Analyzing the Internet Using DNS

Initial Bachelor Talk

Lennart Bader

Advisor: Dr. Oliver Hohlfeld

http://comsys.rwth-aachen.de

Aachen, 2016-12-12

Motivation

Domain name system

Motivation

Domain name system

- Central part of the Internet

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI
- Internet evolution: IPv6 deployment

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI
- Internet evolution: IPv6 deployment
- Security, Validity, ...

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI
- Internet evolution: IPv6 deployment
- Security, Validity, ...
- Some aspects already in current research:

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI
- Internet evolution: IPv6 deployment
- Security, Validity, ...
- Some aspects already in current research: DNSSec Analysis [IMC'14]

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI
- Internet evolution: IPv6 deployment
- Security, Validity, ...
- Some aspects already in current research:
  DNSSec Analysis [IMC'14], Webserver Security [IMC'14]

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI
- Internet evolution: IPv6 deployment
- Security, Validity, ...
- Some aspects already in current research:
  DNSSec Analysis [IMC'14], Webserver Security [IMC'14],
  DNS Validity [IMC'15], ...

Motivation

Domain name system

- Central part of the Internet
- Allows analyses on several aspects

Possibilities

- DNS "Junk" Analysis
- Support for ZMap Port Scans to enable SNI
- Internet evolution: IPv6 deployment
- Security, Validity, ...
- Some aspects already in current research:
  DNSSec Analysis [IMC'14], Webserver Security [IMC'14],
  DNS Validity [IMC'15], ...

Goal: Collect DNS data periodically for many domains

Seems easy...

foo.org

`dig`

Seems easy...

foo.org

AAAA
`dig`

Just needed $\approx 36ms$

Seems easy...

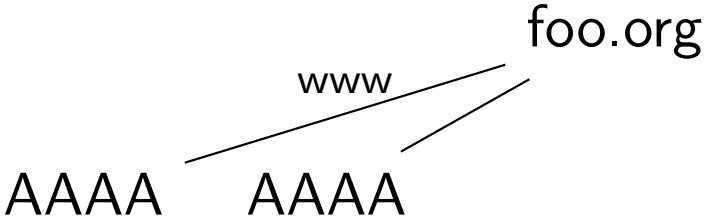
foo.org

AAAA
`dig`

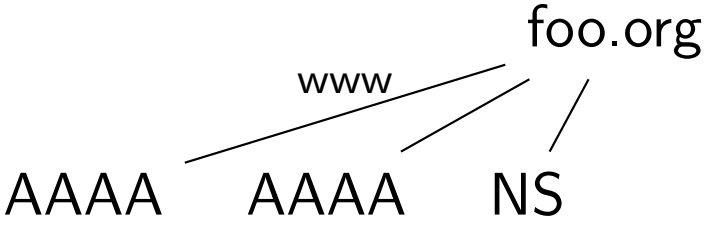
Just needed $\approx 36ms$

Where is the challenge?

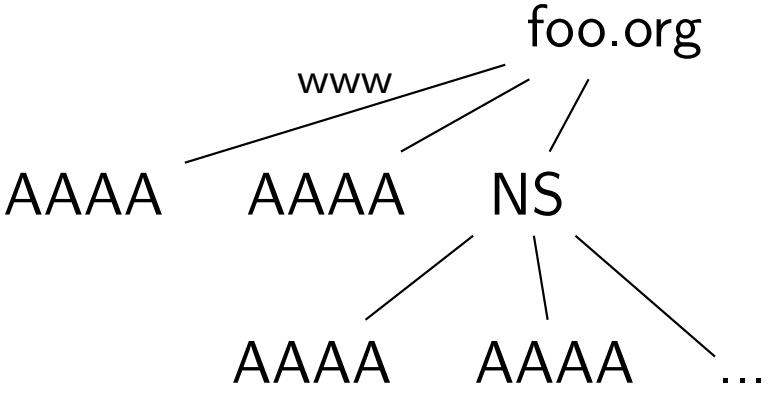
Seems easy...



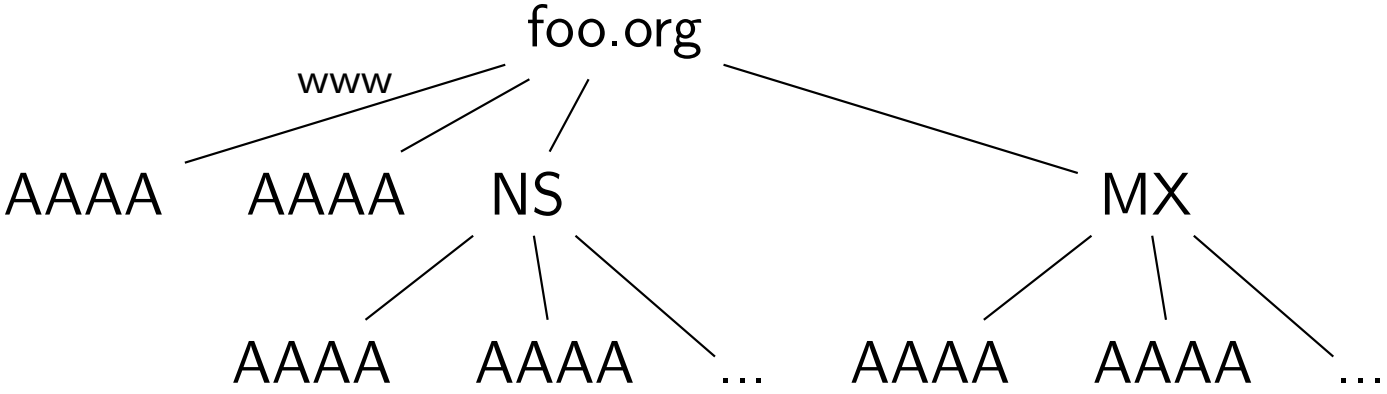
Seems easy...



