Evaluating the Security of a DNS Query Obfuscation Scheme for Private Web Surfing

Dominik Herrmann · Max Maaß · Hannes Federrath

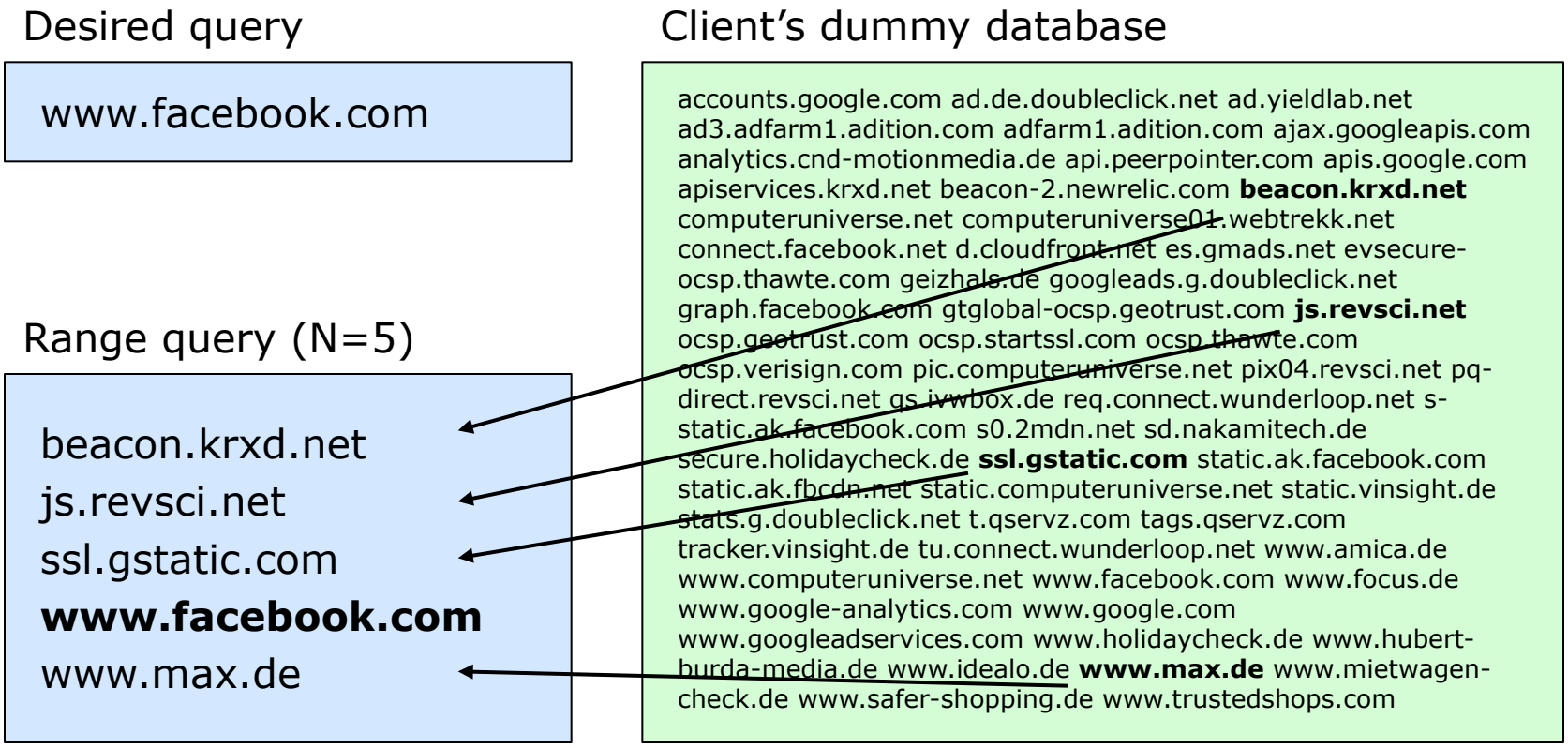
http://svs.informatik.uni-hamburg.de/

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

IFIP SEC 2014, Marrakech, June 3, 2014

