ALTO for Querying LMAP Results
draft-seedorf-lmap-alto-02

Jan Seedorf
David Goergen
Radu State
Vijay Gurbani
Enrico Marocco

IETF 88, Vancouver
ALTO WG
November 2013

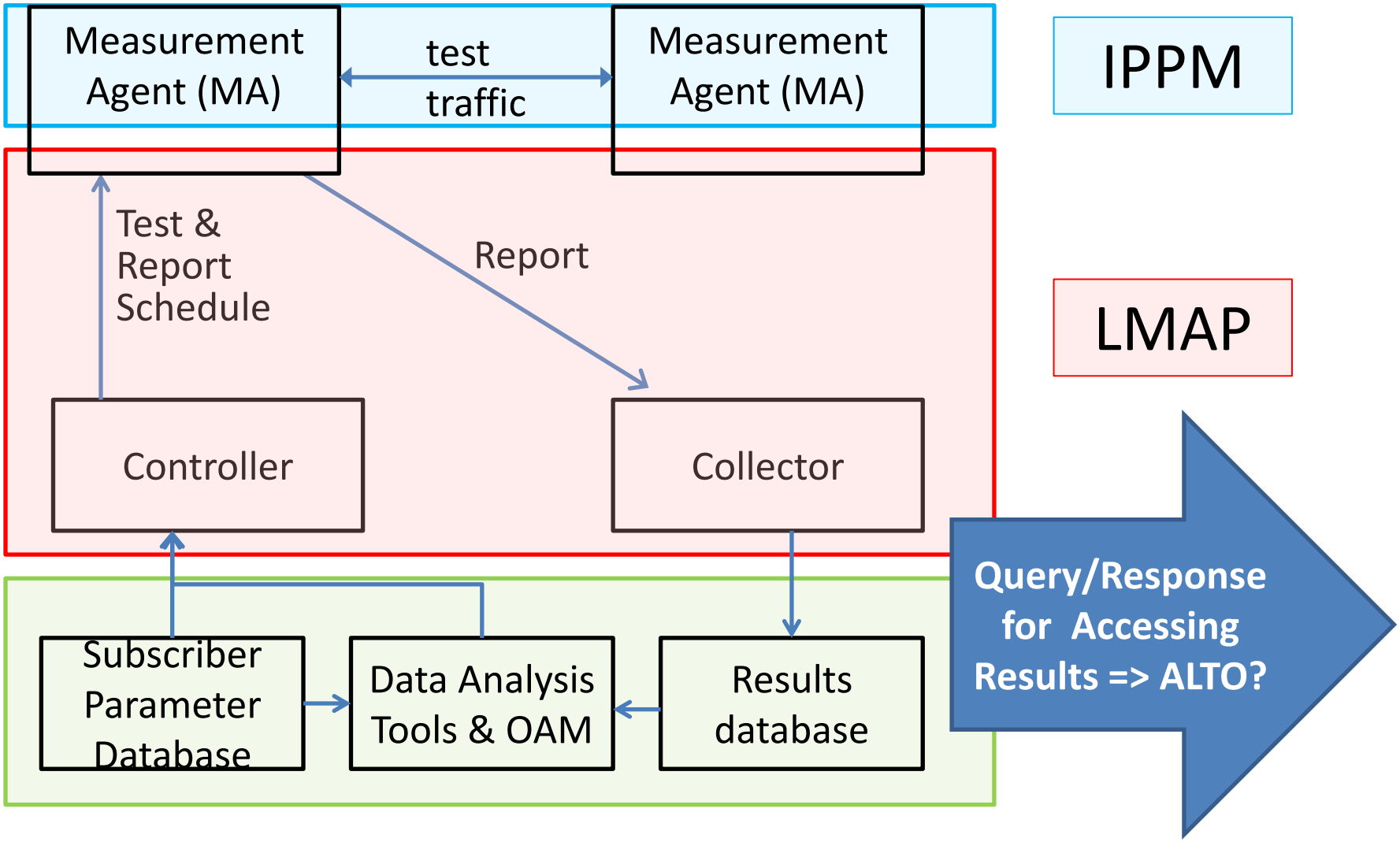
Overview & Motivation

How to make "Large-Scale Measurement of Broadband Performance" measurement results available?

