Tracking im Web

So werden Sie beim Surfen verfolgt

WITECHWI

Arbeitskreis

Wissenschaft – **Tech**nik - **Wi**rtschaft

19. und 26. Januar 2022

Gunter Eger

Inhalt

- Was bedeutet Tracking?
  - Was sind die Ziele?
  - Was geschieht beim Tracking?
- Was sind Cookies?
  - Was ist mit 1st Party, 3rd Party Cookies gemeint?
  - Sind Cookies böse?
- Demo
- Exkurs: Wie funktioniert Progammatische Werbung?
- Skripte und moderne Alternativen zu Cookies
- Tracking durch Google Analytics
- Was ist Browser Fingerprinting?
  - Browser Fingerprinting ist böse
- FLoC – Googles Alternative zu Cookies
- Apples opt-in Alternative
- Ist Werbung unerlässlich für ein freies Internet?

Was bedeutet Tracking?

Tracking ist das Sammeln von Daten und ihre Auswertung bzgl. des Verhaltens von Besuchern auf Websites und von Nutzern von Apps

um daraus Rückschlüsse auf Ihre:

- Interessen
- Neigungen, Schwächen
- finanziellen Verhältnisse
- aktuellen Bedürfnisse
- Alter, Geschlecht
- politischen Einstellungen
- religiösen Haltungen
- Wohnort, Mobilität
- Bildungsabschluss
- …

zu ziehen.

Ziel ist es, im Laufe der Zeit mithilfe Ihrer Webseitenbesuche ein vielfältiges und realistisches Persönlichkeitsprofil zu erstellen, um Ihnen möglichst passgenaue Werbung zu präsentieren

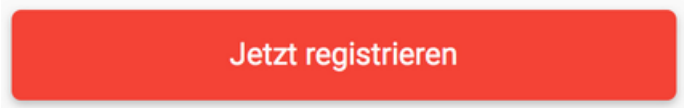
Ihre persönlichen Daten im Mittelpunkt des Interesses

Jeder Webseitenbetreiber oder jeder Dienst im Web möchte seine Besucher kennenlernen und möglichst viel über Sie erfahren

- vor allem soll der Benutzer wiedererkannt werden, wenn er die Webseite das nächste Mal aufruft
- Es gibt eine Reihe personenbezogener Daten, die besonders begehrt sind, sogenannte PII personal identifiable information
- Name
- Anschrift
- E-Mail
- Telefonnummer
- Kontodaten
- Geburtsdatum

die bei jeder Bestellung abgefragt werden

Passend dazu die penetrante Aufforderung:



Was geschieht beim Tracking?

Tracking ist das Zusammenführen möglichst vieler Informationsbruchstücke, die Sie als Anwender

- freiwillig oder
- gezwungenermaßen
- mit und ohne ihr Wissen

preisgeben, wenn Sie bestimmte Dienste in Anspruch nehmen

- Informationen werden über Monate und Jahre gesammelt und zusammengeführt

Besonders „elegant“:
Die Anmeldung bei Dritten mit Ihrem facebook, twitter oder Google Konto

Mit diesen Schaltflächen verfolgt facebook Ihre online-Aktivitäten

auch wenn Sie diese Knöpfe niemals anklicken

Welche Ziele werden mit Tracking verfolgt?

Offiziell und beschönigend: Optimierung der Website

- Dahinter verbergen sich legitime Ziele wie
  - Verbesserte Benutzererfahrung
    - intuitivere Bedienung der Website
  - Steigerung der Häufigkeit von Besuchen
  - Vermehrung von Seitenaufrufen
  - möglichst lange Verweilzeit auf den Plattformen
  - Newsletter-Abonnements
  - Verführung zu Käufen (Conversion)
  - ….
- Dazu werden sog. 1st – party cookies verwendet

- Die Ziele im Hintergrund
  - Platzierung möglichst passgenauer Werbung
  - Abschätzung und Beeinflussung des Kaufverhaltens
  - gläserner Kunde
  - Weiterverkauf der gewonnenen Daten
  - ….
- Hier kommen 3rd – party cookies zum Einsatz

