Tobias Scheible, M.Eng.

Deanonymisierung

Gefahren und Chancen im Web

Tobias Scheible, M.Eng.

Hochschule Albstadt-Sigmaringen
Albstadt-Sigmaringen University

- 1999 GeoCities Website, 2000 eigene Domain, 2001 erste Projekte
- 2009 bis 2012: Softwareingenieur im Bereich Web Development
- Seit 2012: Wissenschaftlicher Mitarbeiter an der Hochschule Albstadt-Sigmaringen Im Bereich IT-Sicherheit & Forensik

Forschungsprojekt SEKT (BMBF) IT Security & Smart Textiles	Grundlagen der digitalen Forensik Masterstudiengang IT GRC Management	Digitale Forensik (Übungen) Bachelorstudiengang IT Security
Internet Grundlagen Masterstudiengang Digitale Forensik	Cybersecurity (Seminar + Praktikum) Bachelorstudiengang IT Security	Internettechnologien Hochschulzertifikatsprogramm
Cloud Technologies and Cloud Security Masterstudiengang IT GRC Management	Betriebssystemforensik Masterstudiengang Digitale Forensik	Informationssicherheit (Praktikum) Bachelorstudiengang Wirtschaftsinformatik
Blog scheible.it Cyber Security & IT-Forensik Artikel	Zeitschriftenartikel Web & Mobile Developer, <kes>-Magazin, …	Vorträge & Workshops Unternehmen und u. a. für den VDI und die IHK

Agenda

- Anonym und Pseudonym
  - Pseudonymisierung
  - Anonymisierung
  - Deanonymisierung
- Beispiele aus der Praxis
  - Krankmeldung
  - Webbrowser Fingerprinting
  - VPN-Zugang
- Beispiele aus der Forschung
  - Netflix und IMDb Verknüpfung
  - New Yorker Taxifahrer
  - Zuordnung mit Merkmalen
- Einsatzszenarien im Darknet
  - Tor Netzwerk
  - Kontrolle von Exit-Nodes
  - Analyse des Netzwerkverkehrs
  - Ausnutzung von Schwachstellen

Hinweis

Die kompletten Folien der Präsentation mit Zusatzinformationen werden im Blog unter *www.scheible.it* bereitgestellt.

Anonym und Pseudonym

Pseudonymisierung

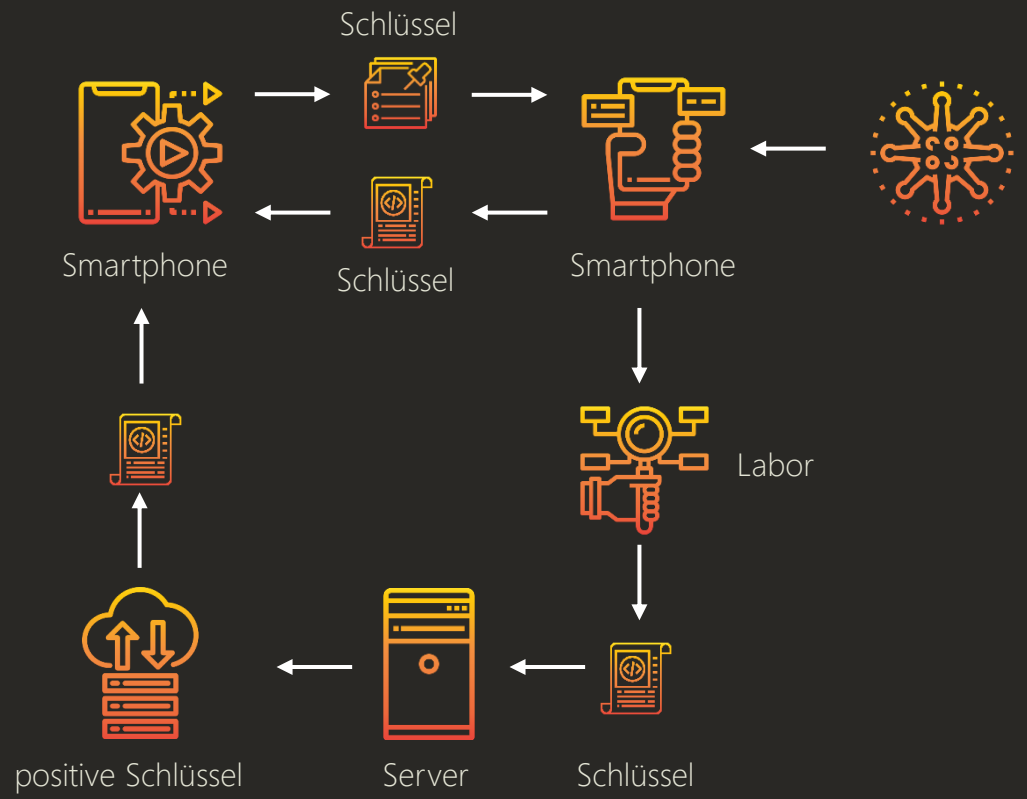
Datensatz

Beispiel Pseudonymisierung

Das Tracing in der Corona Warn-App kann nur pseudonymisiert stattfinden, da ansonsten keine Warnungen mehr möglich wären.