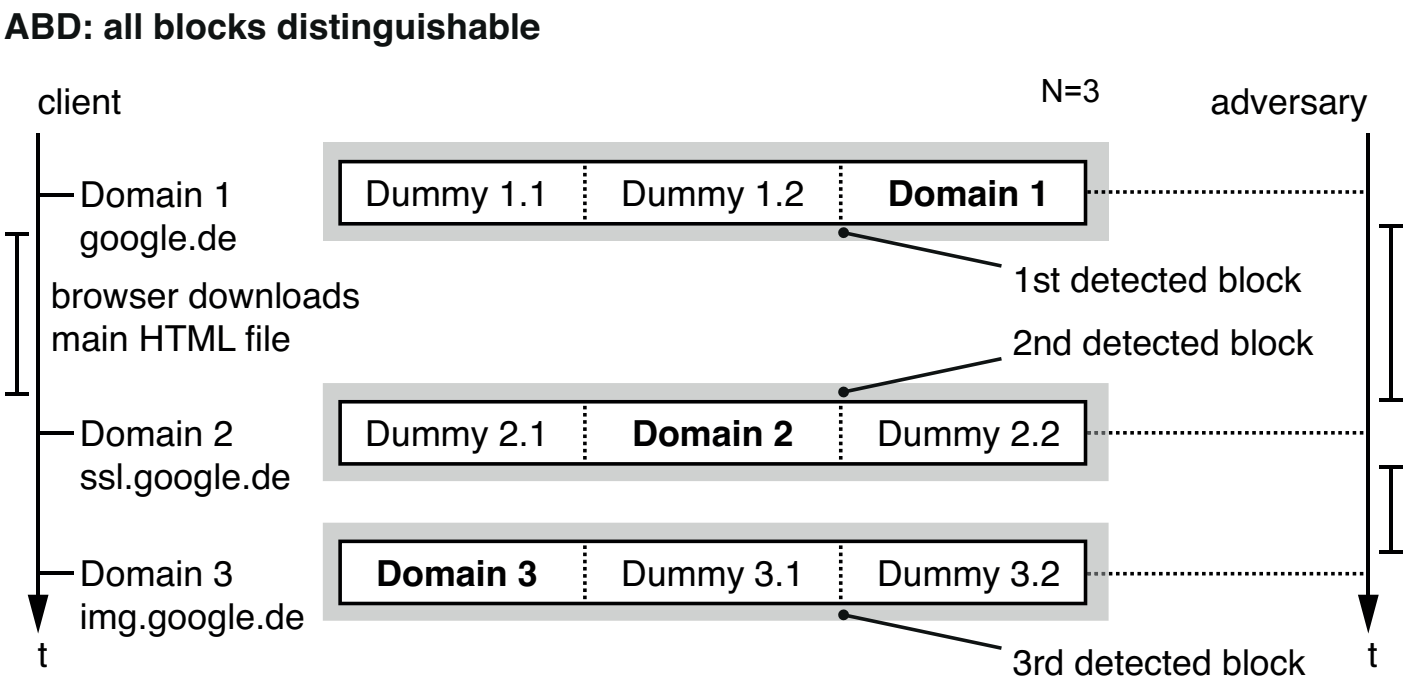
The Range Query scheme by Zhao et al. (2007)

- What? Protect DNS users from a curious DNS resolver, e.g. Google Public DNS (adversary)
- How? User wants to resolve 1 DNS name and randomly adds N-1 dummy queries from a dummy database



