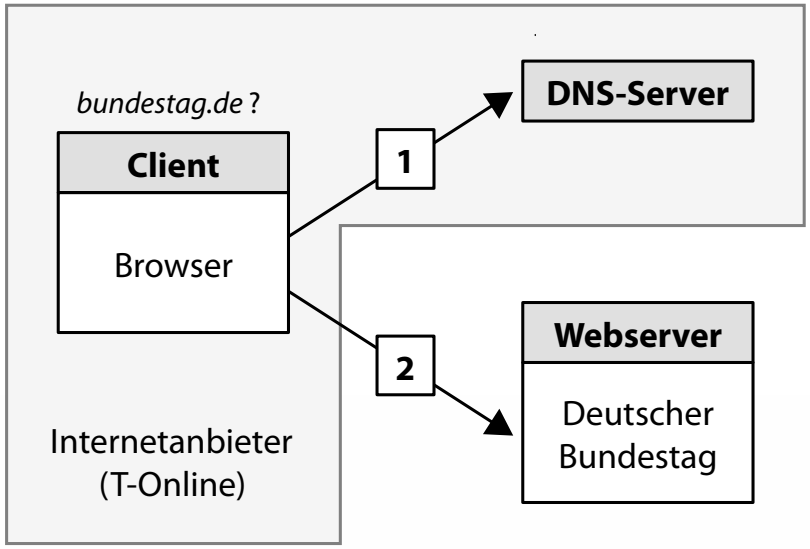
Unerfreulich auskunftsfreudig

Das Internet-Adressbuch bedroht unsere Privatsphäre

Dr. Dominik Herrmann
Universität Hamburg, Universität Siegen

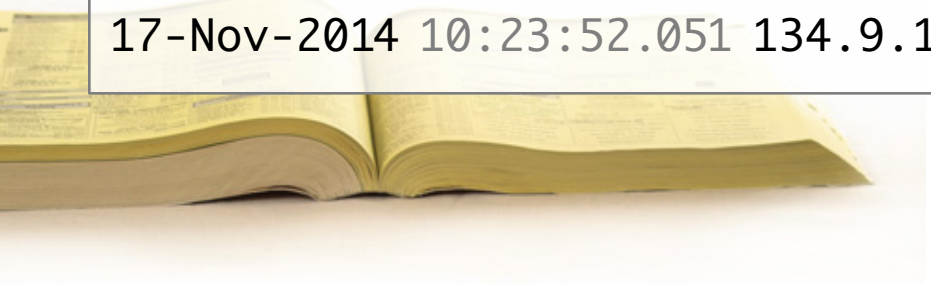
Folien zum Download:
http://dhgo.to/dns-auskunftsfreudig

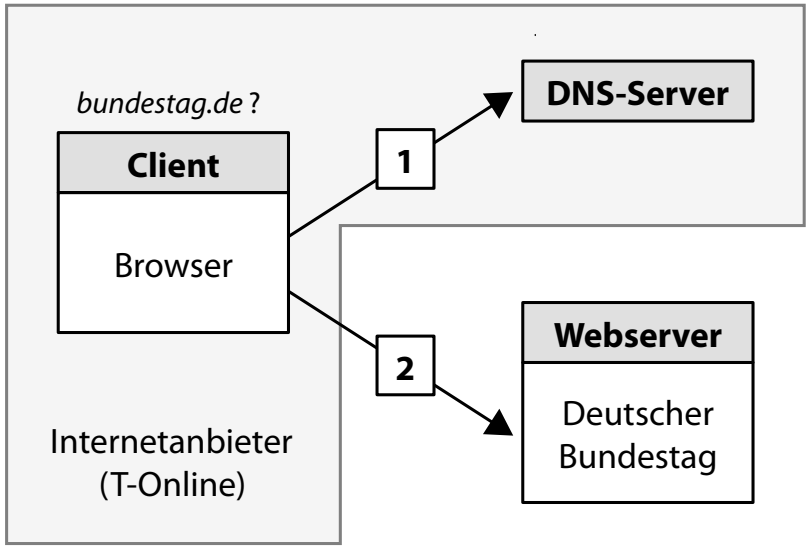
Das Domain Name System löst Domains in IP-Adressen auf.



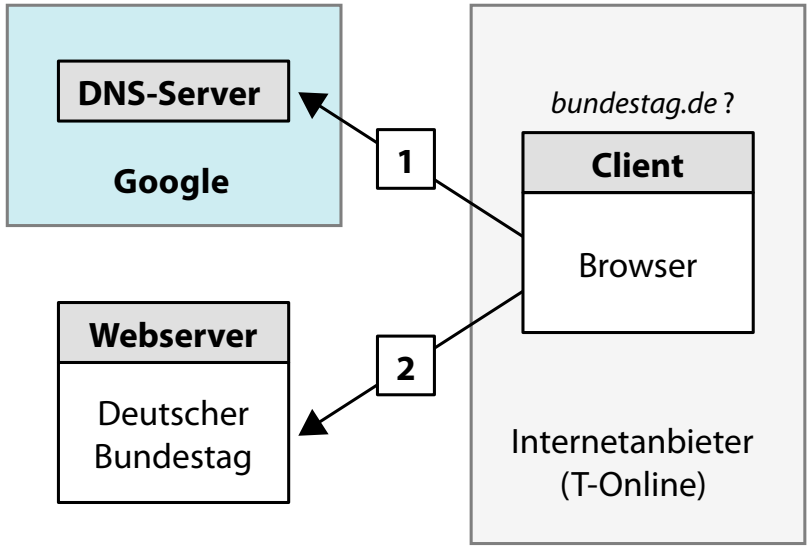
Vertraulichkeit?
Brauchen wir nicht.

```
17-Nov-2014 10:23:49.770 189.11.9.16 #15619: www.google.de  IN A +
17-Nov-2014 10:23:51.622 42.81.144.1 #12191: wikipedia.org  IN A +
17-Nov-2014 10:23:52.051 134.9.15.51 #13170: www.spiegel.de IN A +
```





sicherer,
zuverlässiger



Google DNS-Server 8.8.8.8
>150 Mrd. Anfragen pro Tag (2013)

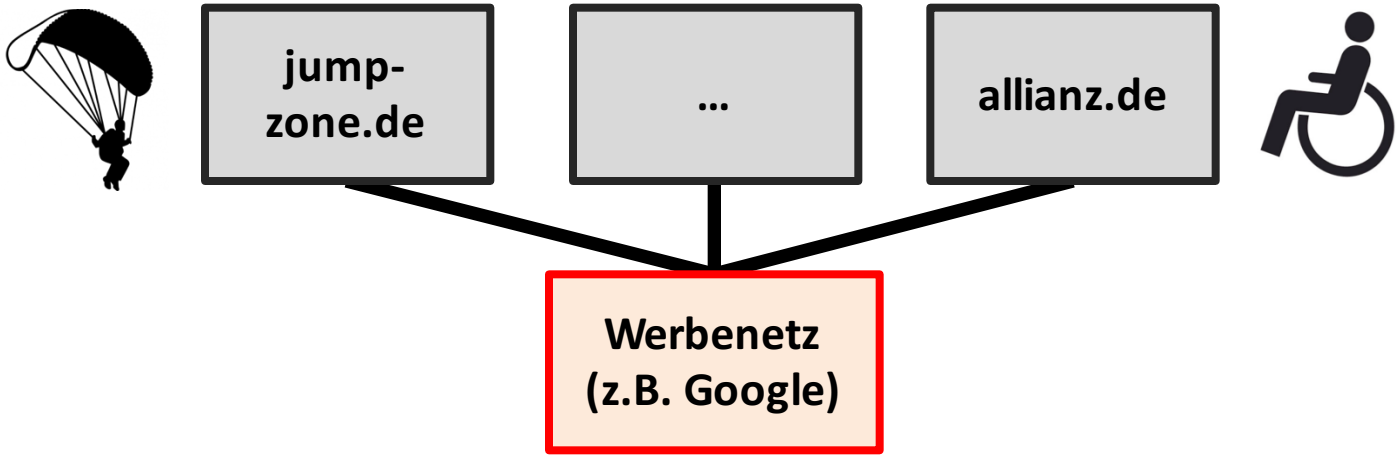
oder doch?

c't magazin

SPIEGEL ONLINE

Tracking ohne Cookies

Überwachung von Internetnutzern
anhand ihrer DNS-Anfragen



Interessenskonflikt

Nutzer vs. Internetwirtschaft

Aktuelle Techniken sind unzuverlässig bzw. erkennbar.

