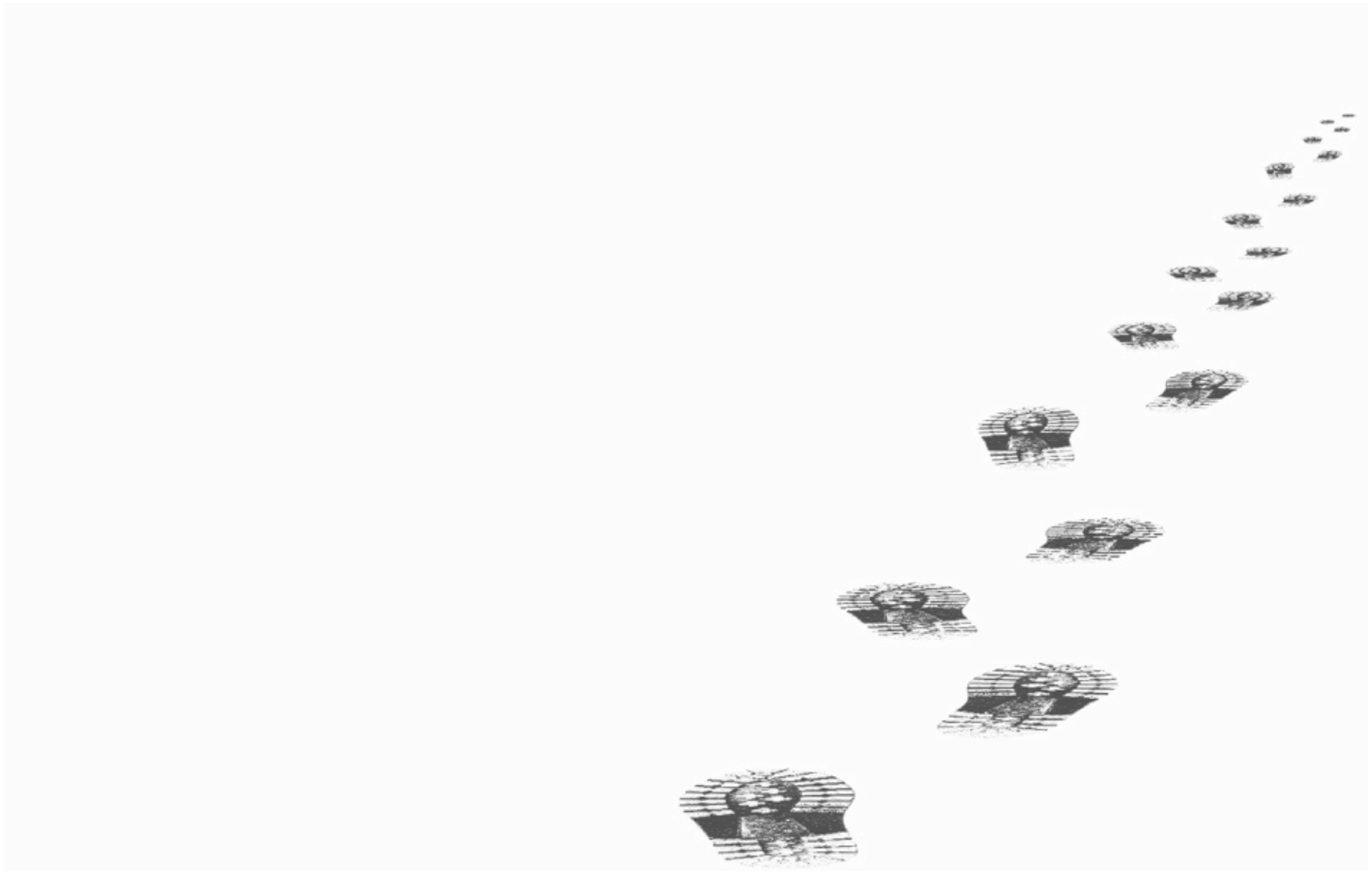
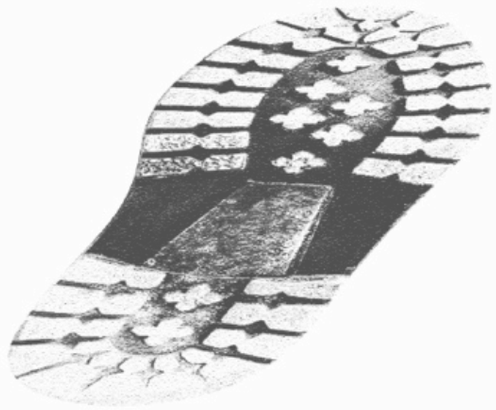
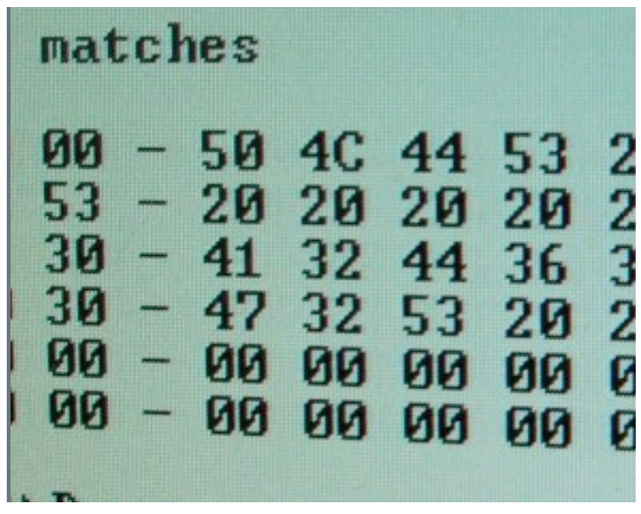


# ANGRIFFE AUF DIE PRIVATSPHÄRE
## DURCH TRAFFIC-ANALYSEN IM INTERNET

**Dominik Herrmann und  Christoph Gerber**

Universität Regensburg
LSt Management der Informationssicherheit
Prof. Dr. Hannes Federrath







**MOTIVATION UND SZENARIO**

**WEBSITE FINGERPRINTING**

**USER LINKABILITY**

**KONSEQUENZEN FÜR DIE PRAXIS**

**Ziel: Sensibilisierung**

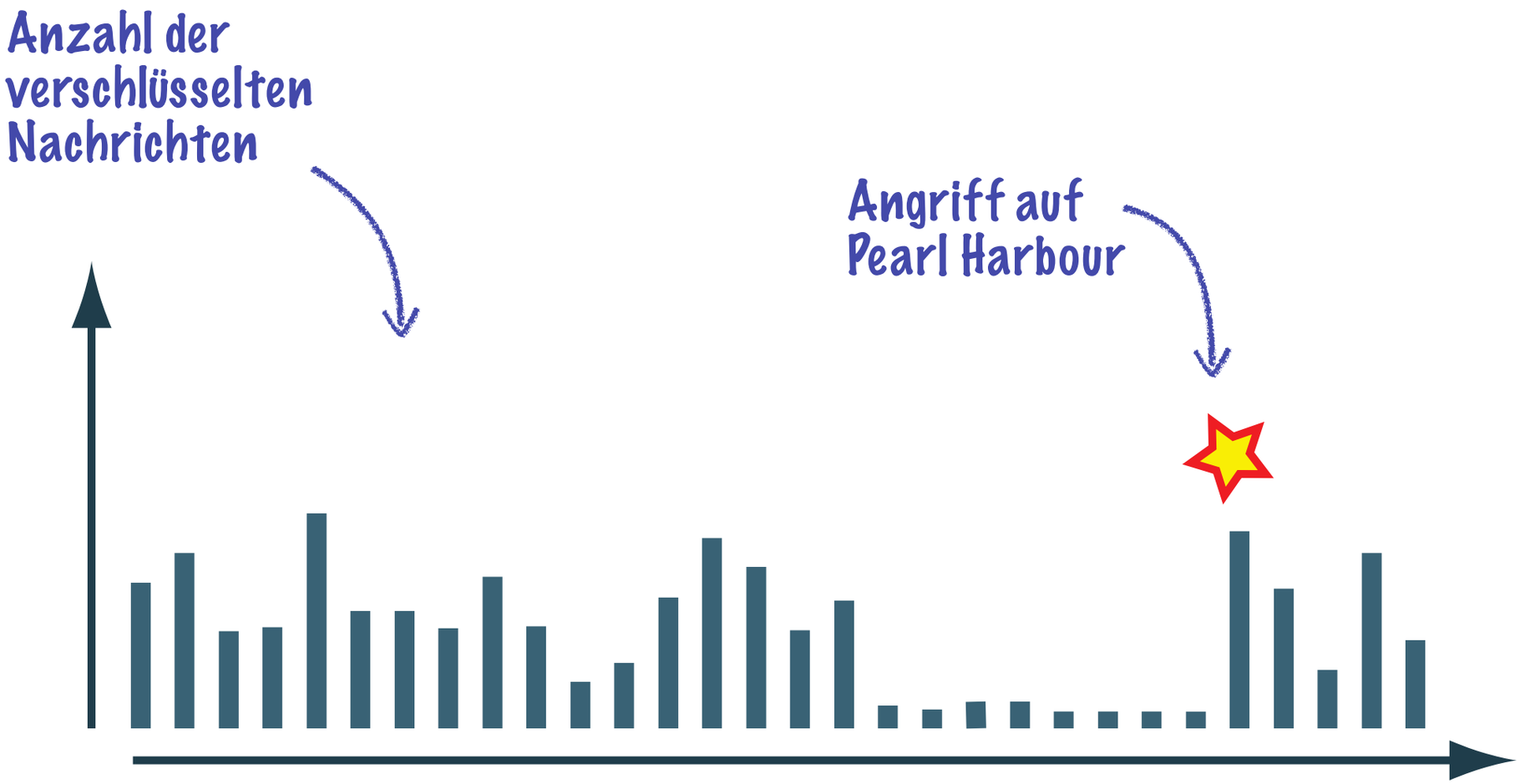
# MOTIVATION UND SZENARIO

# TRAFFIC-ANALYSE

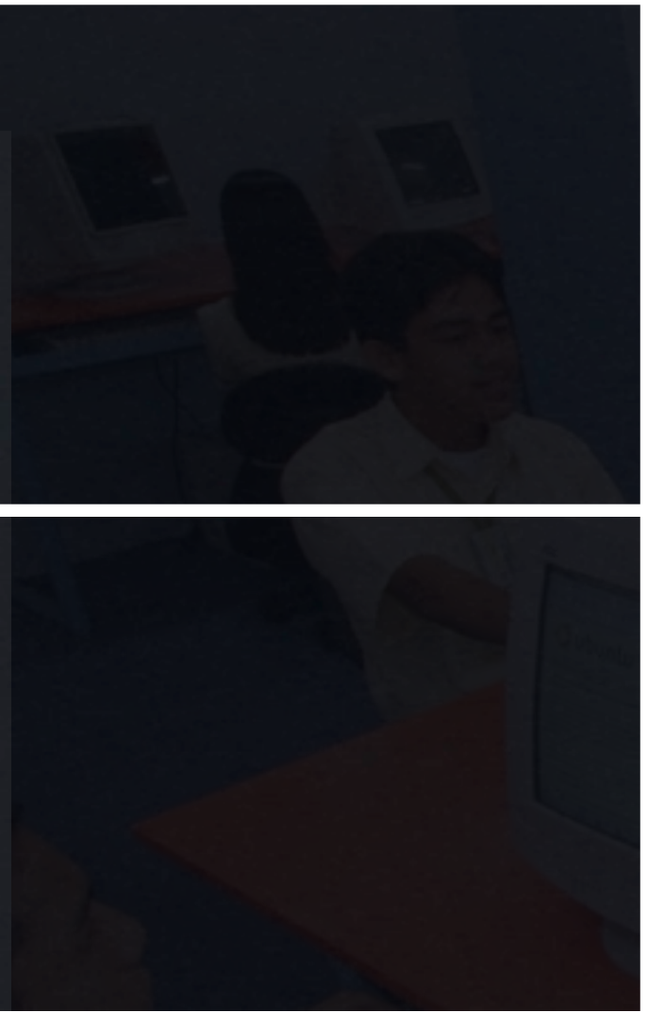
Beobachtung übermittelter Nachrichten

Rückschluss auf Inhalt

Ursprung: Militär


Anzahl der verschlüsselten Nachrichten
Angriff auf Pearl Harbour

# TRAFFIC-ANALYSE IM INTERNET



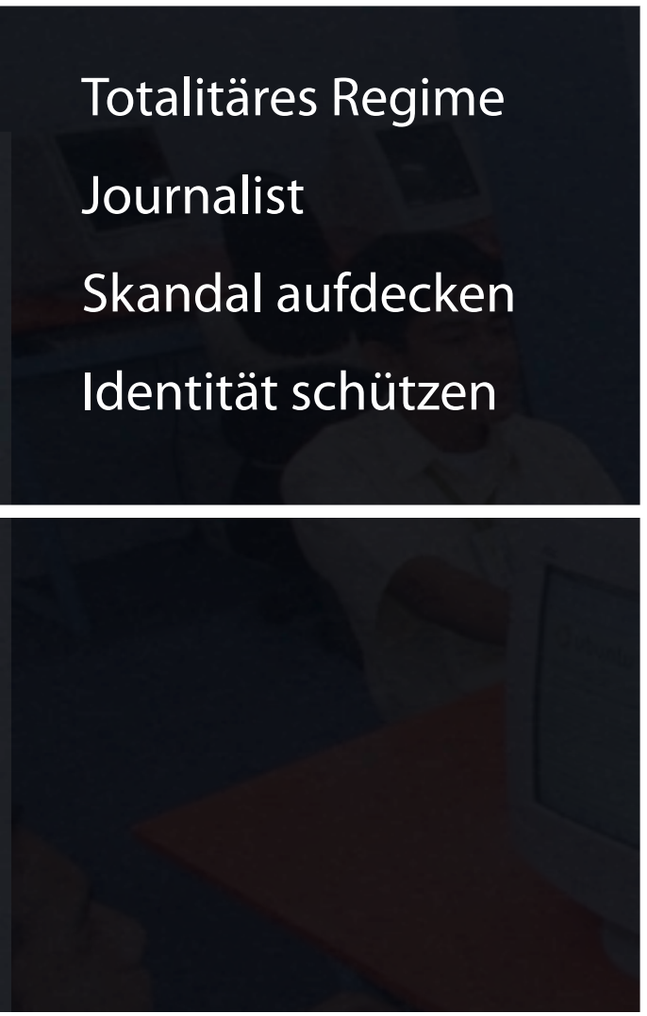
# SZENARIO

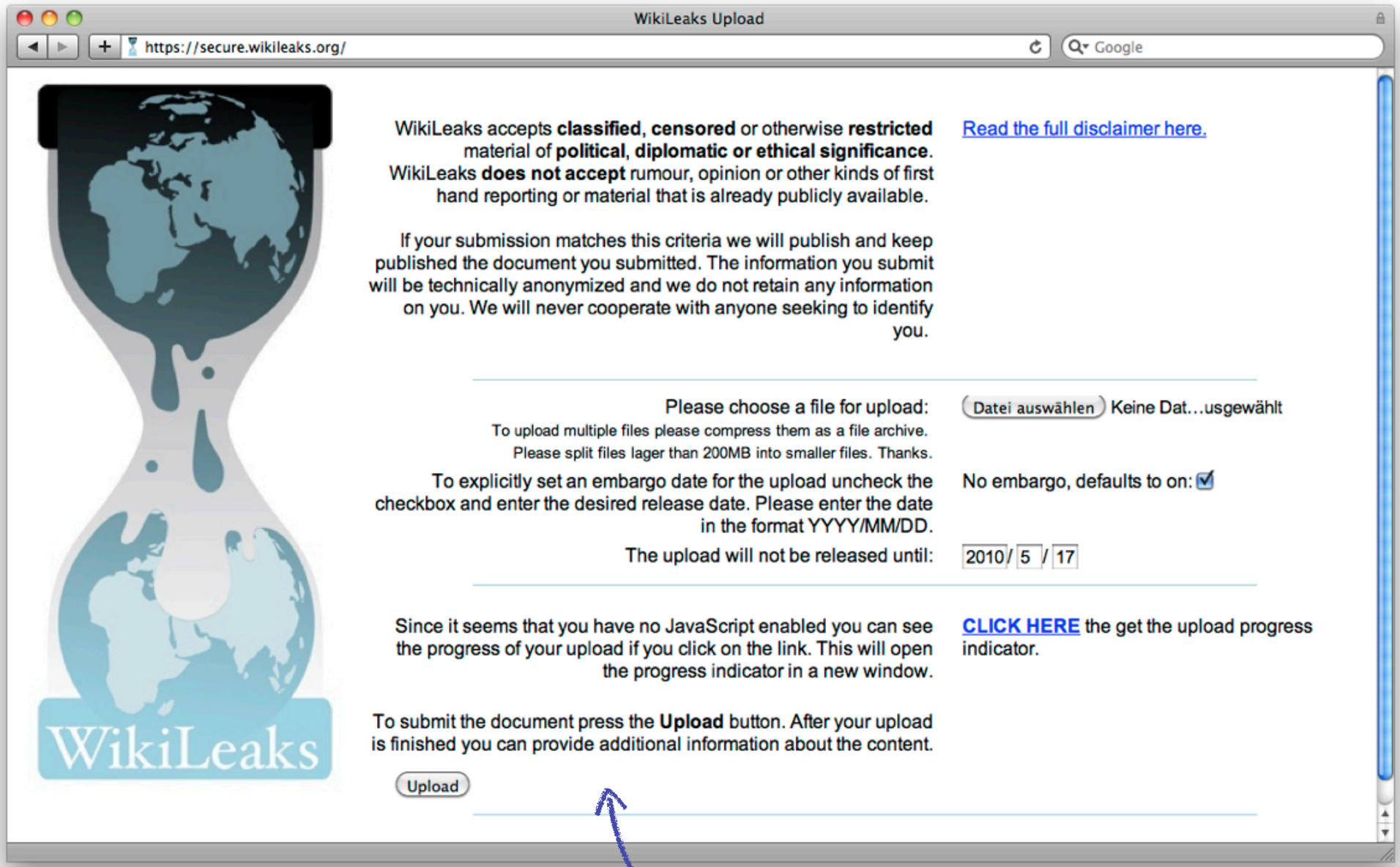
Totalitäres Regime

Journalist

Skandal aufdecken

Identität schützen

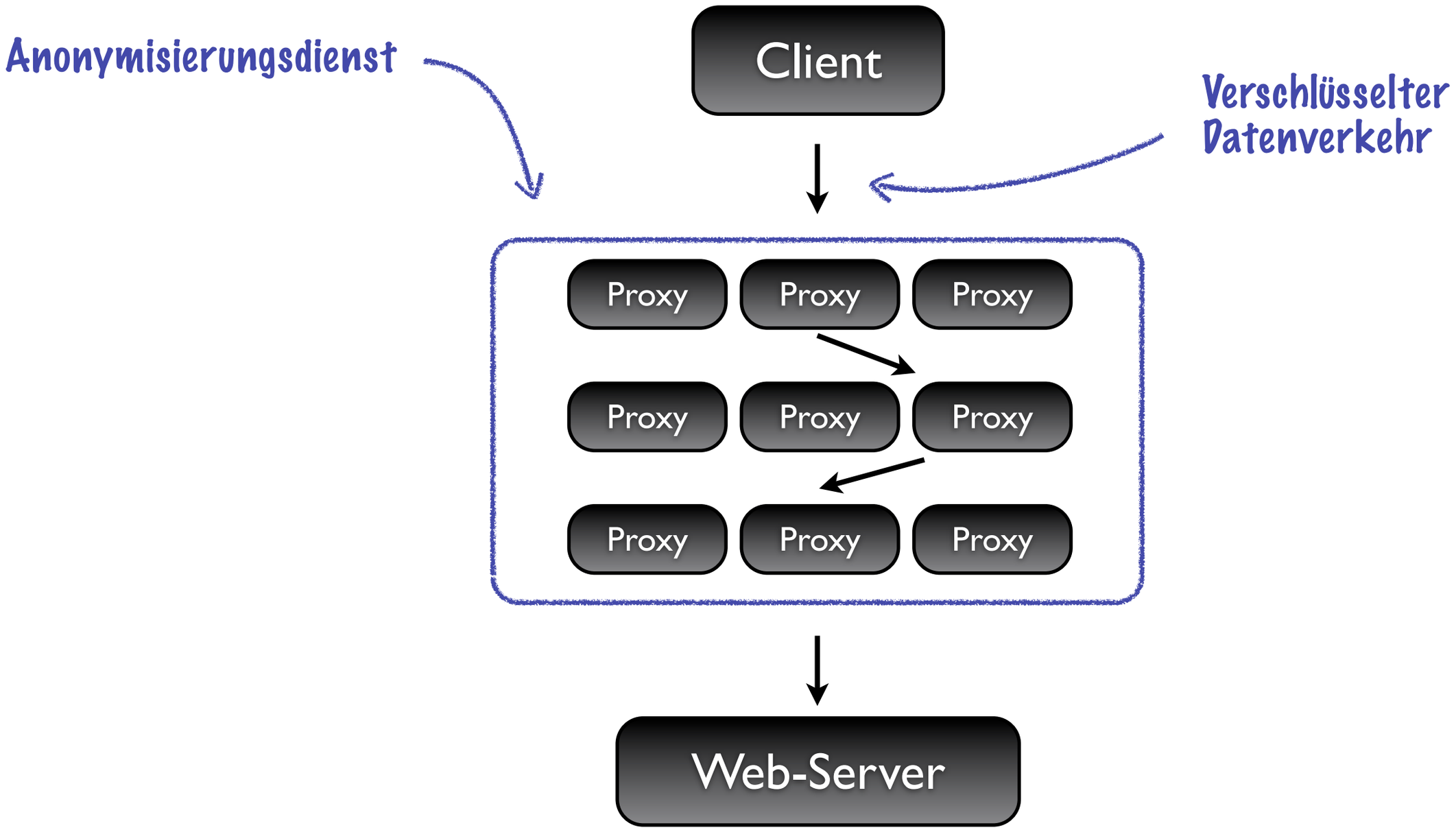




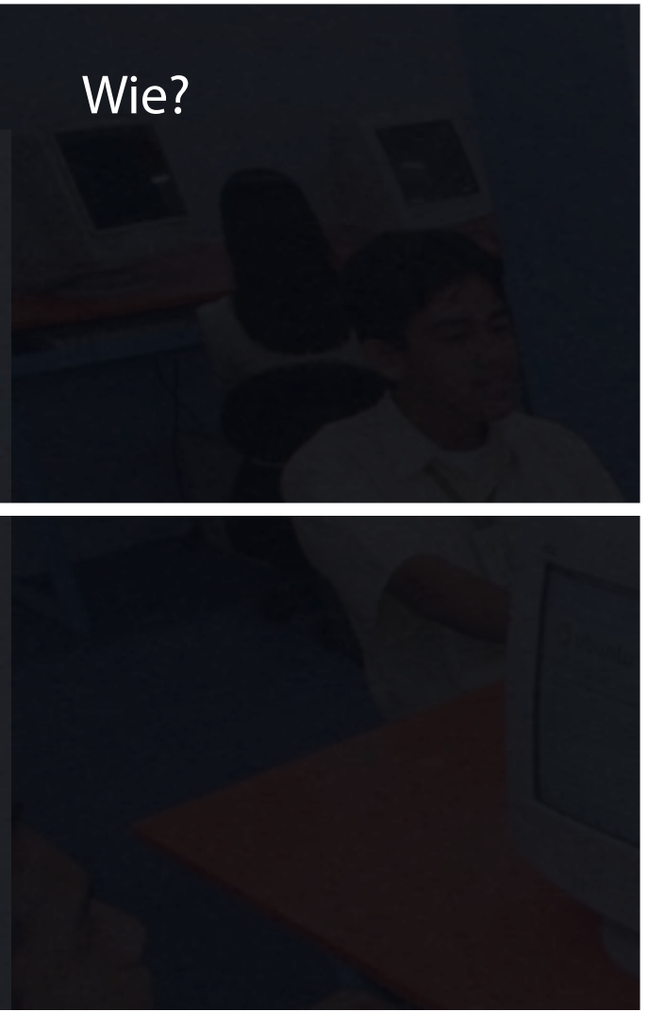

Anonymer Upload bei Wikileaks

# VORSICHTSMASSNAHMEN

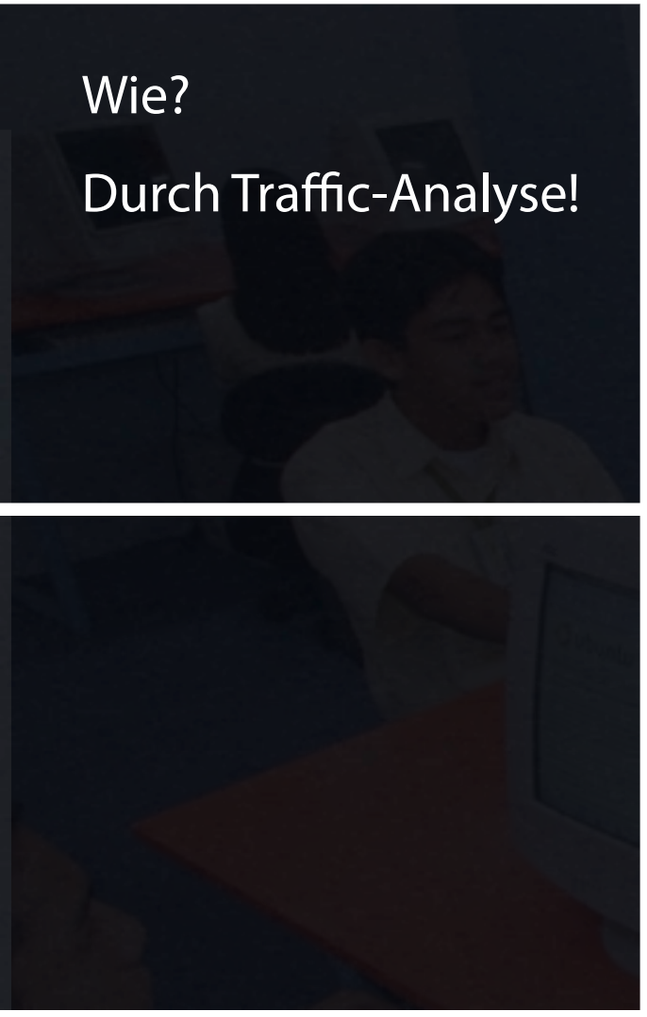
- Wikileaks
- Internet-Café
- Dummy-Traffic
- Anonymisierungsdienst


