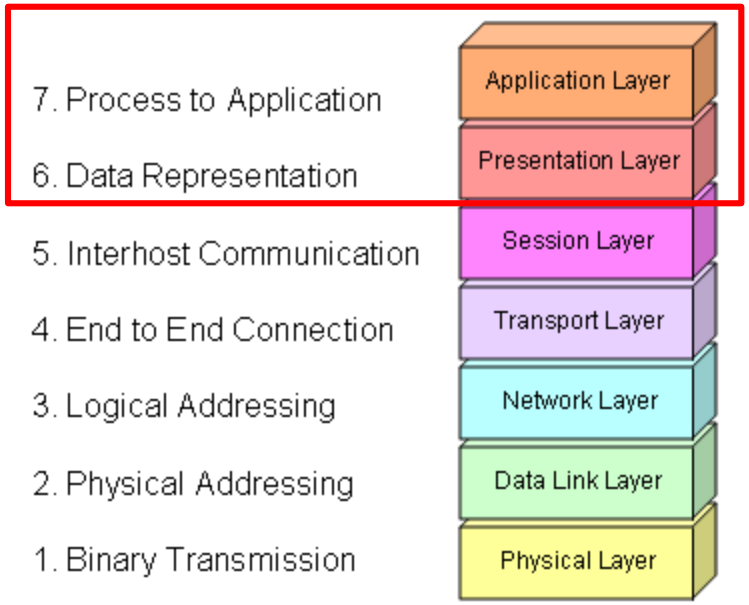
IT-SICHERHEIT UND DATENSCHUTZ

KAPITEL 7 - WEB-DATENSCHUTZ

buchmann@hft-leipzig.de

INHALTE UND LERNZIELE DIESES KAPITELS

- Welche Daten fallen bei der Internet-Kommunikation an?
  - Zugehörige Datenschutzprobleme
- Nutzertracking im Internet
  - Cookies, Web-Bugs, Fingerprinting
- Lösungsoptionen
  - Do Not Track, Cookie-Management



- Lernziele
  - Sie können die Funktionsweise des Internets und des WWW beschreiben und aufzeigen, welche persönlichen Daten dabei wo und mit welchen Verfahren gesammelt werden.
  - Sie können erläutern, welche Datenschutzprobleme daraus entstehen, wie schwerwiegend diese sind und was sich dagegen tun lässt.

ZIELE DES DATENSCHUTZES

- *1 grundlegendes Gewährleistungsziel* **Datenminimierung**
  - *Umfang* der verarbeiteten Daten, *Zahl* der zugreifenden Stellen, *Ausmaß* der Verfügungsgewalt
- *6 elementare Gewährleistungsziele*
  - **Verfügbarkeit**: Daten für vorgesehene Zwecke verfügbar
  - **Integrität**: Daten halten Spezifikation aus Zweckbindung ein
  - **Vertraulichkeit**: Kein Zugriff durch Unbefugte
  - **Nichtverkettung**: Kein Kombinieren von Daten ohne Erlaubnis
  - **Transparenz**:Betroffene, Betreiber, Kontrollinstanzen können erkennen, wo welche Daten zu welchem Zweck vorliegen
  - **Intervenierbarkeit**: Betroffenenrechte (Benachrichtigung, Auskunft, Berichtigung, Sperrung, Löschung)

WEB-ANWENDUNGEN: WER ERFÄHRT WAS?

LADEN EINER WEBSEITE

Wiederholung

Input lines are **bold**
2x line break
is important!

/home/buchmann> **telnet dbis.ipd.uni-karlsruhe.de 80**
Trying 129.13.182.137...
Connected to dbis.ipd.uni-karlsruhe.de.
Escape character is '^]'.

request

GET / HTTP/1.0
Referer: http://localhost/index.html
User-Agent: Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.8.1.1) Firefox/2.0.0.1
Host: localhost
Accept: text/html, image/gif, image/jpeg, image/pjpeg, */*
Accept-Language: en, de

(here follows the response from the server)

ANTWORT VOM WEBSERVER

Wiederholung

Input lines are **bold**
2x line break
is important!

/home/buchmann> **telnet dbis.ipd.uni-karlsruhe.de 80**
Trying 129.13.182.137...
Connected to dbis.ipd.uni-karlsruhe.de.
Escape character is '^]'.
GET / HTTP/1.0

response

HTTP/1.1 200 OK
Date: 11 April 2018 08:46:39 GMT
Server: Apache/2.2.8 (Unix) DAV/2 mod_ssl/2.2.8 OpenSSL/0.9.8h PHP/5.2.6
X-Powered-By: PHP/5.2.6
Set-Cookie: PHPSESSID=74e6a1911499e9578534f31814c357d4; path=/
Expires: 19 Nov 2018 08:52:00 GMT
Cache-Control: no-store, no-cache, must-revalidate, post-check=0, pre-check=0
Pragma: no-cache
Connection: close
Content-Type: text/html

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" <…>
</html>*Connection closed by foreign host.*

