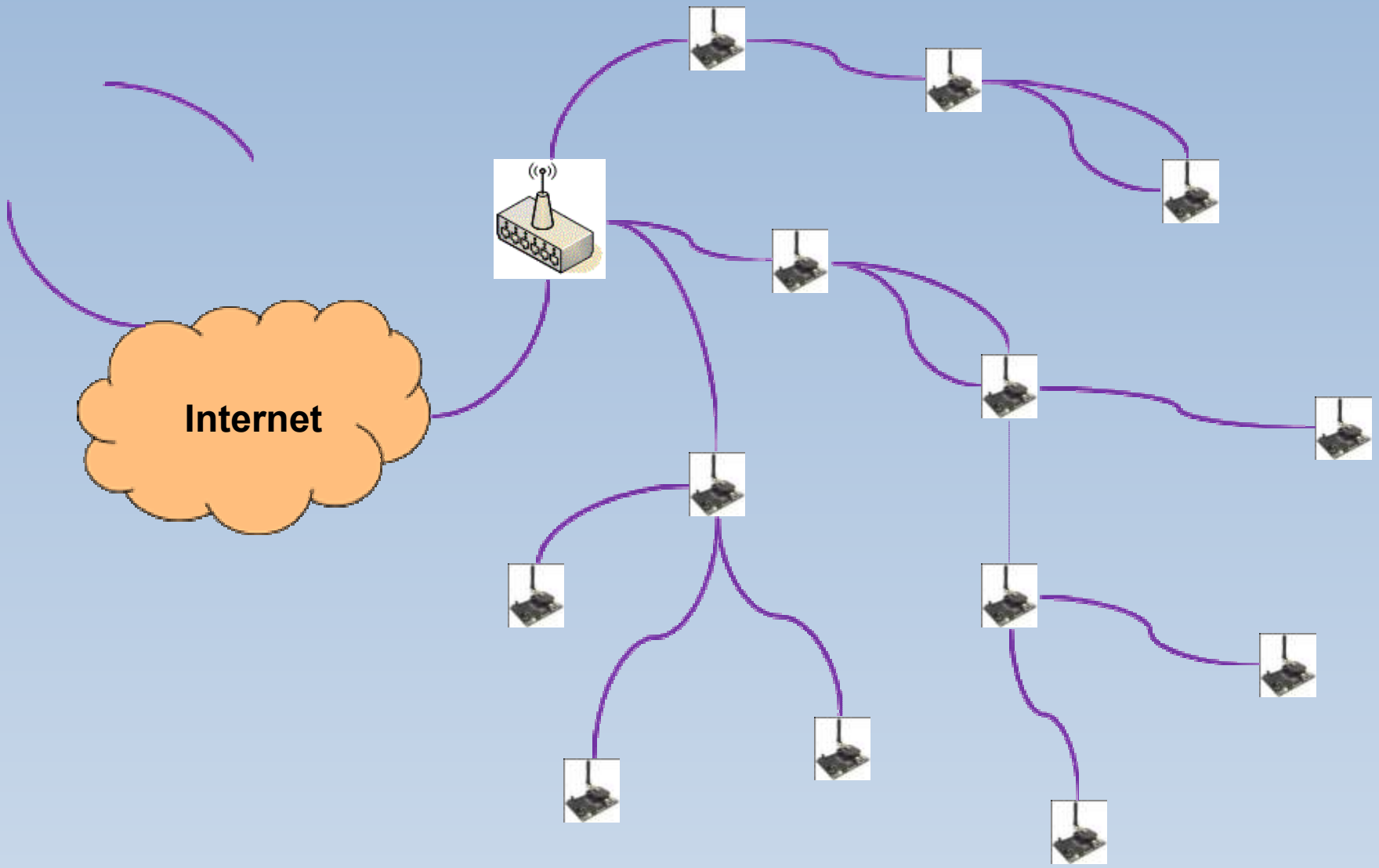
Histogram and Other Aggregate Queries in Wireless Sensor Networks

Khaled Ammar and Mario A. Nascimento
University of Alberta
Canada
Research supported by NSERC Canada
Kammar@cs.ualberta.ca
SSDBM 2011

Outline

- (Very) Quick introduction to WSN
- Histogram in WSN:
  - State of the Art: TAG Algorithm
  - Our Contribution: HIU Algorithm
  - Experimental results
- Using Histogram for other queries:
  - Approximate answers for FREE
  - Inexpensive accurate answers
- Conclusions
- Current Research

Wireless Sensor Networks (WSNs)



Motivation

- WSN is used to observe some phenomena and report data about them.
- It is used in many applications like:
  - Industrial: machine health monitoring, pressure monitoring, structural health monitoring, …
  - Environment: Animals habitat monitoring, Climate change monitoring, …
  - Health monitoring.

