OpenINTEL

an infrastructure for long-term, large-scale and high-performance active DNS measurements

An update and ongoing efforts

Mattijs Jonker[†], Roland van Rijswijk-Deij[†‡], Anna Sperotto[†]
[†]University of Twente, [‡]SURFnet bv

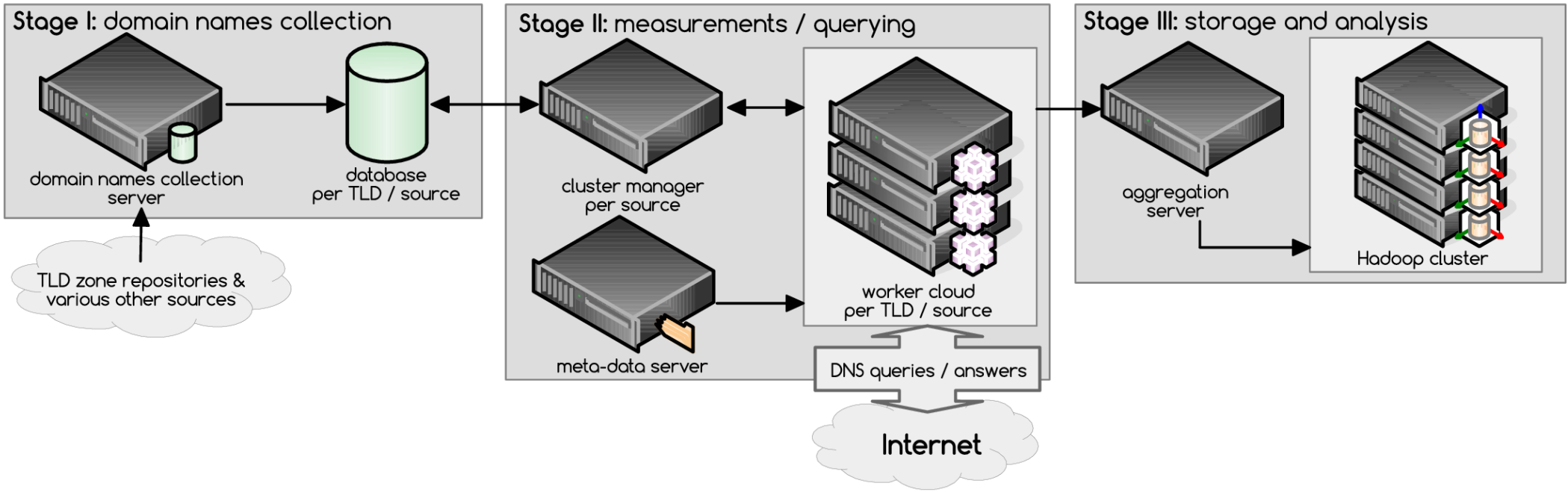
Why measure DNS?

- Measuring **what is in the DNS over time** provides information about the **evolution of the Internet**
- As we will see, longitudinal DNS data also has **security applications**

