Optimizing Data Acquisition in Wireless Sensor Networks

Bhaskar Krishnamachari
Assistant Professor
Department of Electrical Engineering-Systems
University of Southern California
bkrishna@usc.edu

September 29, 2003

Overview of Talk

- Wireless sensor networks can be viewed as providing an interface between the real and virtual worlds. In its simplest form this is a one-way interface involving the acquisition of information from the physical environment.

- Given the severe energy constraints, however, the traditional networked system design methodology (based on intuitions, heuristics, and experience) must necessarily be complemented with mathematical modeling and optimization *a priori*.

- We will present three case studies where analysis provides useful design insights.

Acknowledgements

The work described in this talk has been performed in collaboration with several faculty and students at USC:

- R. Govindan, A. Helmy, F. Ordonez
- N. Sadagopan, S. Pattem

Case study I

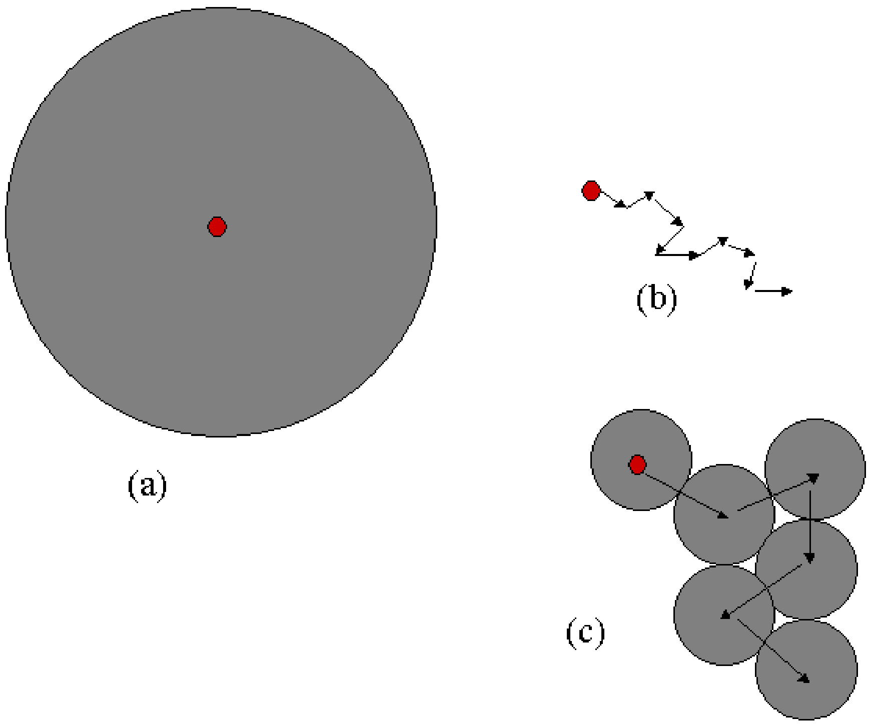
ACQUIRE: Active Query Forwarding in Sensor Networks

- The choice of query dissemination and response technique matters most for one-shot queries (for long-term flows, even the cost of a flooded query can be amortized).

- Flooding (FBQ) is one extreme; sending the query on a (random / predetermined / gradient-based) trajectory (TBQ) is another.

- ACQUIRE provides a flexible tradeoff between these two extremes with a controlled d-hop flood at each step of a trajectory-based query forwarding. ($d = 0 \Rightarrow$ TBQ, $d = D \Rightarrow$ FBQ).

ACQUIRE: Active Querying in Sensor Networks



(a) Flooding-based querying (b) Trajectory-based querying (c) ACQUIRE

Analysis of ACQUIRE

Can analytically derive the expected number of transmission required to query for the information and obtain a response back to the sink using ACQUIRE with a d-hop look-ahead:

