Multiple Distributed Indexing Scheme for Supporting Energy-efficient Range Query in Data-Centric Storage Sensor Networks

XuanTung Hoang, Younghee Lee

Information and Communications University. Korea

Introduction

Single sensors → Application specific Sensor Networks → General purpose Sensor Networks

Simple computing application:
❖***Small-scale of sensing***
❖***Rough sensing data***

Collaborative sensing
❖Large scale and fine-grain sensing

Application specific
❖***Less flexible***

Shared sensor networks among applications
❖***Introducing more flexibility***

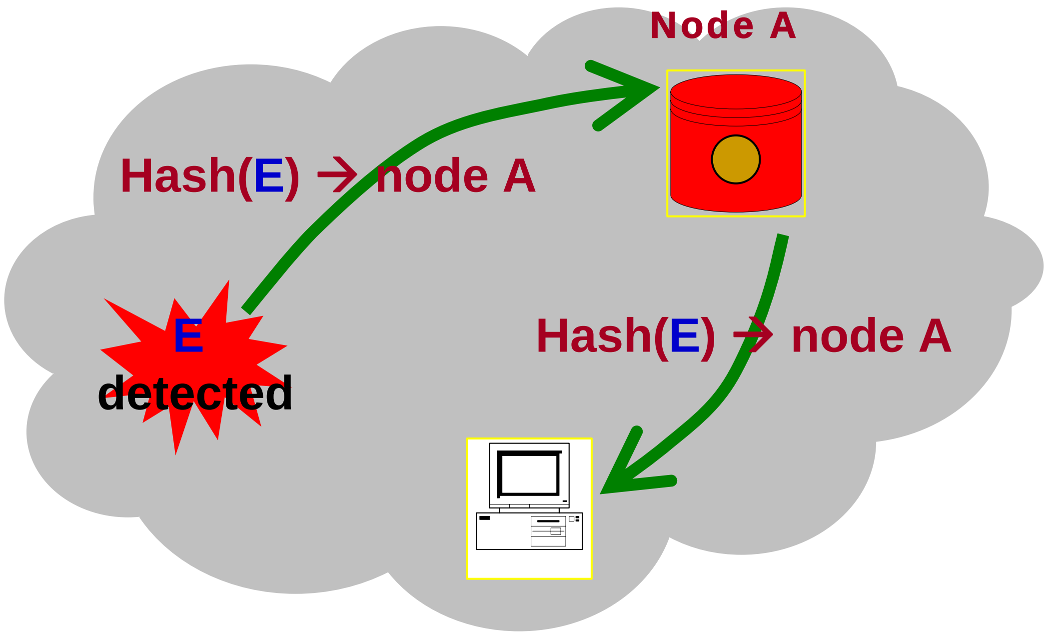
- Collaborative sensing: capability of in-network processing (aggregating, summarizing, in-network query processing, etc.) on sensor readings.
- General purpose sensor networks target to flexibility in terms of collaborative sensing

Our focus

- resolving *multi-dimensional range query* for general purpose Sensornets.
  - High level events inferred from various sensor readings
  - Multi-attribute events

Preliminary and Related Work

Data-Centric Storage (DCS) Sensor Networks



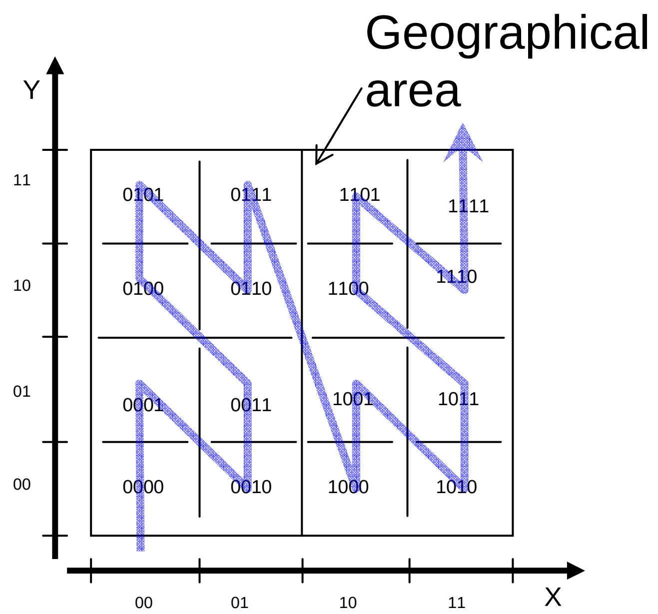
Query Event
Insert Event

- Data-dependent reservoir
- Hash function determines organization of data
- Efficiency depends on the underlay routing
  - GPSR is used