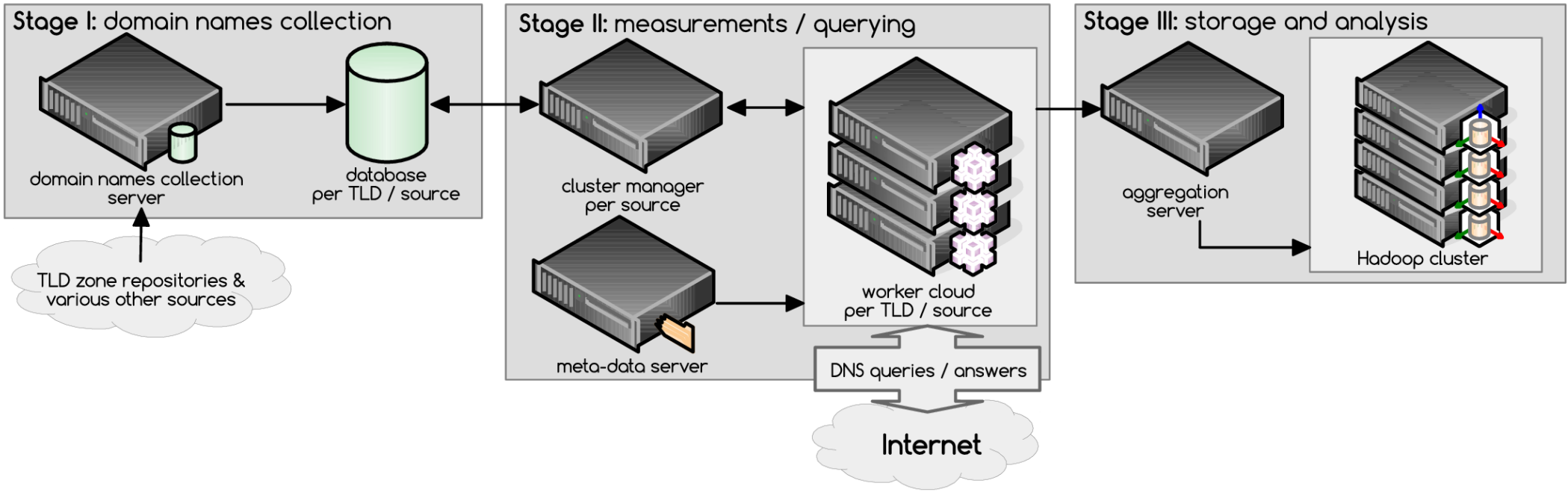
Goals and Challenges

- We send a c**omprehensive set of DNS queries for every name** in a TLD, **once per day**
- We do this at **scale**, our current measurement covers over **60% of the global DNS namespace:**
  - com, .net, .org, .info, .mobi
  - .nl, .se, .nu, .ca, .fi, .at, .dk (more being negotiated)
  - 1183 new gTLDs (e.g. .berlin, .xxx, .xyz, ...)
  - Alexa Top 1 million
  - in total around **200M domain names**

High-level architecture

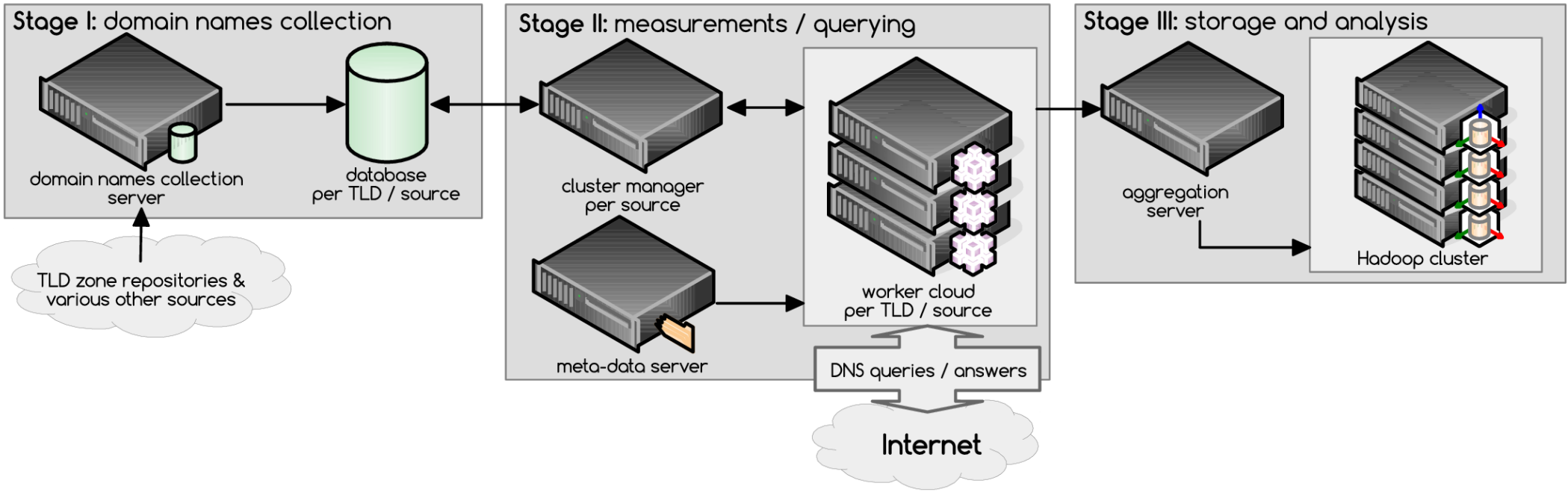


High-level architecture



- **We measure all names from the previously outlined sources**
- **We acquire the full zones through contracts with various registry operators**