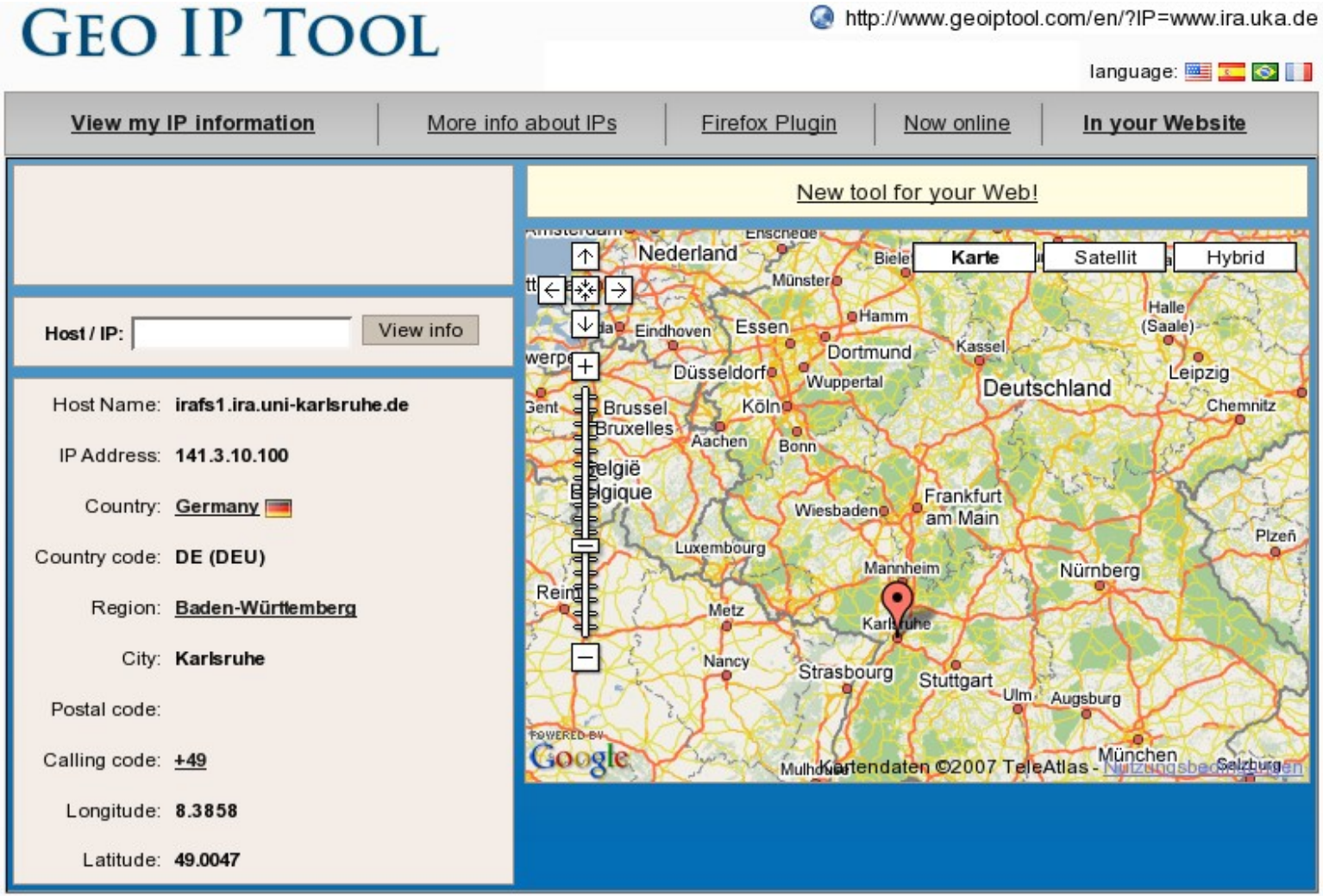
WAS LOGGT DER WEBSERVER?

- Direkt ablesbar:
  - Wo bin ich, wer bin ich? (IP-Adresse)
  - Was will ich? (URL)
  - Wo komme ich her? (Referrer, nicht im Beispiel enth.)
  - Wann habe ich die Seite abgerufen?
  - Welche Sprache spreche ich?
  - Welche Systemsoftware setze ich ein?

→ **Ungewöhnliche Kombinationen sind Quasi-Identifier!**

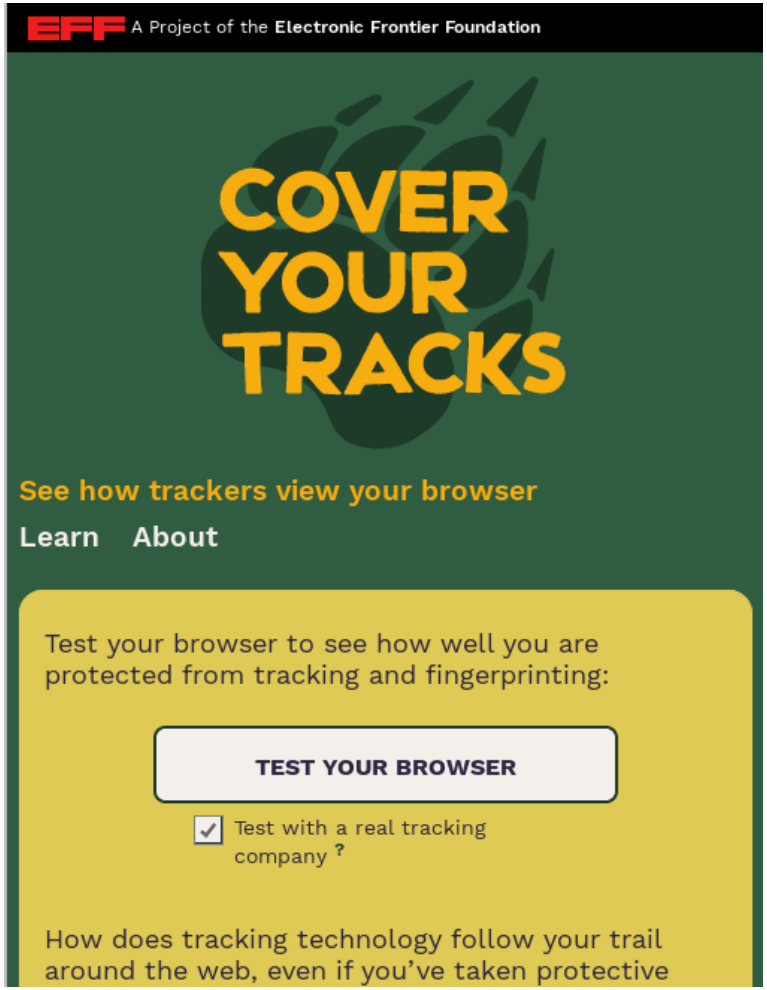
/home/buchmann# **tail -1 /var/log/apache2/access_log**
123.4.5.6 - - [11/May/2018:11:01:42 +0200]
"GET / HTTP/1.0" 200 234
"http://dbis.ipd.uni-karlsruhe.de/index.html"
"Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.8.1.1) Gecko/20121208
Firefox/2.0.0.1"

WAS VERRÄT DIE IP-ADRESSE?



WAS VERRÄT MEIN BROWSER?

