



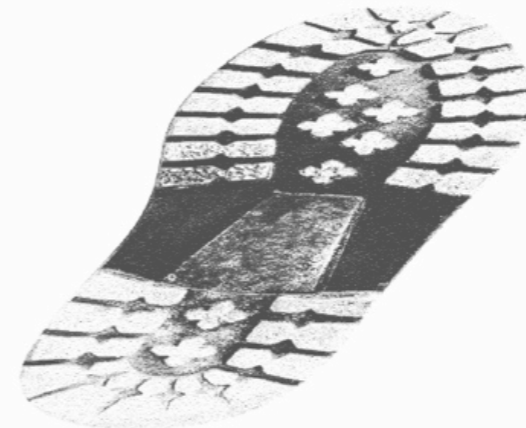
ANGRIFFE AUF DIE PRIVATSPHÄRE DURCH TRAFFIC-ANALYSEN IM INTERNET

Dominik Herrmann und Christoph Gerber

Universität Regensburg

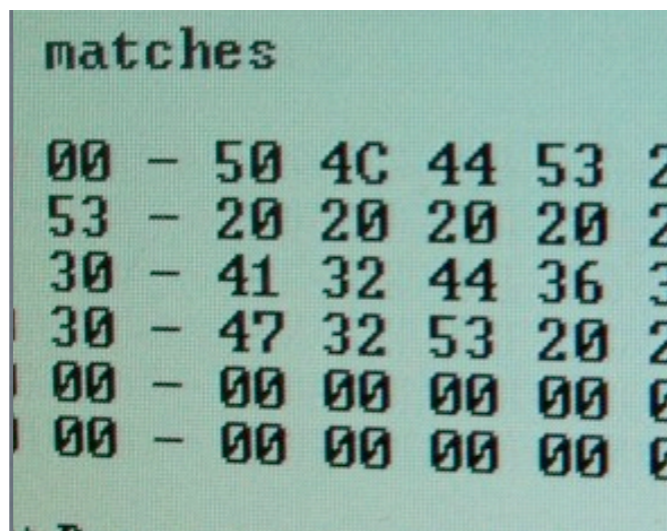
LSt Management der Informationssicherheit

Prof. Dr. Hannes Federrath





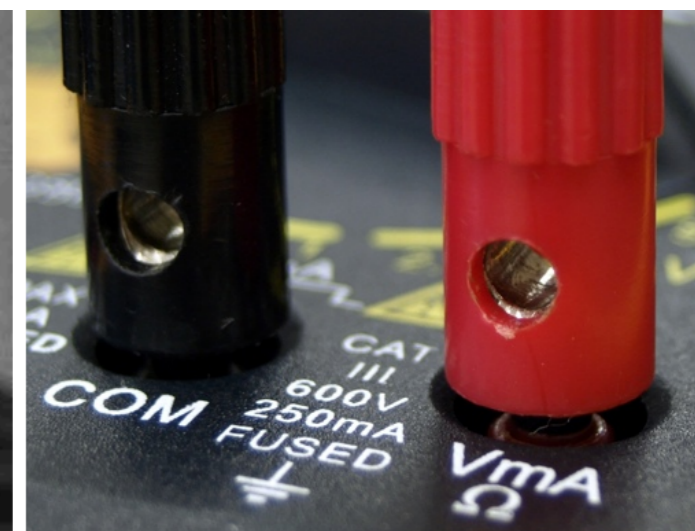
**MOTIVATION
UND SZENARIO**



**WEBSITE
FINGERPRINTING**



**USER
LINKABILITY**



**KONSEQUENZEN
FÜR DIE PRAXIS**

Ziel: Sensibilisierung



MOTIVATION UND SZENARIO

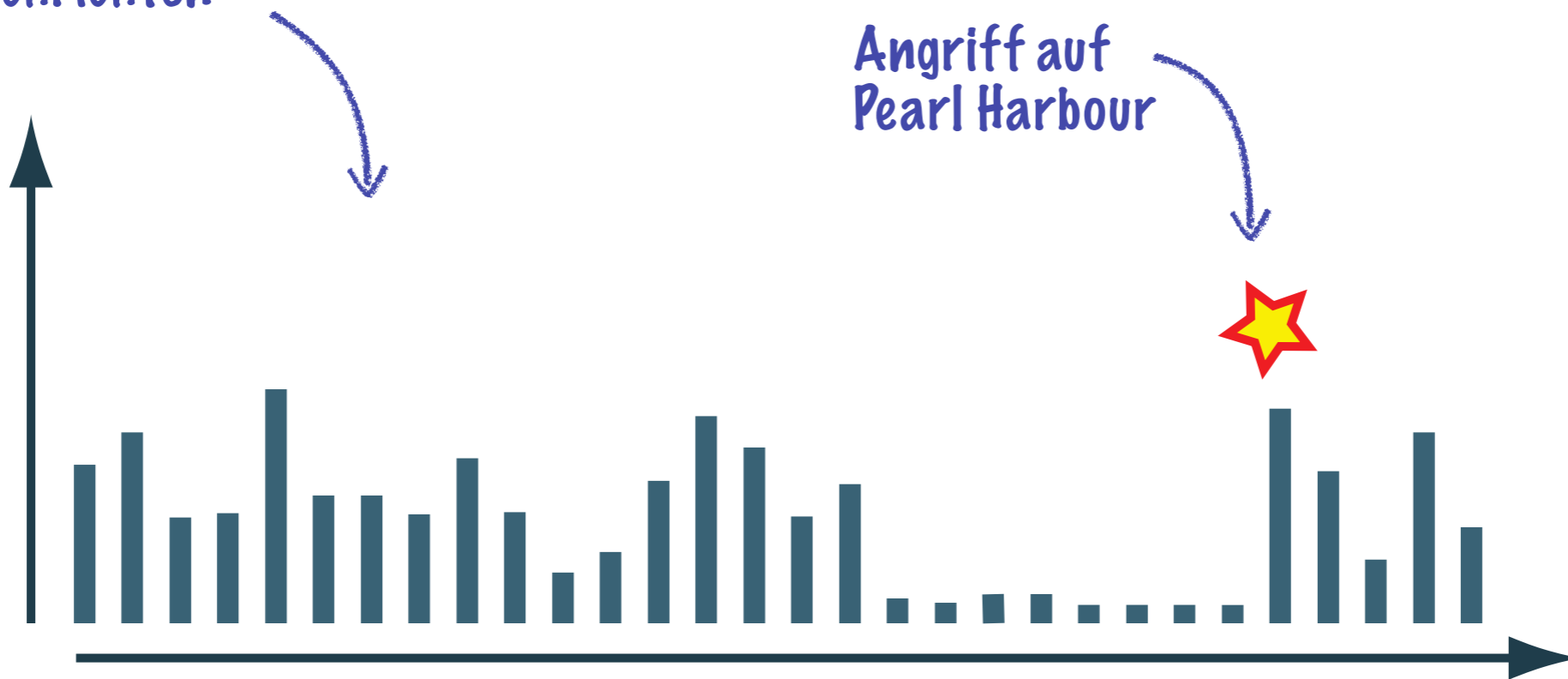
TRAFFIC-ANALYSE

Beobachtung über-
mittelter Nachrichten
Rückschluss auf Inhalt
Ursprung: Militär

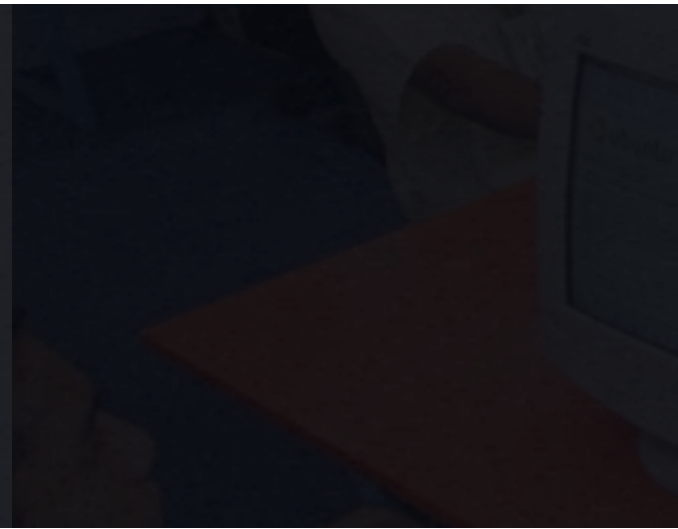
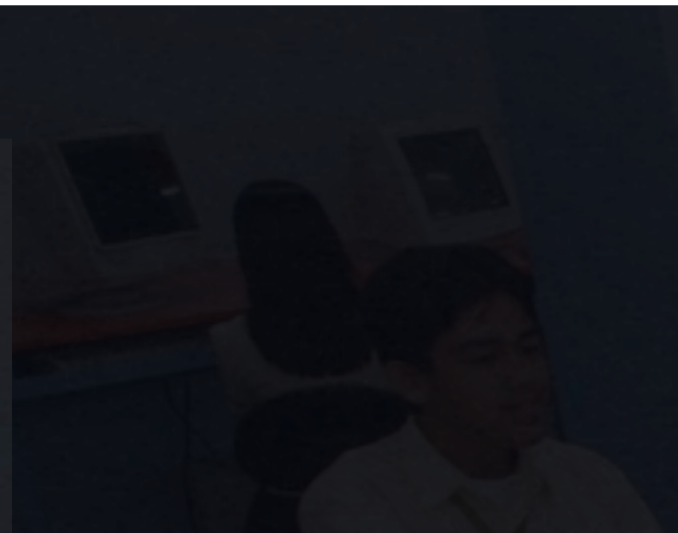


Anzahl der
verschlüsselten
Nachrichten

Angriff auf
Pearl Harbour



TRAFFIC-ANALYSE IM INTERNET



SZENARIO

Totalitäres Regime

Journalist

Skandal aufdecken

Identität schützen



WikiLeaks Upload

https://secure.wikileaks.org/

WikiLeaks accepts **classified, censored** or otherwise **restricted** material of **political, diplomatic or ethical significance**.
WikiLeaks **does not accept** rumour, opinion or other kinds of first hand reporting or material that is already publicly available.

[Read the full disclaimer here.](#)

If your submission matches this criteria we will publish and keep published the document you submitted. The information you submit will be technically anonymized and we do not retain any information on you. We will never cooperate with anyone seeking to identify you.

Please choose a file for upload: Keine Dat...usgewählt

To upload multiple files please compress them as a file archive.
Please split files larger than 200MB into smaller files. Thanks.

To explicitly set an embargo date for the upload uncheck the checkbox and enter the desired release date. Please enter the date in the format YYYY/MM/DD.

No embargo, defaults to on:

The upload will not be released until: 2010/ 5 / 17

Since it seems that you have no JavaScript enabled you can see the progress of your upload if you click on the link. This will open the progress indicator in a new window. [CLICK HERE](#) the get the upload progress indicator.

To submit the document press the **Upload** button. After your upload is finished you can provide additional information about the content.

Anonymer Upload bei Wikileaks

VORSICHTSMASSNAHMEN

Wikileaks

Internet-Café

Dummy-Traffic

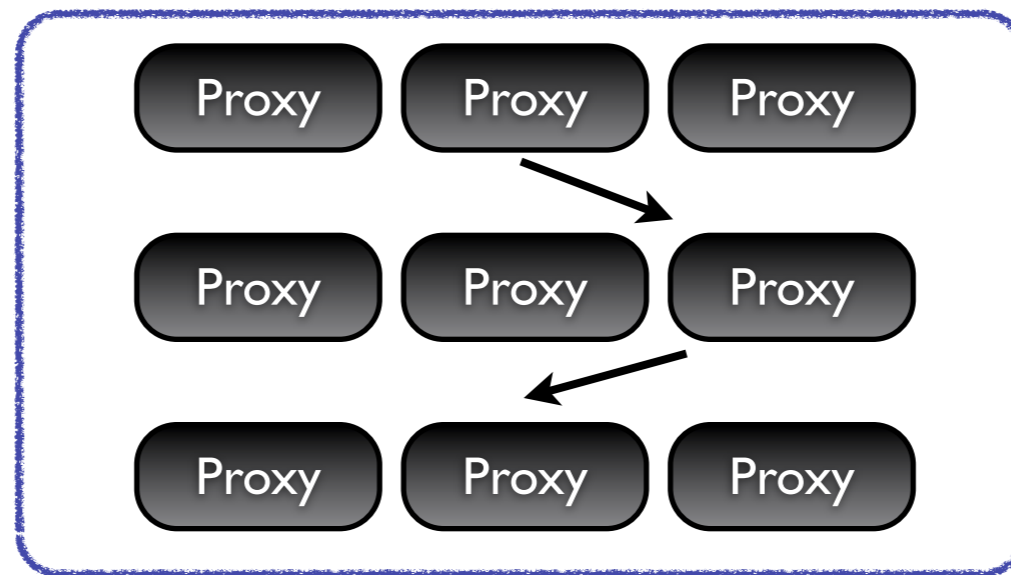
Anonymisierungsdienst



Anonymisierungsdienst

Client

Verschlüsselter
Datenverkehr



Web-Server

JOURNALIST WIRD TROTZDEM ENTTARNT

Wie?



