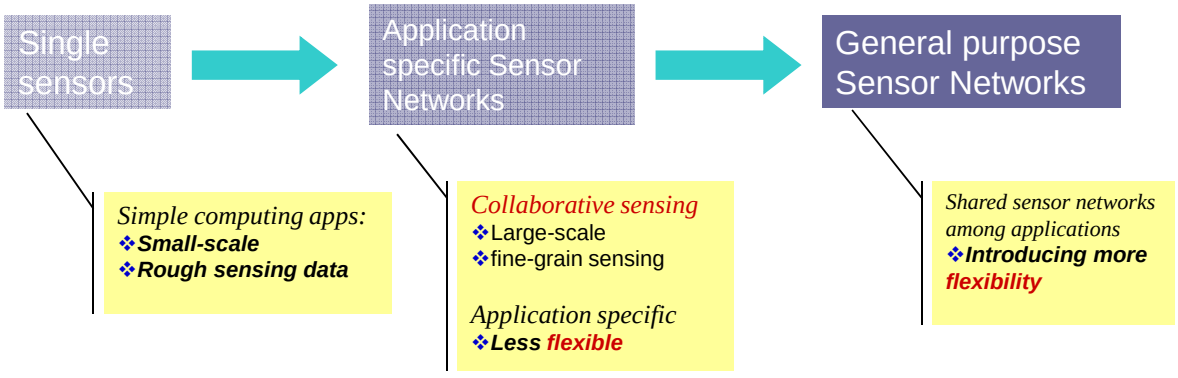
Multiple Distributed Indexing Scheme for Supporting Energy-efficient Range Query in Data-Centric Storage Sensor Networks

XuanTung Hoang, Younghee Lee
Information and Communications University, Korea

Computer Networks Lab

Introduction

- Sensor networks could participate in the Future Internet
  - Sensor networks as single computing entities
  - *"Browsing the physical world"* → Network of Sensornets



Collaborative sensing: capability of in-network processing (aggregating, summarizing, in-network query processing, etc.) on sensor readings.

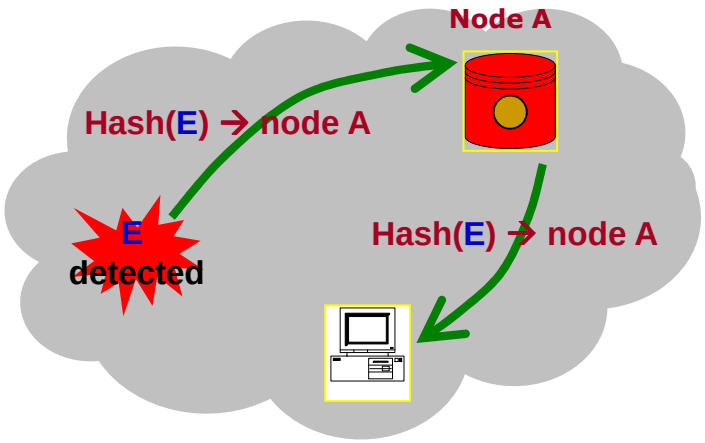
Computer Networks Lab

Our focus

- *Multi-dimensional range query* is important for general purpose Sensornets.
  - Sensor networks as ubiquitous accessed distributed databases
  - High level detected events inferred from various sensor readings → Multi-attribute events
- A realization of *Data-Centric Architecture*
  - … which seems to be the dominant trend of the Future Internet

Preliminary and Related Work

Data-Centric Storage (DCS) Sensor Networks

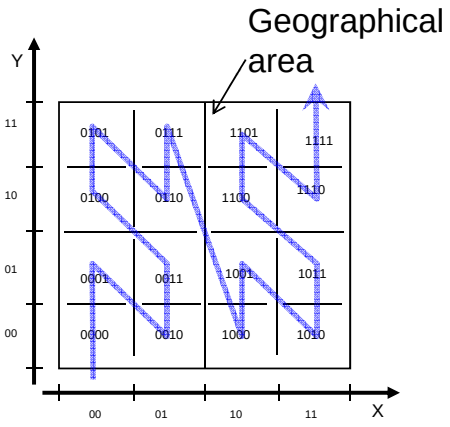


- Data-dependent reservoir
- Hash function determines organization of data
- Efficiency depends on the underlay routing
  - E.g.: GPSR (*Greedy Perimeter Stateless Routing*)