$$E_{avg} \approx \frac{\widehat{\sigma}}{f(d)}(c(f(d) + g(d)) + 2d) \tag{1}$$

$f(d)$: expected number of nodes within d hops of an active node

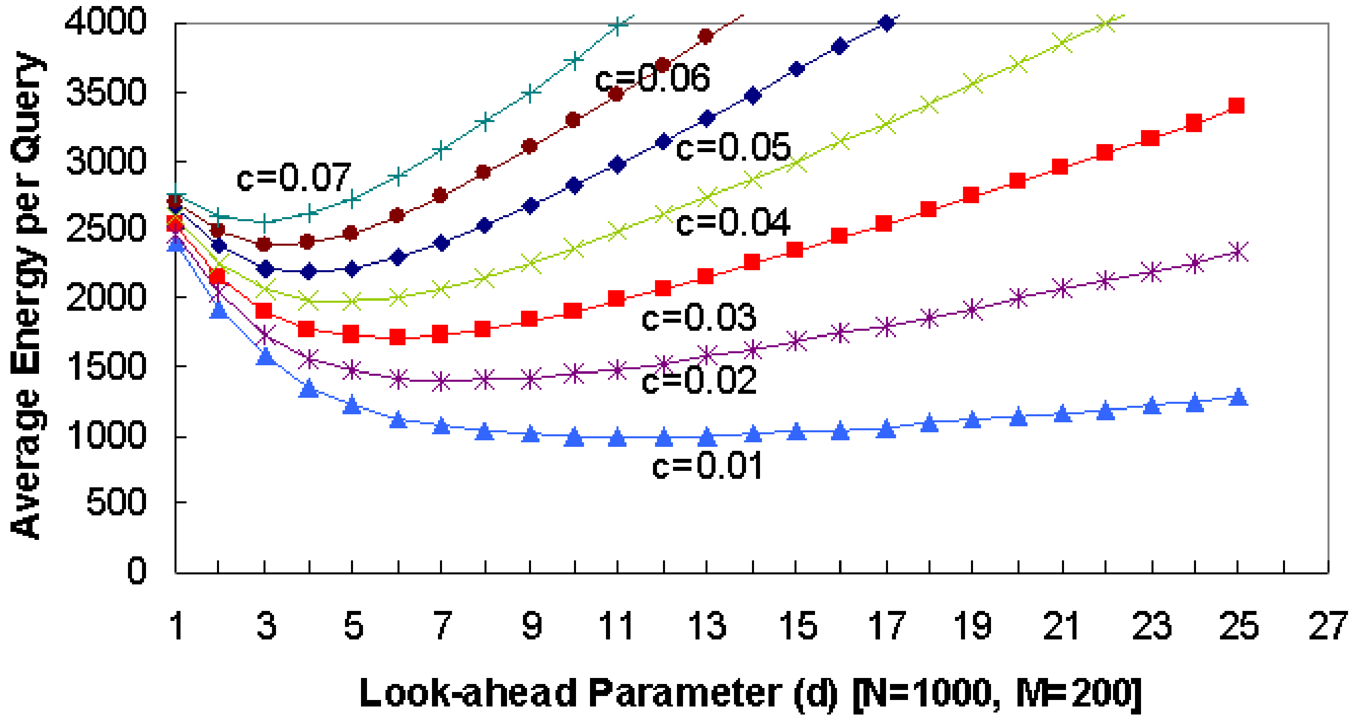
$g(d)$: expected number of messages required for all nodes within d hops to respond

$\widehat{\sigma}$: expected number of unique nodes that must be queried

c: Expected number of updates per query ($c < 1$ with caching)

