An Evaluation of Multi-Resolution Storage for Sensor Networks

SenSys'03 Paper by Deepak Ganesan, Ben Greenstein, Denis Perelyubskiy, Deborah Estrin, and John Heidemann

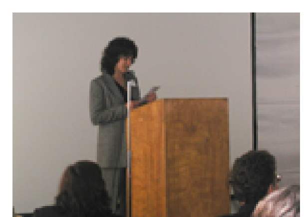
CPSC 538A Presentation: Georg Wittenburg

partly based on slides by Deepak Ganesan

Background of the Paper

- Authors:
  - Deepak Ganesan – Ph.D. Candidate, UCLA
  - Ben Greenstein – Ph.D. Candidate, UCLA
  - Denis Perelyubskiy – Completed M.S., UCLA
  - Deborah Estrin, Ph.D. – Professor of CS, UCLA; Director, Center for Embedded Networked Sensing (CENS); Associate Editor, ACM Transactions on Sensor Networks
  - John Heidemann – Assistant Professor, USC


?




The Truth about Sensor Networks

The one big, huge, fundamental truth about sensor networks is:

The Truth about Sensor Networks

The one big, huge, fundamental truth about sensor networks is:

"Resources are limited so don't waste them!"

(Just in case someone missed that… ☺)

Motivation

- So which resource do we concentrate on this time?

Storage

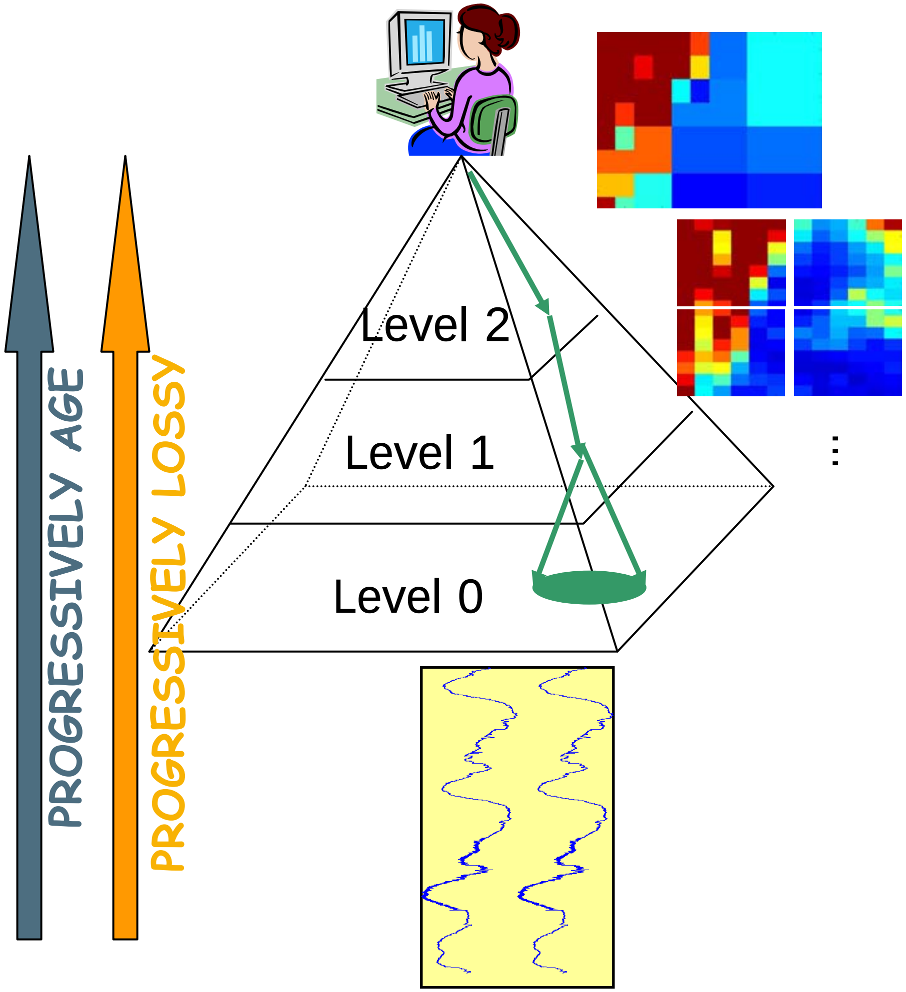
- Setting:
  - A lot of data will be generated by the sensor network over time, i.e. continuous measurements rather than discrete events.
  - At the time of deployment, no knowledge exists exactly what kind of queries will be performed.

