

Location-Aware In-Network Data Aggregation for Sensor Networks



Alexandros Labrinidis and Panos Chrysanthis

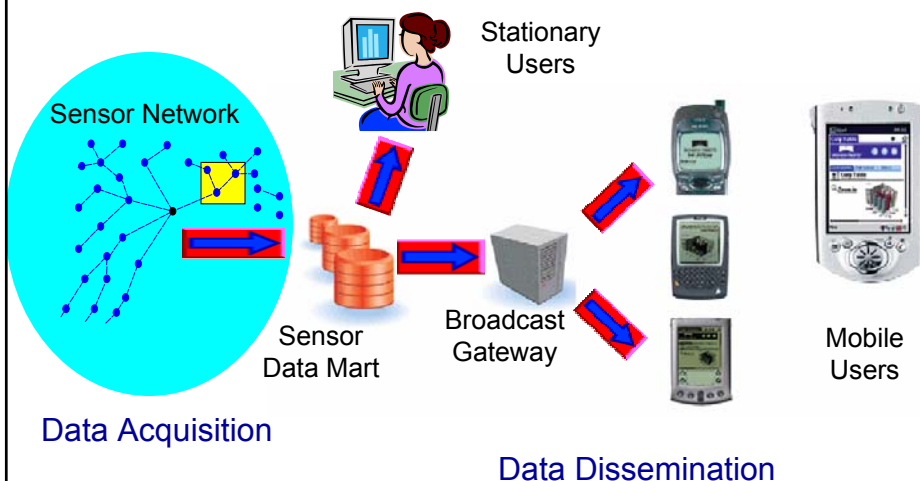
Advanced Data Management Technologies

Department of Computer Science

University of Pittsburgh

<http://db.cs.pitt.edu>

Data Aggregation in Pervasive Environments



Sensor Networks

- Applications:
 - Habitat Monitoring (Storm Petrel, Redwoods, Zebras)
 - Monitoring Buildings, Bridges, etc
 - Environmental Monitoring (toxic substances, weather)
 - Precision Agriculture
 - ...
- Limitations of Sensor Nodes
 - Small Memory Storage
 - Communication (Low bandwidth, Unreliable transmission)
 - Low Processing Power
 - **Limited Energy**

Reducing Energy Consumption

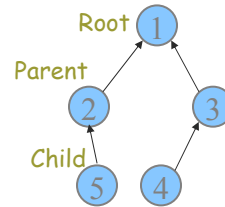


- Why?
 - Sensors often in remote areas – cannot recharge
- Where to start?
 - Be in *sleep mode* for as much as possible
 - **Reduce communication** (= principal energy consumer)
- How to reduce communication at sensor nodes?
 - Trade communication for computation
 - 1 bit of communication = 200-800 CPU instructions
 - How: **In-network query processing (aggregation)**

Query Processing in Sensor Networks

Two processing approaches:

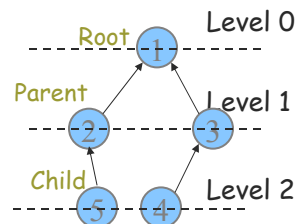
- Centralized at the **Base Station (BS)**:
 - Sensor nodes do sampling and routing
 - BS gathers sensor readings
 - Query evaluation at BS
- Decentralized: **In-network query processing**
[directed diffusion, MobiCom'00; TAG, OSDI'02; Cougar, CDIR'02]
 - Intermediate nodes are not simply routers
 - Query is evaluated incrementally at intermediate/parent sensor nodes
 - TAG experiments:
 - one order of magnitude less communication



In-Network Synchronization

- Sending/receiving must be synchronized
 - A parent node has to hear from all its children before doing the aggregation
- Synchronization in TAG
 - EPOCH specifies the arrival rate of new results
 - Divide EPOCH into shorter communication slots
 - During one slot, nodes at:
 - One level of the tree are sending
 - Parent level are listening
 - Other levels are sleeping

```
SELECT floor, Avg(temp)
FROM Sensors
GROUP BY floor
EPOCH 30Sec
```



Roadmap

- Introduction
- **TiNA Framework**
- Multi-Resolution, Location-Aware Data Aggregation
- Location-Aware Network Configuration
- Conclusions / Challenges

