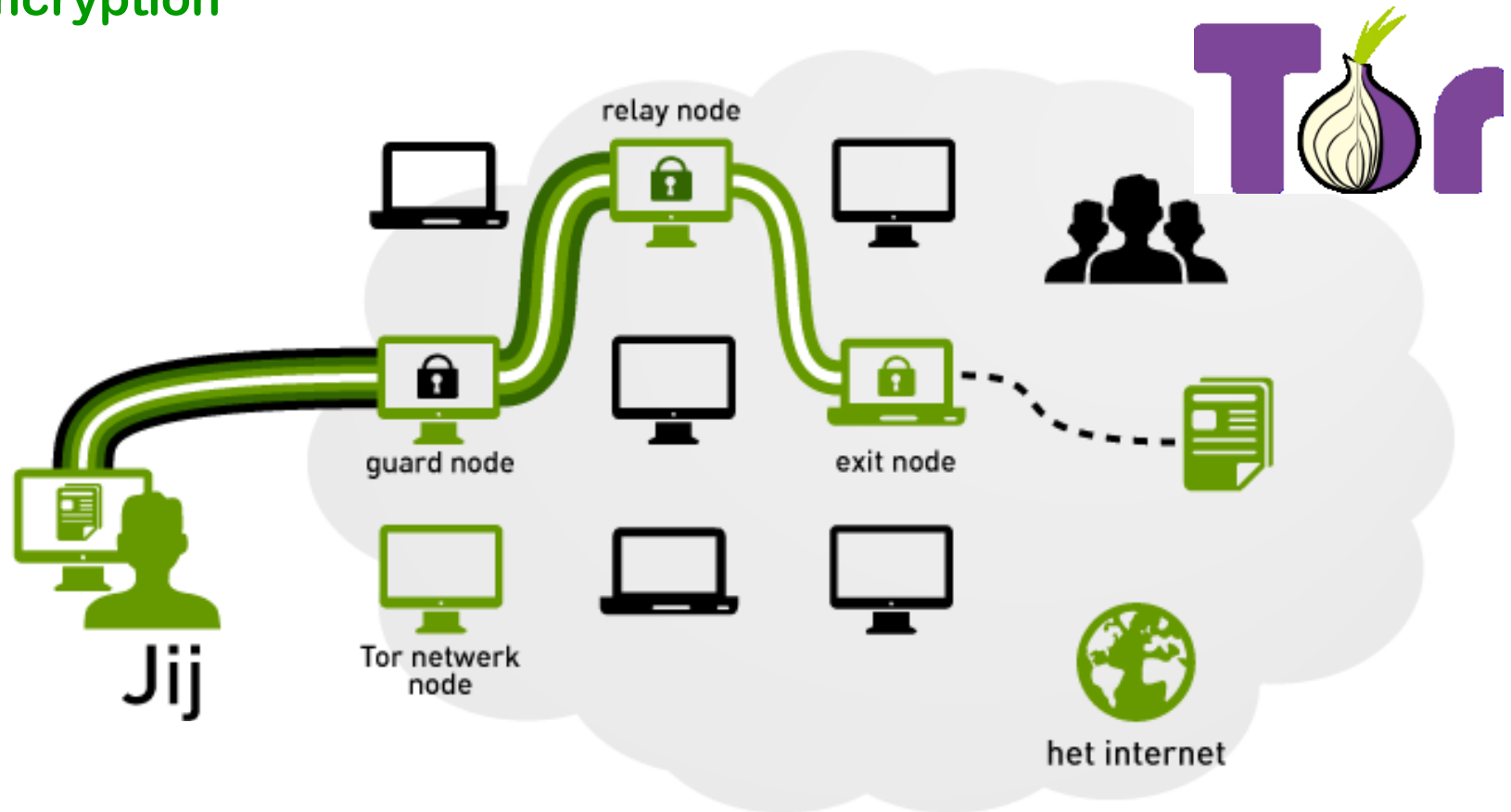


Downside?

You have to trust the proxy!

Countermeasure: Tor

Tor works with layered encryption, with traffic relayed via multiple nodes, with each node peeling off one layer of encryption



Tor

- Tor (The Onion Router) networks aims to provide **anonymity** on the internet:

No single node knows both source & destination IP address

- Started by US Naval Research Laboratory
- **Not immune to all attacks!** eg
 - **traffic analysis** (eg *end-to-end correlation*)
 - **eavesdropping at the exit node**
 - for example using SSL stripping
 - **weaknesses of user's browser or other user actions on that machine**
 - which could still leak IP address
 - for instance, serving a PDF document with a web beacon





Cookies & other ways to track users

Cookies, esp. 3rd party cookies

Most websites will include 3rd party content from eg

- social networks
- advertising networks
- web analytic services (eg google-analytics)
- ...

Borders between categories above are vague/non-existent.


Very little 3rd party content is actually useful to users, apart from google-maps?