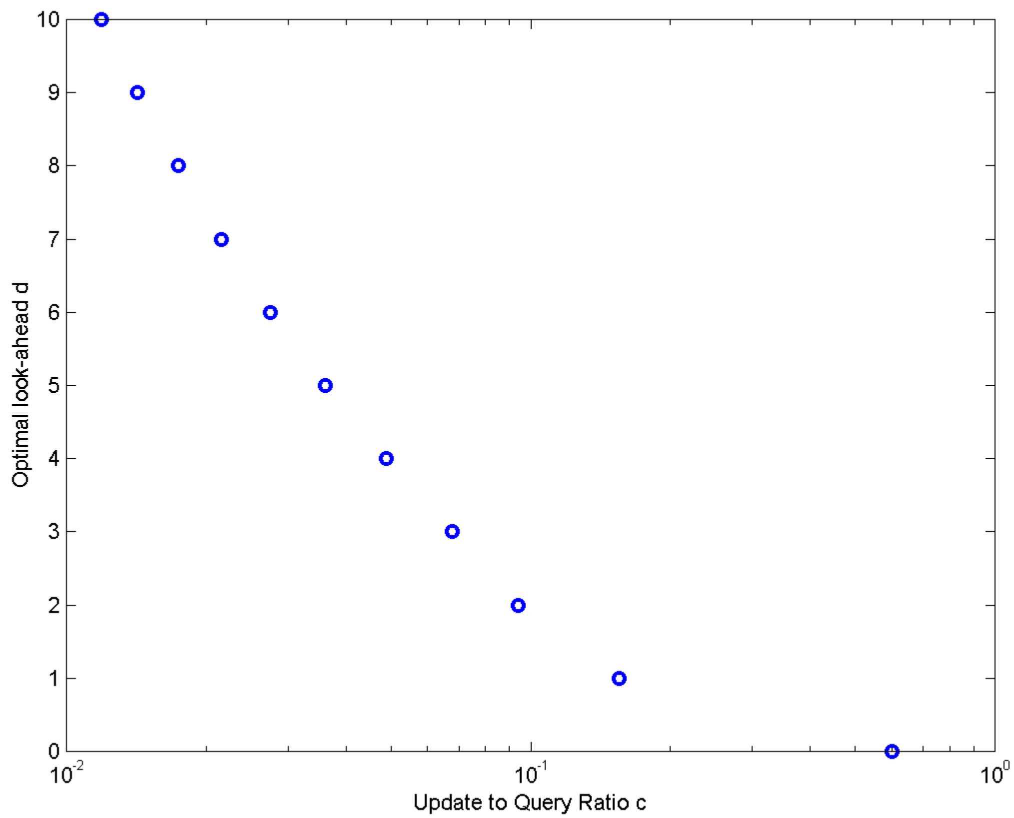
Analysis of ACQUIRE

There exists an optimal, intermediate value of d which depends on the data dynamics, as measured by the factor c.



Analysis of ACQUIRE

When c is high, the optimal value of d is low (more like TBQ) and *vice versa*



Case study II

Impact of Spatial Correlations on Routing with Compression

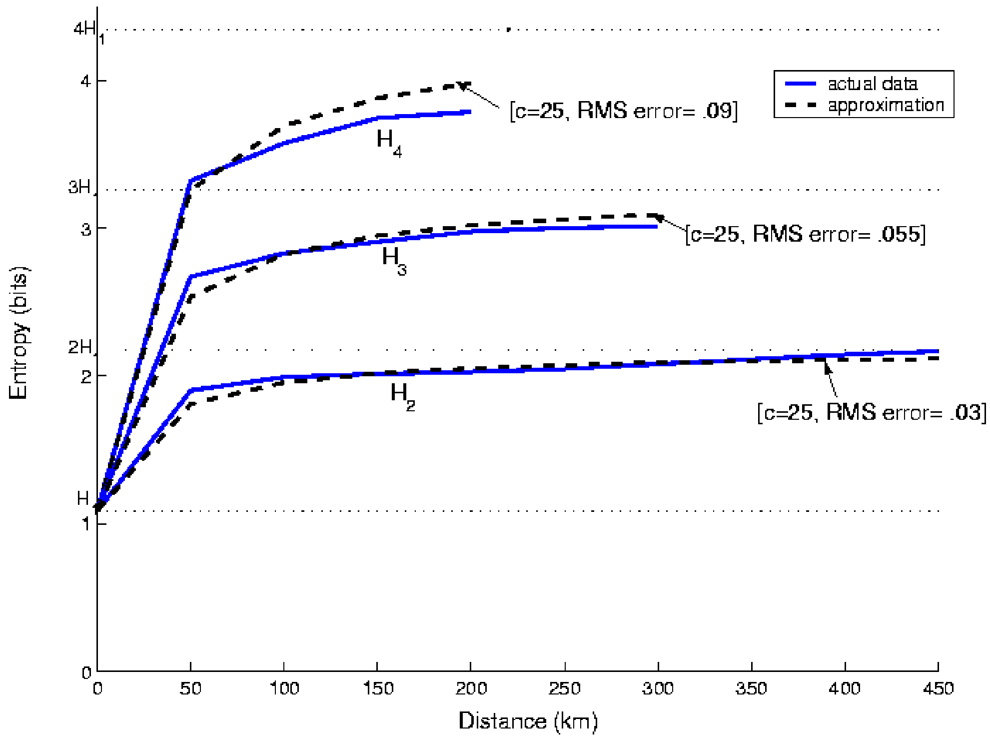
- One of the unique features of sensor networks is the possibility of data-centric routing mechanisms which combine routing with in-network data aggregation.

- Consider an application that involves the gathering of spatially correlated information from a set of nodes in the network. The total bits transported can be minimized if data from correlated sources is compressed enroute.