WSNs Limitations

- WSNs nodes have:
  - Slow CPU
  - Low storage
  - Access to limited energy supply
- It is crucial to minimize the energy cost of query processing by reducing amount/size of the sent or received messages.

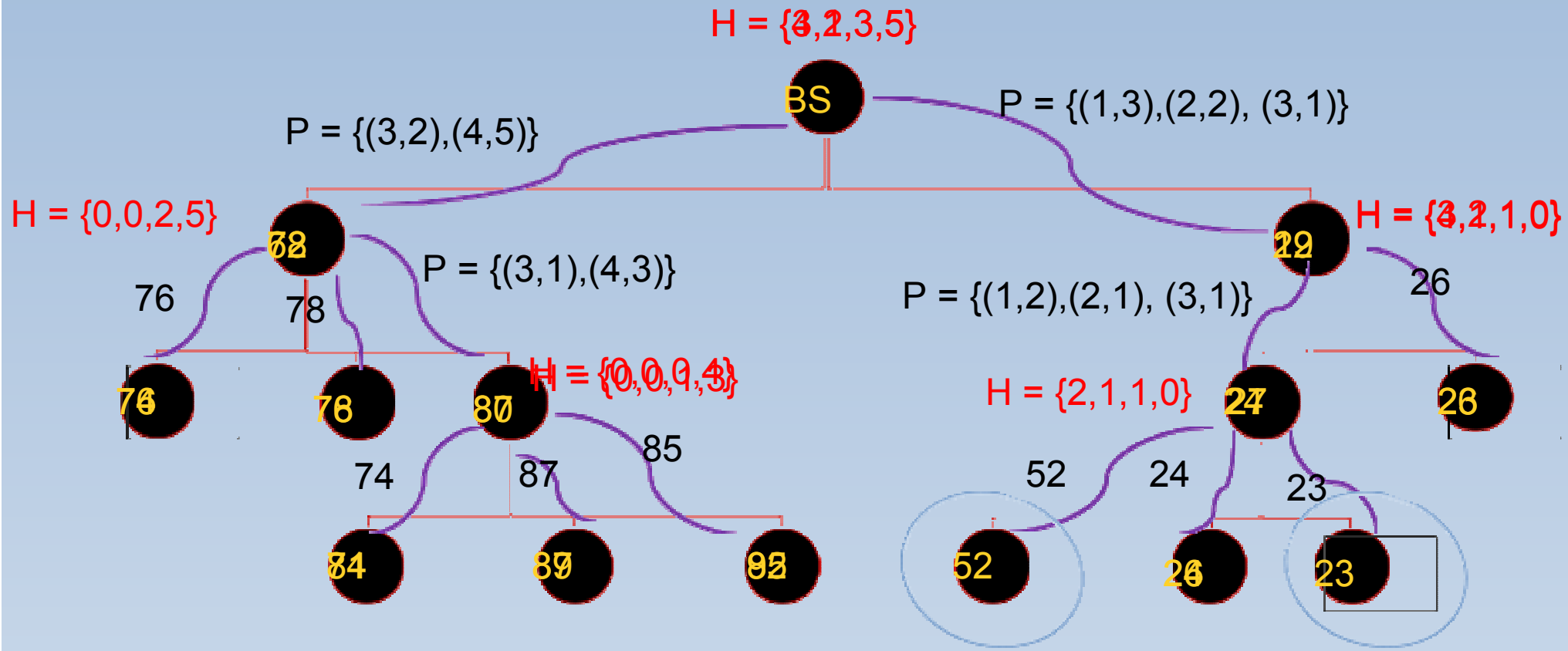
Histogram Definitions

- **Query:** *Q(Lb,Ub, b1, b2, b3,...bB, epoch)*
  - *bi = [Lbi , Ubi [where 1 ≤ i < B and bB = [LbB,UbB]*
  - *Ubi ≤ Lbj i < j and {bi} = [Lb,Ub]*
  - *Lb1 = Lb and UbB = Ub*
- **Answer:** *H = (h1, h2, h3,...hB)*

TAG Algorithm for Histogram Queries

Q(0, 100, [0,25), [25,50), [50,75), [75,100] , T)