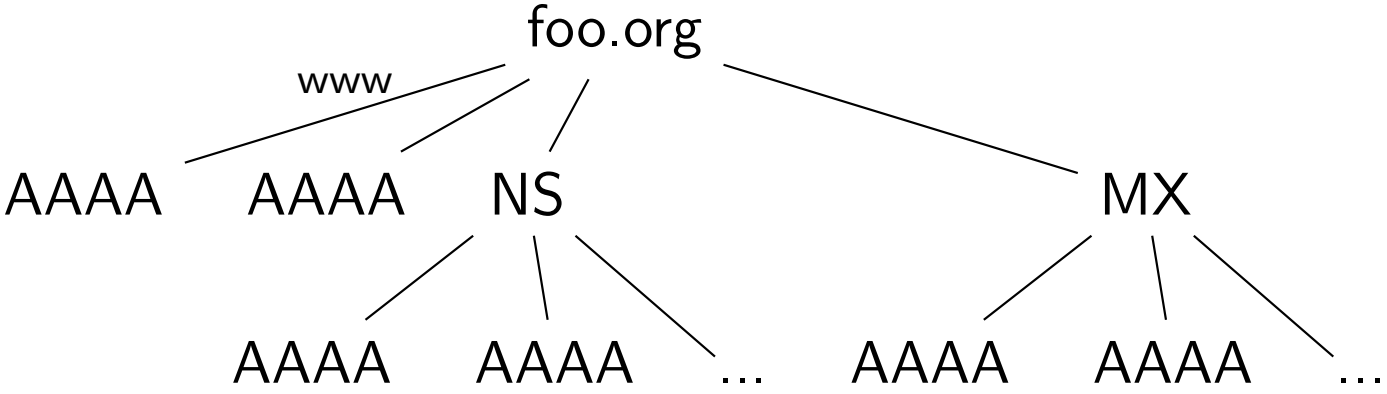
Seems easy...



Seems easy...

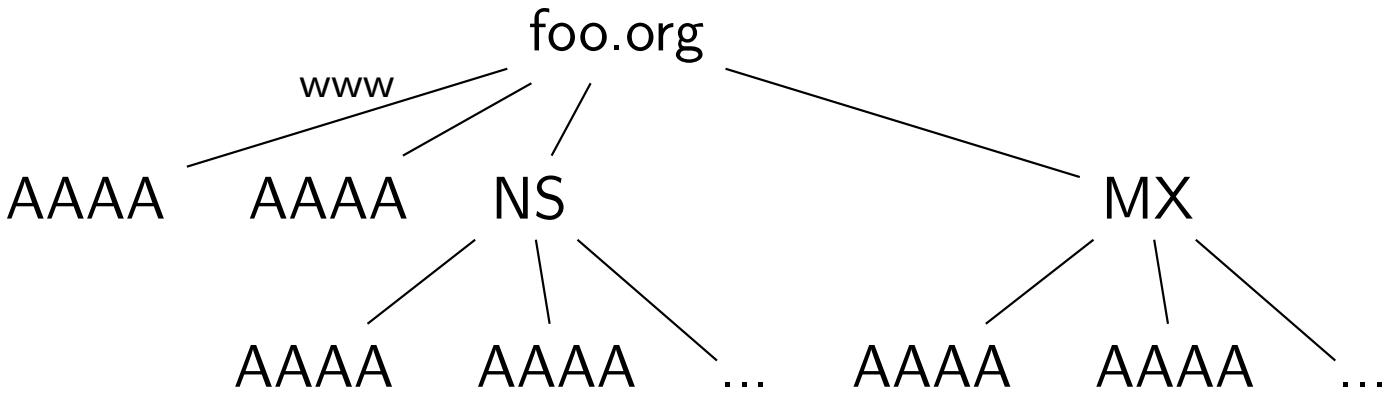


Seems easy...



Internet is huge

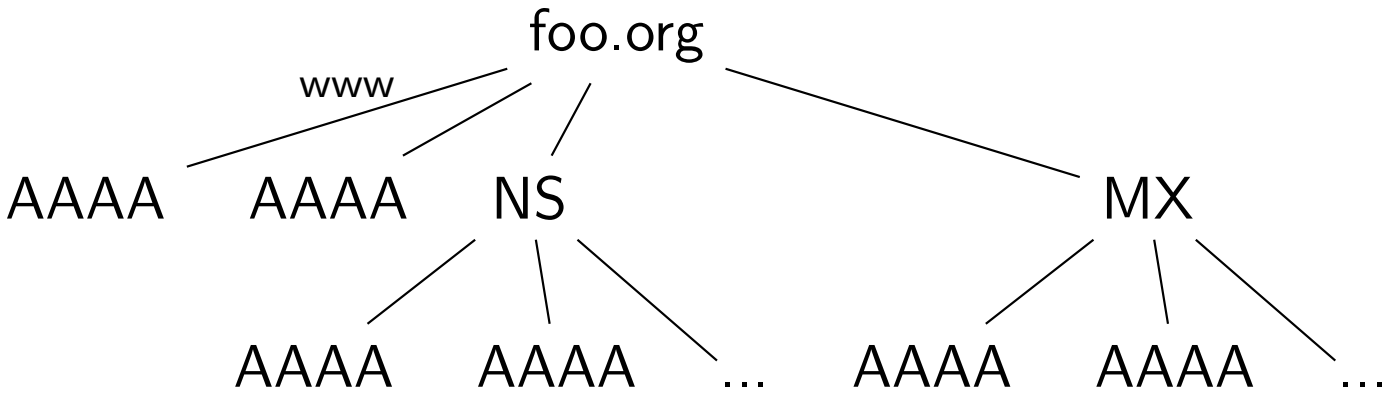
Seems easy...



Internet is huge

- Millions of domains
- Different analyses (not only IPv6)

Seems easy...