Modeling Spatial Correlations

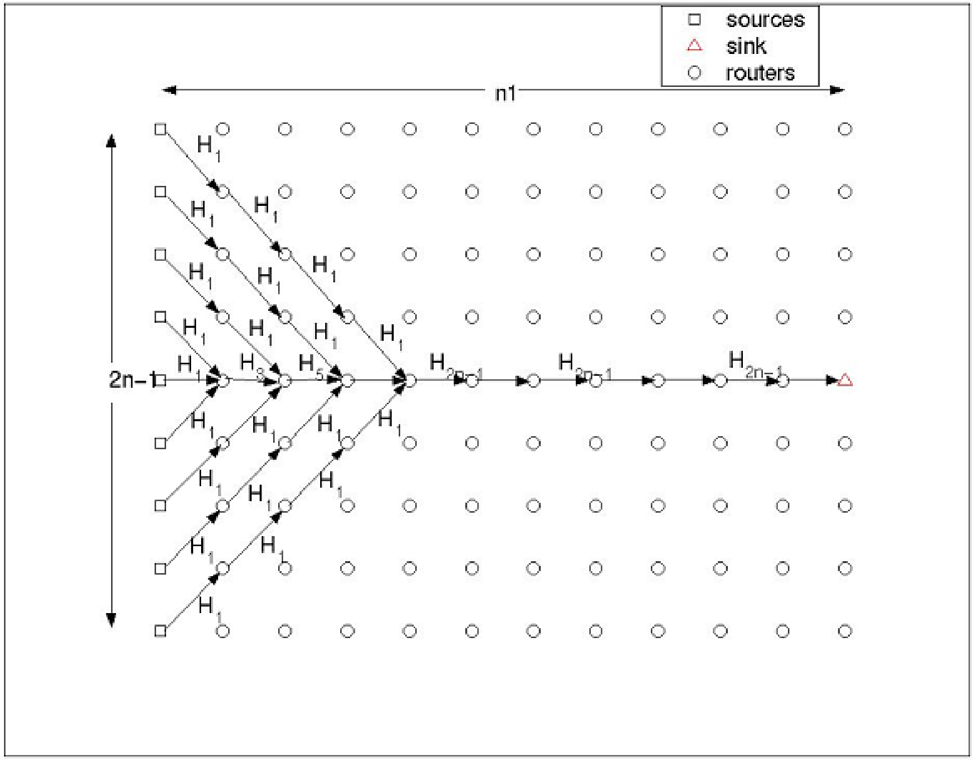
An empirically-founded parameterized model for the joint entropy H_n of n sources that are placed linearly at a distance d from each other:

$$H_n(d) = H_1\left(1 + \frac{(n-1)d}{d+c}\right) \tag{2}$$



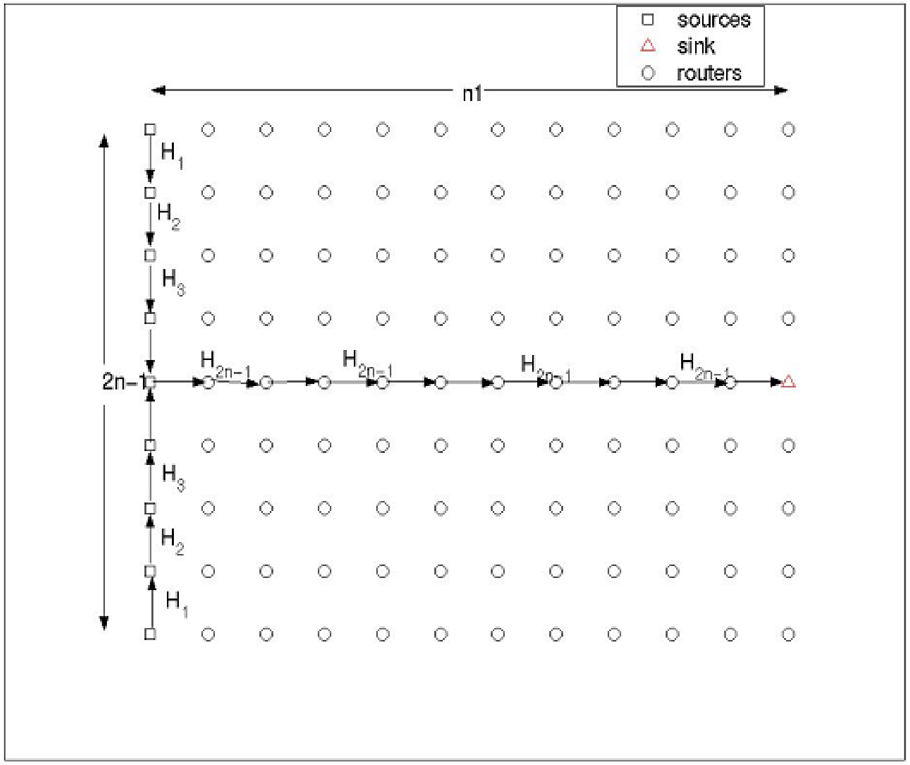
Routing Driven Compression

Route information from each source along shortest paths to sink, and compress wherever the routes happen to overlap.