TiNA Framework

- Motivation:
 - Everything's OK alarm (from *The Simpsons* ☺)
- Basic idea:
 - Exploit temporal correlation between readings
 - Cut down on number and/or size of messages
 - Suppress readings that add no significant new information to the query results
 - **Thus: reduce energy consumption**
- Allow user to specify what is significant
- Work with any synchronization scheme that supports in-network aggregation

Continuous Queries in TiNA

- Sample query using TiNA:

Select **room, Avg(temperature)**

From **sensors**

Where **temperature < 65**

Group By **room**

Epoch **30Sec**

Suppress Values Within temporal-coherency-tolerance (*tct*)

- WHERE clause acts as an input result filter
- TiNA adds the **Suppress Values Within** Clause for specifying amount of temporal coherency tolerance (*tct*)
- **Suppress Within** is an output transmission filter
 - **tct** = 0 → report any changes
 - else → report if change is over +- *tct*%

How does TiNA work?

- Leaf Node:

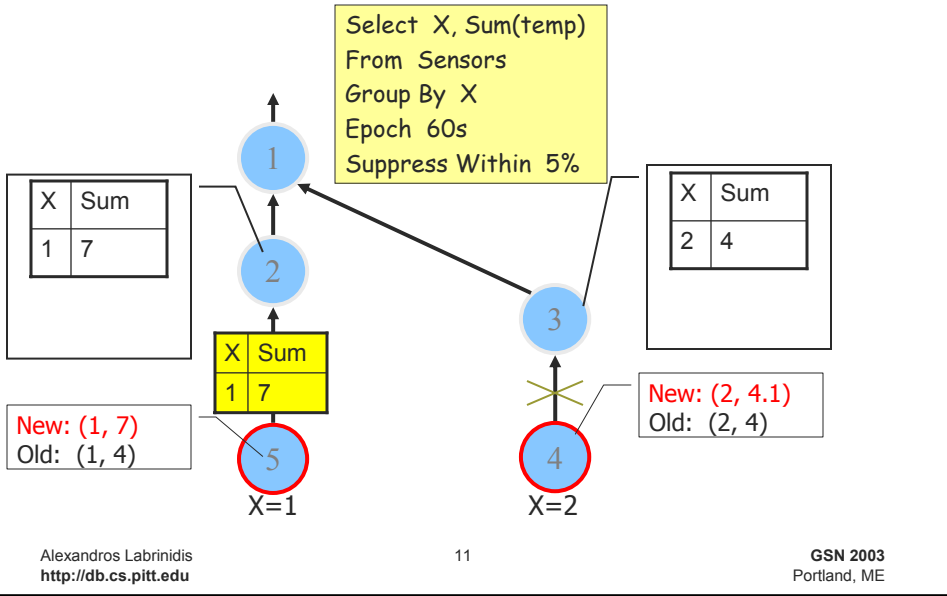
- Reports reading S to its parent iff $(S - S_{old})/S > tct$
- Sends invalidation bit if S_{old} is no longer valid
- **Heartbeat** is sent periodically for fault-detection

- Internal Node:

- Receives a Partial Result (PR) from its child
- If a child PR^i is missing or incomplete, it substitutes from PR_{old}^i unless it was invalidated
- Aggregate PRs from all children into one PR_p
- It will include its own reading if:
 - It violates *tct*
 - It can be aggregated within a group already in PR_p