JOURNALIST WIRD TROTZDEM ENTTARNT

Wie?

Durch Traffic-Analyse!



METHODEN

Statistische Analyse

Mustererkennung

Data-Mining

Machine-Learning

$$\int \vec{v} \cdot d\vec{v} = m \cdot R_s$$

$$(v_2) - m \cdot R_s \cdot \ln(v_1)$$

METHODEN

Text Mining

vgl. Spam-Filter

Analyse der
Worthäufigkeiten

* 60% Off All Laundry Detergent Sales & In-Store
* Free Consultation with a Debt Professional
* To: fdhgf.sss
* Pharmacy Discount for fdhgf.com

* Re candy
* Superstar stock report
* Fw:



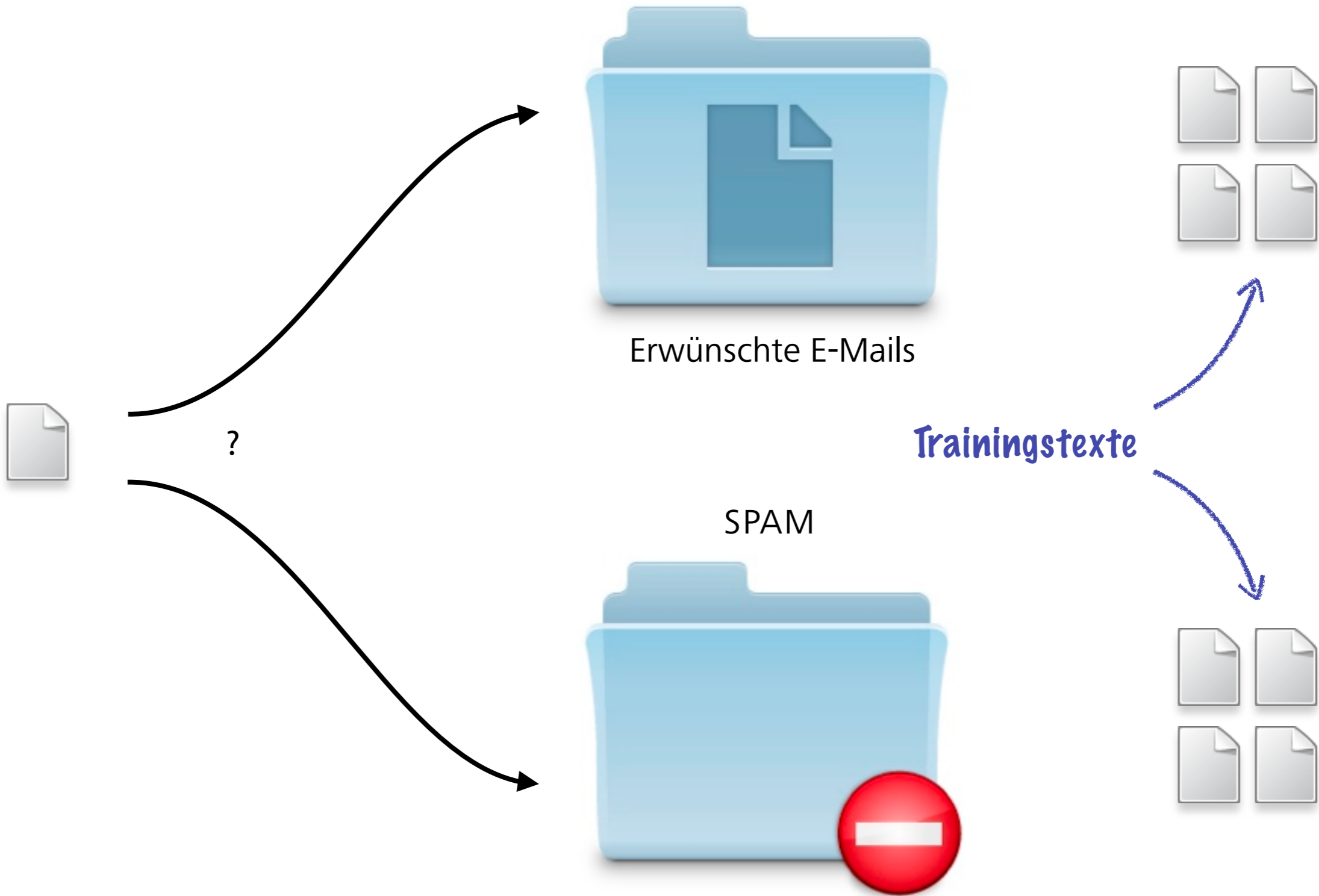
?



Erwünschte E-Mails

SPAM





extract no.2

matches

WEBSITE FINGERPRINTING

32 5B 00 00
31 36 44 32
35 30 43 41

00 - 50 40 44 53 20 20 20 20
53 - 20 20 20 20 20 20 20 20
30 - 41 32 44 36 30 30 43 40

38 31 31 30
00 00 00 00
00 00 00 00

30 - 47 32 53 20 20 20 00 00
00 - 00 00 00 00 00 00 00 00
00 - 00 00 00 00 00 00 00 00

saved to C:

Documents and Settings

HERAUSFORDERUNG

Opfer nutzt
Krypto-Tunnel

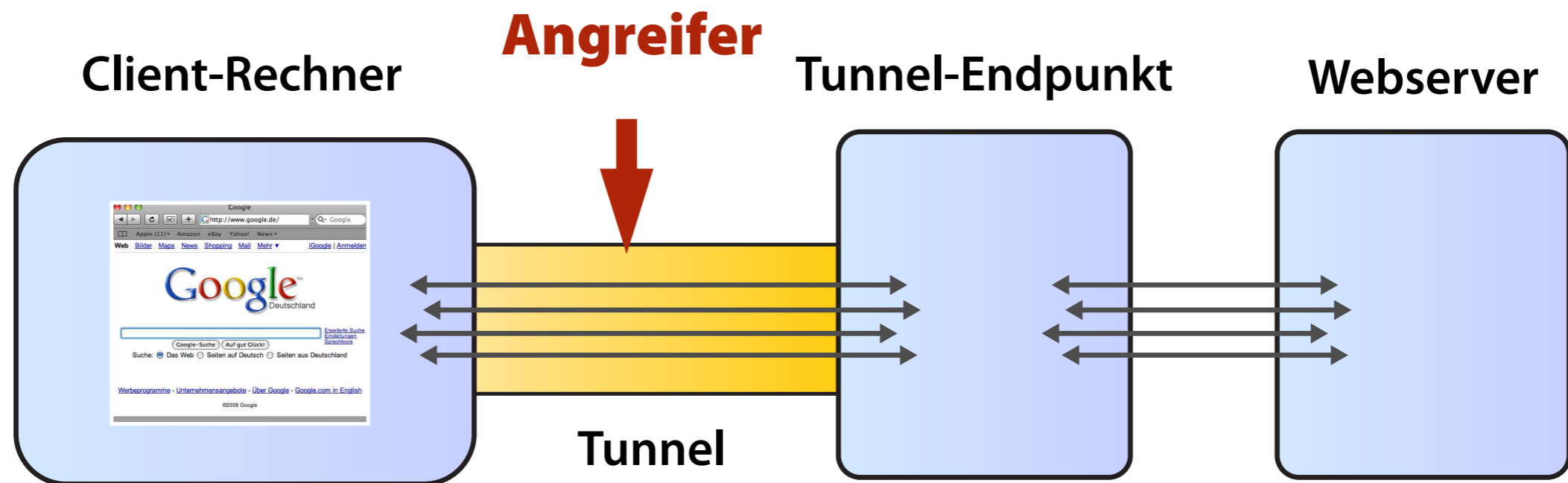
Angreifer versucht,
besuchte Webseiten
zu ermitteln

matches

| | | | | | | | | | |
|----|---|----|----|----|----|----|----|----|----|
| 00 | - | 50 | 4C | 44 | 53 | 20 | 20 | 20 | 20 |
| 53 | - | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 30 | - | 41 | 32 | 44 | 36 | 30 | 30 | 43 | 47 |
| 30 | - | 47 | 32 | 53 | 20 | 20 | 20 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |

Documents and Settings

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

matches

| | | | | | | | | | |
|----|---|----|----|----|----|----|----|----|----|
| 00 | - | 50 | 4C | 44 | 53 | 20 | 20 | 20 | 20 |
| 53 | - | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 30 | - | 41 | 32 | 44 | 36 | 30 | 30 | 43 | 47 |
| 30 | - | 47 | 32 | 53 | 20 | 20 | 20 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |



saved to C:\Documents and Settings

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

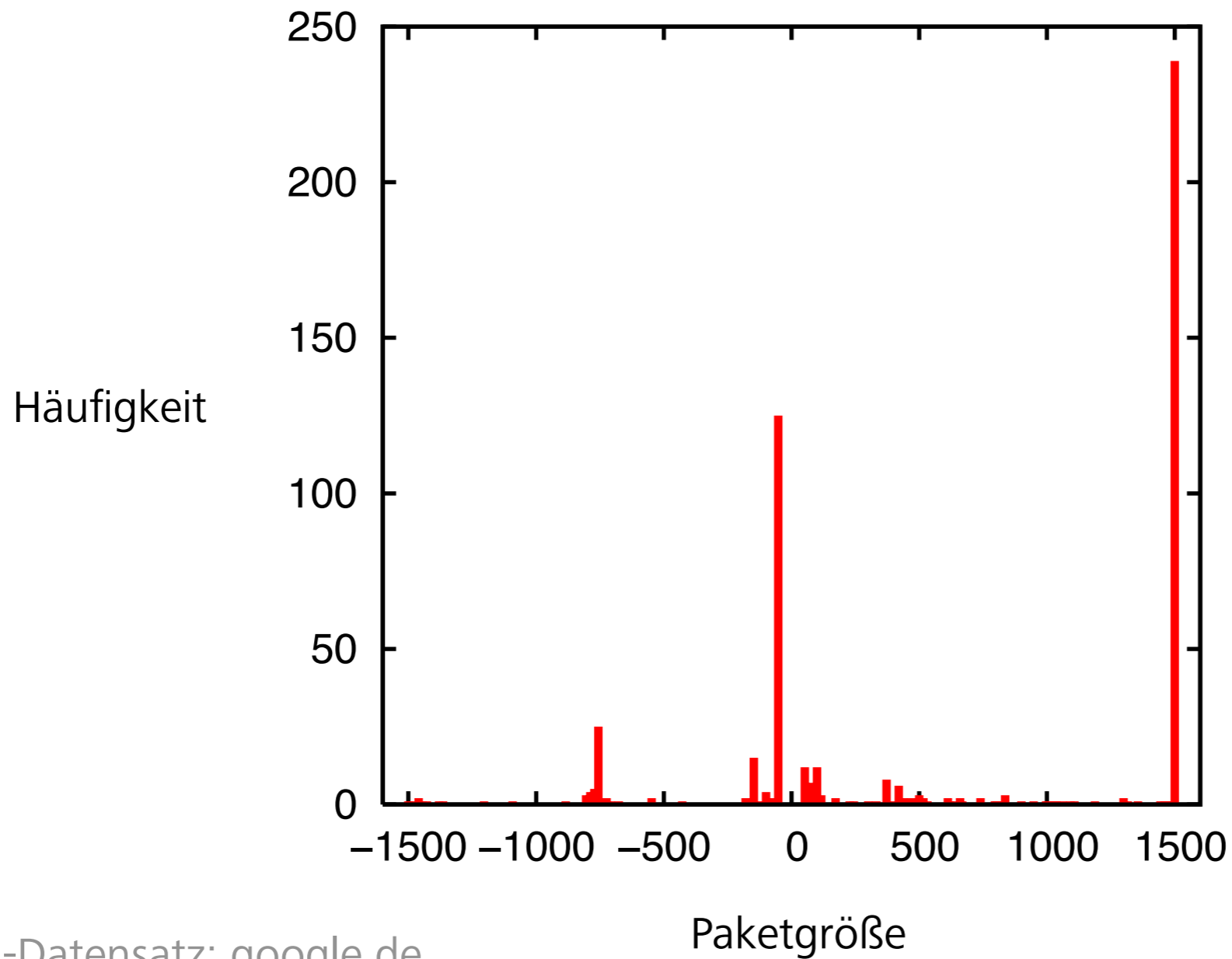
Deterministische
Fragmentierung
bei TCP/IP

matches

| | | | | | | | | | |
|----|---|----|----|----|----|----|----|----|----|
| 00 | - | 50 | 4C | 44 | 53 | 20 | 20 | 20 | 20 |
| 53 | - | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 30 | - | 41 | 32 | 44 | 36 | 30 | 30 | 43 | 47 |
| 30 | - | 47 | 32 | 53 | 20 | 20 | 20 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |

saved to C:\Documents and Settings\...