CORONA WARN-APP

(1)

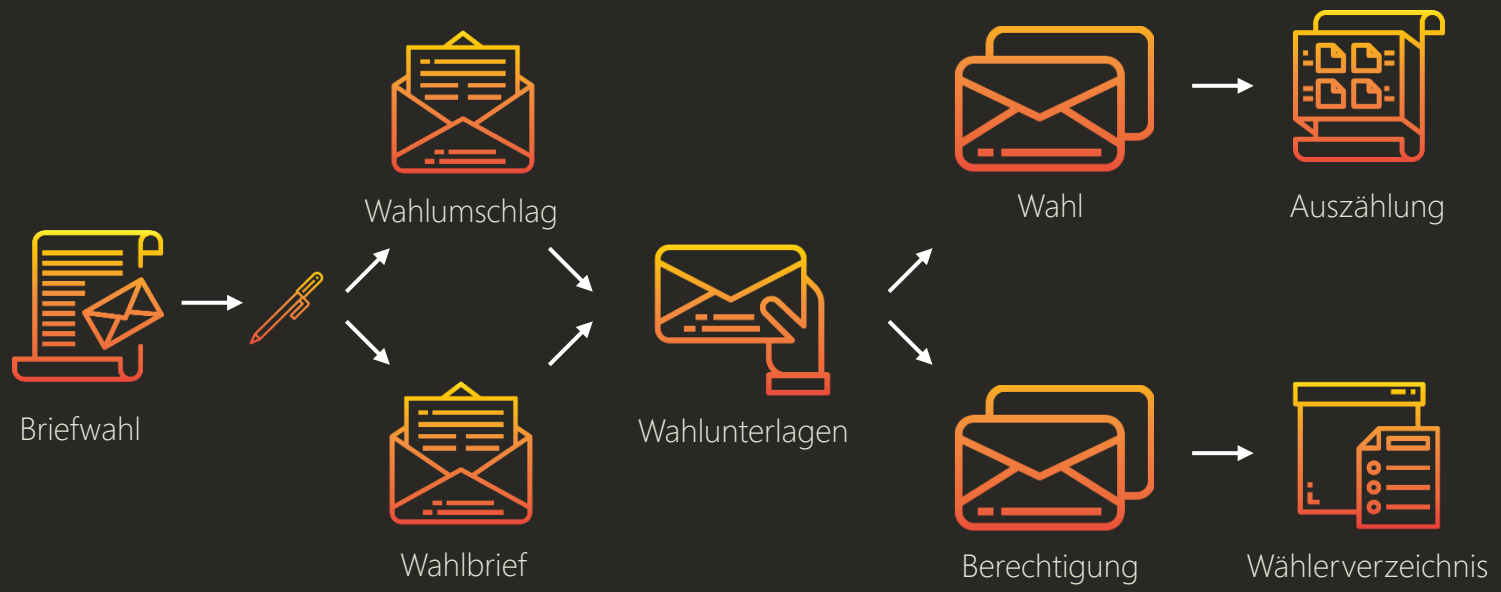


Anonymisierung

Datensatz

Beispiel Anonymisierung

Bei einer geheimen Wahl ist nachvollziehbar, wer gewählt hat, aber eine Zuordnung zwischen Wahlzettel und Wähler ist nicht möglich.