Using cookies, these 3rd party web sites can track users across the web

- eg Facebook tracks you on all sites that have a like button

Browser plugins such as Ghostery, LightBeam, Privacy Badger ... provide insight in the large numbers of 3rd parties that are following your browsing!

Example 3rd party content: Facebook Like button

- Facebook tracks *members* across sites with Like or Share buttons 
 - because the Facebook cookie that identifies user is included with all requests to facebook.com
 - Note: this happens *before* the user clicks the Like button.

- Facebook even tracked *non-members* 
 - The Connect button *installed* a cookie, with a life time of 2 years
 - when button is shown, not only after it is clicked
 - If non-member joins facebook later, histories can be linked
 - Similar, if facebook members surf anonymously (for Facebook), because they are not logged on, their browsing can be linked as soon as they do log in

Example 3rd party content: Facebook Like button

- German website heise.de came up with **privacy-friendly two-click Like button**: 1st click downloaded real like button; 2nd click clicked it



- Facebook claimed this violated their copyright, because it used logo's based on Facebook logos

Google Chrome to end support f x +

cnbc.com/2020/01/14/google-chrome-to-end-support-for-third-party-cookies-within-two-years.html

CNBC SIGN IN PRO WATCHLIST MAKE IT ↗ SELECT ↗ SEARCH QUOTES Q

MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV USA INTL

TECH DRIVERS

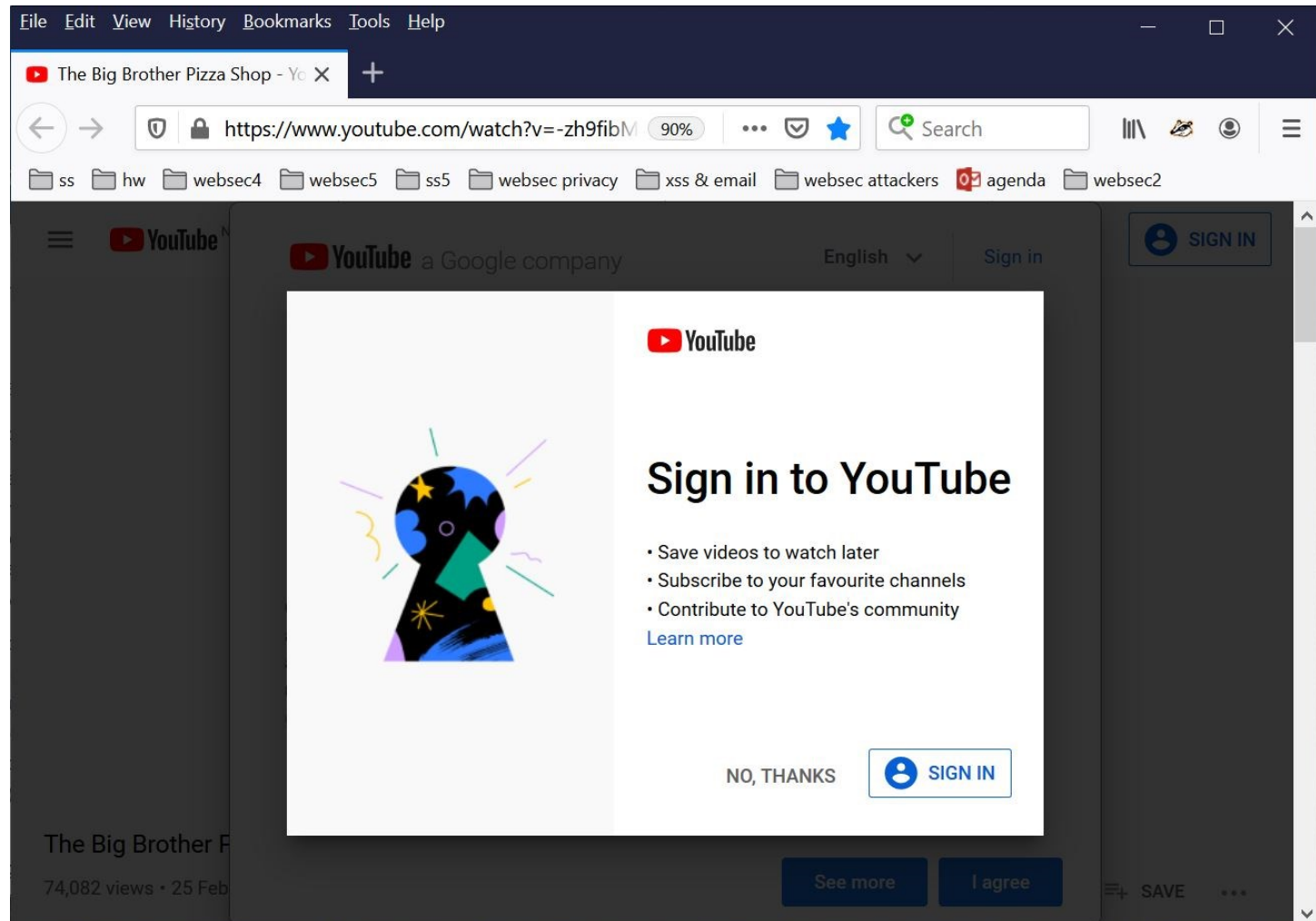
Google plans to kill support for third-party cookies that track you all over the internet