Round # i+1

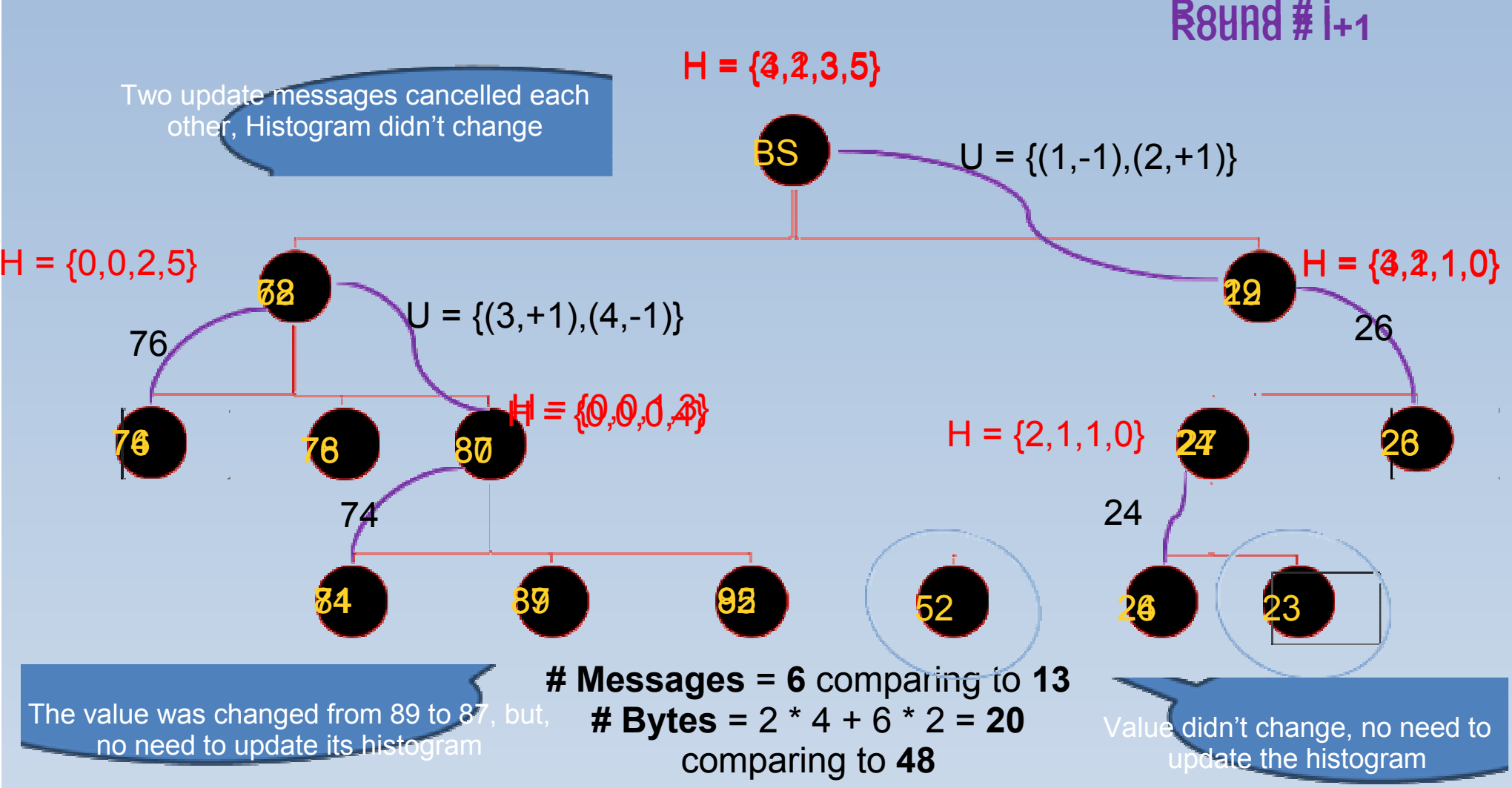


 # Messages = 13**# Bytes** = 2 * 9 + 6 * 2 + 9 * 2 = 48

Histogram Incremental Updates (HIU) Algorithm by Example

In-node Caching

Q(100, [0,25), [25,50), [50,75), [75,100] , T)



It Works

- In this example a “send all” approach (e.g. TAG) would send 48 bytes whereas HIU sends only 20 bytes.
- The main idea behind HIU is

