Web Privacy

194.144 Privacy-Enhancing Technologies

Dr. Markus Donko-Huber

Outline

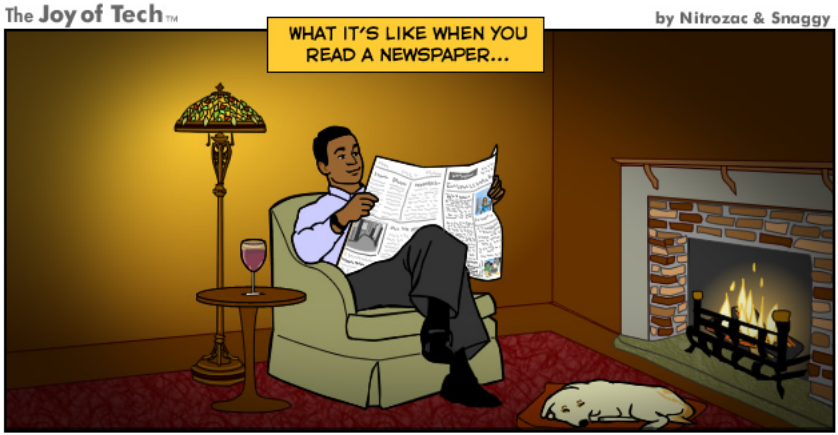
Network Layer Tracking

Web Tracking

Tracking Protection

Beyond the Browser

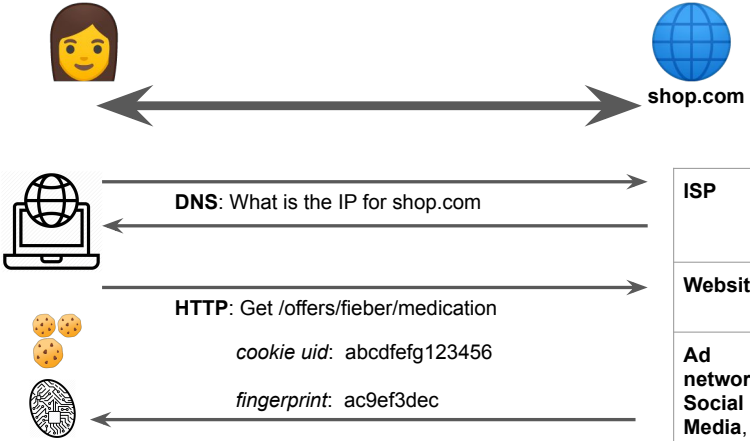
Conclusio


The Joy of Tech ™
by Nitrozac & Snaggy
WHAT IT'S LIKE WHEN YOU READ A NEWSPAPER...

© 2014 Geek Culture

joyoftech.com

Network Layer Tracking



ISP	Domains visited (+ site content) by specific customer
Website	Site content visited by specific user
Ad networks, Social Media, ...	Content on different Websites visited by specific User [web tracking]

Figure: Online requests and tracking

Domain Name Service (DNS) Leaks

- DNS is a plaintext protocol (UDP port 53)
  - Requests are **visible** within same WiFi, to ISP, and in transit
  - **DNS hijacking**
    censorship, injecting advertising[1], attacks
- Monitoring independent of DNS provider with deep-packet inspection for Public DNS (e.g. 8.8.8.8), recursive DNS resolver.