Preliminary and Related Work

Distributed Index for Multidimensional data (DIM)

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

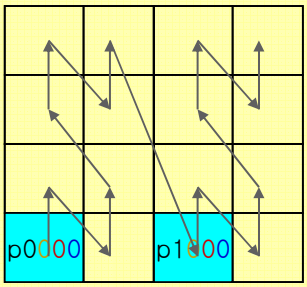


Locality Preserving Hash =	Nearby events in k-D space are nearby in 2D area

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

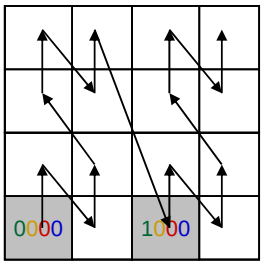


- A range query may be split into far away zones (and far away nodes)
  - Communication overhead (and energy waste) since routing in sensor networks depends on relaying via intermediate nodes

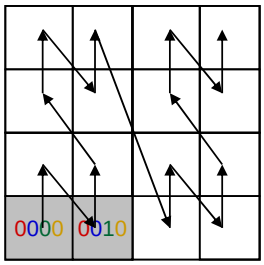
Our solution: MDI

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

DIM1: A1, A2, A3, A4



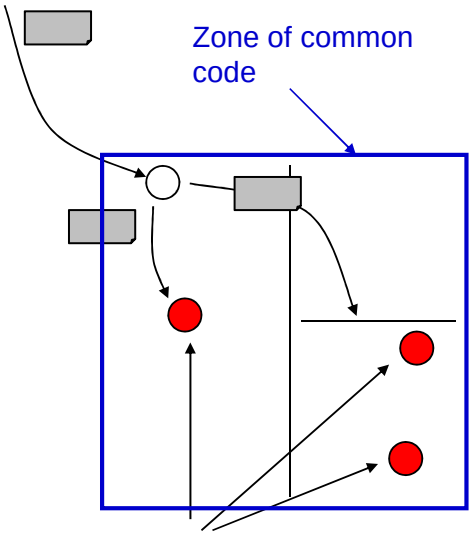
DIM2: A3, A4, A1, A2



Our solution: MDI (cont.)

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

Common code of event codes = longest common prefix



Our solution: MDI (cont.)

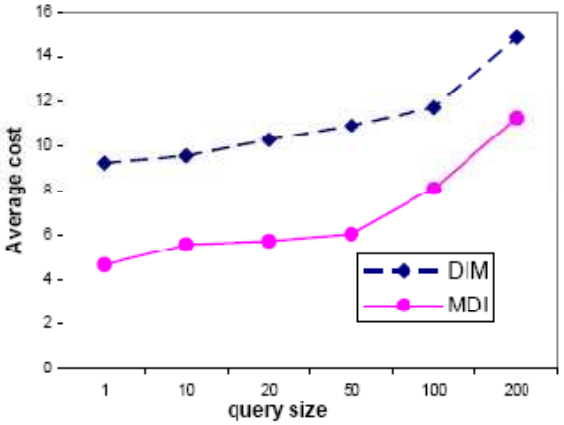
- Query processing
  - Adopt and modify DIM query routing
  - *Concurrently lookup* on ⌈k/2⌉ DIMs
  - *Greedy query processing*: Forward query message to the closest destination zones

Implementation and Simulation

- Based on DIM source code and ns2 simulation
- Simulation settings
  - 8-d data model (4 DIM indices); 1024 values per dimension
  - 150 nodes on 200mx200m sensor field
- Performance metrics
  - Insertion cost
  - Query cost 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 satisfy 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 DIMs.
  - need further evaluation