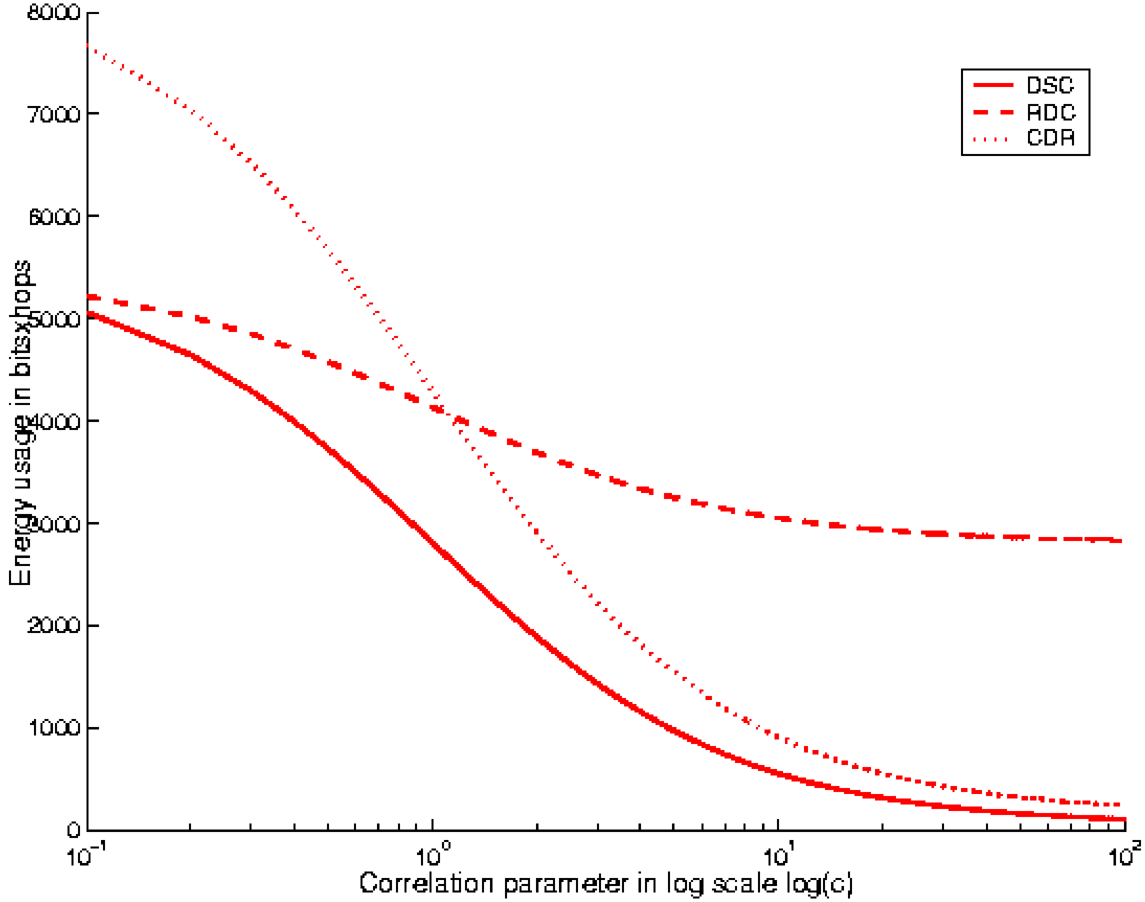


Compression Driven Routing

Route the data so that compression occurs as close to the source as possible (the routes need not be shortest path to sink)



Comparison based on Spatial Correlation



Cluster-based compression

- One way to combine RDC and CDR is to generate clusters of size s — perform CDR within the cluster, and RDC from the cluster-heads to the sink. This can be analyzed for a simple scenario involving n linearly placed sources on the grid.

Analysis of Cluster-based compression

- Total energy cost, using a cluster size s:

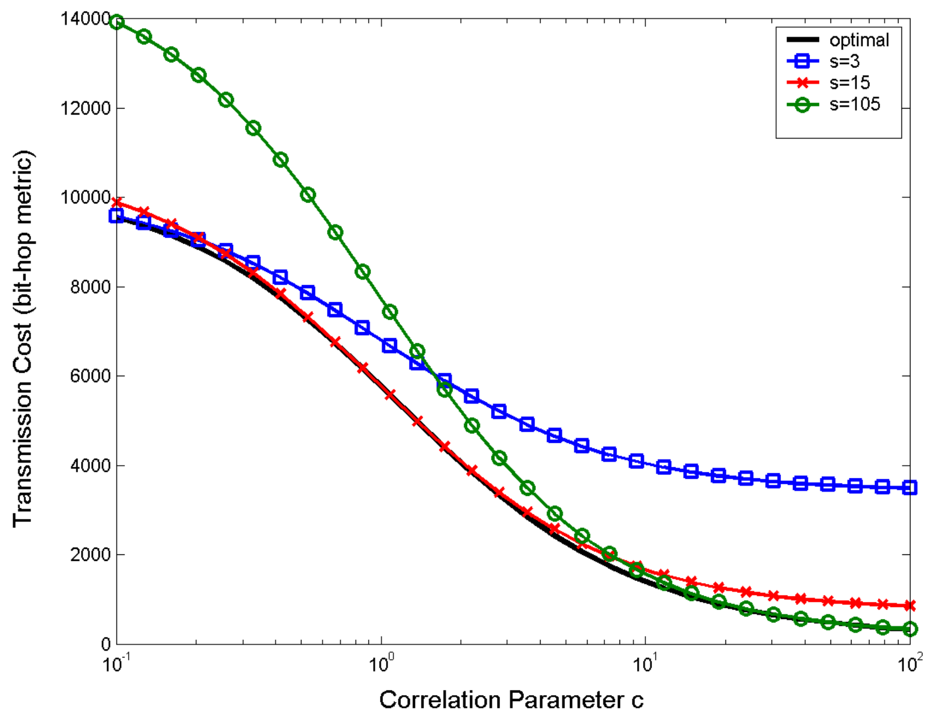
$$E_s(c) = nH_1[1 + \frac{(s-1)}{2(1+c)} + \frac{n}{s} + \frac{(s-1)n}{(s)(1+c)}] \tag{3}$$

- By setting derivative to zero, can determine the optimal cluster size as a function of the spatial correlation c:

$$s_{opt} = \sqrt{2nc} \tag{4}$$

Analysis of Cluster-based compression

105 linearly placed sources at one end of a 105x105 grid. The following figure suggests that there may be a "near-optimal" cluster size that works well for a wide range of correlation values.



Near-Optimal Clustering?

- We can formalize the notion of a near-optimal cluster size as follows: find a cluster size $s = s_{no}$ that minimizes the maximum-difference metric:

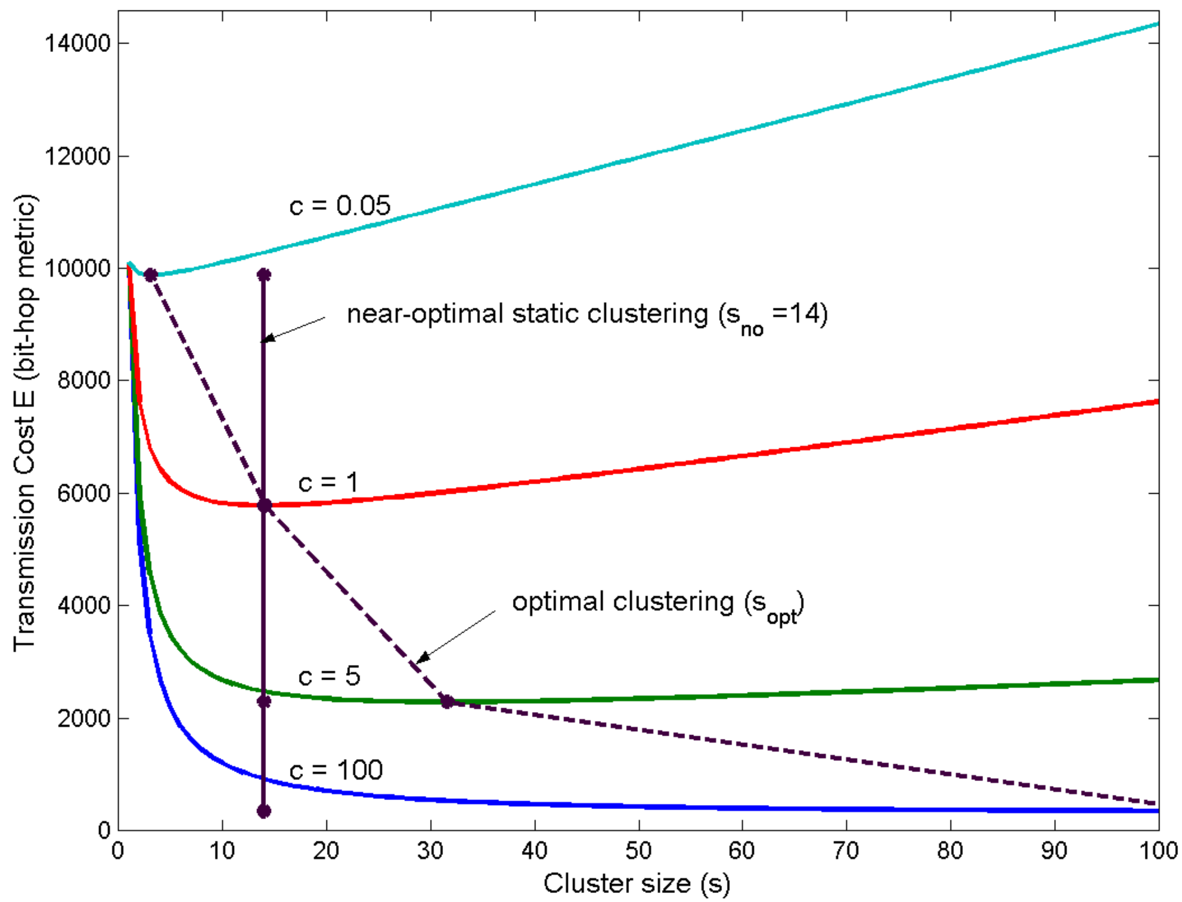
$$\min_{s} \max_{c \in [0,\infty)} |E_s(c) - E_{s_{opt}}(c)| \tag{5}$$