Anonymisierungsdienst
Client
Verschlüsselter Datenverkehr
Proxy
Proxy
Proxy
Proxy
Proxy
Proxy
Proxy
Proxy
Proxy
Web-Server

# JOURNALIST WIRD TROTZDEM ENTTARNT




# JOURNALIST WIRD TROTZDEM ENTTARNT

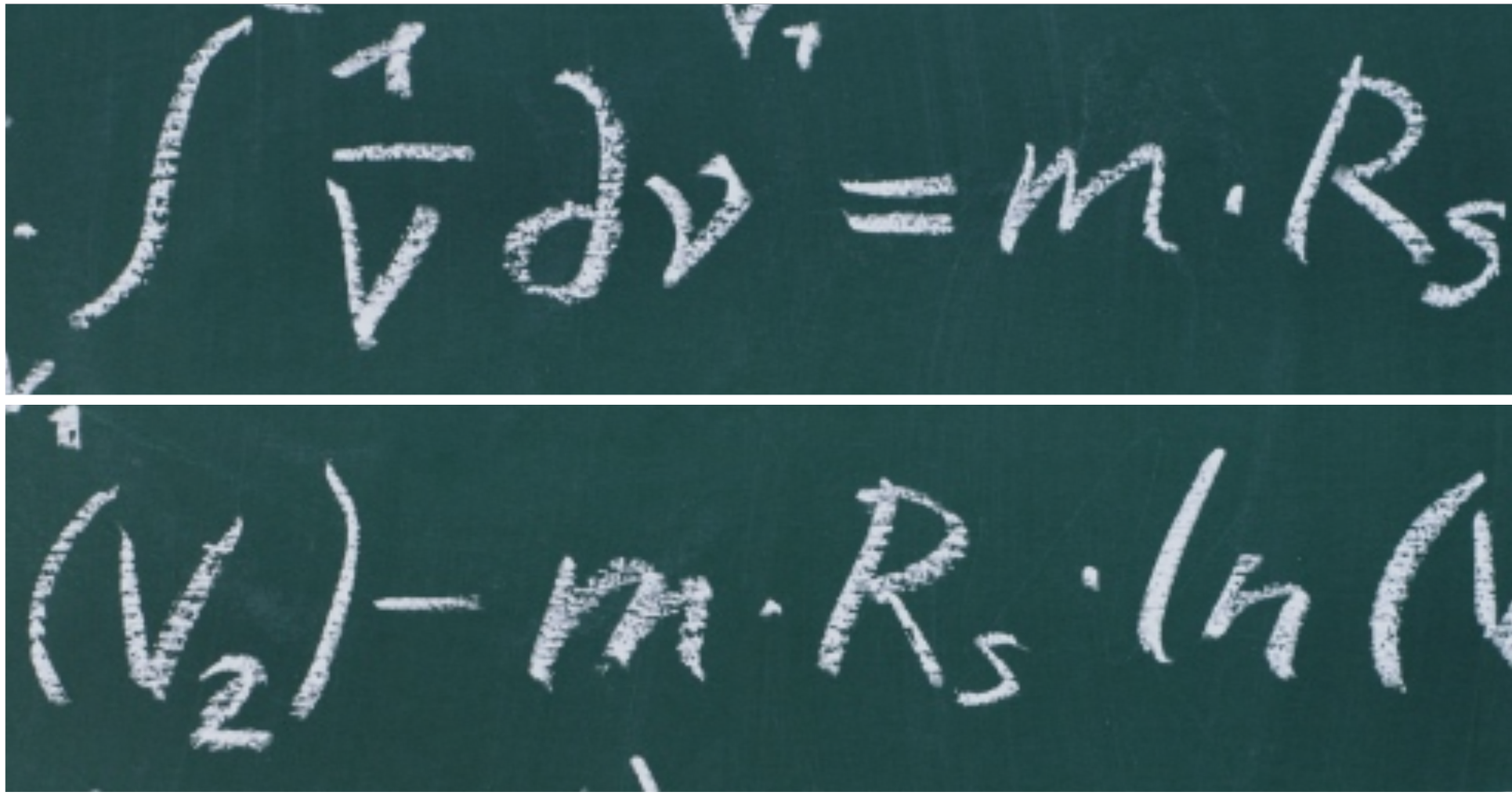


Wie?

Durch Traffic-Analyse!

# METHODEN

- Statistische Analyse
- Mustererkennung
- Data-Mining
- Machine-Learning

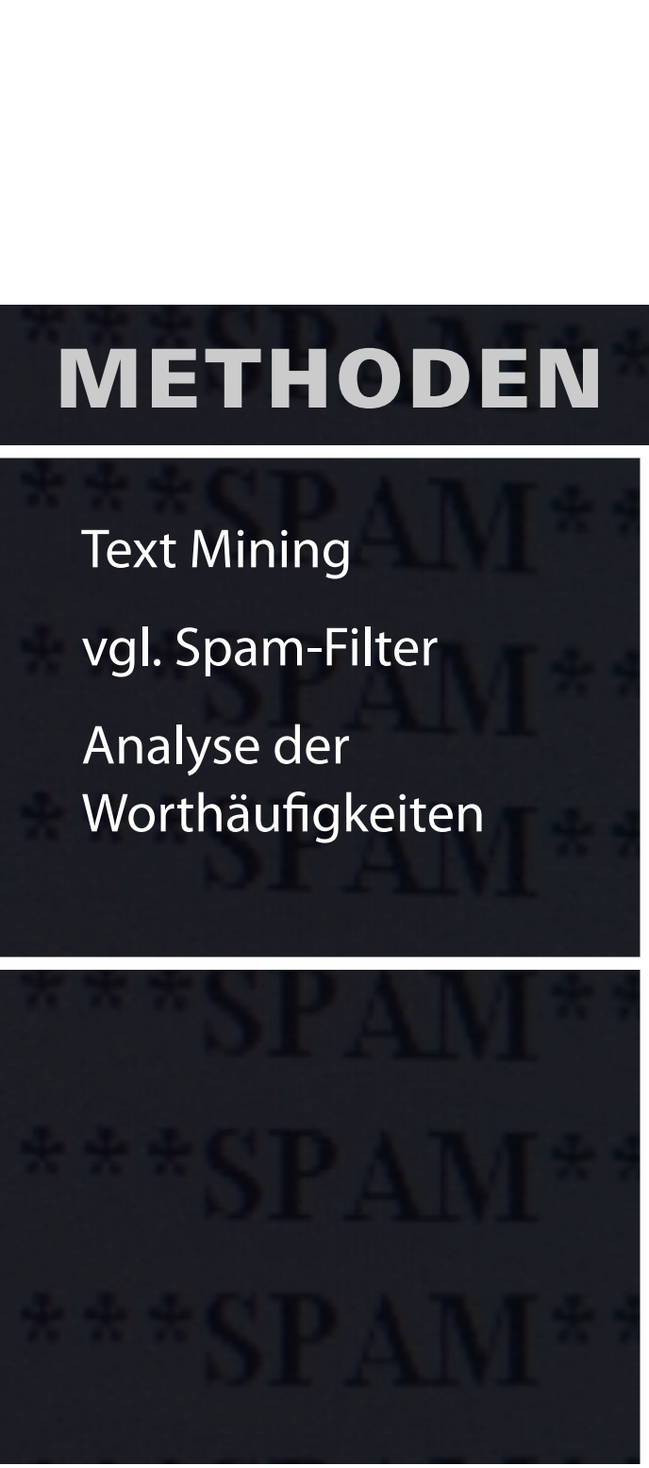


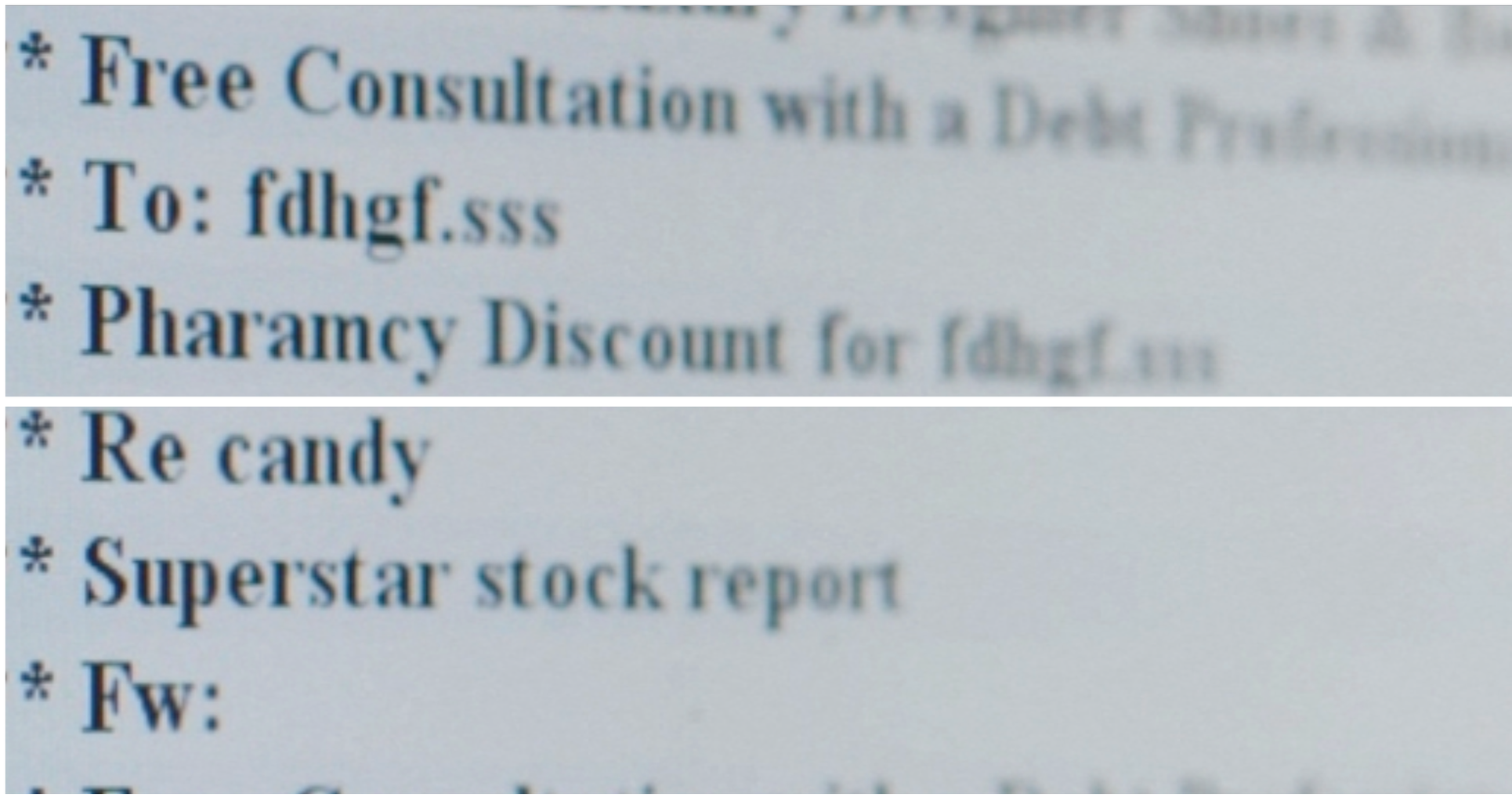
# METHODEN

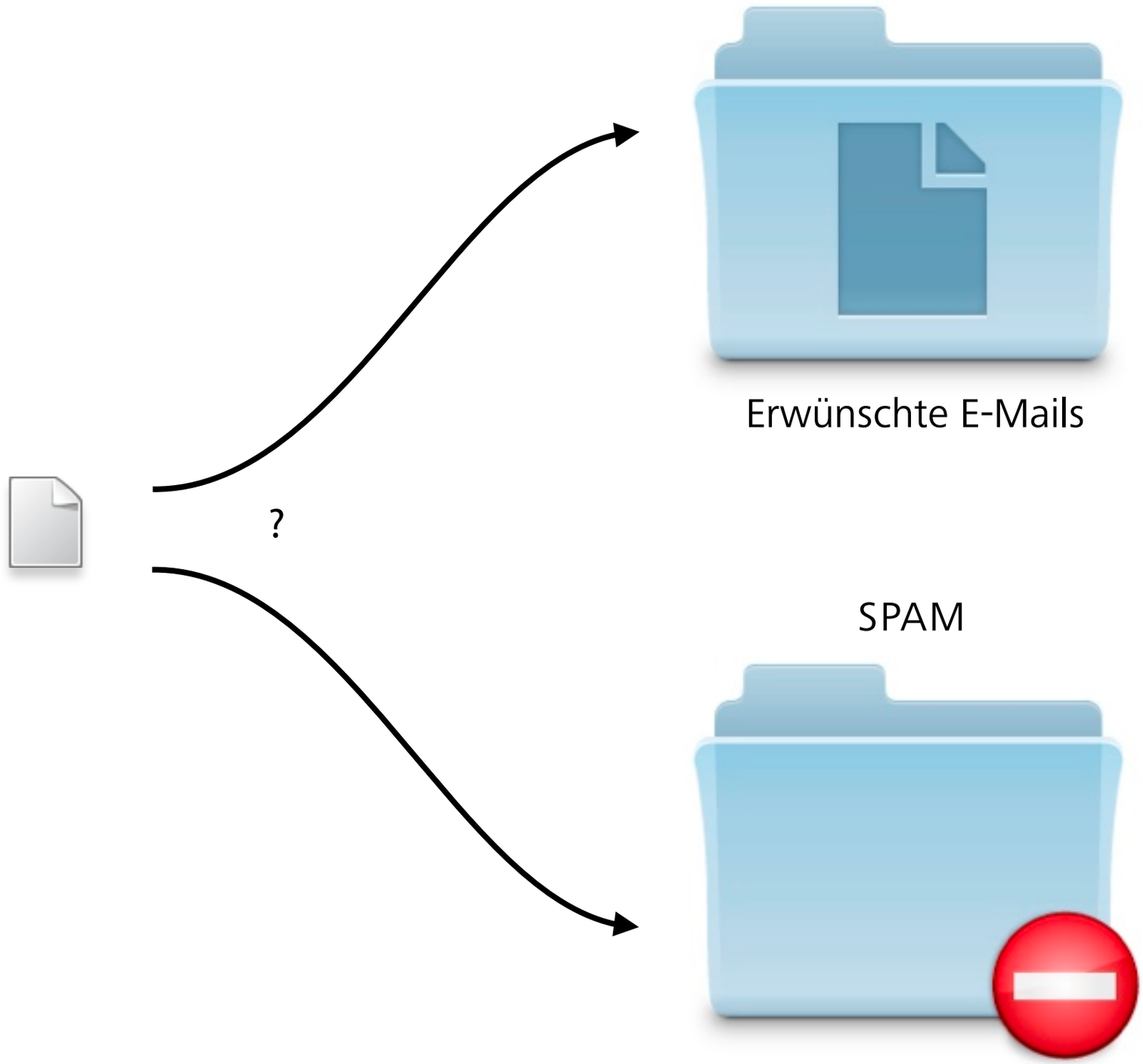
Text Mining

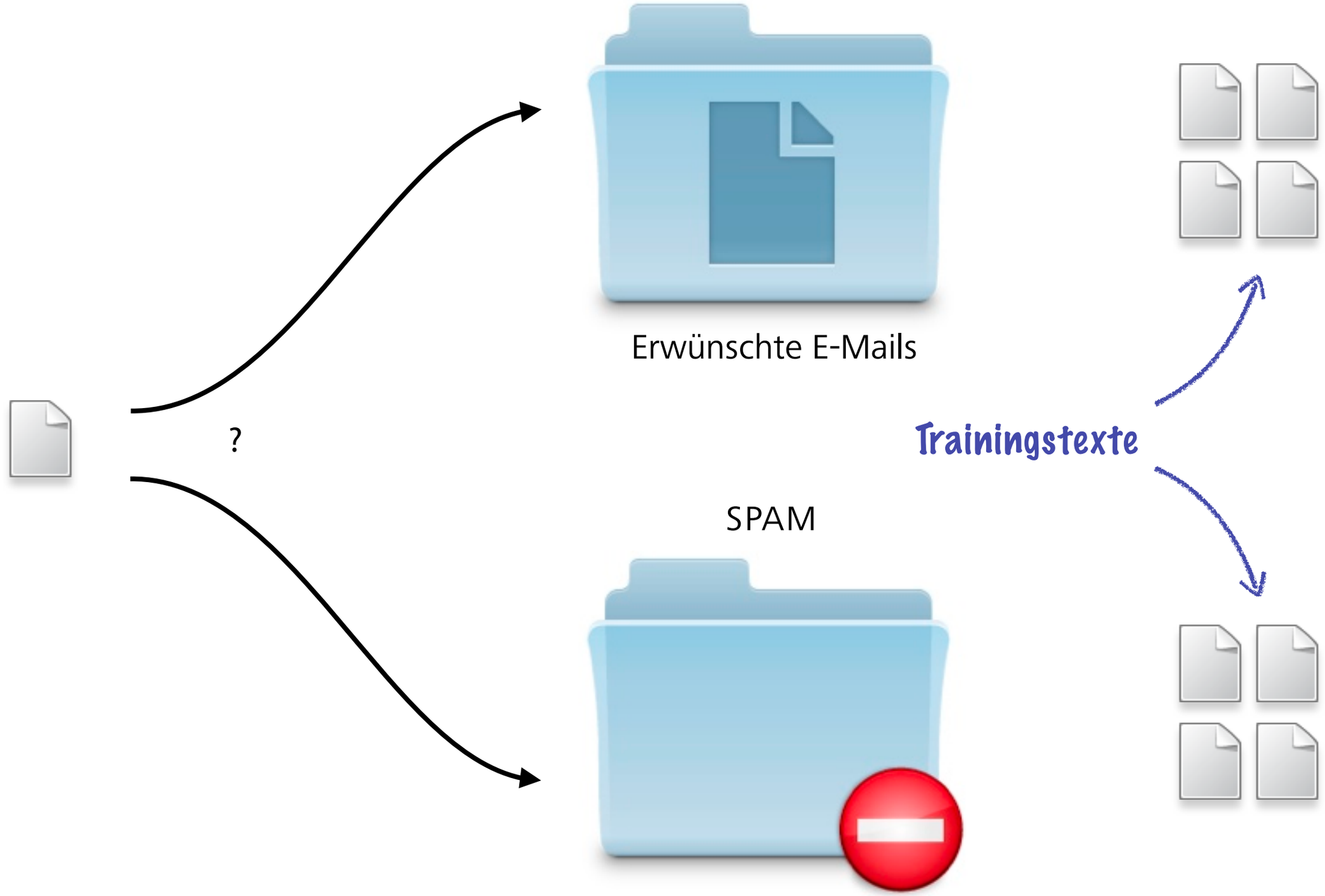
vgl. Spam-Filter

Analyse der Worthäufigkeiten




Erwünschte E-Mails
?
SPAM

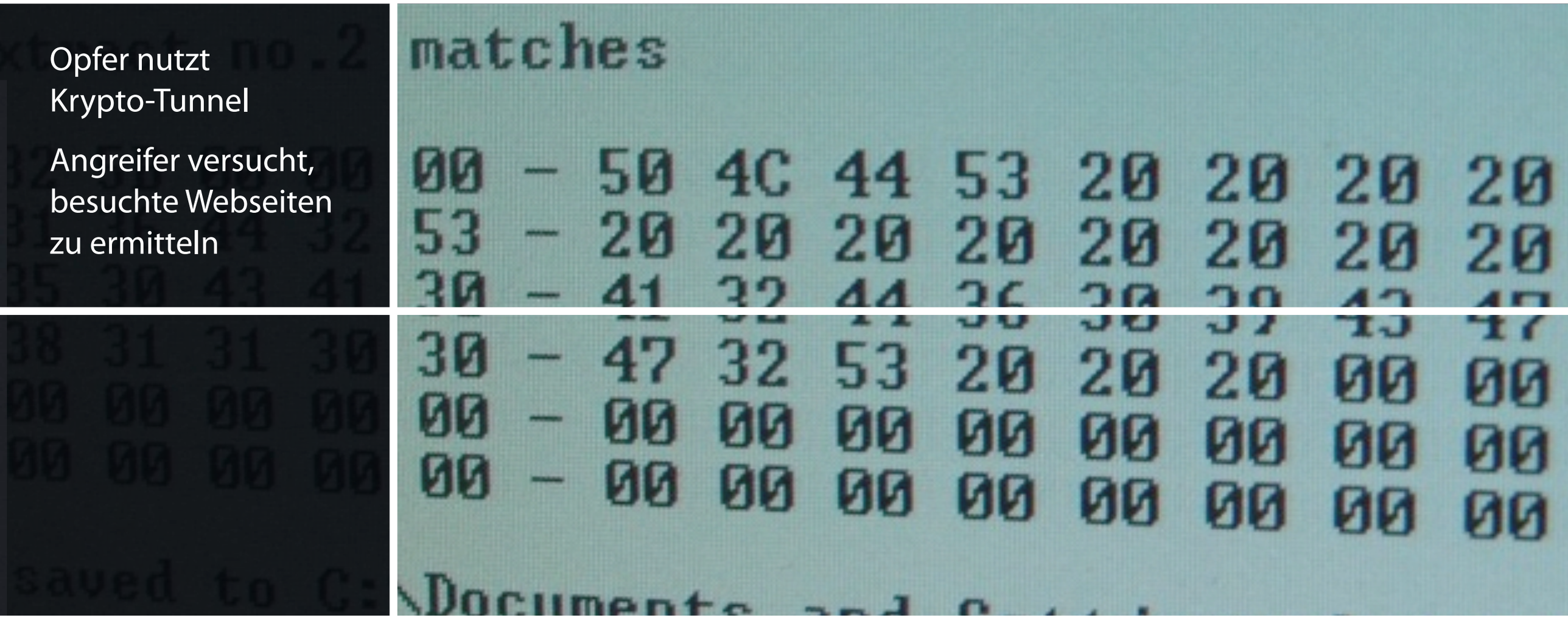

Erwünschte E-Mails
?
Trainingstexte
SPAM


WEBSITE
FINGERPRINTING

# HERAUSFORDERUNG

Opfer nutzt Krypto-Tunnel

Angreifer versucht, besuchte Webseiten zu ermitteln



z.B. ISP, Administrator, Strafverfolger, ...