- Status quo (of measurement result publication):
  - very high level human-readable format (pdf, jpeg, ...)
  - finest granularity level (csv)
- Question: Are there use cases that benefit from an intermediate way to provide access to large-scale network measurement results?
  - flexible enough to allow for querying of specific and possibly aggregated data

→ If answer to question is yes, ALTO may be a good candidate

Overall LMAP Measurement Framework



draft-seedorf-lmap-alto-02

- Merger of drafts
  - draft-seedorf-lmap-alto-01 (Sections 1-4 of draft-seedorf-lmap-alto-02)
  - draft-goergen-lmap-fcc-00 (Section 6 of draft-seedorf-lmap-alto-02)
  - + new Section 5 on useful ALTO extensions for LMAP use case

- Both drafts presented at IETF-87 in LMAP WG
  - Well received with interest
  - Feedback however that accessing lmap results is currently not on lmap charter

Example Use Cases

- **Video Streaming Service Provider**
  - LMAP results regarding a particular end user's access network provider could help the service provider to optimize/parametrize its HTTP adaptive streaming service
- **Website Front End Optimization**
  - Statistics about e.g. average download speeds for a given end user request can be useful for dynamically adapting HTML/CSS/JavaScript content
- **Display estimation of service quality / total download time to users**
  - Using LMAP results e.g. to indicate to the user what Quality-of-Experience to expect when clicking on a given link, or the estimated total download time for given content
- **Troubleshooting**
  - In general, any service on the Internet may be interested in LMAP data for troubleshooting

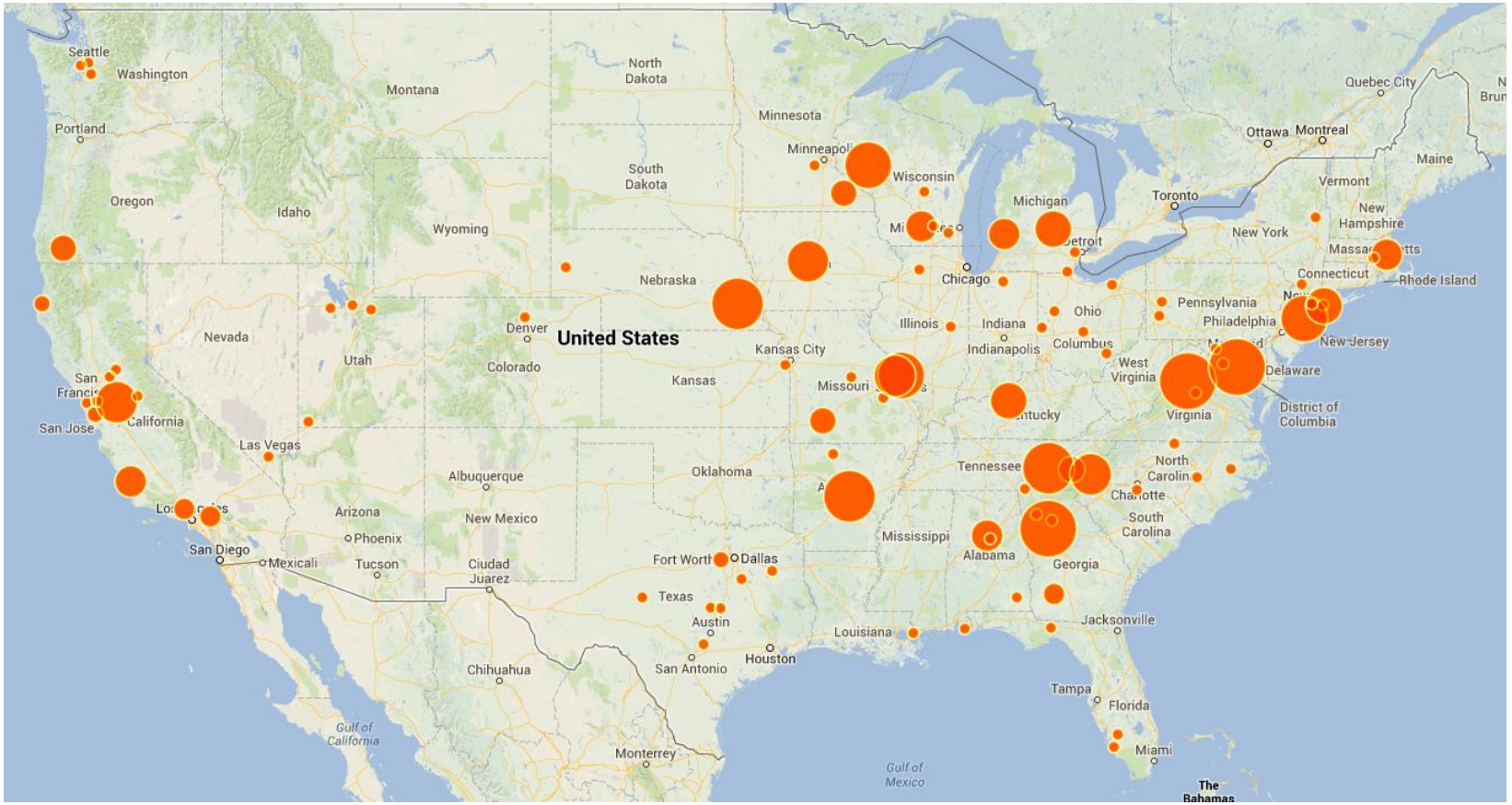
Useful ALTO Extensions for LMAP Use Case