PUBLISHED TUE, JAN 14 2020•12:13 PM EST | UPDATED TUE, JAN 14 2020•4:35 PM EST

 **Megan Graham**
@MEGANCRAHAM

SHARE f t in ✉

Of course, Google likes 1st party cookies 😊



Why tracking & profiling

- targeted content delivery
 - eg to keep people on youtube.com, facebook.com, ...
- targeted advertising
- targeted pricing
 - eg online shop asking higher prices from rich people
or slowly in/decreasing price to see how customers react
- targeted offering of products and services
 - eg online shops *not* offering products to certain people,
say insurance to people in certain neighbourhoods, ...

***if you are not paying for it,
then you are the product being sold***

All 'free' services (gmail, facebook, twitter, WhatsApp..) are paid with ads and collecting personal information for marketing

Even if you do pay, you may still be one of the products ...

- ***Car satnav system maker TomTom sold customer data to police for optimal placement of speed cameras...***

What's going on here?

The screenshot shows the 'Headers' tab in a web browser's developer tools. The 'General' section displays the following information:

- Request URL:** https://ssum-sec.casalemedia.com/usermatchredir?s=183716&cb=https%3A%2F%2Fbeacon.krxd.net%2Fusermatch.gif
- Request Method:** GET
- Status Code:** 302 Moved Temporarily
- Remote Address:** 23.209.176.169:443
- Referrer Policy:** no-referrer-when-downgrade

The 'Response Headers' section shows the following headers:

- Cache-Control:** max-age=0, no-cache, no-store
- Connection:** keep-alive
- Content-Length:** 283
- Content-Type:** text/html; charset=iso-8859-1
- Date:** Fri, 17 Nov 2017 23:11:21 GMT
- Expires:** Fri, 17 Nov 2017 23:11:21 GMT
- Location:** https://beacon.krxd.net/usermatch.gif?partner=casale&partner_uid=Wg9smdHM4WsAAGz0I9oAAAC2&26787
- P3P:** policyref="/w3c/p3p.xml", CP="NOI DSP COR DEVa TAIa OUR BUS UNI"
- Pragma:** no-cache
- Server:** Apache
- Set-Cookie:** CMPRO=787;domain=casalemedia.com;path=/;expires=Thu, 15 Feb 2018 23:11:21 GMT
- Set-Cookie:** CMSC=Wg9smVoPbJkA;domain=casalemedia.com;path=/;expires=Sat, 18 Nov 2017 23:11:21 GMT
- Set-Cookie:** CMSC=Wg9smQ*;domain=casalemedia.com;path=/;
- Set-Cookie:** CMDD=:domain=casalemedia.com;path=/;expires=Sat, 18 Nov 2017 23:11:21 GMT
- Set-Cookie:** CMID=Wg9smdHM4WsAAGz0I9oAAAC2;domain=casalemedia.com;path=/;expires=Sat, 17 Nov 2018 23:11:21 GMT
- Set-Cookie:** CMPS=2425;domain=casalemedia.com;path=/;expires=Thu, 15 Feb 2018 23:11:21 GMT

Blue arrows point to the 'Status Code' and the 'CMID' cookie value.

Cookie value set by casalemedia.com is sent as parameter in request to beacon.krxd.com given via a 302 redirect

Cookie syncing

Cookie of one domain is passed on as parameter to other domain

General

Request URL: https://ssum-sec.casalemedia.com/usermatchredir?s=183716&cb=https%3A%2F%2Fbeacon.krxd.net%2Fusermatch.gif%3Fpartner%3Dcasale%26partner_uid%3D_UID__&C=1

Request Method: GET

Status Code: 302 Moved Temporarily

Remote Address: 23.209.176.169:443

Referrer Policy: no-referrer-when-downgrade

Response Headers

Cache-Control: max-age=0, no-cache, no-store

Connection: keep-alive

Content-Length: 283

Content-Type: text/html; charset=iso-8859-1

Date: Fri, 17 Nov 2017 23:11:21 GMT

Expires: Fri, 17 Nov 2017 23:11:21 GMT

Location: https://beacon.krxd.net/usermatch.gif?partner=casale&partner_uid=Wg9smdHM4WsAAGz0I9oAAAC2_26787

P3P: policyref="/w3c/p3p.xml", CP="NOI DSP COR DEVA TAIa OUR BUS UNI"

Pragma: no-cache

Server: Apache

Set-Cookie: CMID=Wg9smdHM4WsAAGz0I9oAAAC2; domain=casalemedia.com; path=/; expires=Sat, 17 Nov 2018 23:11:21 GMT

Set-Cookie: CMSC=Wg9smQ**;; domain=casalemedia.com; path=;

