Hochschule
Albstadt-Sigmaringen
University of Applied Sciences

Fakultät Informatik

Deanonymisierung – Grenzen und Möglichkeiten

Internet Security Days 2017

Tobias Scheible, M.Eng.

Tobias Scheible, M.Eng.

- Studium Kommunikations- und Softwaretechnik, Fachrichtung Kommunikationstechnik, Hochschule Albstadt-Sigmaringen
- 2009 bis 2012: Softwareingenieur im Bereich Web Development
- Seit 2012: Wissenschaftlicher Mitarbeiter an der Hochschule Albstadt-Sigmaringen Im Bereich IT-Sicherheit & Forensik

Praktikum IT Security 2
IT Security (Bachelor) – 2. Semester
Prof. Holger Morgenstern

Seminar IT Security 2
IT Security (Bachelor) – 2. Semester
Prof. Holger Morgenstern

Digitale Forensik
IT Security (Bachelor) – 5. Semester
Prof. Holger Morgenstern

Projektstudium
IT Security (Bachelor) – 5. Semester
Prof. Holger Morgenstern

Einführung in die Informatik
Digitale Forensik (Master) – 1. Semester
Prof. Dr. Martin Rieger

Internet Grundlagen
Digitale Forensik (Master) – 1. Semester
Prof. Dr. Martin Rieger

Betriebssystemforensik
Digitale Forensik (Master) – 3. Semester
Prof. Dr. Martin Rieger

Vorträge & Workshops
zu aktuellen Themen der IT-Sicherheit, u. a. für den VDI und die IHK

Blog scheible.it
Blog rund um meine Aktivitäten
https://scheible.it

Agenda

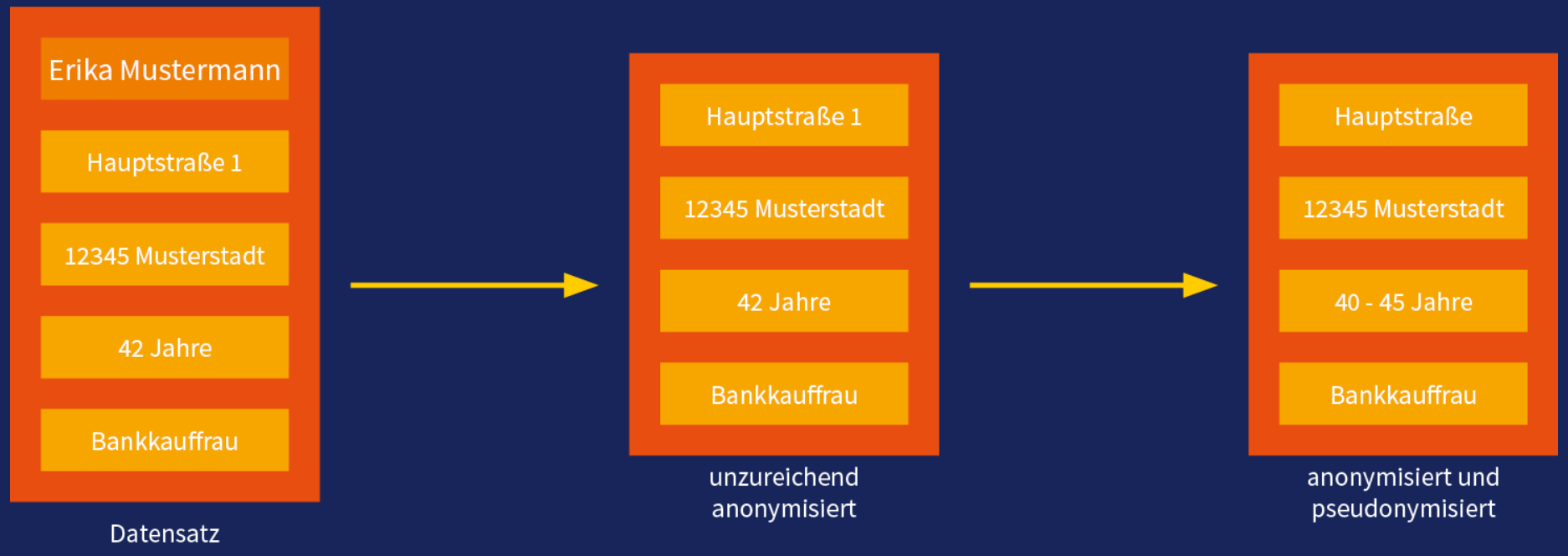
- Deanonymisierung
- Beispiele aus der Forschung
- Einsatzszenarien im Darknet
- Projekt Website Profiling
- Zusammenfassung