Proposed Solution *("Paper-on-a-Slide")*

- Organize sensor nodes hierarchically and summarize the data gathered at each level.
  - This allows for "drill-down queries" that retrieve data from the network when requested, while still providing interesting information at the top level.
- Allow for graceful degradation in quality of replies to queries by aging summaries.
  - Older data is gradually removed from the network.
  - More useful summaries are retained longer.

DIMEMSIONS Architecture

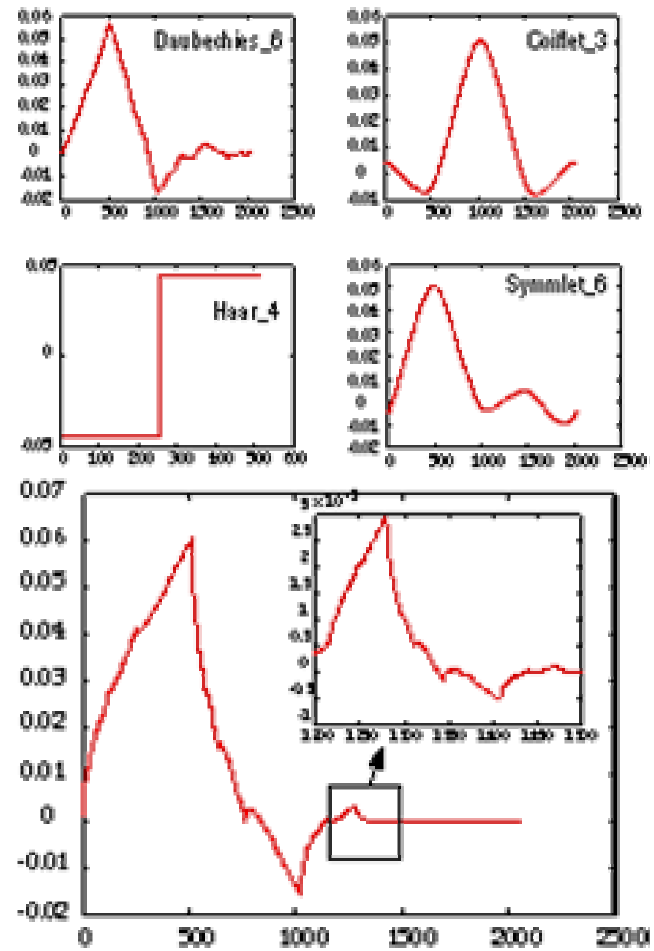
- Construct distributed load-balanced quad-tree hierarchy of *lossy wavelet-compressed summaries* corresponding to different resolutions and spatio-temporal scales.
- Queries *drill-down* from root of hierarchy to *focus search* on small portions of the network.
- *Progressively age* summaries for long-term storage and graceful degradation of query quality over time.