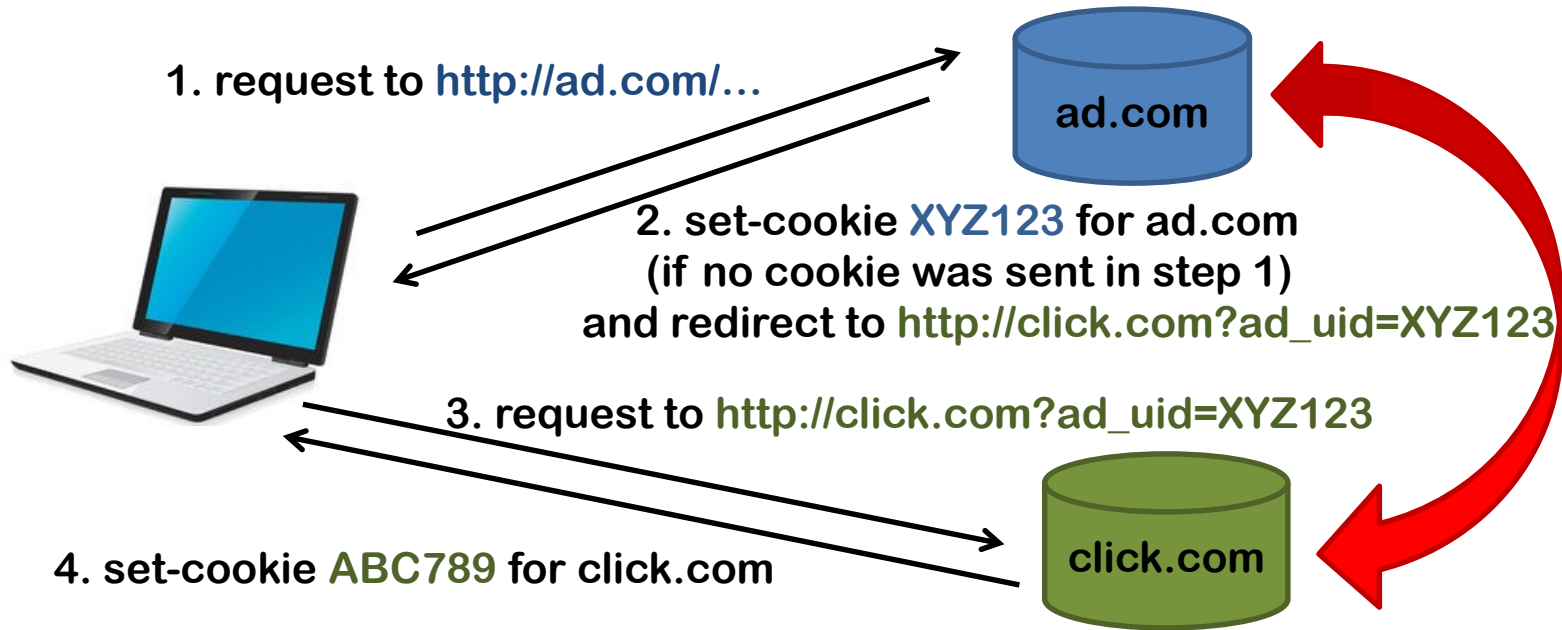
Set-Cookie: CMDD=:domain=casalemedia.com; path=/; expires=Sat, 18 Nov 2017 23:11:21 GMT

Set-Cookie: CMPS=2425; domain=casalemedia.com; path=/; expires=Thu, 15 Feb 2018 23:11:21 GMT

Cookie value set by casaledge.com is sent as parameter in request to beacon.krxd.com, given via a 302 redirect

This allows domains to synchronise cookies, by-passing the SOP restrictions

Cookie syncing



Ad.com and **click.com** can now **exchange (or sell) information** they have on user **XYZ123** and **ABC789**

- This **by-passes the SOP restrictions** that prevent **click.com** from reading **ad.com**'s cookies & v.v.

Websites commonly use (a hash of) a known email address of the user as cookie value, to make matching of user across sites, and even across multiple devices, easier

Countermeasures to tracking by cookies

- **Deleting cookies** regularly
- **Blocking 3rd party cookies**
- **Privacy plugin** (eg **Privacy Badger**, **DisConnect**, **uBlock** or **Ghostery**)
 - These not only block cookies, but they go further, e.g, by
 - blocking all traffic from black-listed sites
 - preventing other tracking mechanisms we discuss later.
- **Use ad blocker**
 - to block 3rd party content (incl. associated cookies)
- Some browsers have built-in support to counter tracking and to support opt-out initiatives like **Do Not Track** (<http://donottrack.us>)
 - using a HTTP header DNT
 - complying with DNT is voluntarily for trackers

Tracking countermeasure: Private Browsing?