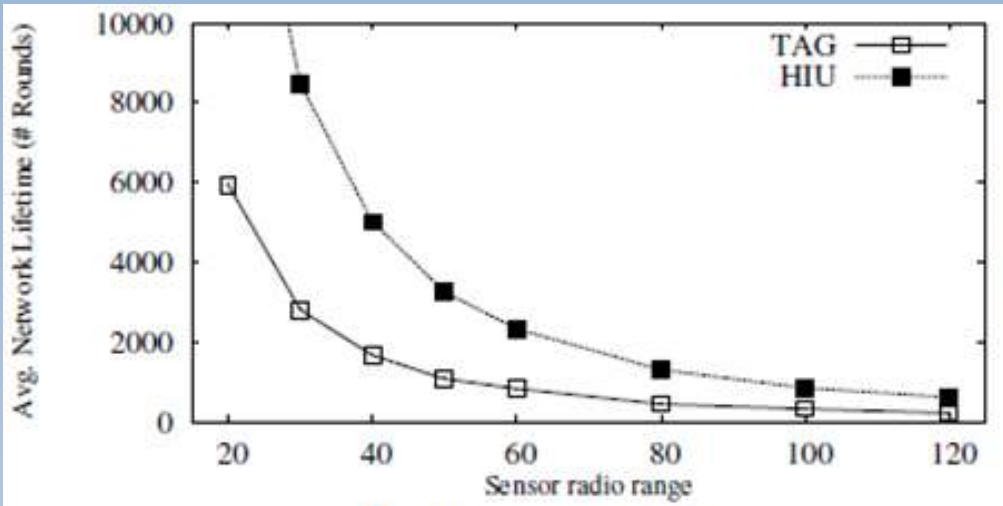
  In-node Caching

Performance Evaluation

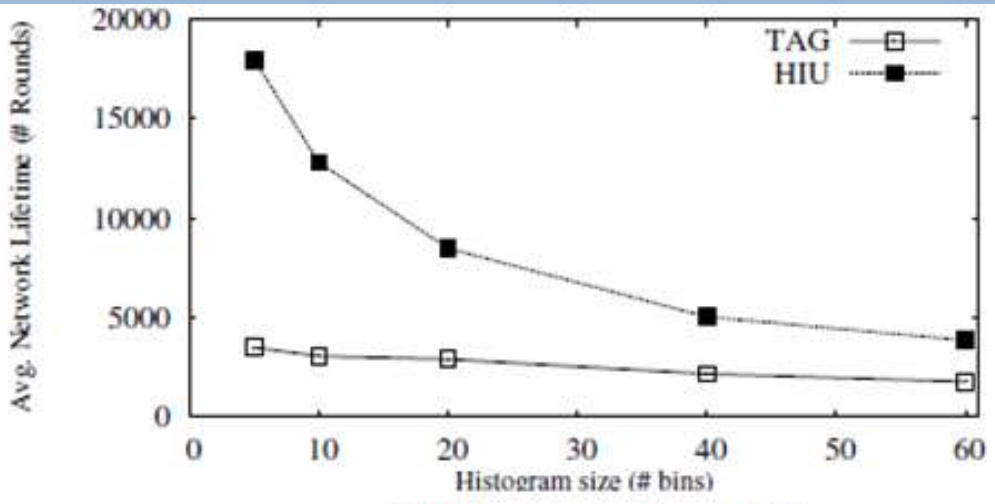
Size of monitored area: 200 m x 200 m
Nodes' locations are static and uniformly distributed
Nodes' values are uniformly distributed [0, 216]
Monitored value consumes 2 bytes
Packet size: 128 bytes
Competitor: TAG (Madden et al, 2002)
Performance Indicator: Network Lifetime

Parameters investigated	Used values
Radio range [m]	20, 30, 40, 50, 60
Histogram Size [# bins]	5, 10, 20, 40, 60
N [# nodes]	1000, 2000, 3000, 4000, 5000
Average Amount of Change	1%, 25%, 50%, 75%, 100%
Probability of Change	1%, 25%, 50%, 75%, 100%