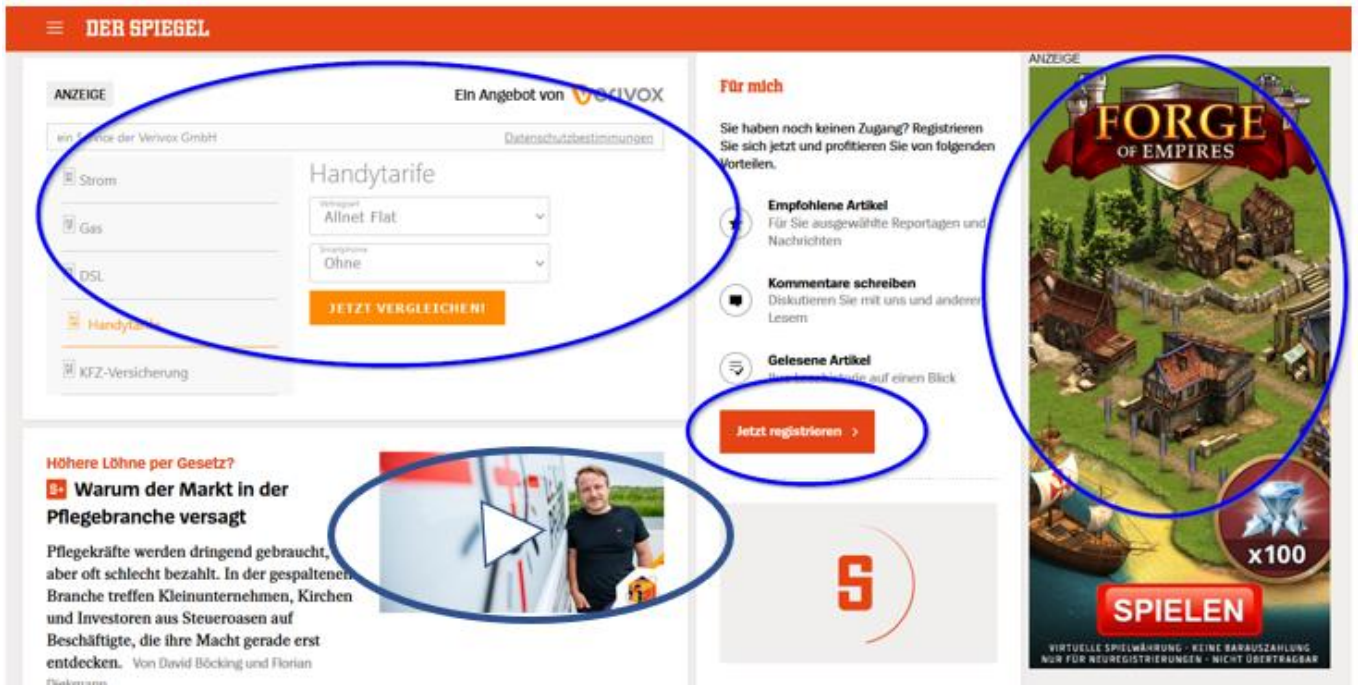
Elemente zur Aktivitätenverfolgung

zeit.de mit Werbung

Um der Nutzung mit Werbung zuzustimmen, muss JavaScript in Ihrem Browser aktiviert sein.

Können sich in nahezu jedem Bestandteil einer Webseite verbergen

- Werbebannern
- Formularen
- Anmeldefeldern
- Kommentaren
- Schaltflächen, Buttons
- unsichtbare Pixel
- Java Script
- Videos und Fotos
- …
- Links



Hyperlink auditing

- In einem Hyperlink können gleichzeitig zwei Webseiten aufgerufen werden
  - Ping Attribut
- einmal die Seite, die Sie sehen, wenn sie mit dem Cursor über den Link fahren und
- und eine zweite, unsichtbare Adresse - meist ein Skript - bei einem dritten Server, das registriert
  - was Sie aufrufen wollen
  - von welcher Seite sie kommen

Hübsche Ferienwohnung in Sylt

```
<a
 href="http://www.ferienwohnungen_in_sylt.com/page.html"
 ping="https://www.tracker.com/tracking.php"
>
 Hübsche Ferienwohnung in Sylt
</a>
```

- Hyperlink auditing findet sich in den Links
  - der Suchergebnisse von Google
  - bei Werbung, die Sie in E-Mails zugeschickt bekommen

Was sind Cookies?