Der Vortrag „Deanonymisierung – Grenzen und Möglichkeiten“ gibt einen Überblick über das Themenfeld Deanonymisierung und beschreibt Gründe, warum diese eingesetzt werden. Anhand von Beispielen wird aufgezeigt, wie vermeintlich anonymisierte Daten durch die Kombination aus verschiedenen Quellen und mit Analysen wieder aggregiert werden können.

Deanonymisierung

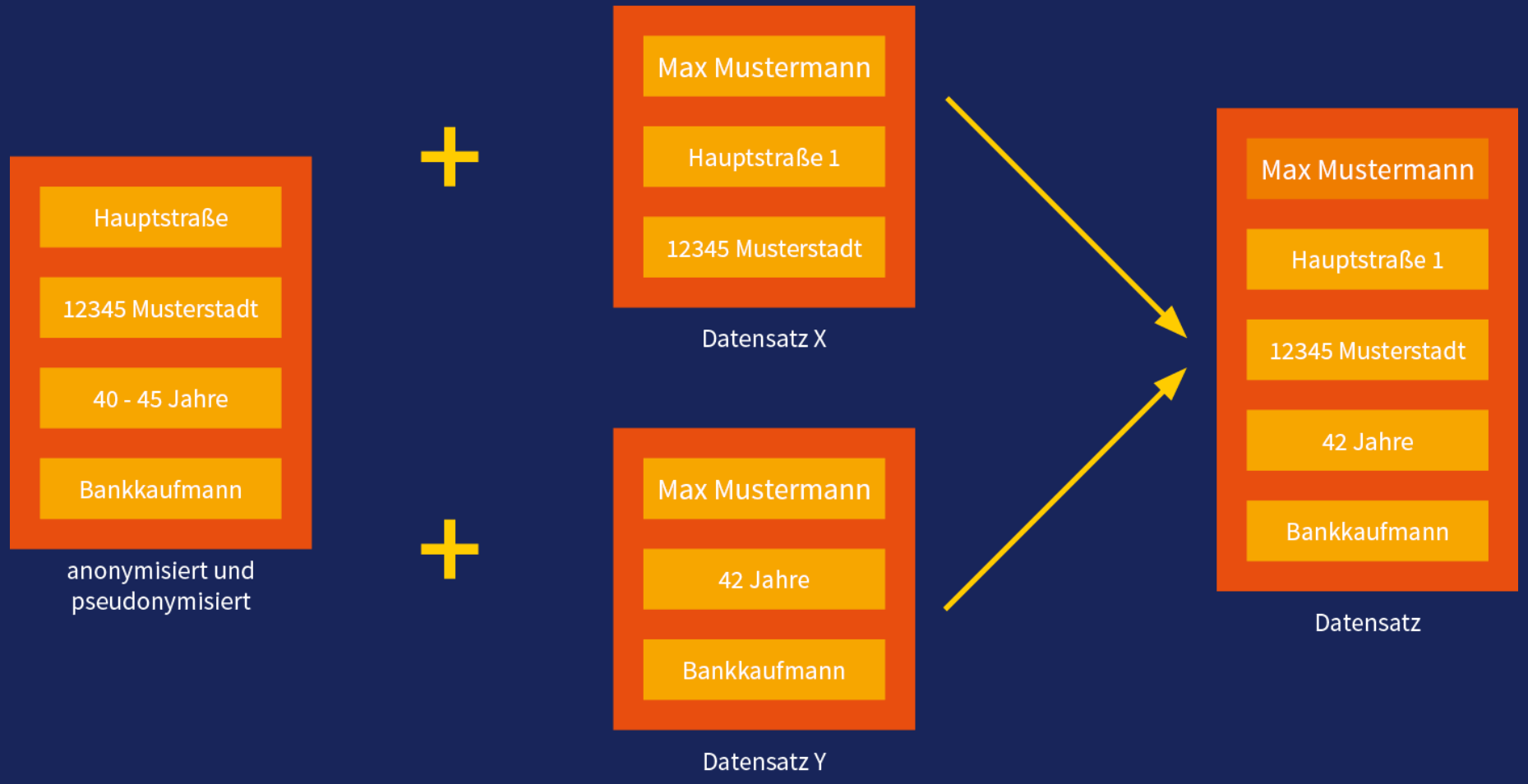
Pseudonymisierung

Anonymisierung