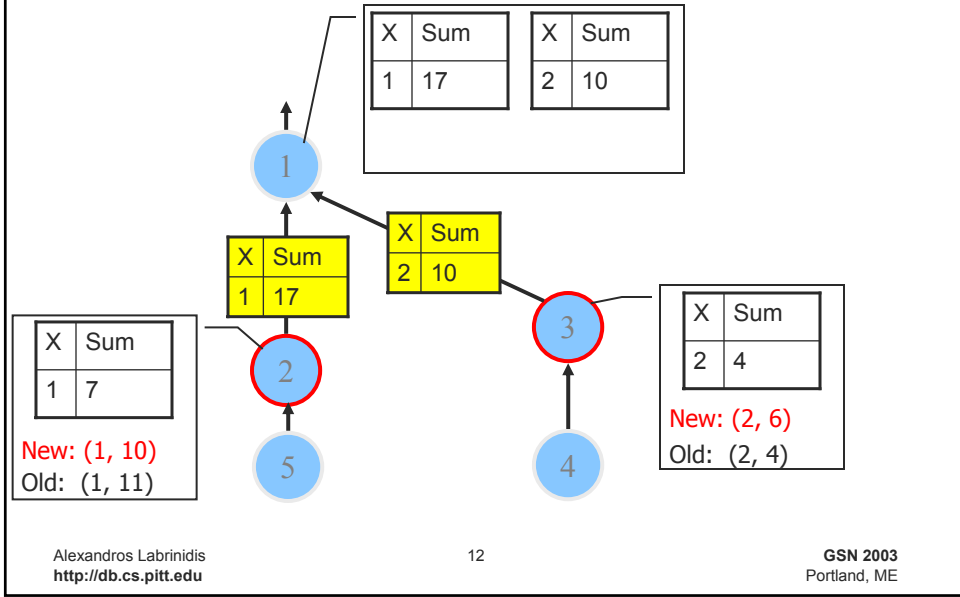
TiNA in Action

1



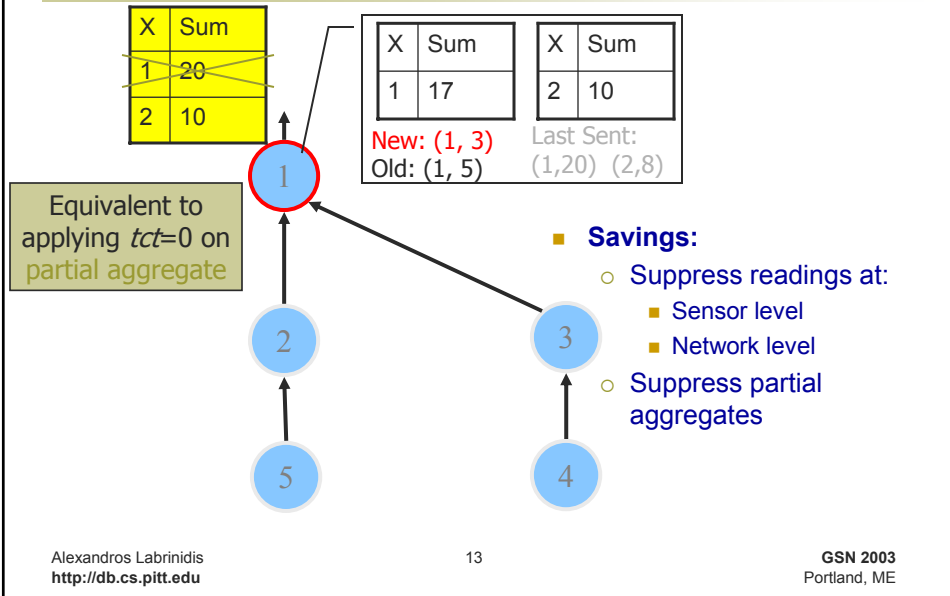
TiNA in Action

2



TiNA in Action

		3
--	--	---



Experimental Setup

- Created a simulation environment using CSIM
- Simulated contention-based MAC Protocol (PAMAS) that avoids collisions
- Base case is the TAG protocol with the optimization of a child cache
- Measured
 - Energy Consumption and
 - Quality of Data (Compared to complete answers)

Simulation Parameters

Parameter	Value	Default
Grid Size	10x10 - 45x45	15x15
Aggregate	MAX, SUM, AVG	SUM
Group By Attributes	0, 1, 2	1
<i>tct</i> (tolerance degree)	0% - 30%	10%
Randomness Degree	0.0 - 1.0	0.5
Random Step Size	0% - 25%	10%
Sensor Values	1 - 100	
Epoch Durations	280 - 1700 mSec	1500
Number of Epochs	100	100

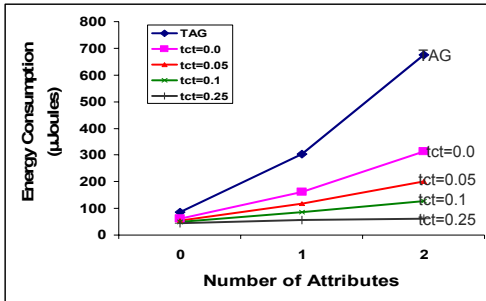
Random Walk Model

- Readings generated from a domain of values [1-100]
 - To simulate temperature readings