	Time	Source	Destination	Protocol	Length	Info
3	0.038043164	192.168.10.254	192.168.10.44	DNS	243	Standard query response 0xaf2d A pbs.twimg.com CNAME cs196.wac.edgecastcdn.net CNAM
4	0.047955804	192.168.10.254	192.168.10.44	DNS	264	Standard query response 0x5794 AAAA pbs.twimg.com CNAME cs196.wac.edgecastcdn.net C
5	1.583696907	192.168.10.44	192.168.10.254	DNS	75	Standard query 0x091e AAAA ssl.gstatic.com
6	1.620240805	192.168.10.254	192.168.10.44	DNS	103	Standard query response 0x091e AAAA ssl.gstatic.com AAAA 2a00:1450:400d:802::2003
7	14.523565820	192.168.10.44	192.168.10.254	DNS	72	Standard query 0x36e4 A tuwien.ac.at
8	14.523672571	192.168.10.44	192.168.10.254	DNS	72	Standard query 0x316d AAAA tuwien.ac.at
9	14.553388078	192.168.10.254	192.168.10.44	DNS	88	Standard query response 0x36e4 A tuwien.ac.at A 128.130.35.76
10	14.565765898	192.168.10.254	192.168.10.44	DNS	132	Standard query response 0x316d AAAA tuwien.ac.at SOA kira.kom.tuwien.ac.at
11	18.799938290	192.168.10.44	192.168.10.254	DNS	73	Standard query 0x071e A www.tuwien.at
12	18.800209989	192.168.10.44	192.168.10.254	DNS	73	Standard query 0xbc6b AAAA www.tuwien.at
13	18.835304040	192.168.10.254	192.168.10.44	DNS	142	Standard query response 0x071e A www.tuwien.at CNAME www.tuwien.ac.at CNAME info.z
14	18.848416045	192.168.10.254	192.168.10.44	DNS	154	Standard query response 0xbc6b AAAA www.tuwien.at CNAME www.tuwien.ac.at CNAME info

[1] DNS Manipulation by ISPs

Encrypted DNS

- Popular standards to encrypt DNS (covered in TLS lecture)
  - DNS over TLS (DoT) from 2016[2]
  - DNS Queries over HTTPS (DoH) from 2018[3]
- **DoT**: DNS wrapped with TLS (new port tcp/853)
  - Potential issue: blocking / detection
  - Android $\geq$ 9 ("Private DNS"), $\geq$ iOS 14, Windows 10, `systemd`
- **DoH**: HTTPS for transporting DNS queries
  - HTTPS is commonly used for web services / browser APIs
  - Supported by major web browsers and mobile apps available

[2] rfc7858

[3] rfc8484

Discussions around encrypted DNS

- ISPA (UK) criticized Mozilla and Google for adapting DoH[4]
  - "Internet Villain award 2019" for undermining blocking efforts
  - Blocking+monitoring still possible
    (HTTP, based on HTTPS certificate)
- **Mozilla defaults to CloudFlare**'s DNS when enabling DoH
  - CloudFlare can link requests to source IP / user agents
- Our Recommendations:
  - `https://appliedprivacy.net/services/dns/`
  - `https://docs.usableprivacy.com/dns/`

[4] UK ISP group names Mozilla 'Internet Villain' for supporting 'DNS-over-HTTPS'

Network HTTP(S) Leaks

- Unencrypted HTTP
  - Website(s) requested **http://shop.com/offers/fieber/medication**
  - Entire Page-content including authentication tokens etc.
  - **Straightforward to monitor** with transparent HTTP proxies
- HTTPS (covered in detail in the TLS lecture)
  - **Hostname leaks** in initial TLS handshake
  - Deep packet inspection can be used to monitor / censor HTTPS
  - https://**shop.com**/offers/fieber/mediacation

Web Tracking

Web Tracking

- Web Tracking = Creation of unique user profiles
- First parties
  - websites
  - mobile applications
- **Third parties**
  - advertisement
  - analytic providers
  - online social networks
- Trackers link people to sensitive information
  - health related, location information ...

Third Parties

Online Advertisement

- Direct sales
  - links to products on websites / on social media (usually no tracking by third parties)
- **Ad networks**: place ads on multiple websites, targeting ads based on:
  - demographics (age, gender, etc.)
  - location based
  - website content (z.b. car-related)
  - user profiles (history of visited websites, customer loyalty programs [Datalogix])
- Ad exchanges
  - Auction of available advertisement spaces
  - Also sell customer information

Targeted Advertisement

Figure: Dani Kapp #creepy #epicfail

- Ad networks collect sensitive information
- Insurances, credit scores etc.