Tracking-Cookies

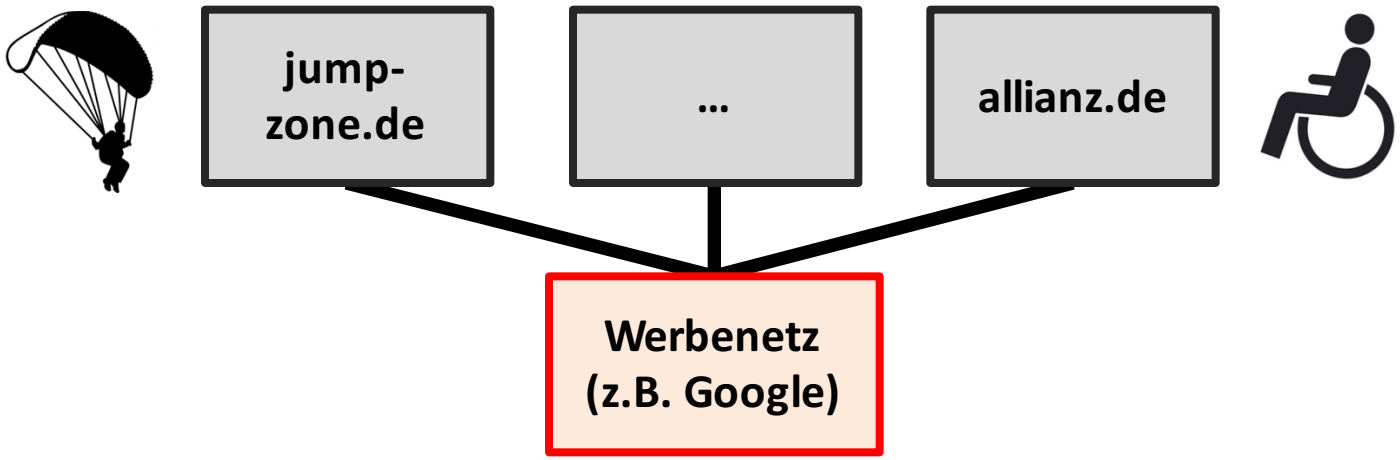
56ac1c08fa5479fd57c4

a5c65861c4ed3ed93ff8

Browser-Fingerprinting

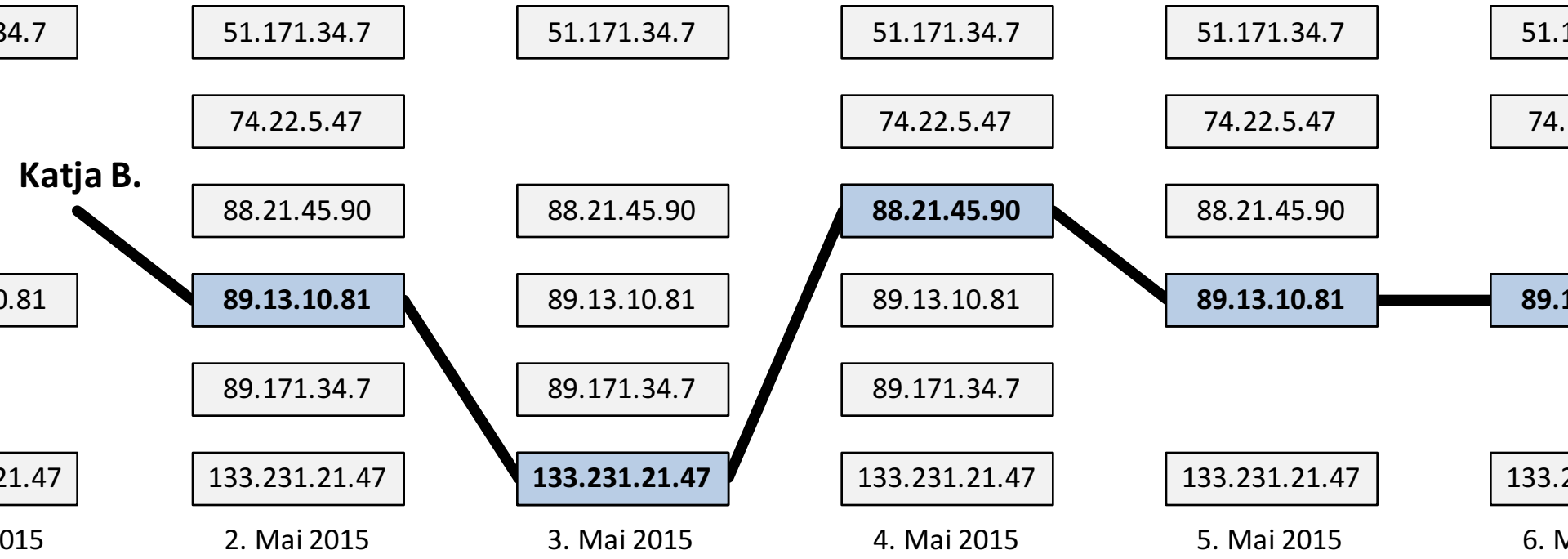
How quickly daft jumping

How quickly daft jumping



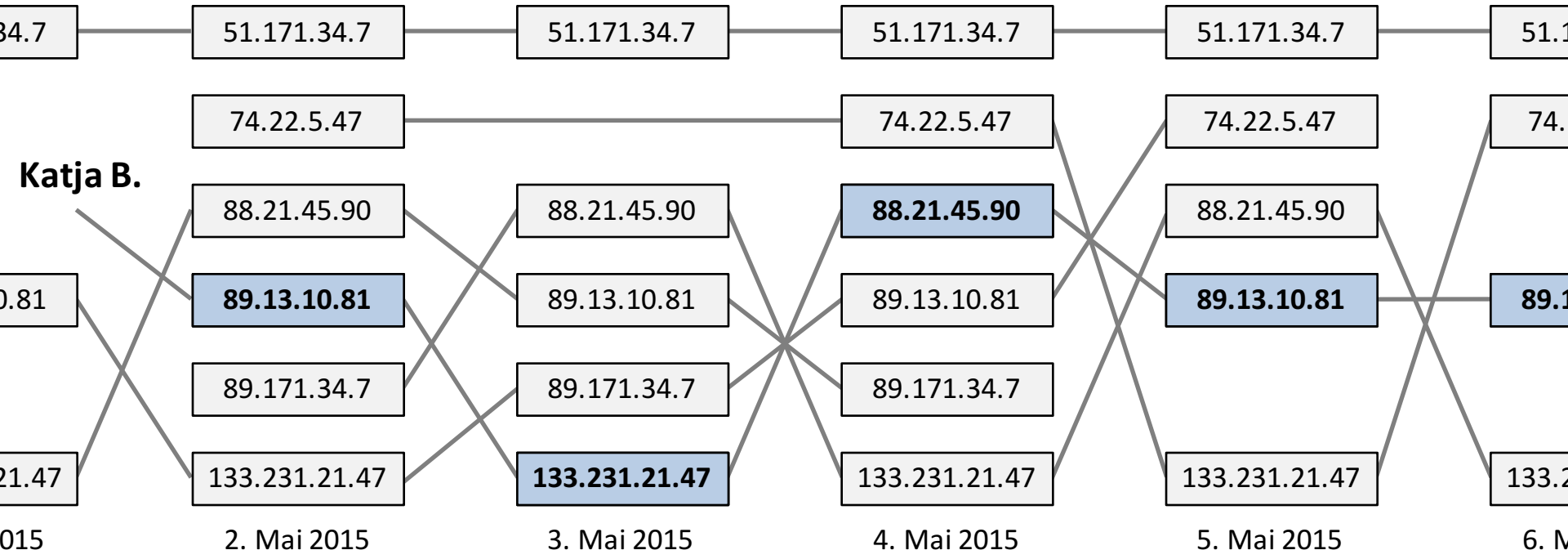
Herausforderung

Wiedererkennung von Nutzern trotz