**Angreifer**

Client-Rechner

Tunnel-Endpunkt

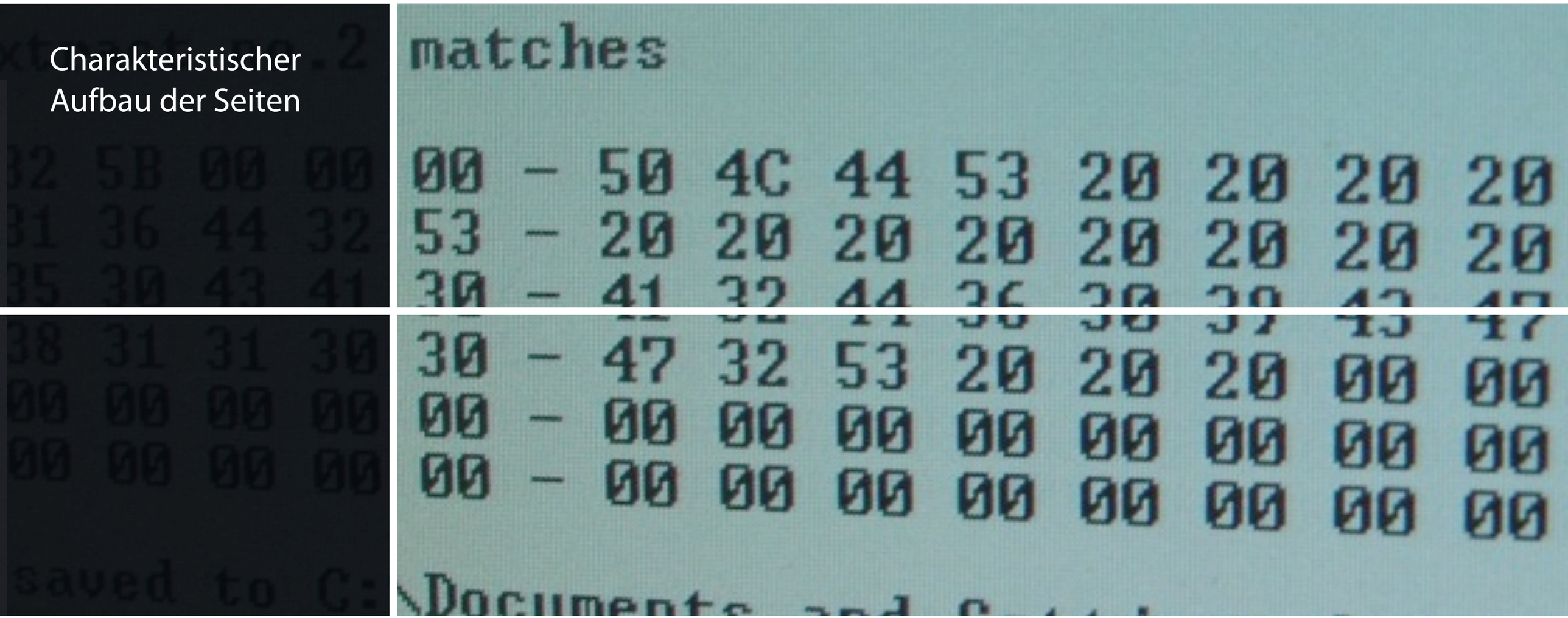
Webserver


Tunnel

z.B. verschlüsseltes WLAN, VPN, OpenSSH, Anonymisierungsdienst, ...

# GRUNDLAGE FÜR DEN ANGRIFF

Charakteristischer Aufbau der Seiten



Apple

http://www.apple.com/

Google

Store | Mac | iPod + iTunes | iPhone | Downloads | Support | Search

# iPhone
## 3G

Twice as fast.

Half the price.

**Hot News Headlines** | Fans purchase and download more than 5 billion songs from iTunes

**WWDC Keynote Address**
Watch Apple CEO Steve Jobs unveil iPhone 3G.

**Introducing MobileMe**
Exchange for the rest of us.


**iPhone in Enterprise**
Push email, contacts, and calendars with Microsoft Exchange ActiveSync.

**iPhone 3G coming soon.**
Watch the new TV ad.

1x 

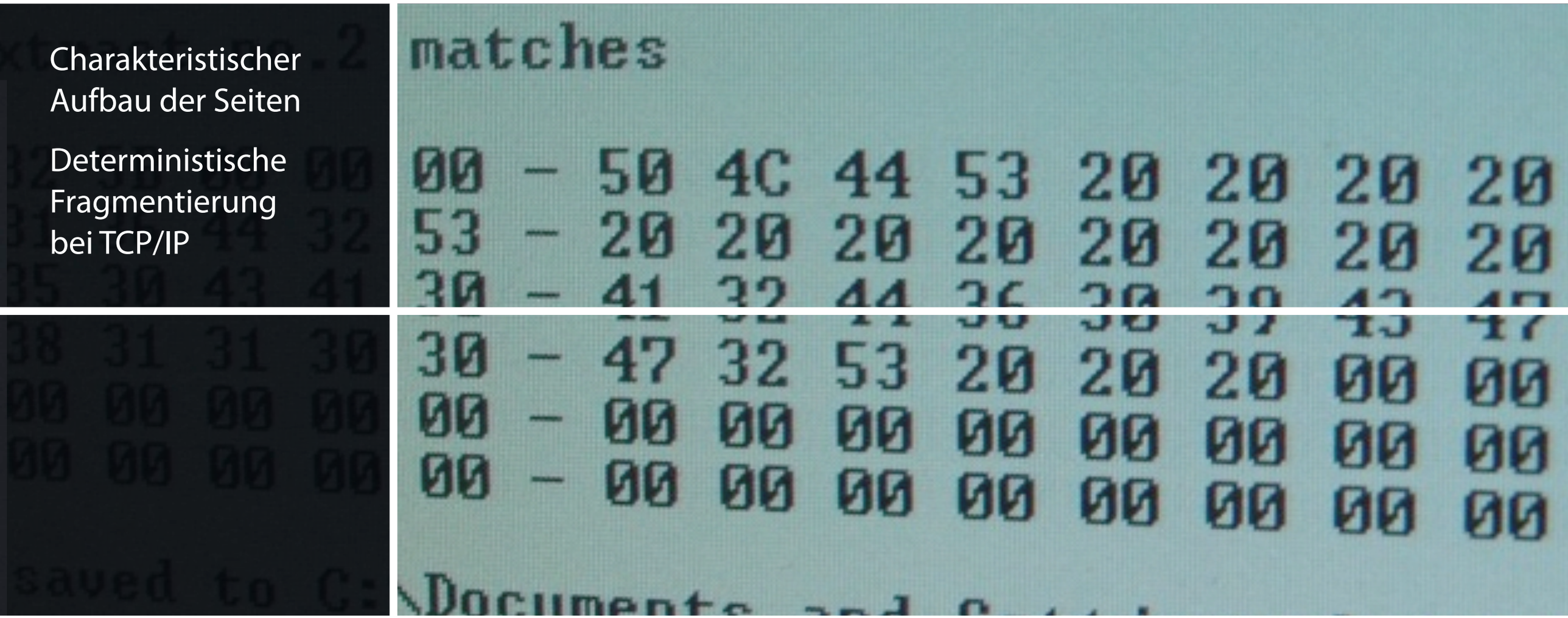
4x 

1x 

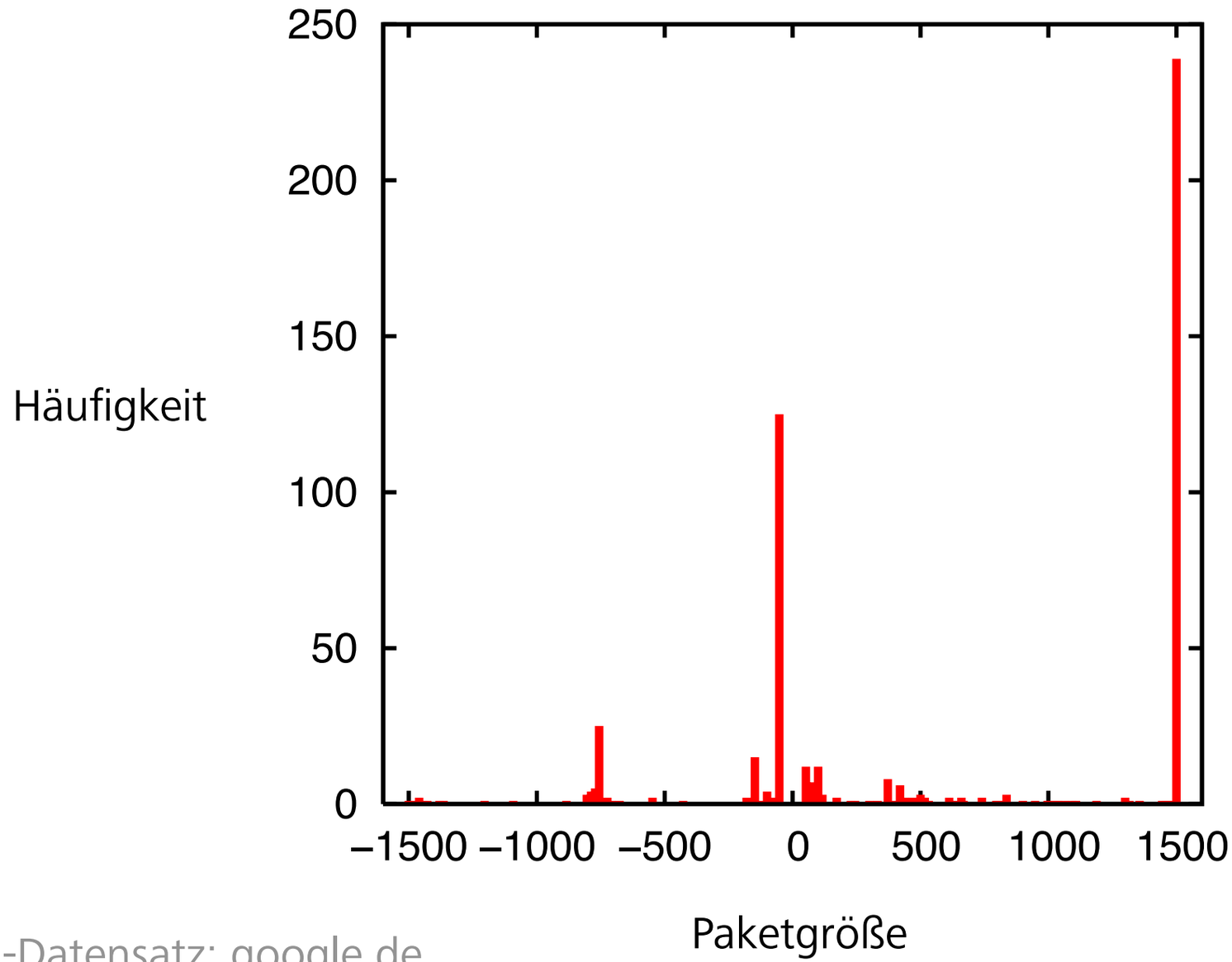
1x 

# GRUNDLAGE FÜR DEN ANGRIFF

- Charakteristischer Aufbau der Seiten
- Deterministische Fragmentierung bei TCP/IP



# Verteilung der sichtbaren IP-Paketgrößen

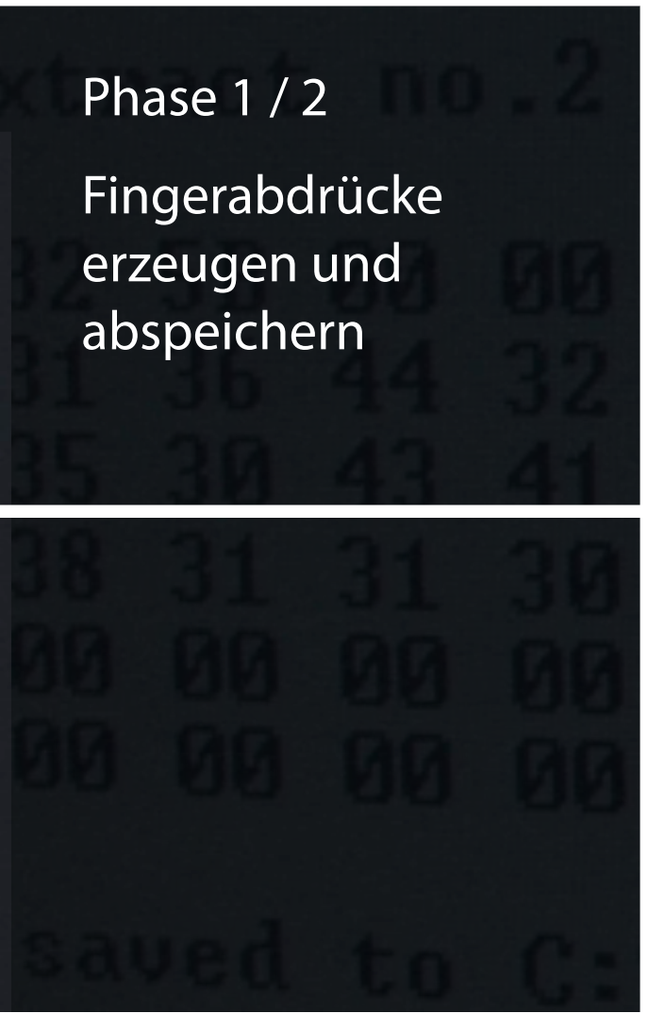


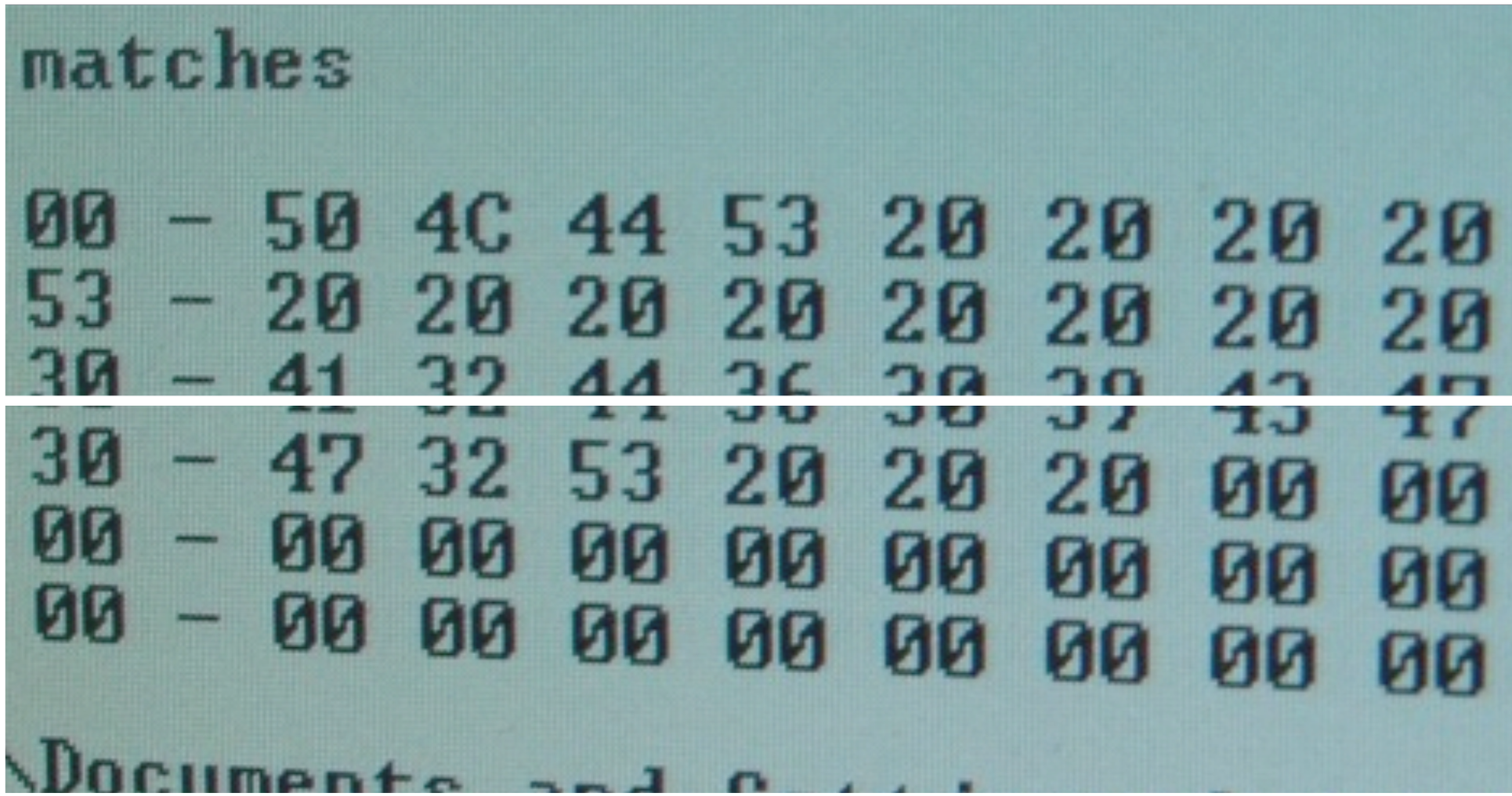

OpenSSH-Datensatz: google.de

# DURCHFÜHRUNG DES ANGRIFFS

Phase 1 / 2

Fingerabdrücke erzeugen und abspeichern



seite1.de

seite2.com

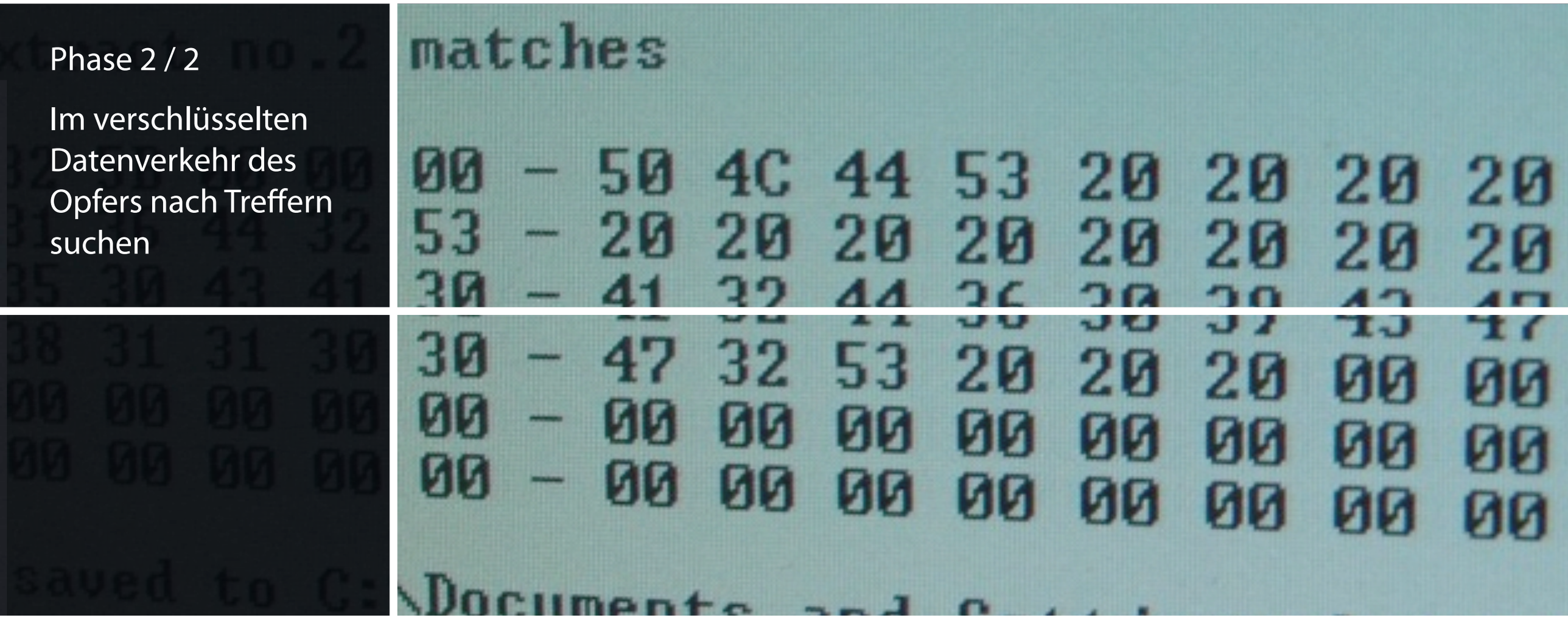
seite3.org

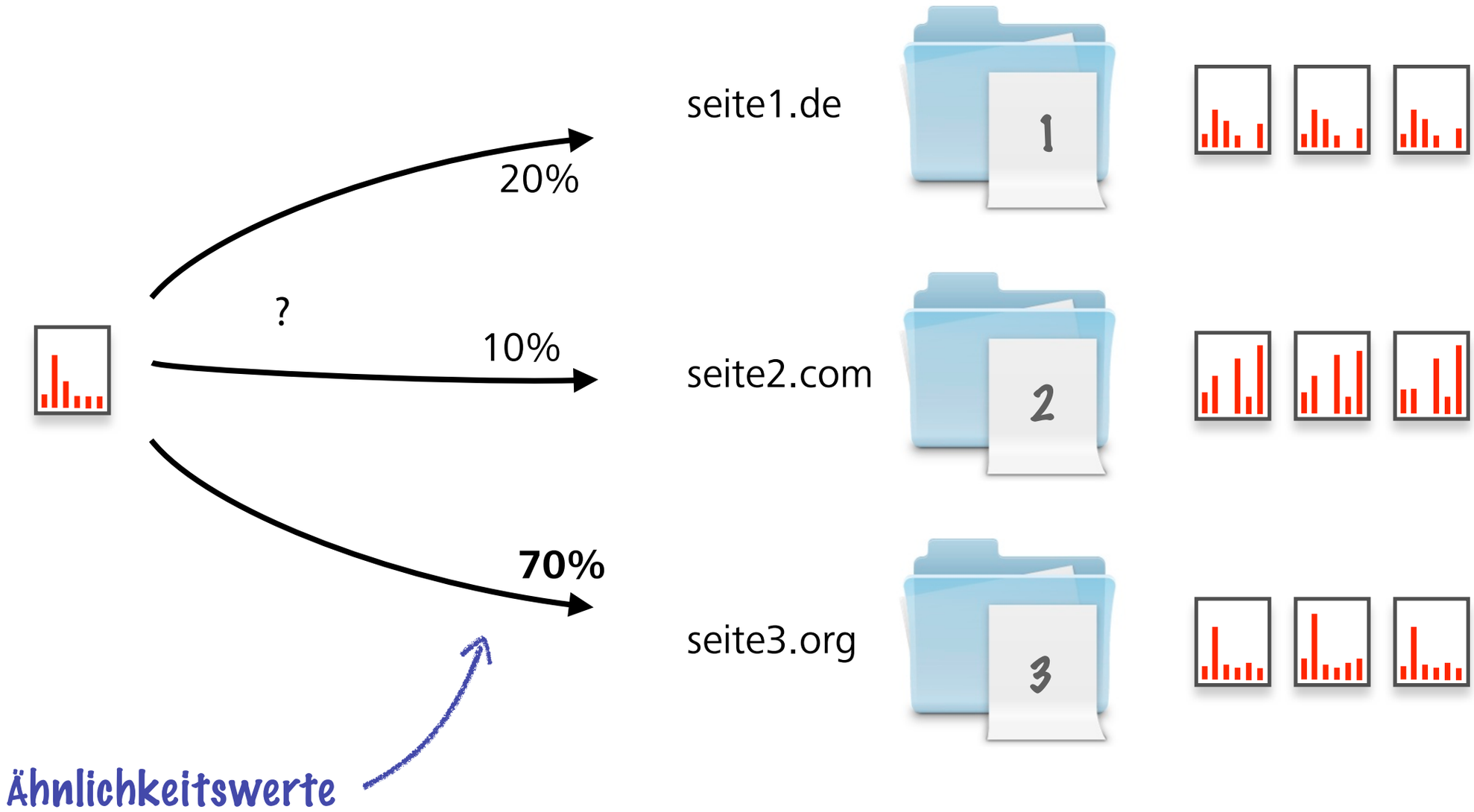
Trainingsdaten

# DURCHFÜHRUNG DES ANGRIFFS

Phase 2 / 2

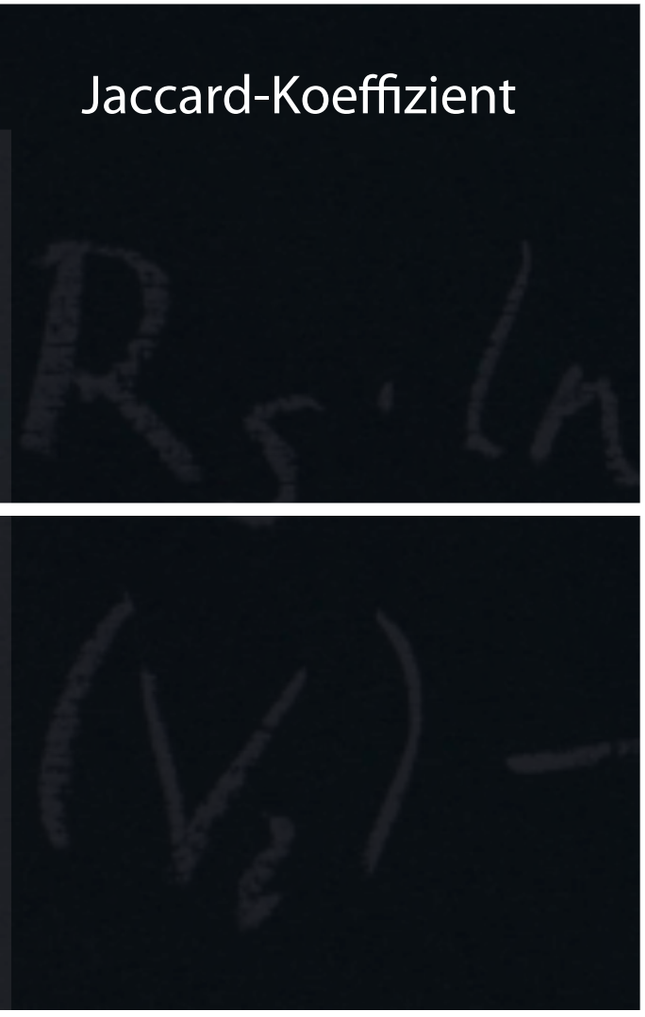
Im verschlüsselten Datenverkehr des Opfers nach Treffern suchen