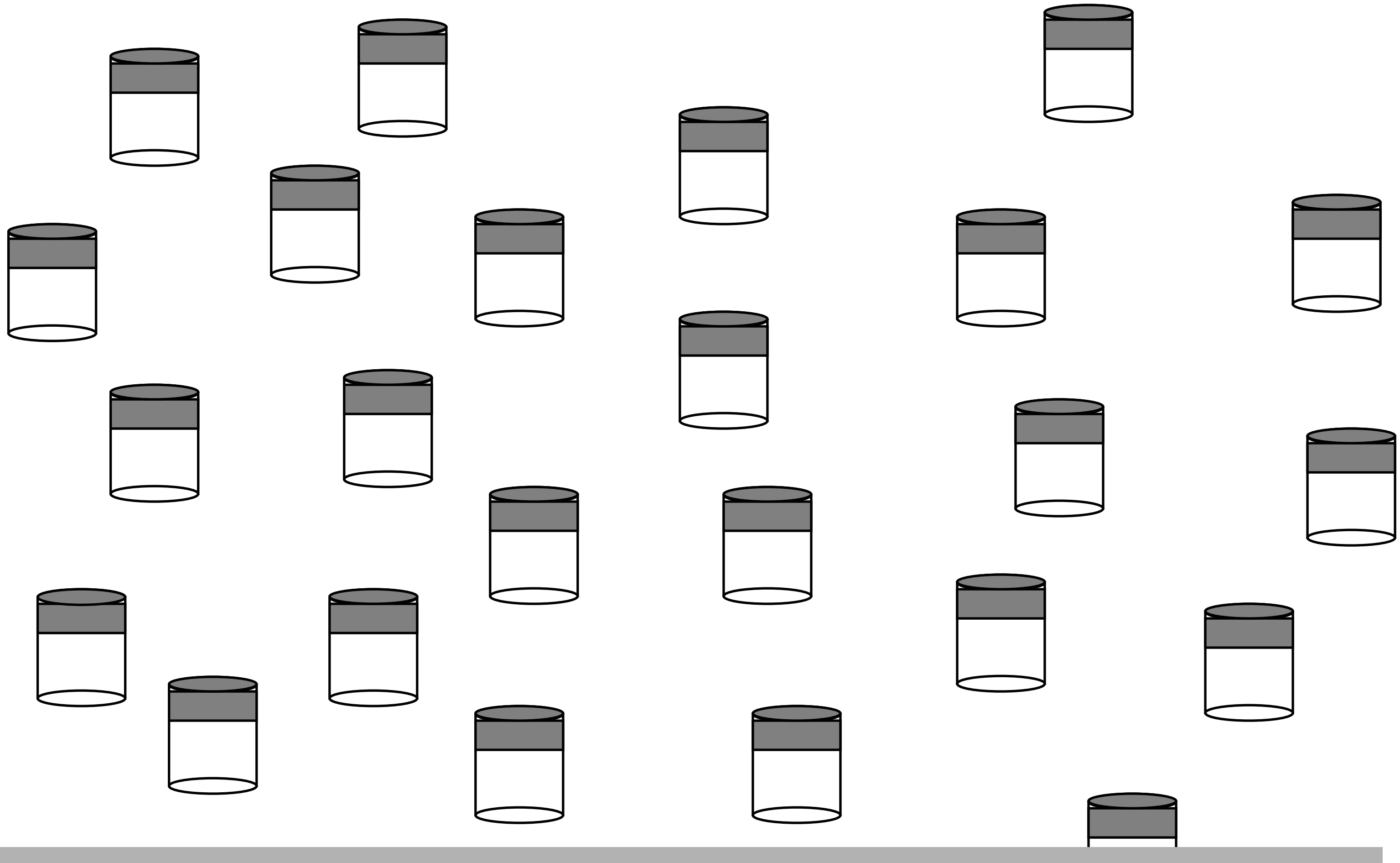
A Word about Wavelets

(from "An Introduction to Wavelets" by Amara Graps)

- Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale.
- They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes.
- See http://www.amara.com/IEEEwave/IEEEwavelet.html