(meist täglich) wechselnder IP-Adressen



Herausforderung

Wiedererkennung von Nutzern trotz
(meist täglich) wechselnder IP-Adressen

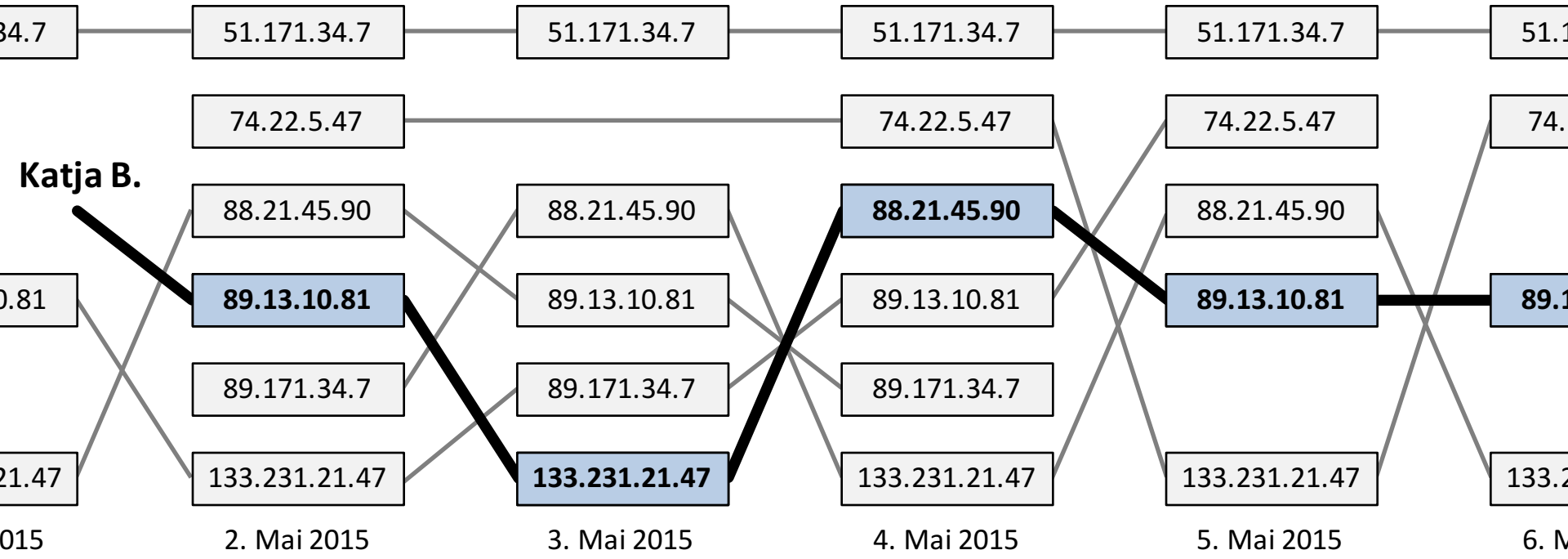


Herausforderung

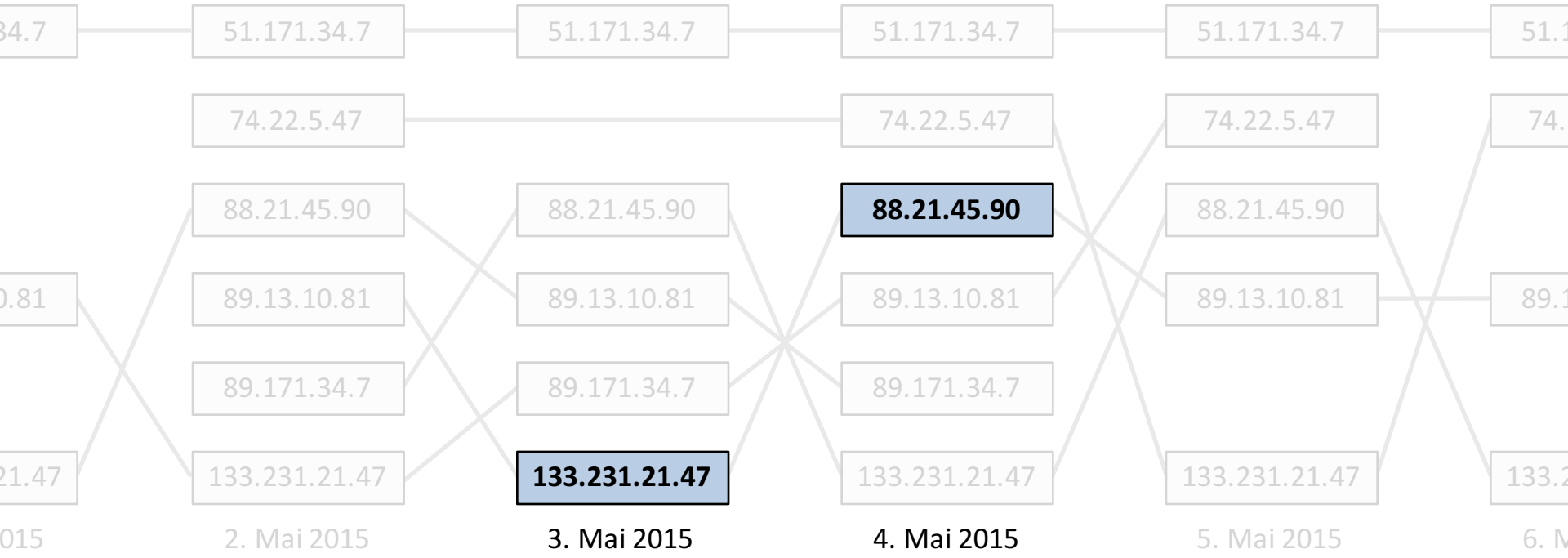
Wiedererkennung von Nutzern trotz (meist täglich) wechselnder IP-Adressen

