Automatic Consistency Assessment for Query Results in Dynamic Environments

Jamie Payton  
University of North Carolina at Charlotte  
payton@uncc.edu

Christine Julien  
The University of Texas at Austin  
c.julien@mail.utexas.edu

Gruia-Catalin Roman  
Washington University in St. Louis  
roman@wustl.edu

6 September 2007  
ESEC/FSE
Queries in Dynamic Pervasive Computing Environments

- Strong interest remains in applications for dynamic pervasive computing environments
  - Mobile ad hoc networks
  - Sensor networks
- Applications rely on information and services in the network
- Queries offer convenient abstractions for developers
  - Requires no knowledge of location, nature of data
Need for a Range of Query Semantics

- Queries in traditional database systems have well-defined meaning
  - Strict, transactional semantics
  - Implemented by locking protocols

- Needs of dynamic pervasive computing applications are not met
  - Mobility often causes failure
  - Streaming data does not translate to locking

- A new perspective is needed
  - Discover, define, reason about a range of query semantics
A Formal Model of Query Execution

Goal: Define execution environment

Provide formal foundation for describing how queries relate to the changing environment
Modeling Query Processing: Effective Active Configurations
Modeling Queries and Results

- A query is defined as a function over a sequence of effective active configurations to a set of host tuples
  - A configuration!
    - Simplifies expression of consistency

- Result configuration $\rho$ subject to constraints
  - Each result element $r$ must be reachable
  - Each host contributes only one host tuple $h$ to result

\[
 h \in \rho \Rightarrow \exists i : 0 \leq i \leq m :: h \in E_i \land \\
 \langle \forall r : r \in \rho - \{h\} :: h \uparrow 1 \neq r \uparrow 1 \rangle
\]
Reasoning About Query Results

Goal: Define degree of consistency
Examine relationship between result configuration and effective active configurations
Consistency Semantics: Immediate

- All results existed at the same time, remain available upon return
- No configuration changes during query execution

\[ \text{IMMEDIATE} = \rho = E_0 \land m = 0 \]
Consistency Semantics: Atomic

- All results existed at the same time
- A snapshot of some configuration

\[ \text{ATOMIC} \equiv \exists i : 0 \leq i \leq m \land \rho = E_i \]
Consistency Semantics: Atomic Subset

- All returned results existed at same time
- Possible that not all available results are returned
  - A partial snapshot

\[ \text{ATOMIC SUBSET} \equiv \exists i : 0 \leq i \leq m \land \rho \subset E_i \]
Consistency Semantics: Qualified Subset

- Slightly stronger than Atomic Subset
  - Provides some information about missing results

\[
\text{QUALIFIED SUBSET} \equiv \exists i : 0 \leq i \leq m \land \rho \subseteq E_i \land |\rho| \geq \alpha |E_i|
\]
Consistency Semantics: Weak

- All results existed during query execution
- Weakest meaningful guarantee

\[
\text{WEAK} \equiv \rho \subseteq \bigcup_{i=0}^{m} E_i
\]
Consistency Semantics: Weak Qualified

- Slightly stronger than Weak
- Provides some information about missing results

\[
\text{WEAK QUALIFIED} \equiv \rho \subseteq \bigcup_{i=0}^{m} E_i \land \left| \rho \right| > \alpha \left| \bigcup_{i=0}^{m} E_i \right|
\]
Practical Support

Goal: Provide self-assessing protocol to implement range of semantics
Examine environmental conditions during execution and report outcome
Protocol Overview

- Controlled flooding to propagate query and replies
  - Controlled by logical connectivity function
  - Practical support for scoping in [KAB07, ROM02]

- Monitoring of changes in values that impact consistency

- Query reply augmented with assessment of achieved semantics
A Controlled Flooding Approach

Phase 1: Establish

Wave 1

Wave 2

Phase 2: Collect

Wave 1

Wave 2
Responding to Environmental Change

- Changes after participants established, before results are returned can impact achieved semantics
  - No changes
    - ATOMIC
  - Departures only
    - ATOMIC SUBSET
    - QUALIFIED SUBSET
  - Departures and additions
    - WEAK
    - WEAK QUALIFIED
  - Value changes ↔ departures
Protocol Description: Membership