Verteilung der sichtbaren IP-Paketgrößen



DURCHFÜHRUNG DES ANGRIFFS

Phase 1 / 2

Fingerabdrücke erzeugen und abspeichern

matches

| | | | | | | | | | |
|----|---|----|----|----|----|----|----|----|----|
| 00 | - | 50 | 4C | 44 | 53 | 20 | 20 | 20 | 20 |
| 53 | - | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 30 | - | 41 | 32 | 44 | 36 | 30 | 30 | 43 | 47 |
| 30 | - | 47 | 32 | 53 | 20 | 20 | 20 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |

saved to C:\Documents and Settings

seite1.de



seite2.com



seite3.org



Trainingsdaten



DURCHFÜHRUNG DES ANGRIFFS

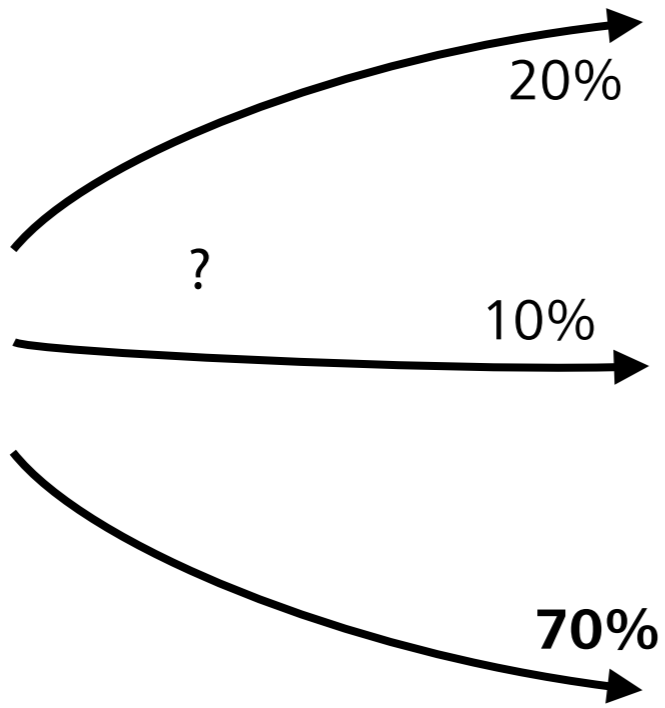
Phase 2 / 2

Im verschlüsselten
Datenverkehr des
Opfers nach Treffern
suchen

matches

| | | | | | | | | | |
|----|---|----|----|----|----|----|----|----|----|
| 00 | - | 50 | 4C | 44 | 53 | 20 | 20 | 20 | 20 |
| 53 | - | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 30 | - | 41 | 32 | 44 | 36 | 30 | 30 | 43 | 47 |
| 30 | - | 47 | 32 | 53 | 20 | 20 | 20 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 00 | - | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |

Documents and Settings



seite1.de



seite2.com



seite3.org



Ähnlichkeitswerte



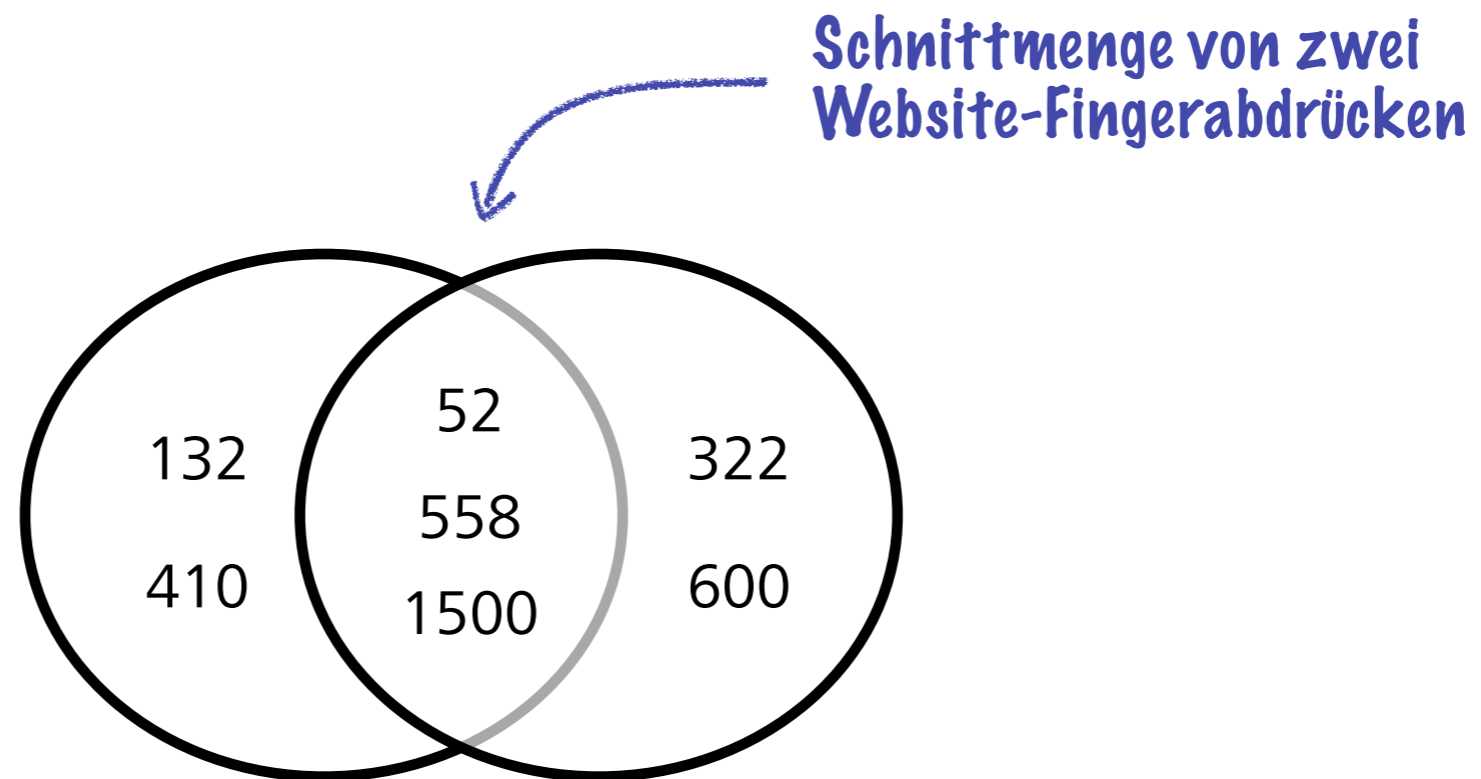
ANALYSE-METHODEN

Jaccard-Koeffizient

$$|V_2| - m \cdot R_s = |h(V_2)|$$

$$|h(V_1)| = m \cdot R_s'$$

Jaccard-Koeffizient



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

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

ANALYSE-METHODEN

Jaccard-Koeffizient

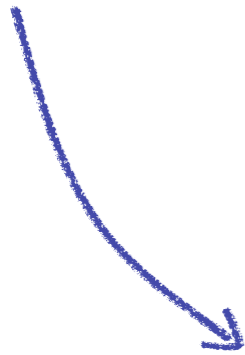
Naïve Bayes

$$\ln(V_2) - m \cdot R_s \cdot \ln(V_1)$$

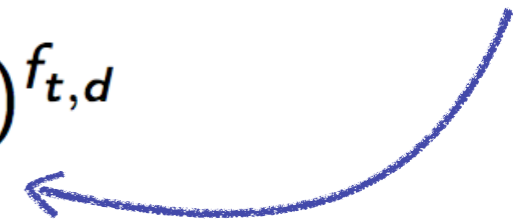
$$\ln(V_1) = m \cdot R_s$$

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.