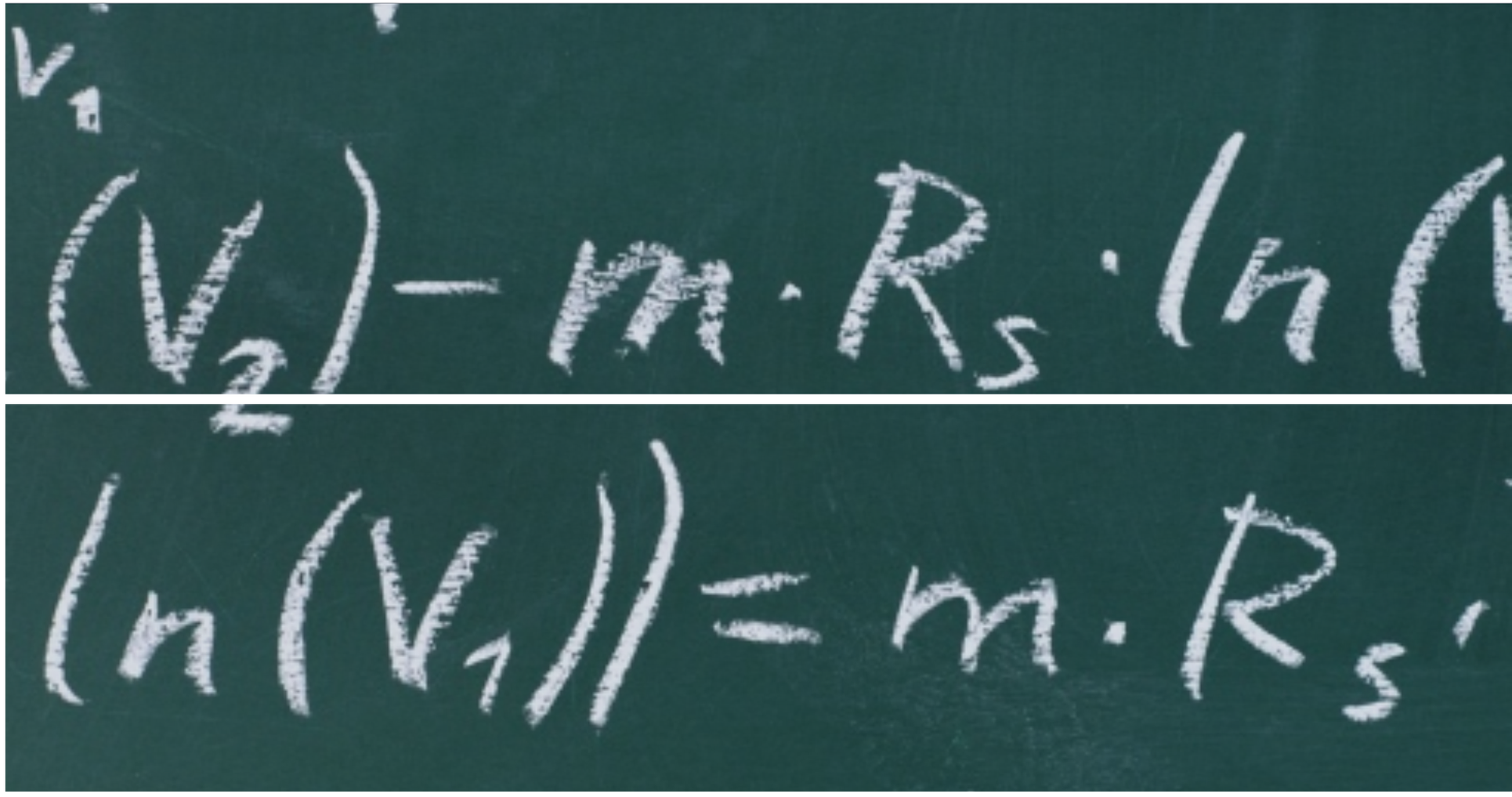



seite1.de
1
20%
?
10%
seite2.com
2
70%
seite3.org
3
Ähnlichkeitswerte

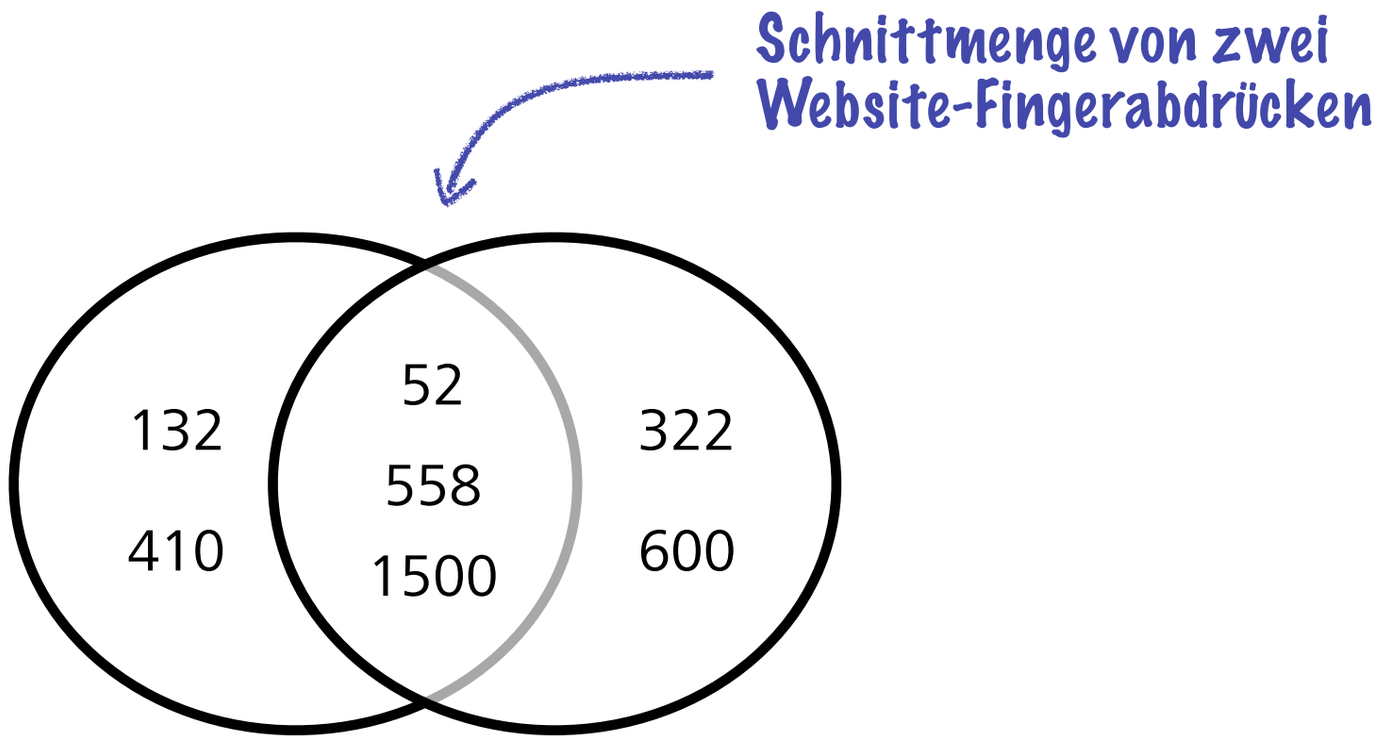
# ANALYSE-METHODEN

Jaccard-Koeffizient





# Jaccard-Koeffizient



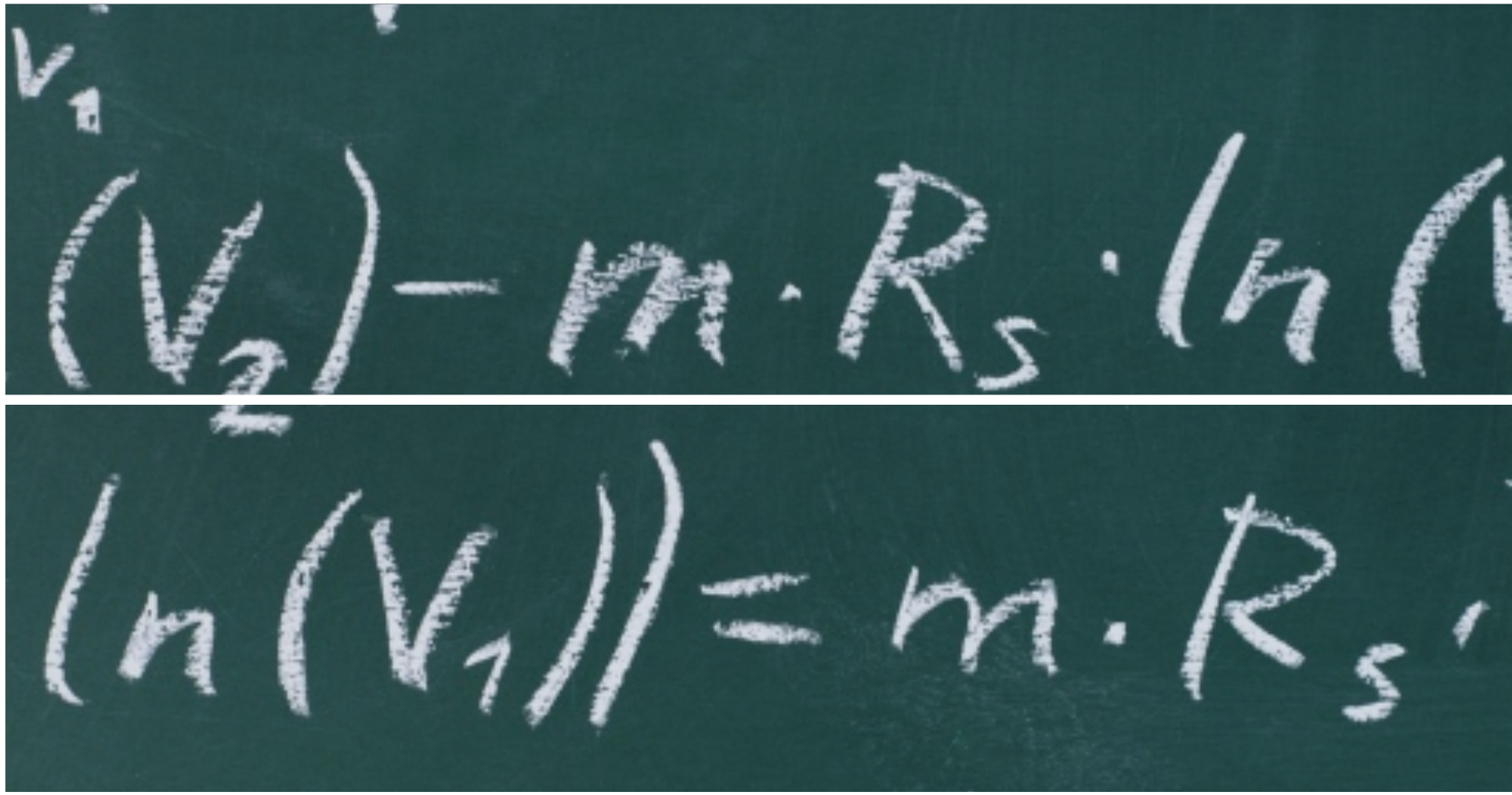

$$A = \{ 52, 132, 410, 558, 1500 \}$$

$$s = 3 / 7 = 0{,}42$$

# ANALYSE-METHODEN

Jaccard-Koeffizient

Naïve Bayes



# Multinomialer Naïve Bayes Klassifizierer

Wahrscheinlichkeit, dass unbekannter Fingerabdruck zu einer gewissen Webseite gehört

Wahrscheinlichkeit steigt, wenn der Fingerabdruck viele Pakete enthält, die häufig zu beobachten sind, wenn die gesuchte Webseite abgerufen wird

$$P(c_i|d) \sim P(c_i) \cdot C_{multi} \cdot \prod_{t \in V} P(t|c)^{f_{t,d}}$$

$$\hat{P}(t|c) = \frac{f_{t,c}}{\sum_{t' \in V} f_{t',c}}$$

# STUDIE

775 populäre Webseiten

7 Übertragungs-verfahren untersucht

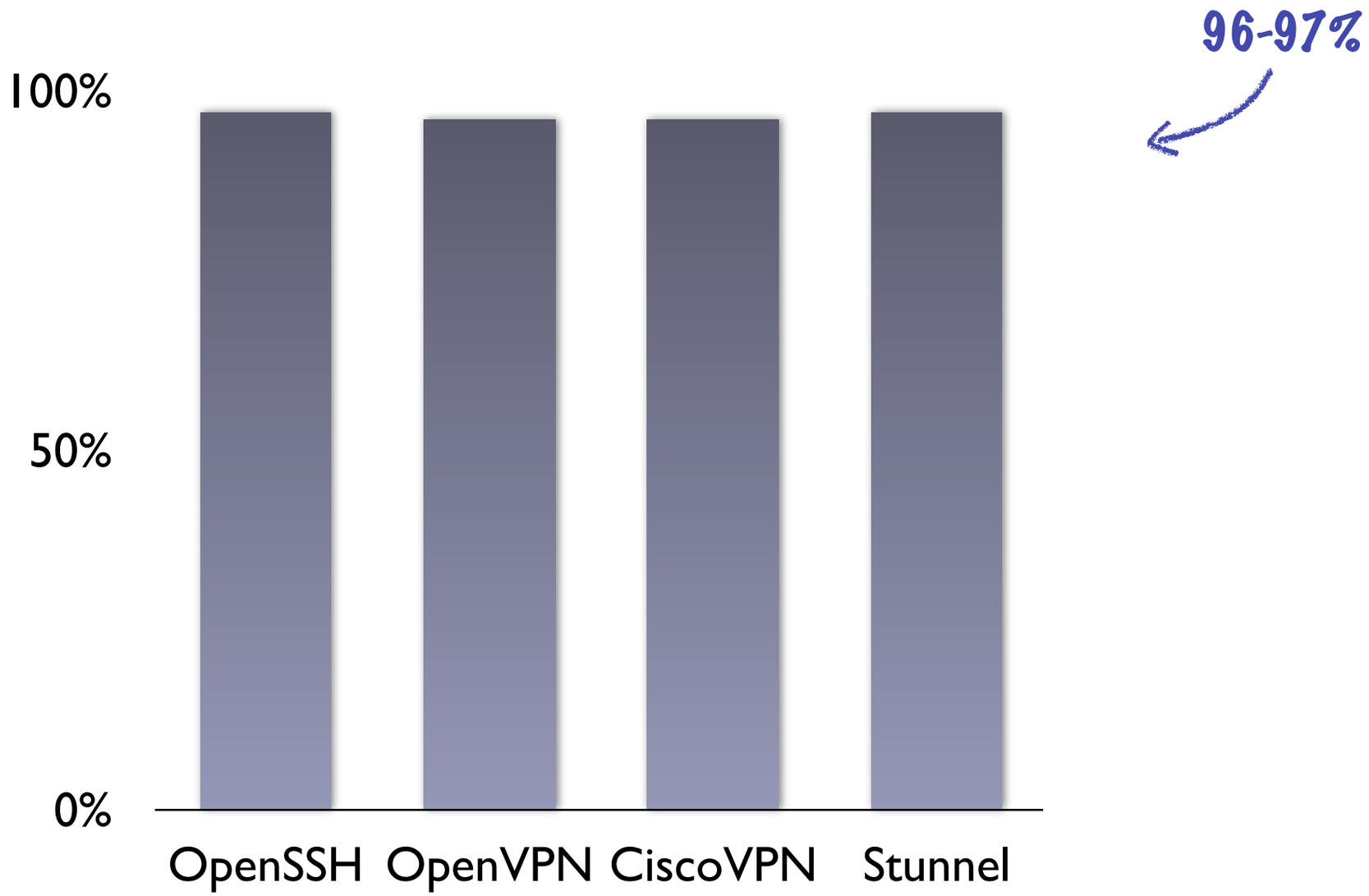
# ERGEBNISSE

Erkennungsraten bis zu 97% bei VPNs etc.