- A reading changes from one round to the next with probability *Randomness Degree (RD)*
 - RD = 0 means values never change
 - RD= 1 means values change every round

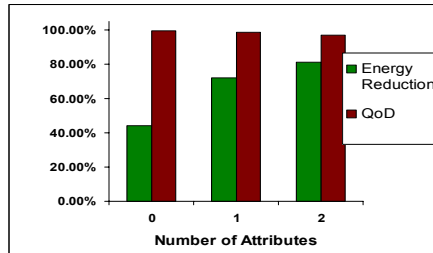
- The magnitude of change is up to a *Random Step Size Limit* (= % of domain range)

Number of Attributes



Energy and QoD versus TAG

for tct=0.10



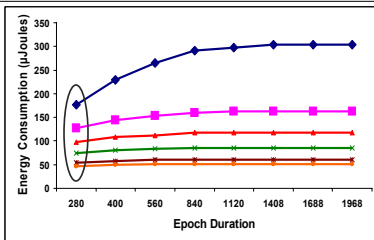
Energy Consumption

- Number of attributes in the Group By clause
 - 0 means there is one group for the entire network
 - 2 means each sensor is its own group

Energy Reduction:

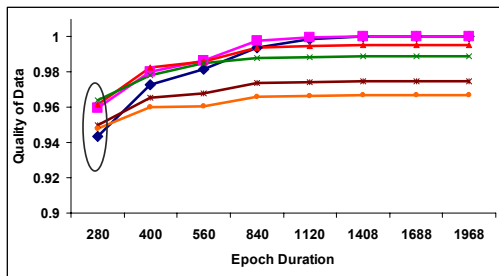
- 40% for 0 attributes
- 80% for 2 attributes

Effect of Congestion



Energy Consumption

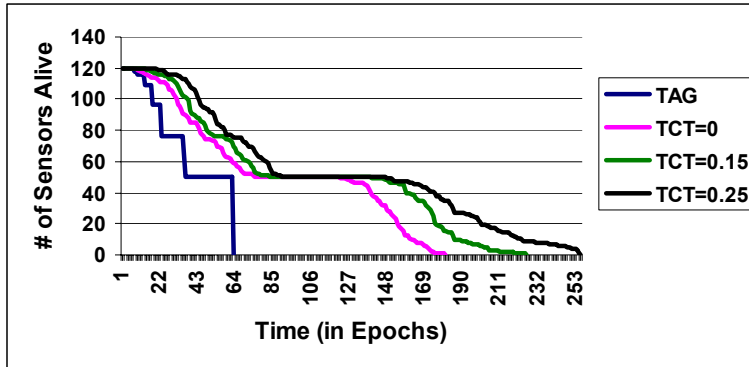
Quality of Data



In congested networks, TiNA:

- Reduces energy
- Improves QoD

Lifetime Savings



TiNA is able to extend lifetime of sensor network by 187%

More experiments

- In our MobiDE 2003 paper:

TiNA: A Scheme for Temporal Coherency-Aware in-Network Aggregation

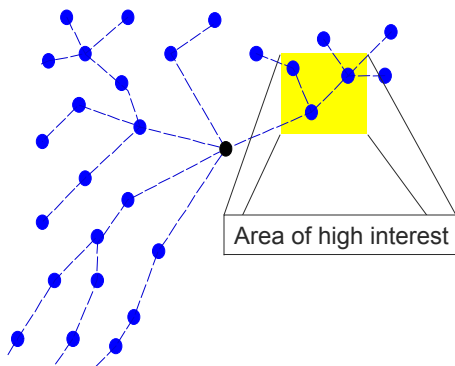
Mohamed A. Sharaf
Alexandros Labrinidis

Jonathan Beaver
Panos K. Chrysanthis

Roadmap

- Introduction
- TiNA Framework
- **Multi-Resolution, Location-Aware Data Aggregation**
- Location-Aware Network Configuration
- Conclusions / Challenges