Deanonymisierung

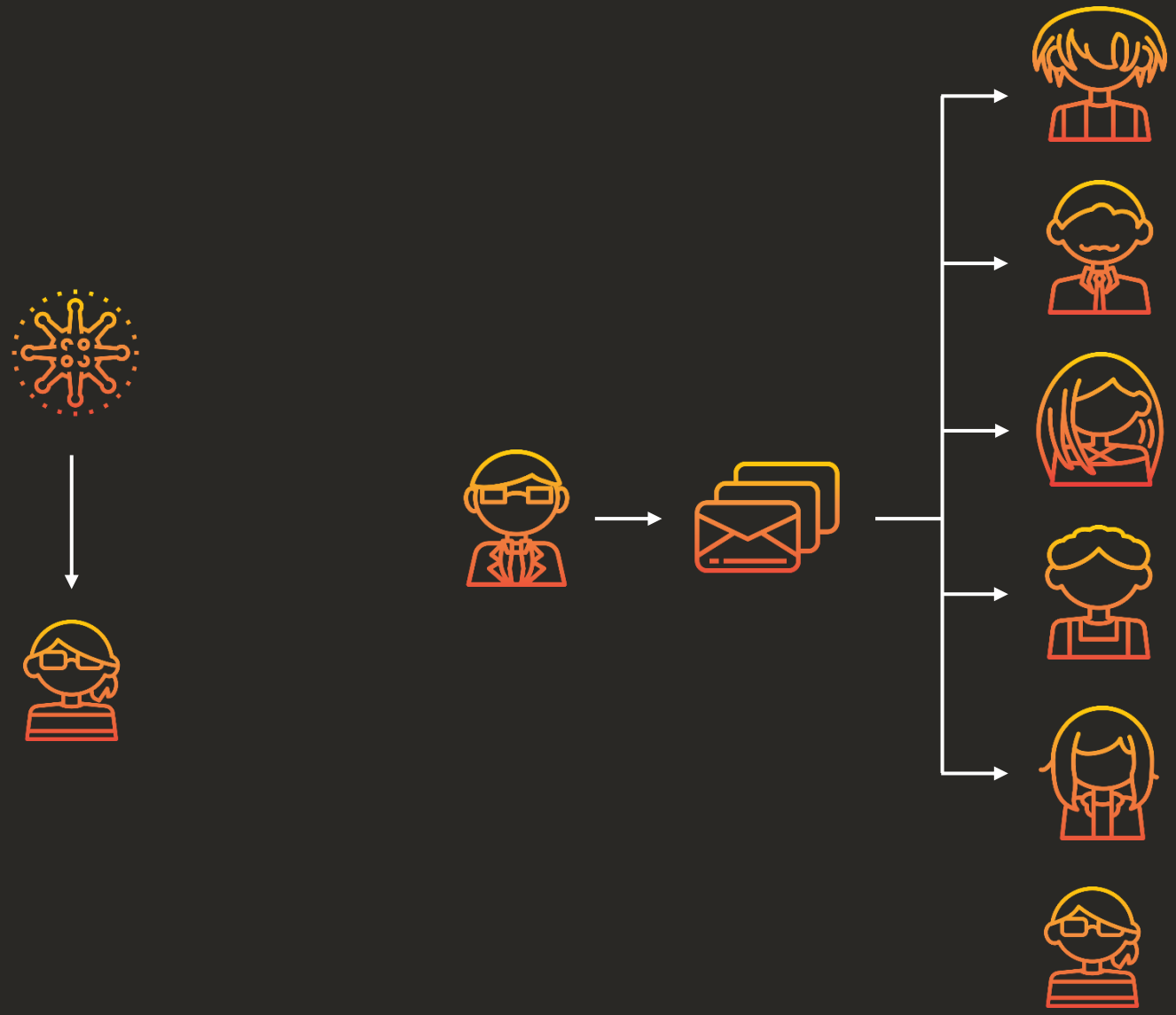
anonymisiert und pseudonymisiert

Gegenmaßnahmen

- Möglichst wenig Daten angeben bzw. preisgeben
- Fake-Daten bei Anbieter ohne Verträge (Kaufabschluss)
- Gezielte Falschinformationen – Generatoren für Profile und Fotos
- Anonyme Wegwerf-Email-Adresse (Trash-Mails) verwenden
- Online SMS Empfangsservice bei Handyzwang nutzen