- It can be shown that

$$s_{no} \approx \frac{\sqrt{8n+1} - 1}{2} \tag{6}$$

- $n = 105 \Rightarrow s_{no} \approx 14$

Static Near-Optimal Clustering



Static Near-Optimal Clustering

- Our simulations show that this behavior applies to 2D scenarios as well.

- This is an analytical result with great *practical* significance. It suggests that that routing paths could be constructed statically with a carefully chosen cluster-size and need not be adapted dynamically with changing spatial correlations.

Case study III

The Maximum Data Extraction Problem

- Consider a network with specified node locations and pairwise communication costs, where each node has a (significant amount of) stored sensor data D_i it wishes to provide to the sink, and each nodes has a specified limited energy resource E_i. We focus on the situation where there is more data stored network-wide than can possibly be extracted.

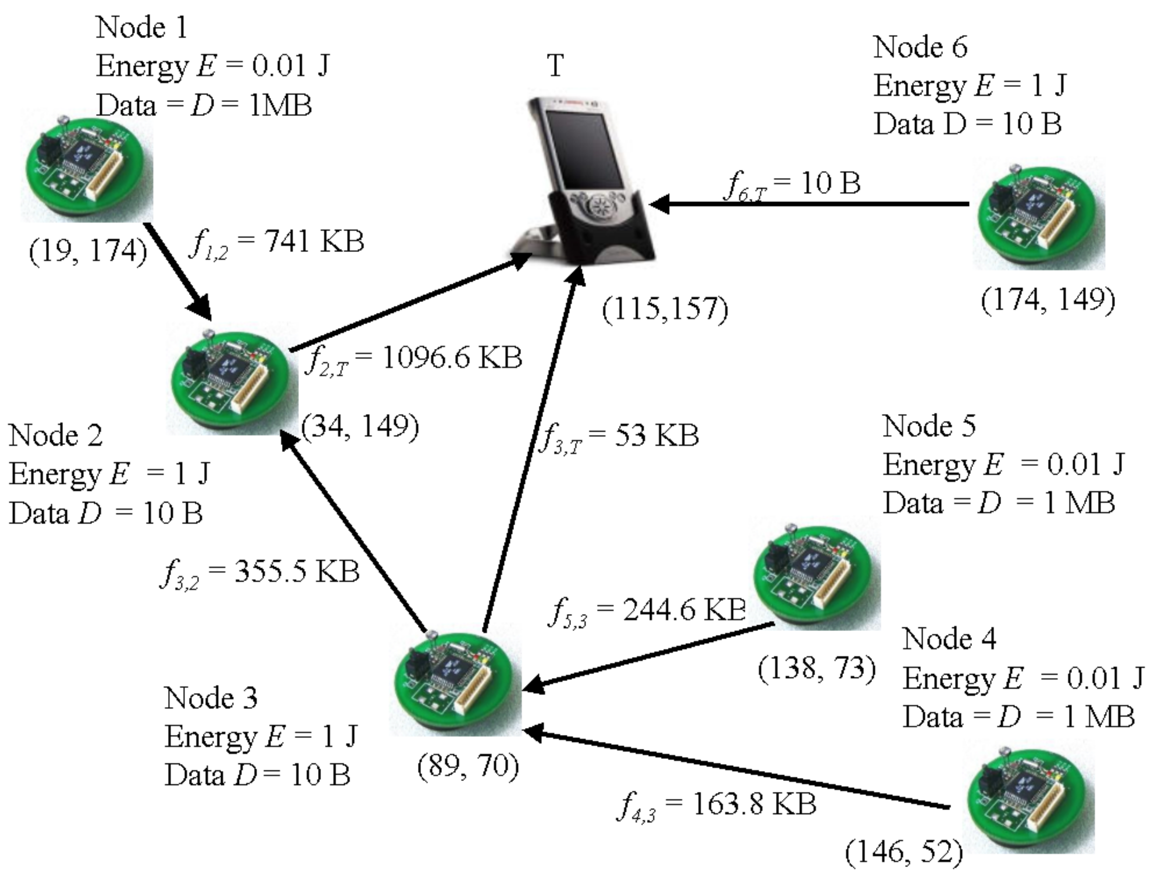
- Question: What multi-hop routing policy will maximize the total amount of data that can be extracted from this network?