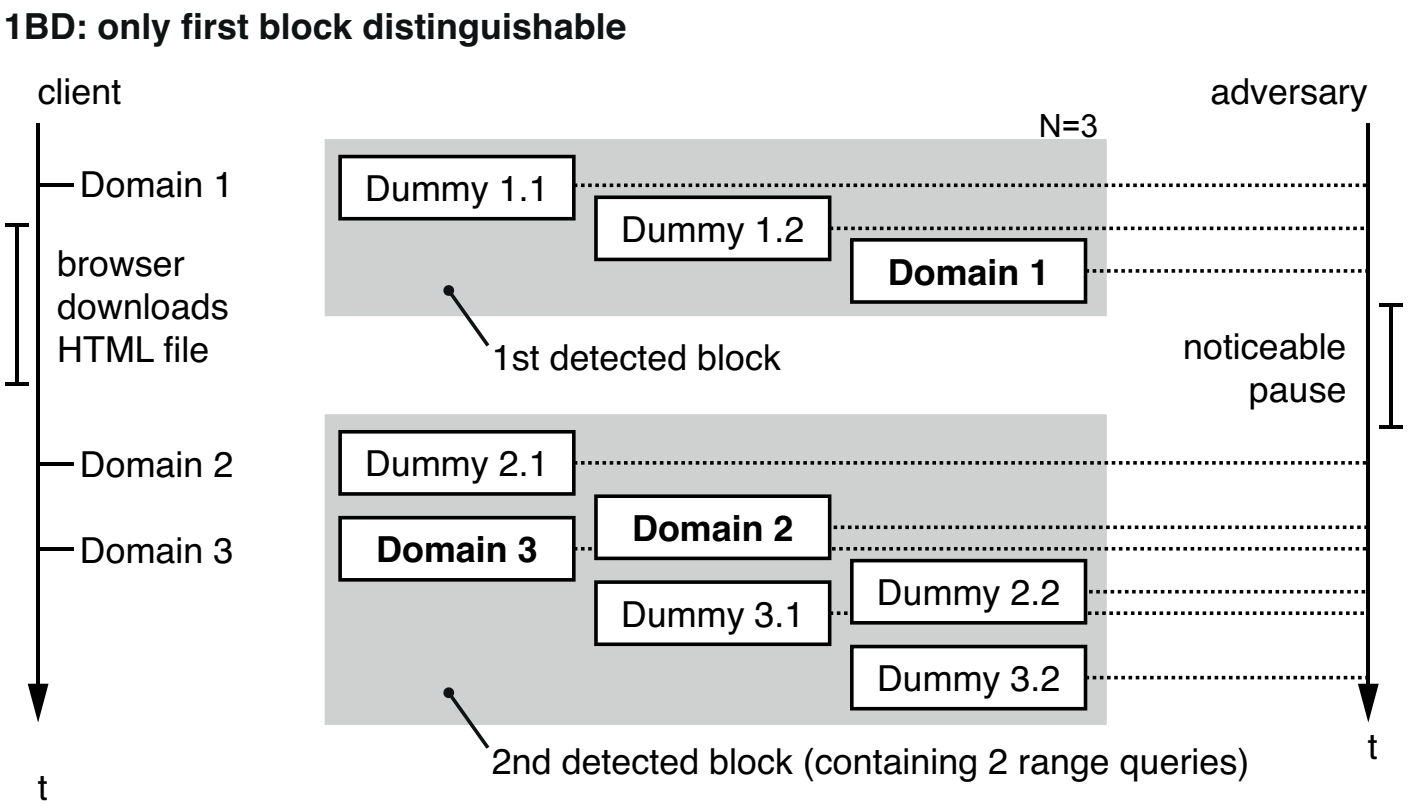
Download of a website from the view of the adversary

- Browsers may send multiple DNS queries after loading the main HTML file (embedded images, CSS, … from different domains)
- Each DNS range query would be issued within a single packet
- Adversary can distinguish individual queries
- This is a oversimplification (ABD model)



Download of a website from the view of the adversary

- Standard DNS can only resolve 1 query per message
- Client has to send multiple DNS queries for each range query
- Consequence: Adversary can only isolate queries from first block
- Remaining queries are mixed up in a second block (1BD model)