Preliminary and Related Work

Distributed Index for Multidimensional data (DIM)

- Follow DCS approach
- Support multidimensional range query
  - K-dimensional event space (K >= 2)
  - An event = 1 point in K-D space
- Mapping K-D space to geographical area
- Geographical area and data space are indexed with z-order Space-Filling Curve
  - Portion of k-D space to geographical zones
  - Mapping through z-value (***<u>code</u>***)
- Locality preserving hash

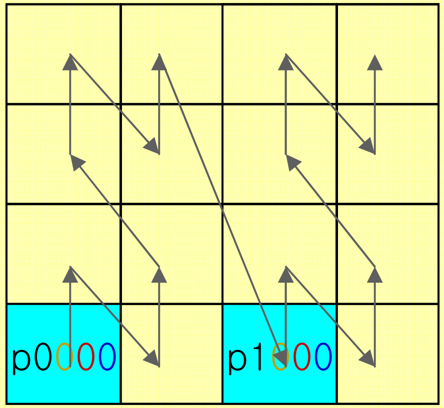


Nearby events in k-D space are nearby in 2D area

Query Splitting problem

- Locality preserving in DIM is not perfect

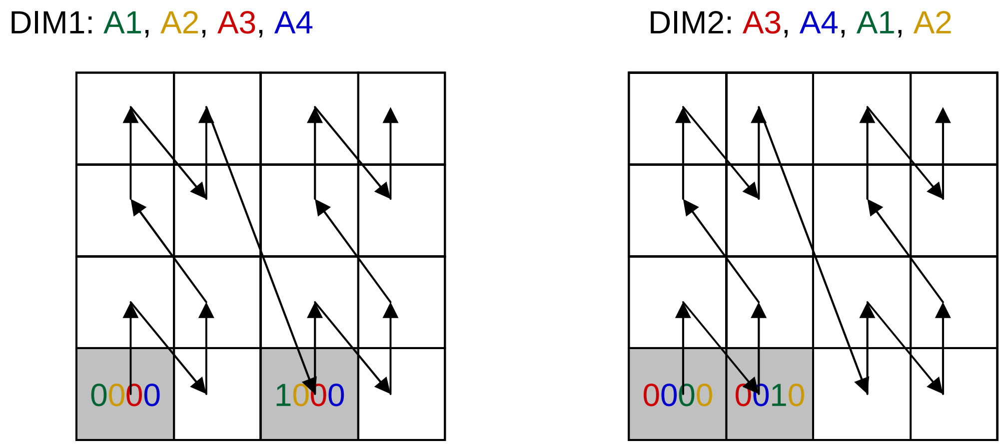
  - 4-D event space
  - Coordinates: A_1, A_2, A_3, A_4
  - Two adjacent data zones code *p*0000 and *p*1000 are mapped to two separated geographical zones
    - *p* is some common prefix



- A range query may be split into far away zones (and far away nodes)
  - Communication overhead (and energy waste) since routing in sensor networks depends on relaying via intermediate nodes