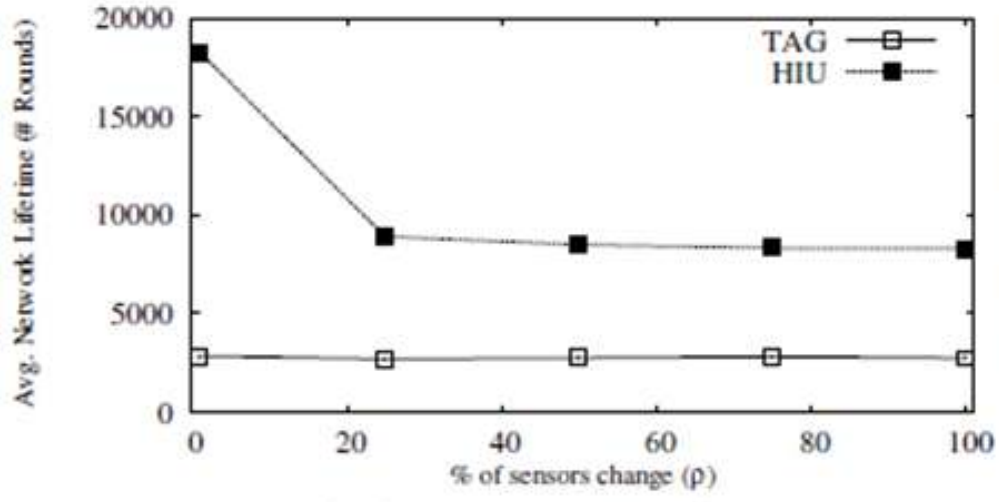
Synthetic Dataset



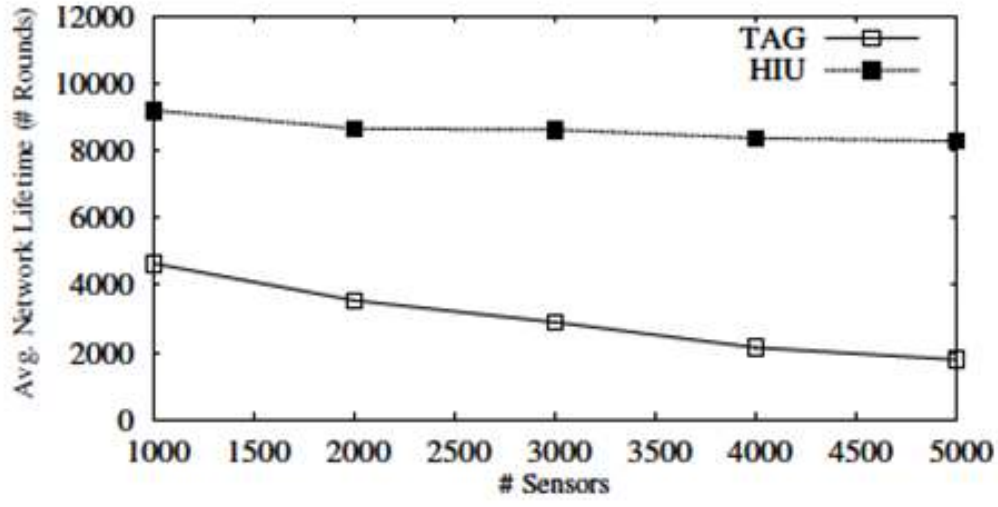
Radio range (R)



Histogram size (B)

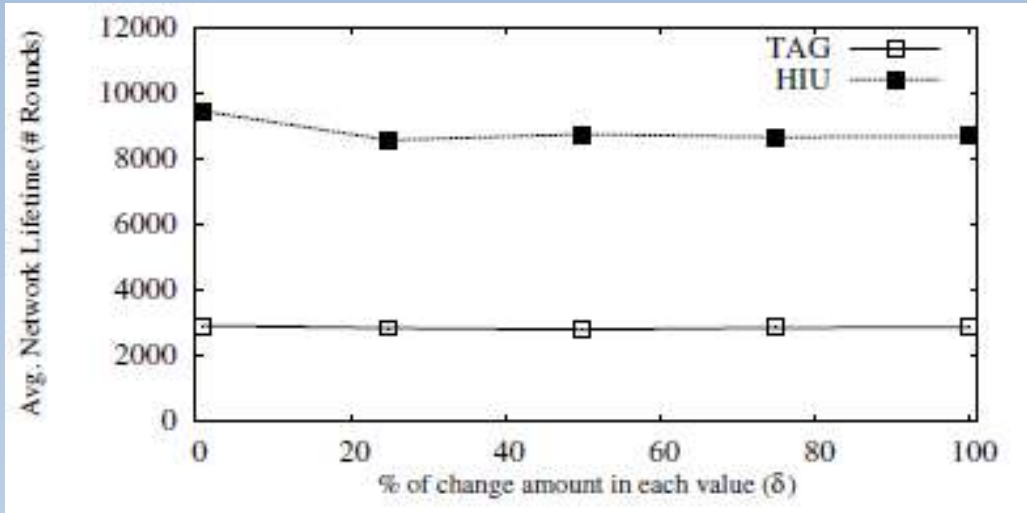


Probability a value changes (ρ)