Cookies sind kurze Textinformationen, die auf dem Endgerät gespeichert werden

Cookies

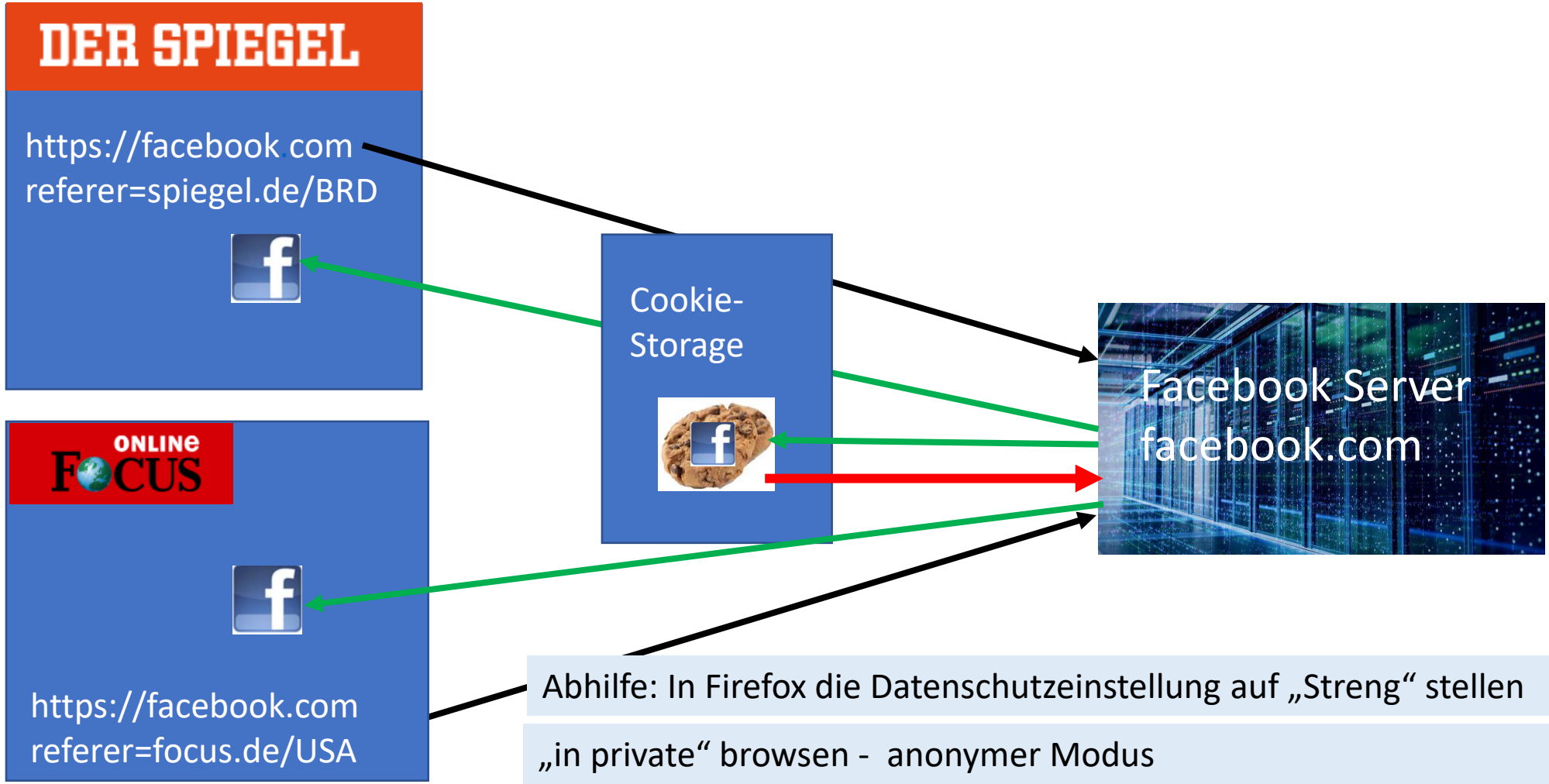
- werden durch Aufruf einer Webseite mithilfe des Browsers
  - angelegt und
  - an die platzierende Website/Domain zurückgeschickt – und nur an diese
  - abgespeichert werden z.B. Datum/Uhrzeit des letzten Besuchs der Website
- bestehen aus einem Namen und einem Wert (key-value pair) mit Attributen (optional)
  - z.B. Verfallszeit oder no expire
- können per html oder Java-Script gesetzt werden
- werden in einer Minidatenbank oder einem speziellen Speicher abgelegt
- Eine Virenübertragung durch Cookies gilt als ausgeschlossen.

Sind Cookies böse?