Targeted Advertisement II

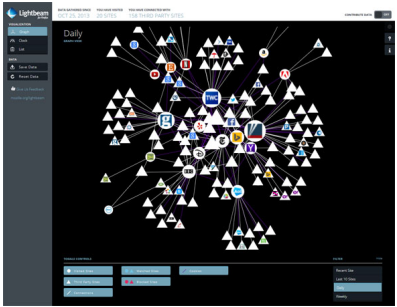
Figure: Minor attribution error ;-) #noflightlist #mistake

Social Networks and CDNs

- Social Plugins aka. “share buttons”
  - Single-Sign-On functionality
  - Twitter Tweet, Facebook Like buttons
  - ShareThis, AddThis $\rightarrow$ collect additional user information
- Content Provider
  - Youtube, Vimeo, Maps, etc.
- JavaScript libraries and fonts
  - Twitter Bootstrap, Google Fonts, vue/react etc.
- Webhoster
  - Akamai, Wordpress.com, ...

Insights with (Mozilla) Lightbeam

- Browser extension for Mozilla Firefox
  - https://addons.mozilla.org/en-US/firefox/addon/lightbeam-chikl/
- Visualization of relation between third-parties



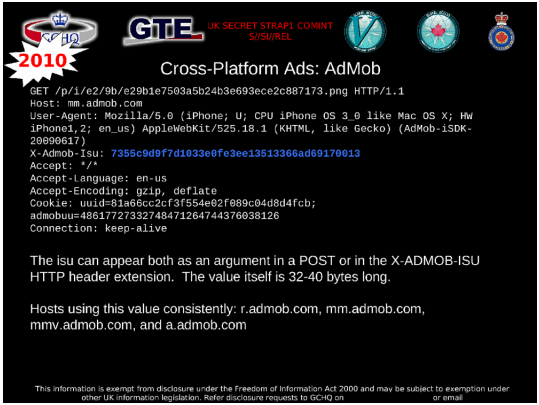
Web Tracking Risks

Scope of tracking

- Third-party is also first party
  - e.g. Users linked across web by Facebook Like button
- First party sells user data
  - personal information is directly sold to e.g. ad networks
- Unintentional sharing of personal information
  - E.g. URI: `http://www.onlinedating.com?profile=markus&gender=male&likes=linux`
- Misuse of security bugs
  - E.g.: XSS, clickjacking, history stealing
- Cross service linking
  - E.g. match users by their profile pictures (Facebook)

Governmental organizations

- NSA piggybacks on Cookies / UUID
- De-Anonymization of Tor users
- Target selection for exploitation



Tracking Technologies

Tracking Technologies

- Tracking via third-party libraries
  - Visited URL is leaked via Referer or submitted directly
- User profiles (traditional): HTTP Tracking Cookie
  - Unique cookie which is set on initial loading of websites
- Supercookies
  - Multitude of storage location for user identifier except HTTP cookie
- Fingerprinting
  - Tracking via unique OS/browser properties

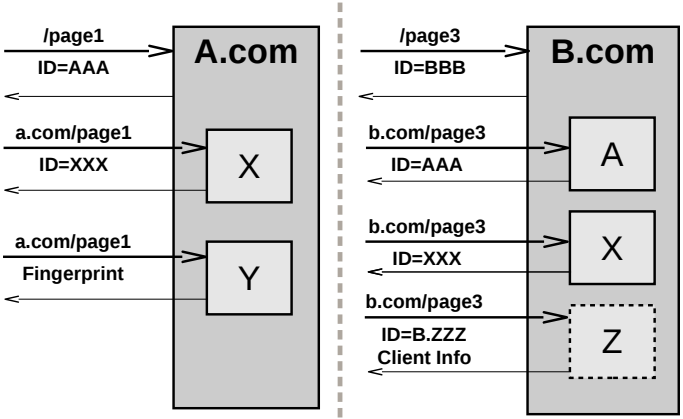


Figure: **A** … First- and Third-party (e.g. Facebook), **X** … advertisement network (e.g. doubleclick), **Y** … uses fingerprints instead of cookies, **Z** … analytics service (e.g. Google Analytics)

