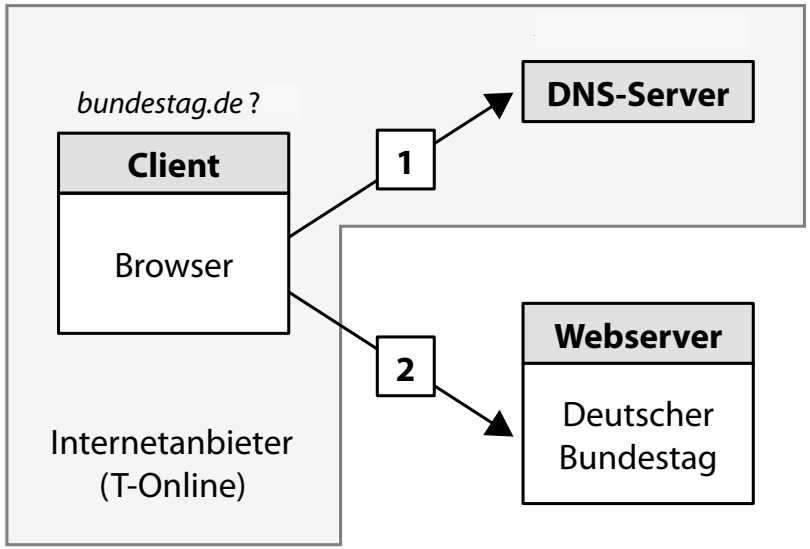
Beobachtungsmöglichkeiten im Domain Name System

Angriffe auf die Privatsphäre und Techniken zum Selbstdatenschutz

Dr. Dominik Herrmann
Universität Hamburg

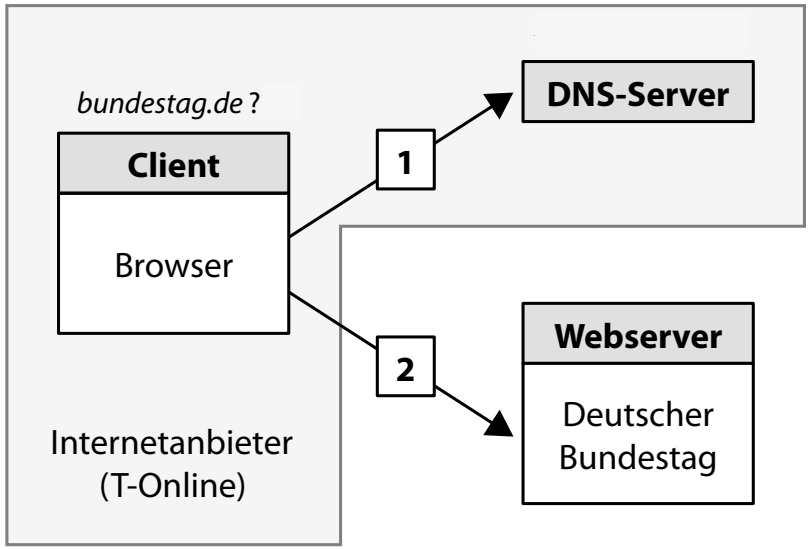
Download Handout und Folien
http://dhgo.to/castfolien bzw. /casthandout

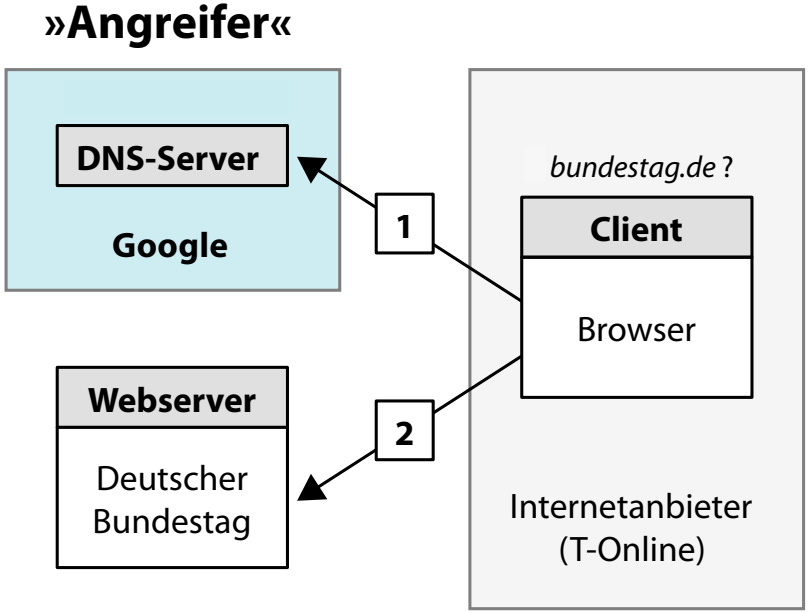
Domain Name System



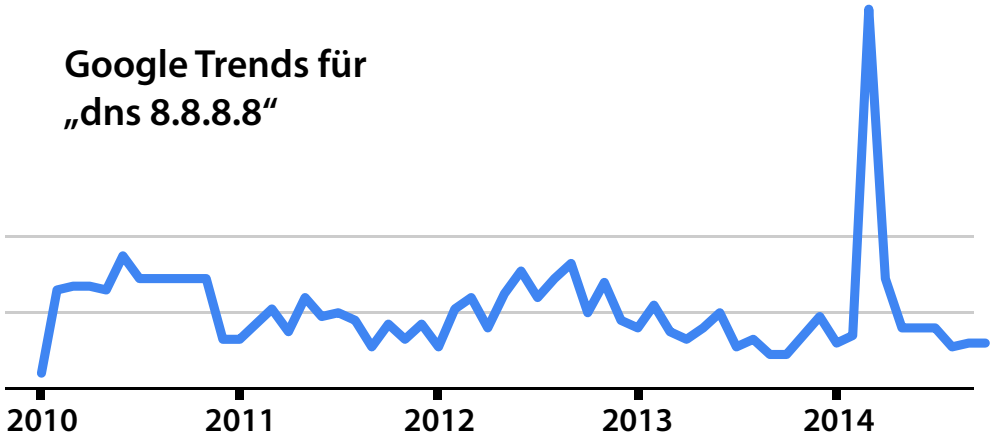
Vertraulichkeit?
Brauchen wir nicht.

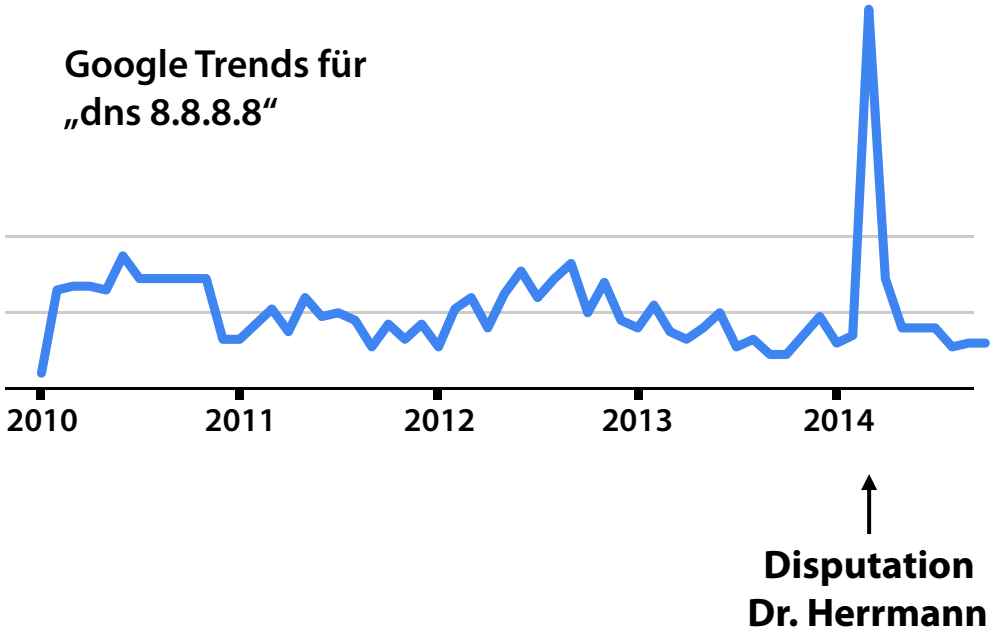
```
17-Nov-2014 10:23:49.770 189.11.9.16 #15619: query:  www.google.de   IN A +
17-Nov-2014 10:23:51.622 42.81.144.1 #12191: query:  wikipedia.org   IN A +
17-Nov-2014 10:23:52.051 134.9.15.51 #13170: query:  www.spiegel.de  IN A +
```


bundestag.de ?
Client
Browser
1
DNS-Server
2
Webserver
Deutscher
Bundestag
Internetanbieter
(T-Online)