# Erkennungsraten

96-97%

# Erkennungsraten

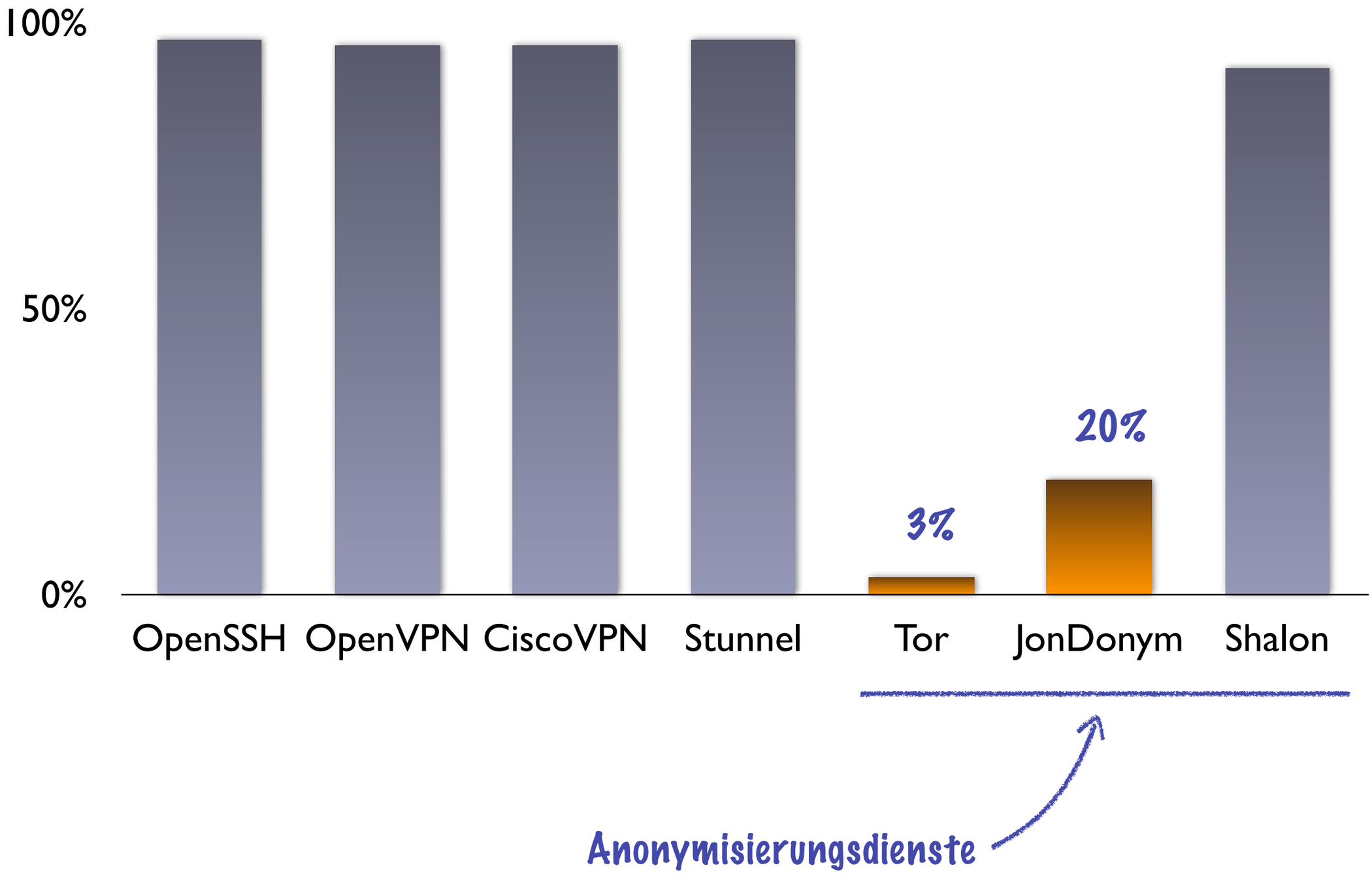


# ERGEBNISSE

Anonymisierungs-
systeme bieten
besseren Schutz

können Angriff nicht
immer abwehren

# Erkennungsraten



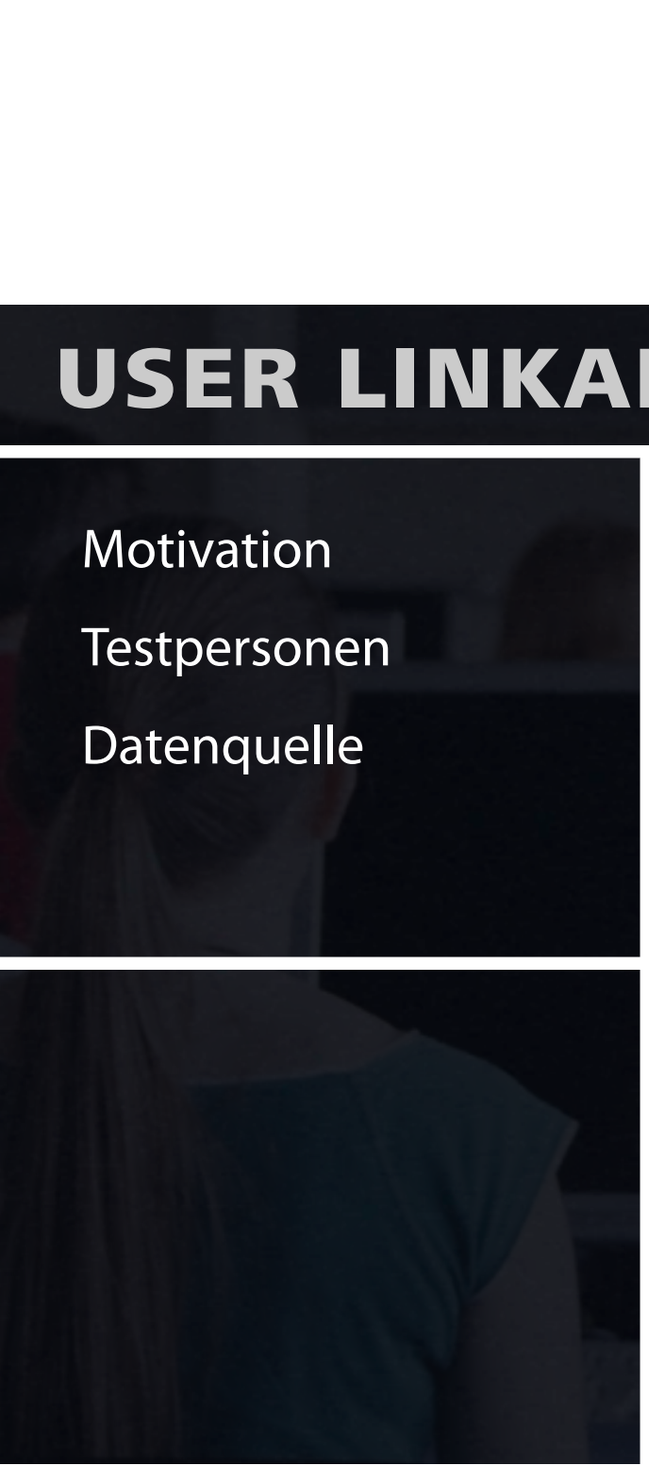
# USER LINKABILITY

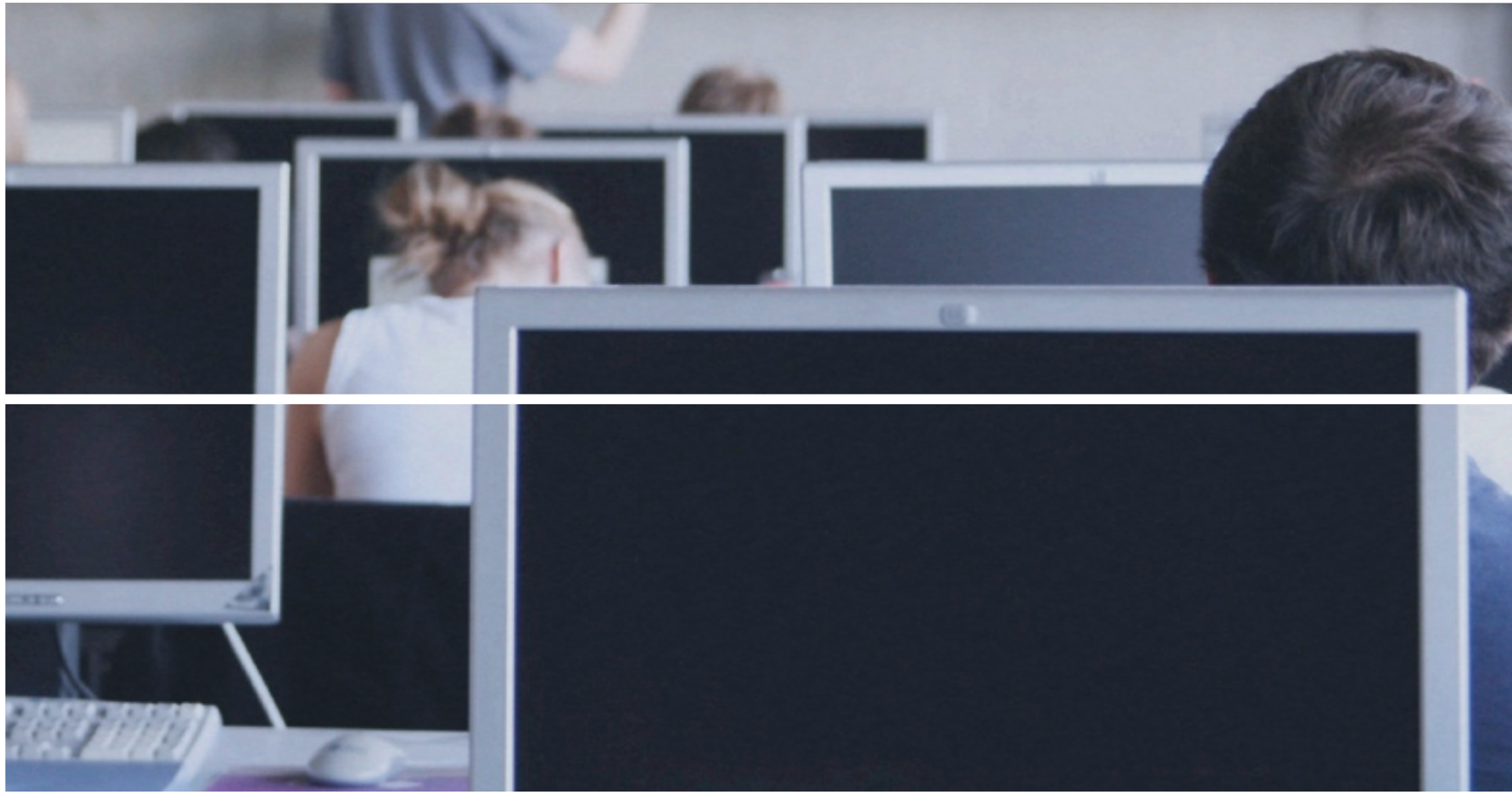
# USER LINKABILITY

Motivation

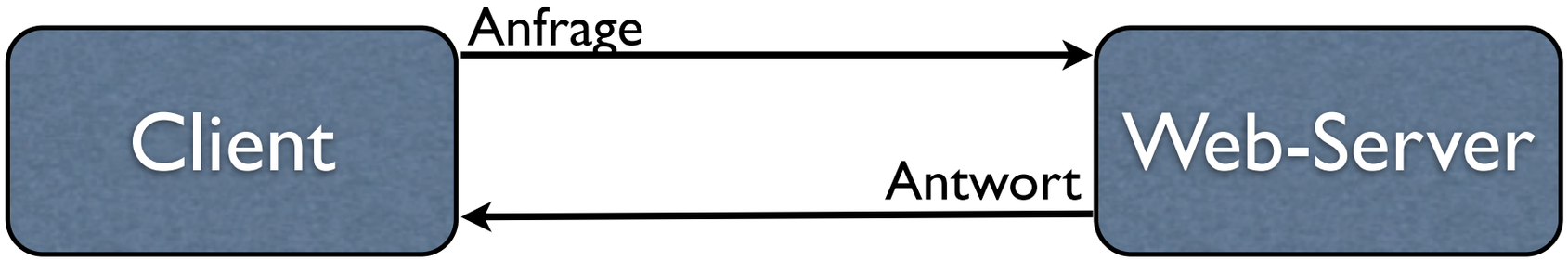
Testpersonen

Datenquelle





# HTTP-Anfrage



# Request Splitting

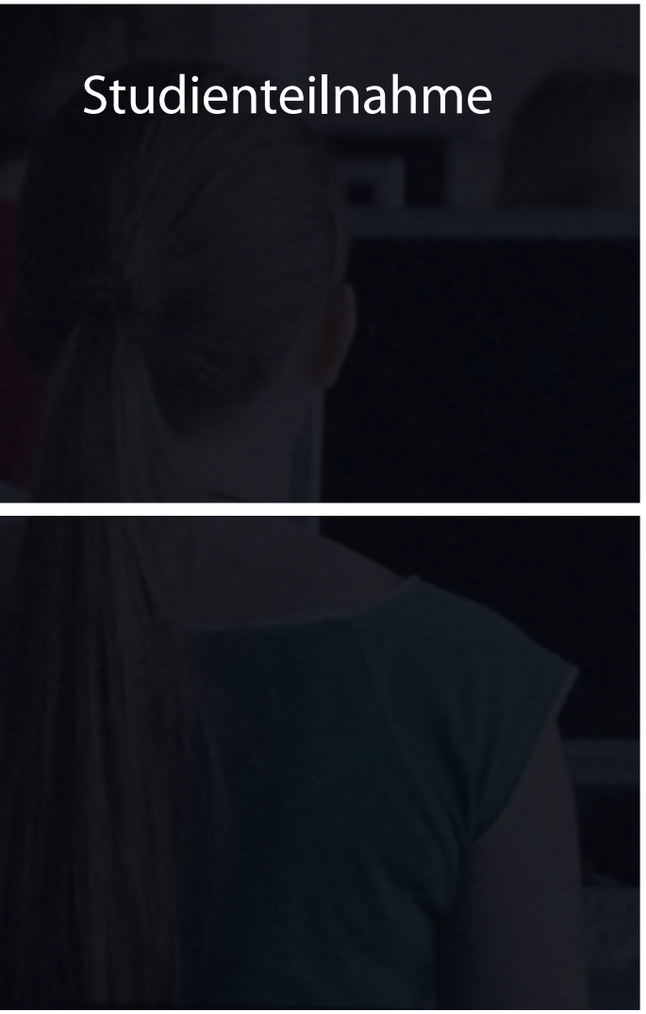
Host www.google.de

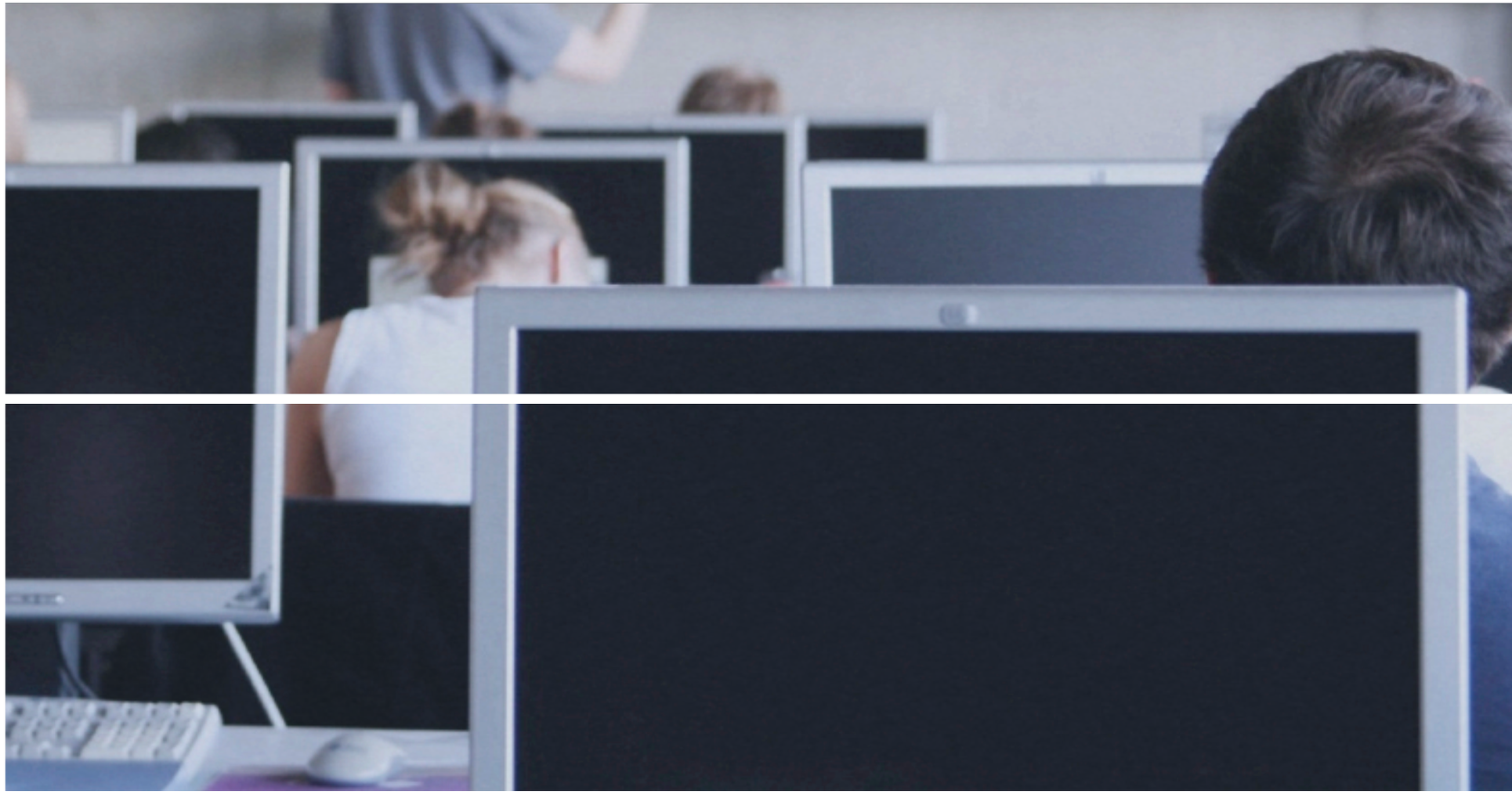
Suffix search?q=informationswissenschaft