Supercookies

- Use alternative storage locations (*cmp. table)*
- e.g. Adobe Flash Cookies
  - http://www.macromedia.com/support/documentation/en/flashplayer/help/settings_manager07.html
- Cookie Resyncing
  - Cookie is restored from one of the many supercookie storage locations
- Evercookie

  https://samy.pl/evercookie/

(a) "Supercookies"

HTTP authentication† [84]
HTTP caching ("cache cookies")
cache control
ETags* ("ETag cookies") [85]
Last-Modified [85] (e.g. [86])
cache content
resource (e.g. JavaScript, HTML, CSS, or media)*
status code
redirect location (e.g. [87])
hits and misses (e.g. [88])
TLS/SSL session ID [89]
browsing history††
userData storage (Internet Explorer only)*
HTML5 storage (session, local, and global)*
HTML5 protocol handlers†
HTML5 content handlers†
W3C geolocation API permission†
`window.name` property* (session only)
HTTP strict transport security [90]
plug-in storage* (e.g. Flash local shared objects, or "Flash cookies")
DNS cache

∗ Observed in use by a third-party website.
† User intervention required.
†† Largely inaccessible in newer browsers, but see [88], [91].

Figure: Supercookie methods[a]

[a] "Third-party web tracking: Policy and technology." 2012 IEEE SP.

Fingerprinting

- Persistent tracking of users without cookies
- Based on unique system properties
- https://panopticlick.eff.org/

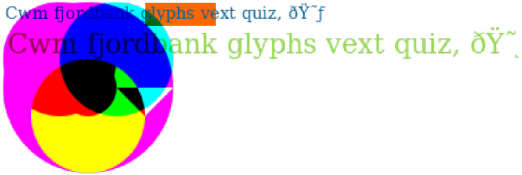
(b) Active "Fingerprinting"
operating system
CPU type
user agent
time zone
clock skew
display settings
installed fonts
installed plugins
enabled plugins
supported MIME types
cookies enabled
third-party cookies enabled

(c) Passive "Fingerprinting"
IP address
operating system
user agent
language
HTTP accept headers

Figure: Fingerprinting sources[5]

Canvas Fingerprinting

- HTML5 Canvas
- Differences between graphic-cards (drivers), OS etc.
- Research from 2022: fingerprinting identical hardware / software stacks[6]



[6] DRAWNAPART: A Device Identification Technique based on Remote GPU Fingerprinting

Tracking Protection

Protection against Tracking

- Website providers (you!)
  - Same-origin policy (dedicated websites for tracking)
  - AnonymizeIp (Google Analytics) or e.g. Matomo
  - Alternatives to standard social plug-ins
- Opt-out
  - Privacy initiatives by industry
  - Opt-out = no targeted advertisement
  - **But trust issue: how is data handled?**
- **Browser + settings / extensions**
  - Settings and features in state-of-the-art browsers
  - Special browser extensions

Opt-out initiatives - industry self regulation

- Special websites to set opt-out cookies
  - http://www.aboutads.info/choices/
  - Issues: validity of cookies, deletion of cookies, trust
- Browser extensions for persistent opt-out cookies
  - TACO, Keep My Opt-Outs
- Do Not Track (DNT) HTTP Header
  - It is up to websites to the honor DNT header or not
  - Was enabled by default (Firefox, IE 10) therefore also ignored
- Global Privacy Control (GPC) HTTP Header
  - Recent alternative to DNT header: https://globalprivacycontrol.org/
  - Feature available in Brave and Firefox

Browsers

Google Chrome

- Advanced security measures (e.g. site isolation)
- Google ad revenue = Chrome is bad for privacy
- First-party tracking across Google products (e.g. search when you logged in to Gmail)

Apple Safari

- Intelligent Tracking Prevention 2.3[7]
  - Separate context for third-party cookies
  - Purging of third-party cookies after 30 days
  - First-party cookies are purged after 7 days

[7] https://webkit.org/blog/9521/intelligent-tracking-prevention-2-3/

Browsers cont.