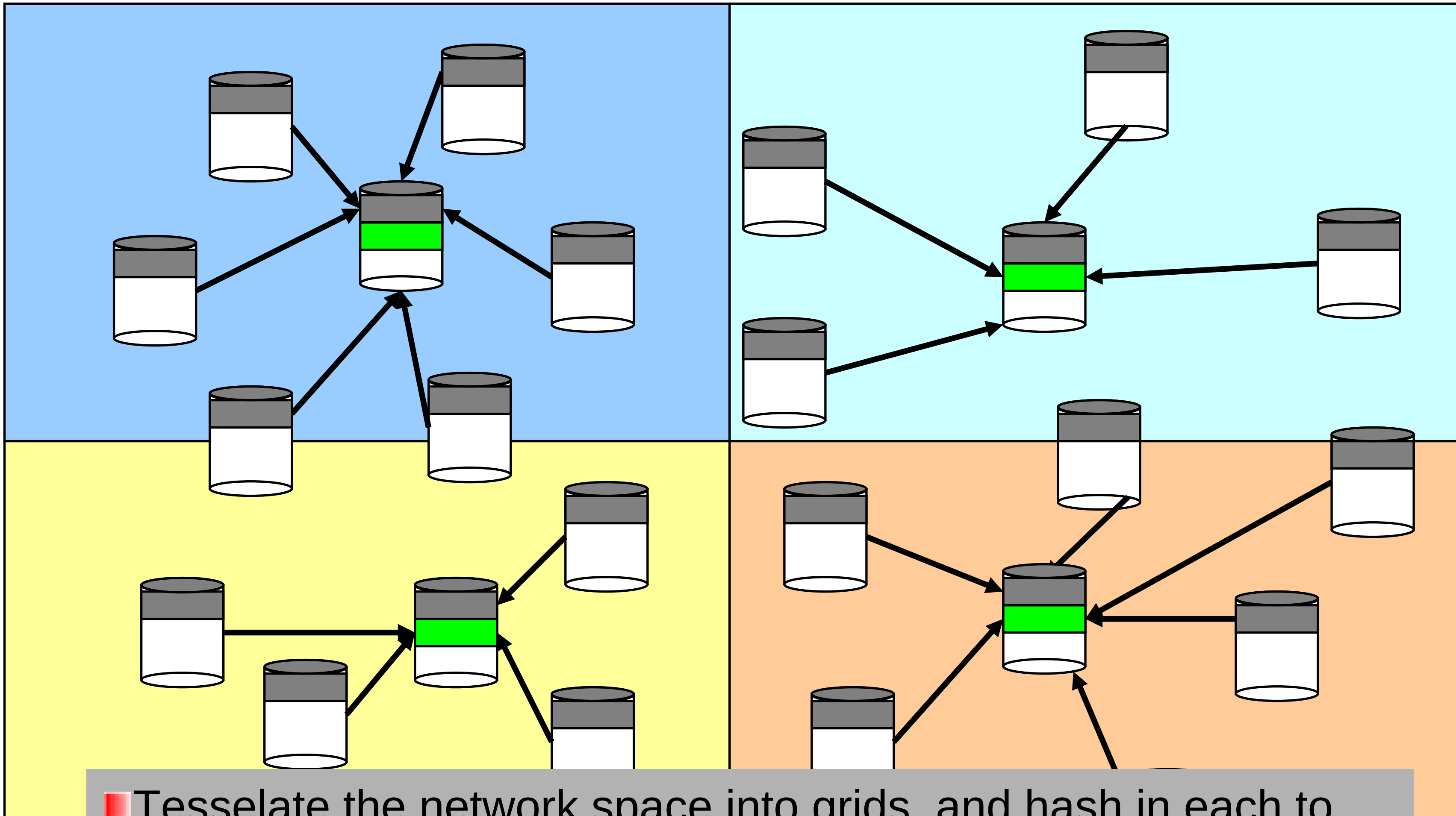


Building the Hierarchy (1)