- https://coveryourtracks.eff.org/
- Siehe auch:
  https://panopticlick.eff.org
  https://amiunique.org



TEST: OPERA, SUSE LINUX

browser characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	13.41	10913.75	Opera/9.80 (X11; Linux x86_64) Presto/2.12.388 Version/12.14
HTTP_ACCEPT	8.58	382.27	text/html, */* gzip, deflate en-US,en;q=0.9
Browser Plugins	21.45+	2859402	Plugin 0: DivX Web Player; DivX Web Player version 1.4.0.233; libtotem-mully-plugin.so; (; video/divx; divx). Plugin 1: Gnome Shell Integration; […]
Time Zone	2.85	7.22	-120
Screen Size and Color Depth	7.1	136.97	1600x1200x24
System Fonts	21.45+	2859402	LM Mono Slanted 10, GFS Baskerville, FreeMono, LM Mono 10, LM Mono 12, Droid Sans, […]
Are Cookies Enabled?	0.42	1.34	Yes
Limited supercookie test	0.99	1.99	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No

ZWISCHENFAZIT

- Wenn *nur das technisch Notwendige* durchgeführt wird
  - Nutzer ist anonym/pseudonym
    - *dynamische IP-Adressen*:
      Änderung nach jedem neuen Einwählen beim ISP
    - *statische IP-Adressen*: können zu Proxies, Hubs, Firewalls oder Gateways gehören; nicht zwingend einem Einzelnen zugeordnet
    - *Gerätesignaturen*: Geräte können von mehreren genutzt werden;
      ein Nutzer kann mehrere Geräte verwenden
  - **Datenspuren sind verteilt**
    - Webseitenzugriff hinterlässt Datenspuren nur auf einen Webserver
    - keine Verfolgung des Nutzers über mehrere Sites

→ Aufbau von umfangreichen Nutzerprofilen kaum möglich

WEB-ANWENDUNGEN: BEISPIEL SUCHMASCHINE

SUCHMASCHINEN: GOOGLE, BING UND CO.

- Suche übermittelt an Suchmaschinenbetreiber:

Kategorie	Attribute
Suchbegriffe (technisch unvermeidlich)	Aneinanderreihung von Schlüsselworten
Browser-Kommunikation (technisch unvermeidlich)	Zeitstempel, IP, Browser, Betriebssystem, Spracheinstellungen, zuletzt besuchte Seite
Browser-Kommunikation (technisch vermeidlich)	ausgewähltes Suchergebnis, (Implementiert als Redirect)
Zusatzinformationen (vom Nutzer vermeidbar)	Cookie-Informationen, Session-ID
	über JavaScript ermittelte Daten, Verweildauer auf der Seite, Bildschirmauflösung
	Nutzer-ID (z.B. Google-Login, Yahoo-ID)

VERKETTUNG VON SUCHANFRAGEN

- Über die **Browser-Kommunikation**, insbes. IP-Adresse
  - dynamische IP-Adressen: überdauern mindestens eine Such-Session
    - Kombination aus Standort des IP-Adressbereichs, Betriebssystem, Sprache, Browser etc. können als Quasi-Identifier ausreichen
  - statische IP-Adressen, z.B. Uni-Netz: länger gültig
- Über **Cookies, Session-IDs, Nutzer-Login**
  - Identifiziert einen Browser (und damit oft dessen Benutzer) über lange Zeiträume eindeutig
  - auch bei wechselnder IP-Adresse
- Über die **Suchterme**
  - z.B. Suche nach eigenem Namen, seltene Hobbies
- **Kombinationen** aus allem

VERKETTUNG MIT ZUSATZDIENSTEN

- Suchmaschinenanbieter oft Anbieter weiterer Dienste
  - Google: Youtube, Maps, Email, Google Docs, Google Earth, News, Google Analytics, Google+
- *separate* digitale Teilidentitäten werden verkettbar
  - über IP-Adresse Nutzerbewegungen über mehrere Dienste hinweg nachvollziehbar
  - oft übergreifendes Login für viele Dienste,
    - Microsoft Passport, Windows Live ID
    - Google Authentication for Web Applications (OAuth)
- Informationen über viele Lebensbereiche
  - Arbeit, Privatleben, Hobbies, Kommunikationspartner etc.

PERSONENBEZUG JA/NEIN?

- kommt ganz drauf an...
  - zur Erinnerung: *"...Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person."*
  - Heute übliche Interpretation: kein Personenbezug, wenn "*für Einzelangaben zu einer Person die Wahrscheinlichkeit, dass diese der Person zugeordnet werden können, so gering ist, dass sie nach der Lebenserfahrung oder dem Stand der Wissenschaft praktisch ausscheidet.*"

Quelle: Roßnagel, A.; Scholz, P.: Datenschutz durch Anonymität und Pseudonymität, MMR 2000

PERSONENBEZUG DER SUCHBEGRIFFE

- auch hier: kommt drauf an...
  - “Katzenfutter billig” → nein;
  - “Erik Buchmann Urlaub Italien” → ja

- Personenbezug ist *abhängig von den Benutzereingaben*; nicht ohne weiteres automatisch vom Betreiber entscheidbar!
  - je umfangreicher die Suchhistorie eines Nutzers, desto wahrscheinlicher kommt eine identifizierende Kombination von Identitätsattributen zusammen
    (Beispiel folgt)

PERSONENBEZUG DER KOMMUNIKATION

- Dynamische IP-Adresse
  - eher nicht personenbezogen
    - ändert sich mindestens täglich
    - *Personenbezug erfordert Mithilfe des Kenners der Zuordnungsregel*, d.h., Internet Service Provider
- Statische IP-Adresse für festen Rechner
  - Suchmaschinenbetreiber kann Suchanfragen über lange Zeiträume einer Person zuordnen
  - mit zunehmender Zahl der Suchvorgänge steigt Wahrscheinlichkeit, dass sich der Suchende offenbart
    (vgl. vorangegangene Folie)
  - daher: oftmals personenbezogen

PERSONENBEZUG VON ZUSATZINFORMATIONEN

- Bildschirmauflösung, Sprache,
  Verweildauer auf der Seite
  - Information für sich allein nicht als identifizierendes Merkmal geeignet, ggf. bildet die Summe aller Informationen aber Browser-Fingerabdruck

- Nutzer-Login, Session-ID, Cookie-Informationen
  - Betreiber hat Login, Session-ID bzw. Cookie selbst dem Nutzer zugeordnet
  - daher: oftmals personenbezogen, da *Suchmaschinenbetreiber kennt auch Zuordnungsregel!*

AUG. 2006: AOL-DATENLECK (1/2)

- AOL Research veröffentlicht für Forschung 20 Mio. Suchanfragen von 650.000 Usern, gesammelt über 3 Monate
  - IDs pseudonymisiert, künstl. Schlüssel

User ID	Search Keywords	Date	Website
4417749	numb fingers	2006-03-06 18:35:02	http://irgendwas.de
...	...	...	...