Beispiele aus der Praxis

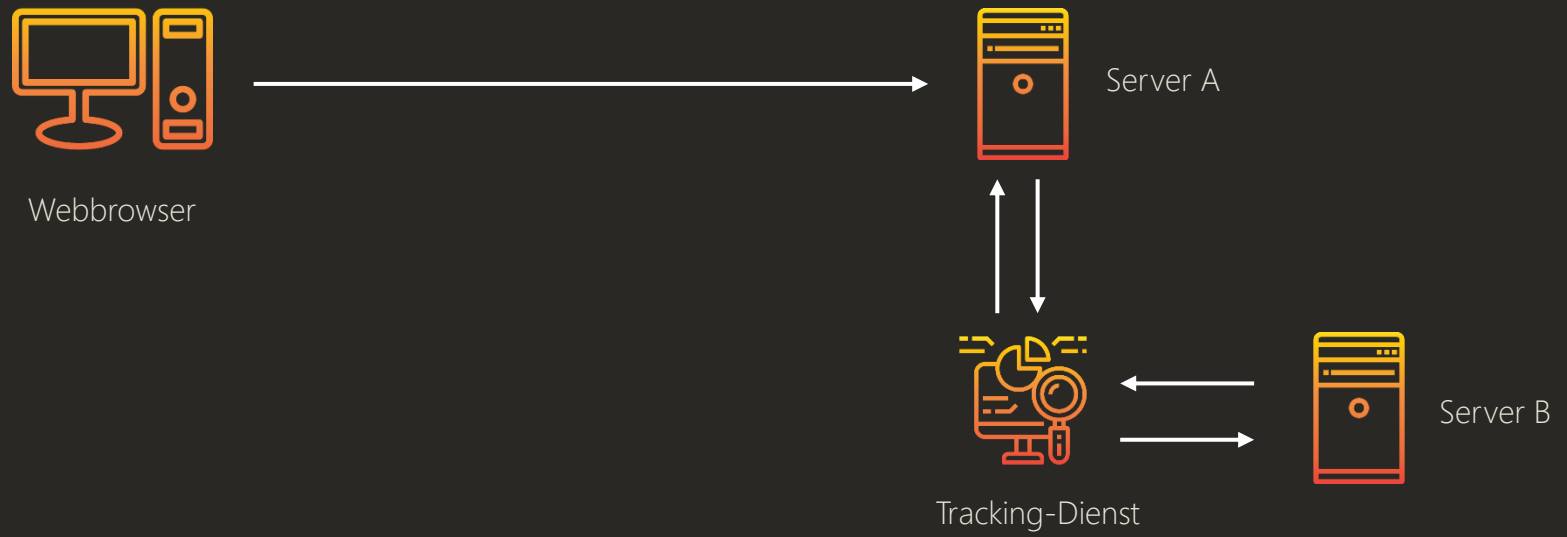
Krankmeldung



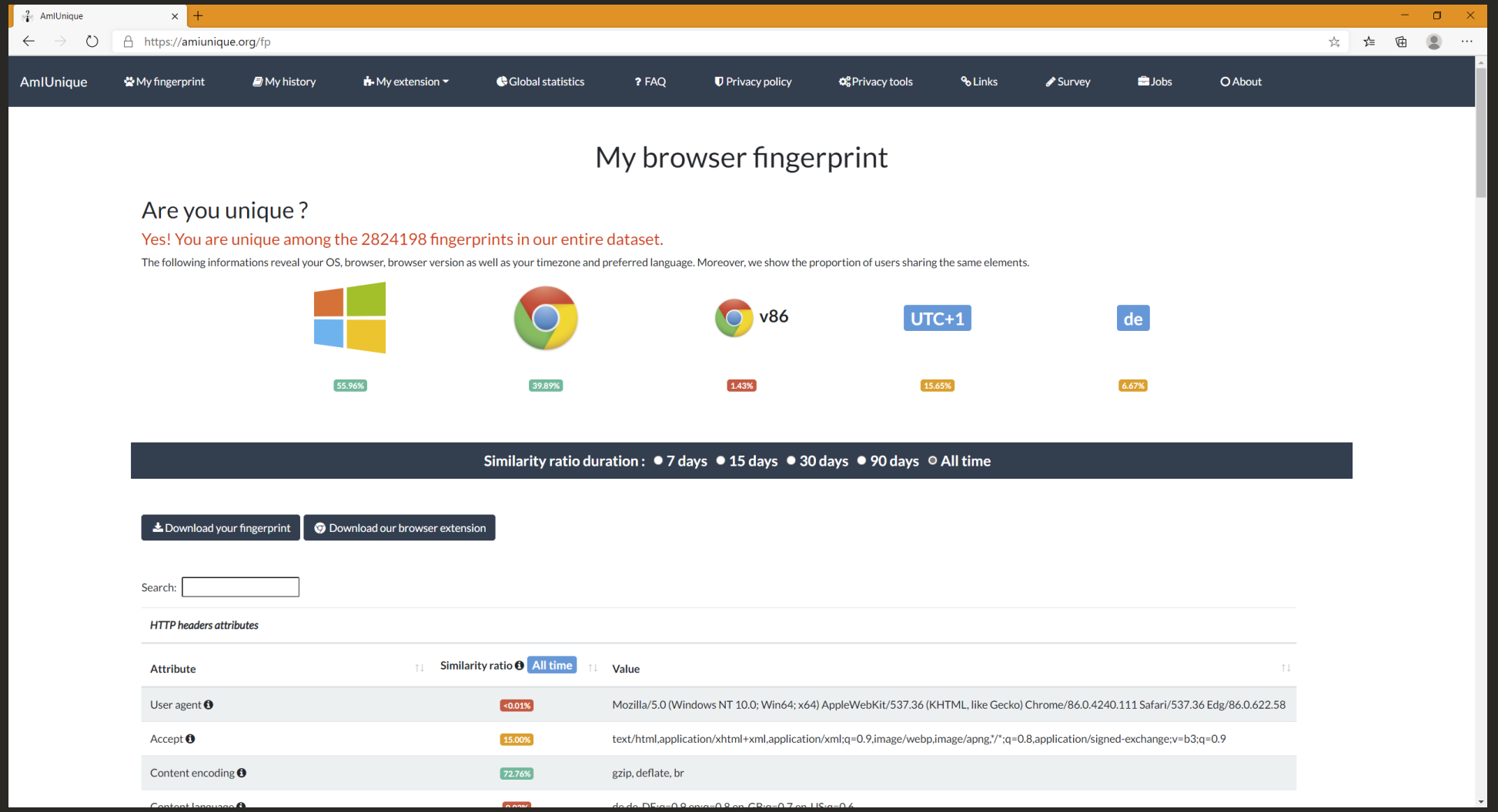
Webbrowser Fingerprinting

Mit Browser-Fingerprints können Browser im Internet eindeutig identifiziert und über mehrere Webseiten hinweg getrackt werden.

- Funktioniert bei ca. 90 Prozent aller Webbrowser
- Dazu werden keine Cookies gesetzt
- Ist mit allen Geräten möglich (Rechner, Smartphone & Tablet)



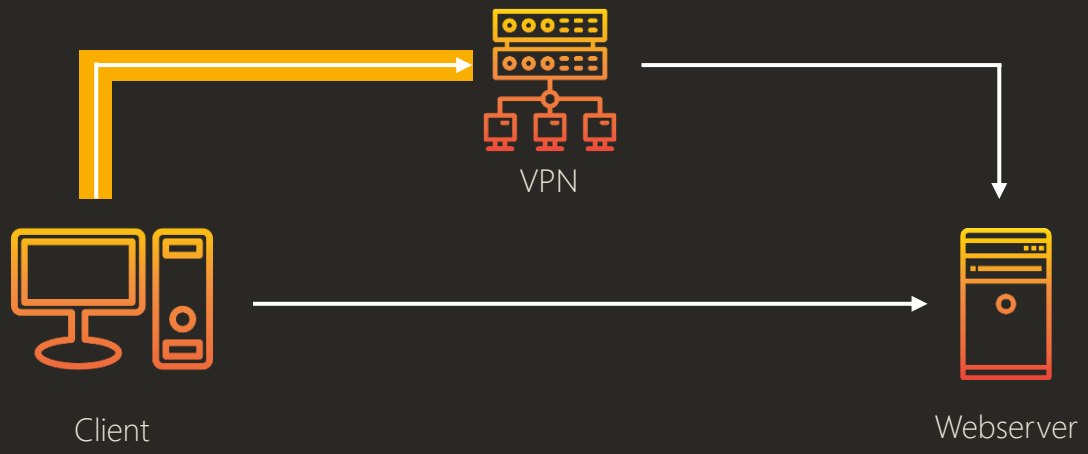
Beispiel AmIUnique Dienst



VPN-Zugang

Die Abkürzung VPN steht für "Virtual Private Network" (virtuelles privates Netzwerk), damit werden verschlüsselte Verbindungen realisiert.

- Verschleiert nur die eigene IP-Adresse
- Der VPN-Anbieter kennt alle Zugriffe und Daten
- Das Werbeversprechen „Anonym surfen“ ist eine Falschaussage



Gegenmaßnahmen

- Zwei-Browser-Strategie
  - Tägliches Surfen: Sichere Konfiguration oder dauerhafter „privater Modus“
  - Dienste mit Login: Aktueller Webbrowser ausschließlich für Seiten mit Logins
- Werbeblocker
  - Werbeblocker ohne kommerzielle Ausrichtung wie z.B. uBlock Origin
  - Pi-Hole - Ein schwarzes Loch für Internet-Werbung
- Integrierter Schutz vor Aktivitätenverfolgung nutzen
- Maximaler Schutz: JavaScript deaktivieren

Beispiele aus der Forschung

Netflix und IMDb Verknüpfung

- Veröffentlichung von 100.480.507 Datensätzen in anonymisierter Form
  - Wettbewerb zur Verbesserung des Empfehlungsservices
  - Alle personenbezogenen Daten wurden erfolgreich entfernt
  - => Keine Deanonymisierung möglich
- Verknüpfung der Datensätze mit Bewertungen der Internet Movie Database (IMDb)
  - Korrelationen zwischen Daten und Inhalten der Bewertungen gleicher Filme
  - Der Fokus lag auf der Anzahl an Bewertungen, die für eine Identifizierung notwendig waren
  - => Mit 8 Bewertungen konnten bereits 99% der Datensätze zugeordnet werden

Quelle: securityfocus.com (2)

