MANCHESTER 1824
The University of Manchester
School of Computer Science

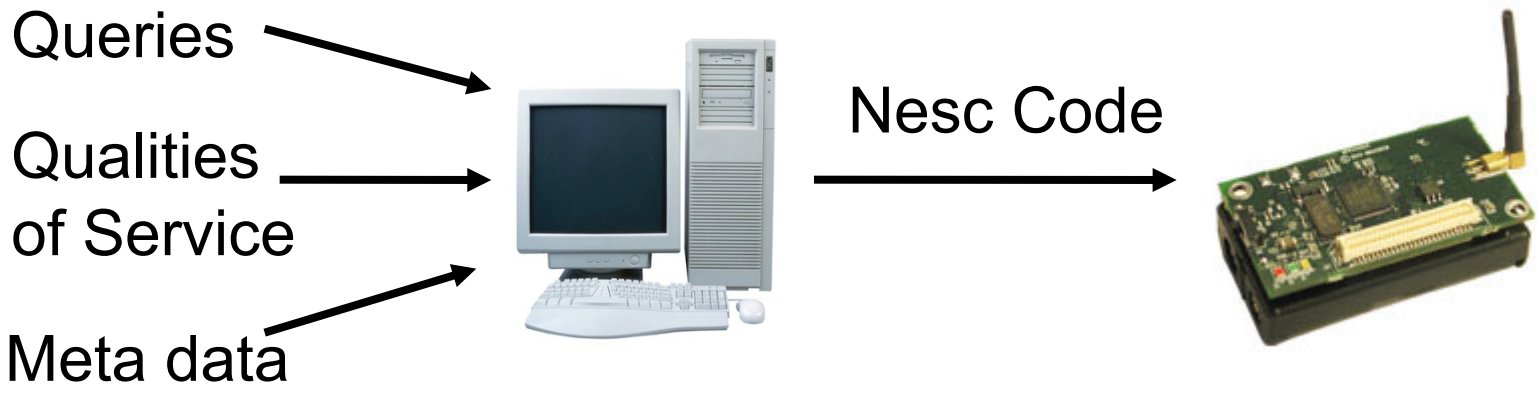
Validated Cost Models for Sensor Network Queries

Christian Y. A. Brenninkmeijer
Ixent Galpin
Alvaro A. A. Fernandes,
Norman W Paton,

Overview

- Introduction
- Motivation
- Example Distributed Query
- Optimization Goals
- Cost and possible decisions
  - Memory
  - Duration
  - Energy
- Conclusion

Introduction



- SNEEql Sensor NEtwork Engine query language
- Using Validated Cost Estimation Models allows the compiler to create an Optimal Query Execution Plan

Introduction and Motivation

- Sensor networks can be seen as Distributed databases.
- Distributed Queries don't do the same thing on every site.
- Query Optimizers decide what operations are done where.
- SNEE Optimizer decides when each operation is done including transmission.
- Query Optimizers depend on Cost Models.
- Transmission between sites must also be costed.

Sensor Networks have Known Input and can be Costed.

- Push stream difficult to cost due to an unknown tuple arrival rate.
- Sensor network users specify an Acquisition Rate and the sites they want to collect data from.
  - Streams therefore have a known tuples arrival rate
  - Streams, there have a known cardinality per acquisition
- Knowledge of input rate allows queries costed and therefore optimized and adapted.