- Der Begriff **Cookie** wird im Datenschutz auch als **Synonym für Datenentnahme**, Datenspeicherung, Datennutzung, Datenverwertung, Datenweitergabe **wie auch Datenmissbrauch** verwendet. (Wikipedia)
- Cookies wurden erfunden, um den Aufrufer einer Webseite wiederzuerkennen, wenn er mehrfach Seiten derselben Website aufruft
- Das eigentliche Webseitenprotokoll (http) kennt keine solche Wiedererkennung des Aufrufs
  - Es ist statuslos (stateless protocol)

- Ohne Cookies wäre kein Warenkorb beim Online-shopping möglich
- Legitimerweise dienen Cookies zur Autorisierung beim Zugriff auf kostenpflichtige Inhalte
- Cookies ermöglichen Komfort beim Surfen, etwa durch das Speichern persönlicher Einstellungen
- Cookies sind aber auch das Einfallstor für das Tracking

So funktionieren Cookies und Cross-Site-Tracking



Abhilfe: In Firefox die Datenschutzeinstellung auf „Streng“ stellen

„in private“ browsen - anonymer Modus

Demo – Tracking mit Cookies

Vierundzwanzig 1st party Cookies beim Aufruf von Spiegel.de (Ausschnitt)

Vier AddOns

- **NoScript**
- blockiert Java Script Aufrufe
- **Privacy Badger**
- blockiert bekannte Tracker
- **Add block plus**
- blockiert Werbung
- **Cookies u. Header Analyser**
- zeigt Cookies an

Wie sehen cookies aus?

COOKIES AND HEADERS ANALYSER

EXPORT DATA SITE: DER SPIEGEL | Online-Nachrichten

COOKIE LIST

Name	Value	Expire	Domain	Path	Secure
sara_user_session	active	no-expire	.www.spiegel.de	/	false
AMCVS_79655FCF5C1D42160A495E15%	1	no-expire	.spiegel.de	/	true
s_cc	true	no-expire	.spiegel.de	/	true
s_sq	%5B%5BB%5D%5D	no-expire	.spiegel.de	/	true
euconsent-v2	CPHilaiPHilaiAGABCDEBdCgAP_AAAAA/ CQlwAFgAVAAyABwADwAIAAZAA0AB5A	no-expire	.spiegel.de	/	false

Sobald NoScript für spiegel.de erlaubt ist

- zeigt NoScript

35 verschiedene Webseitenaufrufe (20) per Java-Script

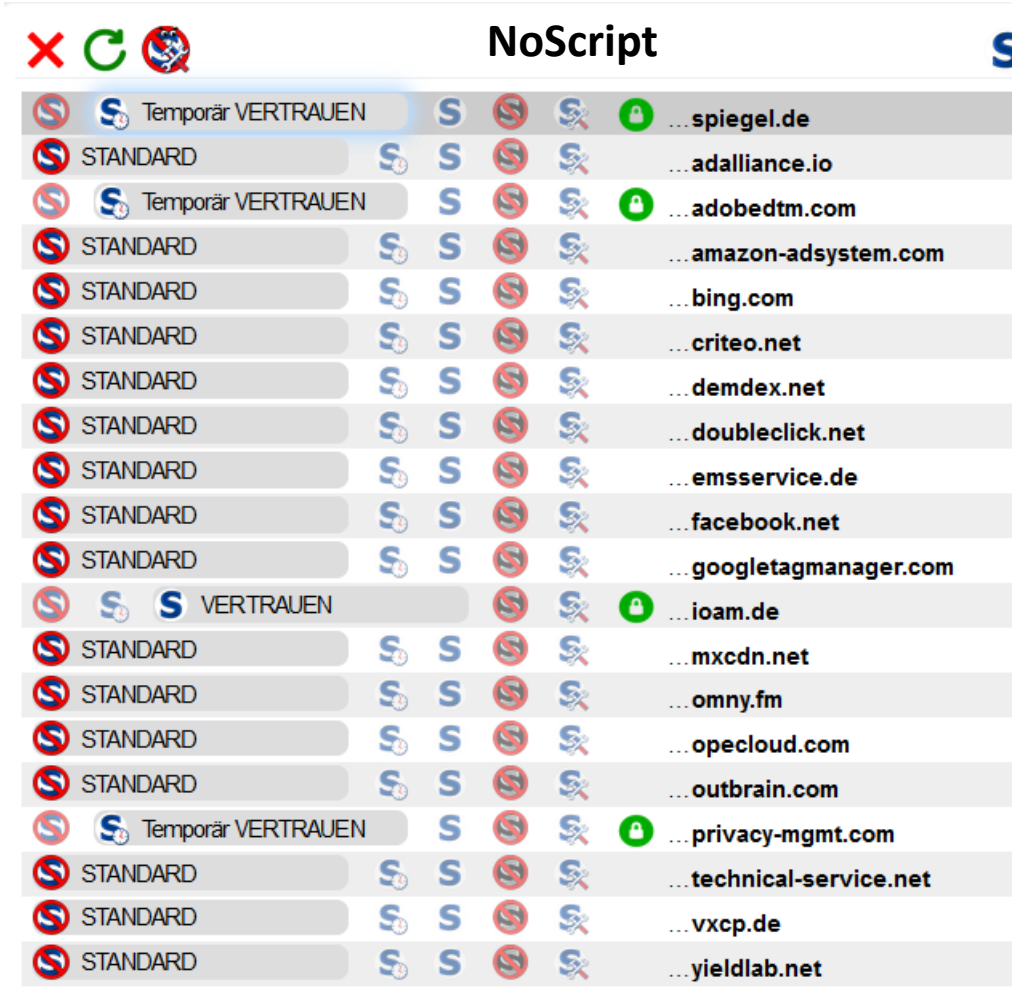
die alle - bis auf die explizit erlaubten-

