Unifying Diverse DNS Data Sources

Alfred Arouna¹,²,³, Mattijs Jonker², and Ioana Livadariu¹

¹ SimulaMet - Simula Metropolitan, Oslo, Norway – (alfred,ioana)@simula.no
² University of Twente, Enschede, the Netherlands – m.jonker@utwente.nl
³ OsloMet – Oslo Metropolitan University, Oslo, Norway

1 DNS Data Sources Diversity

The Domain Name System (DNS) maps names to a value and relies on a reversed tree architecture.

Figure 1: DNS ecosystem overview. DNS data can be collected at-rest, on-the-fly or send onwards from the server (either on the wire or shared with a third party).

Figure 2: DNS datasets usage over the last 10 years. The majority of articles relied on one-time snapshot of the state of (parts of) the DNS. Thus, the combination of DNS datasets provides an overview of real-world DNS behaviour.

2 Limited Perspective of Global DNS Behaviour

Researchers can either create 1) new dataset or 2) collect partially or totally existing datasets or 3) combine datasets.

Research Questions

RQ1: How can we characterize diverse DNS data sources?
RQ2: To what extent can we unify public and/or private DNS data sources?
RQ3: How can we best design a software architecture to unify DNS data and facilitate easy access?

3 Work In Progress

Open Challenges

• Distributed architecture.
• High diversity of actors and datasets.
• Measurement approach artefacts: privacy, confidentiality, coverage, frequency, complexity and availability.
• Support of non publicly available datasets, for instance self-instrumented DNS data.

Takeaway

Unifying diverse DNS data sources will evolve real-world and long-term DNS behaviour characterization.