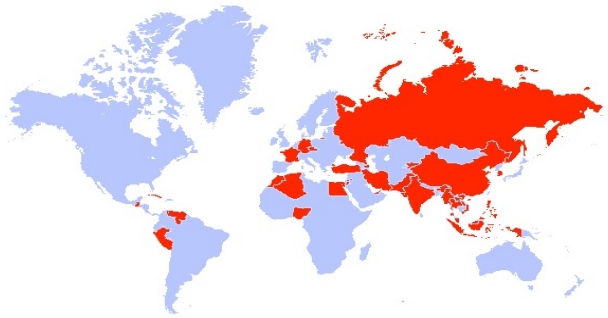

»Angreifer«
DNS-Server
Google
1
bundestag.de ?
Client
Browser
Webserver
2
Deutscher
Bundestag
Internetanbieter
(T-Online)

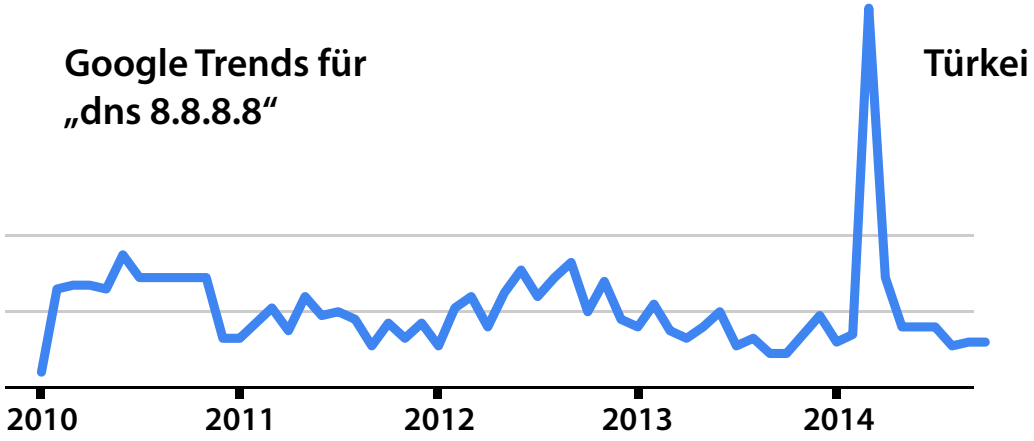
oder doch?


Google Trends für
„dns 8.8.8.8“
2010
2011
2012
2013
2014

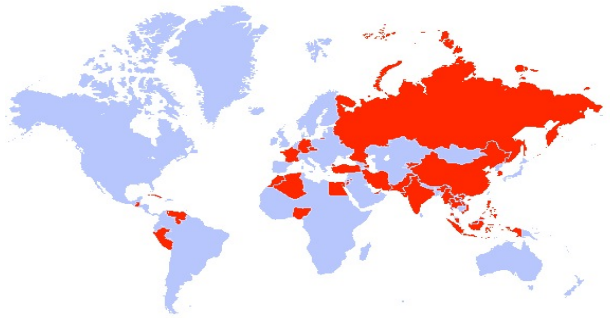

Google Trends für
„dns 8.8.8.8“
2010
2011
2012
2013
2014
Disputation
Dr. Herrmann

Umgehung von DNS-Sperren

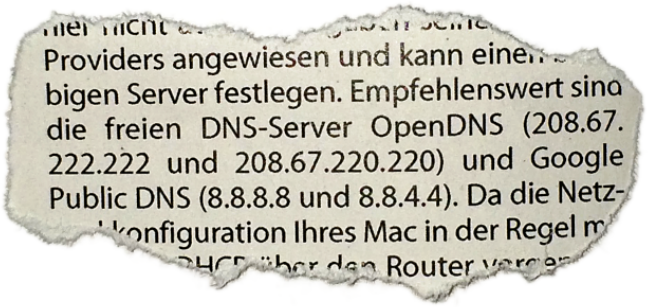


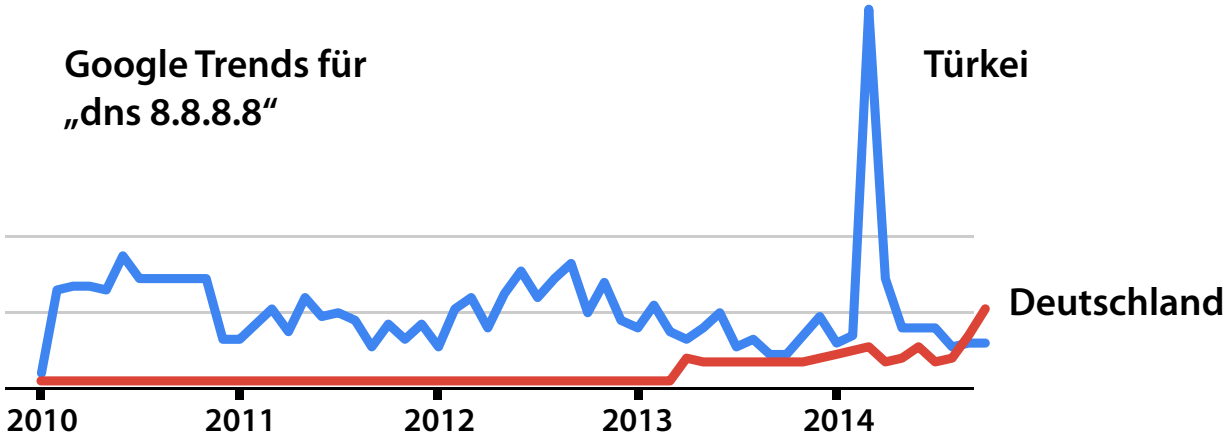


Umgehung von DNS-Sperren

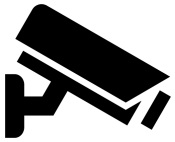


Performanz & Zuverlässigkeit





Motivation



Bedrohung für die Privatsphäre?
Geeignete Schutzmechanismen?

Nutzen für die IT-Forensik?

A 

Ermittlung der
besuchten Webseiten

Motivation: Profiling der Nutzer

Problem 1:
DNS-Server sieht nur Domains, jedoch keine URLs

Problem 2:
Domains korrespondieren nicht mit besuchten Seiten

```
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
www.biospektrum.de                  www.casusconsult.nl
w210.ub.uni-tuebingen.de            www.hta.ac.uk
www.uni-muenster.de                 www.stern.de
ec.europa.eu                        whqlibdoc.who.int
www.alkoholismus-hilfe.de           www.focus.de
```

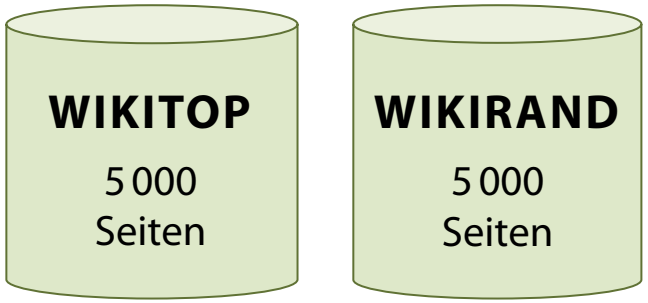
Abrufmuster für
http://de.wikipedia.org/wiki/Alkoholkrankheit

```
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
```

Kann ein DNS-Server (nicht-)besuchte Webseiten anhand charakteristischer Abrufmuster erkennen?