Multi-Resolution Scenario



- Location identifies area of high interest
 - Corresponds to importance or emergency
- We want to
 - increase speed of data acquisition on area of interest
 - Without changing remaining areas

Multi-Resolution Query

- We extend the sensor network query interface to allow multi-resolution, as follows:

```
Select  room, Avg(temperature)
From    sensors
Where   temperature > 75
Group By room
Having  room = reactor core Epoch 2 sec
        room <> reactor core Epoch 60 sec
```

Multi-Resolution support from TiNA

- Parts of sensor network → different EPOCH sizes
 - Multiples of smallest EPOCH size
- How to implement multi-resolution?
 - Must keep track of streams with different “speeds”
 - Should not spend additional energy
- Solution: **TiNA Framework**
 - built to allow non-transmission of values
 - If child does not transmit, parent assumes the value is same with previous
 - would also reduce communication further (if the values from the high-speed stream are not changing often)

General Framework for Quality of Data

- Current measures of Quality of Data (QoD)
 - Accuracy of results (RMS, confidence intervals, etc)

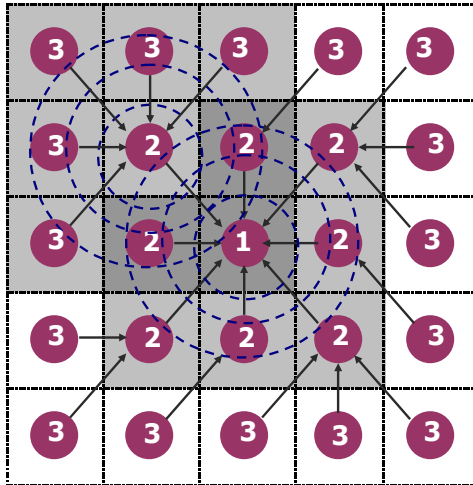
- Must also consider user preferences:
 - Characterize importance based on:
 - Value (sensed or other attribute)
 - Location
 - Latency
 - Loss ratio

- This is open problem

Roadmap

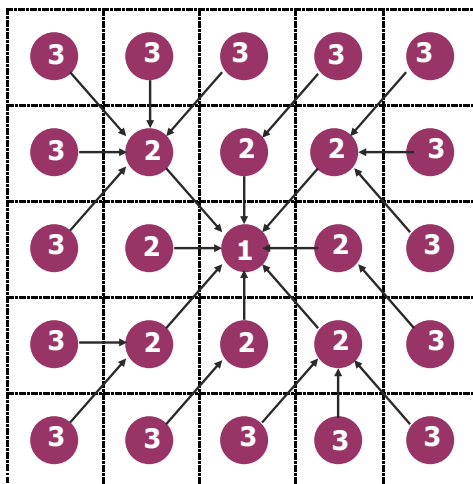
- Introduction
- TiNA Framework
- Multi-Resolution, Location-Aware Data Aggregation
- **Location-Aware Network Configuration**
- Conclusions / Challenges

Traditional Routing Tree Construction



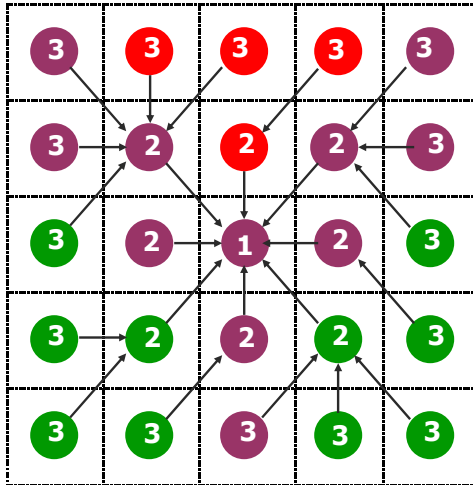
- Node broadcasts message with level#
- Neighbors listen and select parents
 - Based on distance
- Recursively reach entire network

