KI zum Schutz der Privatsphäre und IT-Security von Konsumenten

Projektskizzen

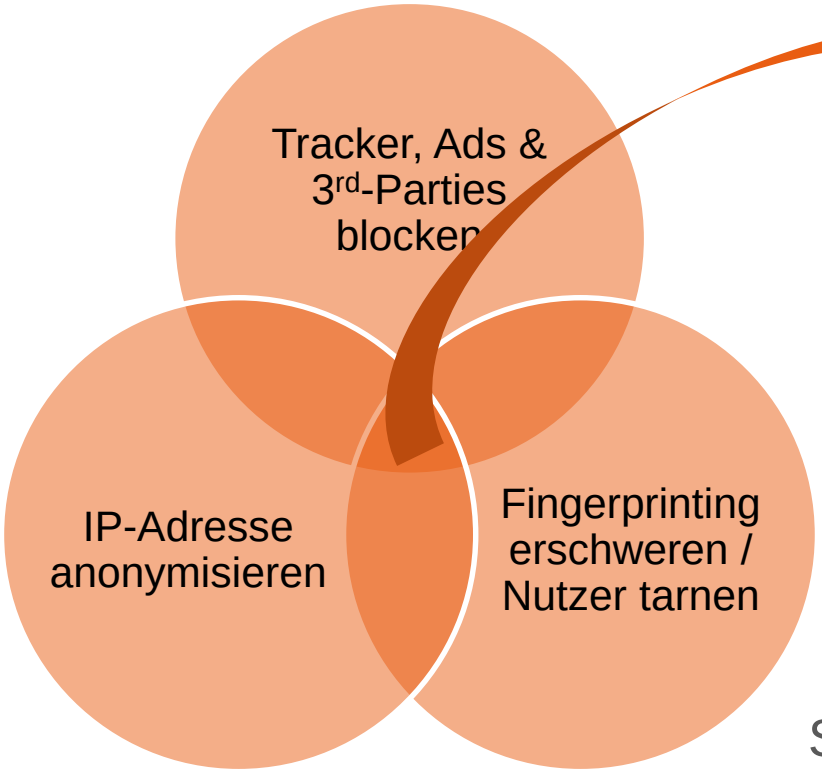
Dipl.-Math. Christian Bennefeld (aka „Benne“)
Gründer und Gesellschafter **etracker**
Gründer und Geschäftsführer **eBlocker**

Über eBlocker

Die **weltweit erste Plug & Play-Lösung** für
Privatsphäreschutz und **Jugendschutz** auf **allen** Geräten.

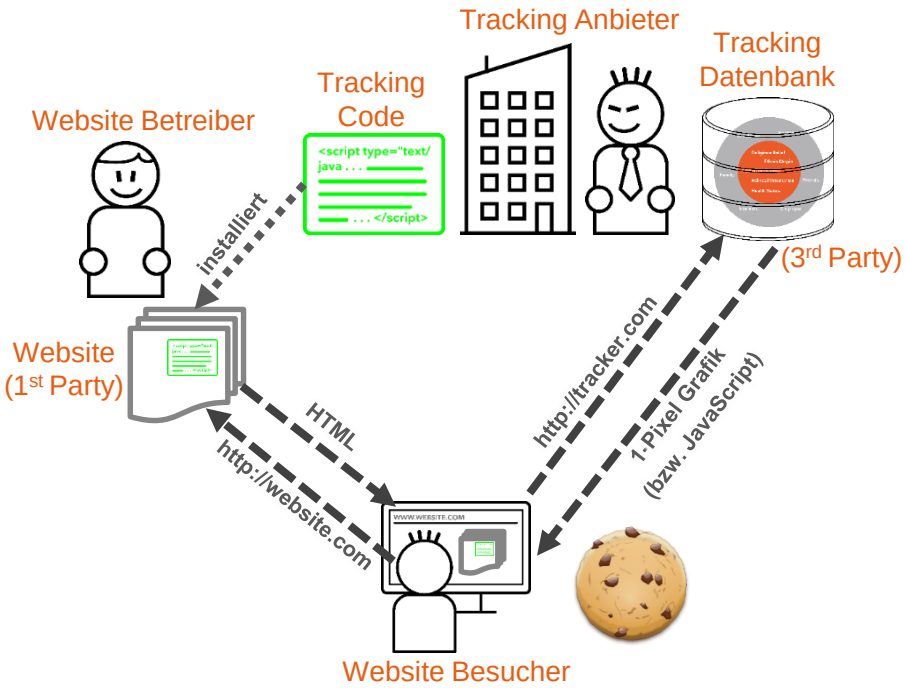
- **eBlocker GmbH**
  - Okt. 2014 von erfahrenen Gründern gegründet; mit 3+ Mio. € Venture Capital finanziert
  - Mit marktreifer Technologie **zehntausende Kunden** und **zahlreiche Innovationspreise** gewonnen
  - Mai 2019 **insolvent**; Hauptinvestor war sehr kurzfristig abgesprungen; **Geschäftsbetrieb eingestellt**
- **eBlocker Open Source UG**
  - Dez. 2019 gegründet (ehem. Gründer der GmbH); **Übernahme der Technologie** vom Insolvenzverwalter
  - **Ziel:** Marktreife eBlocker Plattform jedermann **kostenfrei** zur Verfügung stellen
    - **Non-Profit** auf **ehrenamtlicher** Basis; keine festen Mitarbeiter & keine Büros; Deckung aller Kosten **über Spenden**
    - **Software für Raspberry Pi** (Open Source Mini-Computer) zum **Geräte-Selbstbau**
    - **Open Source Entwicklung** gemeinsam mit Community: eBlockerOS 2.5 **Mitte Oktober 2020 erfolgreich veröffentlicht**
  - **Zukunftsvision: Machine Learning** für verschiedene Anwendungsgebiete nutzen
    - Generell **Mitstreiter** für unsere Vision gewinnen ☺
    - Forschungspartner mit ML-Hintergrund für konkretes **BMBF-Förderprojekt** gesucht (mittel- bis langfristiger Ausblick)