Neues Verfahren

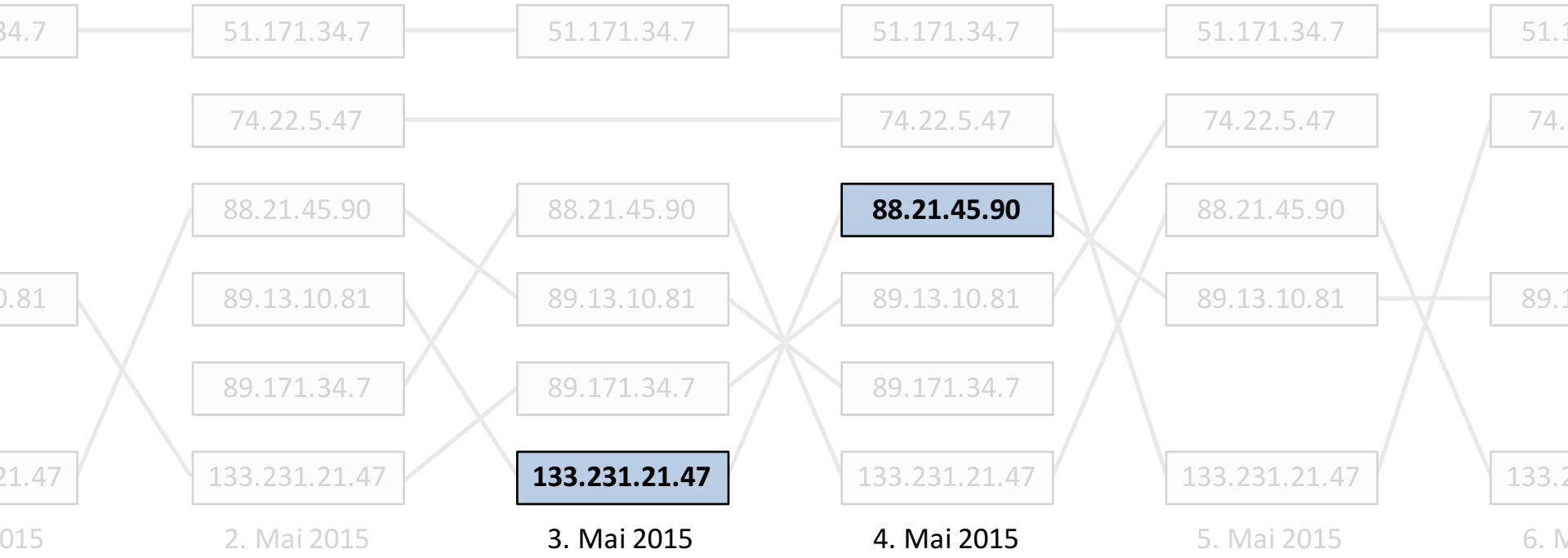
Wiedererkennung anhand der beobachtbaren DNS-Anfragen



Domains vom 3. Mai			Domains vom 4. Mai
spiegel.de	4 x		1 x	spiegel.de
google.de	15 x		9 x	google.de
apple.com	1 x		0 x	apple.com



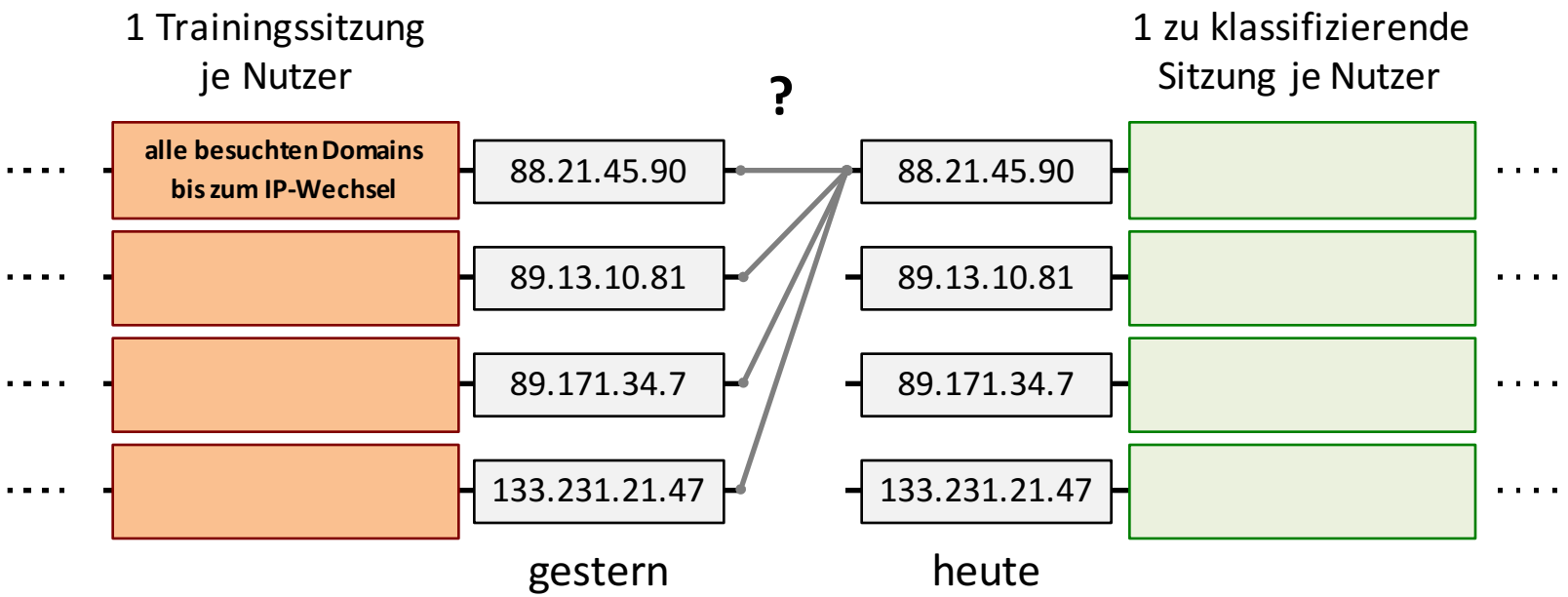
Domains vom 3. Mai			Domains vom 4. Mai
spiegel.de	4 x	1 x	spiegel.de
google.de	15 x	9 x	google.de
apple.com	1 x	0 x	apple.com
airbus.com	**3 x**	**6 x**	**airbus.com**
avalster.de	**2 x**	**3 x**	**avalster.de**