Phase 1
Wave 1

State at A
Membership = true
Monitoring = false
Waiting = {B, C}

State at C
Membership = true
Monitoring = false
Waiting = {E}

Diagram showing the network with nodes A, B, C, D, E, and F.
Protocol Description: Collection

Phase 2
Wave 1

State at A
Membership = true
Monitoring = true
Waiting = {B, C}

State at C
Membership = true
Monitoring = false
Waiting = {}
Protocol Description: Environmental Changes

Phase 2
Wave 2

Not atomic!

Achieved:
atomic subset
Protocol Evaluation

- Simulation using Omnet++
  - Mobility framework extension
- Issue a query from some host in the network
  - Utility of semantics
  - Protocol performance

Experiment Settings
- 1000 X 900 m² area
- Random waypoint mobility model
  - Pause time: 0 sec
- 802.11 MAC protocol
Achieved Semantics in Mobile Networks

![Query Semantic vs. Speed (ttl3, 30nodes)](chart)

- **Weak**
- **Atomic Subset**
- **Atomic**
Conclusions

- Present a new perspective on dynamic pervasive computing environments
  - Model allows for precise definition of query semantics
  - New query semantics provides opportunity for application development
  - Self-assessing protocol provides ability to reason about adaptive query execution

- Long-term goal: theory of query processing
  - Develop model, define semantics of persistent queries
  - Refine model to address peculiarities of sensor networks
Questions?

Jamie Payton  
University of North Carolina at Charlotte  
payton@uncc.edu

Christine Julien  
The University of Texas at Austin  
c.julien@mail.utexas.edu

Gruia-Catalin Roman  
Washington University in St. Louis  
roman@wustl.edu
References

Related Work

- [Pitoura95] focuses on providing consistency among replicas
  - Here: no replicas
  - Here: consistency between returned result and ground truth

- Weakened semantics
  - Split (kangaroo) transactions
  - Isolation-only
  - Toggle transactions
  - Pre-write transaction model
    - Most applicable in nomadic networks
      - Need a powerful fixed node to act as manager
Related Work

- Garcia-Molina et al. define a particular class of consistency semantics
  - Use the notion of single-site validity to achieve atomic semantics from query issuer point of view
    - Our approach applies to the development of several classes of query semantics
    - We provide an adaptive self-assessing protocol
Protocol Description: Membership

Phase 1
Wave 2

State at A
Membership = true
Monitoring = false
Waiting = {B, C}

State at B
Membership = true
Monitoring = false
Waiting = {D} – {D}

State at B
Membership = true
Monitoring = false
Waiting = {}

State at D
Membership = true
Monitoring = false
Waiting = {}

State at E
Membership = true
Monitoring = false
Waiting = {E}
Protocol Description: Non-impact Environmental Changes

Phase 1
Wave 1

State at B
Membership = false
Monitoring = false
Waiting = \{D\} \{F\}
Added = 0
Deleted = 0

State at C
Membership = false
Monitoring = false
Waiting = \{\}\{\}
Added = 0
Deleted = 0

State at A
Membership = true
Monitoring = false
Waiting = \{B, C\}
Added = 0
Deleted = 0

State at C
Membership = false
Monitoring = false
Waiting = \{F\}
Added = 0
Deleted = 0

No impact!
Achieved Semantics in Static Networks

Query Semantic vs. Number of Nodes (ttl3, 0m/s)
Protocol Performance: Increasing Network Density

![Graph showing Overhead and Latency vs. Number of Nodes (ttl3, 20m/s)]
Protocol Performance: Increasing Mobility

![Overhead and Latency vs. Speed Graph]

- **Overhead (bytes)** vs. **Speed (m/s)**
- **Latency (seconds)** vs. **Speed (m/s)**

- Overhead and Latency increase with increasing speed.
Impact of Mobility on Qualified Results

Percentage of Replies vs. Speed (ttl3, 30nodes)