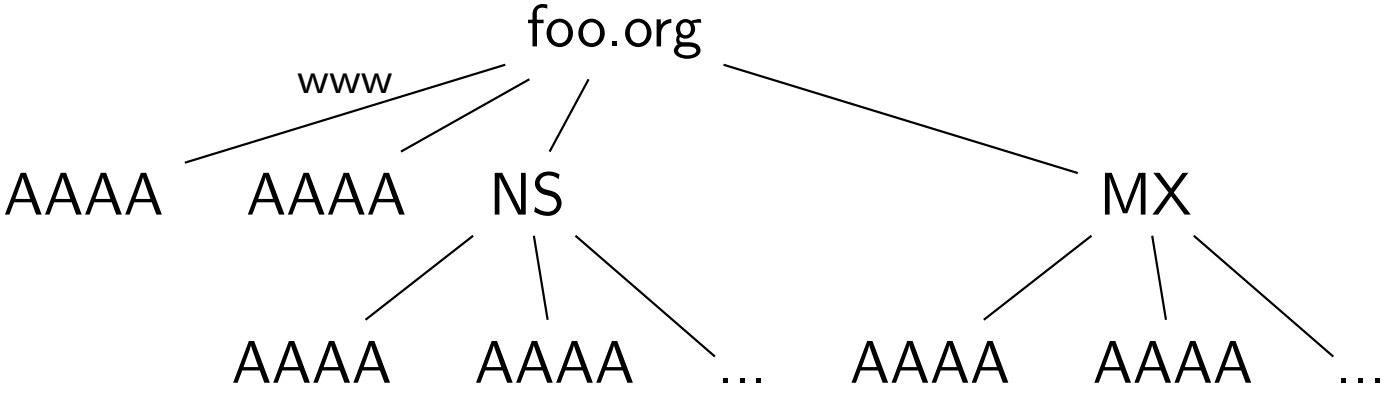


Internet is huge

- Millions of domains
- Different analyses (not only IPv6)

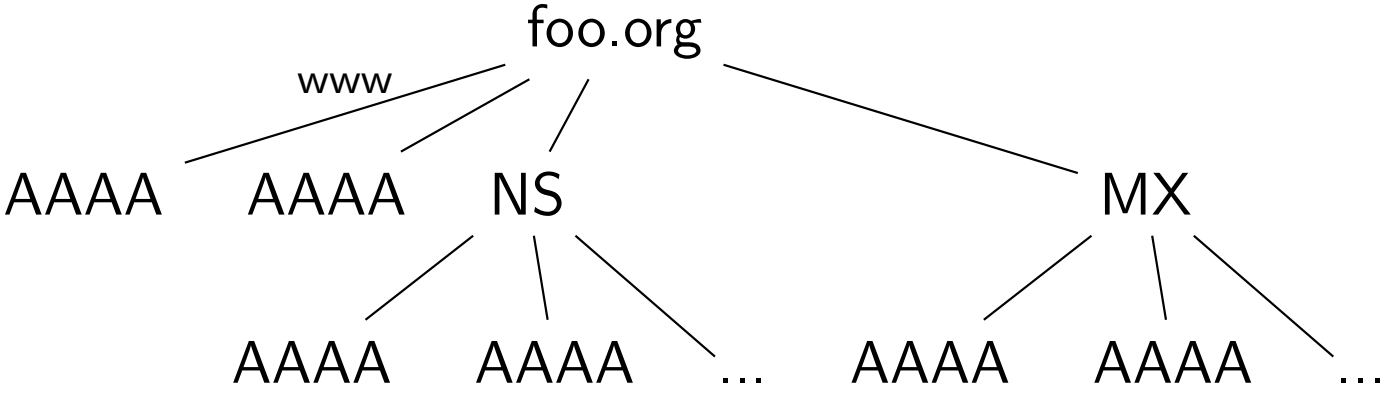
$\Rightarrow$ Multiple Queries per domain + Millions of domains

Seems easy...



$$N_{\text{Queries}} \geq (\underbrace{130M}_{\text{.com}} + \underbrace{15M}_{\text{.net}} + \underbrace{10M}_{\text{.org}} + \underbrace{\ldots}_{\text{Other TLDs}}) * N_{\text{Queries per Domain}}$$

Seems easy...



$$N_{\text{Queries}} \geq (\underbrace{130M}_{\text{.com}} + \underbrace{15M}_{\text{.net}} + \underbrace{10M}_{\text{.org}} + \underbrace{\ldots}_{\text{Other TLDs}}) * N_{\text{Queries per Domain}}$$

$$\Rightarrow N_{\text{Queries}} \geq 1.55 \text{ Billion}$$

Seems easy...

```
dig www.rwth-aachen.de
```

Seems easy...

```
dig www.rwth-aachen.de
```

- `www.rwth-aachen.de.  99054 IN A 137.226.107.63`

Seems easy...

```
dig www.rwth-aachen.de
```

- `www.rwth-aachen.de.  99054 IN A 137.226.107.63`
- Just needed $\approx 36ms$

Seems easy...

`dig www.rwth-aachen.de`

- `www.rwth-aachen.de.  99054 IN A 137.226.107.63`
- Just needed $\approx 36ms$

`dig www.rwth-aachen.de AAAA`

Seems easy...

`dig www.rwth-aachen.de`

- `www.rwth-aachen.de.  99054 IN A 137.226.107.63`
- Just needed $\approx 36ms$

`dig www.rwth-aachen.de AAAA`

- RWTH has no `AAAA` record...

Seems easy...

`dig www.rwth-aachen.de`

- `www.rwth-aachen.de.  99054 IN A 137.226.107.63`
- Just needed $\approx 36ms$

`dig www.rwth-aachen.de AAAA`

- RWTH has no `AAAA` record...
- Again $\approx 36ms$

Seems easy...

`dig www.rwth-aachen.de`

- `www.rwth-aachen.de.  99054 IN A 137.226.107.63`
- Just needed $\approx 36ms$

`dig www.rwth-aachen.de AAAA`

- RWTH has no `AAAA` record...
- Again $\approx 36ms$

Where is the challenge?