Example: Semantic intersection attack (1BD model, N=3)

- User downloads wikipedia.org
- N-1 randomly choosen queries

Adversary's observation

1st block N=3

www.web.de
en.wikipedia.org
www.google.com

2nd block

bits.wikimedia.org
de.yahoo.com
adimg.uimserv.net
c.fsdn.com
upload.wikimedia.org
s.ytimg.com

Adversary's pattern database

web.de

www.web.de cdn.flashtalking.com
ads.ivwbox.de adimg.uimserv.net

slashdot.org

www.slashdot.org beta.slashdot.org
c.fsdn.com c5.zedo.com a.c.ooyala.com

wikipedia.org

en.wikipedia.org bits.wikimedia.org
upload.wikimedia.org

google.com

www.google.com ssl.gstatic.com

...

Example: Semantic intersection attack (1BD model, N=3)

- User downloads wikipedia.org
- N-1 randomly choosen queries

Adversary's observation:

1st block N=3

www.web.de
en.wikipedia.org
www.google.com

2nd block

bits.wikimedia.org
de.yahoo.com
adimg.uimserv.net
c.fsdn.com
upload.wikimedia.org
s.ytimg.com

Adversary's pattern database:

web.de

www.web.de cdn.flashtalking.com
ads.ivwbox.de **adimg.uimserv.net**

2 out of 4

slashdot.org

www.slashdot.org beta.slashdot.org
c.fsdn.com c5.zedo.com a.c.ooyala.com

wikipedia.org

en.wikipedia.org bits.wikimedia.org
upload.wikimedia.org

google.com

www.google.com ssl.gstatic.com

...

Example: Semantic intersection attack (1BD model, N=3)

- User downloads wikipedia.org
- N-1 randomly choosen queries

Adversary's observation

1st block N=3

www.web.de
en.wikipedia.org
www.google.com

2nd block

bits.wikimedia.org
de.yahoo.com
adimg.uimserv.net
c.fsdn.com
upload.wikimedia.org
s.ytimg.com

Adversary's pattern database

web.de

www.web.de cdn.flashtalking.com
ads.ivwbox.de adimg.uimserv.net

2 out of 4

slashdot.org

www.slashdot.org beta.slashdot.org
c.fsdn.com c5.zedo.com a.c.ooyala.com

not part of 1st block

wikipedia.org

en.wikipedia.org bits.wikimedia.org
upload.wikimedia.org

google.com

www.google.com ssl.gstatic.com

...

Example: Semantic intersection attack (1BD model, N=3)

- User downloads wikipedia.org
- N-1 randomly choosen queries

Adversary's observation

1st block N=3

www.web.de
en.wikipedia.org
www.google.com

2nd block

bits.wikimedia.org
de.yahoo.com
adimg.uimserv.net
c.fsdn.com
upload.wikimedia.org
s.ytimg.com

Adversary's pattern database

web.de

www.web.de cdn.flashtalking.com
ads.ivwbox.de adimg.uimserv.net

2 out of 4

slashdot.org

www.slashdot.org beta.slashdot.org
c.fsdn.com c5.zedo.com a.c.ooyala.com

not part of 1st block

wikipedia.org

en.wikipedia.org bits.wikimedia.org
upload.wikimedia.org

full match

google.com

www.google.com ssl.gstatic.com

...

Example: Semantic intersection attack (1BD model, N=3)

- User downloads wikipedia.org
- N-1 randomly choosen queries

Adversary's observation

1st block N=3

www.web.de
en.wikipedia.org
www.google.com

2nd block

bits.wikimedia.org
de.yahoo.com
adimg.uimserv.net
c.fsdn.com
upload.wikimedia.org
s.ytimg.com

Adversary's pattern database

web.de

www.web.de cdn.flashtalking.com
ads.ivwbox.de adimg.uimserv.net

2 out of 4

slashdot.org

www.slashdot.org beta.slashdot.org
c.fsdn.com c5.zedo.com a.c.ooyala.com

not part of 1st block

wikipedia.org

en.wikipedia.org bits.wikimedia.org
upload.wikimedia.org

full match

google.com

www.google.com ssl.gstatic.com

1 out of 2

...