GET „http://www.google.de/search?q=informationswissenschaft HTTP/1.1“

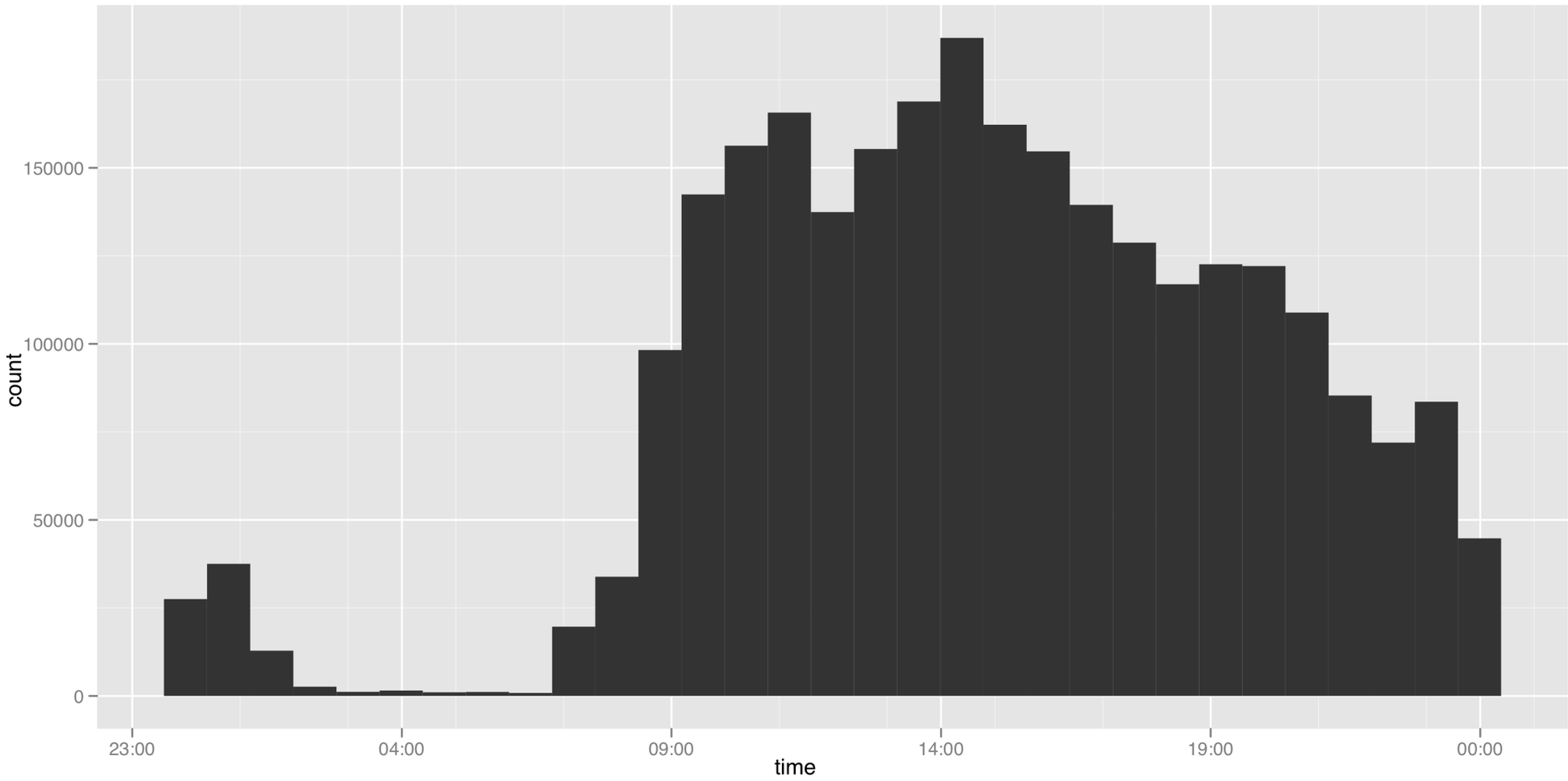
request

# USER LINKABILITY

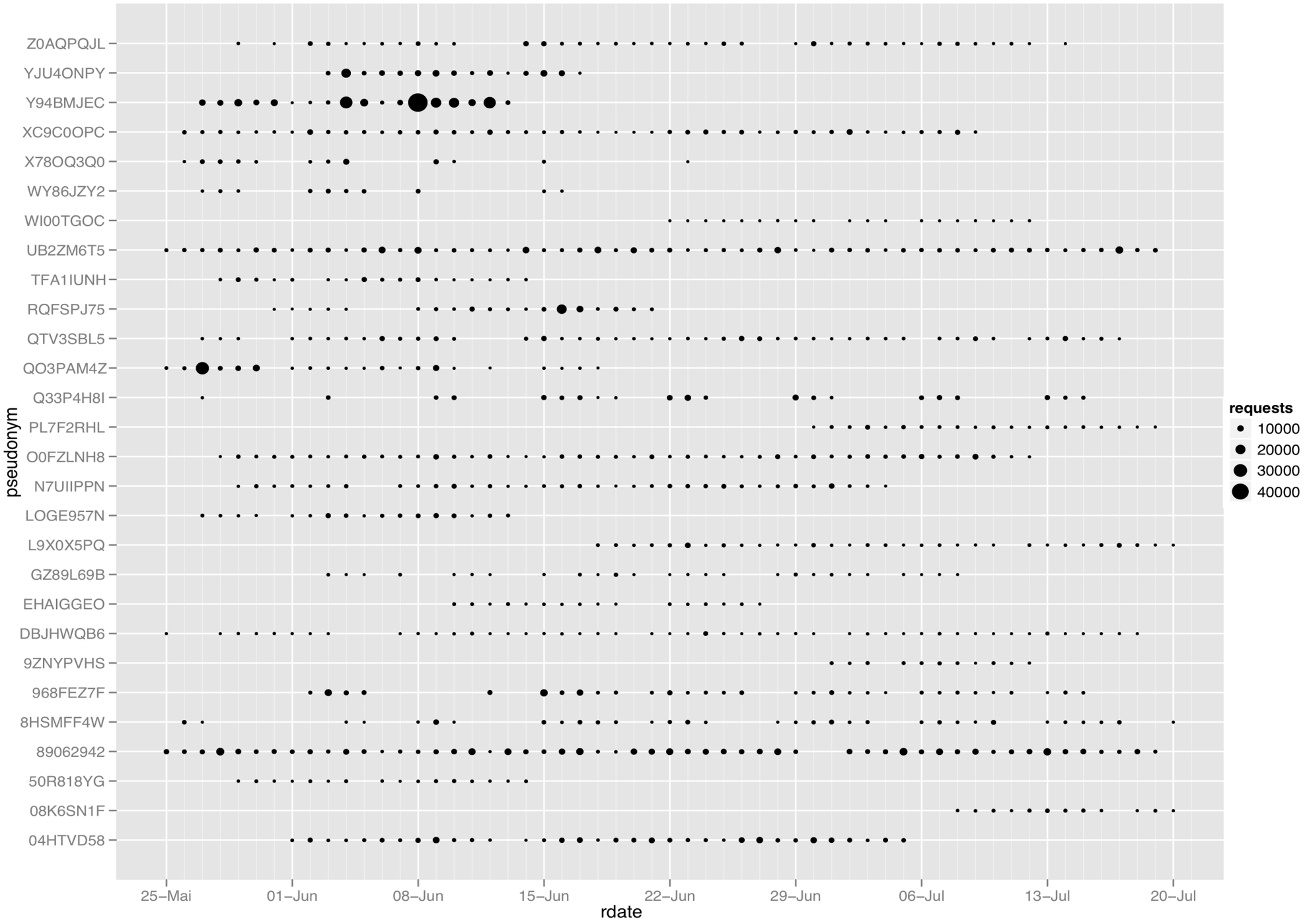


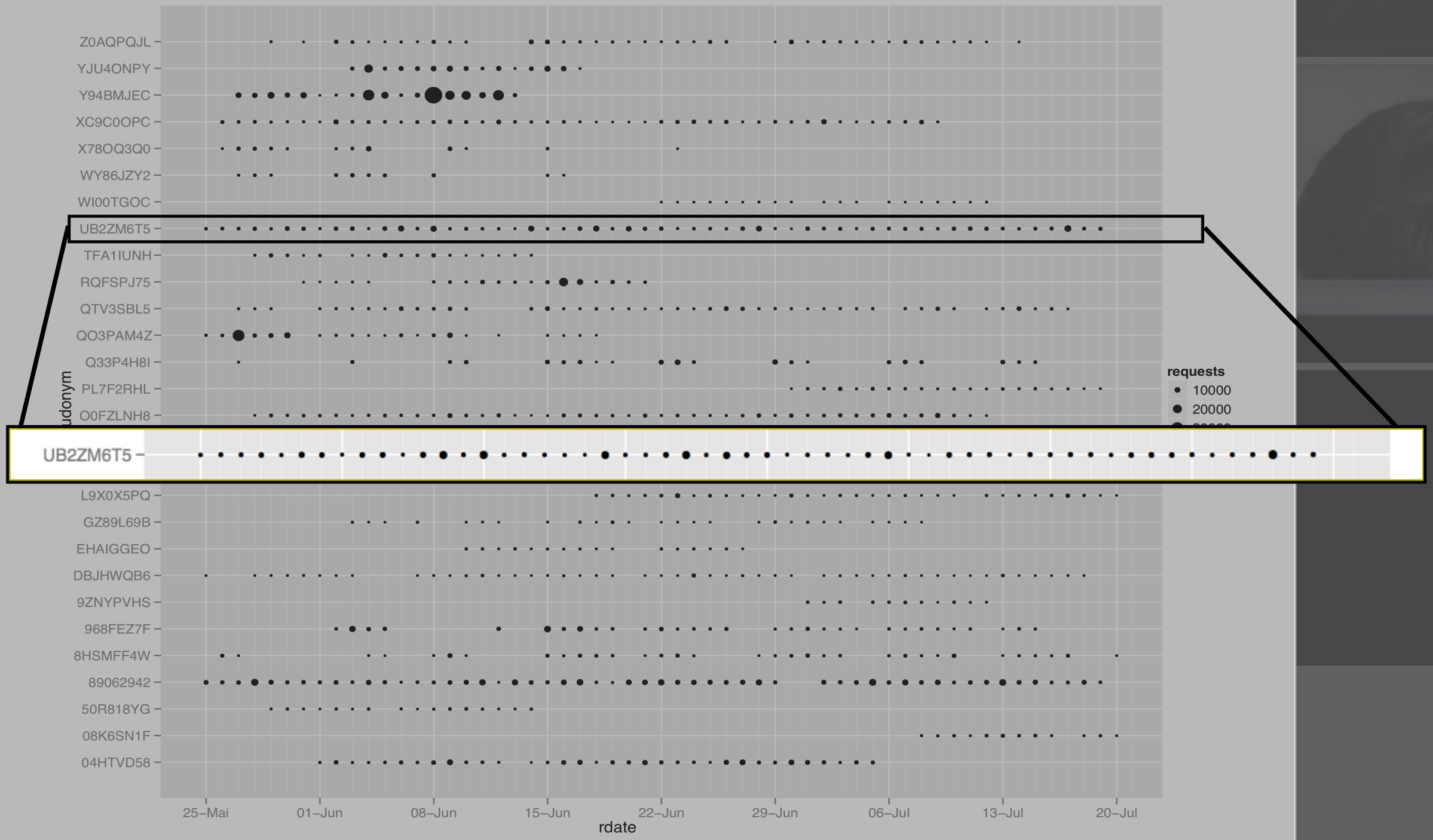
# Aktivität im Tagesverlauf



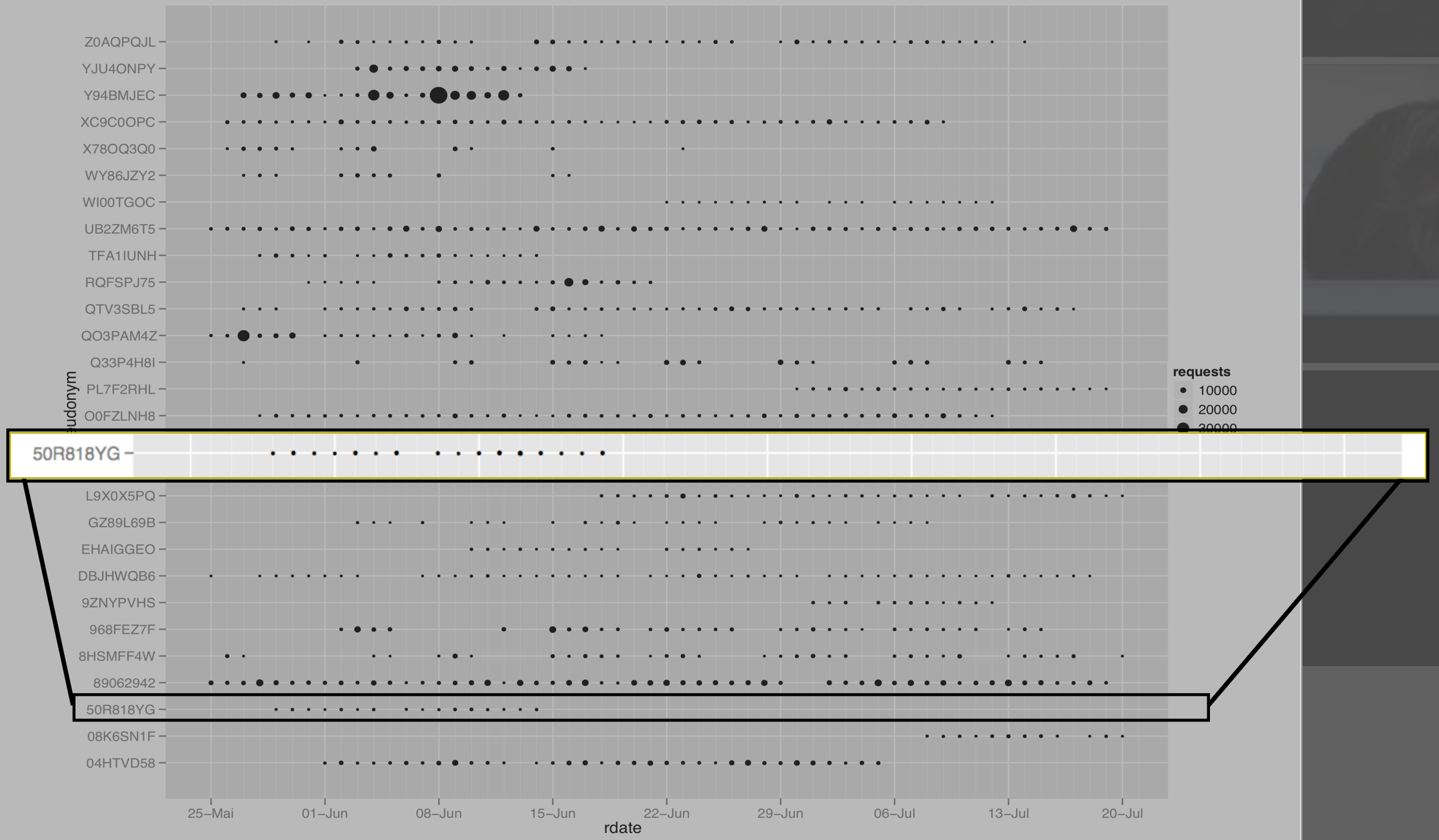
# Tagesnutzung



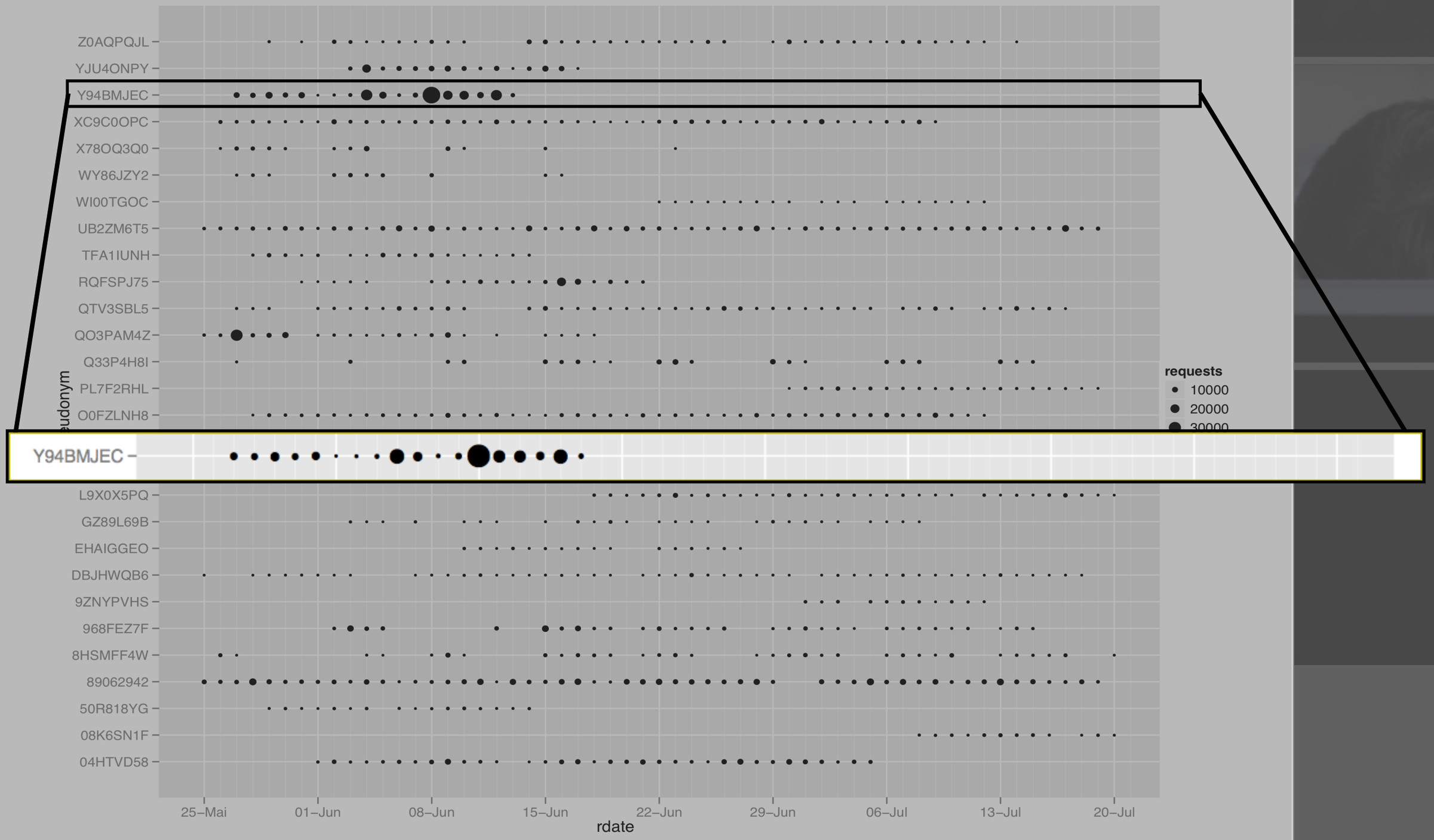
# Tagesnutzung



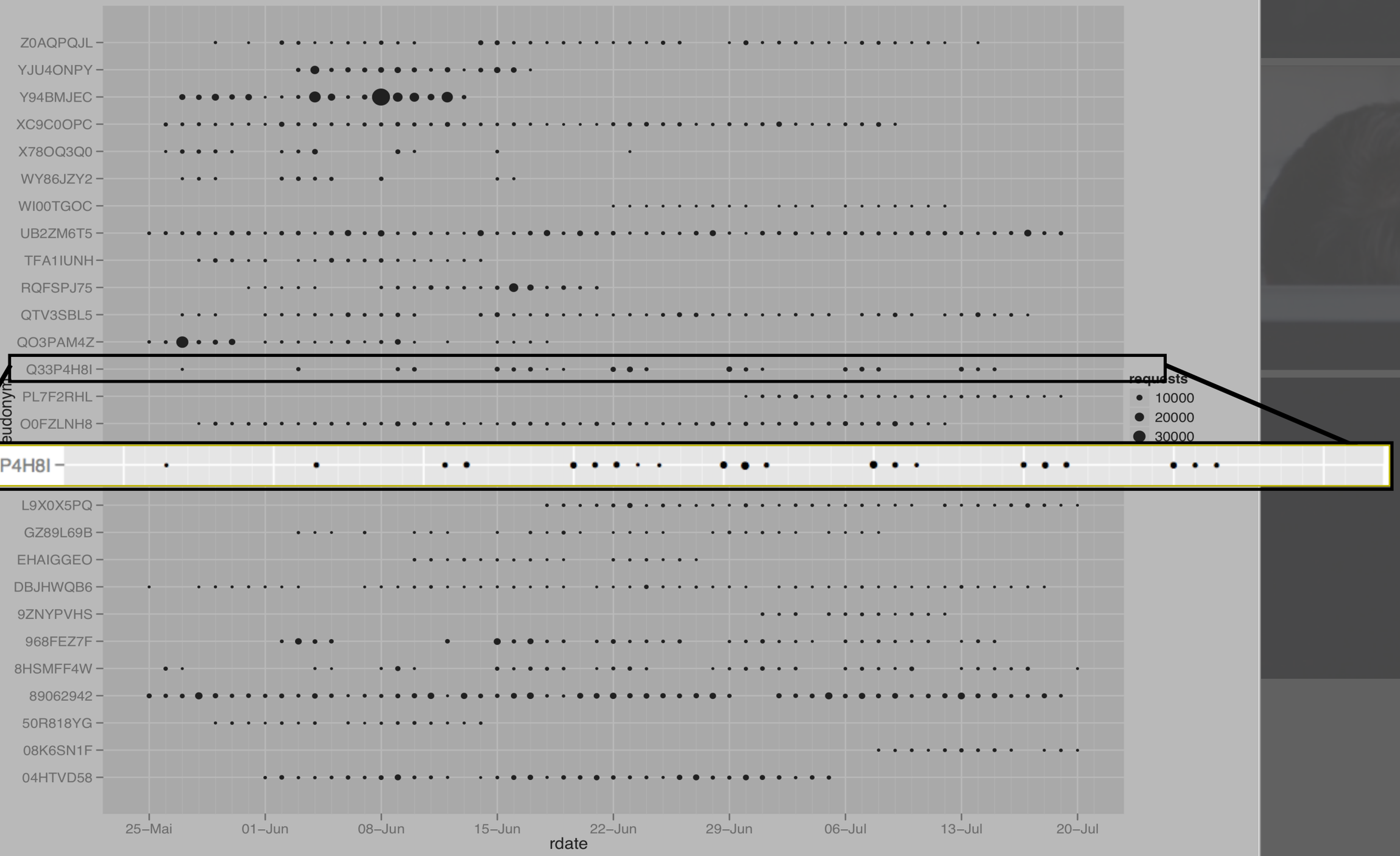
# Tagesnutzung



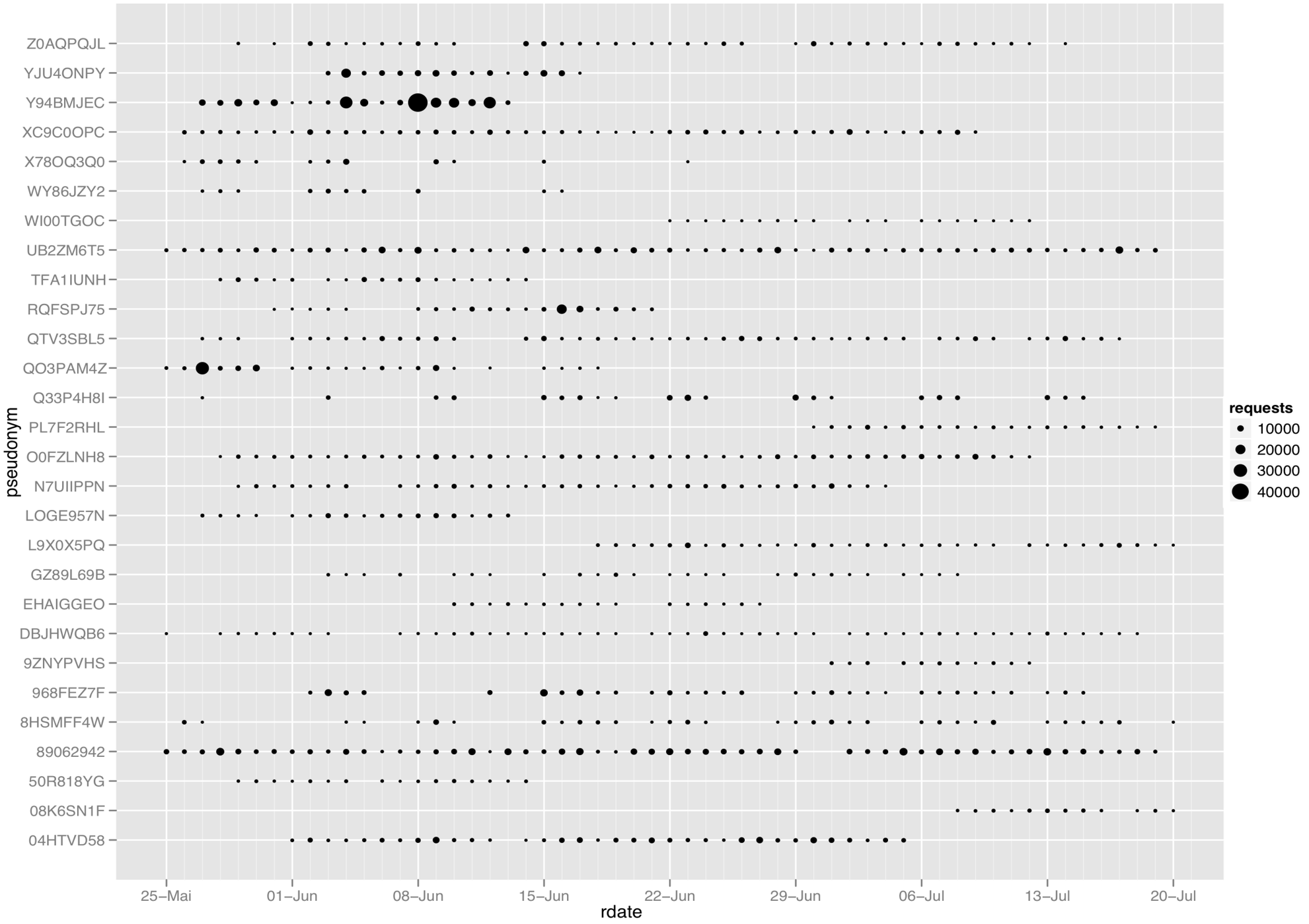
# Tagesnutzung



# Tagesnutzung

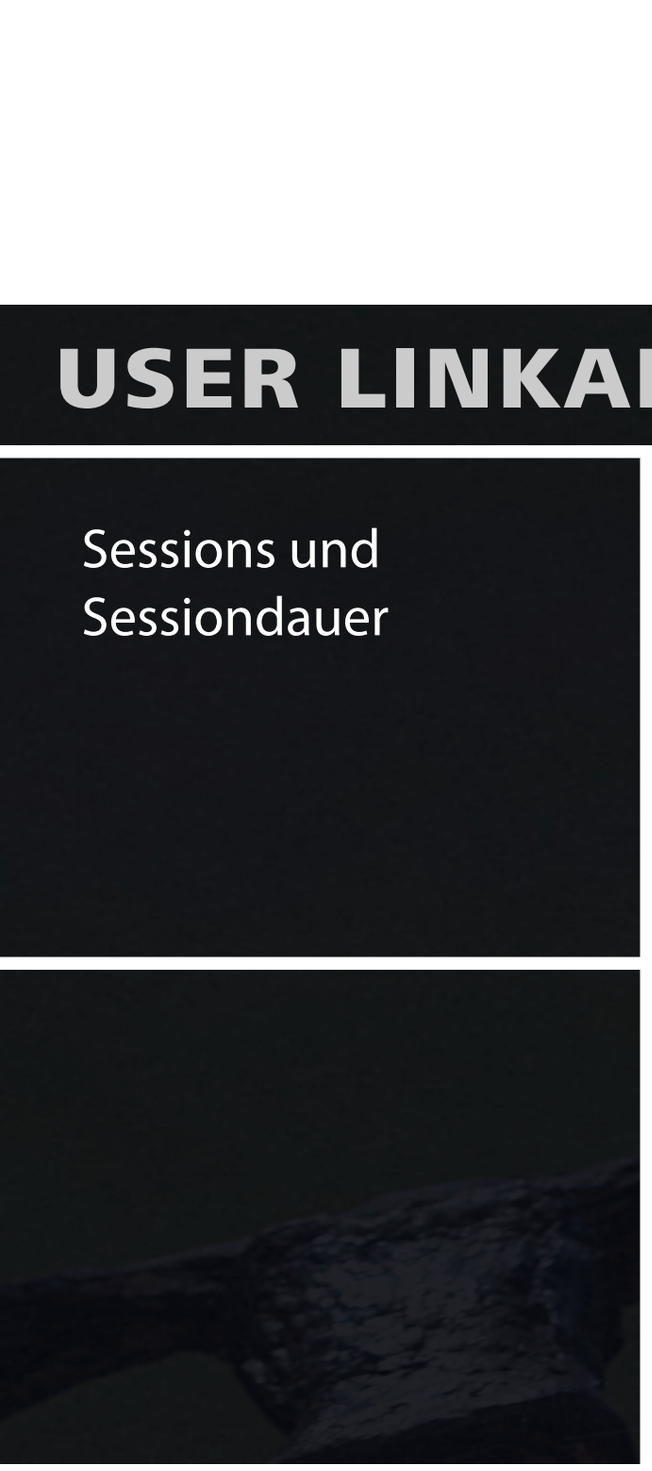


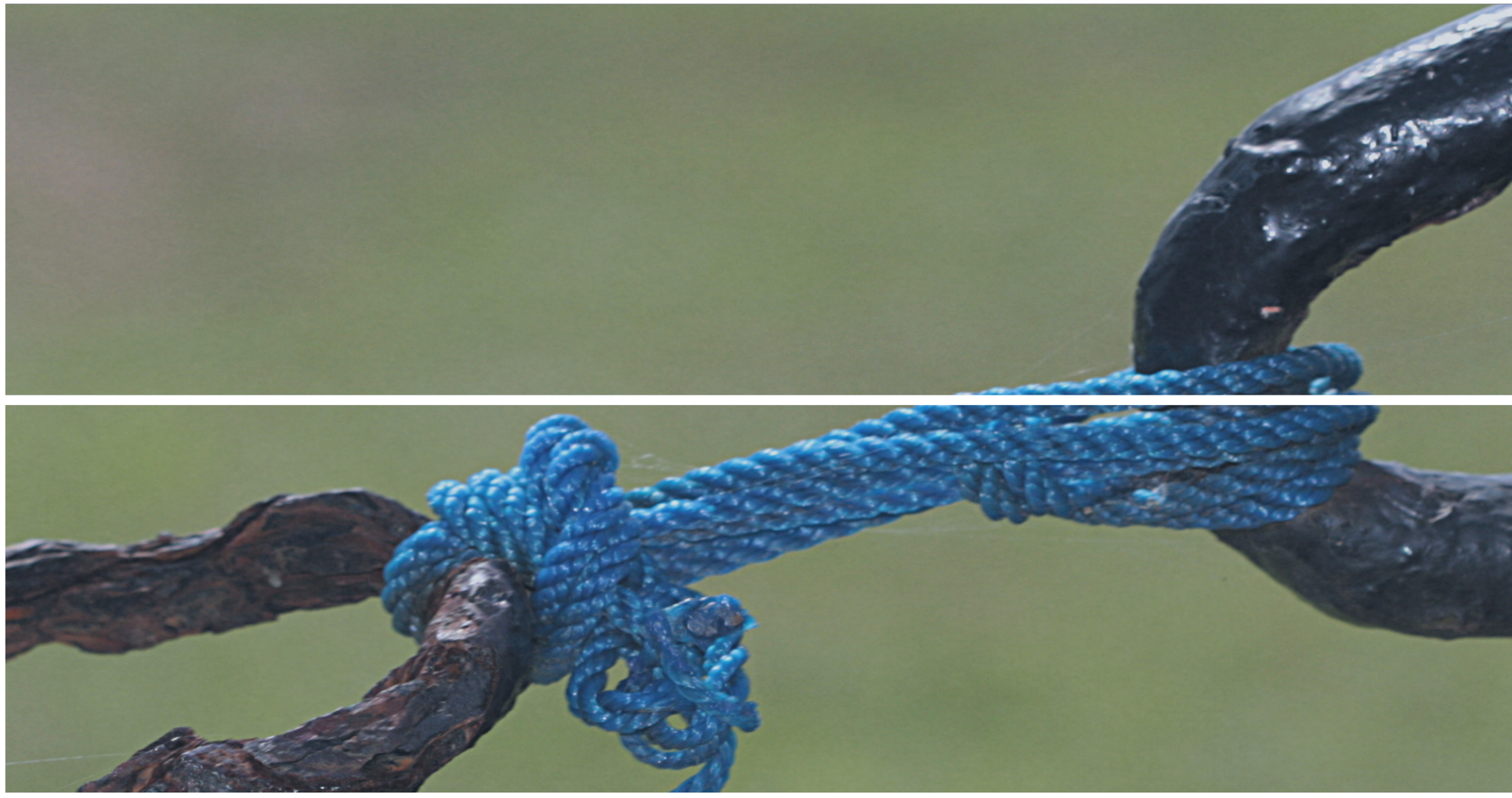
# Tagesnutzung



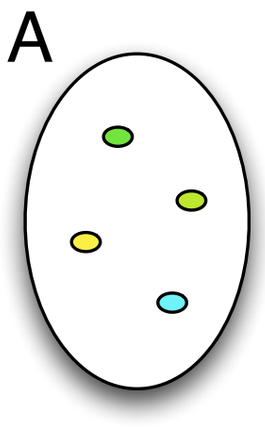
# USER LINKABILITY

Sessions und
Sessiondauer



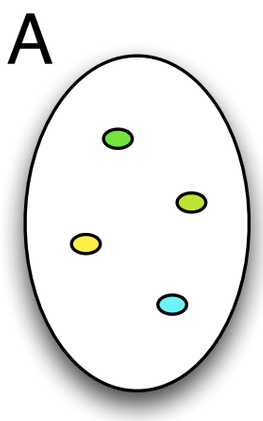


# Sessions

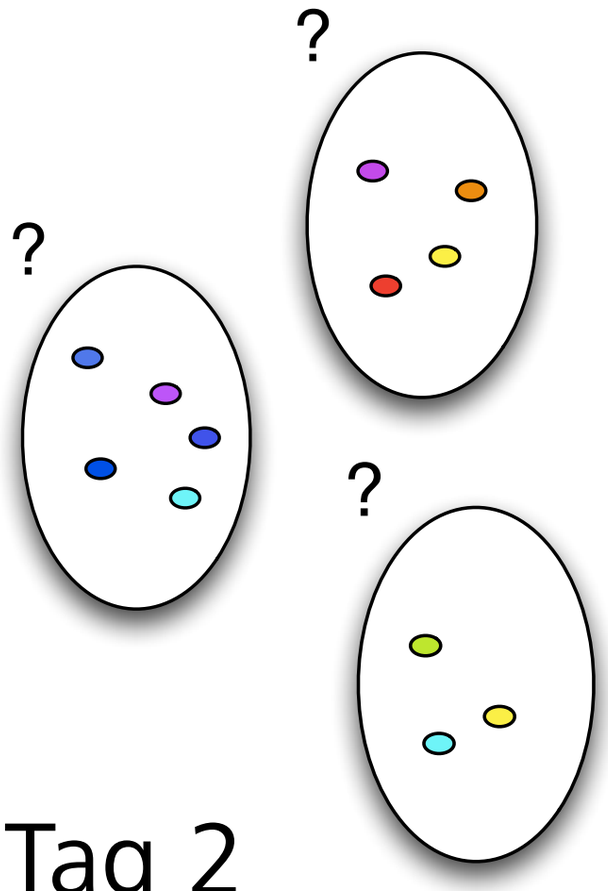



Tag 1

# Sessions




Tag 1

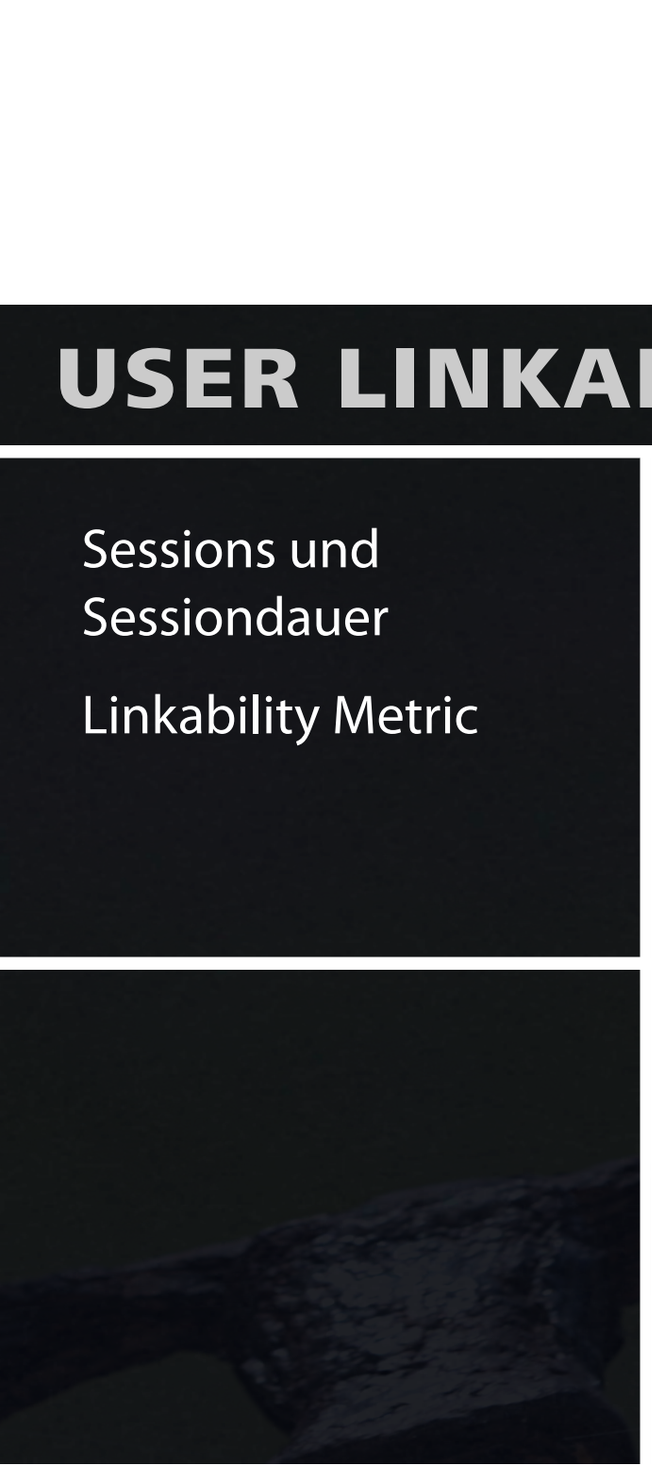


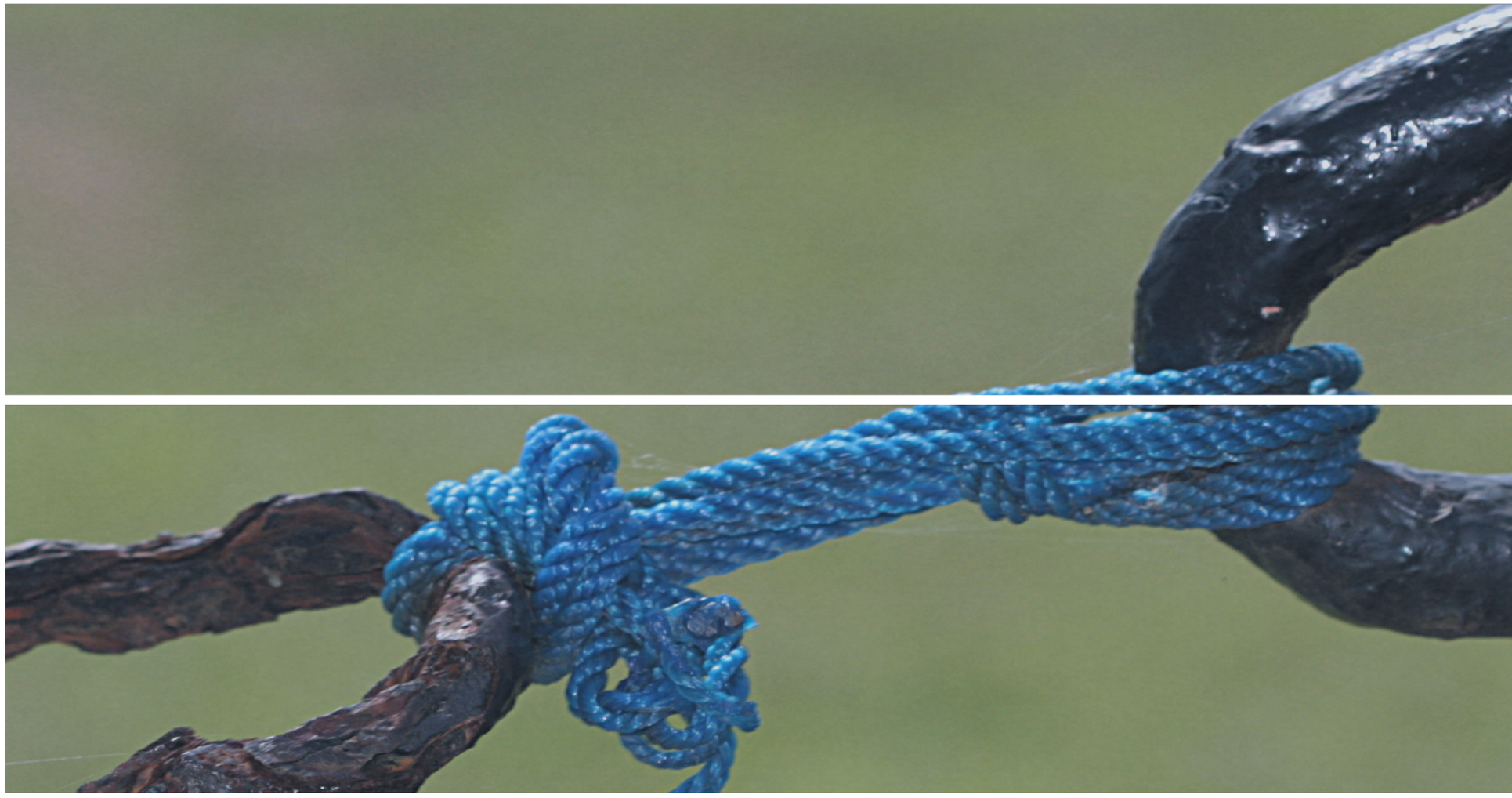

Tag 2

# USER LINKABILITY

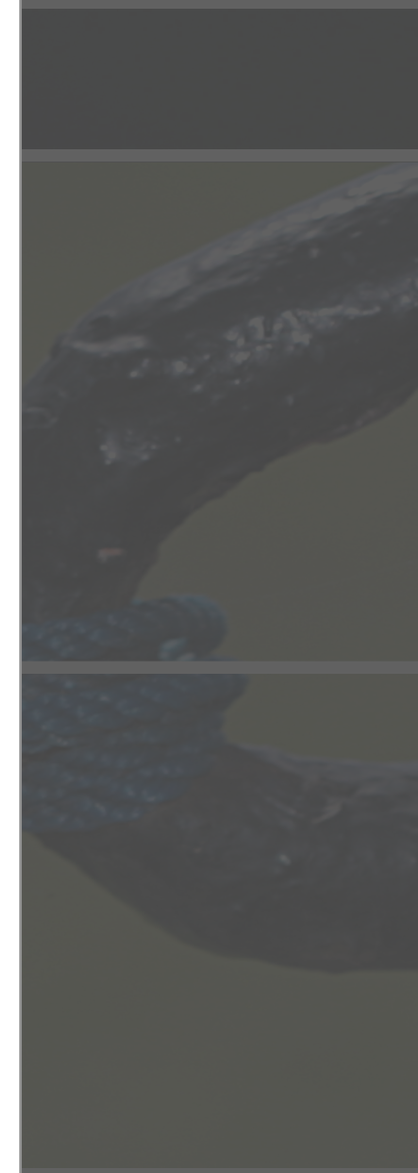
Sessions und
Sessiondauer

Linkability Metric





# Linkability Metric



$$L = \frac{|H_P|}{|H|} * \frac{|S_{P_H}|}{|S_P|}$$