- zwar ist keiner der Datensätze unmittelbar personenbezogen, aber schnell werden einzelne Identitäten und komplette Persönlichkeitsprofile offenbar

- 3 Tage später: Datenbank ist vom Netz, aber schon vielfach in Tauschbörsen kopiert, bis heute verfügbar

AUG. 2006: AOL-DATENLECK (2/2)

- Nutzer 4417749: hunderte Suchanfragen in 3 Monaten
  - "dog urinates on everything": *Hundebesitzer*
  - "60 single man": *einsame ältere Frau*
  - "numb fingers": *körperliche Gebrechen*
  - "homes sold in gwinnett county": *Wohnung*
  - "xxx Arnold, yyy Arnold": *suche nach Verwandten*
  - "school supplies for Iraq children": *karitativ*
  - "best season to visit Italy": *nächster Urlaub*
- identifiziert als Thelma Arnold:
  "My goodness, its my whole personal life!"
  - Privat- und Alltagsleben, Ängste, Gebrechen
  - falsches Selbstbild durch Suchanfragen für Freunde

Quelle: http://www.nytimes.com/2006/08/09/technology/09aol.html

WIE HANDHABT ES GOOGLE? (STAND: FEB. 2020)

Aufbewahrung Ihrer Informationen
Wenn Sie Daten löschen, beginnen wir mit einem Datenlöschvorgang, um dafür zu sorgen, dass die Daten sicher und vollständig von unseren Servern entfernt werden oder nur in anonymisierter Form gespeichert bleiben. Wir bemühen uns, Daten mithilfe unserer Dienste davor zu schützen, irrtümlich oder mutwillig gelöscht zu werden. (...)

Generalisierung von Daten
(...) Um die Privatsphäre dieser Personen zu schützen, nutzen wir die Generalisierung und entfernen dabei Teile der Daten oder ersetzen diese mit einem allgemeinen Wert. Beispielsweise tauschen wir dazu Segmente aller Vorwahlen oder Telefonnummern mit derselben Zahlenfolge aus.
Durch die Generalisierung können wir k-Anonymität erzielen. (…) Wenn alle Einzelpersonen in einem Datensatz denselben Wert für ein vertrauliches Attribut teilen, können vertrauliche Daten preisgegeben werden, indem einfach nur bekannt ist, dass diese Einzelpersonen Teil des entsprechenden Datensatzes sind. Um dieses Risiko zu mindern, können wir l-Diversität nutzen. (…)

Den Daten Rauschen hinzufügen
Bei der Methode der Differential Privacy wird den Daten mathematisches Rauschen hinzugefügt. Bei Differential Privacy lässt sich nicht sicher bestimmen, ob eine bestimmte Einzelperson Teil eines Datensatzes ist.(...)

ALTERNATIVEN

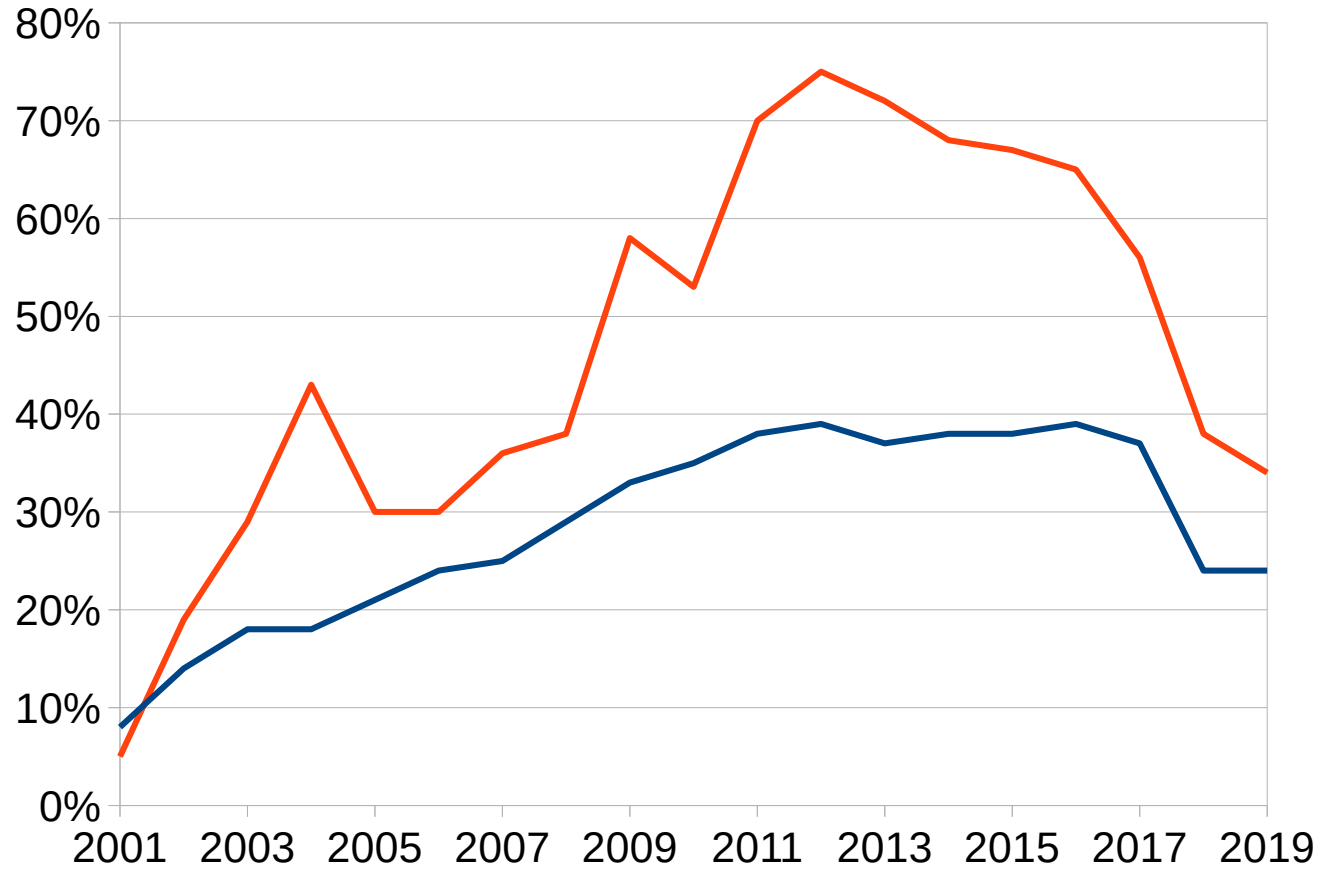
- Es gibt Suchmaschinen, die keine IP-Adressen oder Suchangaben protokollieren