New Yorker Taxifahrer

- Anfrage von Chris Whong auf Basis des „Freedom of Information Act“
  - ca. 170 Millionen Datensätze zu Taxifahrten

	A	B	C	D	E	F	G	H	I	J	K
1	medallion	hack_license	vendor_id	pickup_datetime	payment_type	fare_amoun	surcharge	mta_tax	tip_amount	tolls_amoun	total_amount
2	89D227B655E5C82AECF13C3	BA96DE419E711691B944	CMT	1/1/13 15:11	CSH	6.5	0	0.5	0	0	7
3	0BD7C8F5BA12B88E0B67BED	9FD8F69F0804BDB5549F	CMT	1/6/13 0:18	CSH	6	0.5	0.5	0	0	7
4	0BD7C8F5BA12B88E0B67BED	9FD8F69F0804BDB5549F	CMT	1/5/13 18:49	CSH	5.5	1	0.5	0	0	7
5	DFD2202EE08F7A8DC9A57B0	51EE87E3205C985EF843	CMT	1/7/13 23:54	CSH	5	0.5	0.5	0	0	6
6	DFD2202EE08F7A8DC9A57B0	51EE87E3205C985EF843	CMT	1/7/13 23:25	CSH	9.5	0.5	0.5	0	0	10.5
7	20D9ECB2CA0767CF7A01564	598CCE5B9C1918568DEE	CMT	1/7/13 15:27	CSH	9.5	0	0.5	0	0	10
8	496644932DF3932605C22C79	513189AD756FF14FE670	CMT	1/8/13 11:01	CSH	6	0	0.5	0	0	6.5
9	0B57B9633A2FECD3D3B1944	CCD4367B417ED6634D9	CMT	1/7/13 12:39	CSH	34	0	0.5	0	4.8	39.3
10	2C0E91FF20A856C891483ED6	1DA2F6543A62B8ED934	CMT	1/7/13 18:15	CSH	5.5	1	0.5	0	0	7

- MD5-Hash: „Tobias“ => 2d95188a755067ac95e25f90b6e7c1ab
  - Groß- und Kleinbuchstaben & Ziffern (62 verschiedene Zeichen)
    - Nummernschild: 6 Zeichen | Taxilizenz: 6 Zeichen | 62^{12} = 3.226.266.762.397.899.821.056
  - Schema ist bekannt: Nummernschild: 2 Millionen & Lizenznummern: 22 Millionen
    - Nur 24 Millionen Varianten | Lizenznummern und Namen der Fahrer frei verfügbar
- => Vollständige Deanonymisierung aller Taxifahrten

Quelle: chriswhong.com (3)

Zuordnung mit Merkmalen

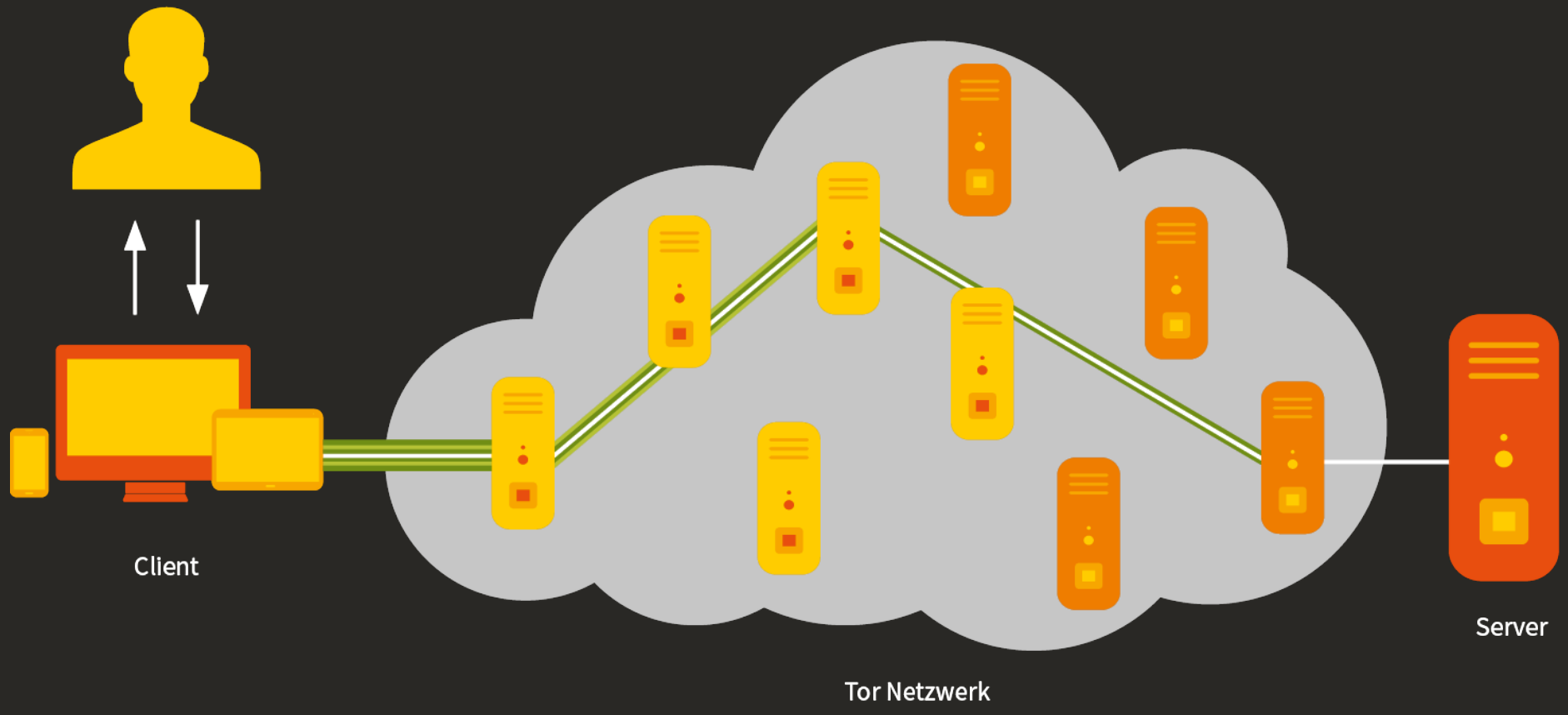
In einer Untersuchung konnten Forscher 99,98 Prozent der US-Amerikaner in jedem Datensatz eindeutig identifizieren.