Deanonymisierung – Grenzen und Möglichkeiten

Deanonymisierung



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

[1]

- 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

Robust De-anonymization of Large Sparse Datasets, Arvind Narayanan and Vitaly Shmatikov, 2008, In Proc. of the 29th IEEE Symposium on Security and Privacy

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

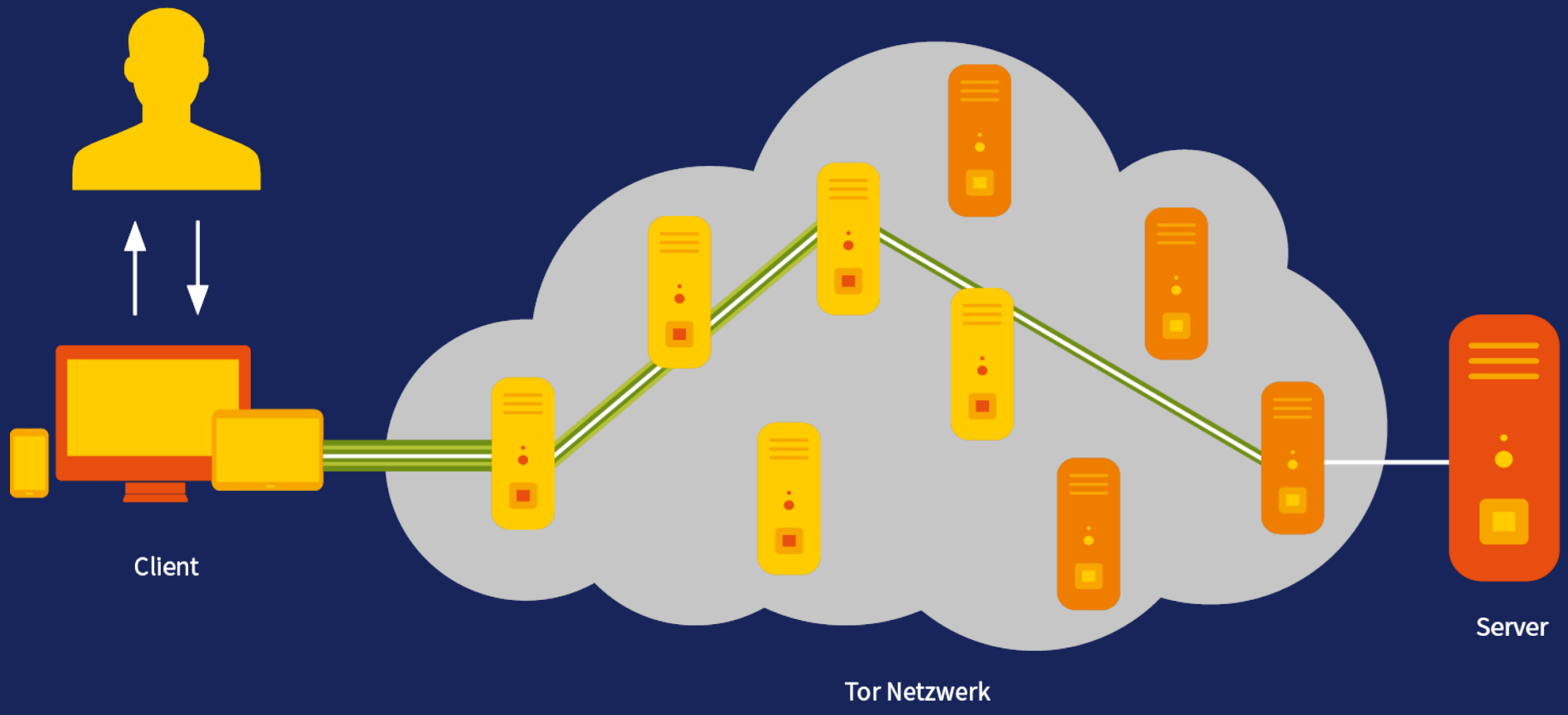
[1]