NUTZER-TRACKING MIT WEB-BUGS, COOKIES, FINGERPRINTING

TECHNIKEN DES ONLINE-MARKETINGS

- **Reichweitenanalyse**
  - Besucherströme auf Webseiten, z.B. Google Analytics
- **Remarketing**
  - Auf Dritt-Websites Anzeigen für Produkte, die der Nutzer kurz vorher in einem Web-Shop angeschaut hat
- **Profiling**
  - Auswertung und Vorhersage von Kundeninteressen, Kaufkraft, Kaufabsicht etc. anhand von Profiling-Daten (Alter, Geschlecht, Ort, soziale Gruppe, betrachtete Websites, etc.)
- **Online Behavioral Advertising**
  - Profiling auf der Basis der Bewegungsdaten im Netz, um das Interesse eines Web-Nutzers für bestimmteWerbebanner zu ermitteln
- **Custom Audiences Marketing**
  - Profiling z.B. auf Basis der Profile in sozialen Medien, zielgruppenspezifische Werbung zu schalten

ANTEIL DER WEBSITES MIT COOKIES/WEB BUGS



*) Zahl der Webbugs ist überschätzt, falls Bugs von mehreren Anbietern auf einer Site

http://www.securityspace.com, Stand 1.4.2019, 4.101.531 untersuchte Sites

NUTZERTRACKING MIT WEB-BUGS

- Ziel: Überwachung des Nutzers, Nachvollziehen seiner Bewegungen
  - auf einer einzelnen Webseite
  - über mehrere Webseiten hinweg
- Idee:
  - Browser ruft **präparierte Datenobjekte** auf verschiedenen Servern ab, und
  - hinterlässt dort **Spuren im Log**
- Methode:
  - Verweise auf Datenobjekte werden in HTML, EMails, PDFs etc. so eingebunden, dass sie der Betrachter automatisch nachlädt
  - dynamisch generierte Namen, damit Web-Bugs nicht aus dem Browsercache geladen werden

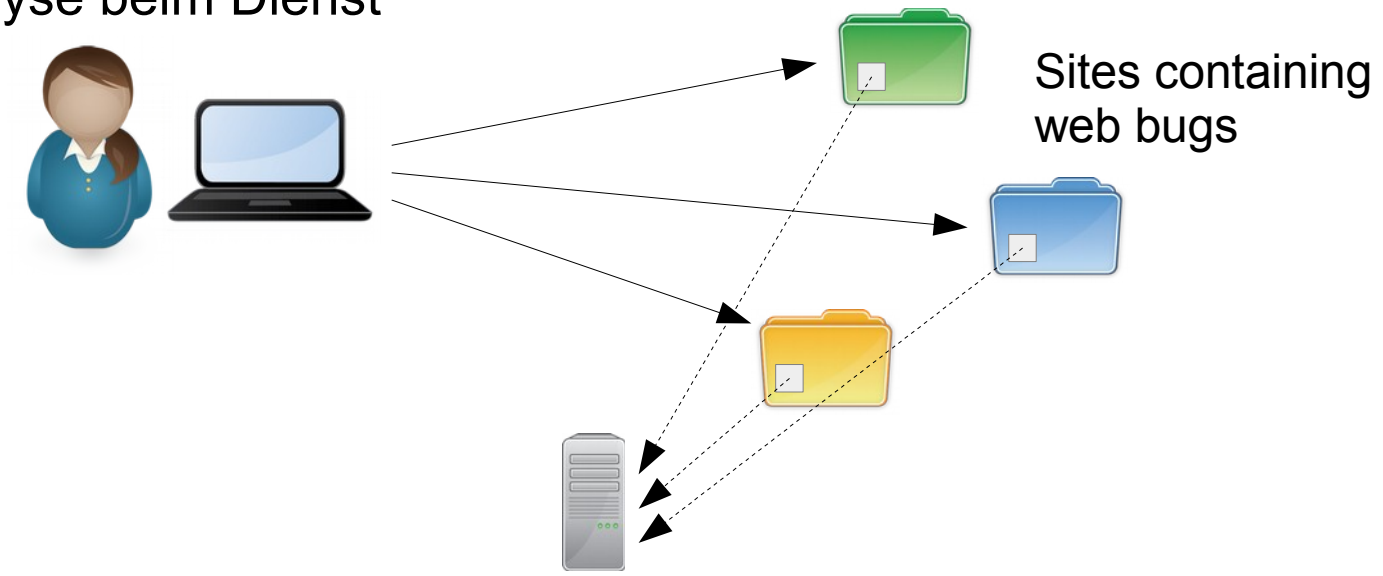
BEISPIEL FÜR EINEN WEB-BUG

- Eingebettet in jede Webseite, die am Web-Bug-Dienst teilnimmt

```
<!DOCTYPE html>
<html lang="de">
<head>
<meta charset="UTF-8"/>
</head>
<body>
...sometext...
<iframe src="http://www.bugger.com/" height="1"
width="1" name="user92374920"></iframe>
...somemoretext...
</body>
```

SO FUNKTIONIERTS

- Viele Webseitenbetreiber nehmen bei Analysedienst teil
  - jeder Webseitenbetreiber bindet den Web-Bug des Dienstes ein
- Webseitenzugriff durch den Benutzer
  - Browser lädt Webseite → *Eintrag im Log des Betreibers*
  - Browser lädt iframe vom Analysedienste → *Eintrag im Log des Dienstes*
- Log-Analyse beim Dienst



WEB-BUGS IM WEBBROWSER

- **Bilder**
  ```
  <img src="http://spy.com/verifyme.cgi?id=X" width="1"
  height="1">
  ```
- **Frames, IFrames**
  ```
  <frame src="http://spy.com/verifyme.cgi?id=X">
  ```
- **Scripts**
  ```
  <script src="http://spy.com/verifyme.cgi?id=X"
  type="text/javascript"></script>
  ```
- **Styles**
  ```
  <link rel="stylesheet" media="screen"
  href="http://spy.com/verifyme.cgi?id=X">
  ```
- **Layer**
  ```
  <layer top="80" left="40"
  src="http://spy.com/verifyme.cgi?id=X">
  ```
- und noch einige mehr...