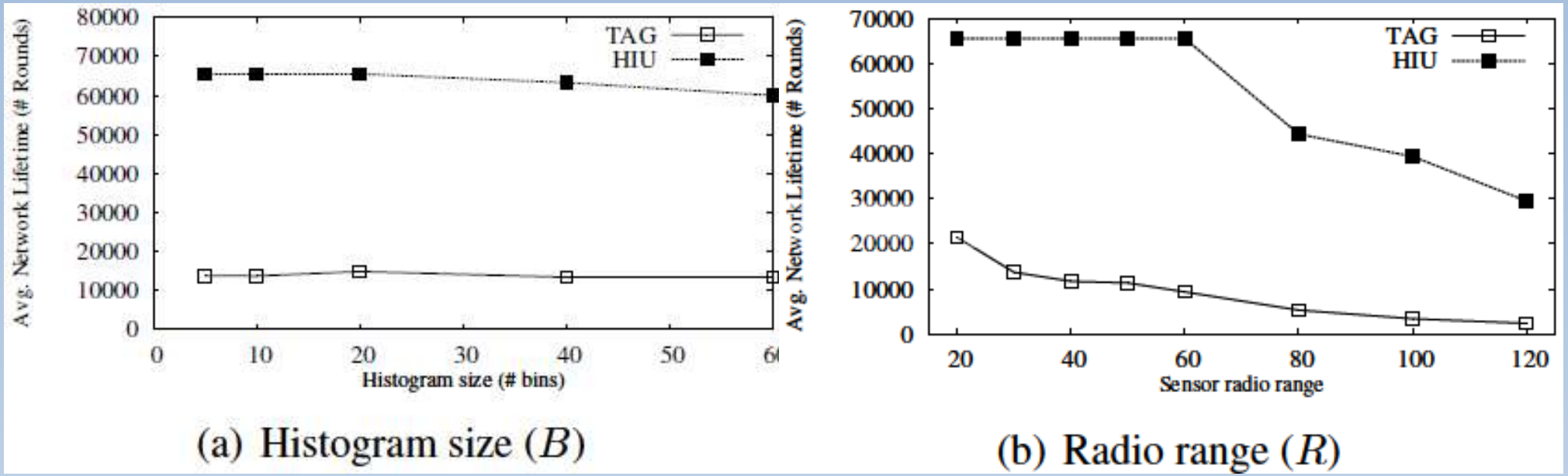
Number of Nodes (N)

Synthetic Dataset



(b) Average amount of change per value (δ)

Real (Intel) dataset



Approximate Answers for FREE

- Histogram is powerful in a sense that gives good approximation for other aggregate queries.
- Q(0, 50, [0,10), [10, 20) , [20,30), [30,40), [40, 50])
- *H = (0, 4, 1, 3 , 2)*
  - Max □ *[40, 50] = 45 ± 5*
  - *Min* □ *[10, 20) = 15 ± 5*
  - *Count = 10*
  - *Sum = 4 * 15 + 1 * 25 + 3 * 35 + 2 * 45 = 280 ± 50*
  - *Average = 280 / 10 = 28 ± 5*

Approximate Answers for FREE

- *Q(Lb,Ub, b1, b2, b3,...bB, epoch)*
- *H = (h1, h2, h3,...hB)*
- *Error Margin depends on the bin size*

Query	Approximate Answer	Error Margin
Max		
Min		
Count		0
Sum		
Average		

Inexpensive Exact Answers

- Motivation:
  - Communication devices in some WSN mandate the sensor to send messages of fixed size only.
- Methodology:
  - **Min/Max**: Each intermediate node sends the value of the sub tree max/min
  - **Sum**: Each intermediate node sends the sum of all values in its sub tree
  - **Average**: Since exact **Count** is available, **SUM** is sufficient to compute the **Average**.

What can be fine-tuned?