Mozilla Firefox

- Tracking prevention based on `Disconnect` ruleset
- Enhanced tracking prevention (separate cookie context)
- Multi-Account **Containers** (e.g. seperate “online life”)

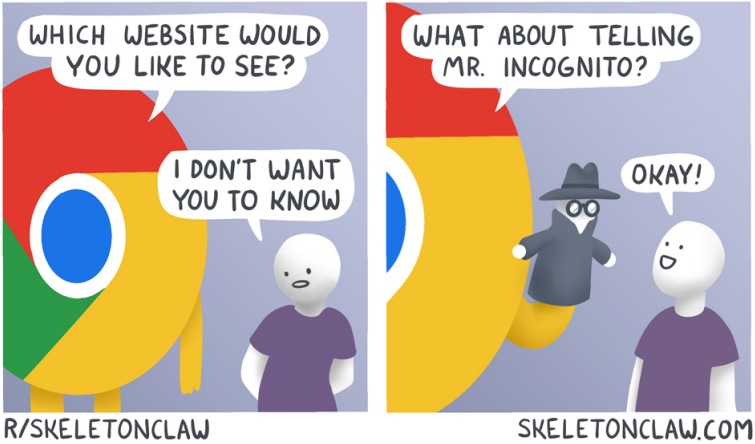
Brave

- Tracking and fingerprinting protection
- Tor-browser tabs (improvement over private mode)
- “Brave Rewards”: privacy-respecting ad ecosystem

Browser-Settings

- Deletion of cookies, cache
  - Manual or once browser is closed
  - Supercookies will survive this setting
  - Loss of settings &active sessions
- Do Not Track / GPC Header
  - Supported by major web browsers
  - Default in Firefox (enable it in other browsers)
- Third-party cookies
  - Can be completely blocked (only default in Safari)
- Private mode
  - No data is locally stored (history, cookies, etc.)

Browser Incognito Mode



Browser-Extensions

- AdBlock Plus (ABP)
  - Was the most popular extension to block online advertisement
  - Advertisement is blocked and set invisible (CSS)
  - Issue: Since 2012 ”Acceptable Ads”
  - Issue: Acceptable Ads are enabled by default
- Ghostery
  - Detection and blocking of web trackers
  - Overlay for social plug-ins
  - Issue: Usability
  - Issue: Business model (now owned by Cliqz)

Browser-Extensions II

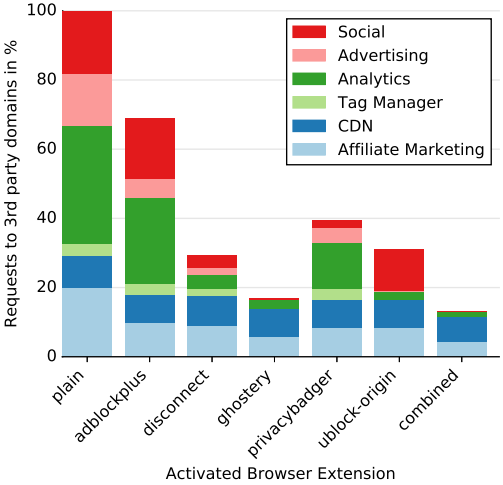
- EFF Privacy Badger
  - Based on heuristics
  - Tests if DNT header is honored
  - Challenge: Maintain whitelist
  - Overlays for social plug-ins
- Disconnect.me
  - Similar to Ghostery but with open-source Ruleset
  - VPN service for mobile devices (paid subscription)
  - Basis for tracker blocking in Firefox private mode
- uBlock (origin)
  - Open-Source “wide-spectrum” blocker
  - Focus on performance
  - Challenge: Overblocking, filterrule maintance