DATEN ZU WEB BUGS

Stand: 01.04.2019, http://www.securityspace.com, 4.101.531 untersuchte Sites

Domain	Sites	Types of Bugs
hztalk.com	130875 (3.2%)	script[100.0%]
google.com	101690 (2.5%)	IMG[0.7%], SCRIPT[0.1%], embed[0.8%], iframe[16.4%], img[11.0%], script[77.6%]
googleapis.com	89261 (2.2%)	img[1.3%], script[98.7%]
youtube.com	57800 (1.4%)	EMBED[0.3%], IFRAME[0.5%], embed[33.1%], iframe[68.5%], img[6.9%], script[1.1%], source[0.1%]
tumblr.com	56916 (1.4%)	frame[0.1%], iframe[66.5%], img[43.6%], input[0.4%], script[95.4%]
baidu.com	56714 (1.4%)	IFRAME[0.2%], IMG[0.2%], SCRIPT[0.4%], embed[0.2%], iframe[2.3%], img[7.8%], script[91.4%]
googlecode.com	53744 (1.3%)	script[99.8%]
gravatar.com	42625 (1.0%)	img[89.6%], script[14.9%]
qq.com	36085 (0.9%)	IMG[2.0%], SCRIPT[0.2%], embed[1.2%], iframe[1.8%], img[77.4%], script[21.6%]
googlesyndication.com	34029 (0.8%)	SCRIPT[0.8%], img[1.6%], script[97.8%]

FÜR WEB-BUGS NUTZBARE SOFTWARE

- Alles, was **automatisch** Inhalte aus dem Web nachlädt
  - Web-Browser, Browser-Plugins
  - Microsoft Office, OpenOffice
  - Mail-Clients (weniger anfällig; hier haben viele Hersteller reagiert)
  - PDF-Dateien, Windows-Hilfedateien
  - sämtliche Produkte mit automatischem Update
  - und viele mehr...

SCHUTZ VOR WEB-BUGS

- Web-Bugs vs. nützlichen Anwendungen?
  Beispiele für Anwendungen der Technik sind:
  - kleinere EMails, Grafiken bei Bedarf vom Server
  - Daten in Spreadsheets automatisch aktualisieren
  - Rechtemanagement für geschützte Inhalte über Authentifizierungsmechanismen des Webservers

- kein wirkungsvoller Schutz möglich, aber Teillösungen:
  - Nachlanden von allen Inhalten aus dem Internet unterdrücken (Mailclients)
  - zusätzliche Inhalte nur aus der Domain des Ursprungsdokuments nachladen (Web-Browser)

NUTZERTRACKING MIT COOKIES

- Ziel: Eindeutige Identifikation des Nutzers, Nachvollziehen seiner Bewegungen („Nutzer-Tracking“)
  - auf einer einzelnen Webseite
  - über mehrere Webseiten hinweg
- Idee:
  - Browser speichert **eindeutige Kennung** auf dem Rechner des Nutzers
- Methode:
  - Cookies
  - 3rd Party-Cookies

COOKIES

- Daten, die ein Webserver auf dem Rechner des Anwenders speichern und jederzeit wieder abrufen darf
  - beliebige kurze Zeichenketten
  - rudimentäre Sicherheitsfeatures
  - Verfallsdatum
- Webbrowser darf Cookies jederzeit löschen
- Cookie-Arten im Webbrowser
  - Cookie im Webbrowser
  - Supercookie (DOM local/session Storage)
  - Internet Explorer User Data
  - Flash Cookie

Diese Webseite verwendet Cookies

Wir setzen auf unseren Internetseiten Cookies und andere Technologien ein, um Ihnen unsere Dienste technisch bereitstellen zu können, Inhalte und Anzeigen für Sie zu personalisieren sowie anonyme Nutzungsstatistiken zu analysieren. Cookies von Drittanbietern setzen wir ein, um Ihnen Funktionen für soziale Netzwerke bereitstellen zu können. Informationen zur Nutzung unserer Dienste werden an unsere Partner für soziale Medien, Analyse und Werbung weitergegeben. Unsere Partner führen diese Informationen möglicherweise mit weiteren Daten zusammen, die Sie ihnen bereitgestellt haben oder die sie im Rahmen Ihrer Nutzung der Dienste gesammelt haben. Durch die weitere Nutzung unserer Internetseite erklären Sie sich mit dem Einsatz von Cookies einverstanden.

Weitere Informationen zu den von uns eingesetzten Cookies oder deren Deaktivierung finden Sie hier

OK

Einstellungen

Cookie-Erklärung | Über Cookies

- Notwendig (124)
- Präferenzen (6)
- Statistiken (48)
- Marketing (225)
- Nicht klassifiziert (161)

IDE	doubleclick.net	Verwendet von Google DoubleClick, um die Handlungen des Benutzers auf der Webseite nach der Anzeige oder dem Klicken auf eine der Anzeigen des Anbieters zu registrieren und	1 Jahr	HTTP

Die Cookie-Erklärung wurde das letzte Mal am 11.01.2019 von Cookiebot aktualisiert

noz

SO FUNKTIONIERTS

- **einmal** gesetzte Cookies werden vom Browser **automatisch** bei **jedem** Klick auf einen Link an Server übermittelt
- In PHP (auf Server-Seite) genügt eine Zeile zum Setzen von Cookies und zwei zum Abfragen:

```
<?php
  if (isset($_COOKIE['PHPSESSID'])) {
    $id = $_COOKIE['PHPSESSID'];
  } else {
    $id = generateId();
    registerInDatabase($id);
    setcookie('PHPSESSID',$id,time()+241920000);
  }
?>
```

- Cookie-Fähigkeit: Webbrowser, Java-Script, Flash
  → Browser-Einstellungen gelten nicht für alle Cookies!

COOKIES VOM WEBSERVER

/home/buchmann> **telnet dbis.ipd.uni-karlsruhe.de 80**
Trying 129.13.182.137...
Connected to dbis.ipd.uni-karlsruhe.de.
Escape character is '^]'.
GET / HTTP/1.0

HTTP/1.1 200 OK
Date: 11 May 2014 08:46:39 GMT
Server: Apache/2.2.8 (Unix) DAV/2 mod_ssl/2.2.8 Open[illegible]SL/0.9.8h PHP/5.2.6
X-Powered-By: PHP/5.2.6
Set-Cookie: PHPSESSID=74e6a1911499e9578534f31814c357d4; path=/;
Expires=Wed, 09 Jun 2021 10:18:14 GMT
Cache-Control: no-store, no-cache, must-revalidate, post-check=0, pre-check=0
Pragma: no-cache
Connection: close
Content-Type: text/html

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" …
</html>*Connection closed by foreign host.*

Werden automatisch bei der Auslieferung beliebiger HTTP -Pakete übertragen

COOKIES AUF DER FESTPLATTE

- Beispiel:

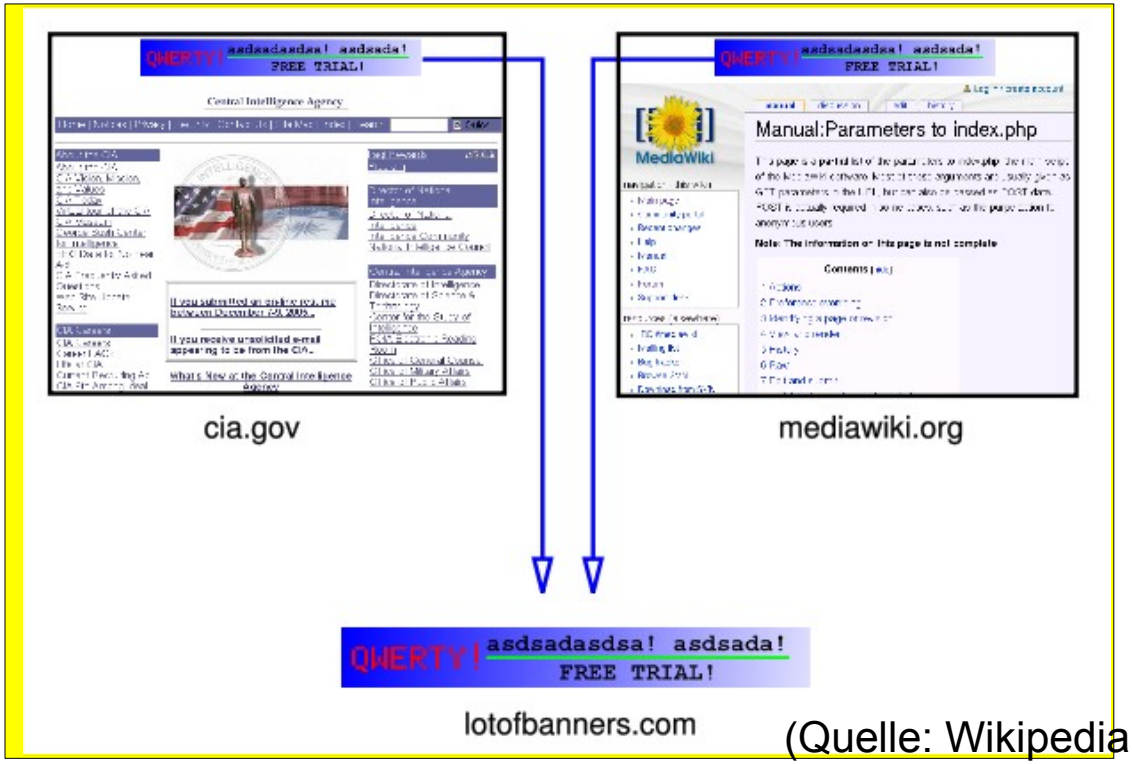
/home/buchmann> **cat ~/.mozilla/firefox/5p4dyjr8.default/cookies.txt**

.advertising.com	TRUE / FALSE	1380216294	BASE	cxqDMdeP80sVzIE!
.advertising.com	TRUE / FALSE	1383850713	C2	1+IFJ14DFMQtFe+h
.affilinet.parship.de	FALSE/ FALSE	1272740302	VID	par-sv-53%3ASRSP
.advertising.com	TRUE / FALSE	1380216294	F1	BYGbejkAAAAAnXo
.dilbert.com	TRUE / FALSE	1609459211	RMAM	01cen8_1006.4Y9ZK
.dilbert.com	TRUE / FALSE	1609459275	OAX	VKP4KEi0ZVYACirH
.uni-karlsruhe.de	TRUE / FALSE	1838923423	PHPID	74e6a1911499e9578

- Spalten
  - Domain, die den Cookie gesetzt hat
  - Zugriff von der ganzen Domain / Zugriff nur bei HTTPS
  - Verfallsdatum (je länger Speicherdauer, desto länger Tracking möglich)
  - Name des Cookies
  - Wert des Cookies

3RD PARTY COOKIES

- Beispiel: lotofbanners.com platziert Banner auf cia.gov und mediawiki.org, erzeugt Cookie mit ID



(Quelle: Wikipedia)

- Besucher beider Webseiten können zugeordnet werden

SCHUTZ VOR COOKIES?

- viele sinnvolle Anwendungen für Cookies
  - Session-Verwaltung auf Webseiten, z.B. für Web-Shops
    → d.h., Zustand speichern (http ist zustandslos)
  - Speichern von Einstellungen im Browser des Anwenders
    → Werkzeug gegen Profiling!
- keine automatische Erkennung von 'guten' und 'bösen' Cookies
  - die gesammelten Informationen sind verborgen im Server des Cookie-Setzers gespeichert
- einzige Abhilfe:
  - Filter im Webbrowser installieren, Regelwerk aufsetzen,
    Filter im Flash-Plugin konfigurieren
    → erfordert Zeit und umfangreiche Kenntnisse