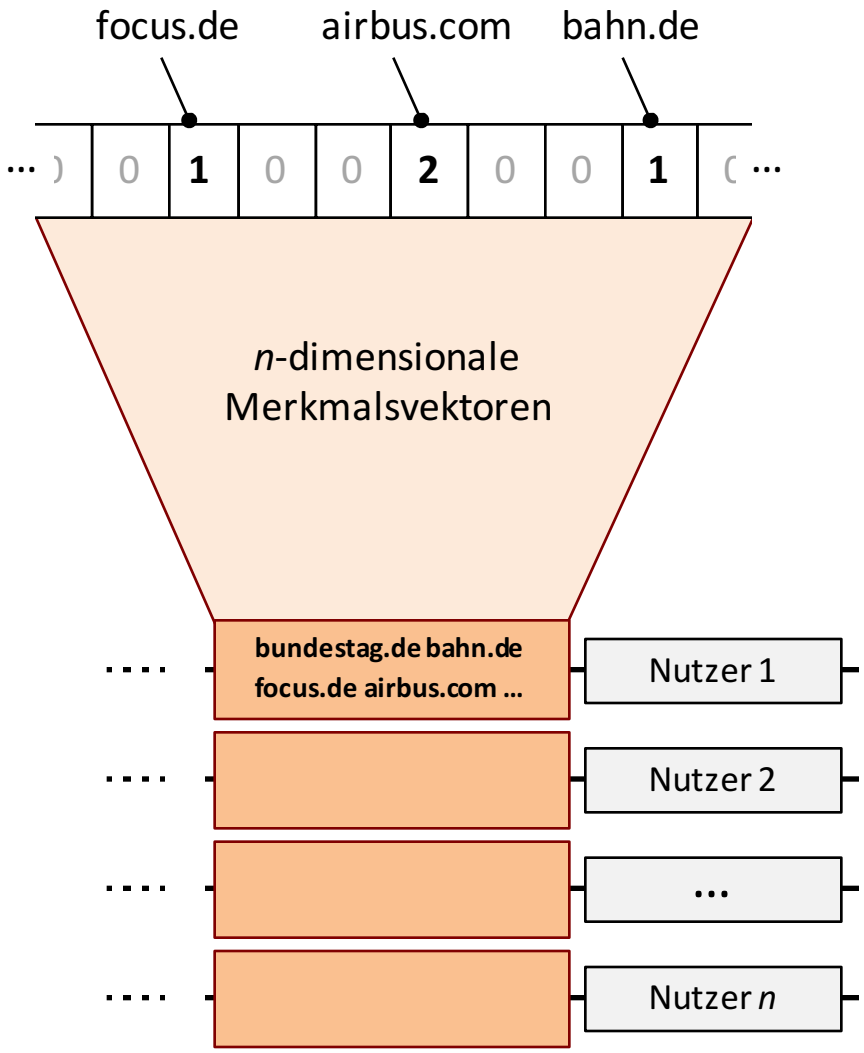
Hypothese

individuelle Vorlieben tägliche Routine

Verkettung durch überwachtes Lernen



Konstruktion des Verkettungsverfahrens

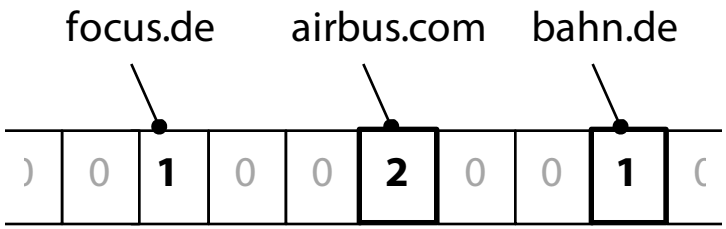


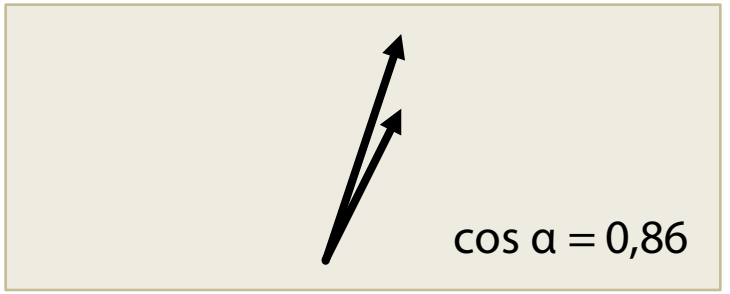
Logarithmierung der Häufigkeiten
Gewichtung mit IDF-Faktor
Normierung der Vektorlänge
Bildung von N-Grammen

1-Nearest-Neighbor-Klassifikator
Cosine-Similarity

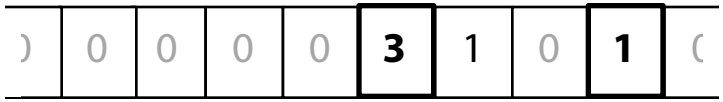
https://github.com/hadoop-dns-tracking

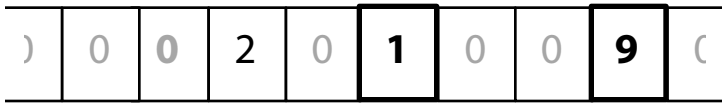
Ermittlung der am besten passenden Sitzung

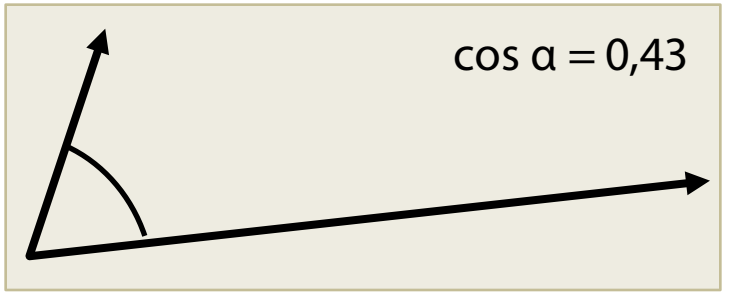




1-Nearest-Neighbor-Klassifikator
mit Cosine-Similarity







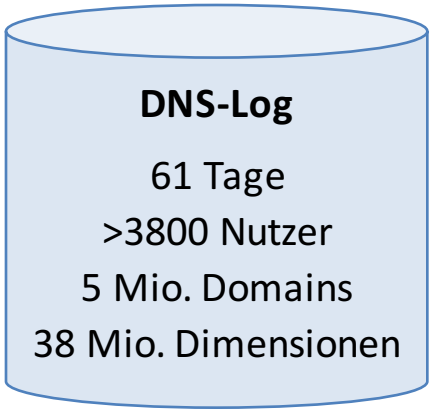
gestern

heute

Empirische Untersuchung

Forschungsfragen:

– Genauigkeit?

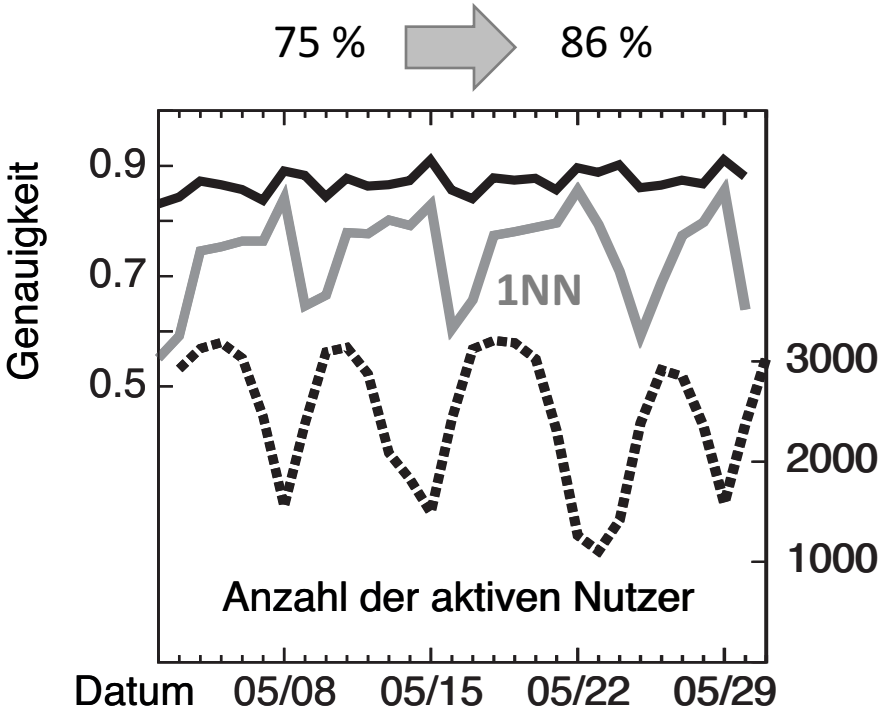


inkl. »ground truth«
(pseudonymisiert)

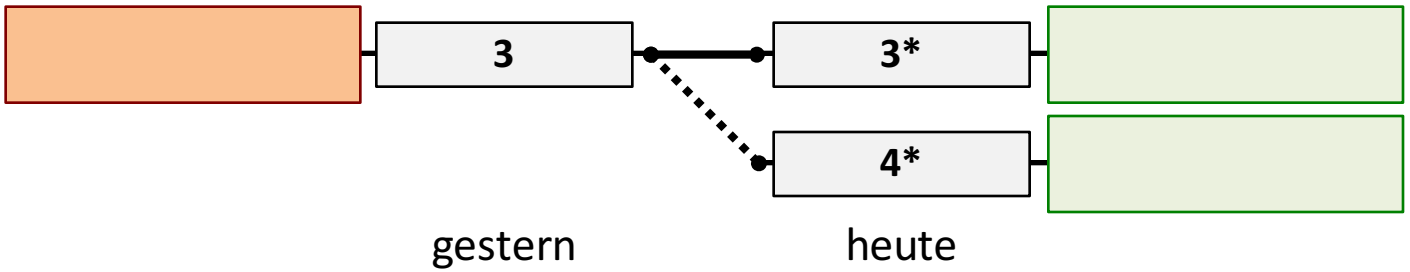
Empirische Untersuchung

Forschungsfragen:

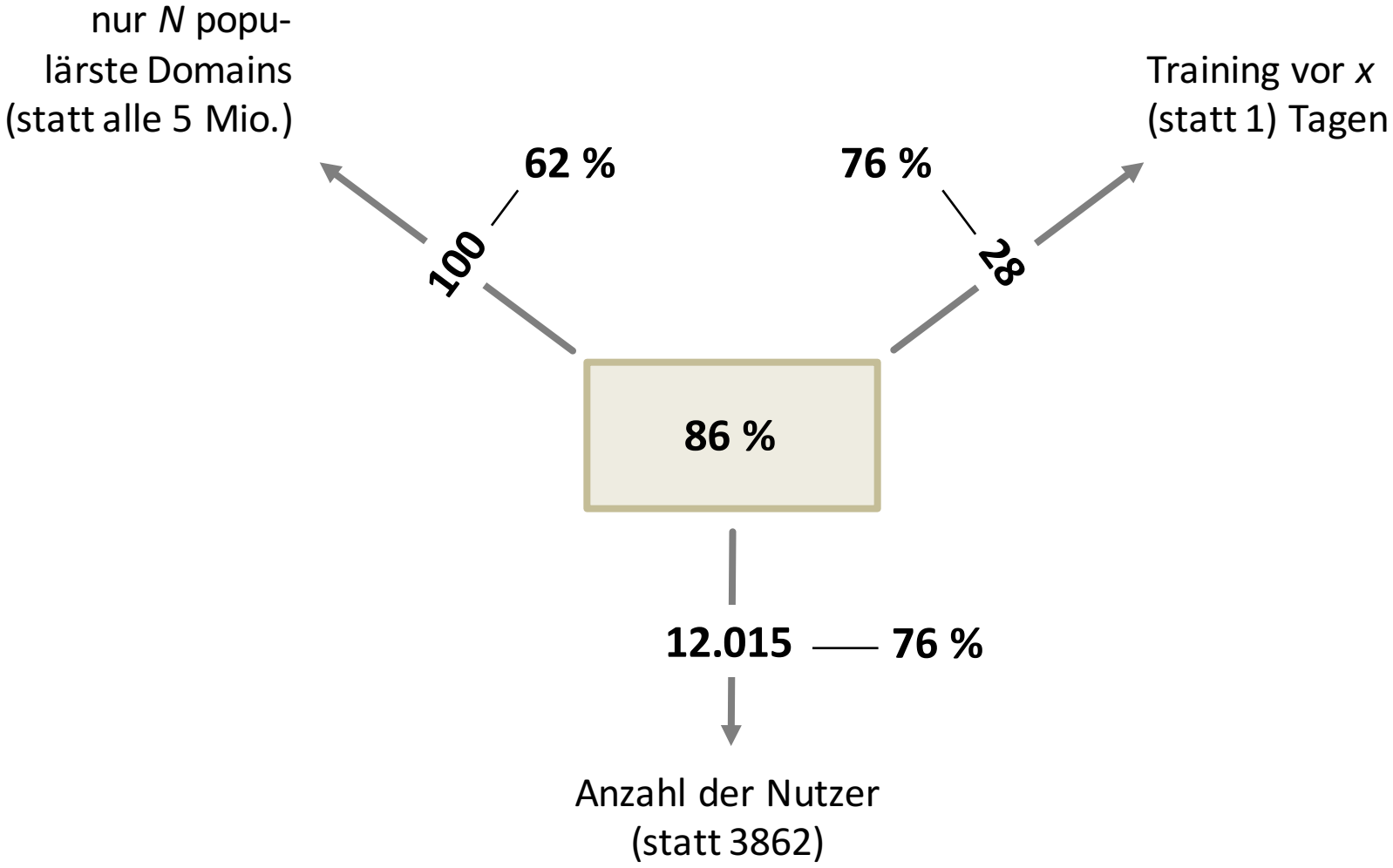
- Genauigkeit?
- Umgang mit Fluktuation?



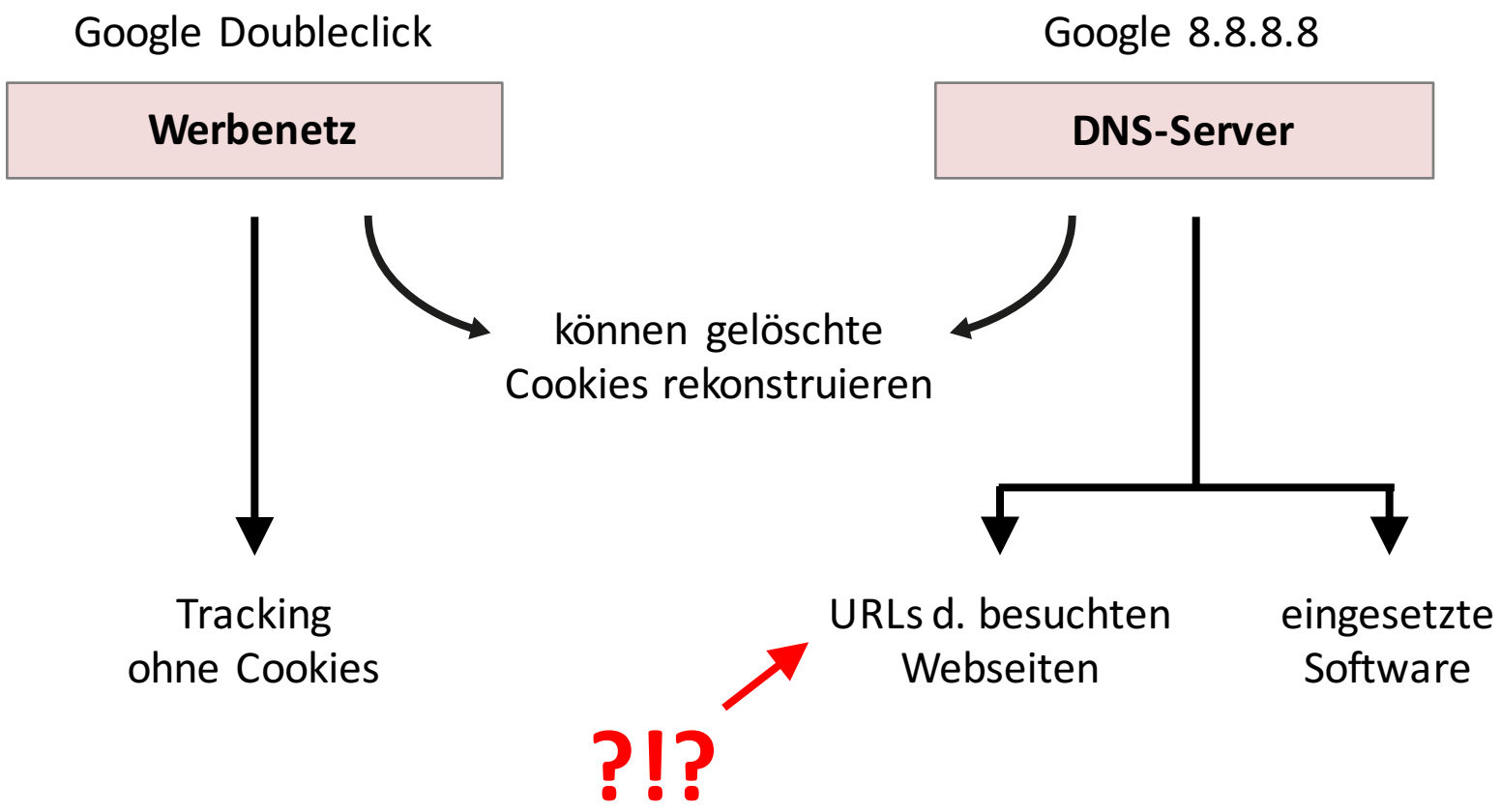
Problem: mehrdeutige Zuordnungen im Open-World-Szenario



Verkettung gelingt auch unter erschwerten Bedingungen.