Analyzing one domain

E.g.: Get information about IPv6 compatibility

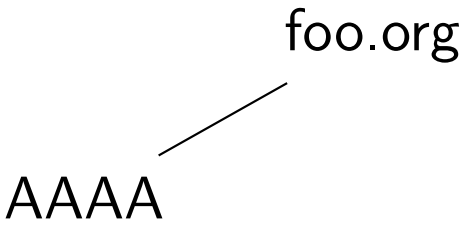
Analyzing one domain

E.g.: Get information about IPv6 compatibility

foo.org

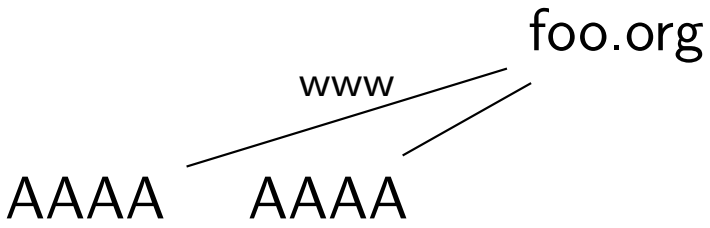
Analyzing one domain

E.g.: Get information about IPv6 compatibility



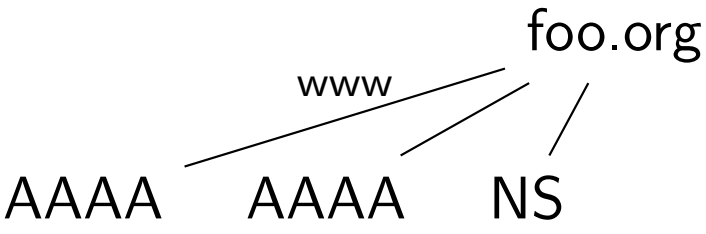
Analyzing one domain

E.g.: Get information about IPv6 compatibility



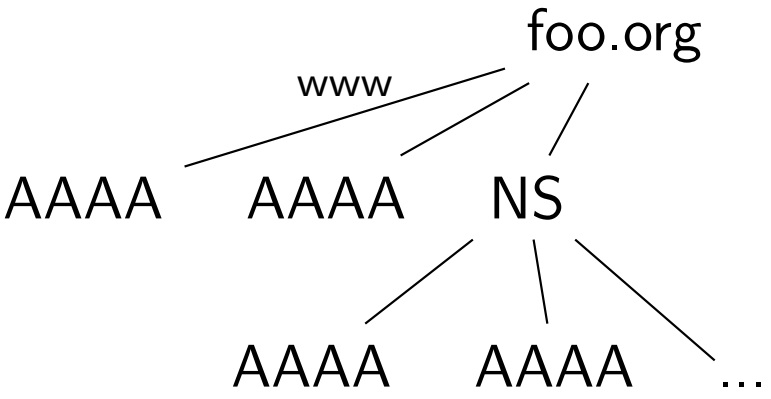
Analyzing one domain

E.g.: Get information about IPv6 compatibility



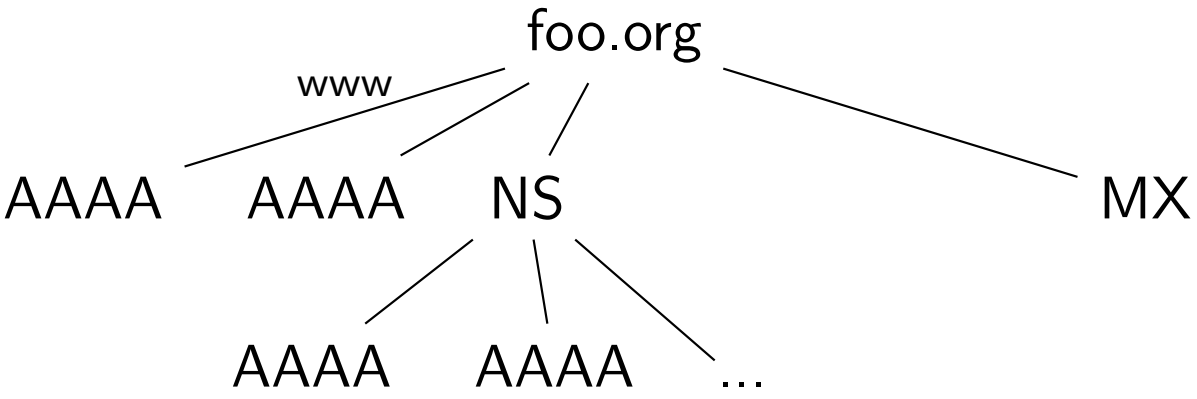
Analyzing one domain

E.g.: Get information about IPv6 compatibility



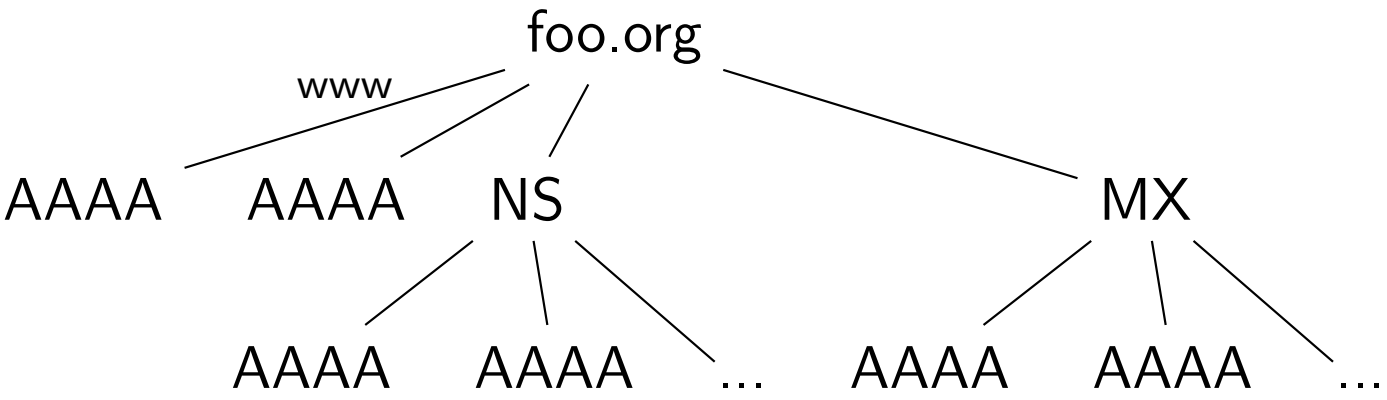
Analyzing one domain

E.g.: Get information about IPv6 compatibility



Analyzing one domain

E.g.: Get information about IPv6 compatibility



$\Rightarrow$ Multiple Queries per Domain

Analyzing the Internet

Internet is huge

- Millions of domains
- Different analyses (not only IPv6)