Example: Semantic intersection attack (1BD model, N=3)

- User downloads wikipedia.org
- N-1 randomly choosen queries

Adversary's observation

1st block N=3

www.web.de ⊘
en.wikipedia.org ✓
www.google.com ⊘

2nd block

bits.wikimedia.org
de.yahoo.com
adimg.uimserv.net
c.fsdn.com
upload.wikimedia.org
s.ytimg.com

Adversary's pattern database

web.de

www.web.de cdn.flashtalking.com
ads.ivwbox.de adimg.uimserv.net

2 out of 4

slashdot.org

www.slashdot.org beta.slashdot.org
c.fsdn.com c5.zedo.com a.c.ooyala.com

not part of 1st block

wikipedia.org

en.wikipedia.org bits.wikimedia.org
upload.wikimedia.org

✓ full match

google.com

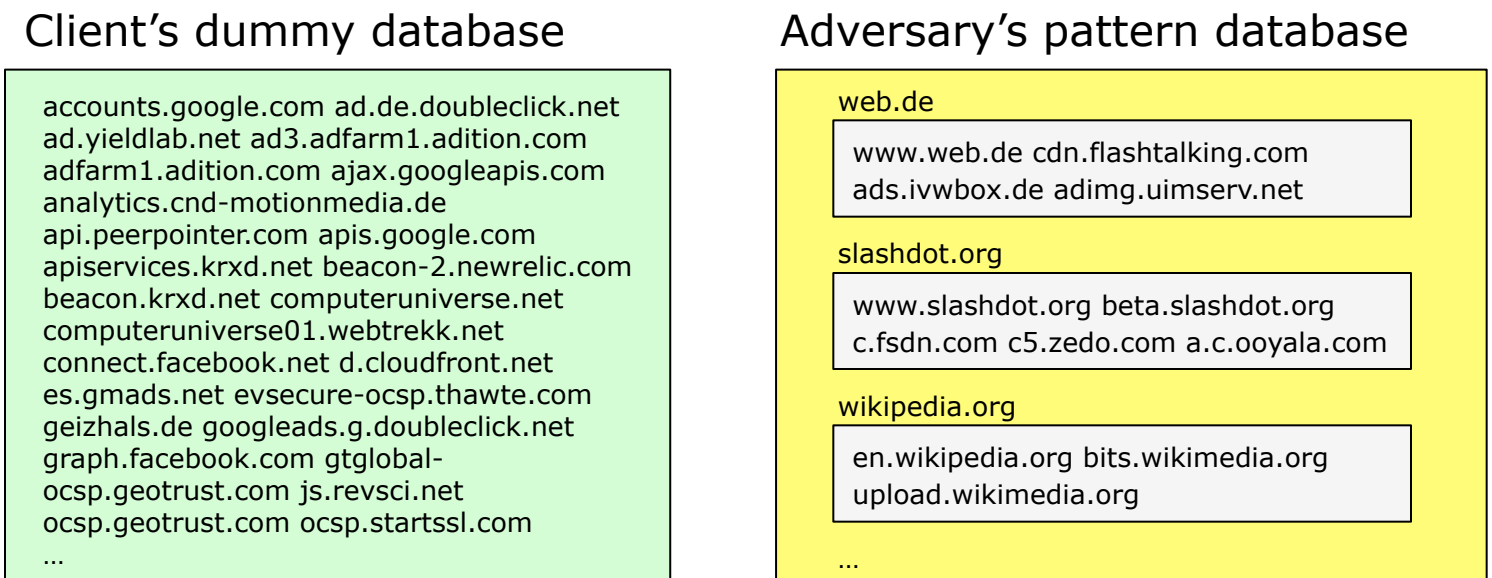
www.google.com ssl.gstatic.com

1 out of 2

...