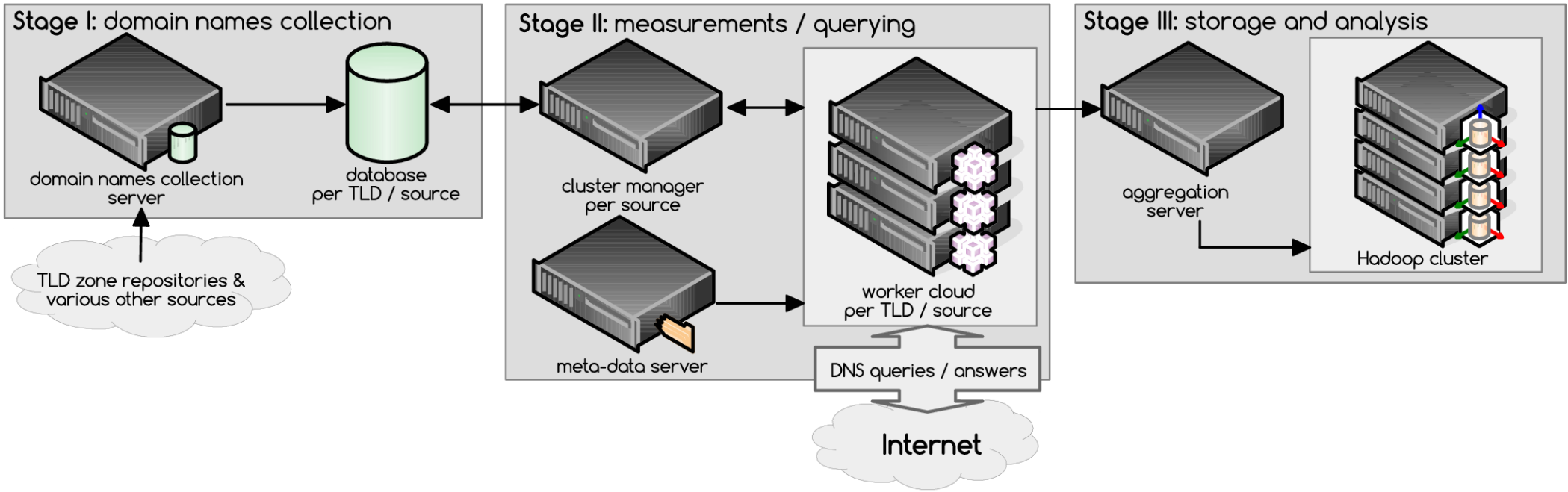
High-level architecture



- We measure all names from the previously outlined sources
- We acquire the full zones through contracts with various registry operators

- We send ~14 queries for various DNS resource record types for each name, every day
- RRs: SOA, A, AAAA, NS, MX, TXT, SPF, DS, DNSKEY, NSEC(3)
- We store valid answers, including full CNAME expansions, RRSIGs, ...
- A / AAAA answers are supplemented with AS numbers (from pfx2as data)

High-level architecture



- We measure all names from the previously outlined sources
- We acquire the full zones through contracts with various registry operators

- We send ~14 queries for various DNS resource record types for each name, every day
- RRs: SOA, A, AAAA, NS, MX, TXT, SPF, DS, DNSKEY, NSEC(3)
- We store valid answers, including full CNAME expansions, RRSIGs, ...
- A / AAAA answers are supplemented with AS numbers (from pfx2as data)

- The data set covers a two-year period for com, net & org ([2015-03-01, 2017-02-28])
- These 3 gTLDs represent ~50% of the global domain name space
- On average, 2.3B data points are stored daily

So what's new?

- Extended coverage of zones
- Added "infrastructure" elements (IPs for NS & MX)
- Released substantial open access data sets
  - Alexa Top 1M (>1 year data)
  - ccTLDs .se and .nu
  - Allows people to do their own analysis
  - We can run "mature" queries "on behalf"

Case study: CEO fraud

- CEO fraud is a highly targeted form of phishing, and a form of Business E-mail Compromise (BEC) scam, involving look-a-like domain names
- n.b.: this content was presented in person at AIMS but is not ready for publication yet. The following 7 sheets have been redacted.

Case study: CEO fraud

Case study: CEO fraud

Case study: CEO fraud

Case study: CEO fraud

Case study: CEO fraud

Case study: CEO fraud

Case study: CEO fraud

Case study illustrates that the measurement data can be used for operational security purposes and forensics

Purpose of talk/what do I want?

- First and foremost: feel free to approach me with suggestions or collaboration ideas
- Secondly, AXFR, seize or steer clear?
  - E.g.: .sv (ccTLD of El Salvador)

Mattijs Jonker
m.jonker@utwente.nl