- MD5-Hash: „Internet Security Days“ => f0bc2757776c80d2cf97a110207b2f1c
  - 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 des Jahres 2013

[1] FOILing NYC’s Taxi Trip Data, Chris Whong, 18.03.2014, http://chriswhong.com/open-data/foil_nyc_taxi/

Einsatzszenarien im Darknet

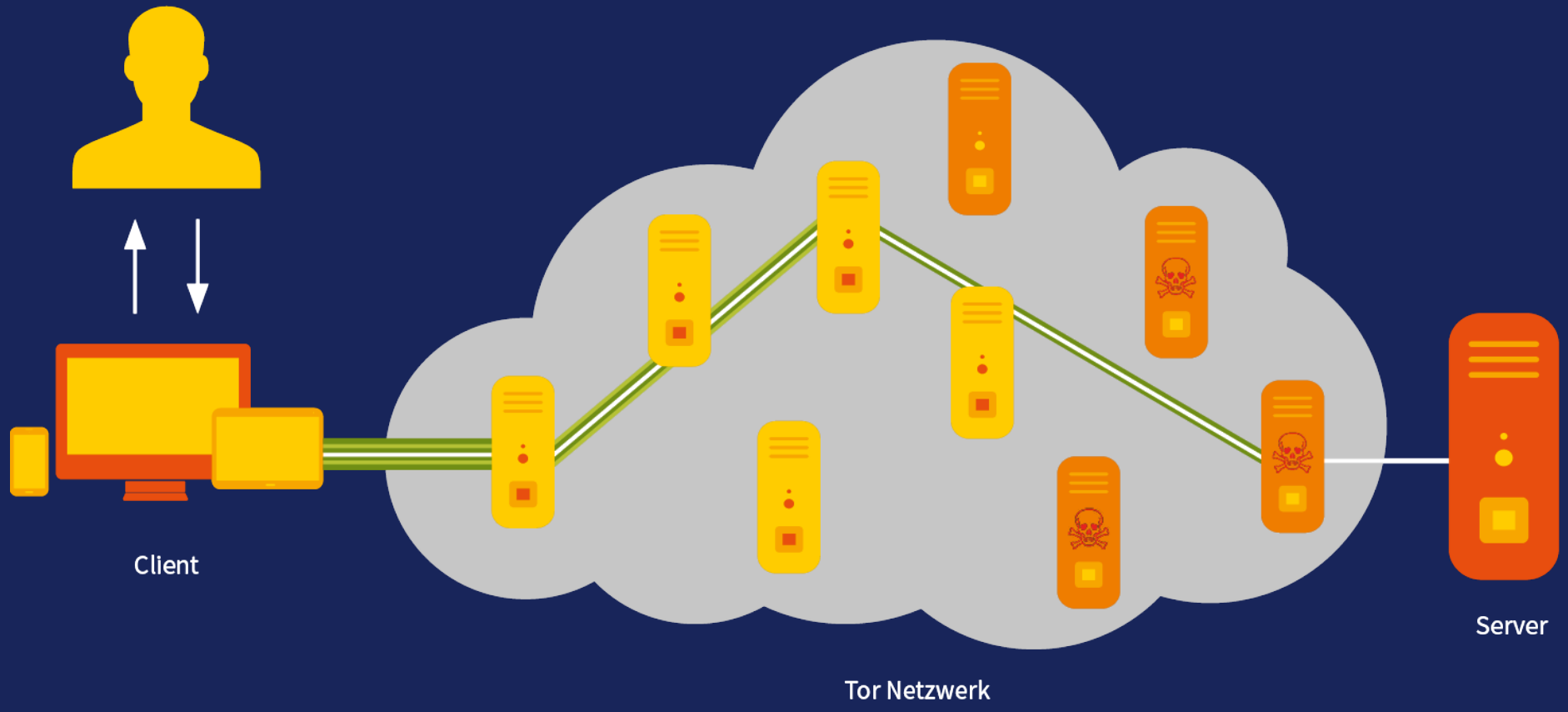
Tor Netzwerk



**Deanonymisierung –
Grenzen und Möglichkeiten**

+ Deanonymisierung
+ Beispiele aus der Forschung
+ Einsatzszenarien im Darknet
  Tor Netzwerk
  Kontrolle von Exit-Nodes
  Analyse des Netzwerkverkehrs
  Ausnutzung von Schwachstellen
+ Projekt Website Profiling
+ Zusammenfassung