- Cookies set in normal windows & tabs mode will not be sent along with requests from private windows
- But: browsing in all private windows & tabs can still be linked! There no isolation *between* private tabs.
 - To confirm this: log into a website in one private tab; open another private tab & check if you are logged on there too
- Further improvements:
 - **Site isolation** in Chrome:
 - Every domain loaded as separate OS process
 - This also ensures (& improves) the Same-Origin Policy (SOP)
<http://www.chromium.org/Homeprocess/chromium-security/site-isolation>
 - **Multi-account containers Add-On** in Firefox:
 - Provides separate private browsing session per account, so interactions with several accounts are segregated from each other
 - This generalizes an earlier Facebook container plugin

By-passing 3rd party cookie blocking?

Suppose browser blocks 3rd party cookie from `tracker.com` on `nu.nl`

This cookie-blocking could be by-passed

1. `nu.nl` redirects user to `tracker.com`
2. `tracker.com` sets 1st party cookie
3. `tracker.com` redirects user back to `nu.nl`

2016 study shows this is used, but rarely.

[Englehard & Narayanan, Online tracking: a 1 million site measurement and analysis, CCS 2016]

Other online tracking techniques

Besides **cookies** there are other techniques to track users

1. Techniques that uses other forms of data stores & exchanged

- **Flash cookies**
- **Web beacons / Tracking pixels**
- **Etags**
- **JavaScript & local storage**
- **Abusing HSTS**



2. Techniques that use inherent behaviour & characteristics of your browser aka **fingerprinting**

- **Canvas fingerprinting**
- **Audio fingerprinting**
- ...



Flash cookies

- aka **LSO (Locally Shared Objects)** or **supercookies**
- Information stored & used by Adobe Flash Player
 - Characteristics
 - stored in hidden folder on the OS file system
 - no expiry date
 - up to 100 Kbyte
 - work across multiple browsers
- Flash cookies have been used to restore deleted HTTP cookies, so-called zombie cookies
- In 2009, 50% of major websites used Flash cookies; Fortunately, Flash is becoming a thing of the past, so this problem should become a thing of the past too

Web beacons

- aka web bug aka tracking pixel
- Invisible 1x1 pixel image included in web page: image will be downloaded from 3rd party when page is accessed
- Can also be used in HTML emails, .docx, PDF, ...
- Web beacons are commonly used in web pages
 - e.g. to gather web statistics
 - if 3rd party cookies are blocked, they cannot directly be used to track visitors *between* websites

Web beacons in email

Web beacons can also be used in HTML email

- HTML emails can include images, but not JavaScript

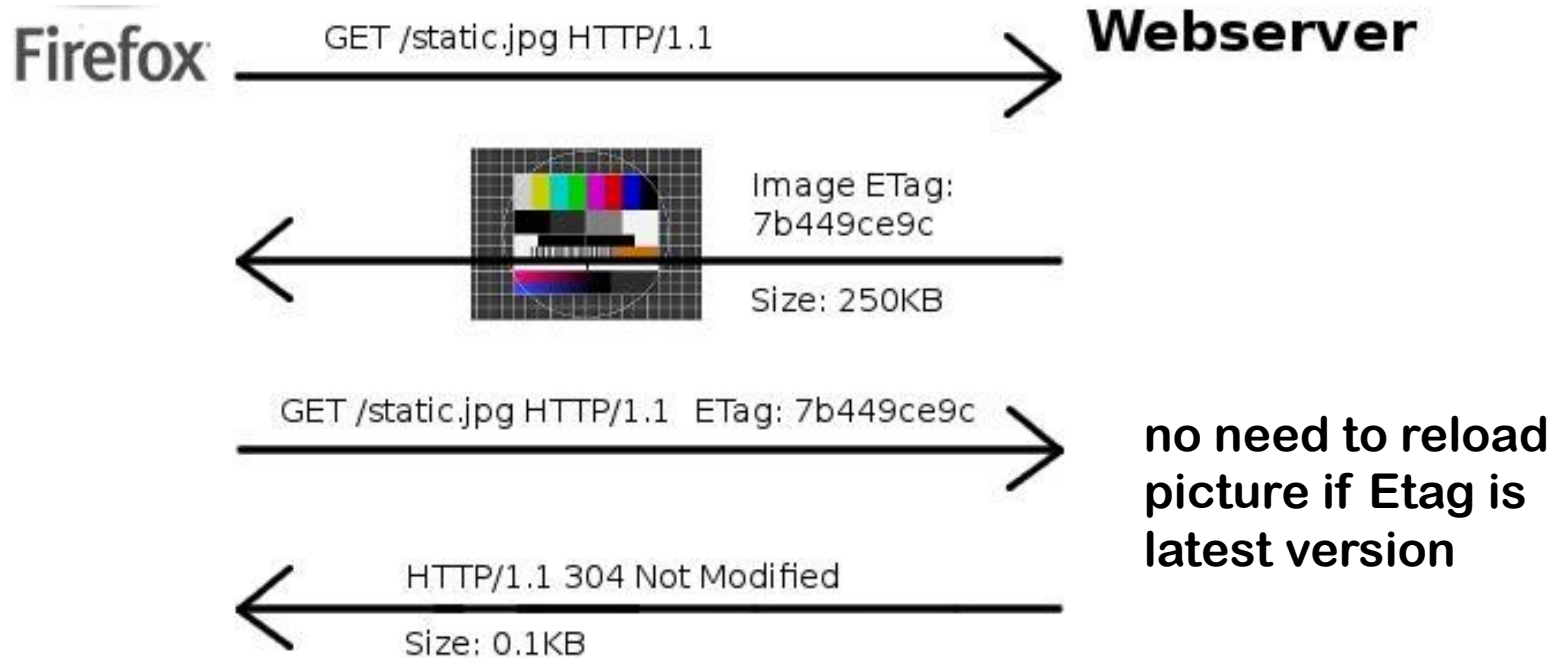
A unique URL for a web beacon in an email allows sender to see