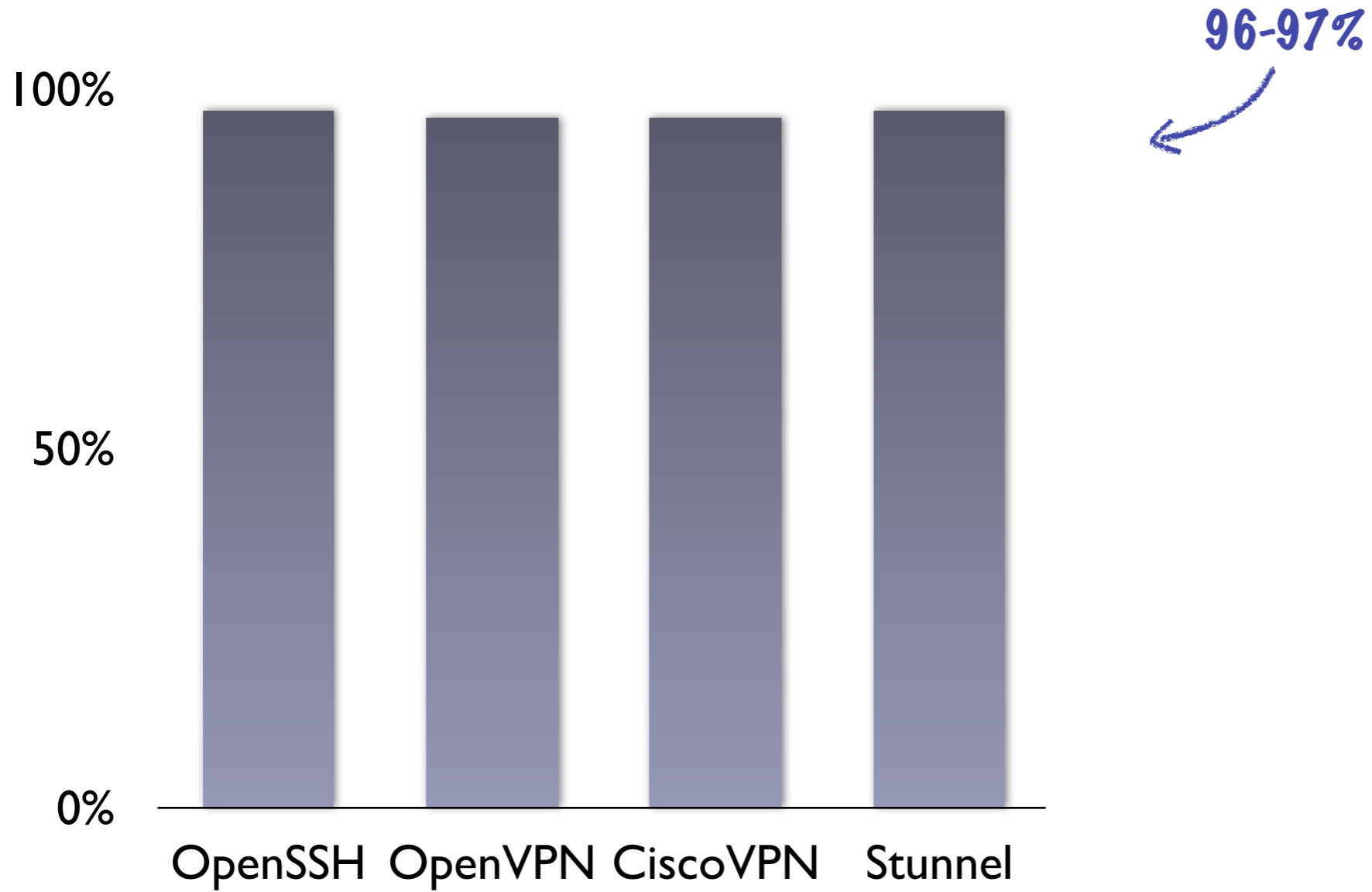
Erkennungsraten

96-97%



TODO

Erkennungsraten

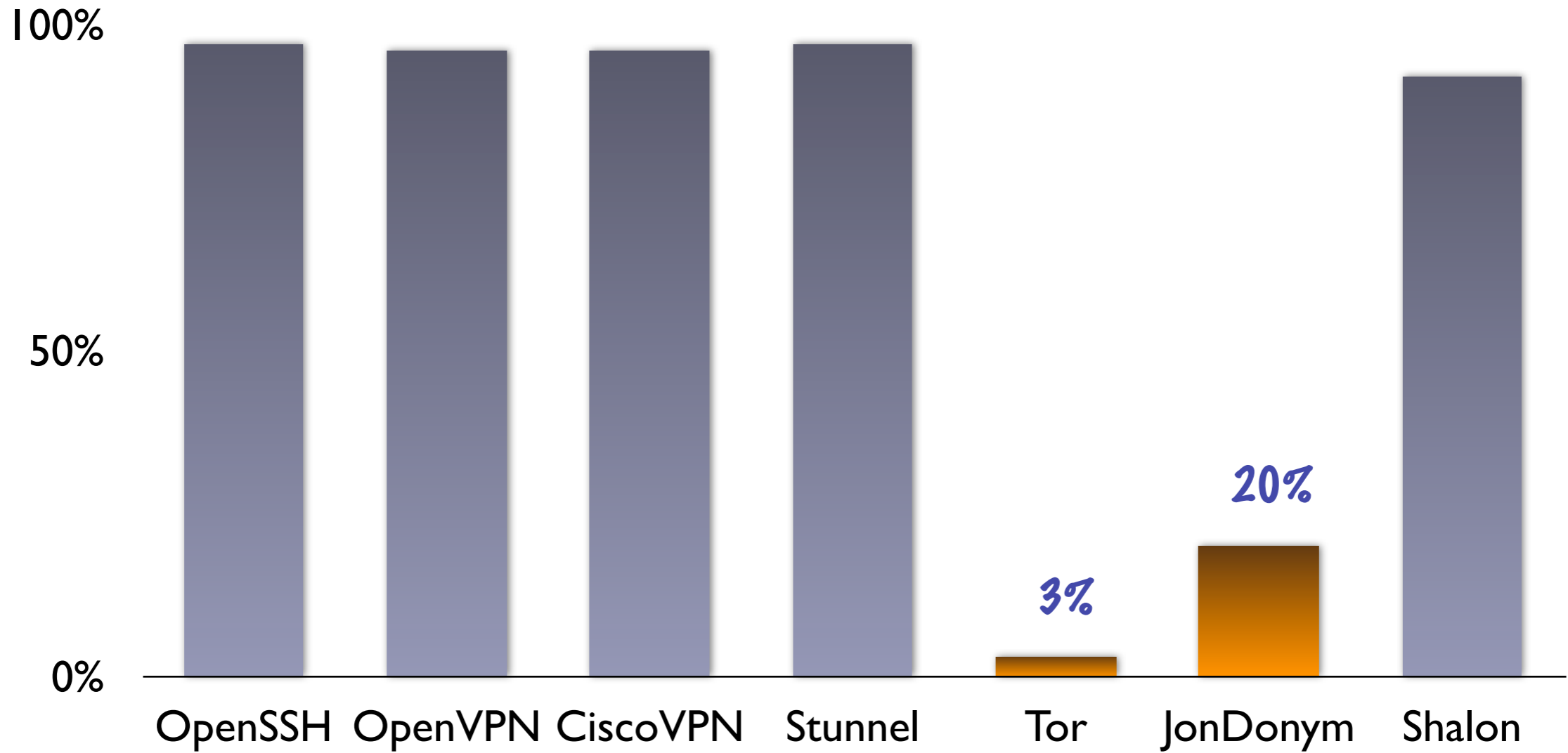


ERGEBNISSE

Anonymisierungssysteme bieten besseren Schutz
können Angriff nicht immer abwehren



Erkennungsraten



Anonymisierungsdienste



USER LINKABILITY

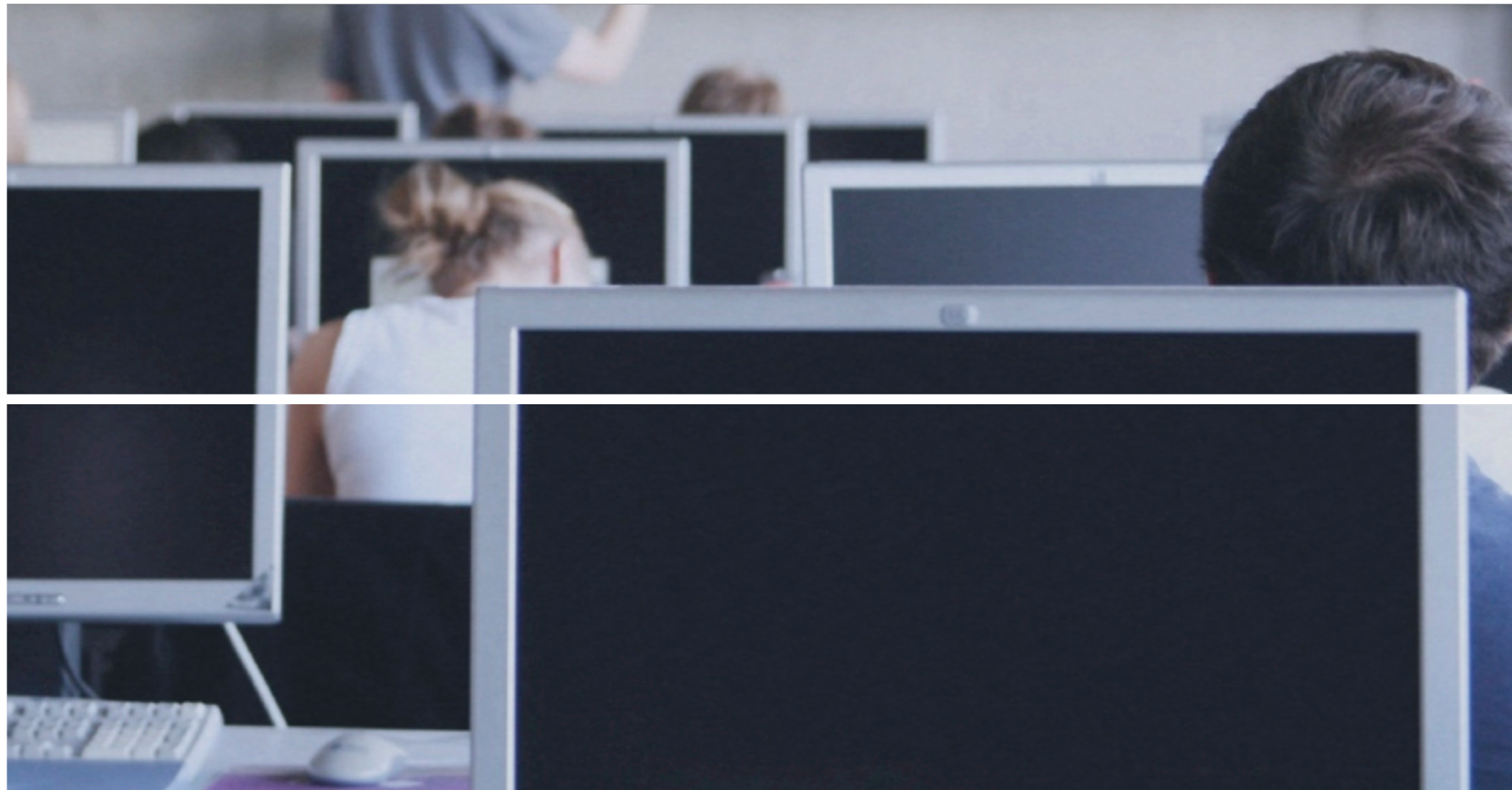


USER LINKABILITY

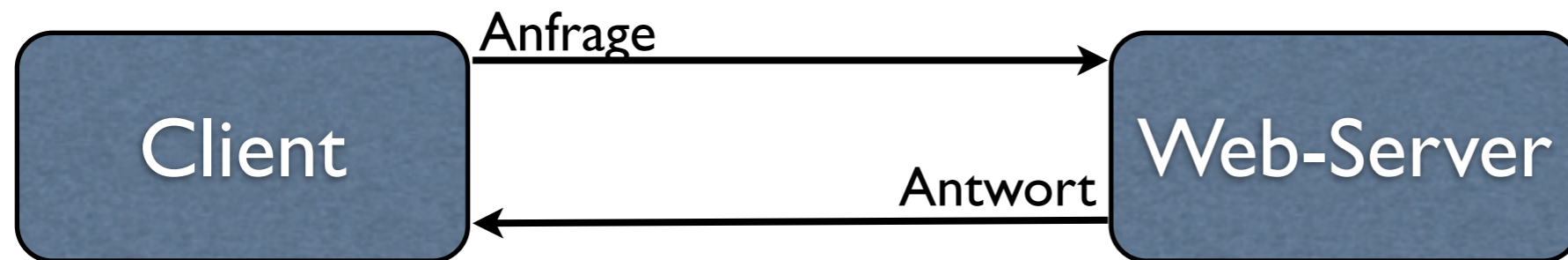
Motivation

Testpersonen

Datenquelle



HTTP-Anfrage



Request Splitting

Host

www.google.de

Suffix

[search?q=informationswissenschaft](http://www.google.de/search?q=informationswissenschaft)