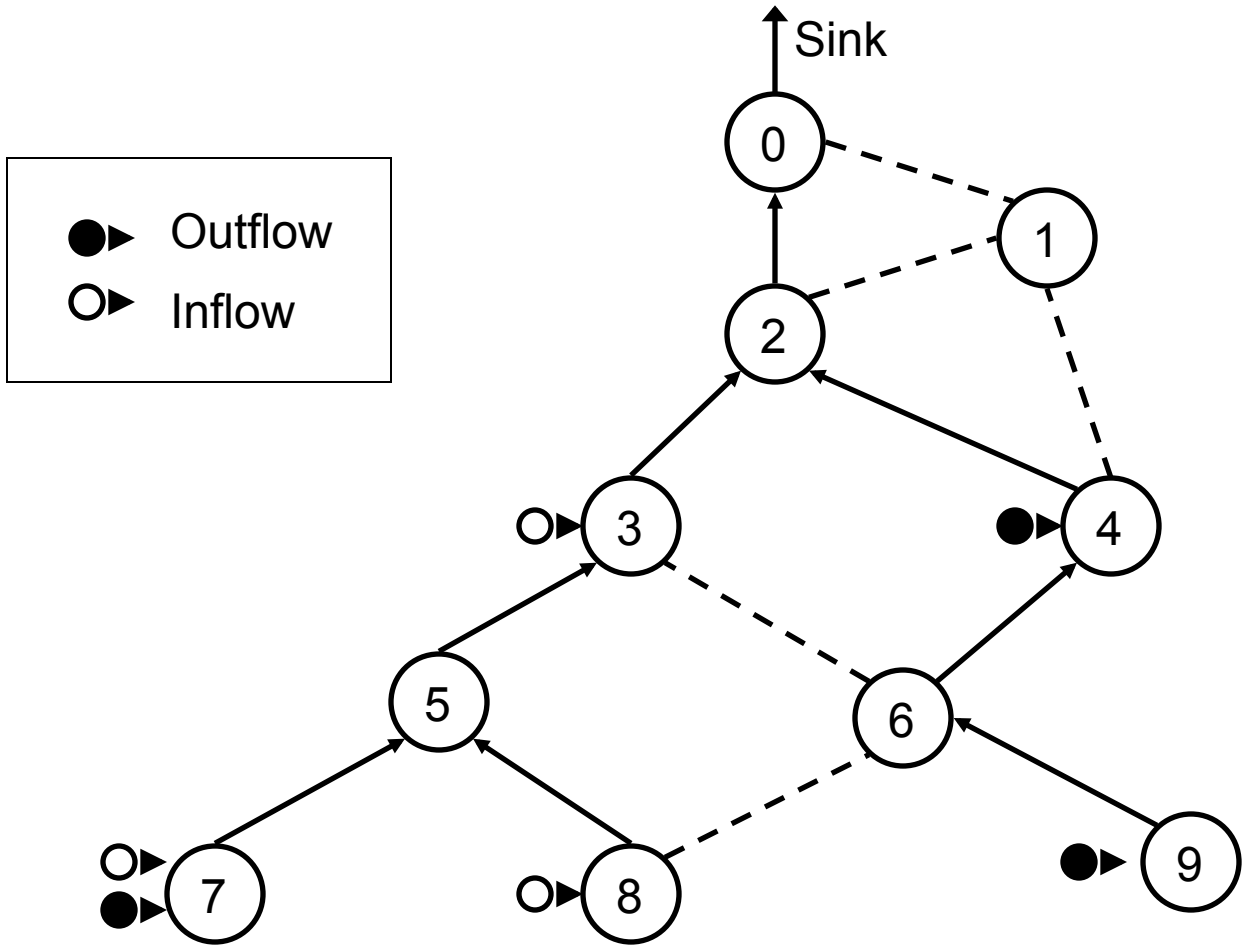
Example Query

schema: outflow (id,time,temp,pressure) sources: {4,7,9}
inflow (id,time,temp,pressure,ph) sources: {3,7,8}