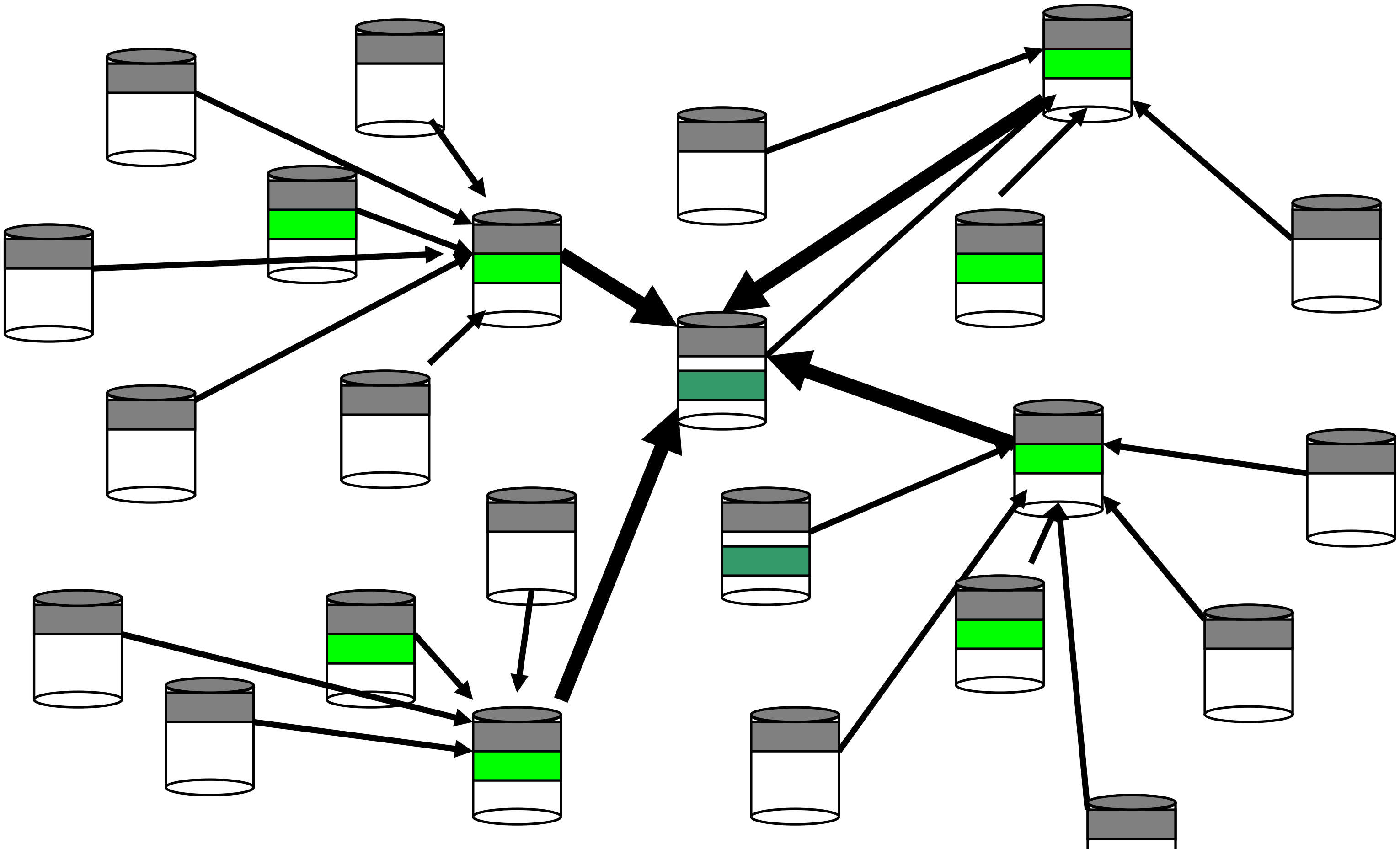
Initially, nodes fill up their own storage with raw sampled data.

Building the Hierarchy (2)



- Tesselate the network space into grids, and hash in each to determine location of clusterhead (ref: DCS).
- Send wavelet-compressed local time-series to clusterhead.

Building the Hierarchy (3)