Analyzing the Internet

Internet is huge

- Millions of domains
- Different analyses (not only IPv6)

Different TLDs

- `.com`, `.org`, `.net`, ...
- Millions of domains per TLD
- Multiple queries per domain

Analyzing the Internet

Internet is huge

- Millions of domains
- Different analyses (not only IPv6)

Different TLDs

- `.com`, `.org`, `.net`, ...
- Millions of domains per TLD
- Multiple queries per domain

$$N_{\text{Queries}} \geq (\underbrace{130M}_{\text{.com}} + \underbrace{15M}_{\text{.net}} + \underbrace{10M}_{\text{.org}} + \underbrace{\ldots}_{\text{Other TLDs}}) * N_{\text{Queries per Domain}}$$

Analyzing the Internet

Internet is huge

- Millions of domains
- Different analyses (not only IPv6)

Different TLDs

- `.com`, `.org`, `.net`, ...
- Millions of domains per TLD
- Multiple queries per domain

$$N_{\text{Queries}} \geq (\underbrace{130M}_{\text{.com}} + \underbrace{15M}_{\text{.net}} + \underbrace{10M}_{\text{.org}} + \underbrace{\ldots}_{\text{Other TLDs}}) * N_{\text{Queries per Domain}}$$

$$\Rightarrow N_{\text{Queries}} \geq 1.55 \text{ Billion}$$

Scalability

`dig` does not scale

- 100.000 queries take about 4 hours
- $\Rightarrow$ Querying 130M domains: $\approx$ 216 days

Scalability

`dig` does not scale

- 100.000 queries take about 4 hours
- $\Rightarrow$ Querying 130M domains: $\approx$ 216 days

Scalability is a big challenge!

Scalability

`dig` does not scale

- 100.000 queries take about 4 hours
- $\Rightarrow$ Querying 130M domains: $\approx$ 216 days

Scalability is a big challenge!

Can we do better?

Goals

Enable performant large scale analyses on the Internet

Goals

Enable performant large scale analyses on the Internet

- Large scale:
  Hundreds of Millions of domains

Goals

Enable performant large scale analyses on the Internet

- Large scale:
  Hundreds of Millions of domains
- Multiple Analyses:
  IPv6 deployment, Port Scans with SNI, . . .

Goals

Enable performant large scale analyses on the Internet

- Large scale:
  Hundreds of Millions of domains
- Multiple Analyses:
  IPv6 deployment, Port Scans with SNI, ...
- Long term evolution:
  Repeat analyses periodically over years

Goals

Enable performant large scale analyses on the Internet

- Large scale:
  Hundreds of Millions of domains
- Multiple Analyses:
  IPv6 deployment, Port Scans with SNI, ...
- Long term evolution:
  Repeat analyses periodically over years
- Performance:
  Be much faster than dig

Roadmap

- Enable performant DNS resolution in large scale

Roadmap

- Enable performant DNS resolution in large scale
  - Investigate performance:
    Time usage

Roadmap

- Enable performant DNS resolution in large scale
  - Investigate performance:
    Time usage
  - Investigate suitability:
    Reliability / success rate

Roadmap

- Enable performant DNS resolution in large scale
  - Investigate performance:
    Time usage
  - Investigate suitability:
    Reliability / success rate
- Analysis Framework:
  - Enable long term and large scale analyses

Roadmap

- Enable performant DNS resolution in large scale
  - Investigate performance:
    Time usage
  - Investigate suitability:
    Reliability / success rate
- Analysis Framework:
  - Enable long term and large scale analyses
  - Analysis: Investigate specific aspects of the Internet

Roadmap

- Enable performant DNS resolution in large scale
  - Investigate performance:
    Time usage
  - Investigate suitability:
    Reliability / success rate
- Analysis Framework:
  - Enable long term and large scale analyses
  - Analysis: Investigate specific aspects of the Internet
- Analyze current IPv6 deployment

Roadmap

- Enable performant DNS resolution in large scale
  - Investigate performance:
    Time usage
  - Investigate suitability:
    Reliability / success rate
- Analysis Framework:
  - Enable long term and large scale analyses
  - Analysis: Investigate specific aspects of the Internet
- Analyze current IPv6 deployment
- (Optional) analyze further aspects (e.g. Junk Analysis)

First Results

Tested performance and suitability of ZDNS (released 5 months ago)

First Results

Tested performance and suitability of ZDNS (released 5 months ago)

- Queried 1 Million domains (Alexa)

First Results

Tested performance and suitability of ZDNS (released 5 months ago)

- Queried 1 Million domains (Alexa)
- Measured needed time (performance)

First Results

Tested performance and suitability of ZDNS (released 5 months ago)

- Queried 1 Million domains (Alexa)
- Measured needed time (performance)
- Investigated success rate

First Results

Tested performance and suitability of ZDNS (released 5 months ago)

- Queried 1 Million domains (Alexa)
- Measured needed time (performance)
- Investigated success rate
- Evaluated result quality:

First Results

Tested performance and suitability of ZDNS (released 5 months ago)

- Queried 1 Million domains (Alexa)
- Measured needed time (performance)
- Investigated success rate
- Evaluated result quality:
  - Completeness of results

First Results

Tested performance and suitability of ZDNS (released 5 months ago)

- Queried 1 Million domains (Alexa)
- Measured needed time (performance)
- Investigated success rate
- Evaluated result quality:
  - Completeness of results
  - Intermediate steps of resolution?

