Efficient Access to Information in Large Sensor Networks

Ramesh Govindan
University of Southern California
ramesh@usc.edu

June 12, 2003

Introduction

The Challenge:

- Sensed data will be stored within a sensor network, for energy-efficiency
- How to scalably access this information?

Introduction

The Challenge:

- Sensed data will be stored within a sensor network, for energy-efficiency
- How to scalably access this information?

Questions:

- What do we mean by *information*?
- What kinds of access? From where? By whom?
- What do we mean by scale?

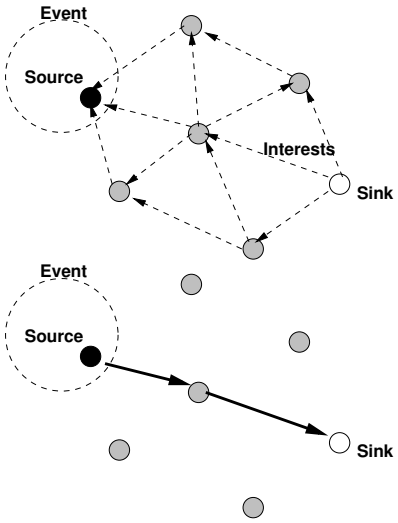
Early Work

Flood-then-respond systems

- Diffusion
- TAG
- Two-tier Data Dissemination

Good for

- Long-lived/continuous queries
- Issued by users outside the network



An Alternative Class of Queries

One-shot queries . . .

. . . issued from *within the network*

- Correlating events at different nodes
- Testing for conditions within the network

These *trigger* actions within the network, and can be expected to be frequent.

Flood-then-respond doesn't scale for this class

Data-Centric Storage

Store information within the network, in such a manner as to enable low-cost *rendezvous* [5]

- Avoid flooding
- . . . but this comes at a cost

What do we mean by information?

- Events
- Features

One Instance of Data-Centric Storage

Geographic Hash Tables (GHTs) [4]

- *Hash* the name of an event to a geographic location
- Use geographic routing to store data at the node *nearest* to the location
- Can retrieve data the same way

Structured Replication (SR)

- Avoids hotspots, trades off increased query latency
- Tesselate the sensor field into sub-fields
- Store data in node within own sub-field
- Query traverses all sub-fields

Generalizing These Ideas

Hashing and geographic tesselation can be used to build a variety of *distributed data structures*, supporting different kinds of information access.

DIFS [2], a distributed index for sensor network features

DIMENSIONS [1], a hierarchical, distributed search structure for wavelet coefficients of sensor network data

This is good, because we can hope to leverage a common software framework for building other kinds of distributed storage structures

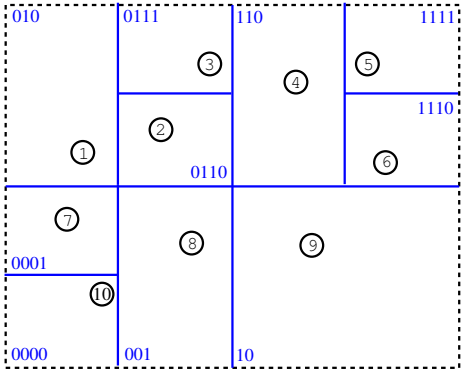
Distributed Index for Multi-dimensional Data

Query: List all events whose temperature lies between 50° and 60°, and whose light levels lie between 10 and 15.

To support this query:

- Use a locality-preserving geographic hash
- Tesselate the sensor-field non-uniformly such that each event maps to a *zone*

Of course, this works well for small query "boxes" [3], but that's what you'll need for data correlations anyway



Future Directions

Data structures for:

- Spatial Indexing
- Computing Aggregates

Systems

- A general purpose programming system for sensor networks, based on these data structures
- Integration with sensor network databases

References

[1] Ganesan, D., Estrin, D., and Heidemann, J. DIMENSIONS: Why do we need a new data handling architecture for sensor networks? In *Proceedings of the ACM Workshop on Hot Topics in Networks* (Princeton, NJ, USA, Oct. 2002), ACM.

[2] Greenstein, B., Estrin, D., Govindan, R., Ratnasamy, S., and Shenker, S. DIFS: A Distributed Index for Features In Sensor Networks. In *Proceedings of the IEEE ICC Workshop on Sensor Network Protocols and Applications* (Anchorage, AK, April 2003).

[3] Li, X., Kim, Y. J., Govindan, R., and Hong, W. Multi-dimensional Range Queries in Sensor Networks. submitted for publication.

[4] Ratnasamy, S., Karp, B., Yin, L., Yu, F., Estrin, D., Govindan, R., and Shenker, S. GHT: A geographic hash table for data-centric storage. In *Proceedings of the ACM Workshop on Sensor Networks and Applications* (Atlanta, Georgia, USA, Sept. 2002), ACM, pp. 78–87.

[5] Shenker, S., Ratnasamy, S., Karp, B., Govindan, R., and Estrin, D. Data-Centric Storage in Sensornets. In *Proc. ACM SIGCOMM Workshop on Hot Topics In Networks* (Princeton, NJ, 2002).