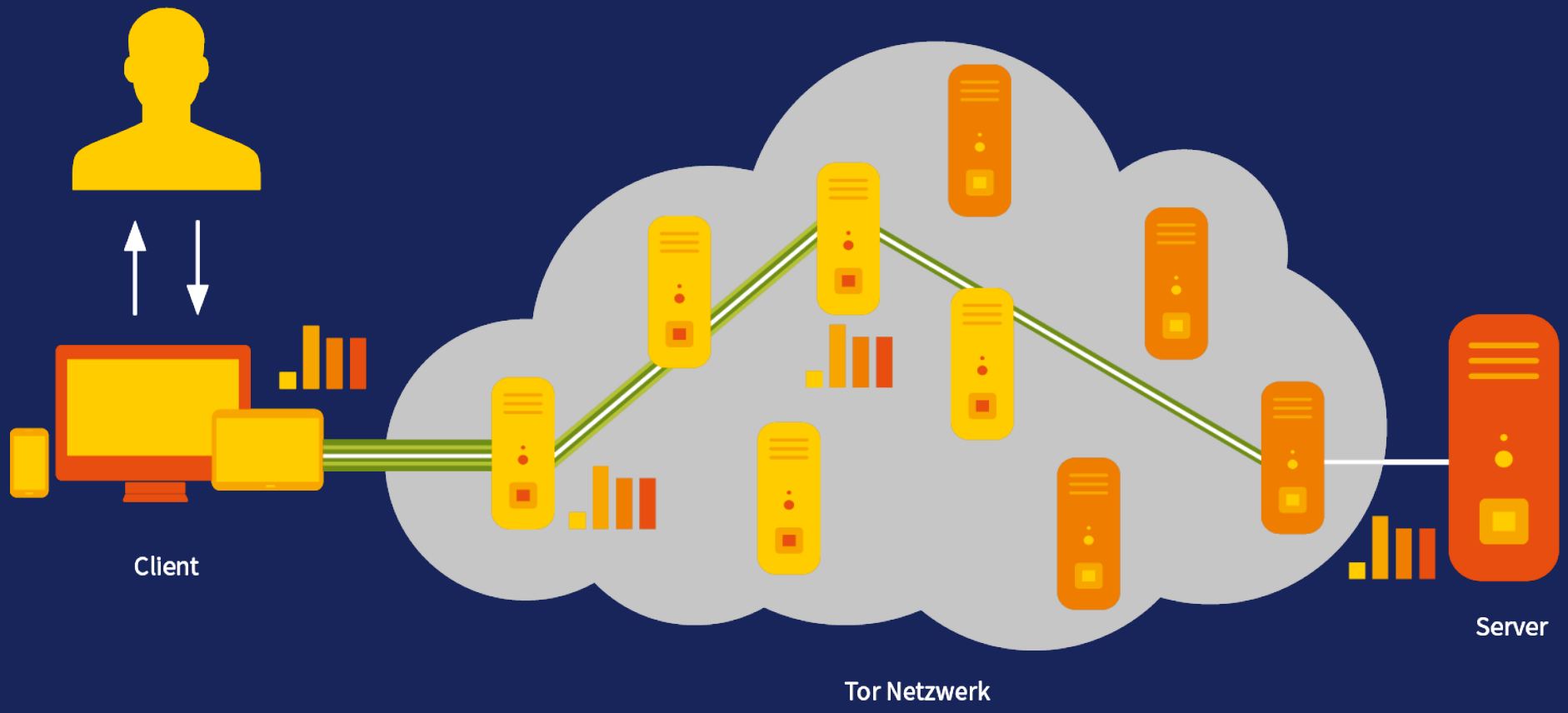
Kontrolle von Exit-Nodes



**Deanonymisierung –
Grenzen und Möglichkeiten**

+ Deanonymisierung
+ Beispiele aus der Forschung
+ Einsatzszenarien im Darknet
  Tor Netzwerk
  Kontrolle von Exit-Nodes
  Analyse des Netzwerkverkehrs
  Ausnutzung von Schwachstellen
+ Projekt Website Profiling
+ Zusammenfassung