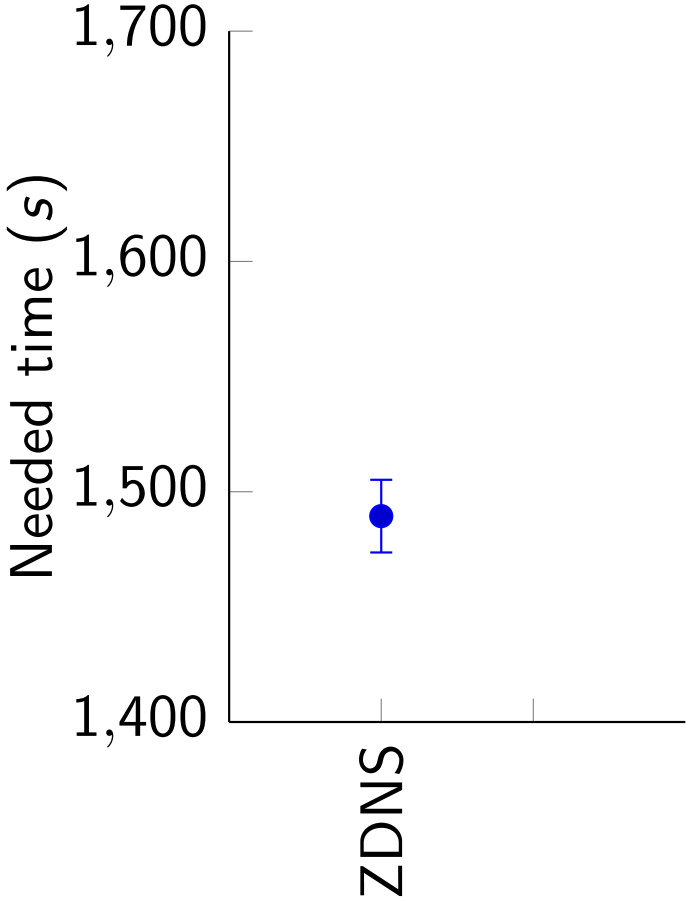
First Results

Tested performance and suitability of ZDNS (released 5 months ago)

- Queried 1 Million domains (Alexa)
- Measured needed time (performance)
- Investigated success rate
- Evaluated result quality:
  - Completeness of results
  - Intermediate steps of resolution?

Compared results with those obtained from own tool.

First Results



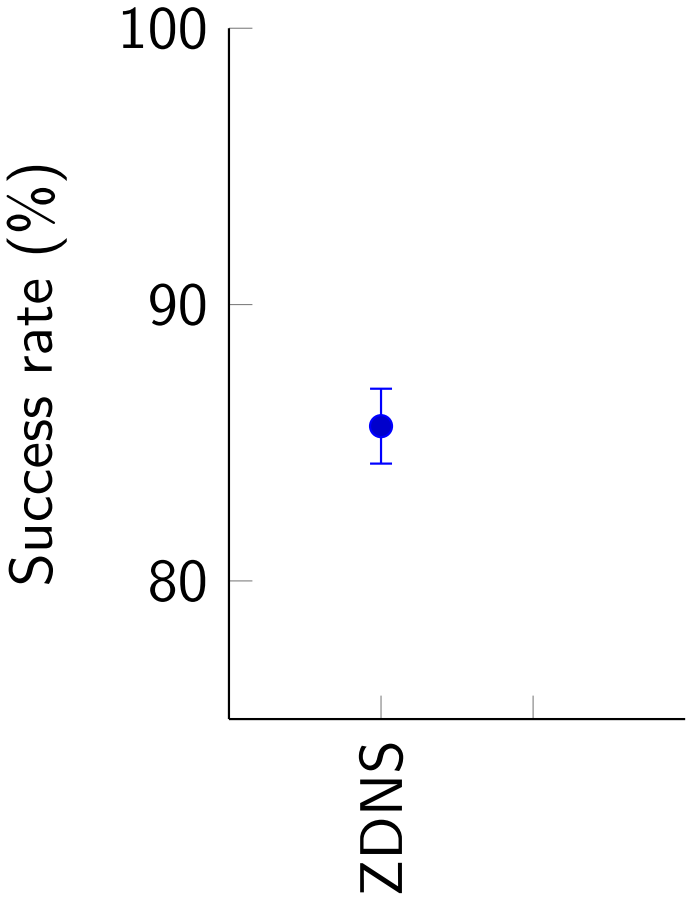


Figure: Test results for ZDNS (1.000.000 queries)