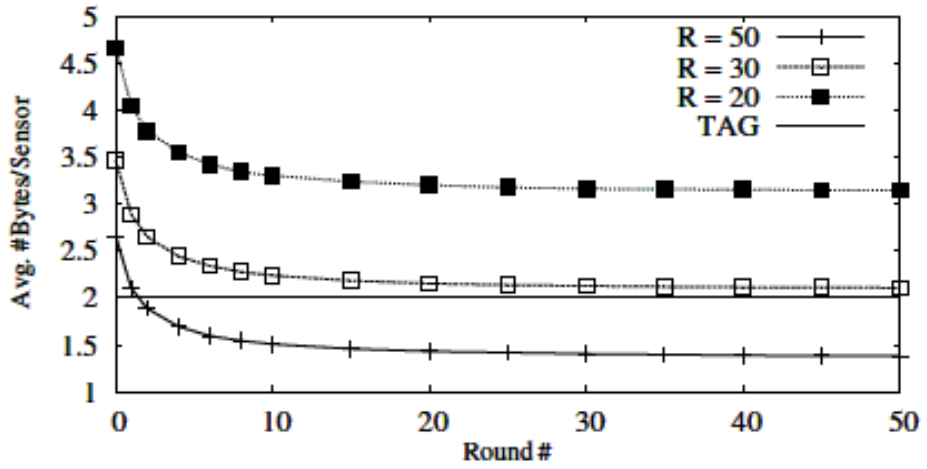
- If the communication device allows variable size messages, then every bit counts.
  - Max/Min: instead of sending the max or min value, we can send (max-LBm) or (min-LBm)
    - Overhead = log2 (Ubm – Lbm) instead of log2 (Ub)
  - Sum/ Average: instead of sending the sum of all values, we can send
    - Overhead = log2 ( ) instead of log2(Ub * count)

Performance Evaluation Settings

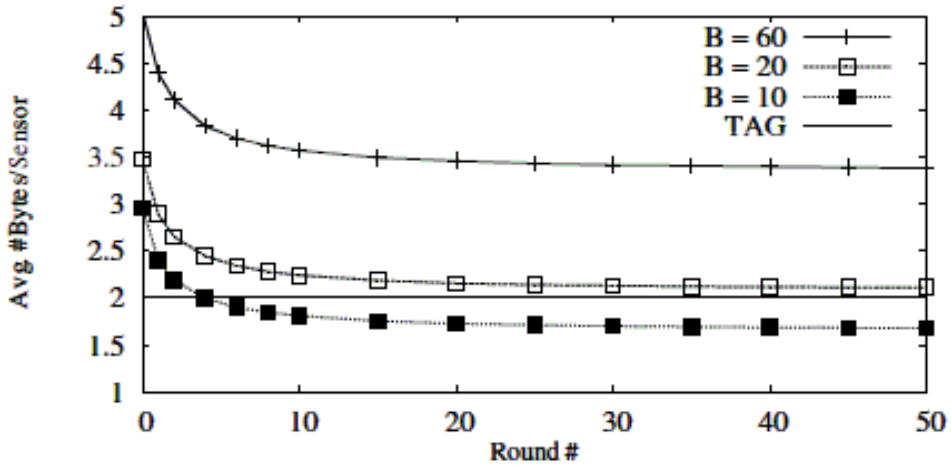
Size of monitored area: 200 m x 200 m
Nodes are static and uniformly distributed
Nodes' values are uniformly distributed [0, 216]
Monitored value consumes 2 bytes
Packet size: 128 bytes
Competitor: TAG (Madden et al, 2002)
Performance Indicator: Transmission cost (bytes)
Aggregate Query: Max

Parameters investigated	Used values
Radio range [m]	20, 30, 40, 50, 60
Histogram Size [# bins]	5, 10, 20, 40, 60
N [# nodes]	1000, 2000, 3000, 4000, 5000
Average Amount of Change	1%, 25%, 50%, 75%, 100%
Probability of Change	1%, 25%, 50%, 75%, 100%