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

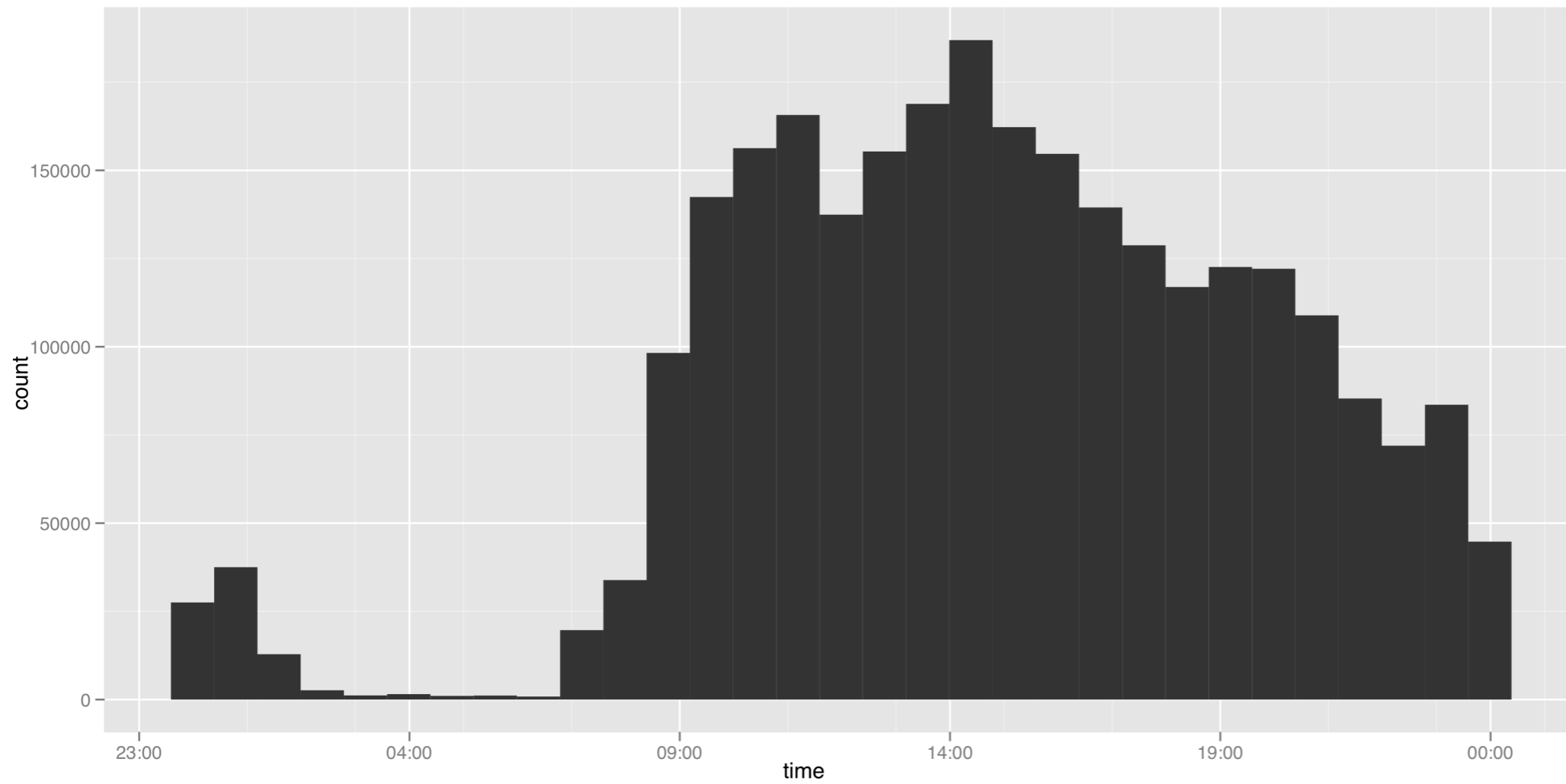
request

USER LINKABILITY

Studienteilnahme

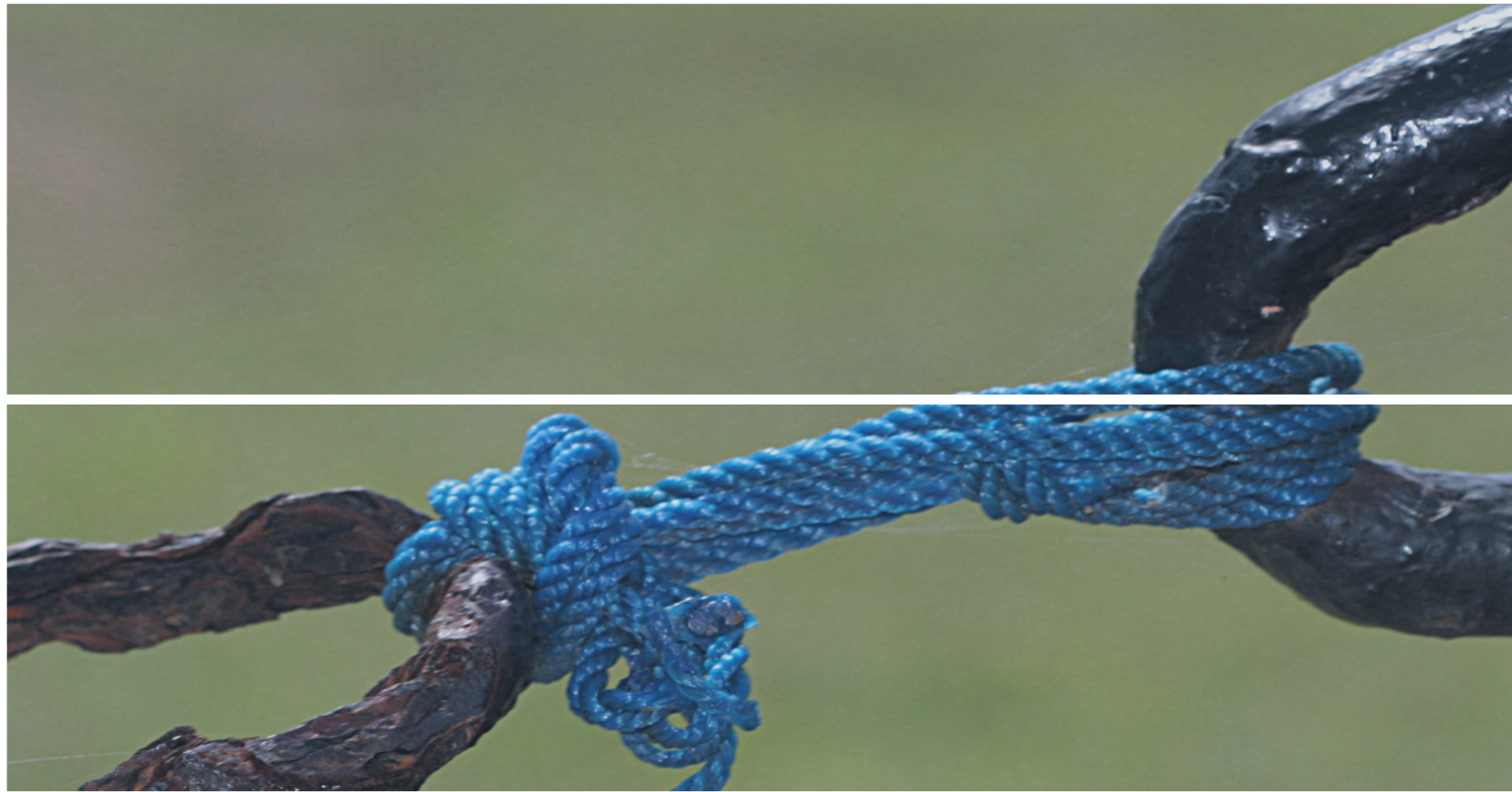


Aktivität im Tagesverlauf

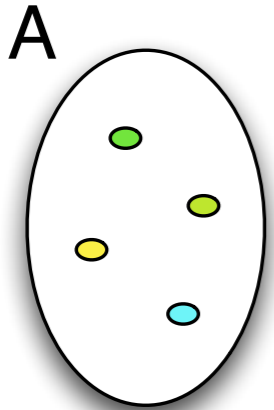


USER LINKABILITY

Sessions und
Sessiondauer

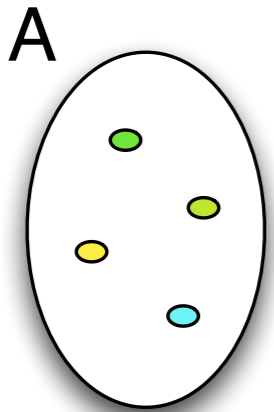


Sessions

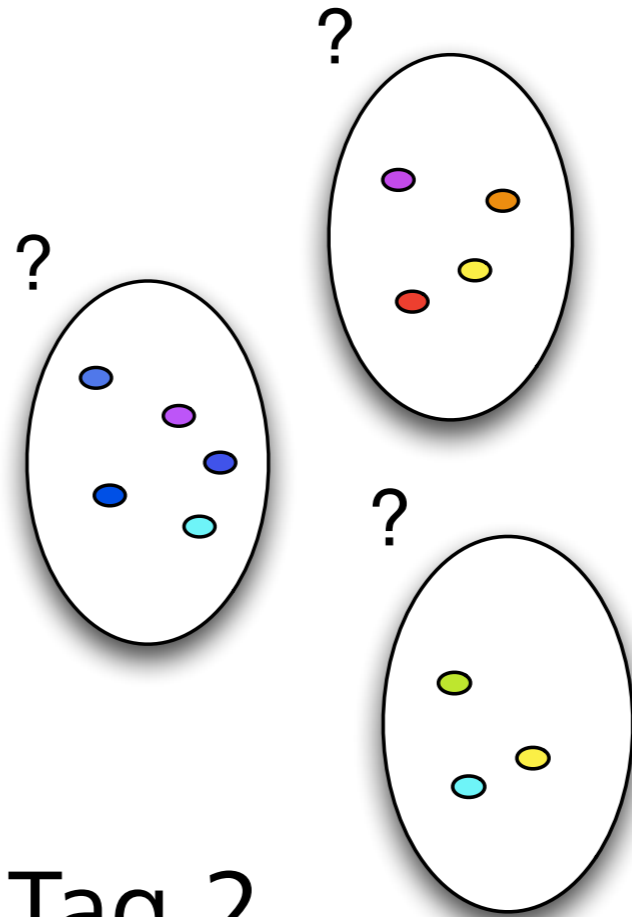


Tag 1

Sessions



Tag 1



Tag 2

USER LINKABILITY

Sessions und
Sessiondauer