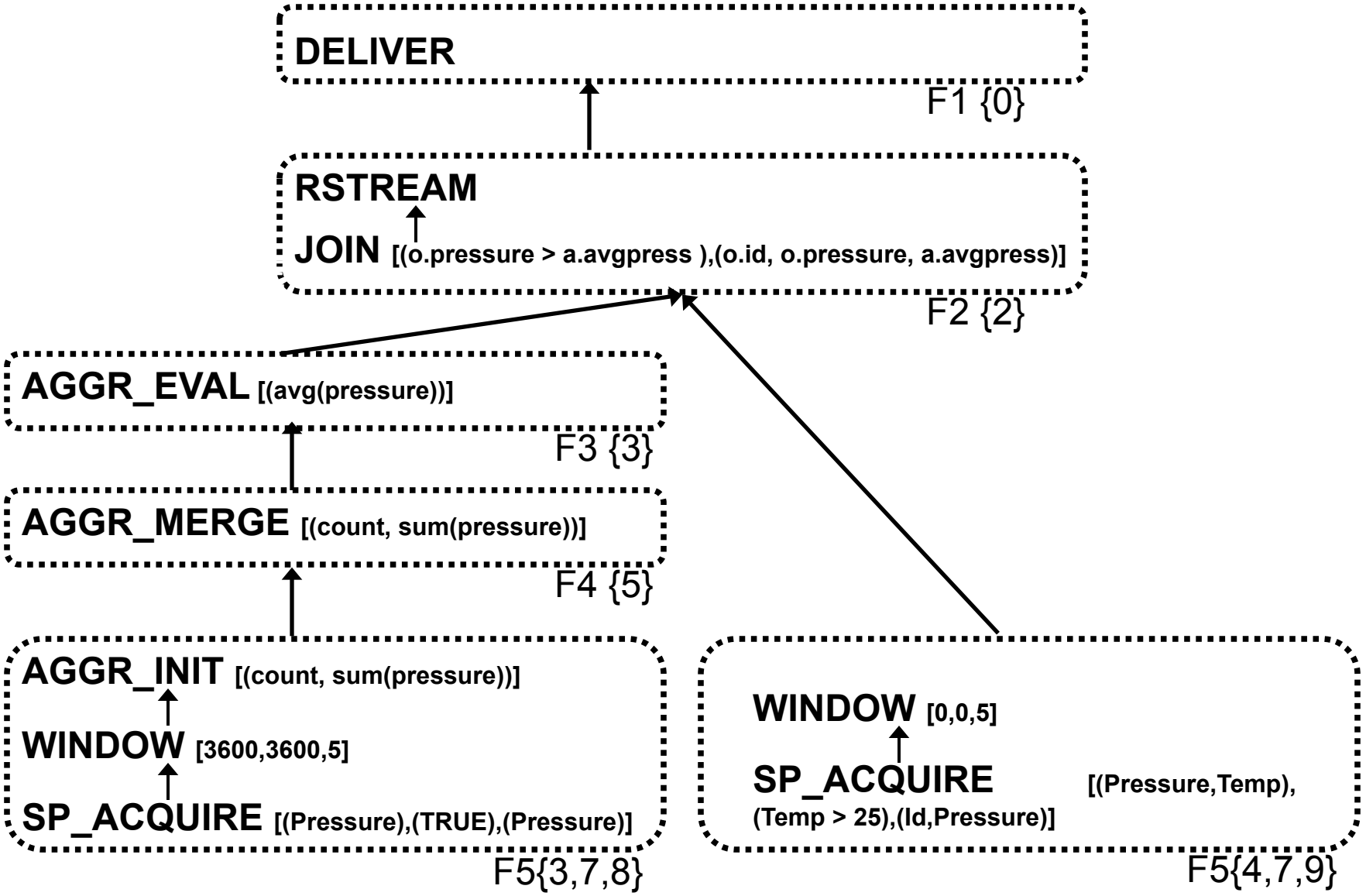
```
SELECT RSTREAM  o.id, o.pressure, a.avgpress
FROM   outflow[NOW] o,
          (SELECT avg(i.pressure) as avgpress
    FROM inflow[RANGE 1 Hour] i) a
WHERE o.pressure > a.avgpress
          AND o.temperature > 500;
```