First Results

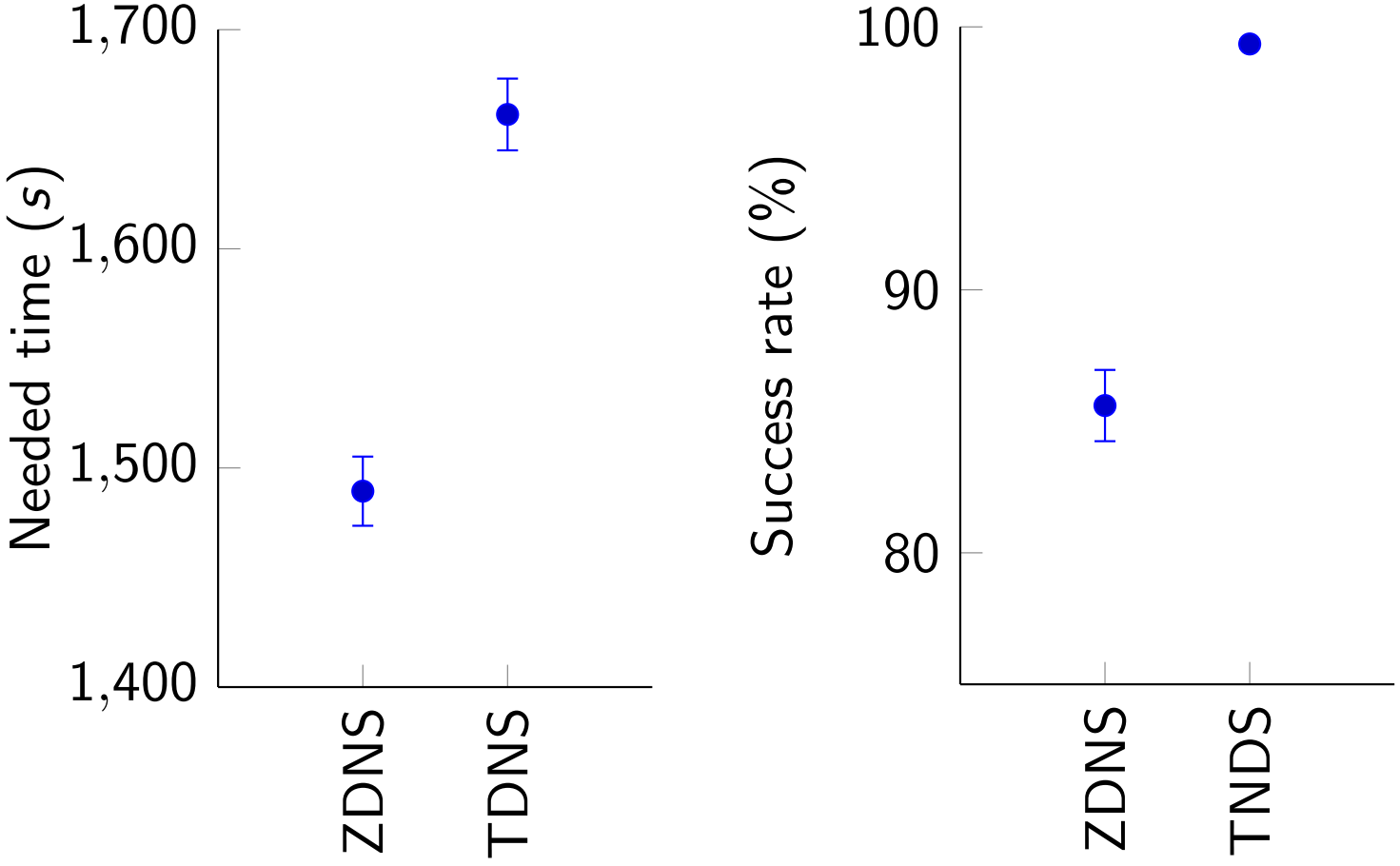


Figure: Test results for ZDNS and TDNS (1.000.000 queries)

Timeline

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Timeline

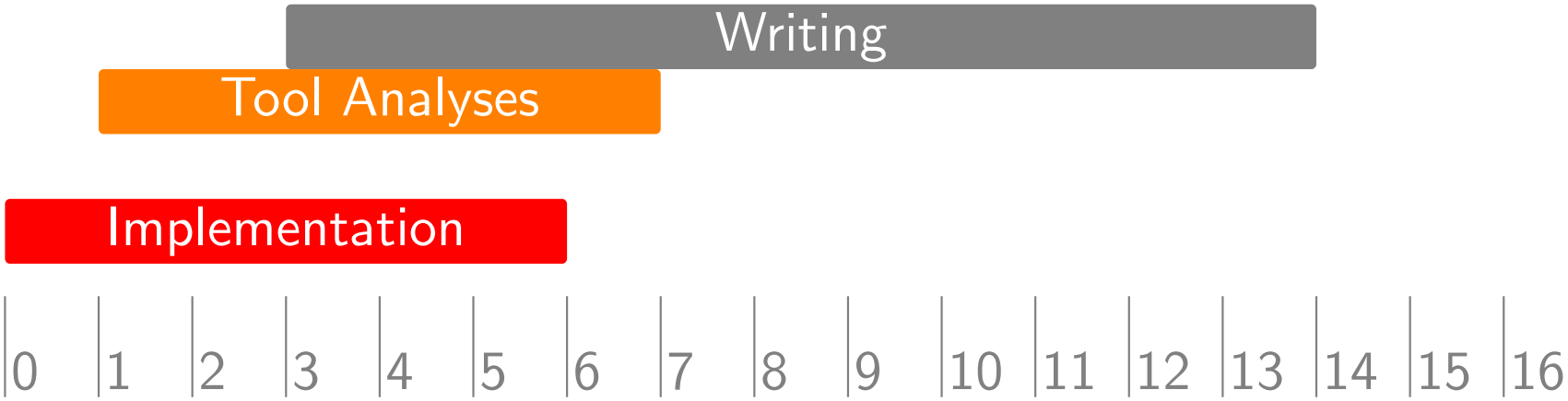
Implementation

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

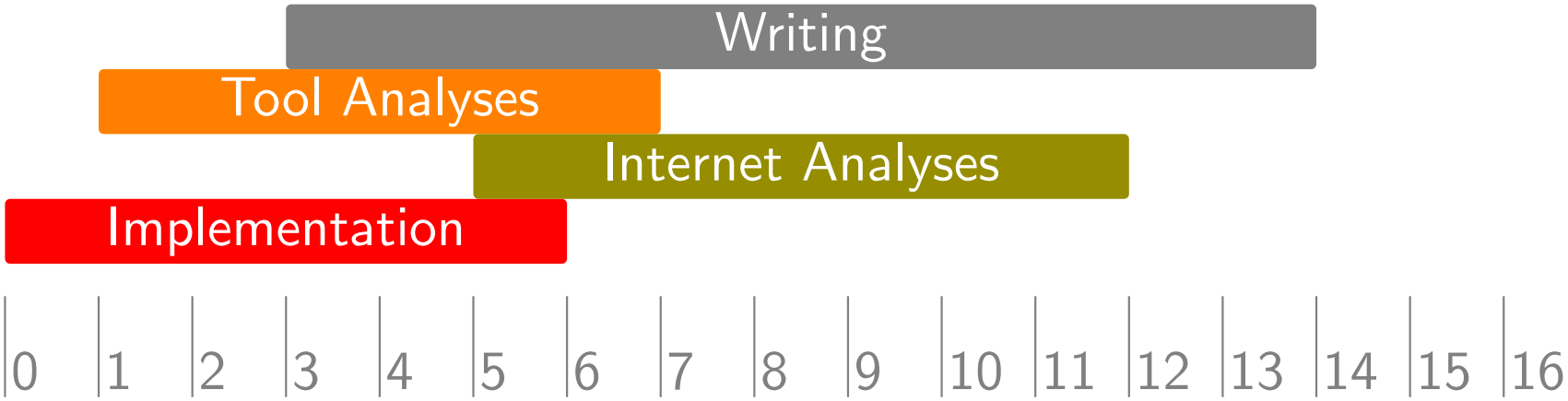
Timeline


Tool Analyses
Implementation
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