Traditional Routing Tree Construction II



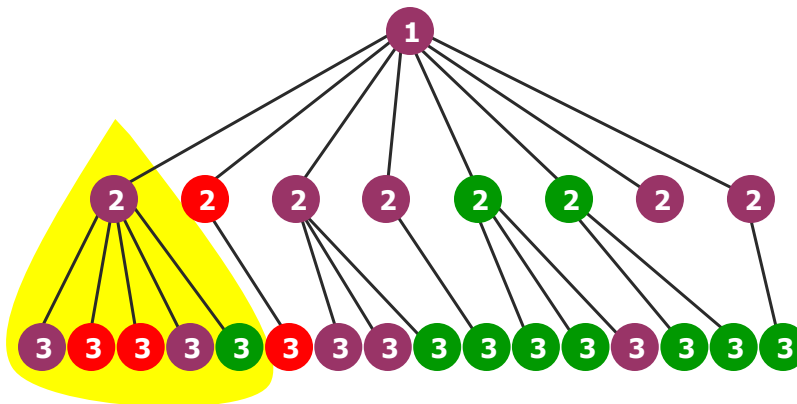
- What happens if nodes belongs to different groups?
- Group ID can be associated with:
 - location
 - value
 - other attributes

Traditional Routing Tree Construction II



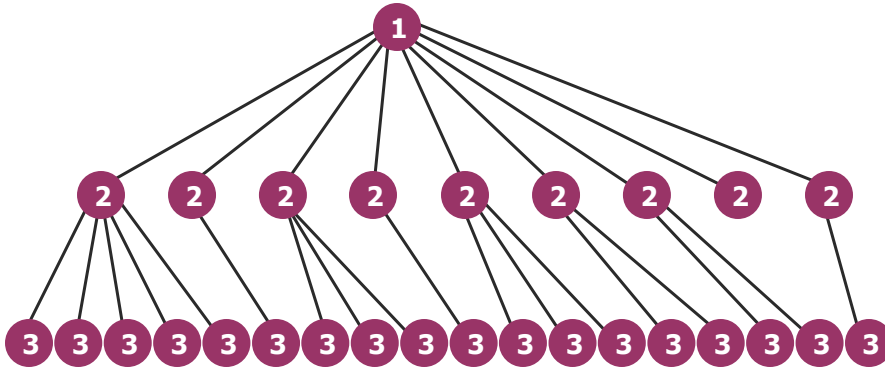
- What happens if nodes belong to different groups?
- Group ID can be associated with:
 - location
 - value
 - other attributes
 - set by query
- E.g., three colors

Routing Tree – with groups



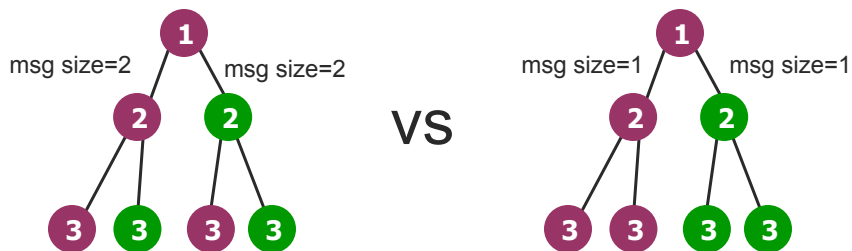
- miss-match of groups means larger messages!

Routing Tree – **without** groups



Location-Aware Network Configuration

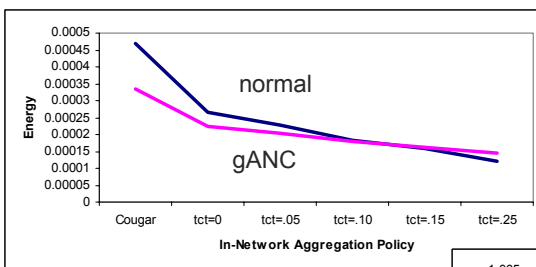
- Basic Idea:
 - Use location/group information **in addition** to distance when constructing routing tree for sensor network



gANC Algorithm

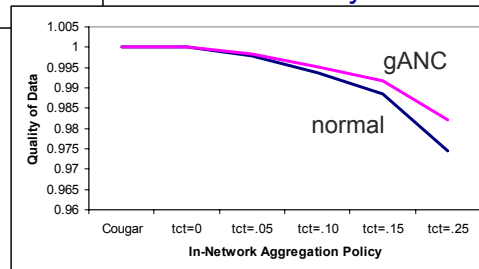
- **Group-Aware Network Configuration Algorithm:**
 - Same as traditional network configuration, except:
 - In addition to their ID and level, nodes also transmit their group ID
 - Children select the candidate parent node with shortest distance (=higher signal strength)
 - If candidate node in the same group then finished
 - If not, then pick candidate (temporarily), until node hears from another candidate parent node that is in the same group.
 - Works with both
 - **TAG** (no need to notify parents) and
 - **Cougar** (children must also notify parents of changes)

gANC Experiments: 15x15 grid (small)