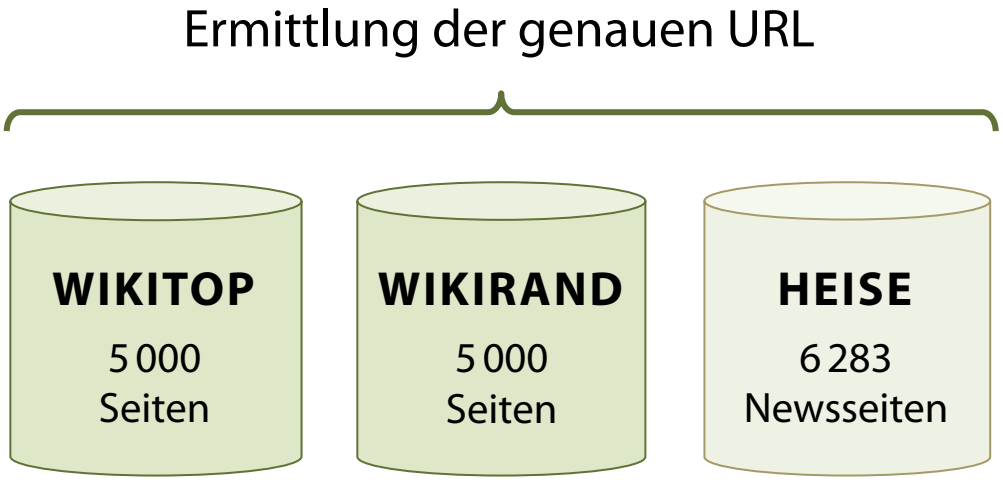
Empirische Untersuchung

1. Abruf von Webseiten
2. Aufzeichnen der DNS-Abrufmusters
3. Bestimmung der *k*-Identifizierbarkeit



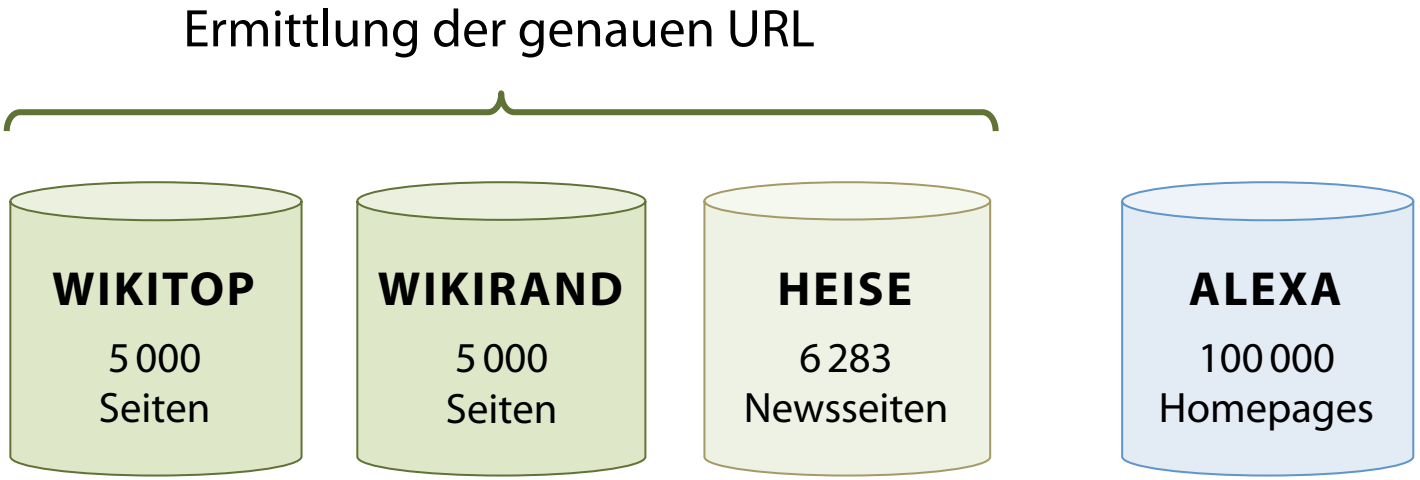
$k = 1$	> 99 %	80 % ←	Seiten mit einzigartigem Abrufmuster
$k \leq 5$	> 99 %	94 % ←	Seiten mit Abrufmustern, die jeweils höchstens auf 5 Seiten auftreten

Empirische Untersuchung



$k = 1$	> 99 %	80 %	63 %
$k \leq 5$	> 99 %	94 %	76 %

Empirische Untersuchung



$k = 1$	> 99 %	80 %	63 %	> 99,9 %
$k \leq 5$	> 99 %	94 %	76 %	> 99,9 %

A

Ermittlung der besuchten Webseiten

genauer als gedacht

B 

Identifizierung der
benutzten Software

Motivation: Nachweis der Beteiligung
an kriminellen Handlungen

Ergebnis: gängige Betriebssysteme und
Browser allein anhand ihres Verhaltens
identifizierbar