Effectiveness of browser extensions

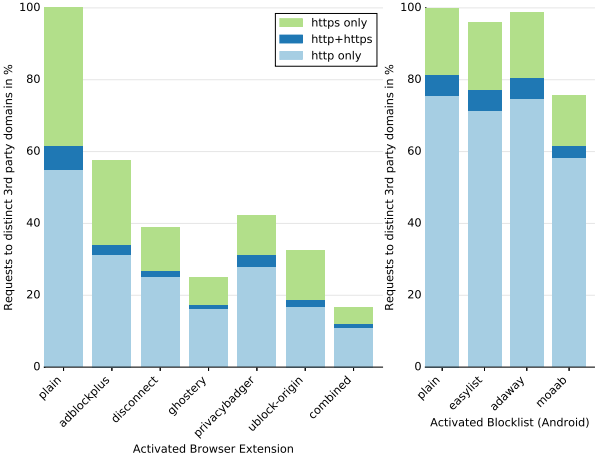
Our Research on Browser Extensions

- **How effective are these browser extensions?**
- Presented at Euro Security and Privacy 2017
- Analysis of Alexa Top 200K websites (0.5 billion requests)
  - `CRAWLIUM` framework
  - Analysis with 7 different browser profiles in $< 12h$
- Analysis of 10,000 Android Apps
- Joined work with Georg Merzdovnik (SBA Research), Nick Nikiforakis (Stonybrook)