NUTZERTRACKING MIT ADVANCED FINGERPRINTING

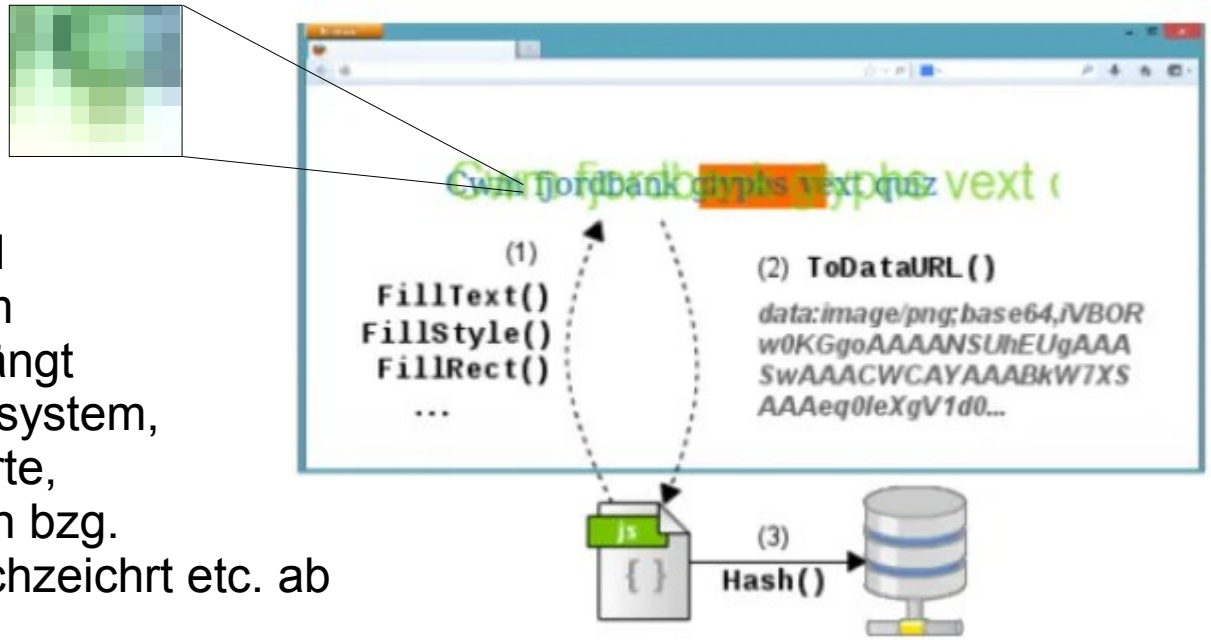
- Ziel: Eindeutige Identifikation des Nutzers, Nutzer-Tracking ohne komplexe Infrastruktur oder Cookies
  - auf einer einzelnen Webseite
  - über mehrere Webseiten hinweg
- Idee:
  - Browser **führt ein Script** aus, das für diesen Browser/Benutzer eine eindeutige Hash-Summe liefert
- Methode:
  - Advanced Fingerprinting, z.B.
    - Canvas-Fingerprinting (rendern von Schrift auf dem Bildschirm)
    - WebGL-Fingerprinting (Umgang mit Kompressionsartefakten in Bildern)
    - AudioContext-Fingerprint (Eigenschaften des Audio-Stacks des Rechners)
    - ...

Bild: https://www.schieb.de

SO FUNKTIONIERTS

- Serverbetreiber bettet Java-Script in Webseite ein
- Browser führt dieses (und viele andere) Scripte aus
- Script bestimmt Kennzahlen, die für diesen Rechner/Browser möglichst einzigartig sind, und sendet diese an den Server zurück

- Beispiel Canvas-Fingerprinting:
  - Wie Schrift in Pixel auf dem Bildschirm umgesetzt wird, hängt von Font, Betriebssystem, Treibern, Grafikkarte, Grafikeinstellungen bzg. Transparenz, Weichzeichrt etc. ab



SCHUTZ VOR ADVANCED FINGERPRINTING

- Seien Sie vorsichtig, wem Sie Java-Script erlauben
  - Bestimmte JavaScript-Elemente haben kaum praktische Anwendungen jenseits von Fingerprinting
  - Browser wie Firefox haben bereits sinnvolle Voreinstellungen
- Browser-Plugins wie UBlock Origin, Adblock Plus, Ghostery, … kommen mit Filterlisten, die bekannte Tracker blockieren
  - *Vollständigkeit und Aktualität der Filterlisten?*

DO NOT TRACK

- Bestandteil des HTTP-Headers,
  den der Browser an den Server übermittelt
  `DNT: 1` *(Do Not Track Enabled)*
  `DNT: 0` *(Do Not Track Disabled)*

- Beispiel

Eingaben sind **fett**
2x Zeilenumbruch
ist wichtig!

GET / HTTP/1.0
Referer: http://localhost/index.html
User-Agent: Mozilla/5.0 (X11; Linux i686; en-US; rv:1.8.1.1) Firefox/2.0.1
Host: localhost
DNT: 1
Accept: text/html, image/gif, image/jpeg, image/pjpeg, */*
Accept-Language: en, de

(hier folgt jetzt die Antwort vom Webserver)

SEMANTIK VON DNT

- Signalisiert der Webseite den Wunsch, dass über das Nutzerverhalten kein Nutzerprofil erstellt wird

- Einhaltung ist (mehr oder weniger) freiwillig
  - Deutschland: Widerspruch im Sinne von § 15 Abs. 3 Telemediengesetz
  - Europa: Opt-Out gemäß Richtlinie 2002/58/EG (Datenschutzrichtlinie für elektronische Kommunikation)
    Aktuell: um ePrivacy-Verordnung wird noch gestritten
  - Welt: vergleichbar mit

VOR- UND NACHTEILE

- Vorteile
  - Leicht implementierbar (alle großen Webbrowser unterstützen es)
  - „richtige“ Semantik: eine Willenserklärung für den Serverbetreiber
    - Sanktionsmöglichkeiten je nach vorhandenen Gesetzen
  - „Privacy by Default“ ist leicht umzusetzen
    - Internet Explorer ab Version 10 sendet in der Standardeinstellung „1“
- Nachteile
  - es gibt keine eine globale Übereinkunft, sich daran zuhalten
    (Anm.: Wie etwa im Fall der robots.txt, hier existiert eine Selbstverpflichtung der Suchmaschinenbetreiber)
  - Keine Kontrollmechanismen auf technischer Ebene
  - Was ist der richtige Default-Wert? Default=1 → welcher Nutzer würde absichtlich „bitte tracke mich“ einschalten?

ABSCHLUSS

ZUSAMMENFASSUNG

- Datenspuren im Internet
  - auf Rechner des Nutzers, beim Internet-Anbieter, beim Weiterleiter, Logs in Web-, DNS-, Mail-Server, Analysedienste von Dritten
  - basieren auf Log-Informationen des HTTP-Protokolls
    Cookies, Web-Bugs, aber auch auf Scripten (Advanced Fingerprinting)

- DoNotTrack
  - Klare Semantik („Bitte nicht tracken!“), aber keine Verpflichtung für den Webseitenbetreiber, sich daran zu halten

MÖGLICHE PRÜFUNGSFRAGEN

- Welche Arten von Daten erhält der Zugangsanbieter, der DNS-Anbieter, die Router im Internet sowie der Email-Dienstleister, wenn Sie Ihre Email über einen Web-Client im Webbrowser abrufen?

- Welche dieser Daten weisen einen Personenbezug auf? Begründen Sie Ihre Antwort

- Welche Vor- und Nachteile sehen Sie bei der Nutzung von DoNotTrack im Vergleich zum Anklicken eines Opt-Out Kontrollfelds auf der Webseite eines Dienstanbieters?

- Bestehen die Datenschutzprobleme im Zusammenhang mit der Verwendung von Cookies auch weiter, wenn Sie die Cookies am Ende der Browser-Session löschen?