Linkability Metric



Linkability Metric

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

Linkability Metric

Aufrufe eines Hosts durch einen Benutzer



Anzahl der Sessions in denen der Benutzer den Host aufgerufen hat

Aufrufe eines Hosts durch alle Benutzer

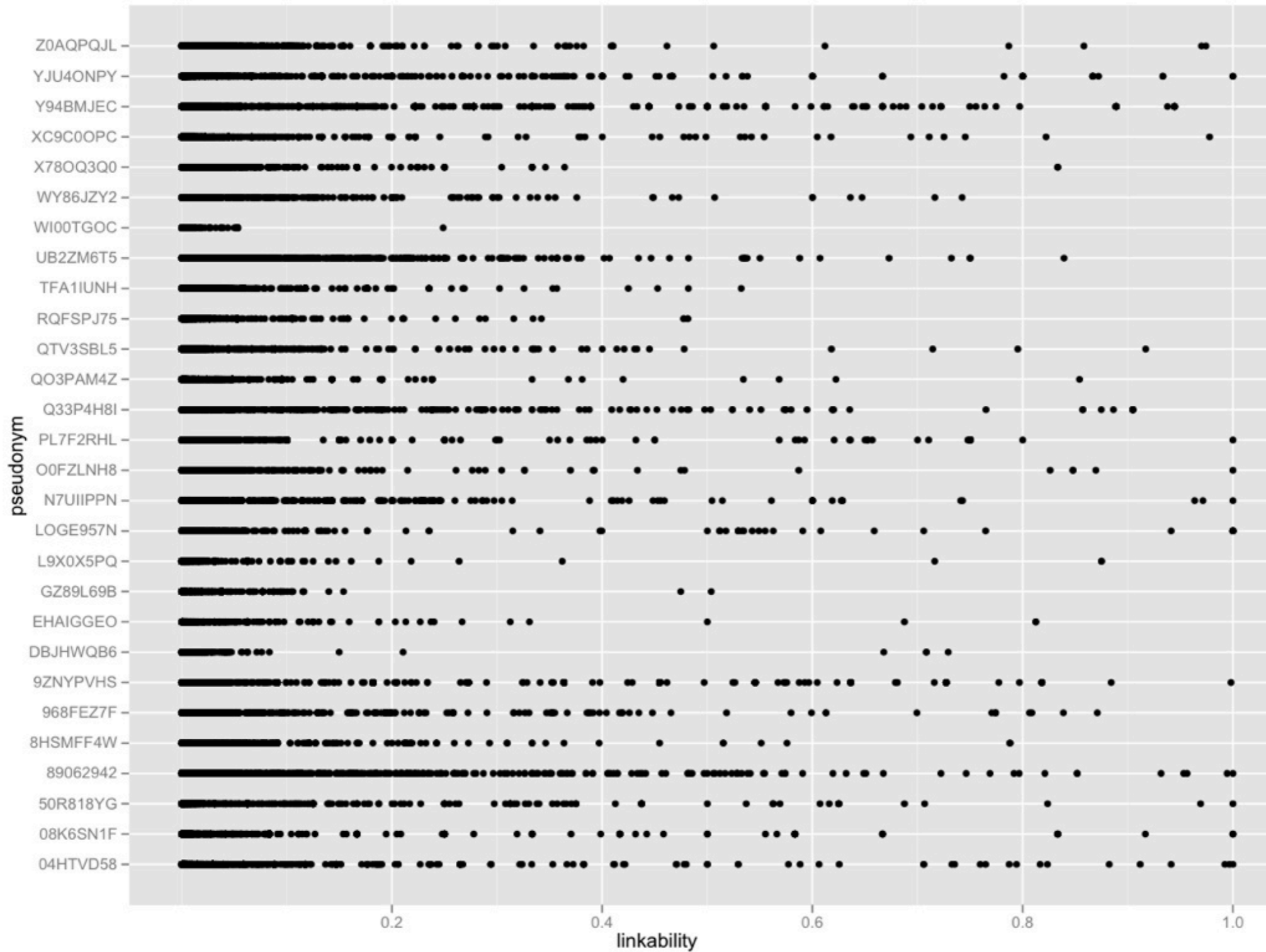
Alle Sessions des Benutzers

Wird 1 wenn der Benutzer den Host täglich aufruft

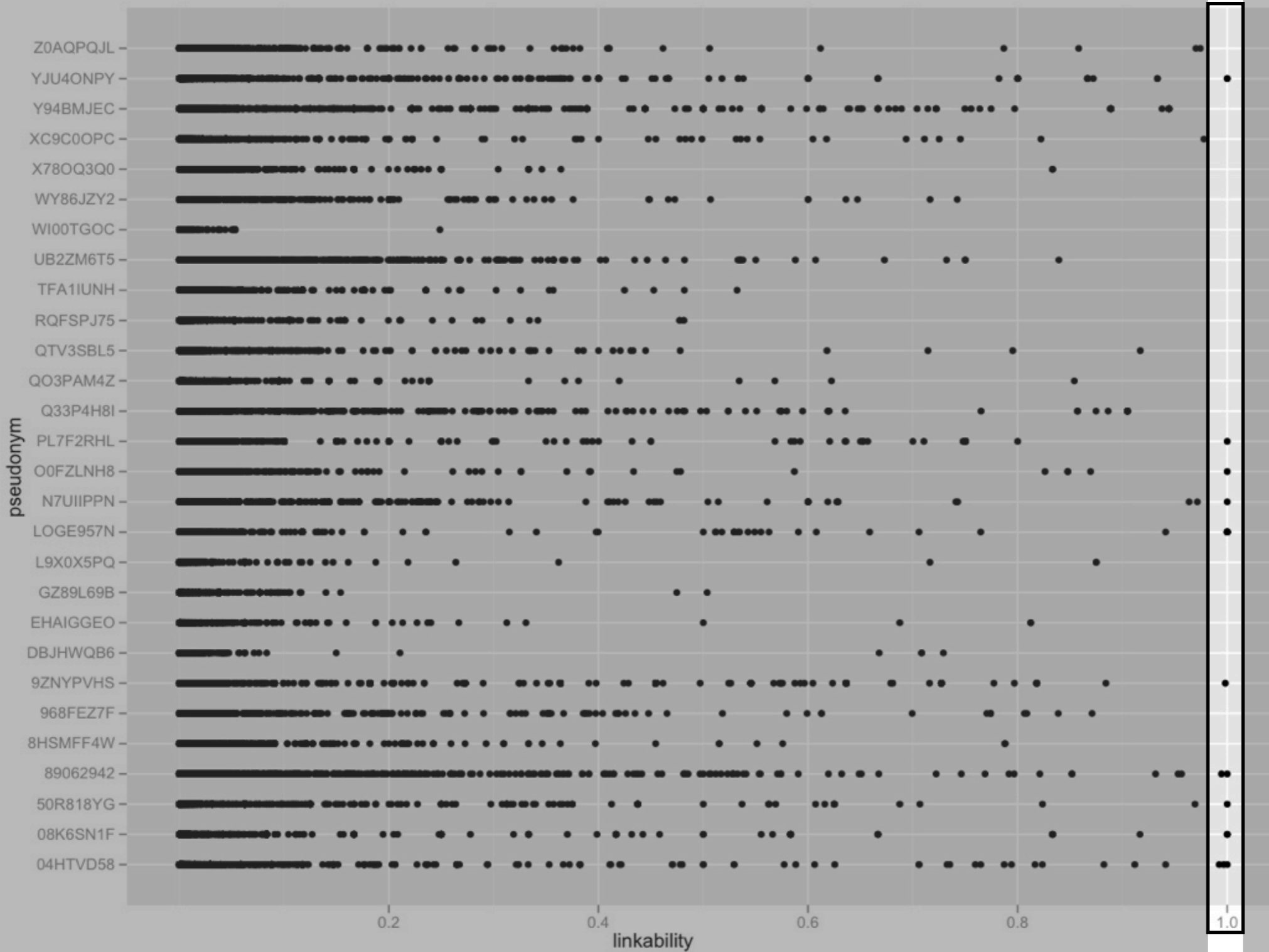
Wird 1 wenn nur der Benutzer den Host aufruft