The Maximum Data Extraction Problem

- This problem can be formulated as a network flow-based linear-program:

$$
\begin{aligned}
\text{Maximize} \quad \sum_{i=1}^{N} f_{i,N+1} \quad & \text{such that} \\
\sum_{j=1}^{N+1} f_{i,j} - \sum_{j=1}^{N} f_{j,i} \; & \leq \; D_i \\
\sum_{j=1}^{N+1} f_{i,j} - \sum_{j=1}^{N} f_{j,i} \; & \geq \; 0 \\
\beta \sum_{(i,j)} f_{i,j} d_{i,j}^2 + \sum_{(j,i)} f_{j,i} \; & \leq \; E_i
\end{aligned}
\tag{7}
$$

Illustration of Sample Solution



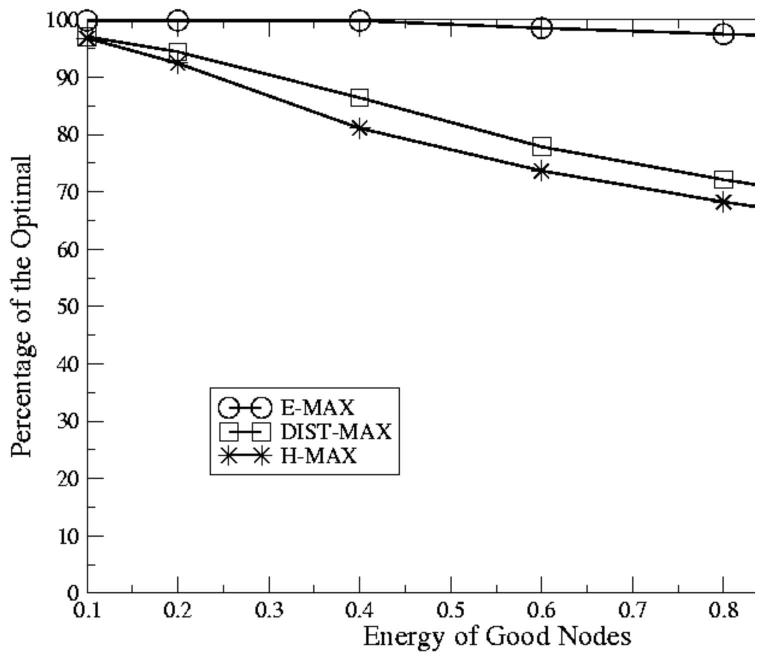
Practical considerations

- Collecting the global state at a central node, solving then distributing the LP solution is infeasible in a distributed practical scenario.

- Can develop a heuristic iterative distributed algorithm (E-MAX) based on the qualities of the observed LP solutions

Distributed Algorithm for MDE

The algorithm (called E-MAX) we developed based on the LP solutions incorporates into its metric both the remaining energy and data level at each node and iteratively draws all possible data from the node with the current shortest-metric path. Simulations suggest that E-MAX performs near-optimally in a wide range of settings.



Conclusions

- We have seen three case studies of analytical approaches that yield practical insight into the design of energy-efficient querying and routing techniques for sensor networks.

- More "constructive" analytical studies are needed to complement the ongoing simulation/implementation work in this space.

References

All papers can be obtained at ceng.usc.edu/~bkrishna/ or by sending email to bkrishna@usc.edu.

- Narayanan Sadagopan, Bhaskar Krishnamachari, and Ahmed Helmy, "Active Query Forwarding in Sensor Networks (ACQUIRE)," accepted in 2003 to appear in the Elsevier journal on Ad Hoc Networks.

- Narayanan Sadagopan, Bhaskar Krishnamachari, and Ahmed Helmy, "The ACQUIRE Mechanism for Efficient Querying in Sensor Networks," IEEE International Workshop on Sensor Network Protocols and Applications (SNPA'03), Anchorage, Alaska, May 2003.

- Sundeep Pattem, Bhaskar Krishnamachari, and Ramesh Govindan, "The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks," in submission, July 2003.

- Narayanan Sadagopan and Bhaskar Krishnamachari, "Maximizing Data Extraction in Energy-Limited Sensor Networks," in submission, June 2003.