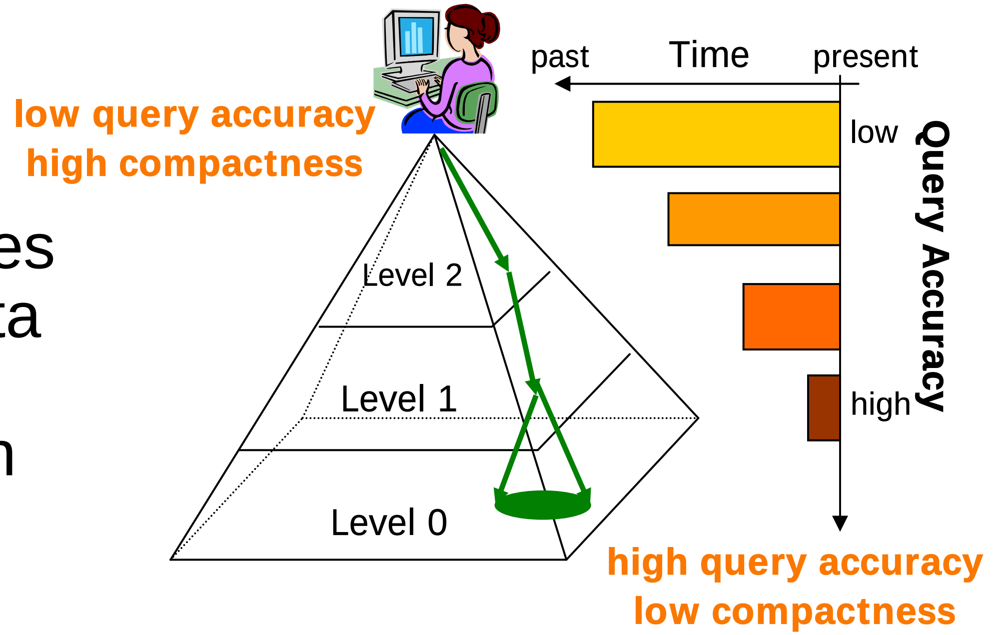
Hash to different locations over time to distribute load among nodes in the network.

In Other Words

- A temporal summary is generated in each sensor.
- Construct grid-based overlay and re-summarize data at each level, compressing it both over space and time.
- Open questions:
  - Are there better hierarchies than the quad-tree?
  - How about only storing the difference to the summary on the next level locally?

Aging the Data (1)

- *Graceful Query Degradation:* Provide more accurate responses to queries on recent data and less accurate responses to queries on older data.