A 

Ermittlung der
besuchten Webseiten

B 

Identifizierung der
benutzten Software

C 

Techniken zum
Selbstdatenschutz

Schutz der Identität
des Nutzers

Verbergen der
wahren Interessen

Range Querys

Existierende Ansätze für DNS
ungeeignet oder unsicher

Schutz der Identität
des Nutzers

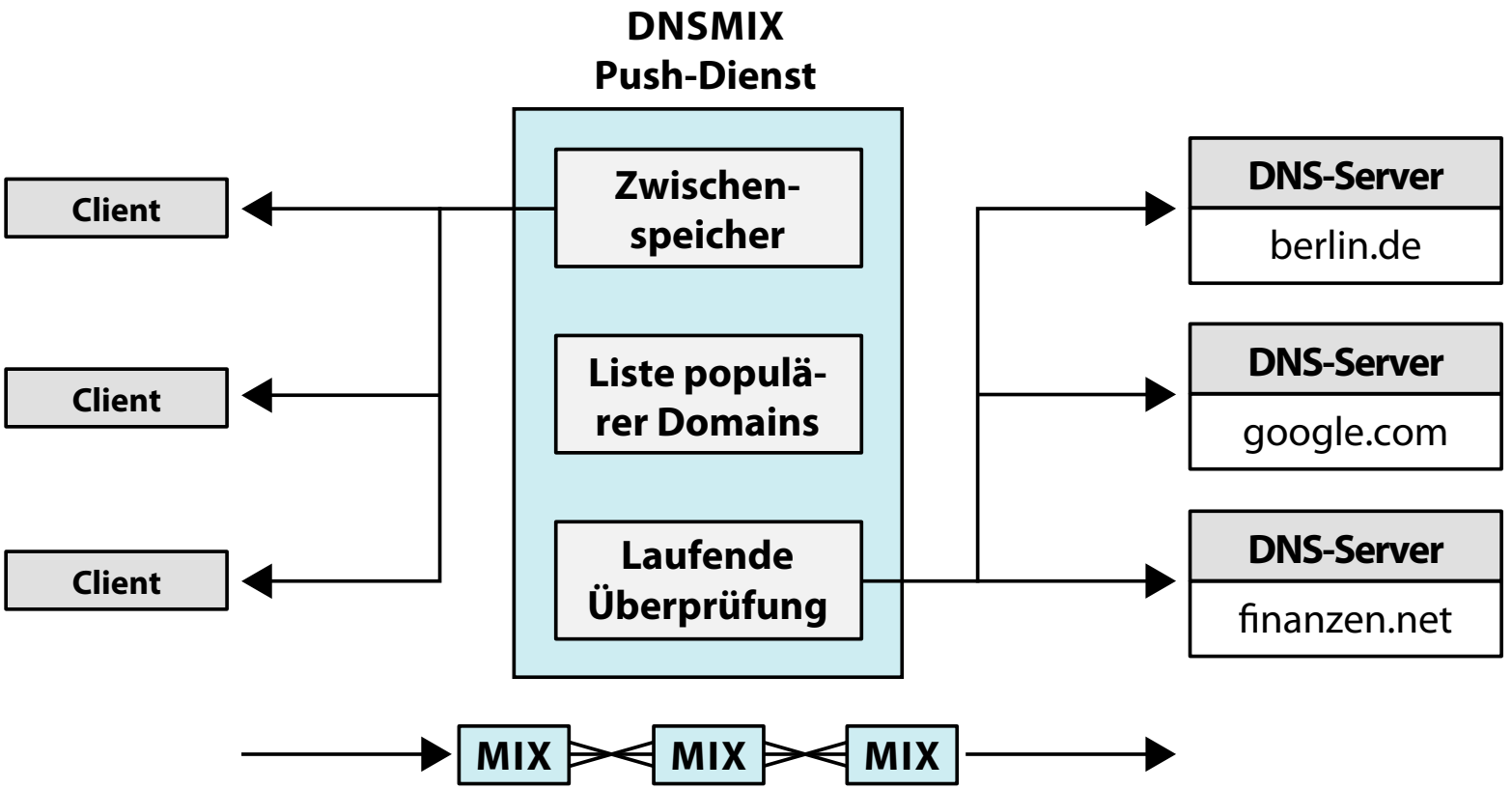
Verbergen der
wahren Interessen

Range Querys

Mit **DNSMIX** lässt sich
Unbeobachtbarkeit erreichen!

