Delta: Scalable Data Dissemination under Capacity Constraints

Konstantinos Karanasos Microsoft* Asterios Katsifodimos TU Berlin*

Ioana Manolescu INRIA

Publish/subscribe paradigm

- Subscribers: subscribe to data of interest
- Publishers: publish data, push them to interested subscribers

Existing approaches for scaling up

- Centralized: focus on filtering
- Distributed: with brokers/servers

What if subscribers help each other?

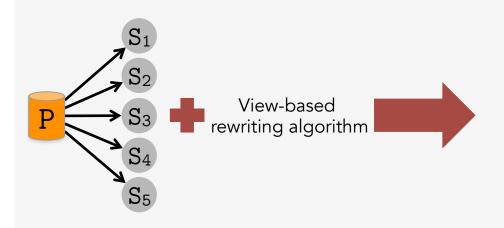
Goal: Organize subscriptions into a network

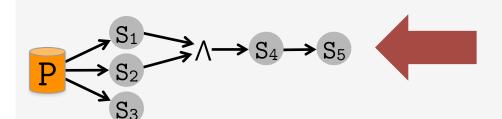
- Minimizing resource utilization and update latency
- Respecting resource capacity constraints

Basic idea:

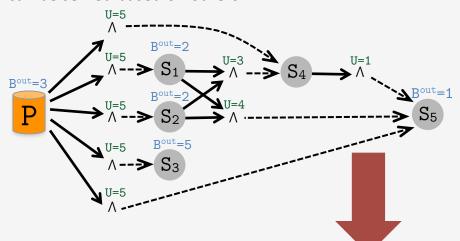
- Subscriptions = views (to be kept up-to-date)
- View-based rewriting algorithm determines how subscriptions can be fed based on others

Overview of Approach





Rewritability Graph: encodes how each subscription can be served based on others



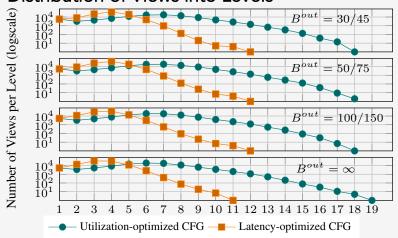
Find Best Network Configuration

- Configuration with minimum resource utilization under capacity constraints (Integer Linear Programming)
- Reduce latency of configuration under capacity constraints (greedy algorithm)

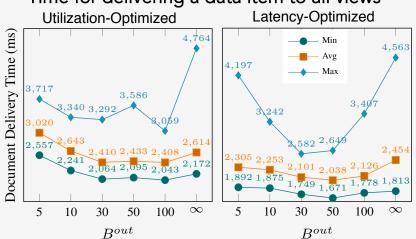
Application & Experiments

- XML subscriptions and documents
- Network of 300 machines across 9 French cities
- Gurobi optimizer for ILP
- Generated 10K subscriptions (using YFilter)

Distribution of Views into Levels



Time for delivering a data item to all views



*Work partly done while A. Katsifodimos was at INRIA & Univ. Pari-Sud and while K. Karanasos was at INRIA and later IBM Almaden.