- Benutzt wurden die Daten aus dem US-Zensus und frei zugängliche Umfragen von Universitäten, die angeblich anonym waren
- Sämtliche direkt identifizierenden Daten wie der Name, die Adresse oder die E-Mail entfernt
- 83 Prozent der US-Bürger können klar identifiziert werden, selbst wenn nur Geschlecht, Postleitzahl und Geburtstag bekannt sind
- Für eine 99,98 Prozent Quote werden 15 Merkmale benötigt

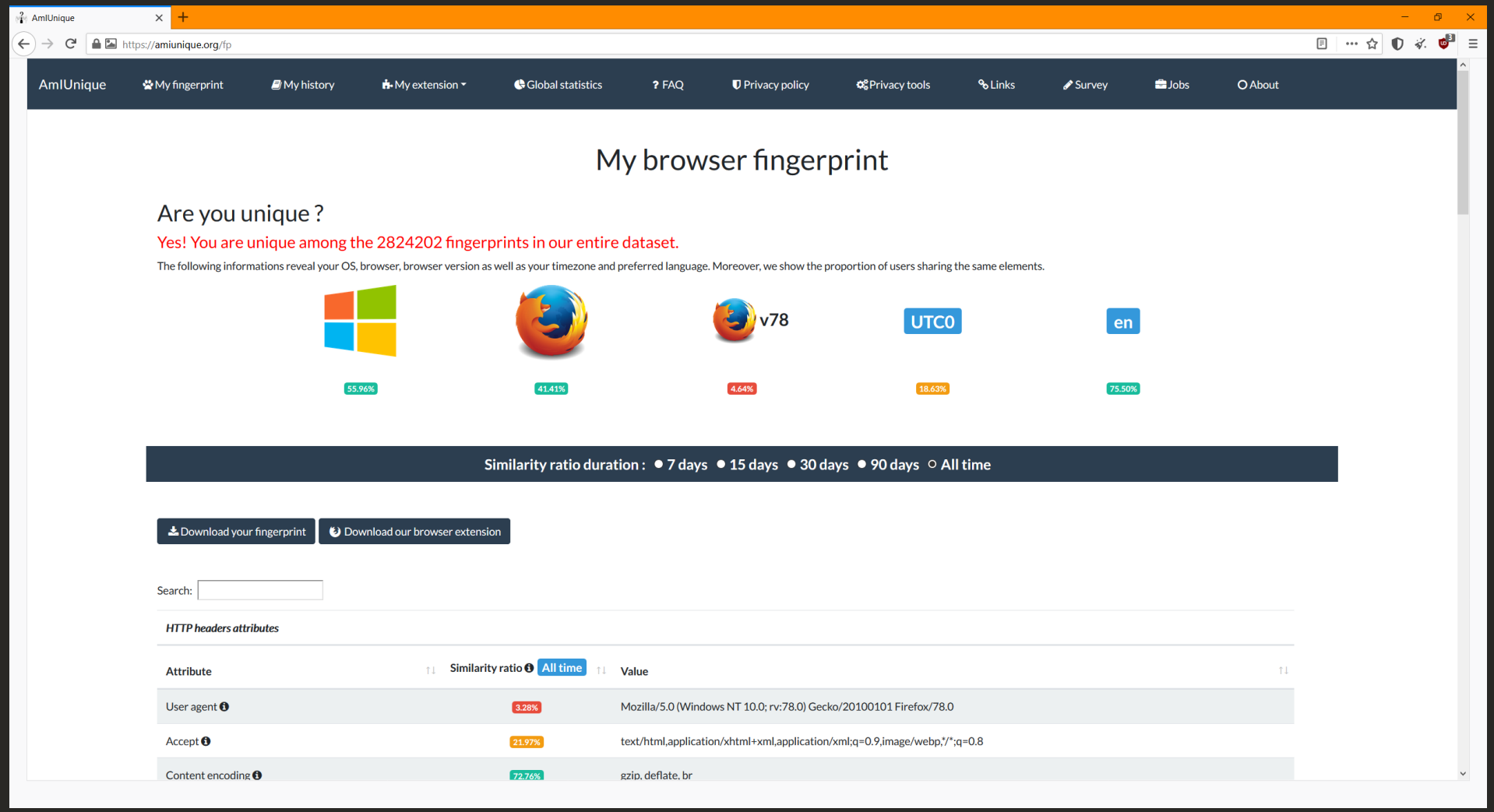
Quelle: nature.com (4)