- **Sensor network:**
 - 15 x 15 grid
 - Group# is function of sensor ID
 - 5 groups

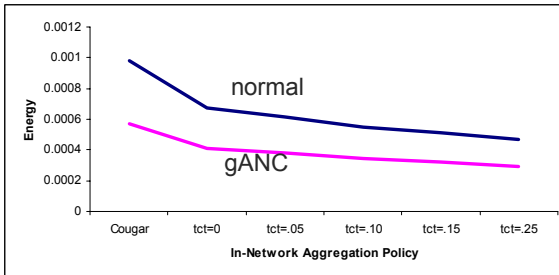
Quality of Data



Energy

- Cougar improves
- TiNA has crossover
- Why did OoD improve?
 - “for free” updates

gANC Experiments: 45x45 grid (normal)

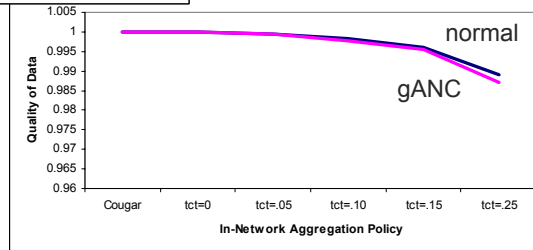


- Sensor network:
 - 45 x 45 grid
 - Group# is random
 - 5 groups

Quality of Data

Energy

- Cougar improves
- no crossover for TiNA
- Larger grid → more options for parents

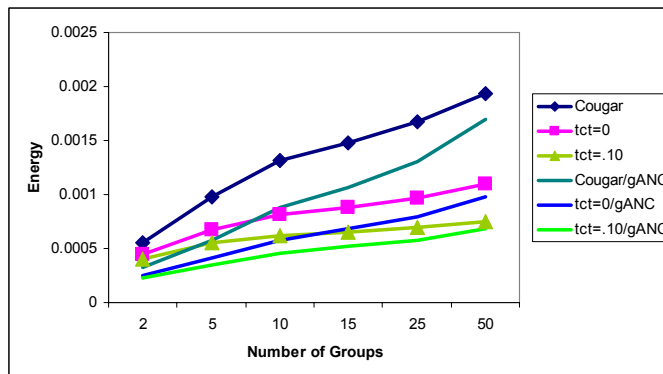


Alexandros Labrinidis
<http://db.cs.pitt.edu>

35

GSN 2003
 Portland, ME

gANC Experiments: 45x45 grid



- Amount of improvement decreases as number of groups increases
 - e.g., Cougar (41% for 2 groups, 12% for 50 groups)

Alexandros Labrinidis
<http://db.cs.pitt.edu>

36

GSN 2003
 Portland, ME

gANC++

- Right now only gANC only “breaks ties”:
 - If parent nodes equidistant, pick one with same group

- Extension:
 - Introduce knob: **distance_factor**
 - Treat as the same distances within a certain range of the minimum distance
 - Select parent based on group ID from those within such range
 - Trade-off:
 - Extra communication energy (to send the signal further)
 - Savings in number and/or size of messages

Conclusions

- TiNA Framework minimizes energy consumption by exploiting end-user tolerance to temporal coherency
 - It decreases the size and number of messages transmitted in a sensor network
 - TiNA works with existing synchronization schemes

- TiNA can facilitate Multi-Resolution, Location-Aware Data Aggregation

- gANC is location-aware network configuration algorithm
 - Energy savings for traditional Cougar/TAG and for TiNA.