blockiert wurden:

Der Aufruf von Skripten kann feingranular eingestellt werden:

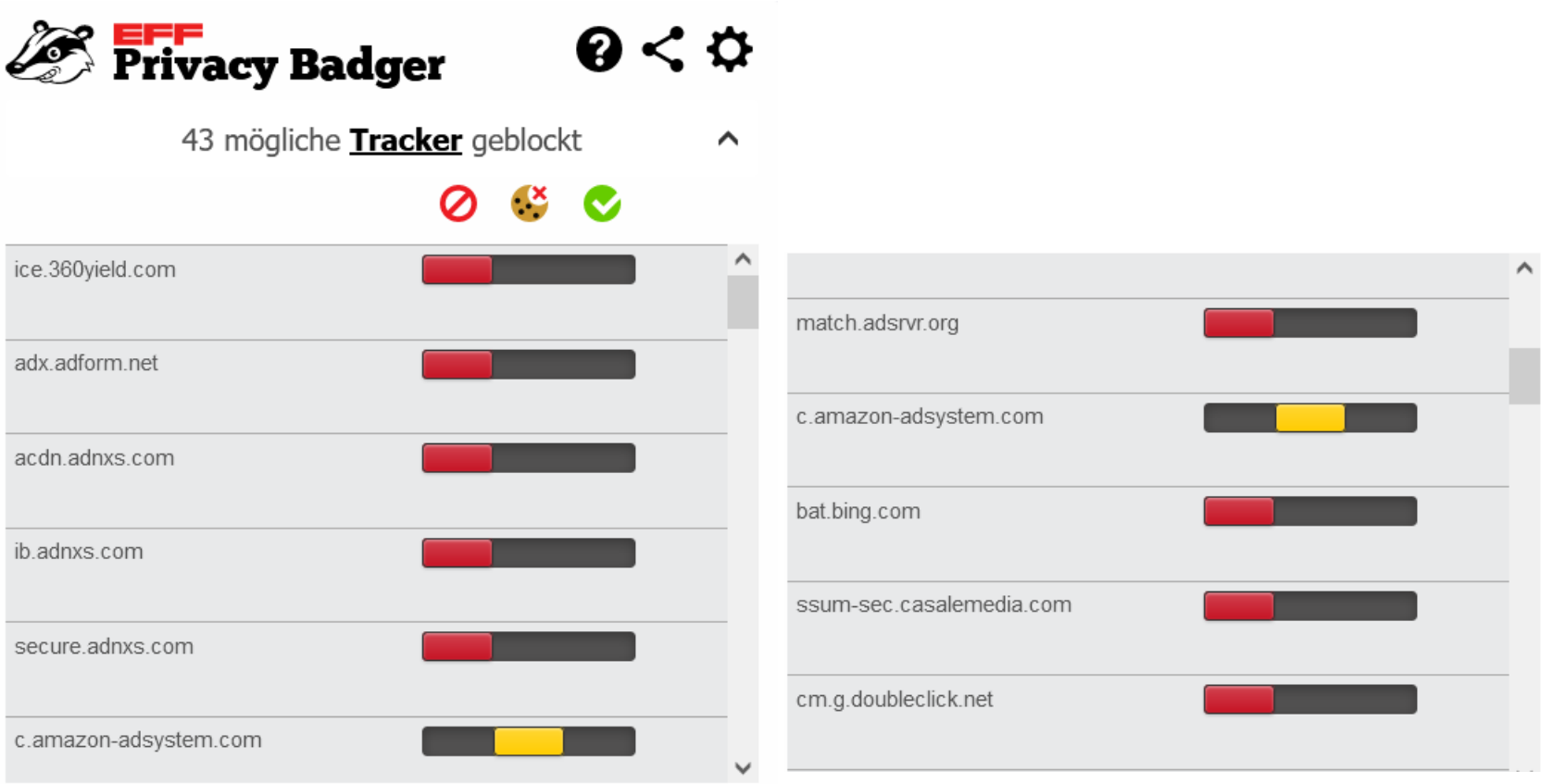
z.B. vertrauen und temporär vertrauen

Standard ist die Blockade



43 blockierte Aufrufe von bekannten Trackern

- Wir finden alte Bekannte, die auf spiegel.de vertreten sind und versuchen, Sie zu tracken
  - Amazon
  - bing
  - facebook
- und dazu die prominente Werbeindustrie



Ad block plus

- blockiert auf Spiegel.de
- sagenhafte 139 Aufrufe
- Sie können bei Hochzählen zuschauen

Moderne Cookies: Webstorage

- Webstorage dient als alternatives Speichermedium für Cookies
  - Speicherung durch Skripte im local storage: ohne Ablaufzeit
  - wiederverwendbar beim nächsten Besuch
  - bis 5 MB groß – keine automatische Übertragung an den Server
  - Entfernen nur durch Löschen des Browsercaches – Empfehlung: oft löschen!
- Skripte
  - Anweisungen, die durch den Browser ausgeführt werden
  - Sie ermöglichen
    - flexible Darstellung von Webseiten
    - unmittelbare Reaktion auf Eingaben des Benutzers (Serverzugriffe im Hintergrund - ajax)
    - Eingaben in Adresszeile oder bei der Suche
- Skripte sind äußerst vielfältig einsetzbar
  - dienen keineswegs ausschließlich der Überwachung der Benutzer
  - eignen sich dazu hervorragend
  - Benutzerverhalten zwischen zwei Eingaben überwachen
    - Wechsel von Browser Tabs - Ermittlung der Verweilzeit
    - Mauszeigerbewegung verfolgen

Exkurs: Was ist Programmatische Werbung?

- Im Rahmen einer Echtzeitauktion werden Werbeplätze passend zu den Interessen des jeweiligen Betrachters versteigert