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

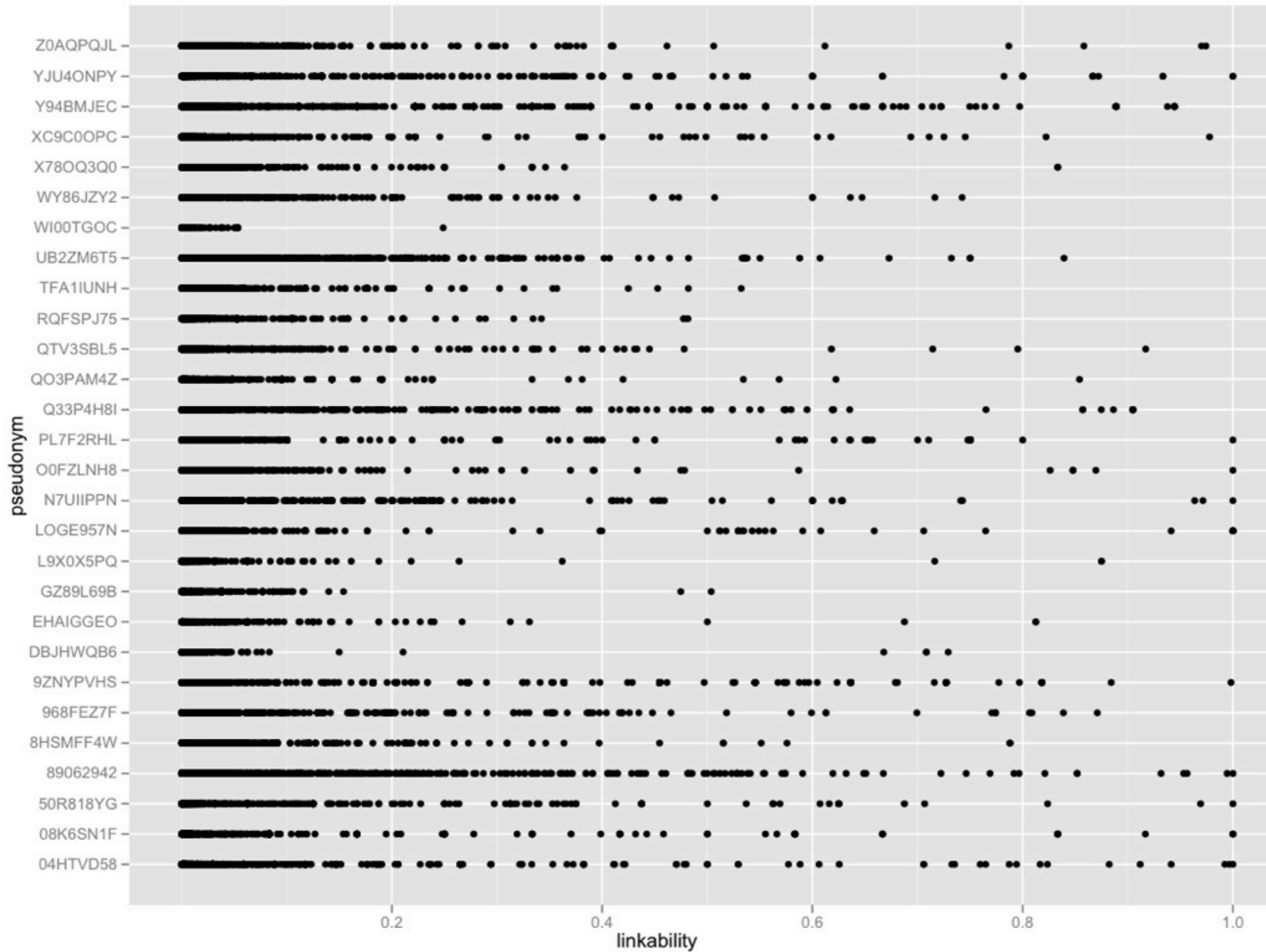
Linkability Metric



Linkability Metric



Linkability Metric

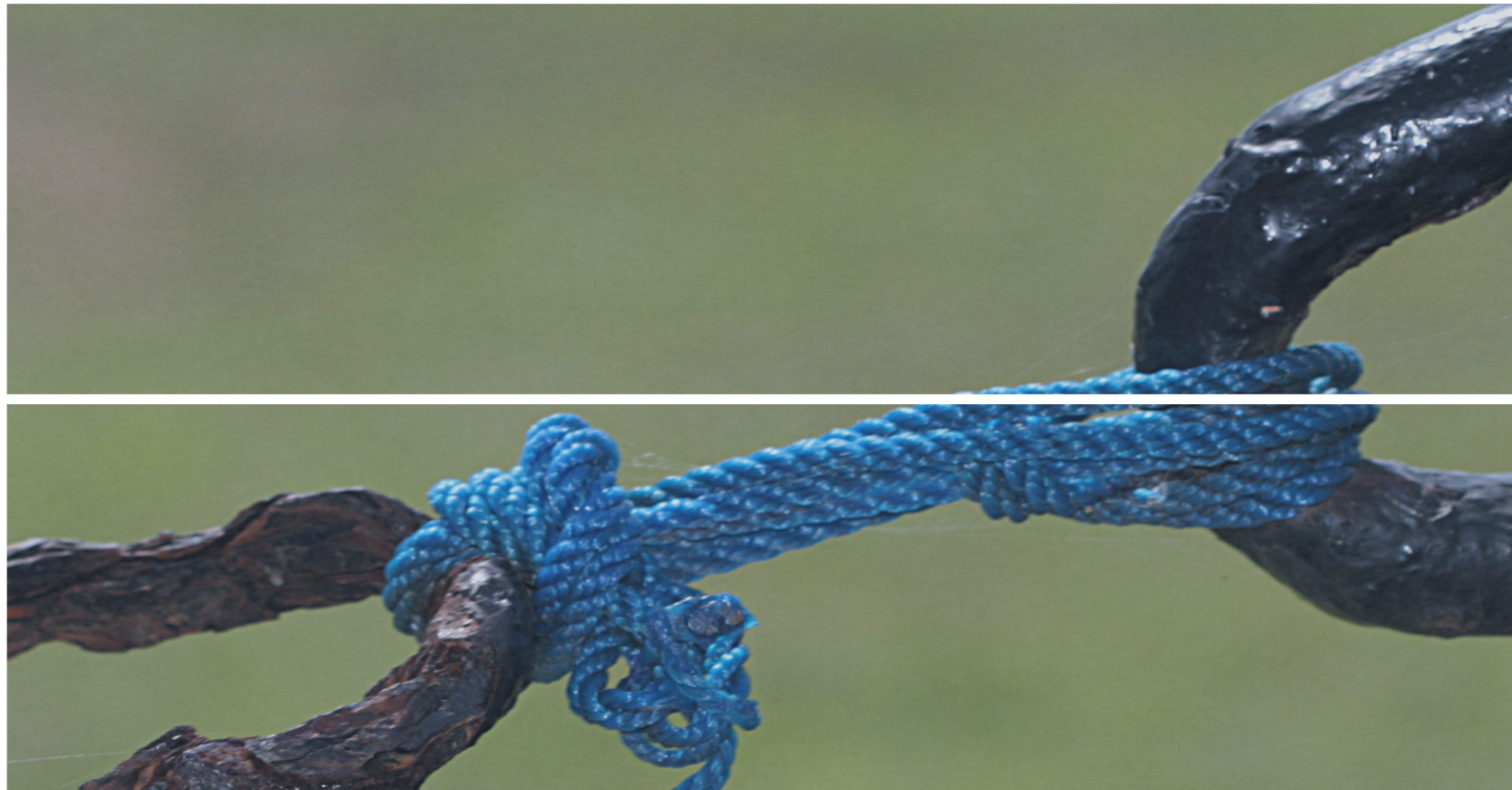


USER LINKABILITY

Sessions und
Sessiondauer

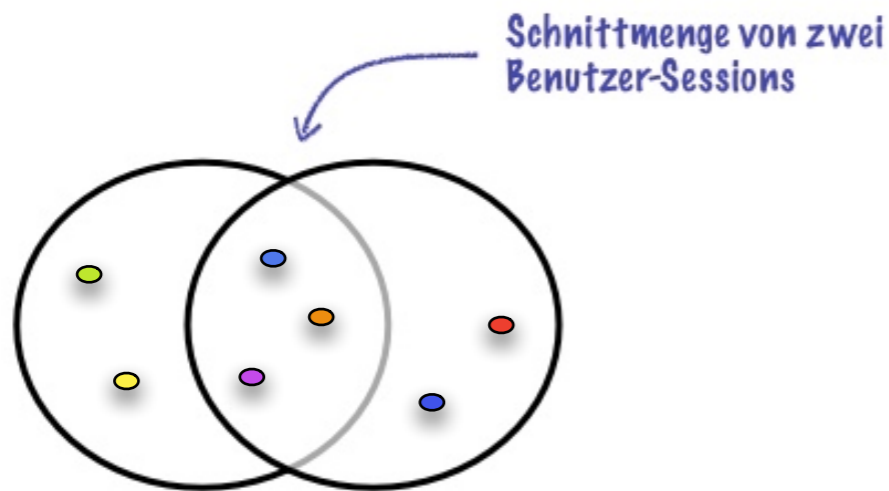
Linkability Metric

Ergebnisse der
Klassifikation



Wiedererkennungsraten

Jaccard-Klassifikation



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

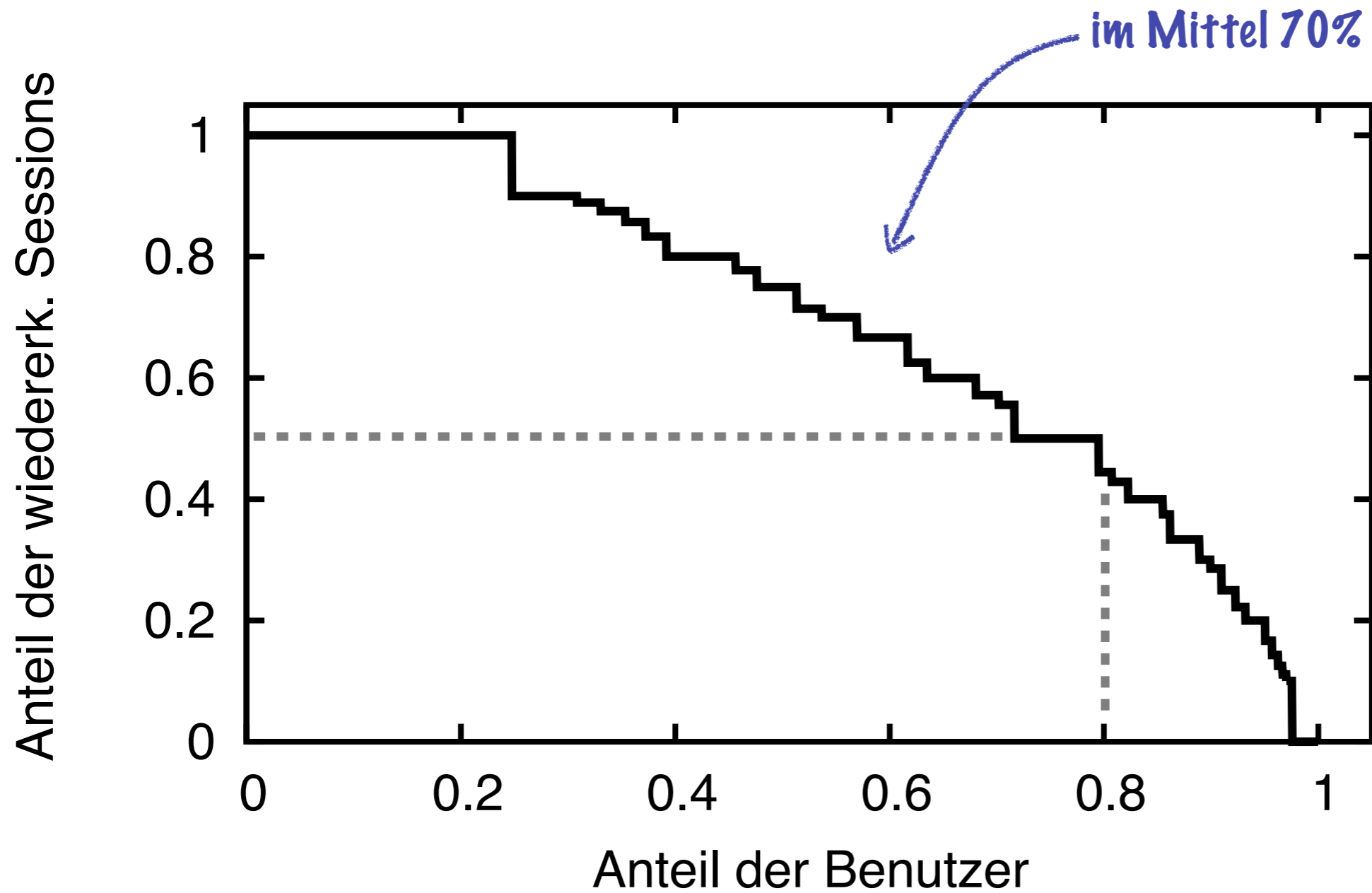
Studie mit DNS-Requests

DNS-Log-File der Uni Regensburg

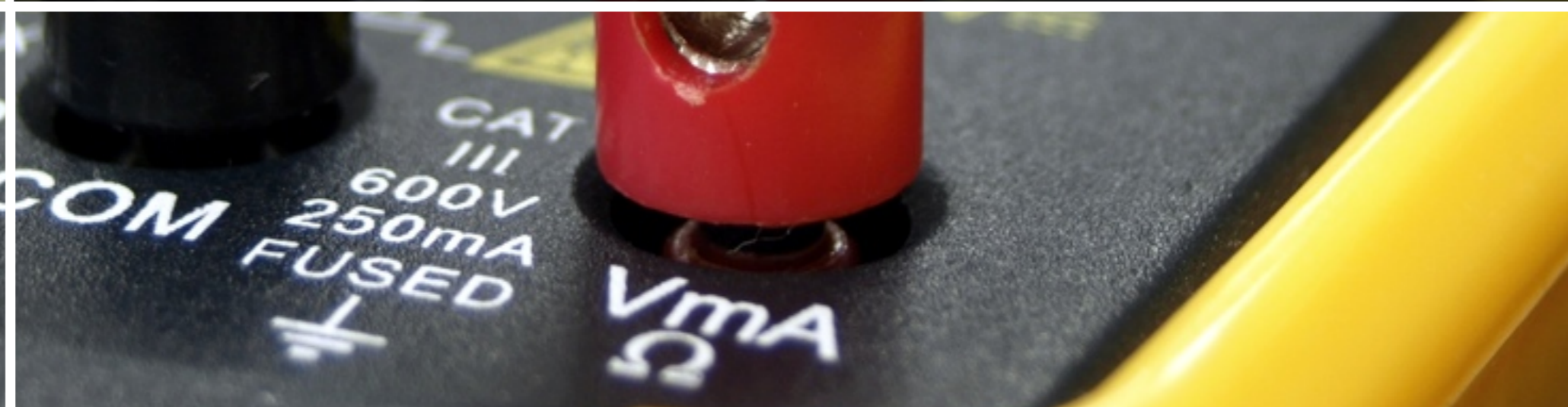
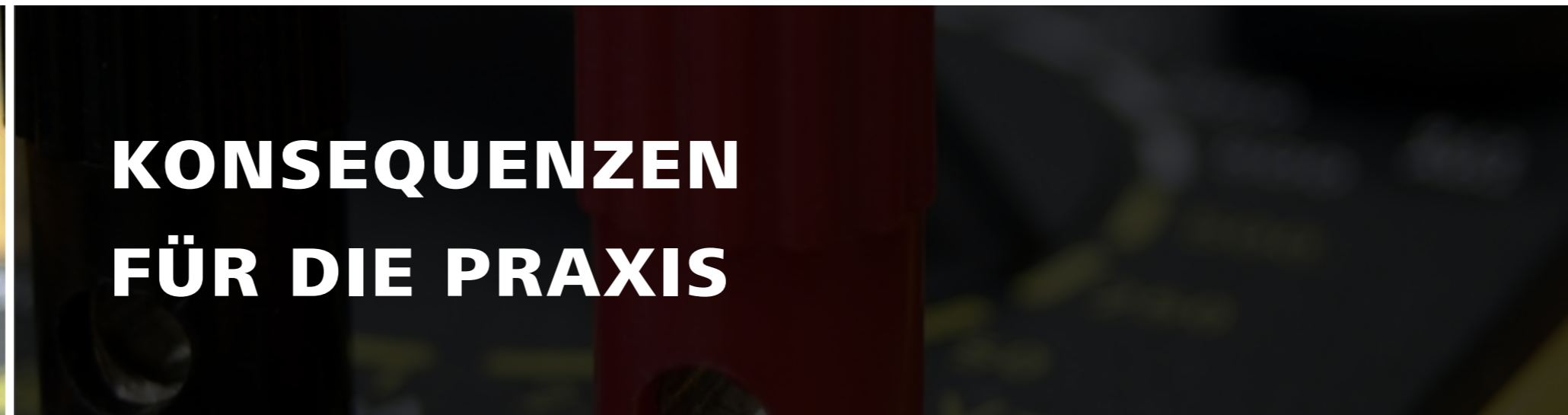
3100 Benutzer