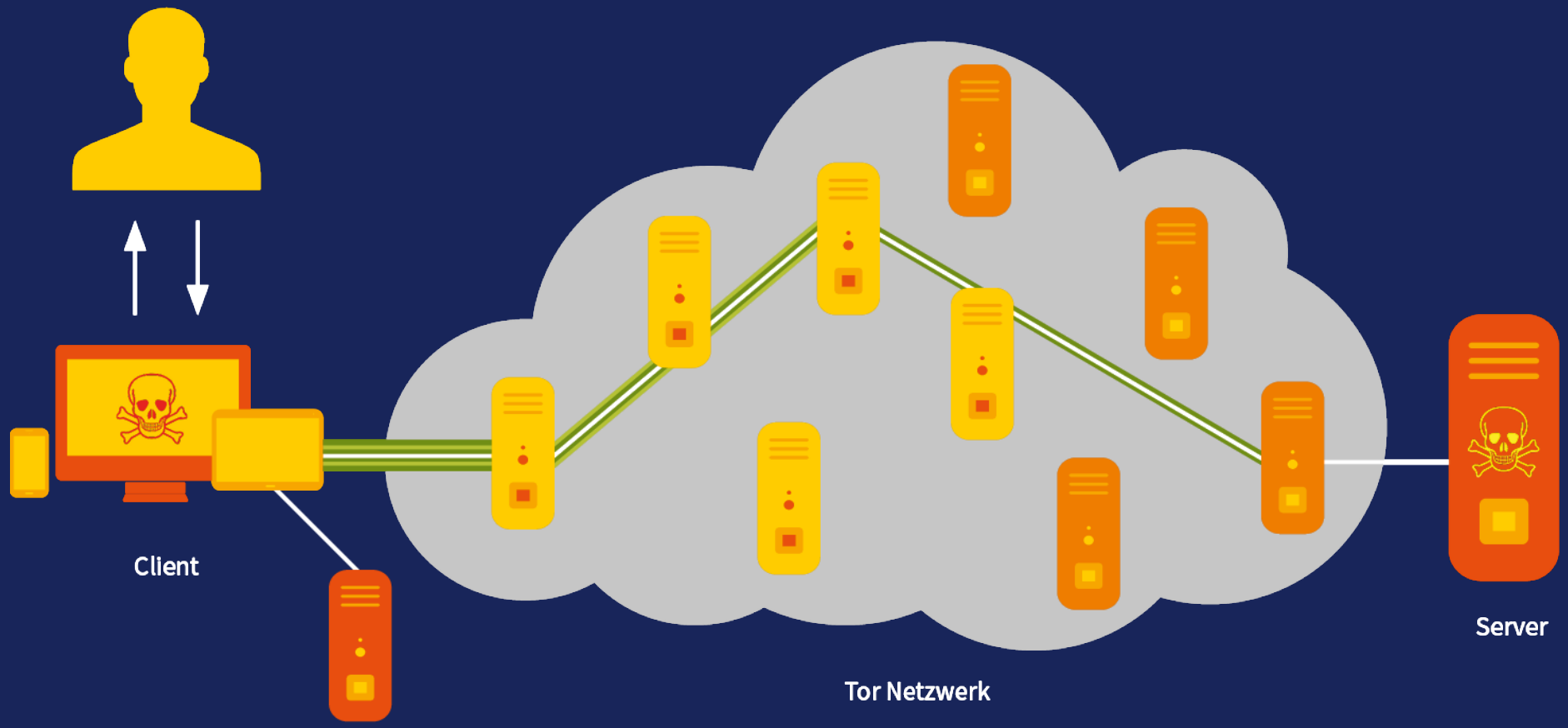
Analyse des Netzwerkverkehrs



**Deanonymisierung –
Grenzen und Möglichkeiten**

+ Deanonymisierung
+ Beispiele aus der Forschung
+ Einsatzszenarien im Darknet
  Tor Netzwerk
  Kontrolle von Exit-Nodes
  Analyse des Netzwerkverkehrs
  Ausnutzung von Schwachstellen
+ Projekt Website Profiling
+ Zusammenfassung

Ausnutzung von Schwachstellen



**Deanonymisierung –
Grenzen und Möglichkeiten**

+ Deanonymisierung
+ Beispiele aus der Forschung
+ Einsatzszenarien im Darknet
  Tor Netzwerk
  Kontrolle von Exit-Nodes
  Analyse des Netzwerkverkehrs
  Ausnutzung von Schwachstellen
+ Projekt Website Profiling
+ Zusammenfassung


Projekt Website Profiling

Analyse der Websites

unbekannte Identität des Betreibers

bekannte Identität des Betreibers

**Deanonymisierung –
Grenzen und Möglichkeiten**

Extraktion der Merkmale

Technische Merkmale

- Verwendetes System (Typo3, WordPress, Magento, …) und Programmiersprachen
- JavaScript-Bibliotheken, CSS-Frameworks, gesetzte HTTP-Header, …
- „Schreibstil“ von Programmier- bzw. Scriptsprachen
- …

Inhaltliche Merkmale

- Dateinamen, URL-Schema, Anzahl verwendeter Zeichen
- Abbildungen und ihre Meta-Daten inkl. Kompressionsalgorithmen
- Fehlerseiten, Standardgrafiken und Quelle der Komponenten
- …

Erstellung des Profils