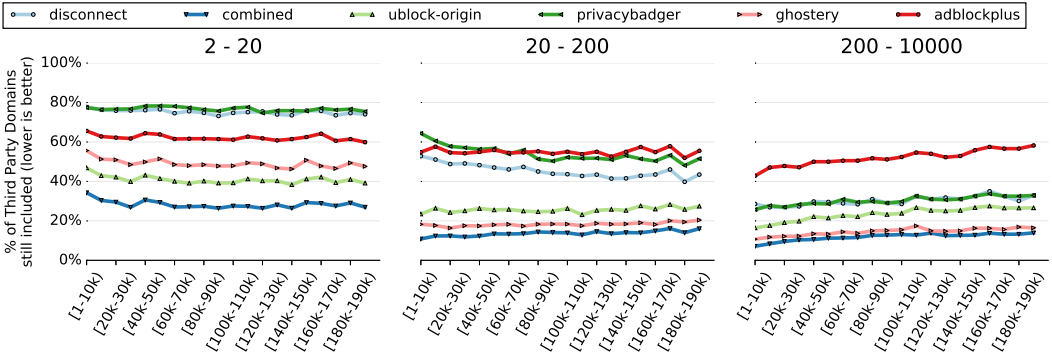
Blocked Trackers



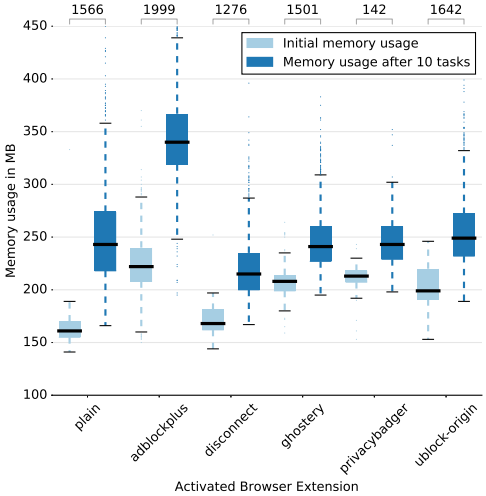
HTTP vs. HTTPS



Small vs. large tracking services



Memory consumption

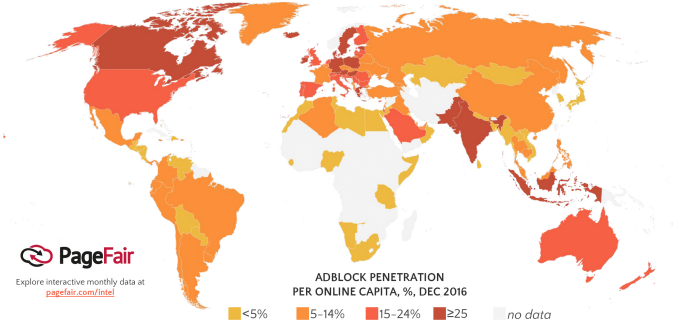


Fingerprinters

	Web Dataset							Android			
	plain	abp	dc	gh	pb	ubo	c	plain	e	a	m
FP-Detective											
IOVATION	98	97	97	1	97	4	1	-	-	-	-
ThreatMetrix	39	39	39	1	37	-	-	149	149	149	-
BlueCava	27	5	-	-	25	-	-	-	-	-	-
...											
OpenWPM: Canvas Font Fingerprinting											
mathid.mathtag.com/d/i\.js	437	374	2	1	2	3	-	-	-	-	-
admicro1.vcmedia.vn/core/fipmin\.js	39	1	-	3	41	1	-	-	-	-	-
.*.online-metrix.net	39	39	39	1	37	-	-	149	149	149	-
...											
OpenWPM: Canvas Fingerprinting											
doubleverify.com/dvtp_src_internal.*\.js	4118	78	8	6	37	9	-	-	-	-	-
ap.lijit.com/www/delivery/fp	826	27	296	3	349	1	-	1	-	1	-
tags.bkrtx.com/js/bk-coretag\.js	631	391	134	-	449	6	-	10	10	10	-
...											
OpenWPM: WebRTC Local IP discovery											
cdn.augur.io/augur.min\.js	111	31	4	43	57	21	2	-	-	-	-
click.sabavision.com/*/jsEngine\.js	78	54	81	84	77	56	46	-	-	-	-
static.fraudmetrix.cn/fm\.js	11	11	11	11	14	11	9	-	-	-	-
...											

Adblock Usage Worldwide

- Main motivation: security and annoyance
- In Asia: mobile browsers pre-configured with adblockers
- Global: more educated users rely on adblockers



Adblock detection

- Baiting: inject (random) `html-tag` check if it's blocked
- Integrity checks: verify if certain scripts are loaded
- 75% of users leave websites with adblock detection (according to PageFair)

The ideal browser

- **Browser** (2023)
  - **Mozilla Firefox** *(+ arkenfox*[8]*)*
  - Brave for privacy per default on a Chrome-basis
- **Extensions**
  - **Ghostery** (activate full blocking!), effective protection
  - **uBlock origin**, smaller trackers + ad networks
  - **Decentraleyes**, covers popular CDNs
  - (Anti-Adblock Killer)
    `http://reek.github.io/anti-adblock-killer/`
  - Use encrypted DNS (DoH/DoT)

[8] ArkenFox GUI

Beyond the Browser

Mobile and “Smart” Devices

Mobile Privacy

- Smartphone apps collect a number of sensitive information
  - Contact data
  - Location data
  - ...
- OS background-services (e.g. Google Play Services)
- Third-party providers (apps, ads, analytics, social SDKs)
  - Access sensitive information[9]
  - Rely on unique device identifiers
    (can be reset on current mobile OS versions)

[9] Angry Birds and Location Data

Tracking accross devices

- Holy grail for marketers
- Probabilistic methods
  (e.g. link devices with the same IP + same account ...)
- Big players (Facebook, Google)
  - identify you, once you authenticate with their SDKs
  - are common third-parties in apps
  - Know your social circle
- “My phone is listening”
  - Not necessary, enough tracking information available
  - “Cross Device Tracking”[10] never gained traction

[10] Privacy Threats through Ultrasonic Side Channels on Mobile Devices

Mobile Privacy Tools

- Anti Web Tracking
  - Since iOS 9 blocking extensions for Safari
  - Mobile Firefox + extensions
  - specialized privacy browsers: bromite, ghostery, etc.
- Custom Android ROMs
  - Require rooting/jailbreaking of device
  - Not feasible for the average user/device
  - GrapheneOS, CalyxOS, DivestOS