  - Dazu dient eine hochspezialisierte Werbe- und Auktionsindustrie im Hintergrund
- Der Webseitenbetreiber signalisiert z.B., dass er Werbung für einen etwa dreißigjährigen Mann ausliefern kann, der sich für Sportartikel interessiert und ein älteres iphone verwendet
  - D.h. Ihre persönlichen Daten werden hier preisgegeben
  - Je mehr Informationen der Webseitenbetreiber zu dem Kunden zur Verfügung stellen kann, um so zielgenauer kann die Werbung angeboten werden
- Diese Informationen werden an Auktionsplattformen weitergereicht, die diese über weitere Stationen an eine Vielzahl von Werbetreibenden weitergeben
- In Sekundenbruchteilen wird der Werbeplatz an den Meistbietenden vergeben und die Werbe-URL wird in die Webseite eingebaut
- Ihr Browser holt die Werbung zusammen mit den 3rd-party Cookies des Werbenden und baut sie in die Webseite ein und der Betreiber bekommt eine Provision

Datenschleuder trotz DSGVO

- Auf die geschilderte Weise werden die verfügbaren Daten eines Anwenders im Internet verteilt und können von einer Fülle von Werbetreibenden zur Auswertung und Ergänzung ihrer eigenen Datenbestände verwendet werden
- **Preisfrage:** Wie oft kommt so etwas pro Tag – nur in der EU - vor?
  - 10 Millionen mal? 100 Millionen mal? 1 Milliarde mal?
- Es sind

84 Milliarden

- Bietervorgänge pro Tag
- Dies ergibt 200 pro Kopf der EU-Bevölkerung und Tag
- Überlegen Sie bitte kurz, welche Serverleistung dafür notwendig ist
- In Deutschland hat dieser Markt 2021 einen Wert von über drei Milliarden Euro
- 70% des Online Werbemarktes werden darüber abgewickelt
  - Quelle: c‘t Ausgabe 25/2021