Ergebnis: neue Beobachtungsmöglichkeiten – nicht nur im DNS



rein passiv und nicht erkennbar

Verlust der informationellen Selbstbestimmung

Abrufmuster für
http://de.wikipedia.org/wiki/Alkoholkrankheit

de.wikipedia.org	www.klinik-dr-fontheim.de
bits.wikimedia.org	www.versorgungsleitlinien.de
meta.wikimedia.org	de.wikiquote.org
counsellingresource.com	drogenbeauftragte.de
upload.wikimedia.org	www.sucht-info.ch
www.izb.fraunhofer.de	www.aafp.org
www.spiegel.de	www.thieme-connect.com
www.stadt-und-gemeinde.de	www.kenn-dein-limit.de

DNS-Prefetching erzeugt charakteristische Abrufmuster, anhand derer sich u.U. die genaue URL ermitteln lässt.

Abhilfe: https verwenden

Datenschutz-Techniken

Überwachung durch DNS-Server und verhaltensb. Tracking verhindern

Schutz der Identität
des Nutzers

Verbergen der
wahren Interessen

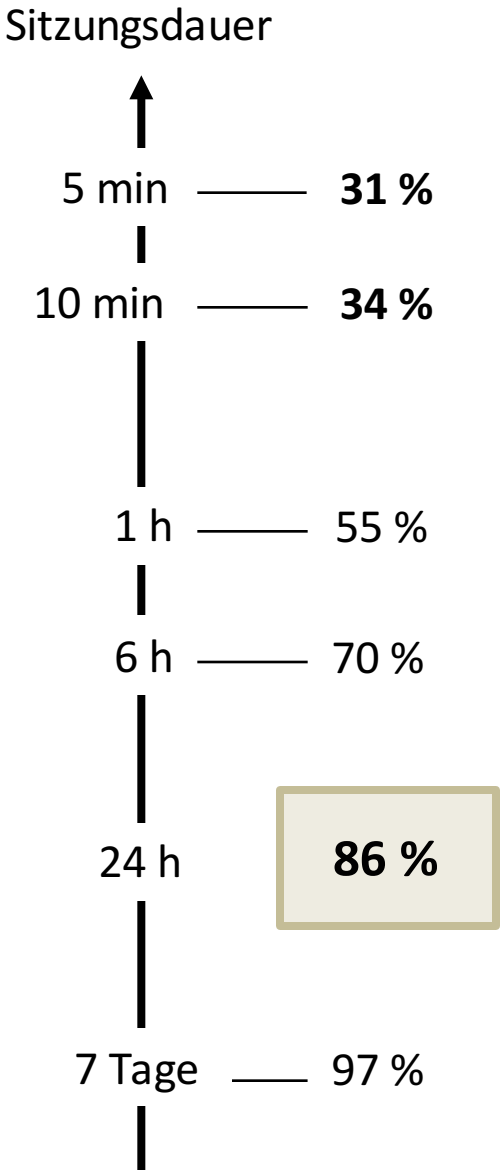
langsam

Range Queries

aufwändig (und unsicher)

existierende Datenschutztechniken für DNS ungeeignet

Praktikabler Schutz vor Tracking wäre leicht umsetzbar.