Evaluation of semantic intersection attack (1BD model)

- Adversary's pattern database
  - Query patterns of top 100 000 websites of »Alexa Toplist«
  - Cleaned dataset: 92 880 query patterns
  - Avg. pattern length: 13 queries
  - Longest pattern: 315 queries
  - Number of DNS names in pattern database: |Q|=216 925
- Client's dummy database
  - assumed a subset of Adversary's pattern database



Evaluation of semantic intersection attack (1BD model)

- Adversary's pattern database
  - Query patterns of top 100 000 websites of »Alexa Toplist«
  - Cleaned dataset: 92 880 query patterns
  - Avg. pattern length: 13 queries
  - Longest pattern: 315 queries
  - Number of DNS names in pattern database: |Q|=216 925
- Client's dummy database
  - assumed a subset of Adversary's pattern database

- Experiment 1: Influence of range query size
  - Number of DNS names in dummy database: S=|Q|=216 925
  - Size of range query: N={10, 50, 100}
- Experiment 2: Influence of dummy database size
  - Size of dummy database: S={2 000, 20 000, 200 000}
  - Fixed size of range query: N=50

Influence of range query size N (1BD model) S=|Q|=216,925

proportion of websites for which the adversary obtained only a single match (the correct website)

the number of candidate sites obtained by the adversary was smaller or equal to median(k) for 50% of the analyzed websites in the experiment

N	1-identifiable	≤ 5-identifiable	median(k)	max(k)
10	62 %	99 %	1	6
50	8 %	88 %	3	14
100	1 %	43 %	6	18

proportion of sites where the adversary obtained at most 5 matches (including the correct website)

max(k) relates to the worst-case level of uncertainty of the adversary in the experiment

Improvement: 1BD-optimized attack to eliminate patterns with wrong length: about 80% of websites are 1-identifiable for N=100

Influence of dummy database size S (1BD model) N=50

proportion of websites for which the adversary obtained only a single match (the correct website)

the number of candidate sites obtained by the adversary was smaller or equal to median(k) for 50% of the analyzed websites in the experiment

S	1-identifiable	≤ 5-identifiable	median(k)	max(k)
2 000	19 %	92 %	3	14
20 000	16 %	95 %	3	11
200 000	9 %	88 %	3	13

proportion of sites where the adversary obtained at most 5 matches (including the correct website)

max(k) relates to the worst-case level of uncertainty of the adversary in the experiment

Countermeasures – DNS privacy improvements

- Instead of independently and randomly drawn dummy domains the client should use whole dummy patterns in consecutive ranges.

1st range query	2nd range query	3rd range query
www.web.de	bits.wikimedia.org	adimg.uimserv.net
en.wikipedia.org	ads.ivwbox.de	upload.wikimedia.org
www.slashdot.org	beta.slashdot.org	c.fsdn.com

- DNS over Tor; high avg. query latency 1.4 s [Fab+10]
- Low latency: Special-purpose DNS Mixes combined with broadcast push service of frequently asked domains [Fed+11]
- Alternative name resolution services based on DHT and P2P architectures [LT10]

[Fab+10] B. Fabian et al., Privately Waiting – A Usability Analysis of the Tor Anonymity Network, AMCIS 2010.

[Fed+11] H. Federrath et al., Privacy-Preserving DNS: Analysis of Broadcast, Range Queries and Mix-Based Protection Methods, ESORICS 2011.

[LT10] Y. Lu and G. Tsudik, Towards Plugging Privacy Leaks in the Domain Name System, P2P2010.

Summary and conclusion

- ≈ 80% of websites are 1-identifiable for N=100
  - 1BD-optimized attack
- ≈ 50% of websites are at least 3-identifiable for N=50
  - no influence: increasing dummy database size > 200 000

- Previously published evaluations of the range query approach may give a misleading sense of security
  - Results apply to web-browsing only