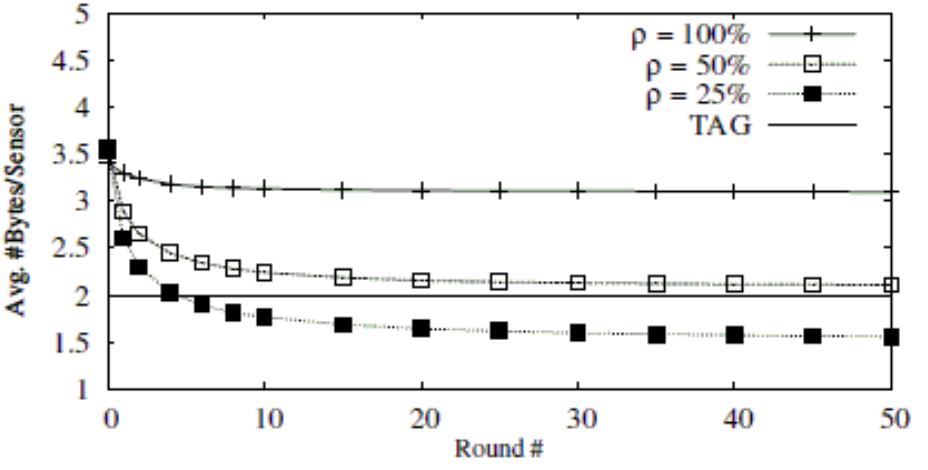
Performance Evaluation



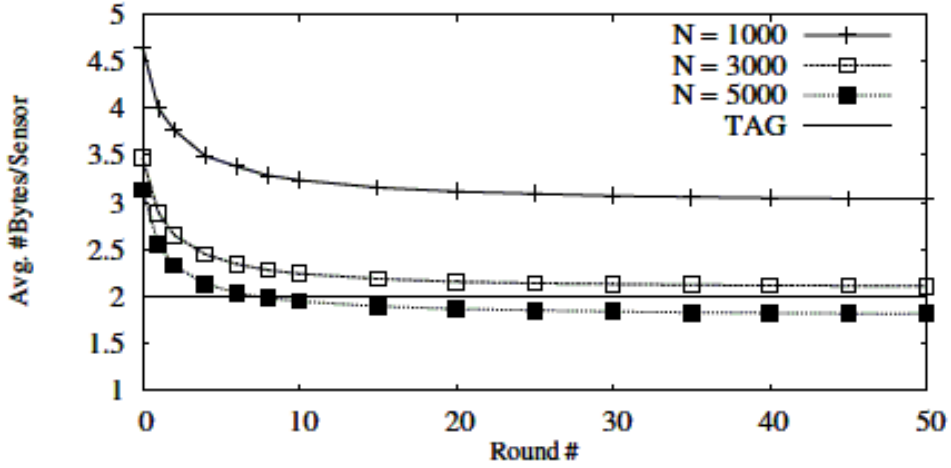
Radio Range (R)



Histogram Size (B)

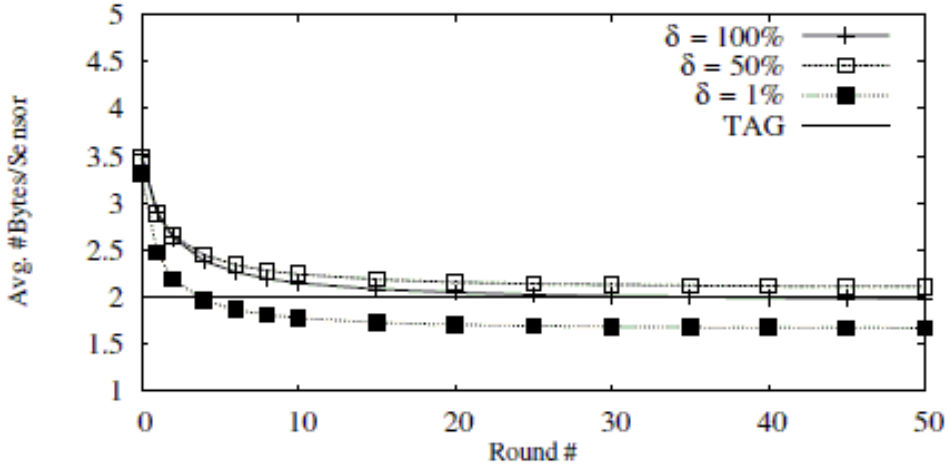


Probability a value changes (ρ)



Number of Nodes (N)

Performance Evaluation (2)



(b) Average amount of change per value (δ)

Conclusions

- HIU is a new algorithm for Continuous Histogram queries.
  - On Average, HIU multiplies the network lifetime about three times comparing to TAG.
- A Histogram also offers approximate free answers for other aggregate queries.
- HIU offers exact (Inexpensive) answers for other aggregate queries.
- HIU also seems to scale well.

Current Research

- What about Holistic queries (e.g. Median) ?
  - We are exploring how to use Histogram to answer harder aggregate queries like Median queries.

Questions ?

Energy Consumption Equations

- Send a message:
- Receive a message:

$$S = 50\ nJ t = 10\ pJ$$

$$r = 50\ nJ$$