**StreamGlobe:**
Adaptive Anfragebearbeitung und Optimierung auf Datenströmen

B. Stegmaier und **R. Kuntschke**

TU München – Fakultät für Informatik
Lehrstuhl III: Datenbanksysteme

http://www-db.in.tum.de/research/projects/StreamGlobe
Outline

- Motivation

- StreamGlobe
  - The StreamGlobe Approach
  - Architecture Overview

- Current and Future Research

- Conclusion
Exemplary Initial Situation

- **Network**
  - Consists of peers
  - Given or grown topology

- **Data Sources**
  - Provide XML data stream
  - Possibly infinite streams (e.g., sensor measurements)

- **User requests**
  - Continuous queries
  - Query language XQuery
  - Registered at a peer
General Traditional Approach

1. Register requests
2. Lookup data sources
3. Establish data transfer
   → Peers may connect arbitrarily
4. Process / Execute requests
5. Routing of streams
General Traditional Approach (ctd.)

- **Drawbacks**
  1. Transmission of useless data
  2. Redundant transmissions
  3. Multiple request evaluation

⇒ Network congestion and processing overhead
Why StreamGlobe?

- Other Systems / previous work
  E.g. Cougar, TelegraphCQ, Multicast techniques:
  - Focus on specific aspects (e.g., query optimization)
  - Tailored to specific domains

- StreamGlobe
  - Contribution is combination of techniques:
    In-network query processing combined with routing
  - Constitutes a generic infrastructure
    - Independent of domain
    - Efficient data stream transformation and distribution
Outline

- Motivation

- StreamGlobe
  - The StreamGlobe Approach
  - Architecture Overview

- Current and Future Research

- Conclusion
The StreamGlobe Approach

Intelligent Routing
- Push query execution into network
- Multicast routing techniques
  - Data Stream Clustering
- Multi-query optimization

⇒ Reduce network traffic
⇒ Reduce processing cost
⇒ Load balancing
Basic Concepts

- P2P Network Topology
  - No arbitrary communication
    → Communication via *transfer paths*
  - No fixed P2P topology

- Classification of peers
  - Thin-Peers
  - Super-Peers
    ⇒ Constitution of a super-peer backbone

- Hierarchical organization
  → *Speaker-peer* responsible for certain subnet
StreamGlobe Peer Architecture

- Based upon *Open Grid Services Architecture (OGSA)*
- Integration similar to OGSA-DAI or OGSA-DQP
- Layers as grid-services
- Availability according to peer capabilities
- Message exchange via RPC and notifications
- Data stream transfer via direct TCP connections

Diagram:
- StreamGlobe Interface
- Optimization
- Query Engine
- Metadata Management
- Globus Toolkit
- XQuery
- Subscriptions
- XML
- Data Streams
StreamGlobe Interface

- Registration of XML data streams
  - Possibilities
    - As individual data stream
    - As part of a virtual data stream
  - Specification of schema with XML Schema

- Registration of subscription rules
  - At a special peer (user device)
  - Specified using XQuery
Metadata Management

- Managed information
  - Network topology
  - Registered subscriptions and data streams
  - Statistics of data streams

- Based upon service-data elements of Globus Toolkit

- Maintenance
  - E.g. peers joining/leaving, change of subscriptions, etc.
  - Using notification mechanisms of Globus Toolkit
Optimization

**Goals**
1. Register arbitrary subscriptions at any peer
2. Reduce network traffic
3. Optimize evaluation of many subscriptions

**Achievement**
- Pushing query execution into the network
  → (1), (2) and (3)
- Multi-query optimization
  → (3)
- Early filtering of data streams resp. evaluation of subscriptions
  → (2)
- Data stream clustering
  → (2)
Multi-Query Optimization

- Performed by speaker-peer
- Analyze subscriptions and streams
  - Common subqueries
  - Re-usability of streams
  - Based on properties of subscriptions / streams
- Computes
  - Filters and queries
  - Data stream clustering
  - Execution locations
Query Execution

- Basic concepts
  - Streaming evaluation and push-based techniques
  - Preclude unbounded buffering by requiring window constraints
  - Extensibility by means of mobile code

- Evaluation of subscriptions with FluX
  - Designed for streaming processing of XQuery
  - Event-based extension to XQuery
  - Usage of schema information for buffer minimization
Outline

- Motivation

- StreamGlobe
  - The StreamGlobe Approach
  - Architecture Overview

- Current and Future Research

- Conclusion
Current and Future Research

■ Current Research
  - Optimization techniques
  - Extension of FluX

■ Future Research
  - Quality-of-Service management
  - Explicit load balancing
  - Load shedding techniques
  - Construction of overlay network
  ...

9. Mai 2006
Conclusion

StreamGlobe

- Exploiting in-network query processing capabilities
- In combination with data stream clustering

→ Minimization of network traffic

- Query execution with FluX
  → Efficient and scalable execution of subscriptions
- Multi-query optimization
  → Parallelization and load balancing in the network
Related Work

- Braumandl, Kemper, Kossmann. “Quality of Service in an Information Economy”. TOIT 2003
- Chandrasekaran, Cooper, Deshpande, Franklin, Hellerstein, Hong, Krishnamurthy, Madden, Raman, Reiss, Shah. “TelegraphCQ: Continuous Dataflow Processing for an Uncertain World”. CIDR 2003
- Krämer, Seeger. “Pipes – A Public Infrastructure for Processing and Exploring Streams”. SIGMOD 2004