QoS: {ACQUISITION RATE >= 5s Minimize;
DELIVERY TIME <= 30s}

Network Topology



Distributed Query Plan



Different Sites have Different Cost

- Different operators
- Different sensors used
- Different sized windows
- Aggregated and raw data
- Different number of tuples processed
- Different sizes of tuples

Example Costs

- Acquiring outflow cost more than inflow
  - Temperature and Pressure vs just Pressure
- Window on inflow cost more
  - Tuples held for one hour.
  - Affects mainly memory to hold 1 hours of tuples

Multiple contrasting Optimization Goal

- Traditional optimizer just consider duration.
- Sensor networks must be optimized for:
  - Memory
  - Duration
  - Energy
- Desiderata trade of against each other.
  - Actions which lower one cost often increase the others.

Memory Cost

- In Nesc all memory preallocated.
  - With only 4K Configuration EEPROM every byte counts
- Costs include
  - Overhead for each operator
  - Memory for temporary variables
    - Sensed data
    - Count and sum to compute average
  - Data stored for later evaluations
    - Ex: Window with 1 hours worth of data
  - Data passed to next operator
    - Size of tuple * Number of tuples
  - Data stored for transmission
    - Size of a packet to send.

Decisions based on Memory

- Is there enough RAM?
  - Will a whole hours worth of tuples fit?
    - If not reduce window or increase acquisition rate
    - Optimizer picks Acquistion Rate of 6 seconds
  - Can operators be grouped together or must they be divided over several sites.
- Can data be stored for later transmission?
  - Example up to 10 acquires fit on site 3
  - Reduces radio overhead
  - Allows more than one tuple per packet

Operator Duration

- Overhead
- Cost of receiving input from previous operator
- Cost of any processing
  - Selecting correct tuples for window
    - Depends on size of winodw
  - Calculating Aggregates
    - Example count and sum or average
  - Evaluating predicates
    - Depends on complexity of predicate
  - Doing the join
- Preparing data for output
  - Number of tuples * complexity of expression

SP_ACQUIRE Duration

- SP_ACQUIRE
  - Acquire
    - Sum of duration cost of each sensor used
    - Number and types on sensor critical
  - Select
    - Depends on number of atomic expressions
  - Project
    - Sum of cost of each expression
      - Depends on number of atomic expression

Transmit Duration

- Receiving data from Child / Storage
- Overhead cost
- Transmitting Tuples
  - Number of tuples
  - Size of each tuple
    - Number and type of attributes
  - Size of the packets
    - Number of tuples combined in a packet

Decision based on Duration

- Can the delivery time be met?
- Is there enough time to do everything required before next data arrives?
  - Does acquisition rate have to be increased?
- Can data be stored for later transmission?
  - Maximum of 5 acquires (6 seconds each) can be delivered
- Selecting the fastest query plan.
  - Reduce the number of relay sites
- Determining when activity takes place
  - Especially coordinating radio traffic between sites
  - Keeping duration of radio slots to an absolute minimum.

Energy Cost

- Based on Duration
- Depends on which physical devices used
  - Sensors
    - Depends on type of sensor
  - CPU
    - Active, Idle (radio on), sleep
  - Radio
    - Receive Mode
    - Transmit Mode (depends on level)
  - Flash

Decisions based on Energy

- How long will the network last
  - Can acquisition rate decreased?
- Where to place operators
  - Balance work between motes as much as possible
- Select the most energy efficient query plan.
  - Decrease transmission cost by increasing the number of relay sites
  - Decreasing time radios are kept on to absolute minimum.
- Trade-off between lifetime and other Q.O.S.
  - Acquisition rate, delivery time, query accuracy
  - Store 5 acquires to save energy

Conclusions

- Cost models allow Sensor networks to be treated as distributed databases
  - Different operators on different sites
  - Each operator and site costs indepenently
- Optimizers can choose the best query plan
  - To fit the limited Mote capacities
  - To meet various Qualities of Service
    - Based on Memory, Duration and Cost