- *when* the email is being read & *from which IP address*
- *if* the email is read; for spammers this is useful, as they can check the email address is real & the email got past the spam filter
- if the email is read in a web browser (aka webmail), then cookies may be sent along too
 - gmail.com will proxy any images in email, so sender does not get user's IP address
- Senders commonly use (a hash of) the email address in the URL for the web beacon

<https://freedom-to-tinker.com/2017/09/28/i-never-signed-up-for-this-privacy-implications-of-email-tracking/>

Cookieless cookies using ETags

ETags (entity tags) are **identifiers added to control caching**



- When browser ask for a resource, it can say which version it already has in its cache, by giving the ETag

This allows server to identify a user by adding **unique Etag**

Try this out at <http://lucb1e.com/rp/cookielesscookies/>

Local Storage & JavaScript

- HTML5 introduced **Local Storage**, where websites can store some info client-side
 - This generalises Flash Local Storage & cookies
- Difference: **cookies are automatically attached to HTTP requests by the browser; data in Local Storage is not.**
- But: JavaScript in a web page can attach info from Local Storage to HTTP requests:
 - gives same functionality as cookies, with a bit more work
- Local Storage is subject to the Same-Origin Policy, just like cookies & other web-content
- Private Browsing model should also treat Local Storage in the same way as cookies

Abusing HSTS?

- HSTS stores persistent information about website in browser
 - not deleted when cookies are deleted
 - but... only **one bit of information for the site**
 - namely, whether HTTPS should be used for that site
- Could HSTS be abused to provide unique fingerprint of a browser?

Yes! a domain could set many HSTS bits for different subdomains
eg `tracker.com` could do this for

`a.tracker.com`

`e.tracker.com`

`g.tracker.com`

`z.tracker.com`

Trying to access all subdomains `a.tracker.com`,
`b.tracker.com`, ... via HTTP, and checking which requests are
turned into HTTPS, reveals the (possible unique) combination of
HSTS bits set

<https://nakedsecurity.sophos.com/2015/02/02/anatomy-of-a-browser-dilemma-how-hsts-supercookies-make-you-choose-between-privacy-or-security>



Fingerprinting

Browser fingerprinting

- Browsers are complex pieces of software that have with many **characteristics**
 - versions, language, OS, screen size, fonts, plugins,...
- These characteristics **leak lots of information**, and **may even uniquely identify a browser**.

See demos at

- <https://panopticklick.eff.org/>
 - <https://amiunique.org/>
- **NB this does not require any storing any info in the client**, unlike cookies, Flash cookies, etc., but uses observable behaviour that already exists
 - hence: harder to counter

Uses of browser fingerprinting (1)

- Re-identify users after they deleted cookies
 - or other explicit info used for tracking