Tracking durch Google Analytics

- Jeder Webserver registriert jeden Zugriff und speichert ihn in einer Protokolldatei
- Google Analytics ist eine Analyse-Software für Webseitenbetreiber, die die Zugriffe auf ihre Website auswertet
- Google Analytics wertet die Zugriffe ohne Umweg über Protokolldateien aus
  - Live Analyse des Webtraffics (Trackingpixel, Analytics Aufrufe per Skript)
  - D.h. alle Daten über Zugriffe werden zu Google übertragen, gespeichert und ausgewertet
- Sehr viele Websites nutzen diese leistungsfähige und z.T. kostenlose Auswertungssoftware von Google
- Webseitenbetreiber und ihre Nutzer werden gleichermaßen ausgeforscht und getrackt
  - Google gewinnt genaue Kenntnis über Nutzung (Verkaufs-Erfolg) jeder Website, die Analytics nutzt
  - Google erhält websiteübergreifende Erkenntnisse über das Verhalten der einzelnen Nutzer

- IP Adresse als identifizierendes Merkmal
  - geräteübergreifende Analyse im WLAN
  - V6 IP Adresse des Routers ändert sich nur einmal pro Tag
  - V4 Adresse bleibt u.U. über Monate stabil
  - ungefähre Standortanalyse wird möglich

Was ist Browserfingerprinting?

- Beim Browserfingerprinting werden **Hardware- und Softwareeigenschaften** sowie Einstellungen Ihres Browsers und Ihres Endgeräts (Computer, Tablets, Smartphone, …) durch Skripte ermittelt
- Mithilfe dieser Informationen gelingt es, ein genaues Profil (einen digitalen Fingerabdruck) von Ihrem Gerät zu erstellen, mit dem Sie auf verschiedenen Webseiten identifiziert und verfolgt werden können
- Der Fingerprint kann (als Hash) gespeichert und mit Anzeigen- oder Vertriebspartnern ausgetauscht werden
- Gegen Browserfingerprinting gibt es strenggenommen keinen Schutz. Es hilft kein
  - In-Private-Browsen oder die
  - Verwendung eines VPNs
  - nur eingeschränkt die Verwendung des TOR Netzwerks
    - obwohl der TOR Browser warnt und versucht einen für alle User identischen Browser darzustellen
      - NSA verfolgt User auch über das TOR Netzwerk (E. Snowden)
  - Die Verwendung eines jungfräulichen Systems (LINUX von DVD gebootet) schützt nicht

Fingerprinting ist böse

Zu den erfassten Merkmale Ihres Geräts bzw. Browsers zählen:

- das Modell Ihres Betriebssystems und Ihres Geräts
- Modell und Version Ihres Browsers
- die installierten Erweiterungen im Browser
- die Bildschirmgröße, Bildschirmauflösung, Farbtiefe, Spracheinstellung, Zeitzone
- Informationen über Ihre Netzwerkverbindung, z.B. Provider, Geschwindigkeit,
- die auf Ihrem Gerät installierten Schriftarten
  - notwendig für korrekte Darstellung
- Einzeln scheinen alle diese Informationen harmlos
- in ihrer Kombination sind sie verräterisch
- Skripte messen die Geschwindigkeit ihrer CPU, Audio Output, Treiber
  - in unsichtbaren Fenstern wird die Leistung ihrer Grafikkarte (GPU) gemessen
  - minimale Leistungsunterschiede werden ermittelt und dienen der Identifikation

Diese Technik eröffnet eine Vielzahl von Kombinationsmöglichkeiten, die in der Summe einen (beinahe) einmaligen Fingerabdruck ergeben.

My browser fingerprint
Are you unique ?

Yes! You are unique among the 3.509.361 fingerprints in our entire dataset.
<*https://amiunique.org/fp*>

Datenkrake Google

Google sammelt Informationen über Sie:

- Websiteübergreifend per Analytics
- Durch Ihre Suchmaschinenaufrufe
  - Google weiß, wonach Sie in den letzten zehn Jahren gesucht haben
- Ihre Klicks in den Suchergebnissen
  - Hyperlink auditing
- Über die Videos, die Sie bei Youtube anschauen
  - Über embedded Videos in den einzelnen Websites
- Über Ihre Standortinformationen (Standardortfreigabe)
  - Google verfolgt Sie buchstäblich auf Schritt und Tritt
  - Lassen Sie sich mal die Timeline zu Ihrem Google Konto anzeigen – Sie werden erschrecken
- Alle Informationen, die Sie unverschlüsselt in der Cloud speichern und
  - Fotos
  - E-Mails
  - Android Handy Backups
- Über den Playstore
- per Google Pay – weltweites Bezahlen