# Linkability Metric

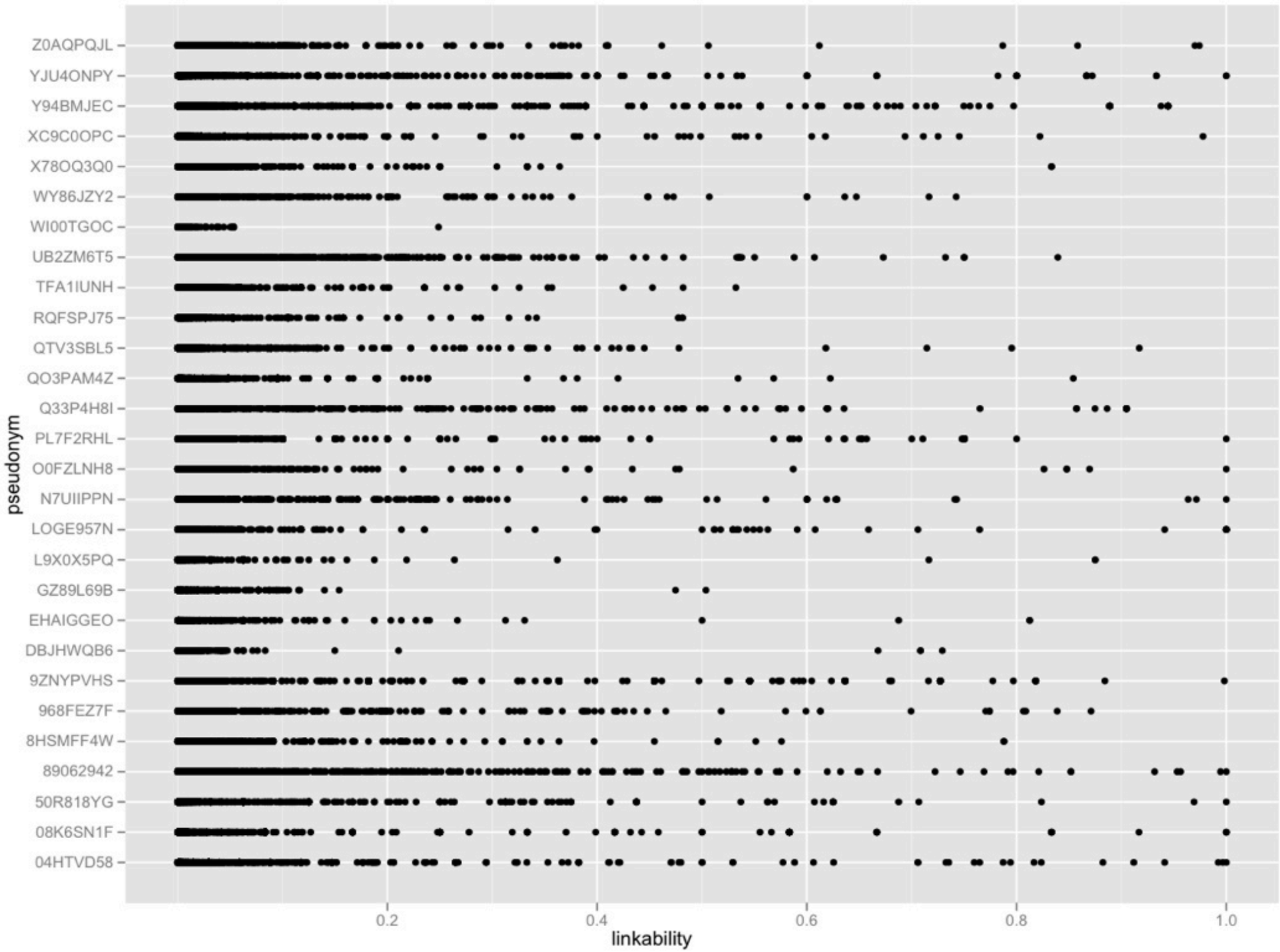
$$\frac{\text{Aufrufe eines Hosts durch einen Benutzer}}{\text{Aufrufe eines Hosts durch alle Benutzer}} * \frac{\text{Anzahl der Sessions in denen der Benutzer den Host aufgerufen hat}}{\text{Alle Sessions des Benutzers}}$$

Wird 1 wenn nur der Benutzer den Host aufruft

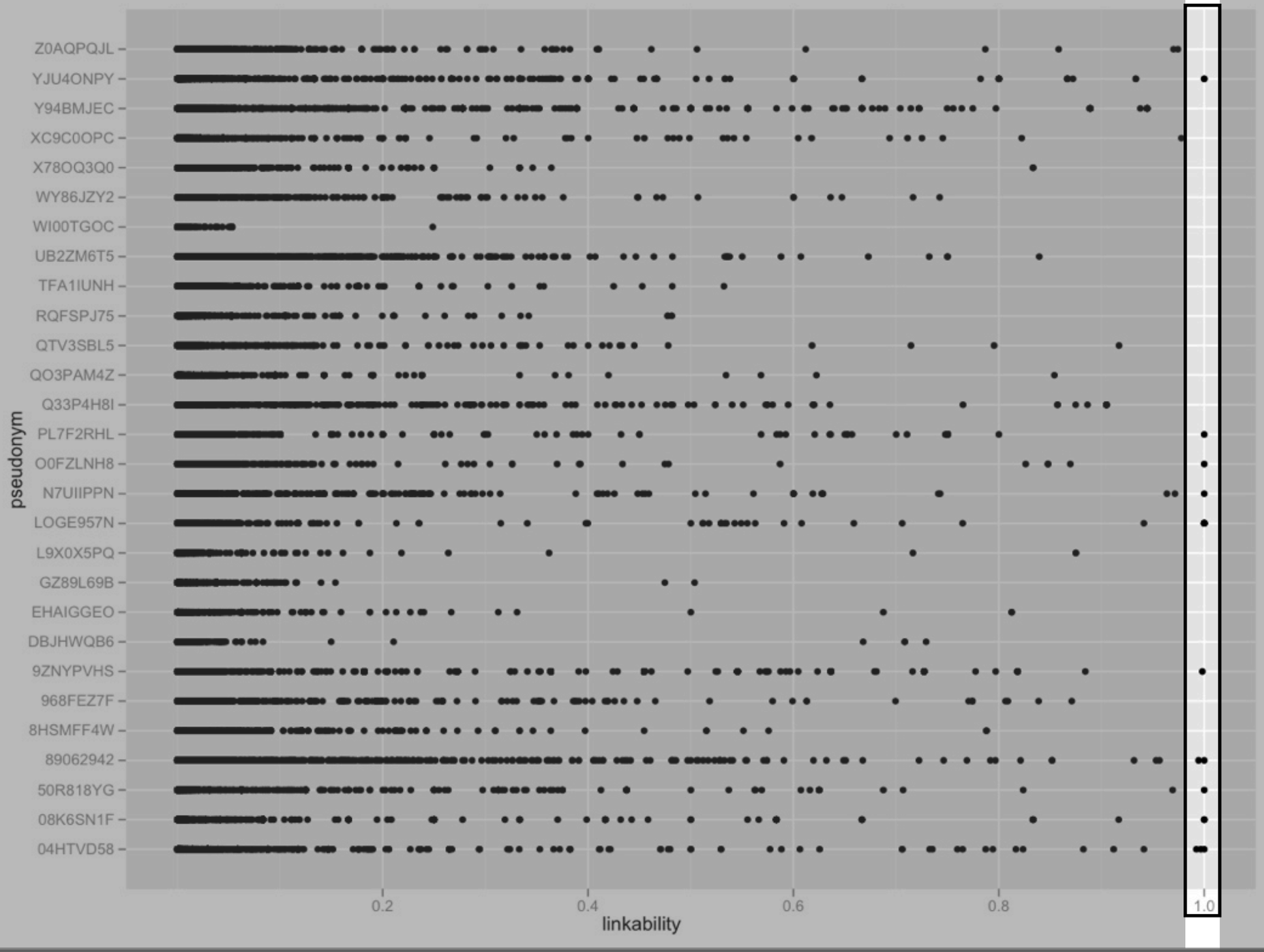
Wird 1 wenn der Benutzer den Host täglich aufruft

$$L = \frac{|H_P|}{|H|} * \frac{|S_{P_H}|}{|S_P|}$$

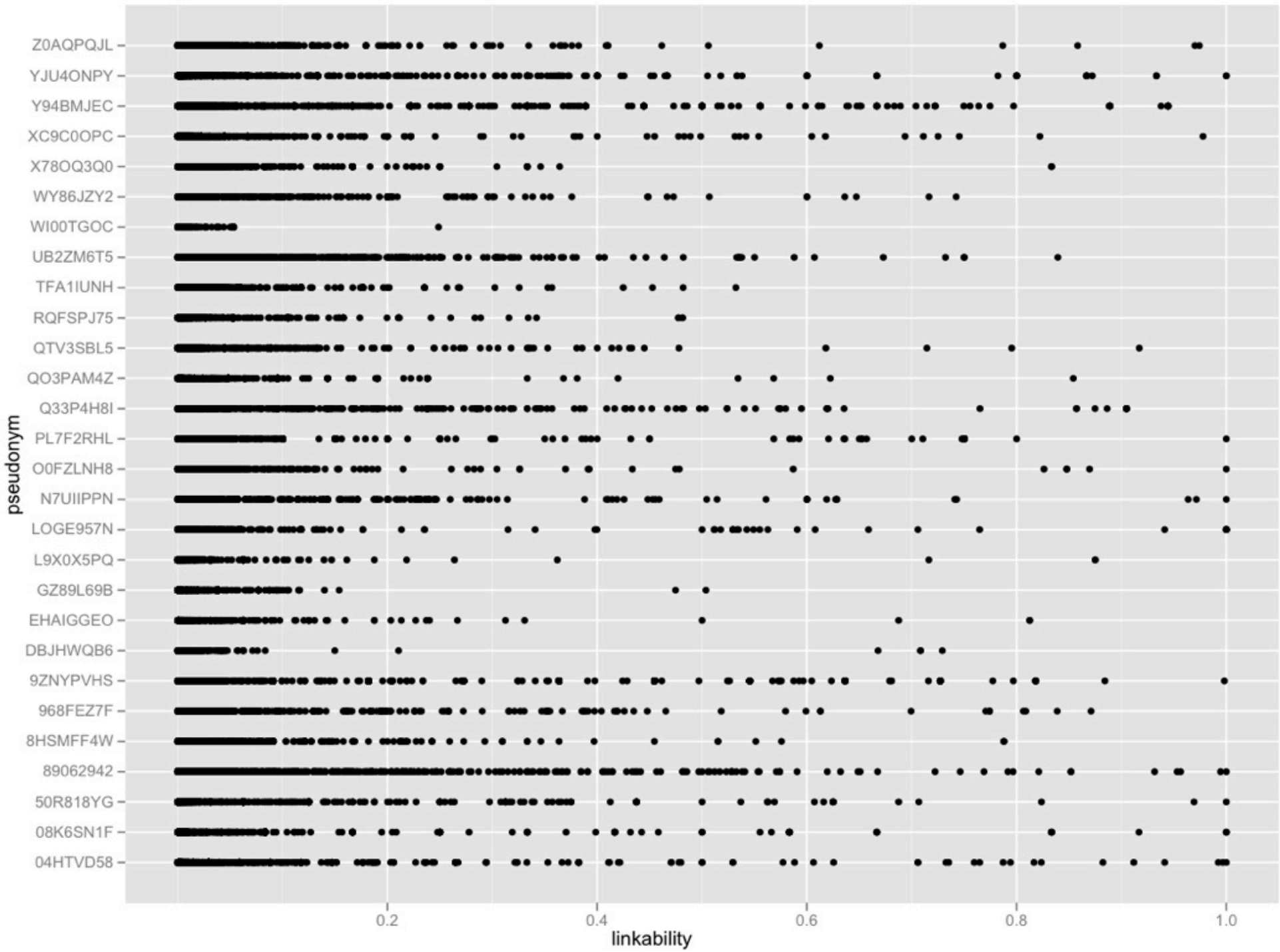
# Linkability Metric



# Linkability Metric

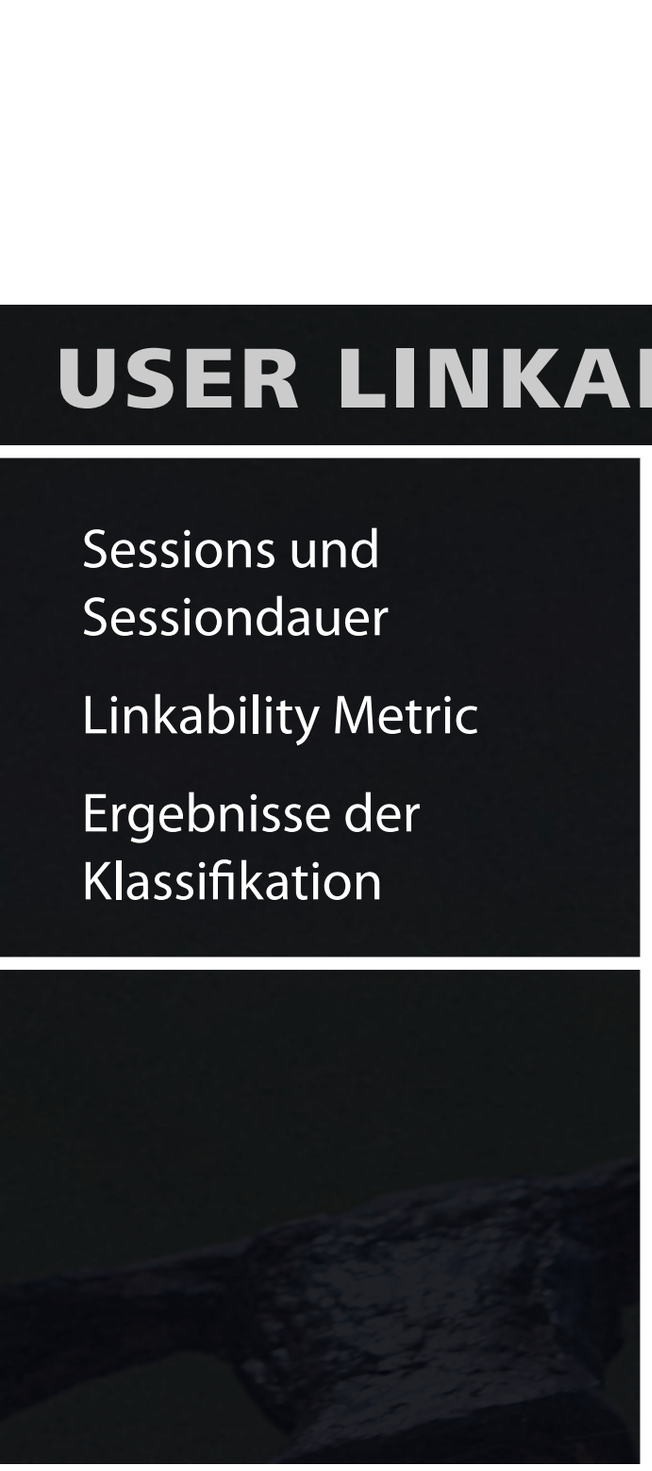


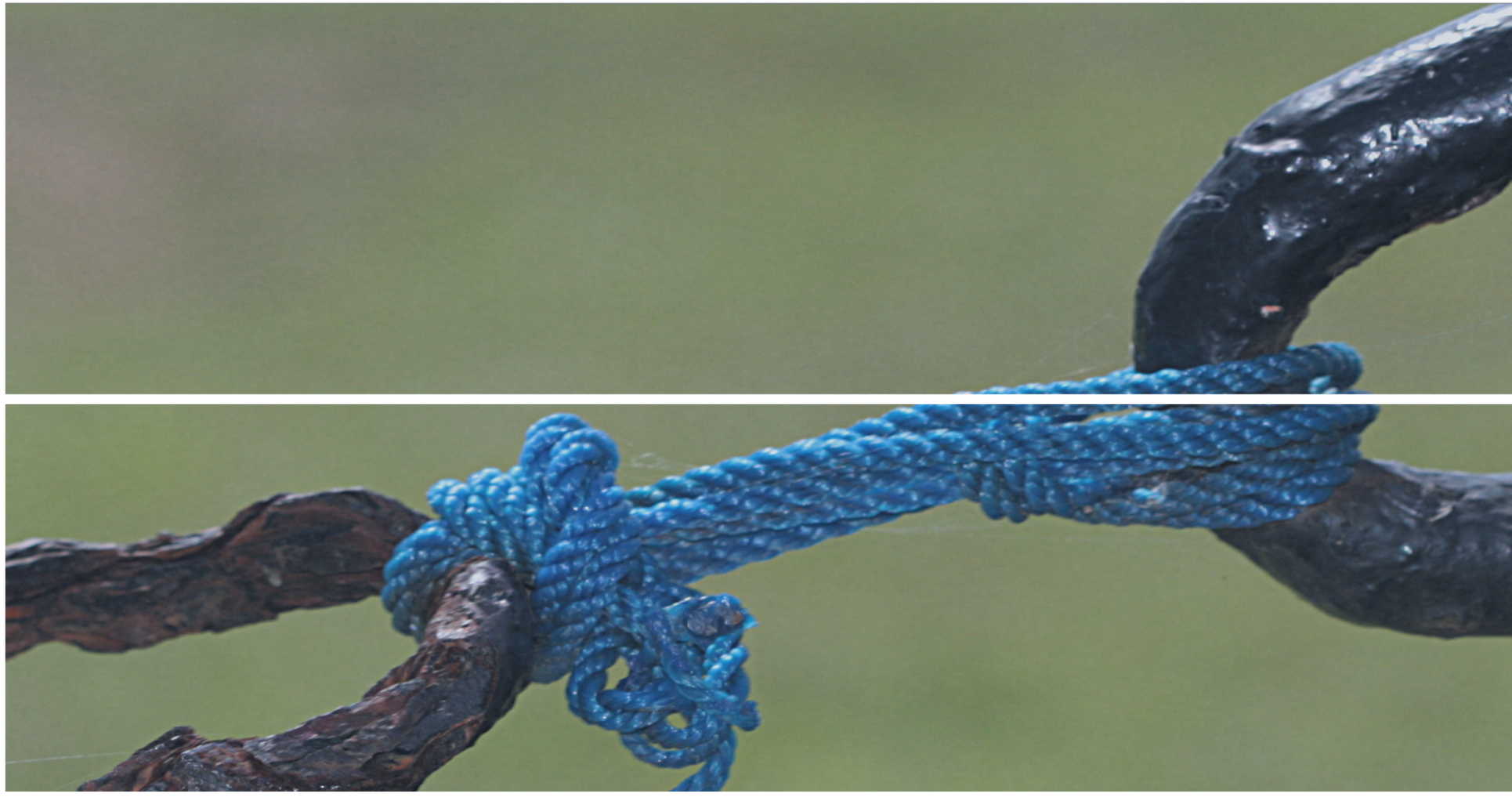
# Linkability Metric



# USER LINKABILITY

- Sessions und Sessiondauer
- Linkability Metric
- Ergebnisse der Klassifikation

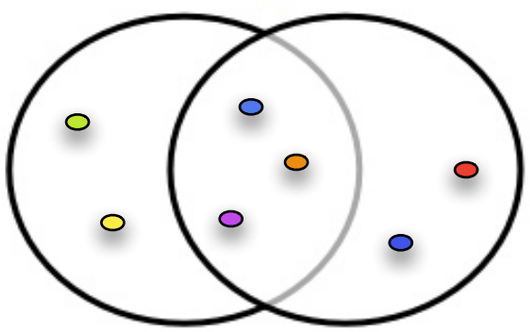




# Wiedererkennungsraten

## Jaccard-Klassifikation

Schnittmenge von zwei Benutzer-Sessions



70%

## Bayes-Klassifikation

Wahrscheinlichkeit, dass eine unbekannte Surfsession von einem best. Benutzer stammt