Idee von DNSMIX

populäre DNS-Einträge automatisch an alle Nutzer senden

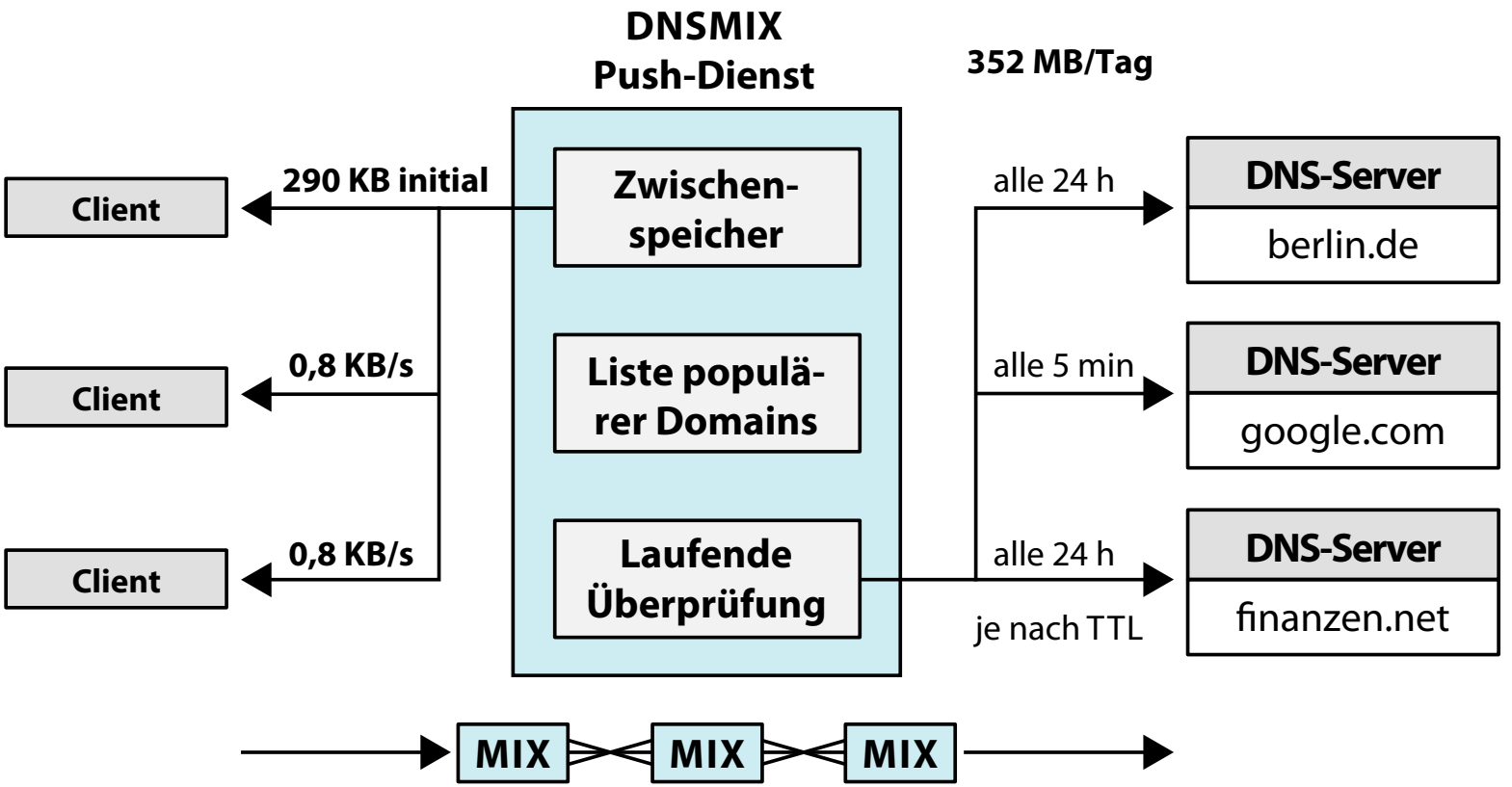


Empirische Untersuchung

Anfragen von 2082 Nutzern

Pushen von 10.000 populären Domains

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



A 

Ermittlung der
besuchten Webseiten

B 

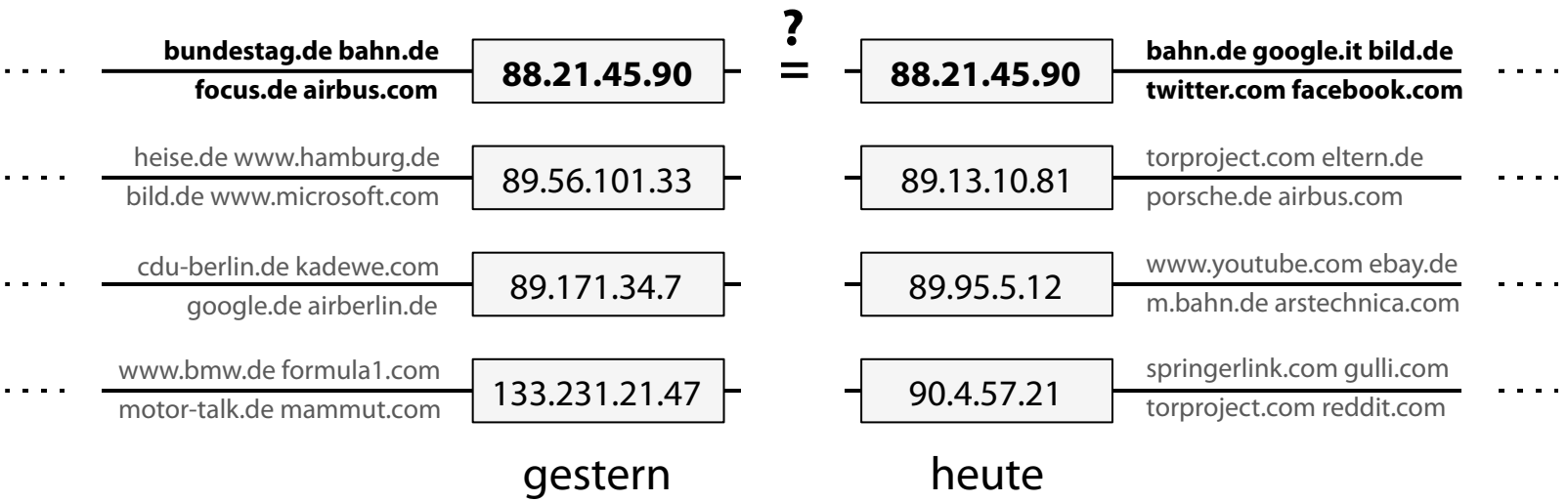
Identifizierung der
benutzten Software

C 

Techniken zum
Selbstdatenschutz

Bislang ignoriertes Problem

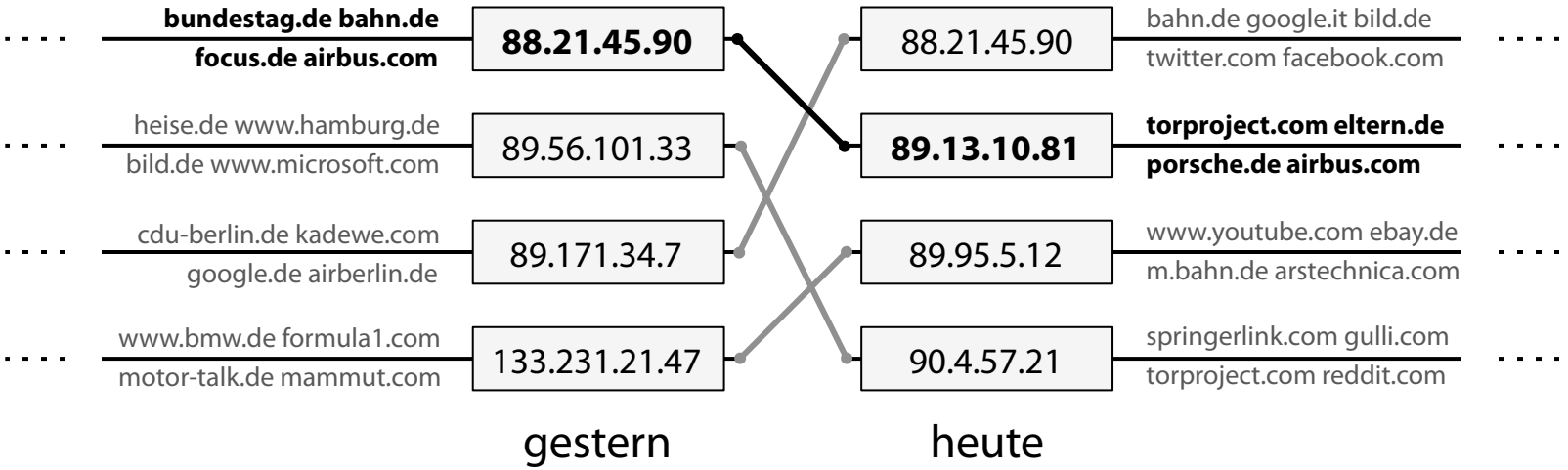
Beobachtung von Nutzern mit dynamischen IP-Adressen



D

Verhaltensbasierte Verkettung von Sitzungen

erlaubt Tracking von Nutzern mit dynamischen IP-Adressen

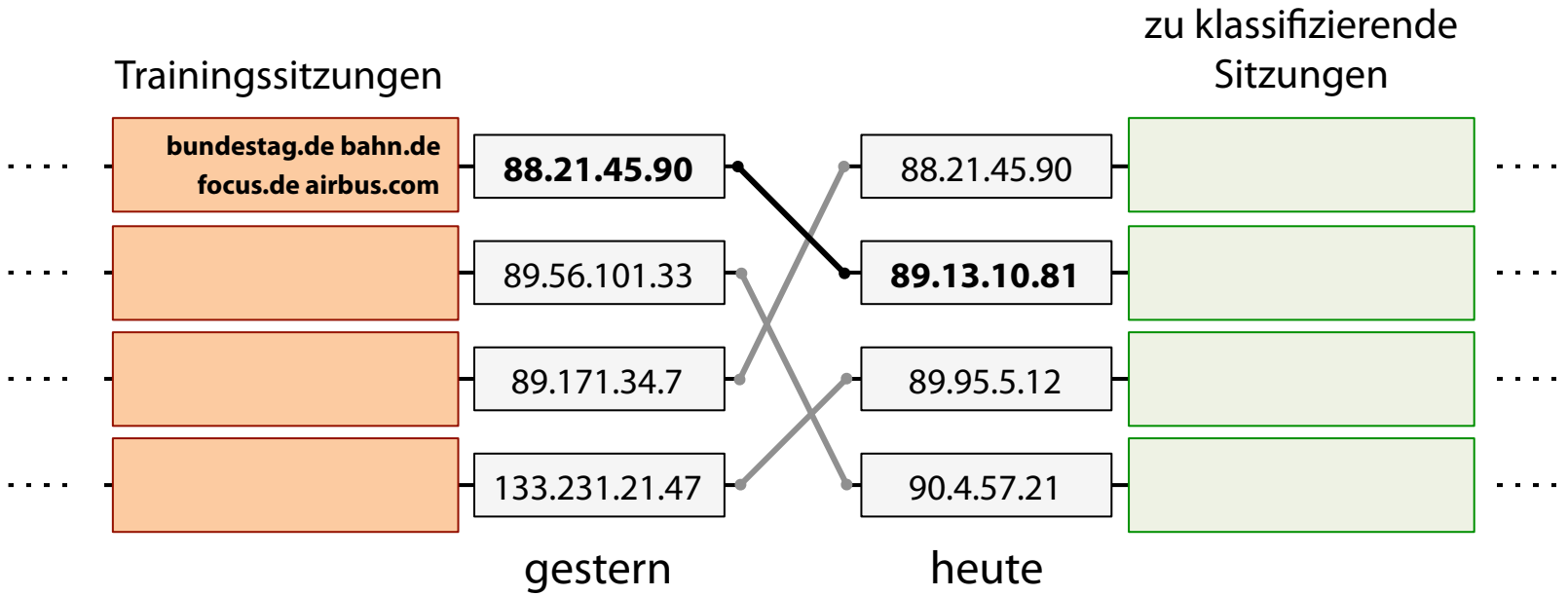


Ausnutzung menschlicher Eigenschaften

individuelle Vorlieben

tägliche Routine

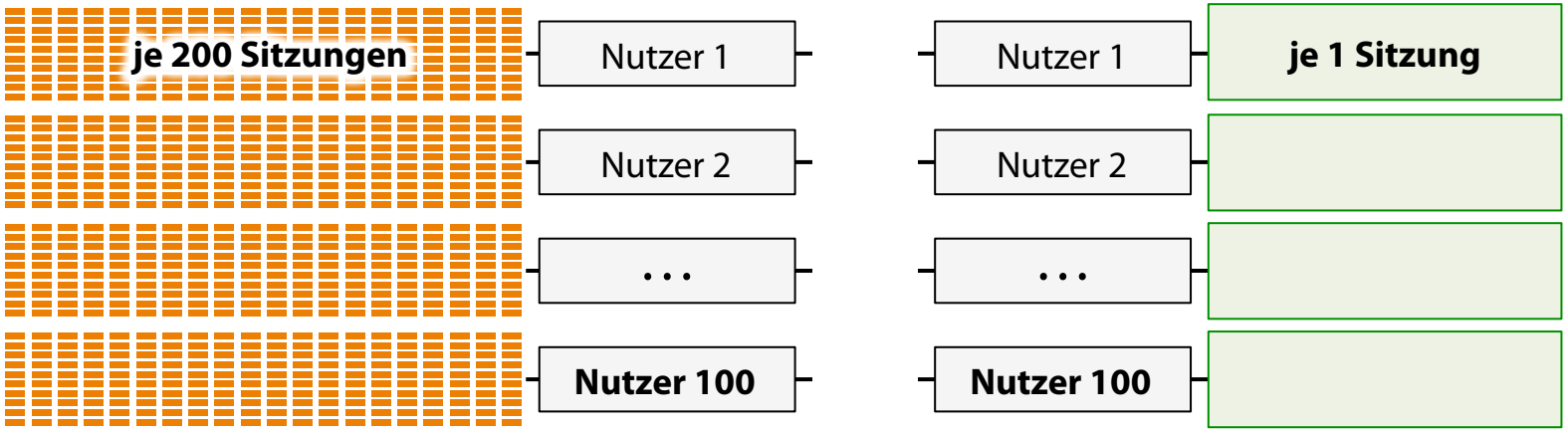
Trainieren eines Klassifikators zur Wiedererkennung



Verhaltensbasierte Verkettung
einzelner Sitzungen im Open-World-Szenario
noch nicht erforscht