Einsatzszenarien im Darknet

Tor Netzwerk



Beispiel AmIUnique Dienst



Anonym und Pseudonym

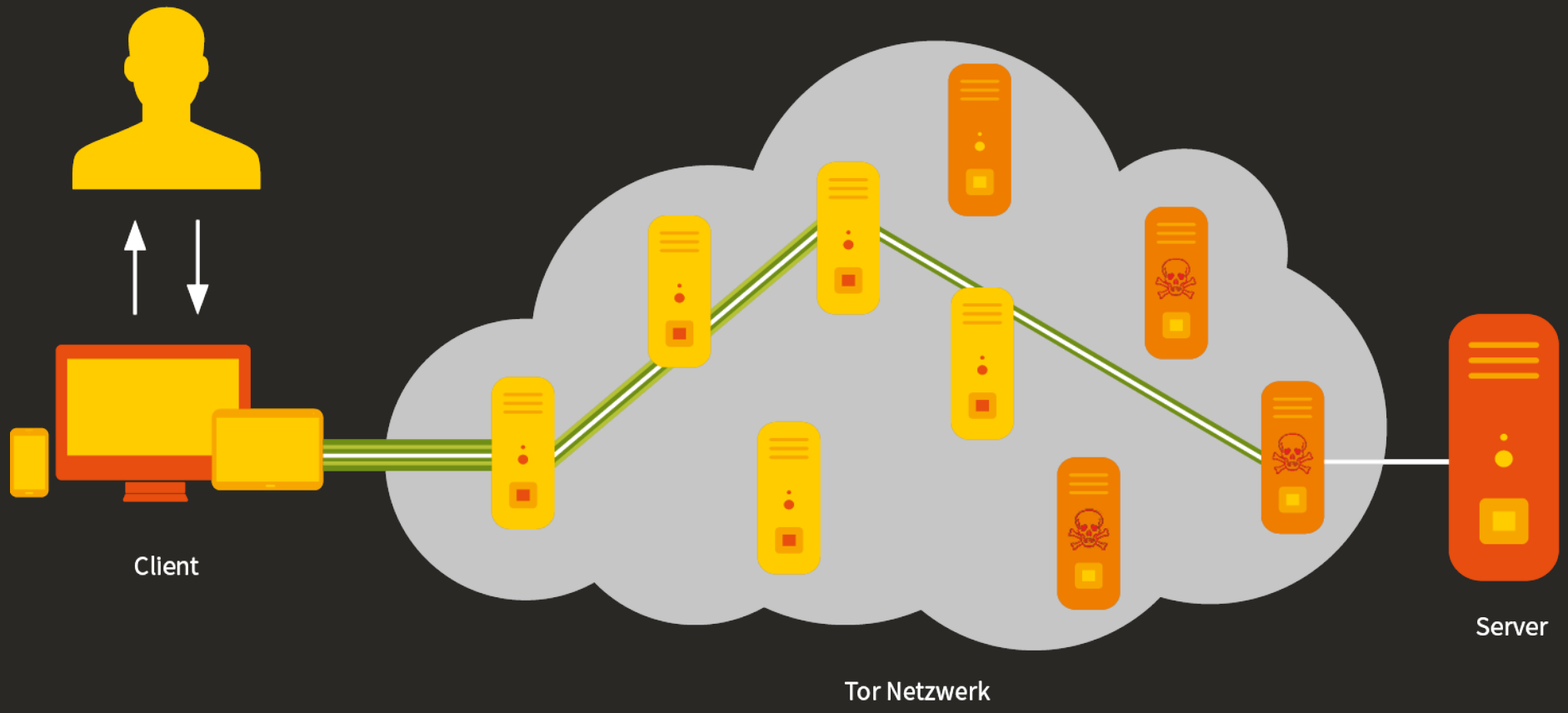
Beispiele aus der Praxis

Beispiele aus der Forschung

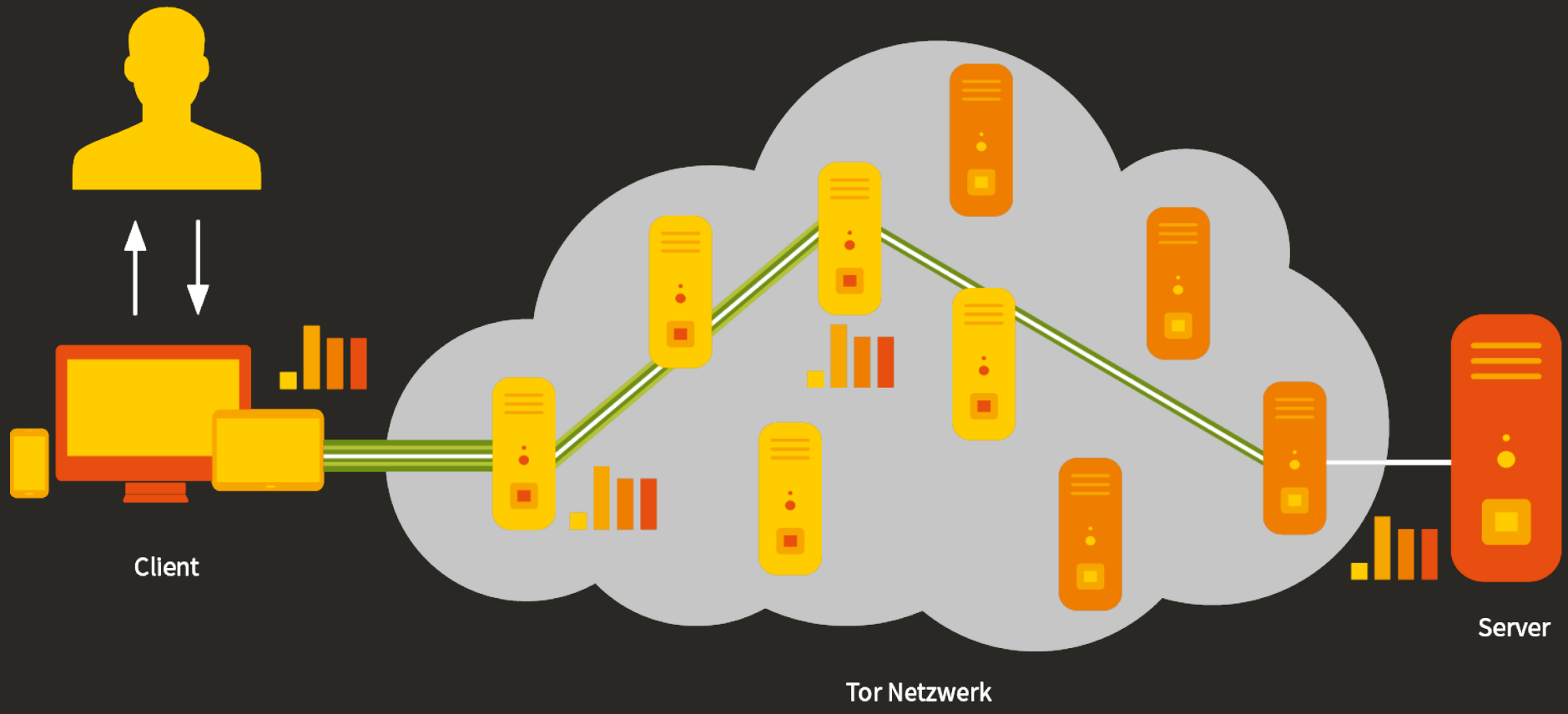
Einsatzszenarien im Darknet

Tor Netzwerk
Kontrolle von Exit-Nodes
Analyse des Netzwerkverkehrs
Ausnutzung von Schwachstellen

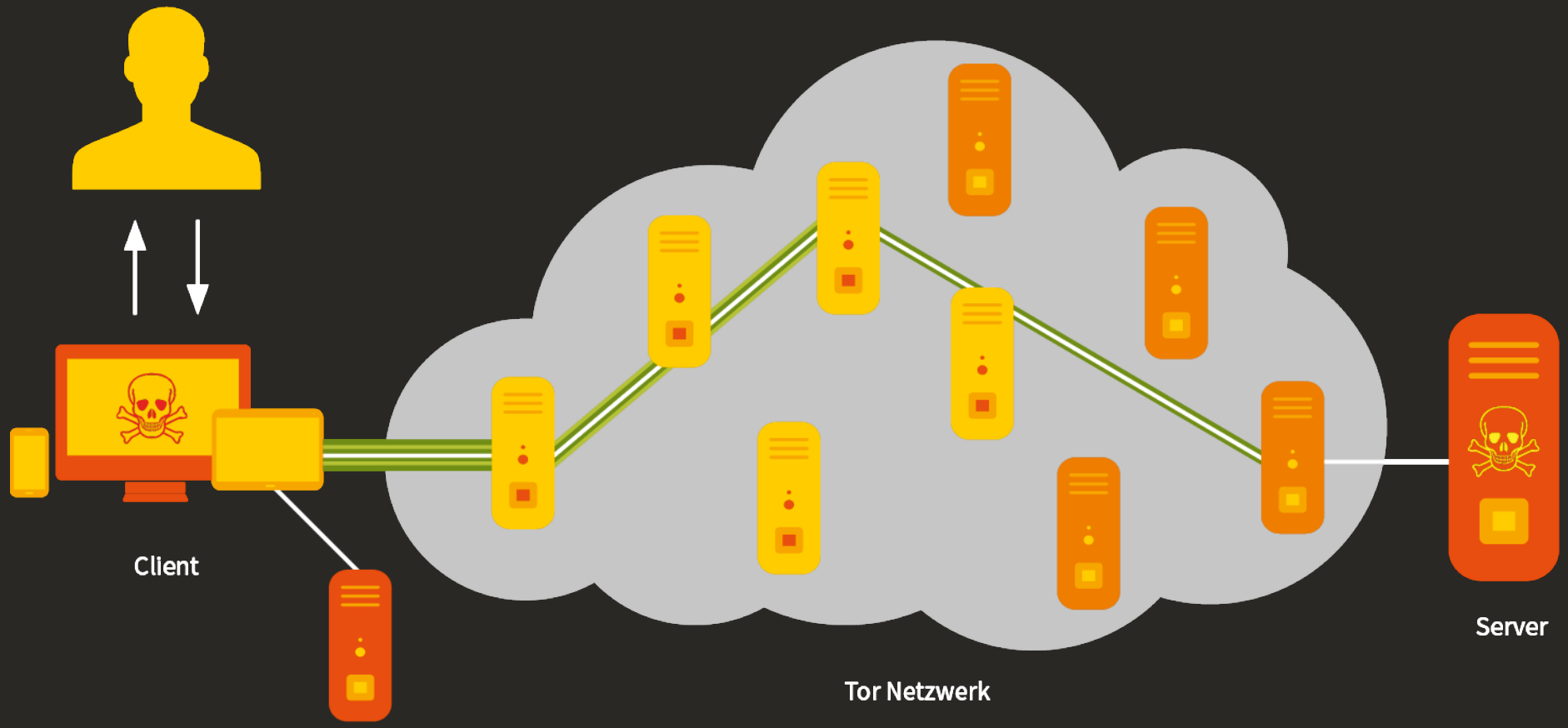
Kontrolle von Exit-Nodes



Analyse des Netzwerkverkehrs



Ausnutzung von Schwachstellen



Gegenmaßnahmen

- Neueste Version des Tor Browsers in der Standardkonfiguration
- Nur sehr wenige Plugins einsetzen - z.B. nur uBlock Origin
- Evtl. Entry und Exit Nodes nach Land festlegen => torrc Konfigurationsdatei
- Evtl. Entry und Exit Nodes fest konfigurieren (CCC, Digitalcourage, Swiss Privacy Foundation, ...) => torrc Konfigurationsdatei


Fragen?
Präsentation online unter: https://scheible.it

Quellen

1) CWA_title.png; Sebastian Wolf; https://github.com/corona-warn-app/cwa-website/blob/master/images/CWA_title.png, abgerufen am 02.11.2020

2) Researchers reverse Netflix anonymization; Robert Lemos, SecurityFocus; https://www.securityfocus.com/news/11497, abgerufen am 03.11.2020

3) FOILing NYC's Taxi Trip Data; Chris Whong; https://chriswhong.com/open-data/foil_nyc_taxi/, abgerufen am 03.11.2020

4) Estimating the success of re-identifications in incomplete datasets using generative models; Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye; https://www.nature.com/articles/s41467-019-10933-3, abgerufen am 03.11.2020