$$P(c_i|d) \sim P(c_i) \cdot C_{multi} \cdot \prod_{t \in V} P(t|c)^{f_{t,d}}$$

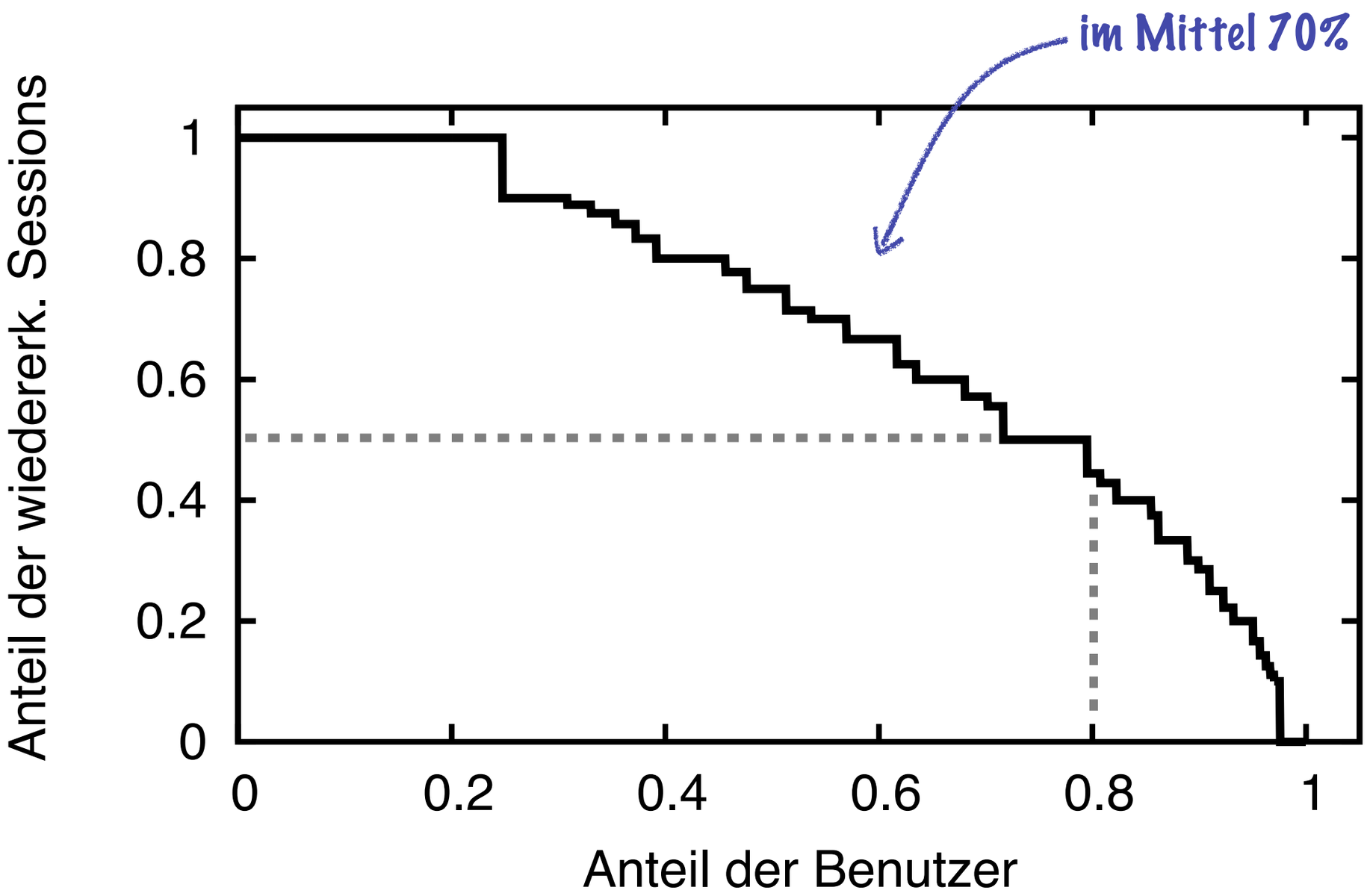
$$\hat{P}(t|c) = \frac{f_{t,c}}{\sum_{t' \in V} f_{t',c}}$$

73%

# USER LINKABILITY BEI DNS

- Studie mit DNS-Requests
- DNS-Log-File der Uni Regensburg
- 3100 Benutzer

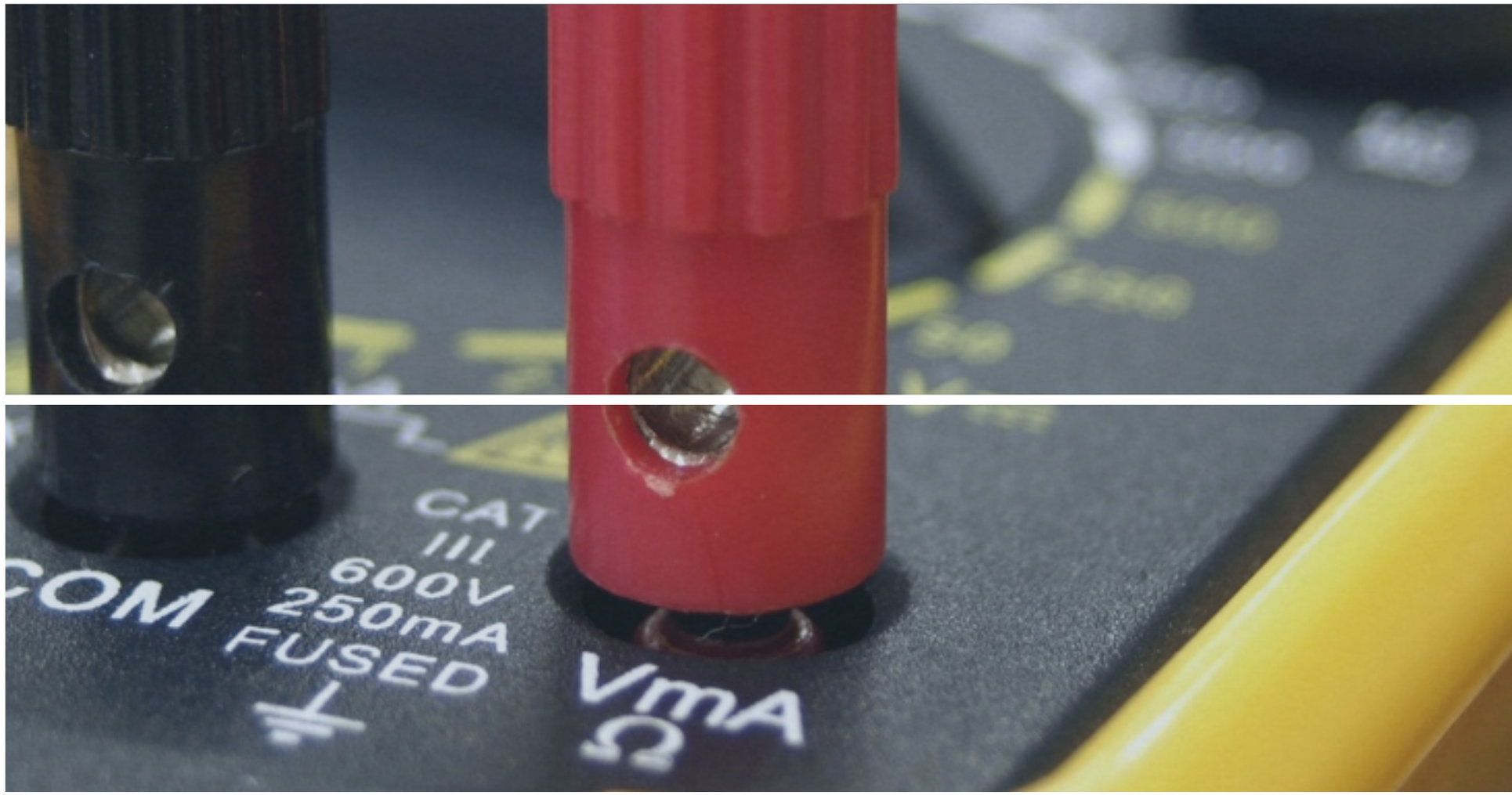
# Wiedererkennungsrate bei DNS




KONSEQUENZEN
FÜR DIE PRAXIS
MAX
10A
FUSED
COM
CAT
III
600V
250mA
FUSED
VmA
Ω

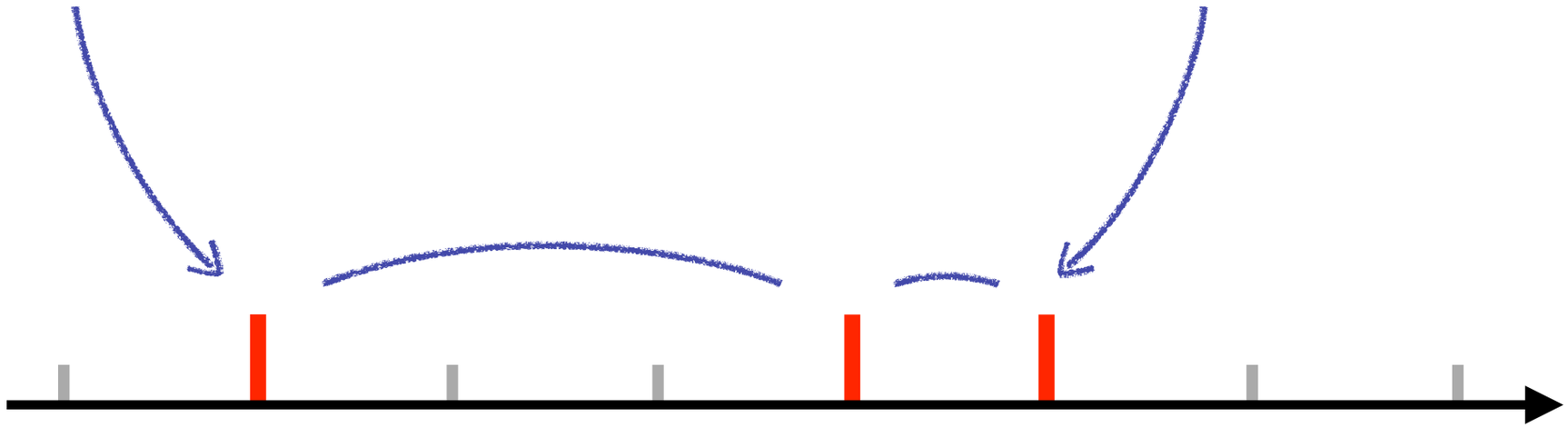
# KONSEQUENZEN FÜR DIE PRAXIS

Eingangsbeispiel



Abruf durch
Website Fingerprinting
aufgedeckt

Sessions durch
User Linkability
verkettet



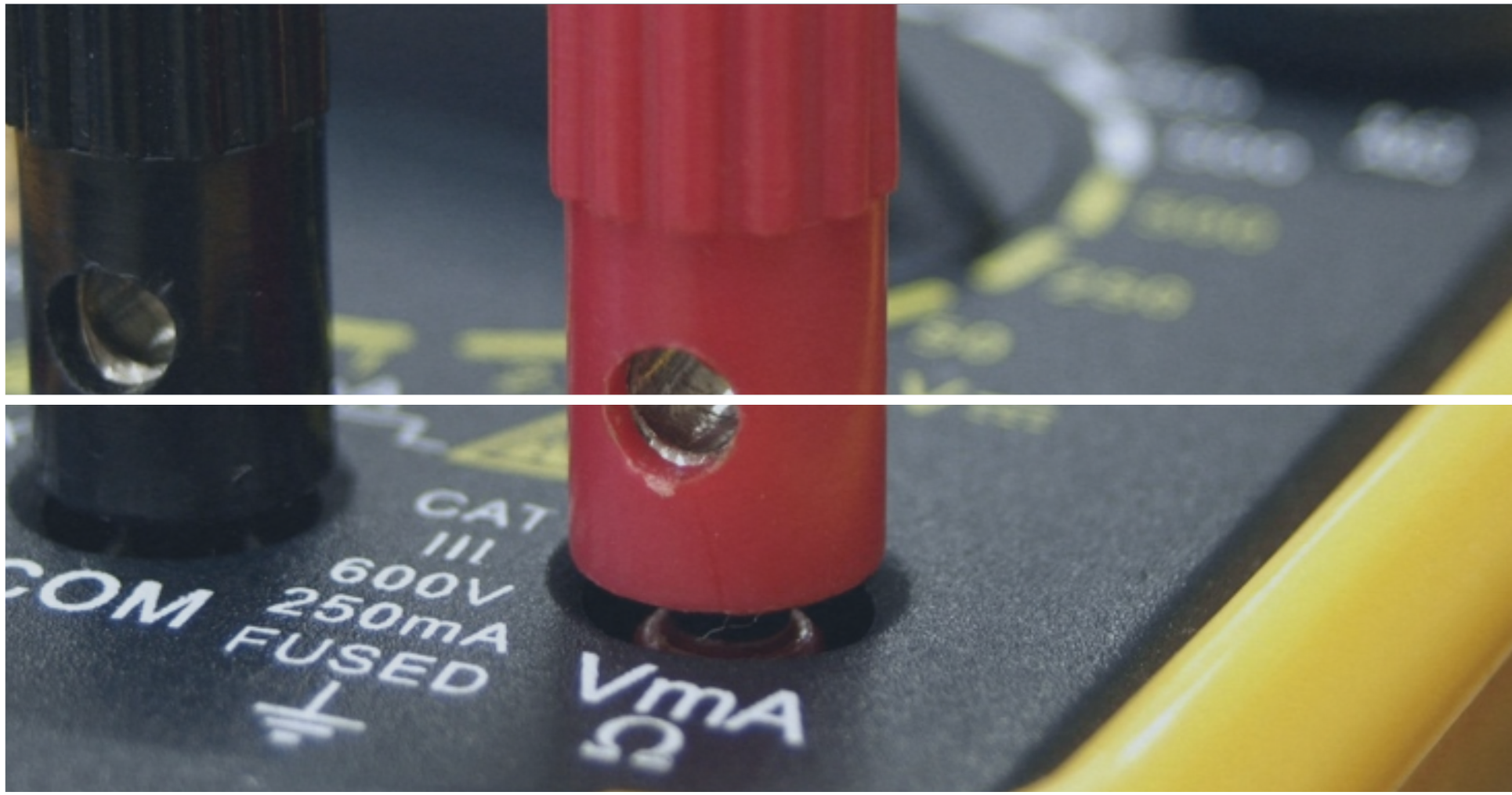

facebook

Identität online
preisgegeben

# KONSEQUENZEN FÜR DIE PRAXIS

Eingangsbeispiel

reale Bedrohung?

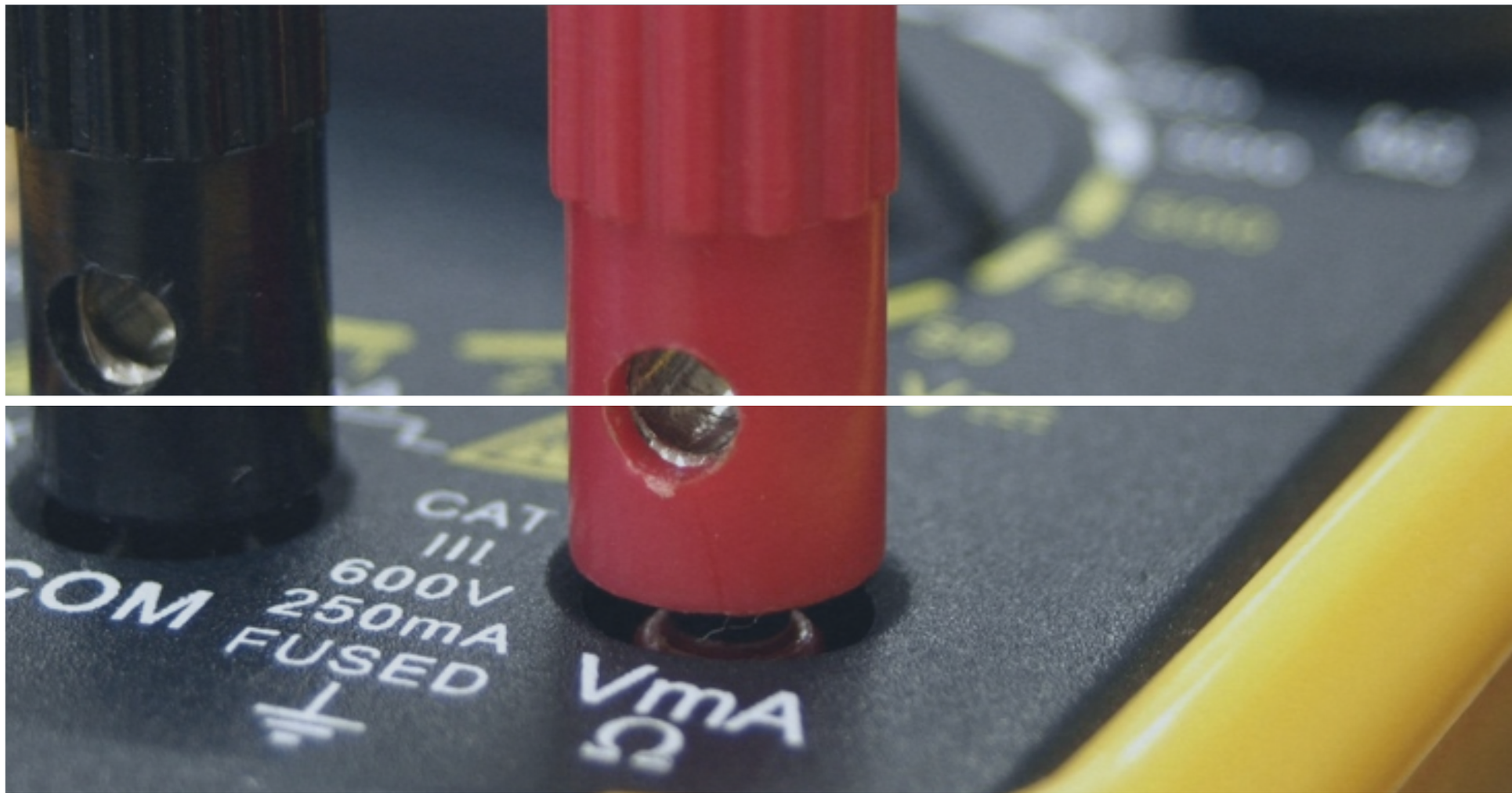


# KONSEQUENZEN FÜR DIE PRAXIS

Eingangsbeispiel

reale Bedrohung?

Abwehrmaßnahmen

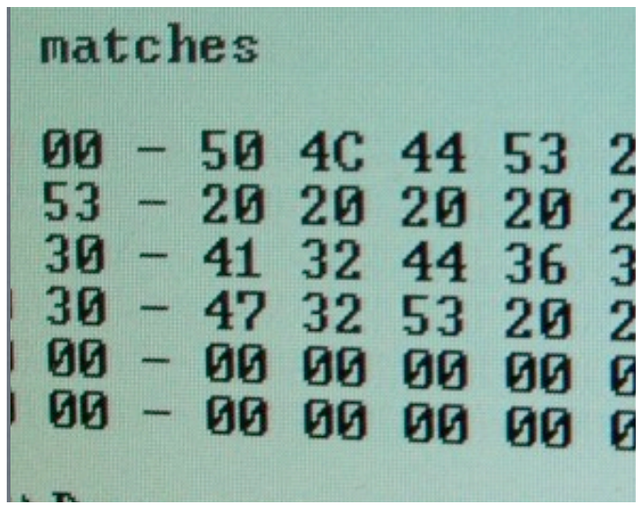


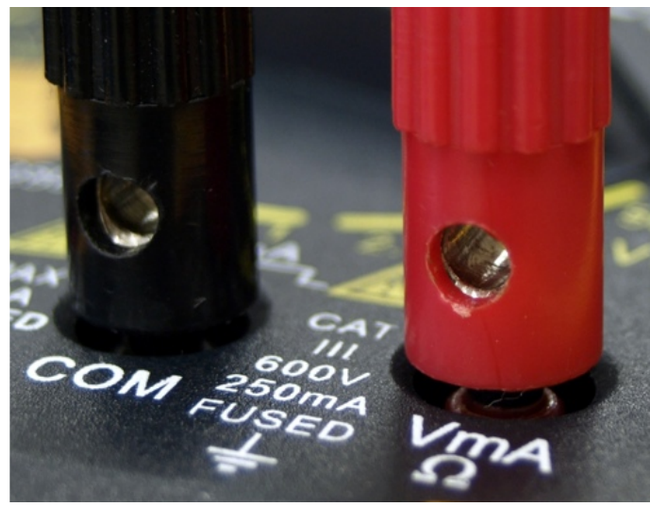
## Zusammenfassung

**Traffic-Analyse:** Auswertung von übermittelten Nachrichten, um auf Inhalte und beteiligte Kommunikationspartner zu schließen

**Website-Fingerprinting:** Ermittlung der URLs der Webseiten, die über einen verschlüsselten Kanal übertragen werden

**User Linkability:** Ausnutzen der charakteristischen Interessen von Benutzern, um mehrere Sitzungen anhand der darin abgerufenen Webseiten zu verketten




**Dominik Herrmann und  Christoph Gerber**

**Universität Regensburg**

Lehrstuhl Management der Informationssicherheit

http://www-sec.uni-regensburg.de