Future Work / Challenges

- TiNA = light-weight trigger system
 - How about “sophisticated” triggers at sensors?
- Sensor Mobility
 - Previous work on Pervasive Catalogs [MDDS'02]
- Multifaceted Quality of Data:
 - Location, Values, User Tolerance (to accuracy, latency)
 - Data Prioritization / Load Shedding
- Fault Tolerance
 - Introduce redundant paths

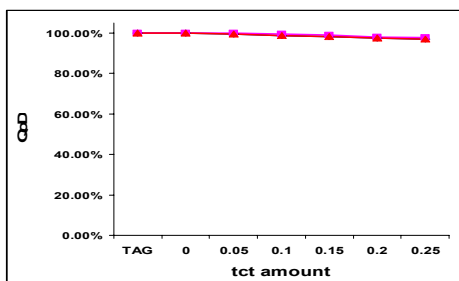
CyberTrust

Acknowledgments

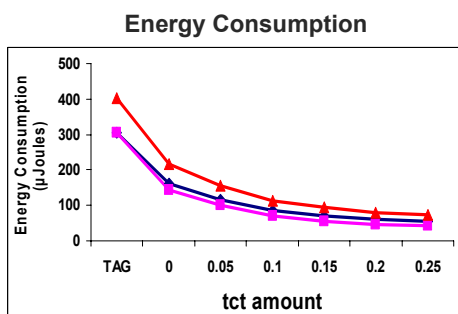
- Advanced Data Management Technologies Group
 - Panos Chrysanthis and Alexandros Labrinidis
 - <http://db.cs.pitt.edu>
- Special thanks to our students:
 - Jonathan Beaver and Mohamed Sharaf
- Special thanks to our sponsor: **NSF**
 - ITR: *Secure CITI: A Secure Critical Information Technology Infrastructure for Disaster Management*
 - <http://www.cs.pitt.edu/s-citi>

Backup

Aggregation Function

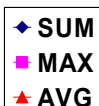


Quality of Data

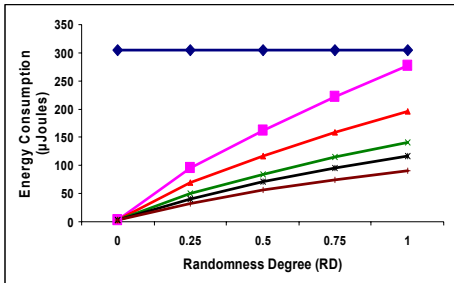


At $tct = 0$, energy reductions are:

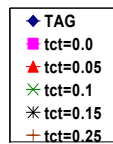
- 46% for AVG & SUM
- 52% for MAX



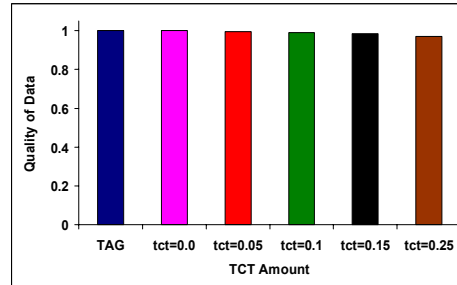
Data Change Rate



Energy Consumption

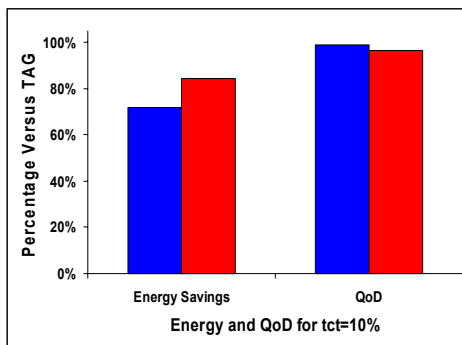


QoD readings for randomness degree = 0.5



- **tct = 0.0** → 45% energy reduction, 100% QoD
- **tct = 0.25** → 80% energy reduction, 97% QoD

Partial Aggregates



tct = 0.1 for readings ■
tct = 0.0 for partial aggregates

tct = 0.1 for readings ■
tct = 0.1 for partial aggregates

tct on readings and partial aggregates ≈ higher tct on readings only

Quality of Data

- Compare with complete answers
- Error E_i for a group G_i defined as
$$E_i = (|m_i - m_i'|) / m_i * 100$$
- Error over entire group-by defined as
$$E_g = 1/n * \sum_{i=1,n} E_i$$
- So, the quality of data over time T for a complete experiment is given as
$$QoD = 1/T * \sum_{t=1,T} (100 - E_{gt})$$