Data Aggregation and Management in Sensor Networks

Alejandro Buchmann
Databases and Distributed Systems
Dept. of Computer Science
Darmstadt University of Technology
buchmann@informatik.tu-darmstadt.de
Data/Event Aggregation

• Diffusion/Percolation (DIMENSIONS et al.)
  – Multi-resolution data storage
  – Spatial & temporal correlations to reduce dimensionality
  – Mostly simple aggregation function w. homogeneous data

• Streaming Queries (Cougar, Fjords, etc.)
  – SQL extensions for continuous querying
  – Tuple aggregation w. windowing relational operators

• Event Graphs (reactive middleware, Hermes, Dream, etc.)
  – Event algebra: \( \text{ANY} 2(e_1, e_2, e_3, e_4) \land (e_5 \lor e_6) \)
  – Event consumption modes (chronological, recent, window)
  – Garbage collection of events
Issues in Data/Event Aggregation

• Where is aggregation performed
• How to deal with time in a distributed and unreliable environment
• How to deal with state/persistence
• How to deal with asynchrony and intermittent communication
• How to deal with (uncontrolled) redundancy and failed sensors
Where to aggregate

- Wherever aggregation takes place, state must be maintained
  - Size
  - Reliability (replication, persistent storage,...)
  - Recoverability
- Different aggregation mechanisms require different resources
  - Heavy duty query processing on simple sensor node?
  - Filters at peripheral nodes, QP on central node?
- Adapting the routing to the aggregation vs. adapting the aggregation to the routing
Dealing with Time in Aggregation

• Absolute temporal consistency: between state of environment and its representation
  – Data is described by a triplet d: (value, avi, ts)
  – Absolute consistency (\(t\text{.current} - d.ts\) \(\leq d.avi\))

• Relative temporal consistency
  – Data used to derive new data form a relative consistency set \(R\)
  – \(R\) is associated with a relative validity interval \(R_{rvi}\)
  – For \(d_k \in R\) \(\forall d_i \in R\) \(|d_k.ts - d_i.ts| \leq R_{rvi}\)
Event detection and composition in distributed systems

- 2g-precedence for closed networks and bounded imprecision

- Accuracy Interval Approach for timestamping events in large-scale, loosely coupled DS

\[
\text{local detection delay - ldd} \quad \text{error bound} \quad \text{NTP (global time service)}
\]
Variable Timestamp Representation

- Must provide interval-based time model with explicit (in)accuracy intervals
- Allen’s interval semantics ==> indeterminacy
- Stable past vs. unstable past and present
- Middleware must expose indeterminacy, resolve through application semantics
Dealing with Stateful Aggregation

- Need persistence for filters and incomplete event/data compositions
- Achieve reliability through replication
- (Selectively) write to persistent medium
- Write asynchronously from secondary copy
Redundancy and Incompleteness

- Must eliminate redundant readings
- May have to avoid overload by dropping events ➔ must be able to give confidence bounds for the result
- Must annotate result with (sub)set of sensors that contributed to answer
There is no silver bullet!
Support for the Elderly
event notification

(stateful) composition - MMDBMS

mass storage

mass storage

client-controlled transaction semantics and QoS

transactional notifications

content/concept-based filtering and routing

self-stabilizing pub/sub

scopes of events

reactive functionality

transactional notifications

event composition

(heterogeneous) event detection

client-controlled transaction semantics and QoS

© A. Buchmann