Komponenten für Privatsphäreschutz



Die eBlocker Idee:
Schutz **aller** Geräte auf Netzwerkebene

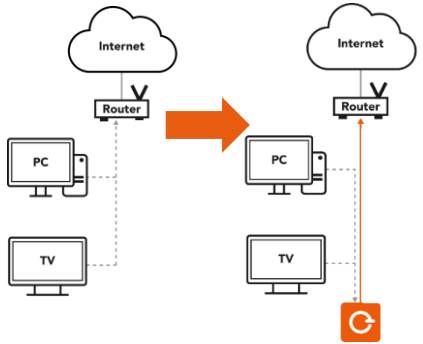
Wie funktioniert Tracking?



Sichtbare 3rd Party „Tracking-Pixel“

- **Social Plugins** wie Facebook Like-Button, Google +1
- **Online-Werbung** wie DoubleClick, Adsense, Taboola, …
- **Eingebundene Inhalte** wie Google Maps, Youtube, Twitter, …

Ausgangssituation & Status quo



- **eBlocker erhält als Gateway alle TCP/IP Pakete**
  - Gesamter In- und Outbound **IP-Verkehr**
    - Quelle (Gerät / MAC-/IP-Adresse im LAN)
    - Ziel (IP-Adresse / Domain)
    - TCP/IP- und HTTP-Header Informationen sowie Payload (Paketinhalt)
    - Antwort des Servers (Header / Payload / Größe etc.)
  - **DNS Anfragen** im Netz (welche Domains angesprochen werden) sind dem eBlocker bekannt
    - auch wenn keine Daten dorthin geschickt werden bzw. der Datenstrom geblockt wird
  - Auch (meist) bekannt bei HTTP(s): **Kontext der aufgerufenen Seite**
    - d.h. auf welcher Webseite werden welche weiteren Third Partys angesprochen (Bsp: auf abc.de sind 25 3rd Party Anbieter nämlich X, Y, Z)
  - „Ungewünschte“ 3rd Partys wie Tracker, Adserver, Malware etc. **werden heute geblockt**
    - **Basis: täglich aktualisierten Listen** mit Mustern der URL oder Domain
  - Alle Datenflüsse sind immer **lokal** im Nutzer-LAN (eBlocker ist kein Cloud-Dienst: Privacy First)
  - Viele **tausend Kunden** haben eBlocker aktiv im Einsatz
    - „eBlocker Insider“-Programm für die **Evaluierung von Innovationen im realen Kontext** möglich