Wiedererkennungsrate bei DNS

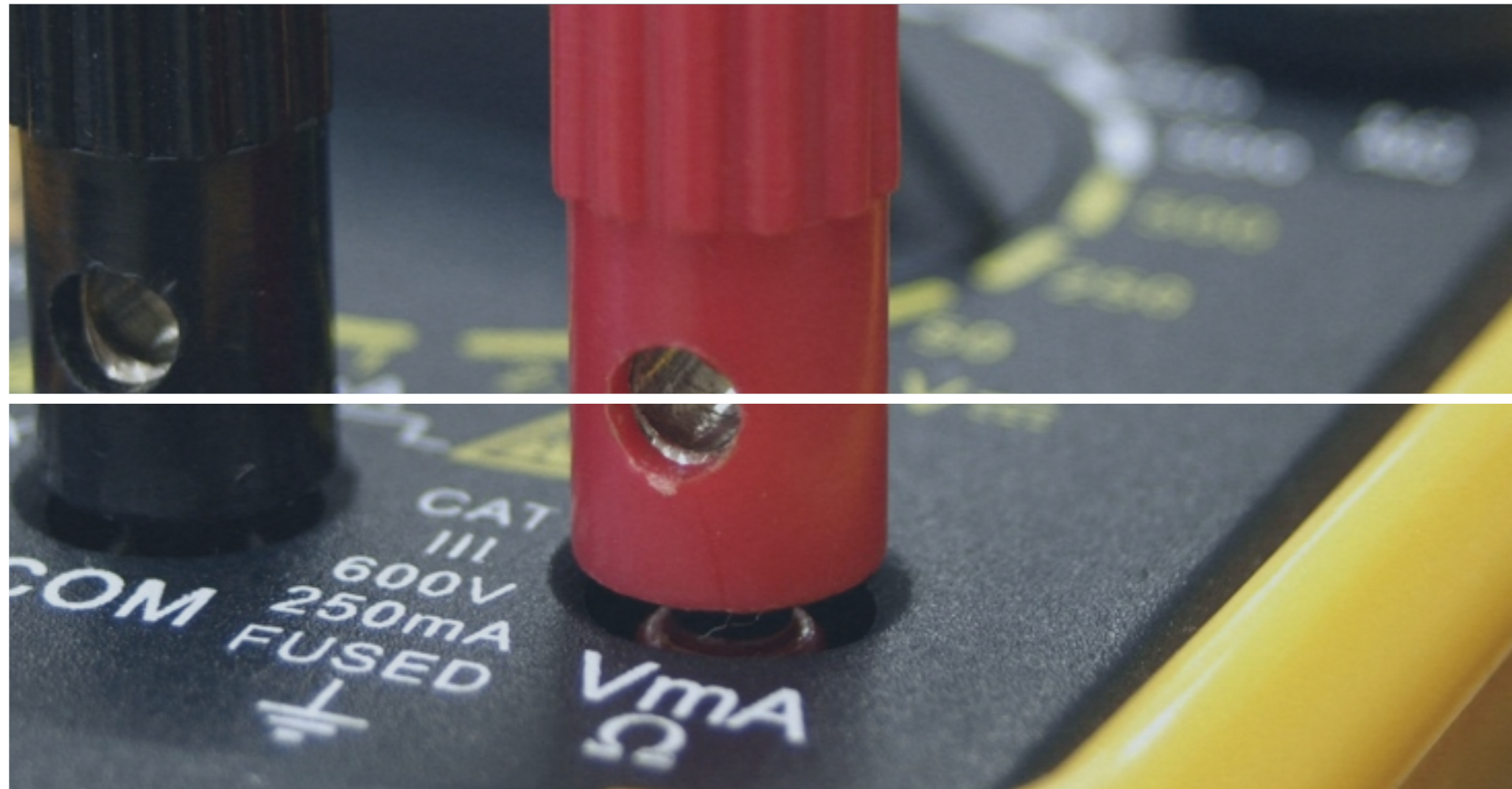


KONSEQUENZEN FÜR DIE PRAXIS



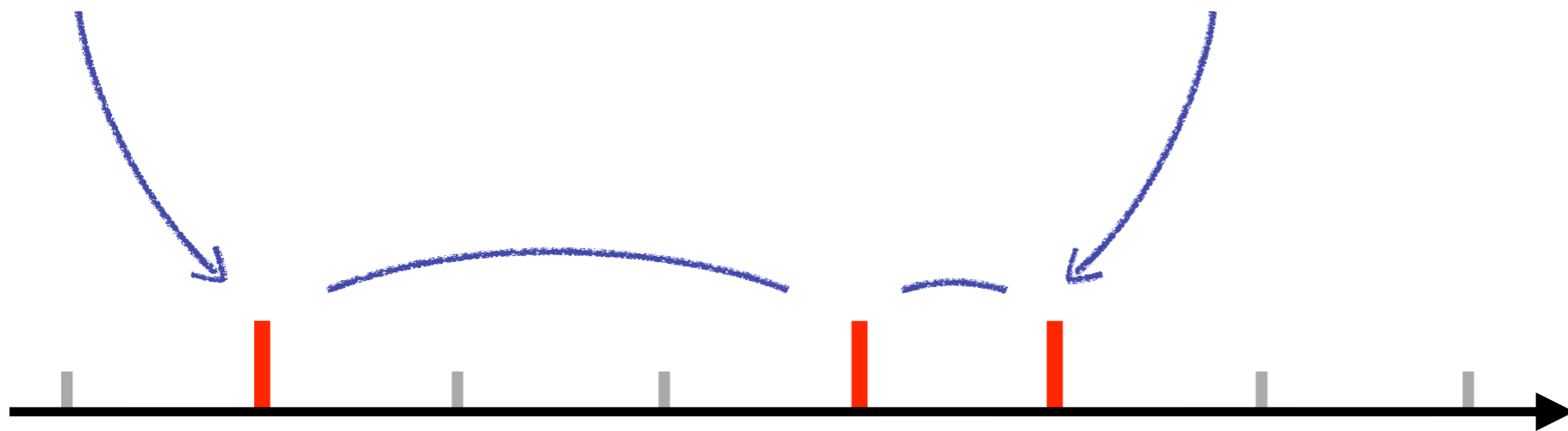
KONSEQUENZEN FÜR DIE PRAXIS

Eingangsbeispiel



Abruf durch
Website Fingerprinting
aufgedeckt

Sessions durch
User Linkability
verkettet

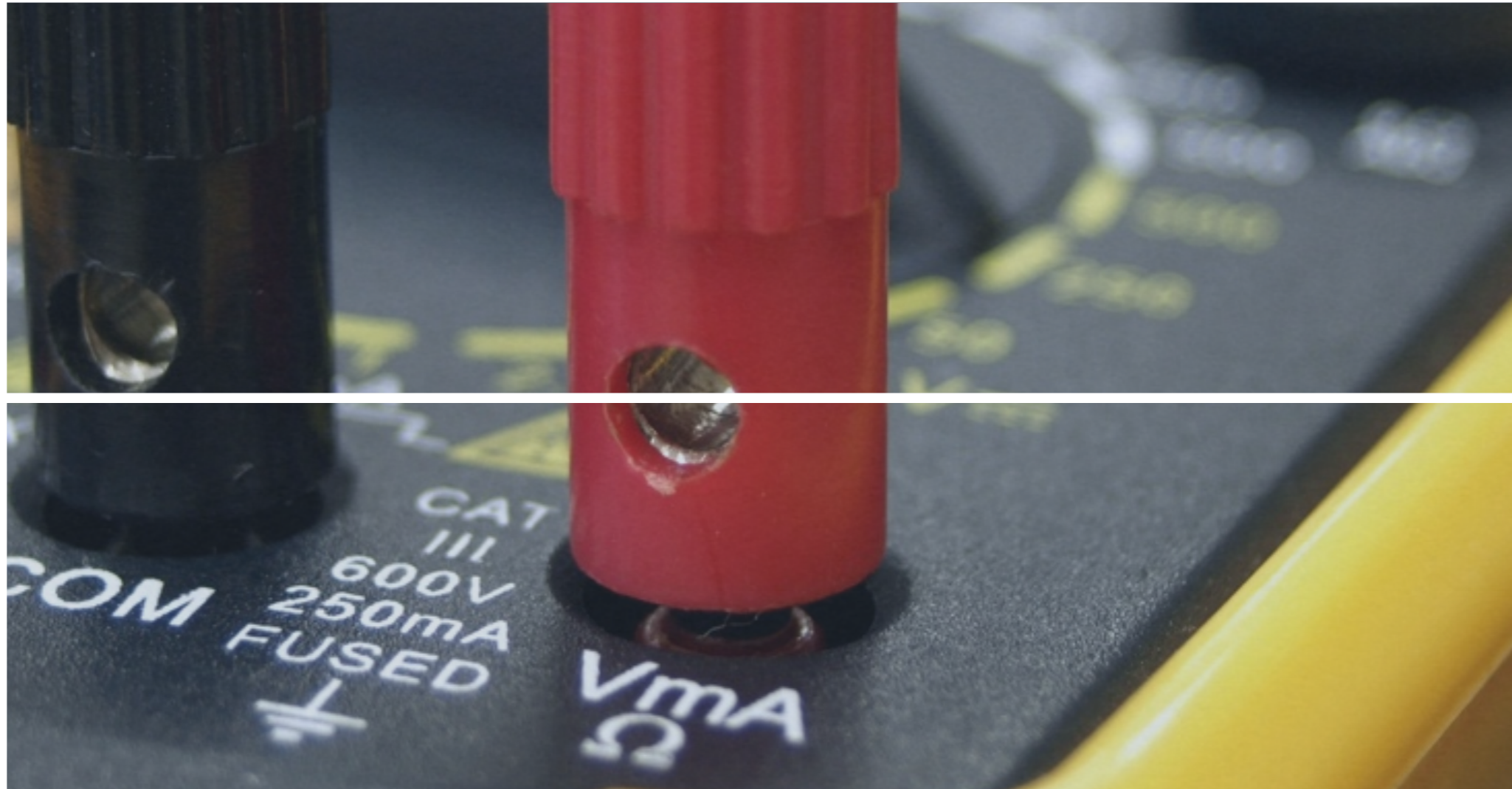


facebook

Identität online
preisgegeben

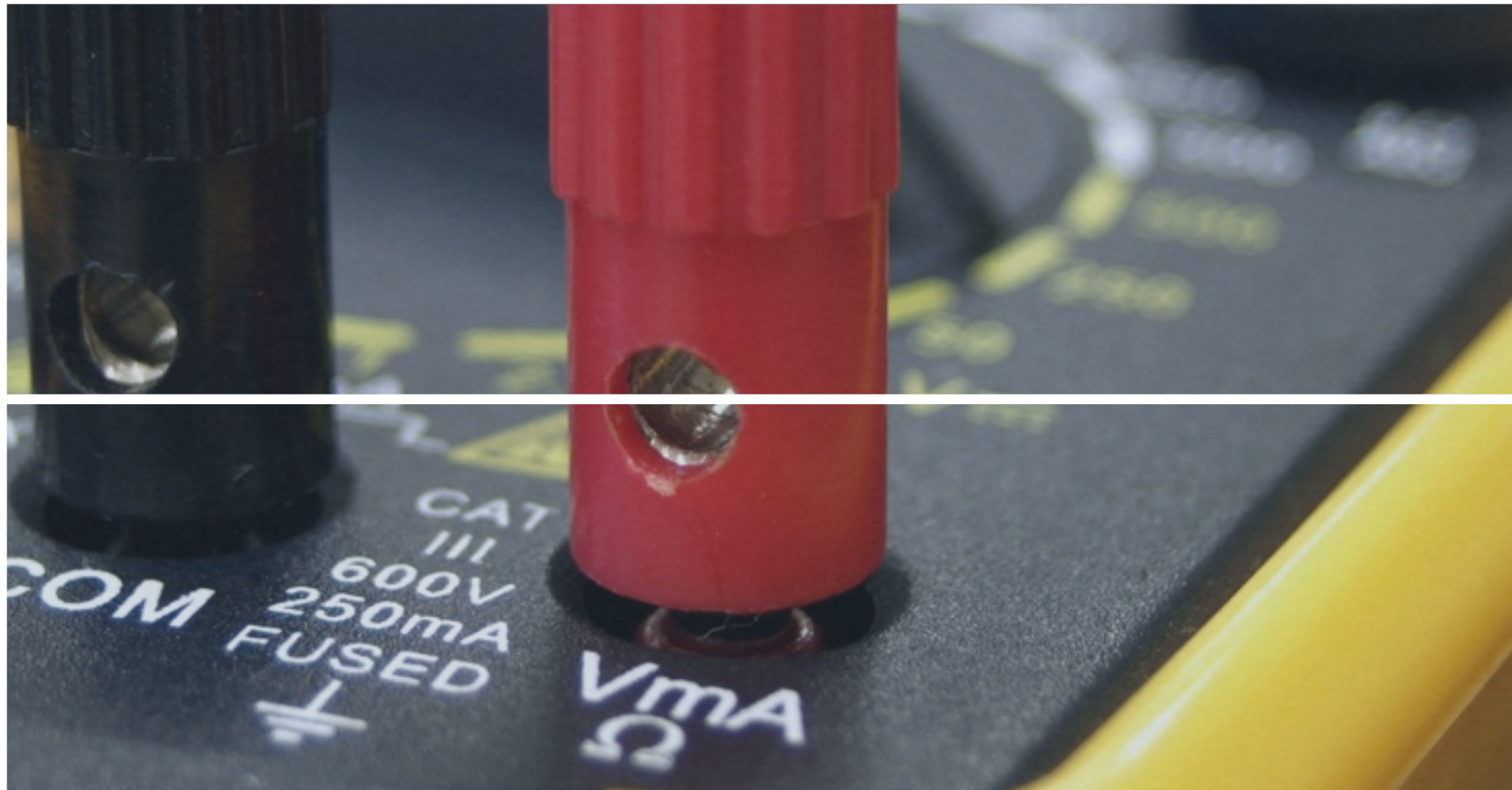
KONSEQUENZEN FÜR DIE PRAXIS

Eingangsbeispiel
reale Bedrohung?



KONSEQUENZEN FÜR DIE PRAXIS

Eingangsbeispiel
reale Bedrohung?
Abwehrmaßnahmen

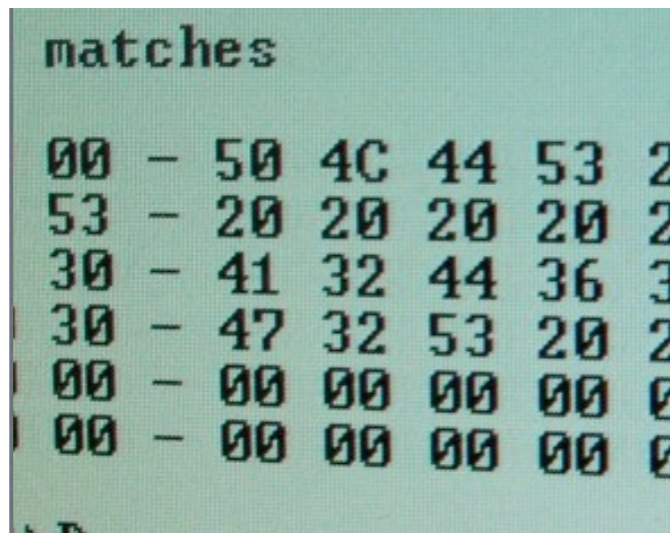
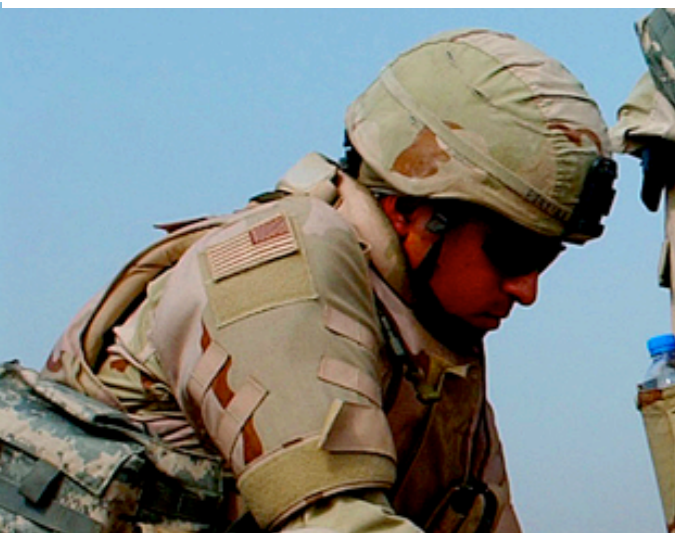


Zusammenfassung

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 verketteten



Dominik Herrmann und Christoph Gerber

Universität Regensburg

Lehrstuhl Management der Informationssicherheit

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