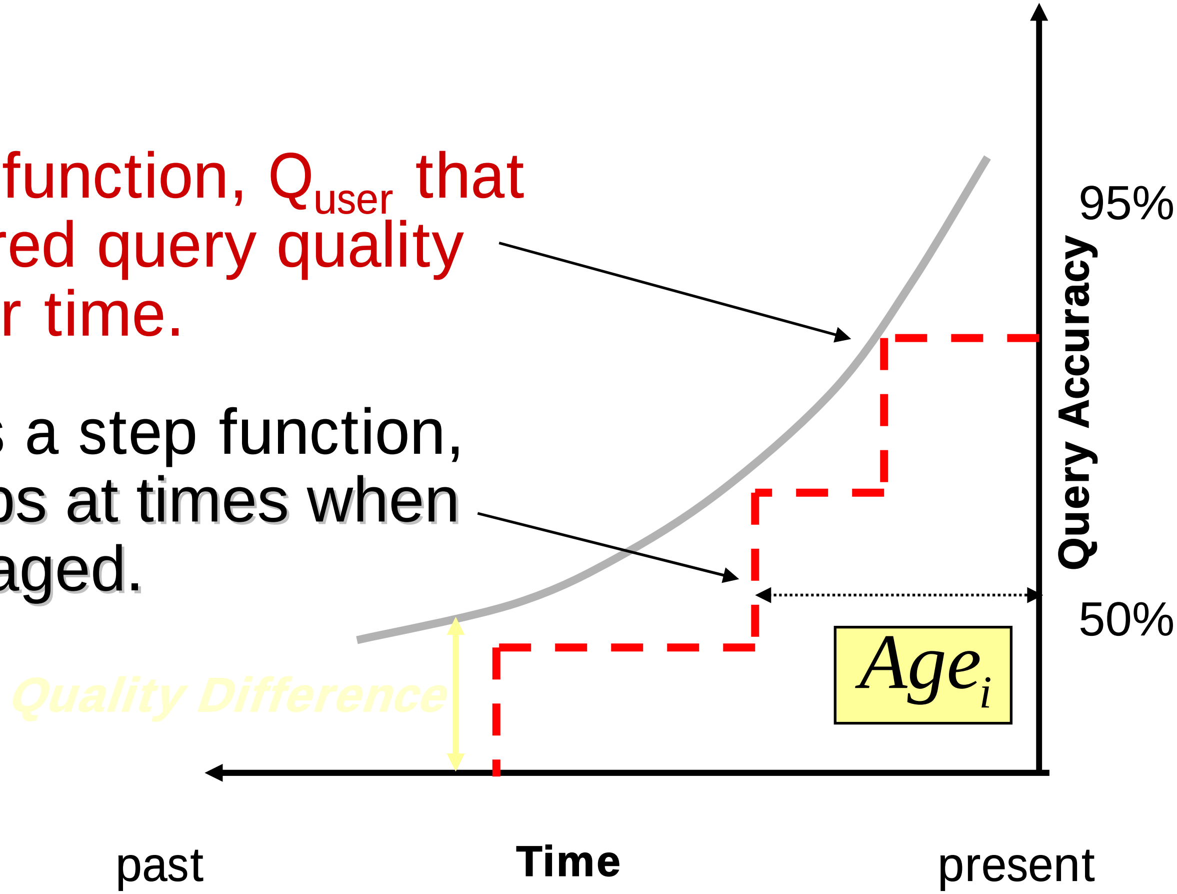


How do we allocate storage at each node to summaries at different resolutions to provide gracefully degrading storage and search capability?

Aging the Data (2)

User provides a function, Q_{user} that represents desired query quality degradation over time.

System provides a step function, Q_{system}, with steps at times when summaries are aged.



Objective: *Minimize worst case difference between user-desired query quality (blue curve) and query quality that the system can provide (red step function).*

Aging the Data (3)

full a priori information

Omniscient Strategy

Baseline. Use all data to decide optimal allocation.

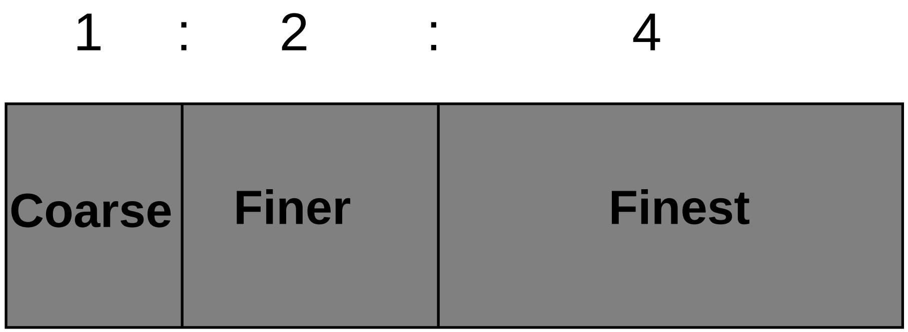
Training Strategy

(can be used when small training dataset from initial deployment).

Greedy Strategy

(when no data is available, use a simple weighted allocation to summaries).

No a priori information



Aging the Data (4)

- Objective: Find $\{s_i\}$, i=1..log4N that:

$$\min_{t=-\infty..0} \max \; Q_{user}(t) - Q_{system}(t)$$

- Given constraints:
  - Storage constraint: Each node cannot store any greater than its storage limit.

$$\sum_{i=1}^{\log_4 N} s_i \le S$$

  - Drill-down constraint: It is not useful to store finer resolution data if coarser resolutions of the same data is not present.

$$Age_{i+1} \ge Age_i$$

In Other Words

- An user-defined aging function is approximated given storage constraints of the network.
- Data may be aged according to different strategies depending on pre-known parameters.
- Open Questions:
  - Is the exponential compression at the root good enough for applications?

Assumptions

- Sensors nodes are arranged in a grid or otherwise uniformly deployed in the physical world for load balancing.
- The network is homogeneous, i.e. sensor nodes have similar capabilities.
- Data needs to be synchronized in time in order to build summaries.
- Summaries at the same level are of equal size, i.e. data is gathered at the same rate in the entire network.

Conclusion

- Experimental evaluation shows that
  - Overhead of communicating summaries can be amortized over many queries.
  - Aging after prior training performs only 1% worse than optimal solution. Greedy aging with “nice” parameters performs 5% worse than optimal solution.
  - A load-balanced hierarchy reduces storage used per node by a factor of three, while having similar communication requirements as a fixed hierarchy.