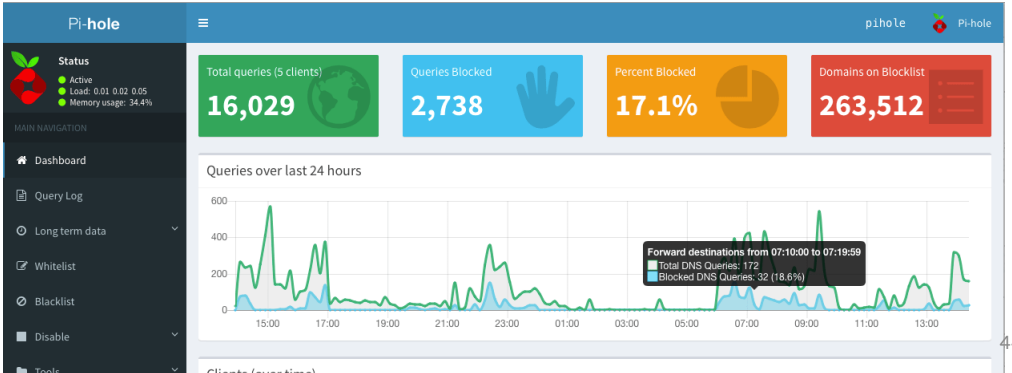
Mobile Privacy Tools: DNS

- DNS based blocking
  - Reply to known tracking domain with `0.0.0.0`
  - Course grained in comparison to browser extensions
  - `ads.facebook.com` (can be blocked DNS-based)
  - `www.facebook.com/ads` (would lead to overblocking)
- Using DNS blocking
  - e.g. specific Android apps: DNS66 (detailed control)
  - external services: special VPN or adblocking DNS resolvers
  - Running your own blocking DNS (e.g.Pi-Hole, upribox)

Pi-hole

Pi-hole project

- Open source project for network-wide Ad Blocking
- Blocking DNS-resolver for Raspberry Pi, Docker, etc.
- `https://pi-hole.net`



Usable Privacy Tools

Usable Privacy Box (*upribox*, UP4)

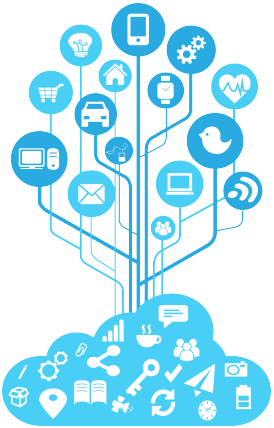
- upribox Open-source project
  - netidee/IPA Privacy Award 2015
  - 2015 - 2023
    (community + features)
- upribox: Blocking, Tor, VPN, IoT
- **UP4**: private DNS box
  (Pi-hole + knot-resolver)
  - UP4 Box Documentation (WiP)

https://upribox.org

@usableprivacy

Usable Privacy DNS

- Public DoH / DoT resolver with ad- and tracker-filtering
- `https://docs.usableprivacy.com/dns`
- Easy to use with Android, iOS



Conclusio

Webprivacy

- **Network-Leaks**:
  DNS, HTTP, HTTPS handshake, (netflows)
- **Third-parties**:
  ads, social networks, analytics
- **Tracking Methods**:
  from cookies to supercookies; fingerprinting
- **Cross-device tracking**:
  feasible for large corps

Technical Challenges

- **Mobile + ”Smartdevices”**:
  - Tracker-Blocking difficult (e.g. requires rooting)
  - current solution: custom blocking DNS service
  - Encrypted DNS as a double-edged sword (apps/devices might come with inbuilt DoT resolvers)

- **arms-race**:
  Adblock-Blocker-Blocker-Blocker et. al.
  new browser protections result in new tracking methods

Overall Tracking Challenges

- **Business models of online services**:
  alternatives to current ad-system necessary
- **Usability**:
  e.g. Ghostery: by default nothing is blocked.
  Desktop vs. Smartphones

“Build systems for people, not companies or states”
(Bart Preneel, CCS16)

Outlook

- Inverted content for web privacy (quiz)
- Lecture recording for VPNs and Doh/DoT
- **Guest lectures in the next two weeks**