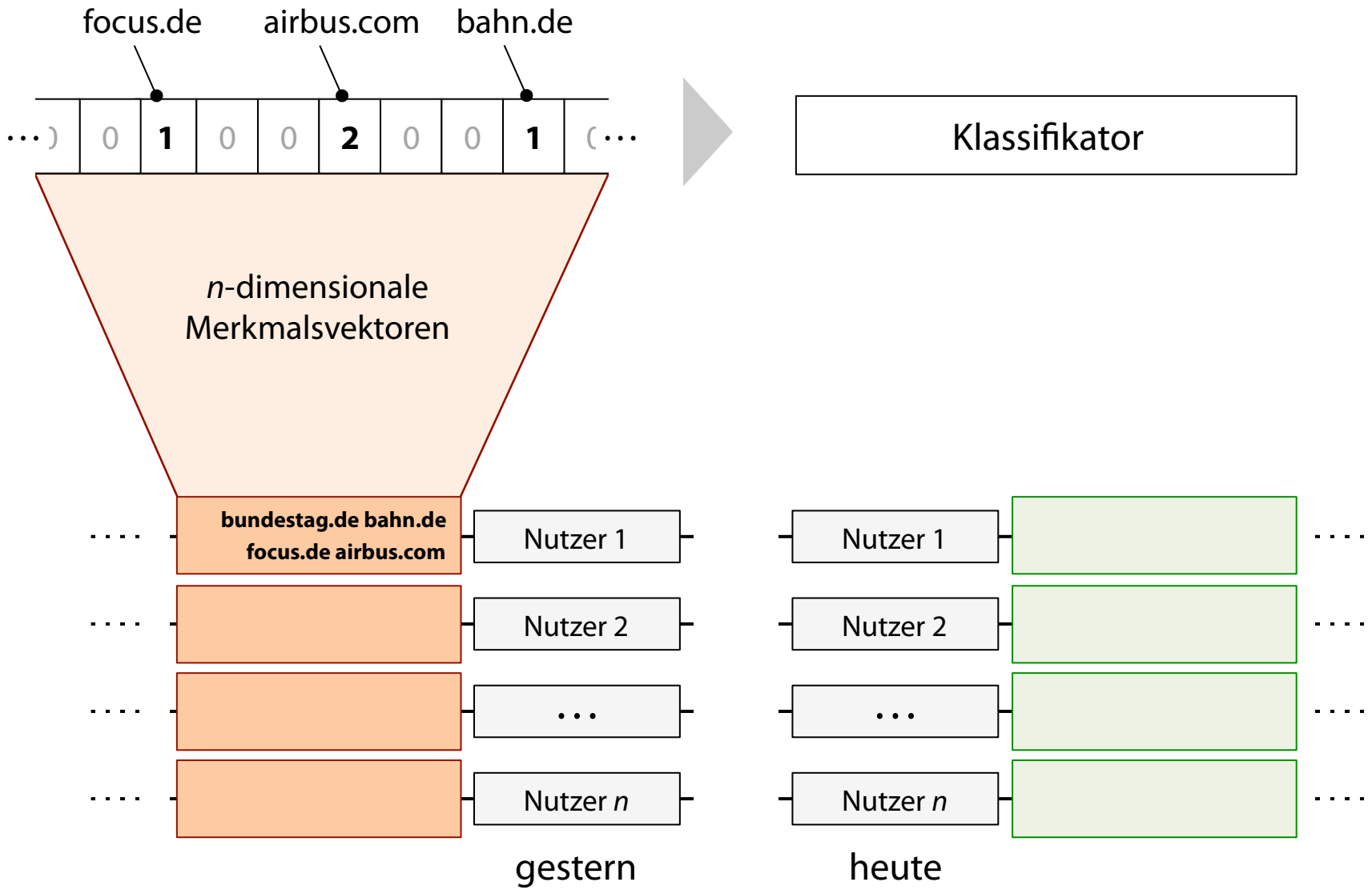
Reicht **eine** Traingssitzung?
Genauigkeit bei großen **fluktuierenden** Nutzergruppen?

Closed-World-Untersuchung von Yang (2010)