Problemstellung

- Automatische Erkennung von **unbekannten Trackern** (ohne statische Listen)
  - **Hintergrund**: Täglich kommen neue Tracker / verändern ihre URLs / werden „gecloaked“
  - Tracker haben **besondere Eigenschaften**
    - Sind auf vielen Websites eingebunden (sprich, sie werden im Kontext von vielen Sites aufgerufen)
    - Werden i.d.R. mit langen URL-Parametern aufgerufen (tracker.com?Kunde=publisher.com&Seite=Home&…)
    - Liefern in der Regel nur wenige Byte zurück (z.B. SVG, Pixel, CSS, …)
  - **Idee**: ML lernt Tracker („Gutfall“) – erkennt andere Tracker, die auch „Gutfall“ aber unbekannt sind
  - **Mögliche Lösung**: Federated Learning von vielen Nutzern, Gesamt-Verarbeitung in RZ
    - Wie gelernte Daten „initial“ auf die Geräte bekommen (nicht bei Null anfangen)
    - Wie Privacy sicherstellen? Differential Privacy?
  - **Alternativer Ansatz**: Zentraler Web-Crawler analysiert Tracker und Datenströme im RZ
  - **Grundfrage**: Macht hier ML überhaupt Sinn, um unbekannte Tracker zu erkennen?

Problemstellung

- Automatische **Klassifikation von Websites** (primär für Jugendschutz)
  - **These**: Im Kontext einer Website-Kategorie (z.B. News, Porno, Gewalt) finden sich ähnliche 3rd Party Elemente (d.h. pornoads.com, schmuddeltracker.de, etc.)
  - **Idee**: Klassifikation mit ML durch Lernen der Kontexte einzelner Kategorien
    - Lernen von „News Websites“, „Porno“ etc.
    - Herausforderungen ähnlich wie auf Folien zuvor
  - **Alternativer Ansatz**: Analyse des Text-Payloads, der ausgehenden Links oder der Bilder
    - Aufbau eines gerichteten Graphen – welche Site linked auf welche anderen, Clustererkennung und Klassifikation
    - **Problem Rechenleistung**: Auspacken / Interpretieren des Payloads (sprich: Bild / Text / HTML) in Real Time / bzw. Speicherung der Inhalte für spätere Analysen ist teuer / problematisch
  - **Fragestellungen**: Ähnlich zuvor ☺

BMBF Förderprojekt

- Unterstützung von Bürgerinnen und Bürgern bei der **privaten IT-Sicherheit**
  - https://www.bmbf.de/foerderungen/bekanntmachung-3160.html
- „Gegenstand ist die Erforschung und Entwicklung von Methoden und Werkzeugen, um Bürgerinnen und Bürger bei der Umsetzung ihrer **privaten IT-Sicherheit und dem Schutz ihrer privaten Daten zu unterstützen**“
- **Unser Ziel:** Partner aus der Forschung für unser Vorhaben gewinnen
  - Konsortialpartner müssen diese Woche „stehen“
  - Bei Interesse: Bitte **kurzfristig melden** ☺

Problemstellung für Förderprojekt

- **Intrusion Detection**: Erkennung von Angreifern oder angegriffenen Geräten
  - **These**: Bei Einbruch ins Netzwerk verändern sich Datenflüsse
    - Z.B. IP-Kamera schickt viele Pakete zu neuen bisher nicht angesurften Domains (Teil eines Bot-Netzes)
    - Hacker/Malware im Netzwerk kontaktieren sog. Command & Control Server (über DNS-Auflösung)
    - Datenausleitungen generieren ungewöhnliche Traffic-Ströme (viele Daten, „anormale“ Payloads/Pakete)
    - Generell: Es entstehen „Anomalien“ im Netzwerk gegenüber „ungehacktem“ Netz
  - **Idee**: Per ML „Gutfall“ lernen, Abweichungen von der „Norm“ erkennen und notifizieren
  - **Stand der Forschung**: Erste Ansätze für ML zur Erkennung von Netzwerkanomalien existieren
    - Primär im Labor und / oder im größere Business-Kontext
  - **Herausforderung**: Geringe Rechenleistung; ggf. auch zu wenig „Gutfall“ Daten in einem LAN
  - *Lösung & Fragestellung wie zuvor (s. Tracker-Erkennung):*
    - ***Mögliche Lösung**: Federated Learning von vielen Nutzern, Gesamt-Verarbeitung in RZ*
      - *Wie sicherstellen, dass nur keine infizierten Datenströme gelernt werden?*
      - *Wie gelernte Daten „initial“ auf die Geräte bekommen (nicht bei Null anfangen)*
      - *Wie Privacy sicherstellen? Differential Privacy?*
    - ***Grundfrage**: Macht hier ML überhaupt Sinn, um unbekannte Abweichungen zu erkennen?*

Vielen Dank!

Lust mitzuwirken?

eBlocker.org
Raspberry Pi Images

github.com/eblocker
Open Source Code

voluntary@eBlocker.org
Kontakt für Unterstützer

eBlocker Core Architecture