Future Work

- Some of the assumptions are too strong for real-world applications.
  - Placing nodes in a structured way (e.g. grid) may not be feasible.
  - Different nodes may produce a significantly different amount of data.
- Hence …
  - … wavelet processing needs to be adapted to cope with irregularities.
  - … the hierarchy needs to adapt the size of the summaries to the regional requirements.

Follow-Up Work

- **Deepak Ganesan**, Sylvia Ratnasamy, Hanbiao Wang and **Deborah Estrin**

  “Coping with irregular spatio-temporal sampling in sensor networks”

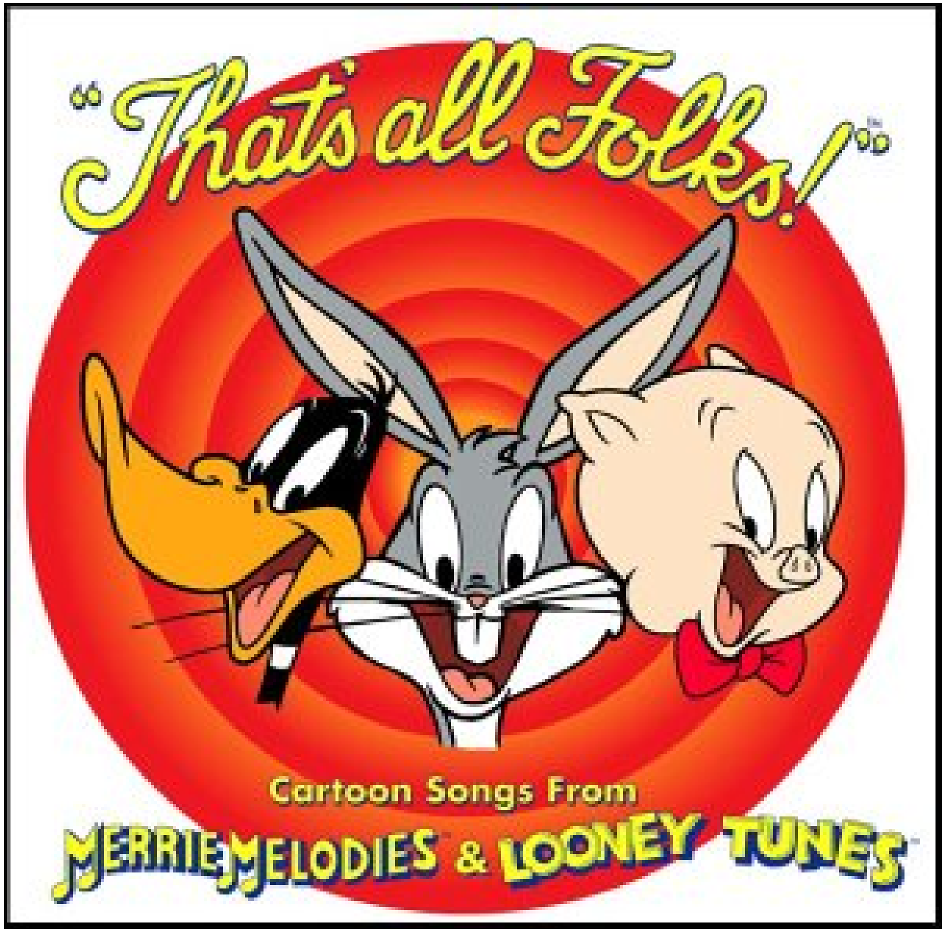
- **Ben Greenstein**, E. Kohler, D. Culler and **Deborah Estrin**

  “Distributed Techniques for Area Computation in Sensor Networks”

Evaluation *("My $0.02")*

- Major contributions are the adaptation of several techniques to the area of sensor networks, especially the aging strategy.
- Some rough edges (assumptions) have been addressed in follow-up papers.
- Further tests are needed as the data set in their experimental evaluation was rather small.

The End


"That's all Folks!"
Cartoon Songs From
MERRIE MELODIES & LOONEY TUNES

Discussion