IP-Adresse häufig wechseln

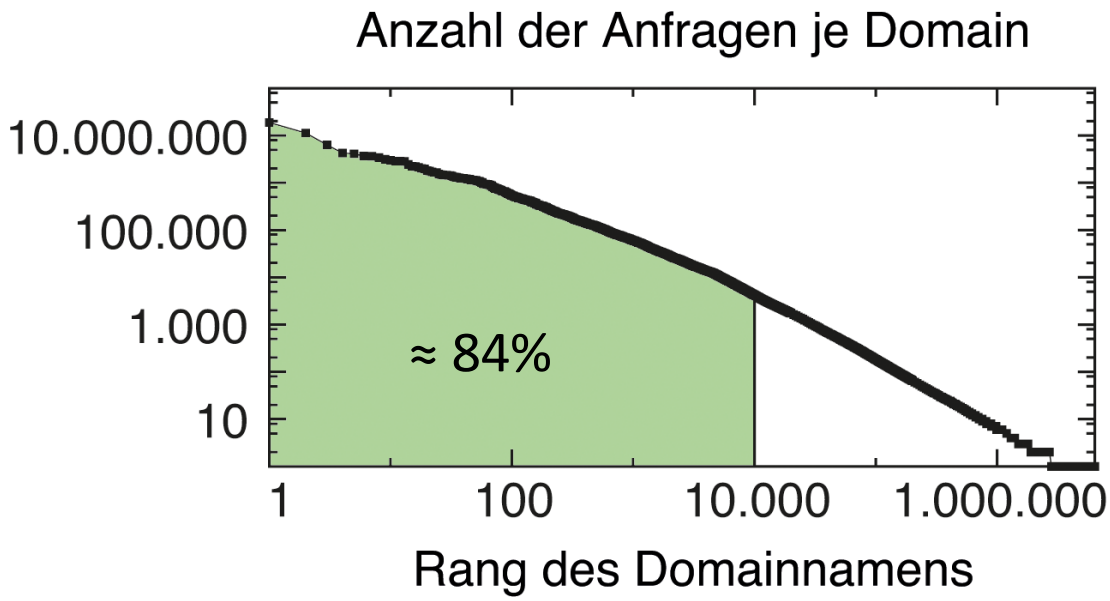
Chance

»Privacy by Default« mit IPv6

Bundesministerium
für Bildung
und Forschung

AN.ON-Next

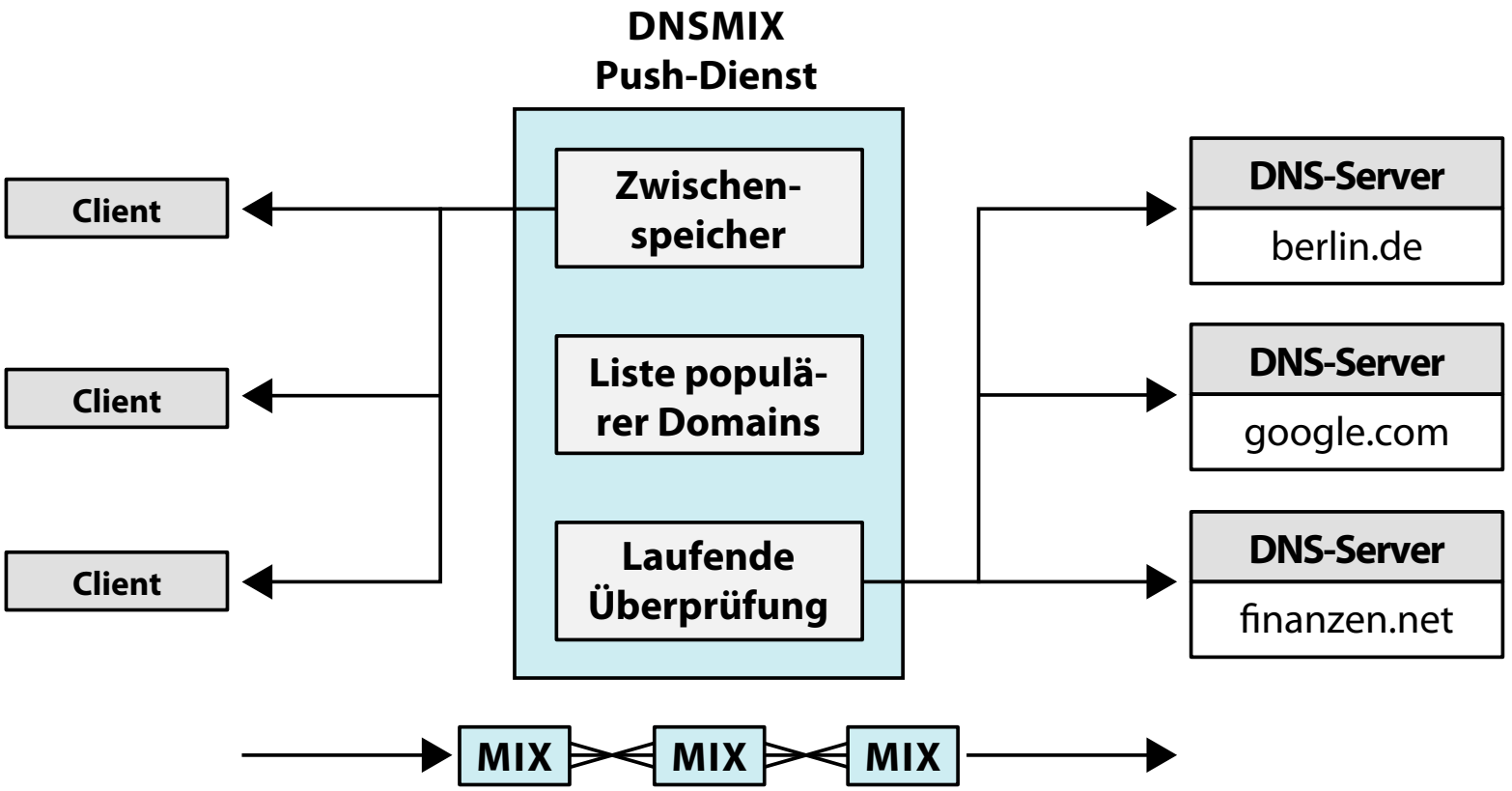
DNSMIX: ein neuer Ansatz zum Schutz vor Überwachung



Idee von DNSMIX

populäre DNS-Einträge automatisch an alle Nutzer senden

Kostet das nicht viel zu viel Bandbreite?

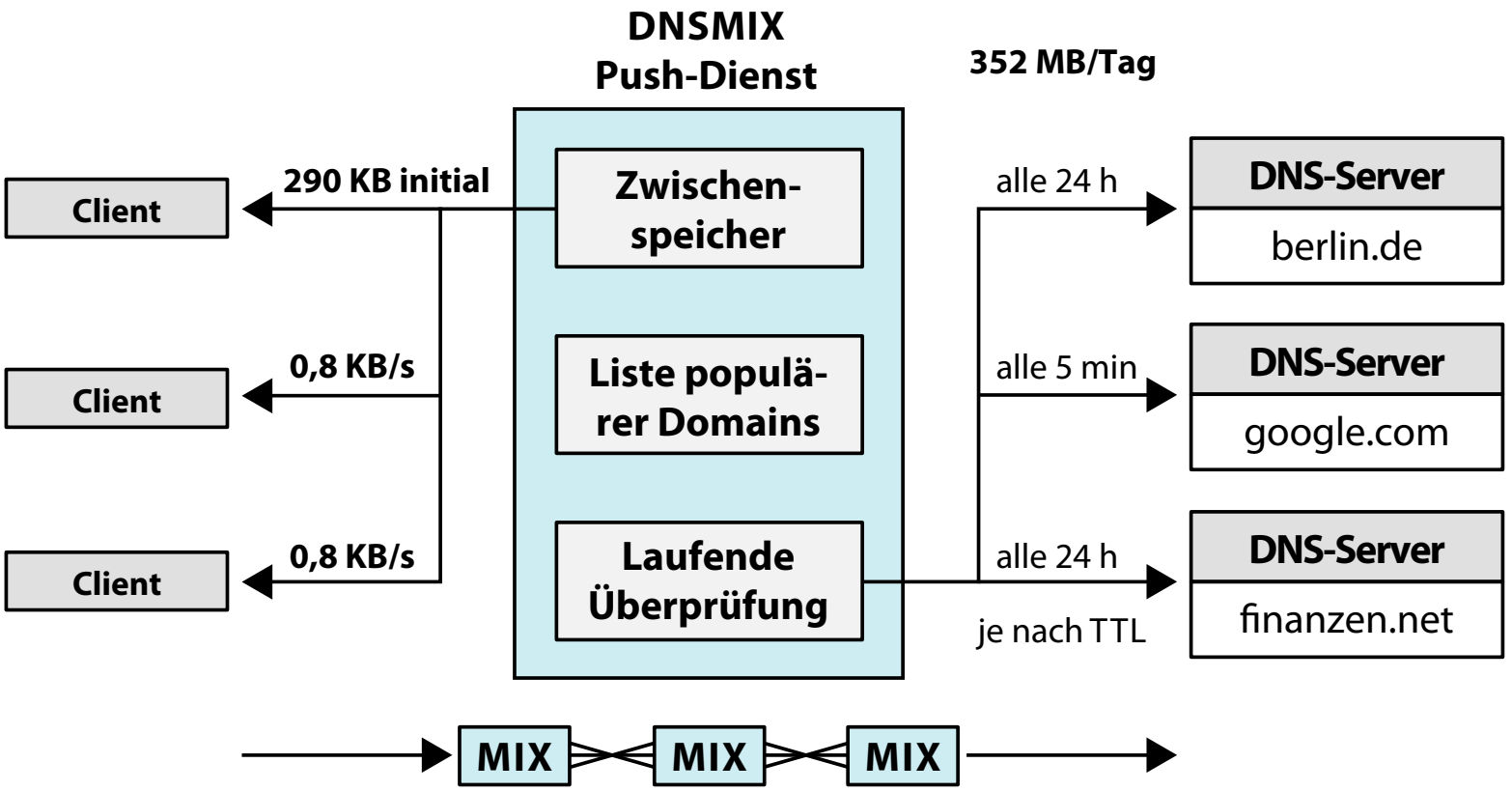


Empirische Untersuchung

Anfragen von 2082 Nutzern

Pushen von 10.000 populären Domains

Auflösung von 84 % der Anfragen **unbeobachtbar** und **unmittelbar**



Beobachtungsmöglichkeiten im DNS

umfangreich, aber bislang vernachlässigt

INFERENZANGRIFFE AUF DIE PRIVATSPHÄRE	TECHNIKEN ZUM **SELBST**DATENSCHUTZ
Verhaltensbasiertes Tracking ohne Cookies	**Häufiger IP-Wechsel** oder längeres DNS-Caching
DNS-basiertes Website-Fingerprinting	**Unbeobachtbarkeit mit DNSMIX-Push-Dienst**
Software-Identifizierung anhand DNS-Verhaltens	Verschleierung mit Range Queries
Sensibilisierung – aber auch in der IT-Forensik anwendbar	Gestaltungsvorschläge für Forschung und Entwicklung