You can erase



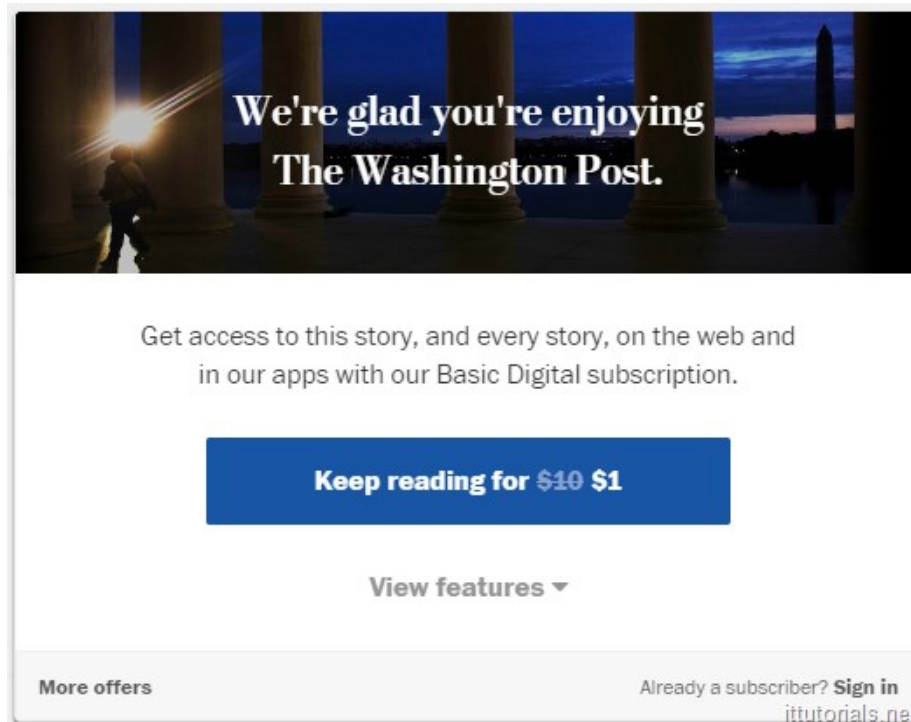
You cannot erase



Uses of browser fingerprinting (1')

Special case of re-identifying users

- **Enforce paywall restrictions**
 - esp. paywalls that are easy to by-pass by clearing cookies



Uses of browser fingerprinting (2)

- Check for suspicious login activity



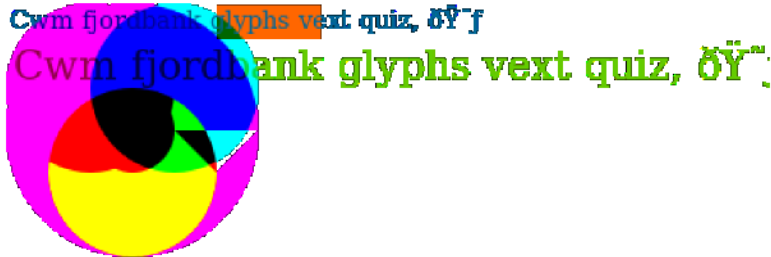
We Detected An
Unusual Login
Attempt

To secure your account, you need to
request help logging in.

Get Help Logging In

Canvas fingerprinting

- Browser is instructed to draw line and/or render text

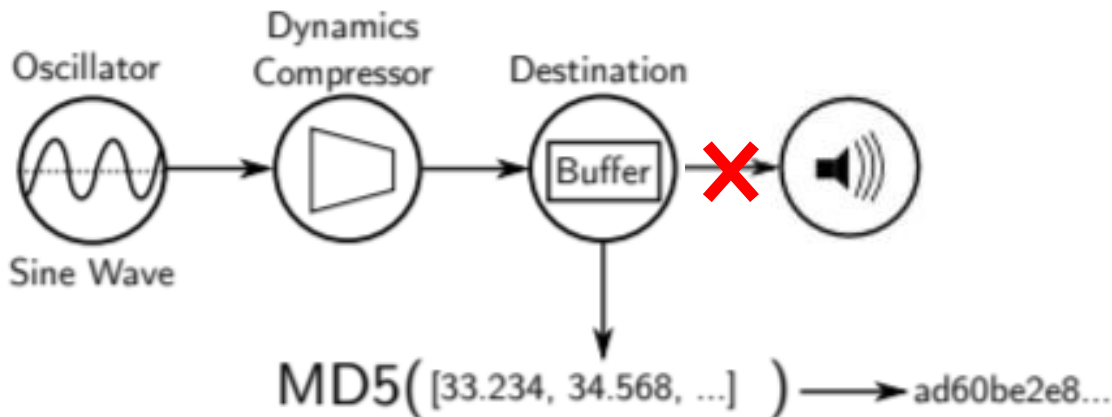


- The result is inspected using the HTML5 Canvas API
- The results depends on the graphics card & fonts loaded, providing a fingerprint of the browser
- Trackers will typically use images invisible to the human eye
- Countermeasures: Privacy plug-ins can stop calls to Canvas API

Audio fingerprinting

Researchers found tracking scripts using the Audio API

- Audio-signal fed through the audio processing software
- Effect observed (without sending it to the speaker) to fingerprint characteristics of the transformation



See <https://audiofingerprint.openwpm.com/>

[Englehard & Narayanan, Online tracking: a 1 million site measurement and analysis, CCS 2016]

Growing possibilities for fingerprinting

Apple declined to implement 16 Web APIs in Safari due to privacy concerns

Apple said these 16 new Web APIs add new user fingerprinting opportunities for online advertisers.

June 28, 2020 -- 16:55 GMT (17:55 BST) | Topic: [Security](#)

- [Web Bluetooth](#) - Allows websites to connect to nearby Bluetooth LE devices.
- [Web MIDI API](#) - Allows websites to enumerate, manipulate and access MIDI devices.
- [Magnetometer API](#) - Allows websites to access data about the local magnetic field around a user
- [Device Memory API](#) - Allows websites to receive the approximate amount of device memory
- [Network Information API](#) - Provides information about the connection a device is using
- [Battery Status API](#) - Allows websites to receive information about the battery status
- [Web Bluetooth Scanning](#) - Allows websites to scan for nearby Bluetooth LE devices.
- [Ambient Light Sensor](#) - Lets websites get the current light level
- [WebHID](#) - Allows websites to retrieve information about locally connected Human Interface Device (HID) devices.
- [Geolocation Sensor](#)
- [User Idle Detection](#) - Lets website know when a user is idle.