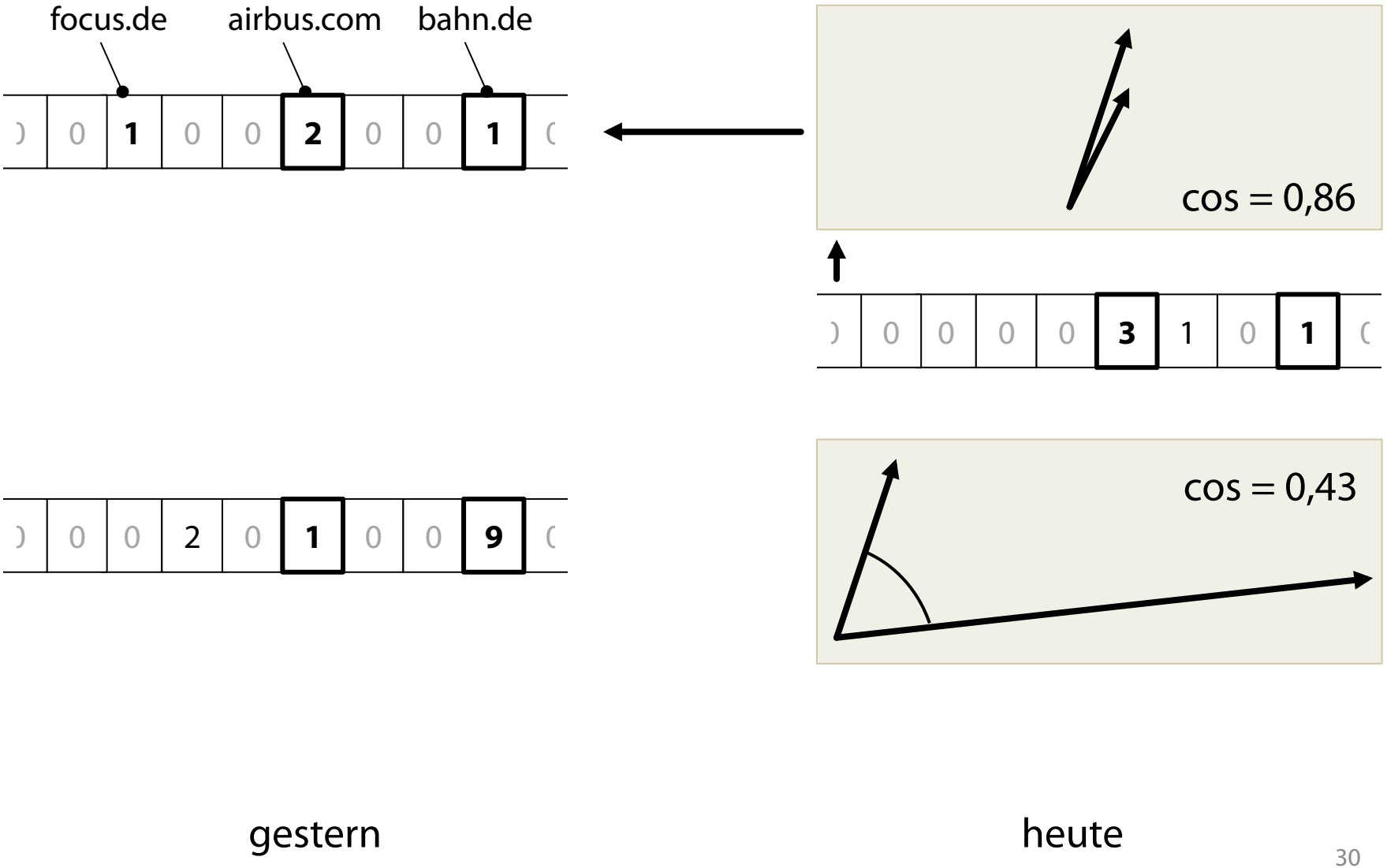


62% Genauigkeit

Konstruktion des Verkettungsverfahrens



Funktionsweise des verwendeten Klassifikators



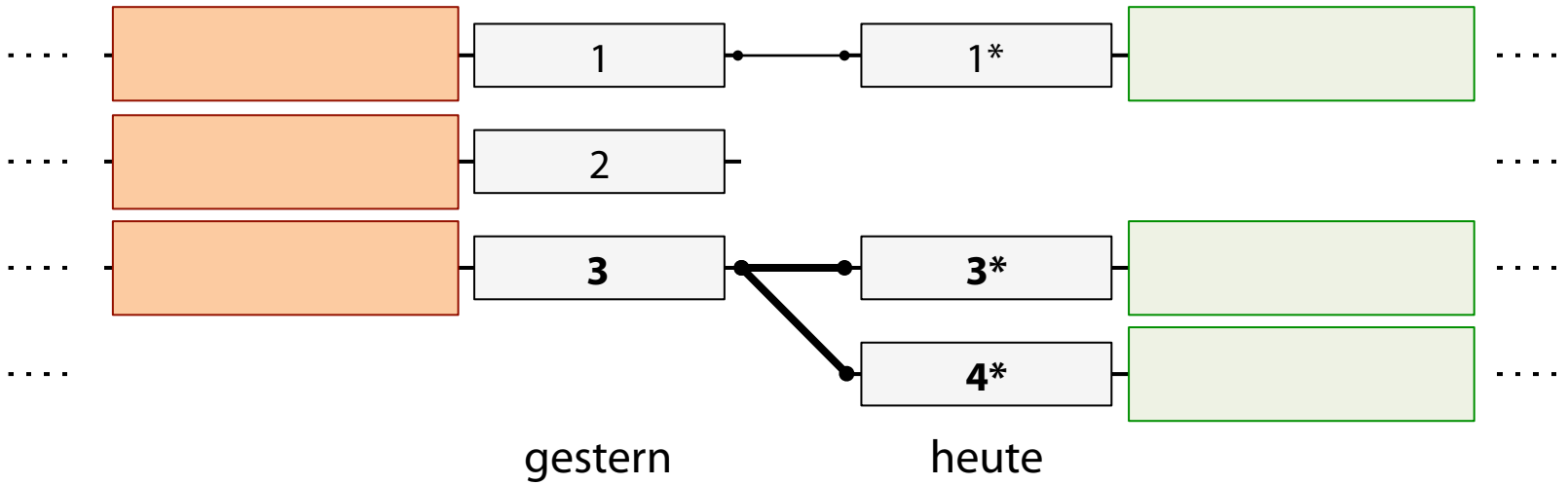
Empirische Untersuchung

1. Implementierung mit MapReduce
2. Erhebung eines DNS-Datensatzes
3. Experimente im Open-World-Szenario

Datensatz 2M

2 Monate
3862 Nutzer
5 Mio. Domains

Beobachtung: Fehlentscheidungen wegen Nutzerfluktuation



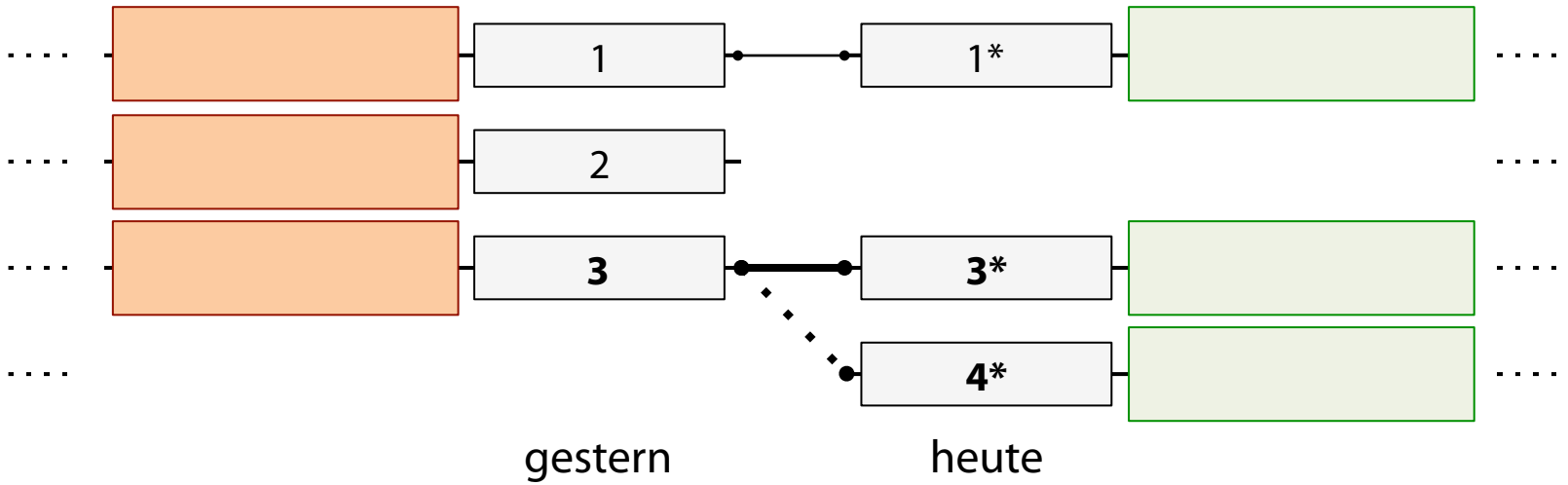
Standardverfahren

75 % Genauigkeit

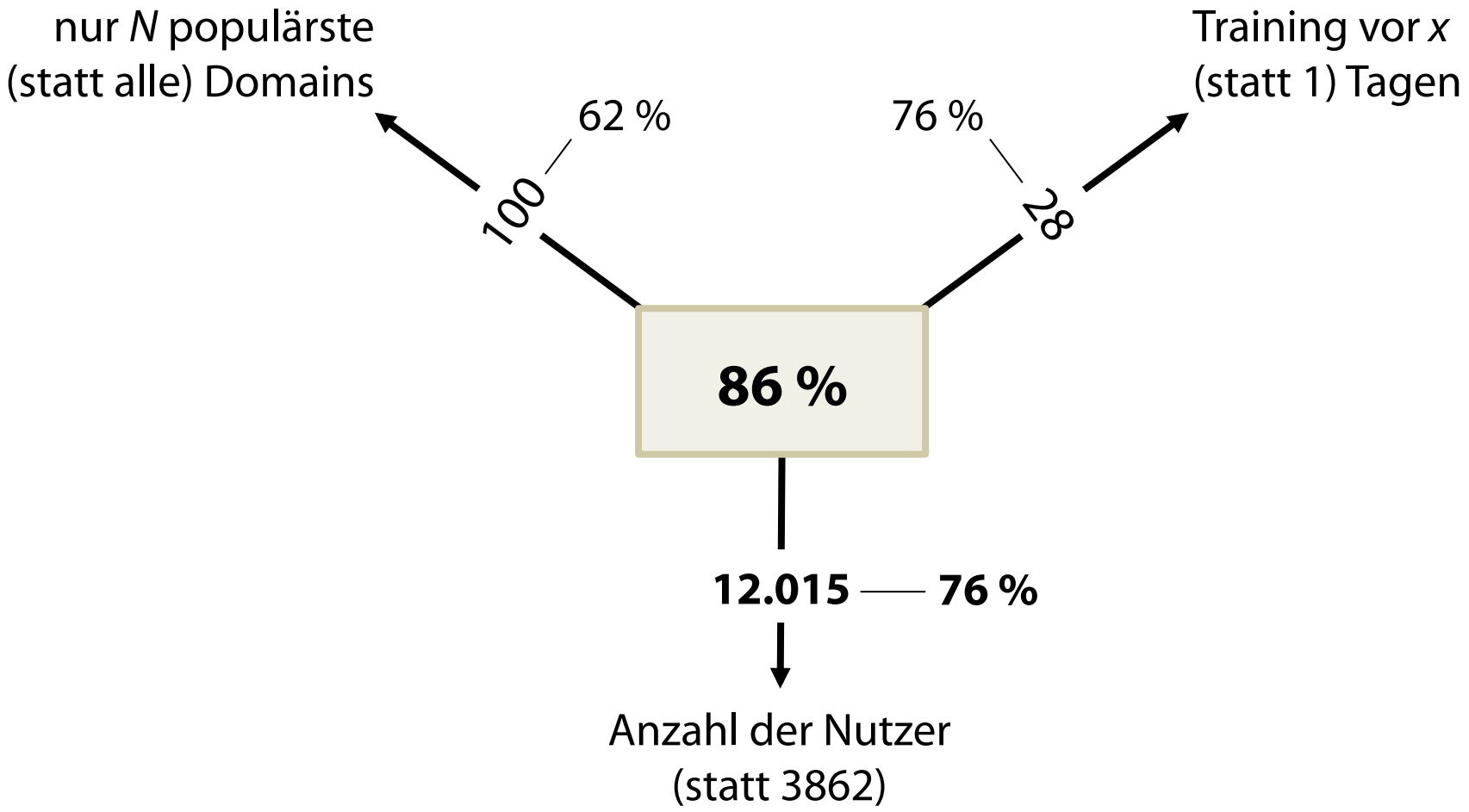
nach Optimierung

86 % Genauigkeit

Optimierung: Ermittlung der ähnlichsten Sitzung



Verkettung erweist sich als robust
gelingt auch unter widrigen Umständen



Einsatz auch unabhängig vom DNS möglich

Verhaltensbasierte Verkettung
erfolgt rein passiv, ist also nicht erkennbar