- Server-initiated Notifications
  - This extension would allow applications to be notified when certain new LMAP measurements are available, such as new measurement results on average download speeds
  - Such new results could then be downloaded and used immediately by applications
- Incremental Updates
  - When ALTO is used for querying LMAP results, the corresponding ALTO maps may potentially be quite large (e.g. when a webservice queries for particular, detailed results regarding a whole ISP)
  - Incremental ALTO updates would be a very useful mechanism to reduce the amount of data that would be needed for transmitting these maps.

Case study: Analyzing a large-scale dataset (1)

- FCC Dataset specification
  - FCC has embarked on a nationwide performance study of residential wireline broadband service
  - Aim is to use the raw datasets from this study for analysis and to create ALTO topology map and a cost map from this dataset
- Processing of FCC datasets
  - Using a canonical Map-Reduce computational paradigm on a Hadoop cluster

Case study: Analyzing a large-scale dataset (2)



Stable unit_id

Case study: Analyzing a large-scale dataset (3)

- Observations
  - Some unit_id are located outside US
    - Assume user has manually configured DNS resolver
  - OpenDNS and Google DNS resolvers were ignored
  - Large convergence to single point (Potwin,KS)
    - Potwin is the geographical center of the US
    - ISPs generally locate their primary or secondary DNS name servers
    - continue to further investigate on minimizing the impact
  - Some unit_id change ISP and/or location

- Next steps
  - Attempt to create network map
    - Rough PID groupings accomplished by unit IDs belonging to same ISP
    - More formal PID groupings for further study (e.g., group by bandwidth speed irrespective of ISP, lowest jitter, …)
  - Attempt to create a cost map
    - Different cost maps for different applications (e.g., use udp latency or jitter as a cost metric for VoIP applications)
  - Cross-reference with other dataset
  - Using stable unit IDs as landmarks in a virtual coordinate system.

Feedback / Discussion

Acknowledgements

- Jan Seedorf is partially supported by the mPlane project (mPlane: an Intelligent Measurement Plane for Future Network and Application Management), a research project supported by the European Commission under its 7th Framework Program (contract no. 318627). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the mPlane project or the European Commission.

Advantages of using ALTO

- A very lightweight JSON-based encoding for network information (set of REST APIs on top of HTTP)
- ALTO “network map“: abstracts physical network topology into an aggregated but logical topology
  - individual hosts are aggregated into a well defined network location identifier called a PID
- ALTO “cost map“: associates “costs” between two network locations (PIDs)
  - Costs can indicate e.g. routing hops, the financial cost of sending data over the link, available bandwidth on the link, or a user-defined cost attribute that allows arbitrary reasoning