<https://www.zdnet.com/article/apple-declined-to-implement-16-web-apis-in-safari-due-to-privacy-concerns/>

The future of tracking ?

- Who will win **the ongoing battle between tracking mechanisms & countermeasures** (in browsers and/or plugins)?
- What will be the **default policies & configurations** of web browsers?
 - eg wrt. **3rd party cookies**
 - Apple's Safari blocks 3rd party cookies by default; Firefox started doing this in June 2019; Google is planning to phase out support
 - Firefox also provides Tracking Protection which blocks traffic to known trackers
- **Apps vs Browser:**
 - If users (have to) use an dedicated app instead of a generic browsers, tracking gets more persistent**
 - You can clear the cookies in your browser, but the app instance on a specific device will always be uniquely identifiable

Example problems

1. Profiling & targeted content on social media
2. AOL search data
3. Abusing Facebook targeted advertising
4. Leaking browser history

1) Profiling & targeting enable by tracking



2. AOL data leak

- In 2006, AOL released 2 Gbyte of anonymised search data for research purposes
 - 20 million search queries for over 650,000 users over a 3-month period
- Research then quickly could identify some users, because the search queries contained personally identifying information.
- It also revealed some amusing, sad, and highly disturbing search histories of individuals.

Moral of the story: **anonymisation is hard!**

<https://www.nytimes.com/2006/08/09/technology/09aol.html>

<https://www.cnet.com/news/aols-disturbing-glimpse-into-users-lives/>

3. Abusing Facebook targeted advertising

- Facebook provides a rich API to target ads to some specific audience on the Facebook site or app.
- You can create
 1. **tracking pixels audiences**

You obtain a tracking pixel from Facebook, say
http://facebook.com/tracking_pixel_for_IPC026.jpg
You add this to your website & Facebook tracks who retrieves this.
 2. **custom audiences**

defined by lists of attributes, such as age or location, or even Personally Identifiable Information (PII), such as email addresses and phone numbers
- Facebook will can only serve ads to audiences of at least 20 people.
- Google & Twitter have similar functionality

Facebook Custom Audiences

Attributes you can use in lists to define audiences:

- **Email address**
- **Phone number**
- **Mobile advertiser IDs**
 - identifier provided by mobile OS; unique per device, but can be reset
- **Name** (First name, Last name)
- **Age** (Date of birth, Year of birth, Age)
- **Sex**
- **Location** (Postal code, City, State/Province, Country)
- **Facebook app user ID & page user ID** - obfuscated versions of user ID

Abusing Facebook custom audiences

- Facebook will tell you the rough size of audiences you create
 - rounded to nearest 10, 100, 1000, 10,000, etc.
- Facebook removes duplicates in audiences

– Eg



jan@student.ru.nl
kees@student.ru.nl
06 - 1245 5673
06 - 3652 1245
06 - 6792 1236

Suppose all these identifiers occurs in Facebook database.

What can the size of this custom audience be ?

- 3, 4 or 5, depending on whether these numbers include the number of Jan and/or Kees

Of course, if some of these identifiers do not occur in the database, the size could be lower.

06-1234 7586
06-3446 9763
...
...
06-5677 6572

rounded size
= 560

06-1234 7586
06-3446 9763
...
...
06-5677 6572
06-7666 7666

rounded size
= 570

This means 06-7666 7666 is in Facebook's database,
and audience size of the list on the left is exactly 564

06-1234 7586
06-3446 9763
...
...
06-5677 6572
jan@student.ru.nl

If size is 560, then Jan's phone number is in the list.
If size is 570, it is not.

Attack: finding phone number for an email address

Make a lists of all Dutch mobile phone numbers starting with a 1
Add some French numbers until you find the place where rounding happens

06-1000 0000
06-1000 0001
...
...
06-1999 9999
+33- 1...
+33- 2...
+33- 3...

rounded size
= 792,000

06-1000 0000
06-1000 0001
...
...
06-1999 9999
+33- 1...
+33- 2...
+33- 3...
+33- 4....

rounded size
= 793,000

Now use the list on the left to find out if first digit in Jan's number is a 1

Repeating this procedure, you find out the whole number.

Abusing Facebook Custom Audiences

- Possible attacks:
 1. find the phone number corresponding to an email address
 - in about 20 minutes
 2. using a Facebook tracking pixel on a website,
 - a) determine if a specific user visited a website
 - b) de-anonymise all visitors to a website
- Attacks only require interaction with Facebook's APIs.
Victim is not involved, beyond being fed the tracking pixel in 2a & 2b
- Countermeasure: Facebook no longer provides size estimates provided for lists containing more than one type of PII

[Privacy Risks with Facebook's PII-based Targeting: Auditing a Data Broker's Advertising Interface, IEEE S&P 2018]

<https://www.youtube.com/watch?v=Lp-lwYvxGpk>