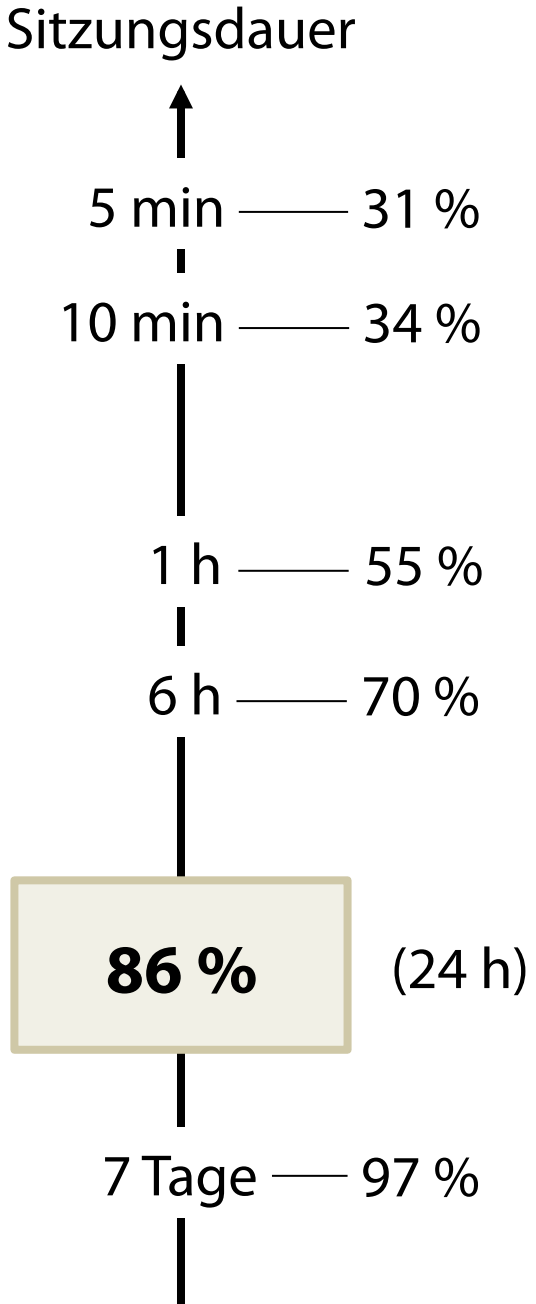
D

Verhaltensbasierte
Verkettung von Sitzungen

E

Schutz vor Tracking durch
verhaltensbasierte Verkettung

ist nicht nur für Nutzer öffentlicher
DNS-Server von Interesse



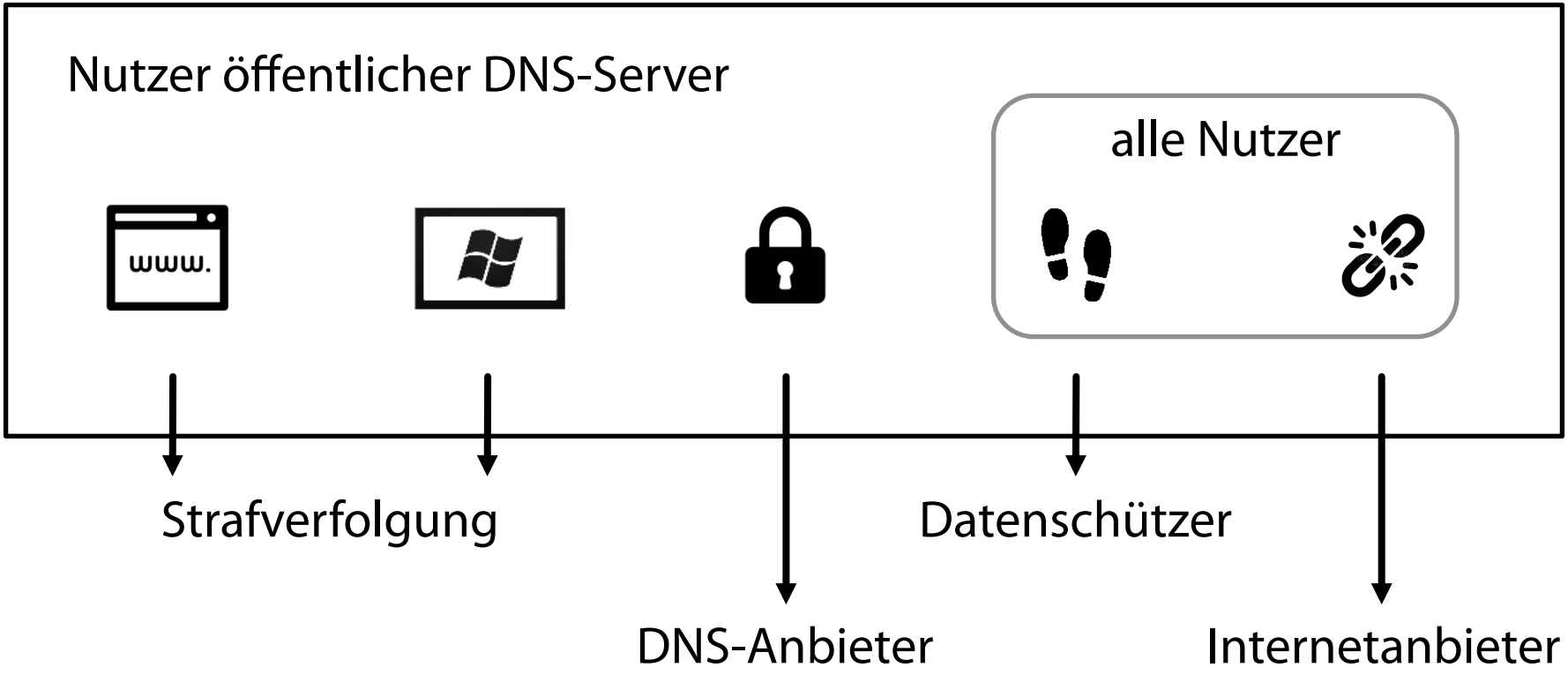
IP-Adresse häufig wechseln

heute: Komforteinbußen

Chance

»Privacy by Default«
mit IPv6

Zusammenfassung und Relevanz der Ergebnisse



Beobachtungsmöglichkeiten
im Domain Name System

… **umfangreicher und genauer** als zu erwarten wäre;

DNSMIX: **dienstspezifische Konzepte** lohnenswert;

unbemerktes Tracking – auch abseits von DNS – möglich;

komfortabler Schutz durch **häufigen Adresswechsel** mit IPv6.




Generalisierbarkeit der Ergebnisse?

Konkrete Handlungsempfehlungen?

Weitere Experimente und Future Work?

http://dhgo.to/castfolien bzw. /casthandout