FLoC Googles Alternative zu 3rd party Cookies

FLoC: Federated Learning of Cohorts – Privacy Sandbox

Flock - Schafherde

- Der einzelne Nutzer soll nicht mehr persönlich getrackt werden
- Googles Browser Chrome berechnet aus der Browserhistorie des Nutzers kontinuierlich eine Art Hash-Wert
- Speicherung und Verarbeitung nur auf dem Endgerät – Löschung?
- alle Surfer mit denselben Interessen werden zu einer Kohorte zusammengefasst
  - Einzelpersonen verschwinden effektiv „in der Menge“
  - k-anonymity threshold
- Sturm der Entrüstung seitens der Werbeindustrie
  - Tests von FLoC zeigen, dass Werbetreibende mindestens 95 Prozent der Conversions pro eingesetztem Werbe-$ zu erzielen wie vorher mit Cookies
- Verspricht auf den ersten Blick einen Zugewinn an Anonymität

Gefahren durch FLoC

Es drohen vielfältige Gefahren

für den Wettbewerb

- der Browserhersteller
- der Werbeindustrie
  - Stärkung der Monopolstellung von Google (Search) und Google Ads
  - ganze Teile der Werbeindustrie könnten durch die neue Technik obsolet werden
- der Werbetreibenden
  - Vorteile für Datensammler, die bereits über große Bestände verfügen
  - erschwerter Marktzugang für neue Wettbewerber
  - verstärkter Verdrängungswettbewerb

für den Datenschutz der Nutzer

- Preisgabe sensitiver Informationen
  - Wer schon über Nutzerdaten verfügt, bekommt mit FLoC viele neue Informationen auf dem Silbertablett, ohne jahrelanges Tracking
  - Hoch interessant auch für Geheimdienste
  - Besonders kritisch für Dissidenten in undemokratischen Staaten

Apples opt-in Initiative

Platzierung von Cookies nur nach ausdrücklicher Zustimmung - gem. DSGVO

- Do not track als Voreinstellung auf Apple Geräten
  - gilt das auch für Apple selbst?
- Steht wirklich der Datenschutz im Vordergrund?
- oder will sich Apple Wettbewerbsvorteile durch Datenschutz sichern?
- Konfrontation mit Facebook, twitter, …
- Apple hat Hardware und Software gleichermaßen unter Kontrolle und kann effektiv blockieren
  - Anonymität im Netz gibt es nur mit Safari-Browser
  - Verschleierung der IP-Adresse nur mit dem Abo von iCloud+
- Es gibt bereits die ersten Werbetreibenden, die staatliche Eingriffe gegen Browserbeschränkungen fordern

Datenschutz bei Googles Pixel 6 – Android 12

Don‘t be evil – Do the right thing

- Google verspricht
  - das sicherste Handy
  - bestmöglichen Schutz für personenbezogene Daten und Fotos
  - volle Kontrolle über Kamera und Mikrofon
  - Apps können Rechte entzogen werden
- Grob irreführend
  - die Bewegungsdaten sind nicht sicher vor Googles Zugriff
  - ebenso die Surfverläufe und die Google Suche
  - E-Mails bei gmail
  - alle unverschlüsselten Daten und Fotos im Cloud-Store von Google
- Alle diese Daten werden von Google ausgewertet
  - Google versteht es Daten zu aggregieren und auszuwerten

Ist Werbung unerlässlich für ein freies Internet?

„Kostenlose“ Dienste?

- Die Bereitstellung von Content (Qualitätsjournalismus) und
- der Betrieb einer Plattform oder einer Serverfarm sind alles andere als kostenlos
- also muss diese Dienstleistung auch irgendwie bezahlt werden.
- Ohne (Werbe-) Einnahmen keine "kostenlosen"
  - Zeitungen
  - Nachrichten
  - Wettervorhersagen
  - Social Networks
  - Web-Speicher
  - Instant Messenger
  - …

Ohne Werbung verschwindet alles hinter einer Bezahlschranke

was nicht bedeutet, dass kein Tracking mehr stattfindet

Die Lösung?

- Datenschutzkonforme Werbung - notgedrungen hinnehmen
- aber: Tracking, Nein danke!
- Hoffnung auf den EU Privacy Act

Danke für Ihr Interesse

Fragen, Kommentare, Ergänzungen?

Hinweise

- Artikelserie zum Thema Tracking in der Zeitschrift c‘t des Heise Verlags
  - Ausgaben 12, 24 - 26 in 2021

  - Browsertests, Browserauswahl
  - Empfehlungen für
    - Browsereinstellungen
    - Add-Ons für mehr Sicherheit

- Netflix-Film zum Thema Überwachung und Beeinflussung durch Soziale Medien:

  - Social dilemma
  - Deutsch: Das Dilemma mit den sozialen Medien
    - Sehenswert - packend
    - Leider nur mit deutschen Untertiteln