Our solution: MDI

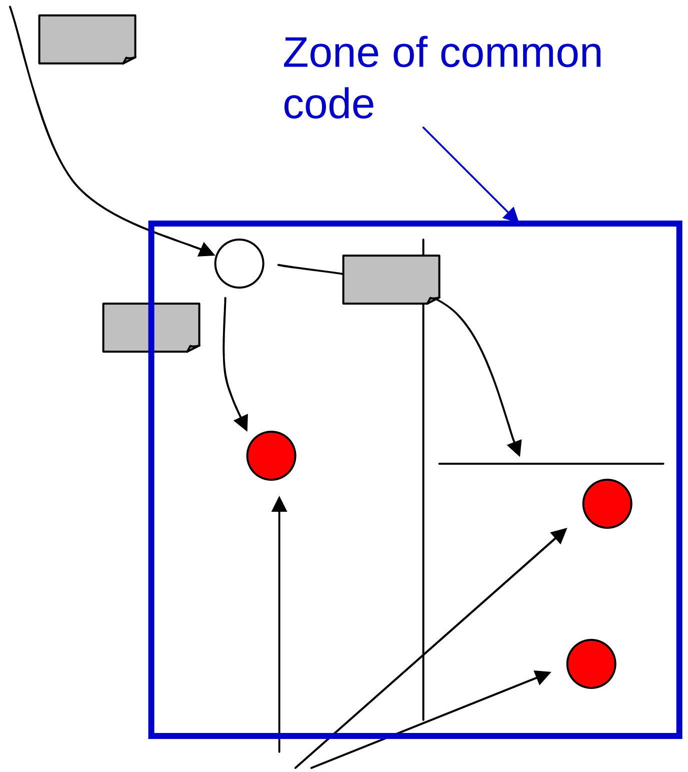
- Multiple distributed indices to enhance locality preserving
  - Use ⌈k/2⌉ DIMs to index data
  - Circularly shift the order of coordinates by 2 positions after each time of indexing
    - DIM1: A_1, A_2, A_3, …, A_k
    - DIM2: A_3, A_4, …, A_k, A_1, A_2
    - …



Our solution: MDI (cont.)

- Using multicast routing for event insertion to $\lceil k/2 \rceil$ DIM indices.
  - Route event message to *<u>common code</u>* of event replicas
  - Duplicate event messages (and partition replica set) when arrive to the *<u>zone of common code</u>*
  - Incrementally refine event codes (similar to DIM event routing)

Common code of event codes = longest common prefix



Our solution: MDI (cont.)

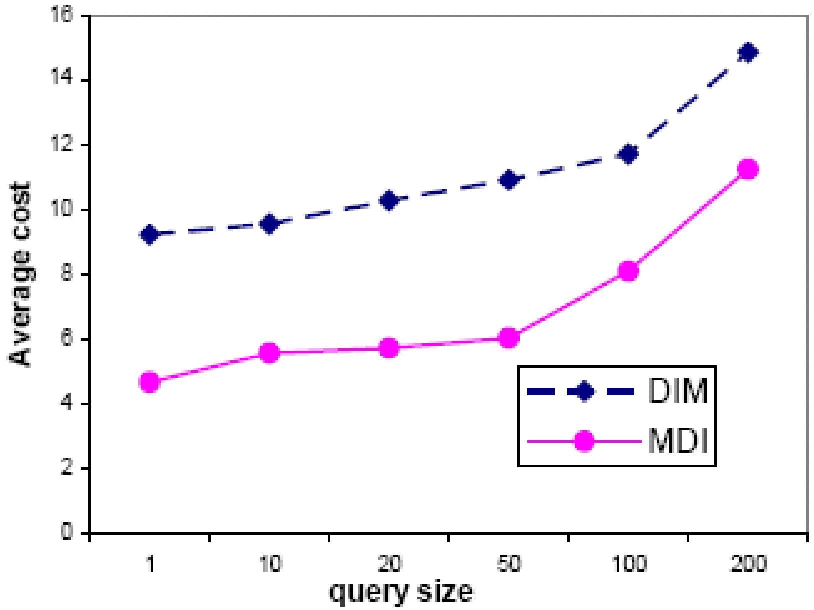
- Query processing
  - Concurrently lookup on $\lceil k/2 \rceil$ DIMs
  - Greedy query processing: Forward query message to the closest destination zones
  - Adopt query routing of DIM

Implementation and Simulation

- Based on DIM (Distributed Index for Multidimensional data) and ns2 simulation
- Simulation settings
  - 8-d data model (4 DIM indices)
  - Network size: 150 nodes on 200mx200m
- Performance metrics
  - Insertion cost
  - Query cost (number of messages for a query) for various range sizes

Result of pilot simulations

- Insertion cost:
  - Expected result: 4 times of insertion cost of DIM
  - Simulation result: only about 2 times
- Query cost:
  - 0.5 times of point query cost in DIM
  - Reduction in query cost decreases when query range increases



Conclusions

- Multiple indices and Shift-and-circulate attributes increase locality preserving property
- Reduce query cost but sacrifice insertion cost
  - Beneficial when there are few insertions and many queries (static environment)
- Increase reliability
  - data redundancy
  - DIM sub-indices are independent: data lost (because of node failure) in one DIM does not affect other DIM.
  - need further evaluation