- Die relevanten Merkmale müssen gespeichert und bewertet werden
- Eine Metrik wandelt die Merkmale in mathematische Werte um
- Alle Merkmale zusammen ergeben ein individuelles Profil

- Entweder wird eine Website mit einer anderen verglichen oder eine ganze Datenbank wird für den automatischen Abgleich verwendet

- => Generierung einer eindeutigen Kennung einer Website

Grenzen des Website Profiling

- Die Methode kann nur auf Websites mit individuellen Merkmalen angewendet werden
- Alle Techniken müssen zu verarbeiten sein und korrekt interpretiert werden
- Gegenmaßnahmen müssen erkannt und entsprechend behandelt werden

- => Es kann nur eine Website einer anderen zugeordnet werden
  - Eine Identifizierung einer Person ist nicht möglich

Zusammenfassung

Zusammenfassung und Ausblick

- Deanonymisierung
  - Immer wieder kommt es zu Fällen von Deanonymisierung
  - Datensätze dürfen beim Thema Anonymität nicht isoliert betrachtet werden
  - Mögliche rechtliche Konsequenzen bei Deanonymisierung
  - Ermittlungsbehörden setzen z.T. Methoden zur Deanonymisierung ein
  - Echte Anonymität ist nur sehr schwer möglich

- Website Profiling – nächste Schritte
  - Entwicklung der Metrik und Validierung der Methode
  - Abschlussarbeiten und Projekte von Studierenden

Vielen Dank für Ihre Aufmerksamkeit

Präsentation unter: https://scheible.it