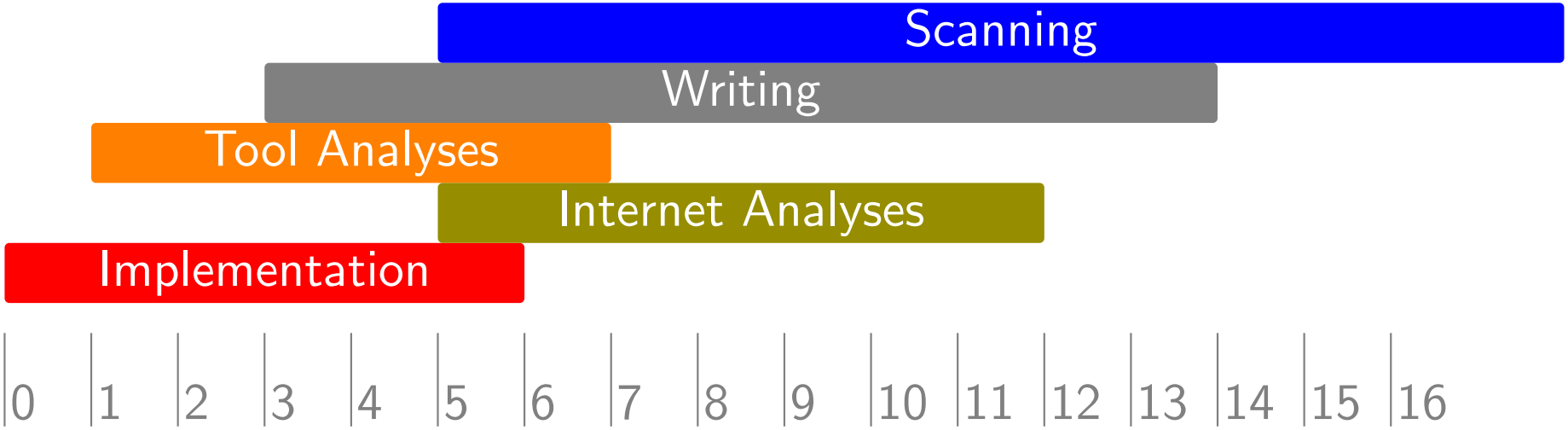
Timeline


Writing
Tool Analyses
Implementation
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

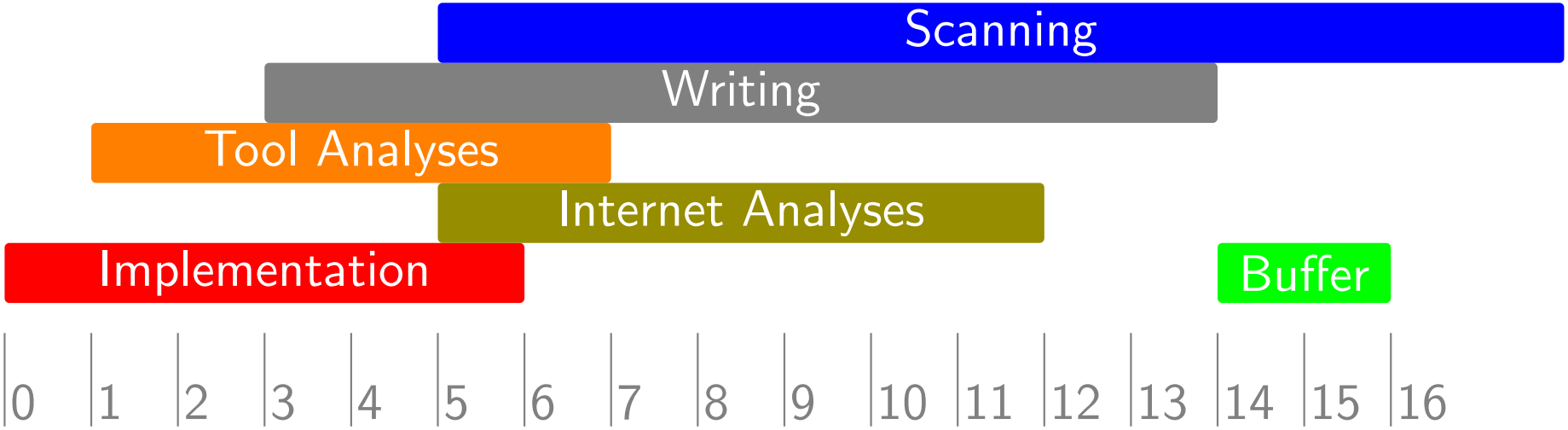
Timeline


Writing
Tool Analyses
Internet Analyses
Implementation
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Timeline


Scanning
Writing
Tool Analyses
Internet Analyses
Implementation
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Timeline


Scanning
Writing
Tool Analyses
Internet Analyses
Implementation
Buffer
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

References

- John S. Otto, Mario A. Sánchez, John P. Rula, and Fabián E. Bustamante.
  Content delivery and the natural evolution of dns.
- Tom van Goethem, Ping Chen, Nick Nikiforakis, Lieven Desmet, and Wouter Joosen.
  Large-scale security analysis of the web: Challenges and findings.
- Roland van Rijswijk-Deij, Mattijs Jonker, Anna Sperotto, and Aiko Pras.
  A high-performance, scalable infrastructure for large-scale active dns measurements.
- Roland van Rijswijk-Deij, Anna Sperotto, and Aiko Pras.
  Dnssec and its potential for ddos attacks.