Wireless Sensor Networks: Research Challenges

Anandha Gopalan
09/12/11

Lecture 2: Outline

- Research Challenges
  - Data Storage
  - Data Dissemination
  - Power Management
- Conclusion

Data Storage

- Three important questions
  - What data needs to be stored ?
    - About the environment being monitored
    - Events generated
    - Generally application specific
  - Where should it be stored ?
    - Nodes in the network
    - Base station
  - How long should it be stored ?
    - Depends on how "fresh" the data needs to be
    - Directly proportional to how often new data is generated

Data Storage

- Base Station Storage
  - Events are sent to base stations where queries are issued and evaluated
  - Best suited for continuous queries
- In-Network Storage (INS)
  - Events are stored in the sensor nodes
  - Best suited for ad-hoc queries
  - Most INS schemes are Data-Centric Storage (DCS) schemes
    - Event data are "named" and stored by name

In-Network Storage

- Treat sensor network as distributed database
  - Use a simple SQL-like language to query the WSN (e.g.: TinyDB/TinySQL)
  - Allows for efficient in-network aggregation and retrieval of query results
- Examples include:
  - Geographic Hash Table (GHT)
  - Distributed Index for Multi-Dimensional data in Sensor Networks (DIM)
  - Spatio-Temporal Data-Centric Storage for Real-Time Sensornet applications (STDCS)

GHT

- Hashes event names to a unique geographic location for storage and retrieval
- Built on geographic routing

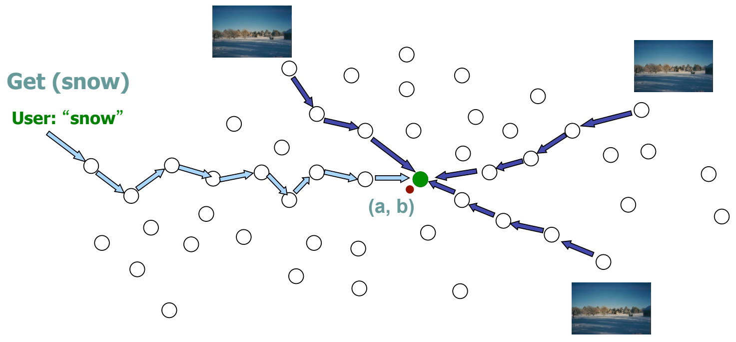
GHT

- Two operations:
  - Put(*k, v*) stores value *v* (observed data) according to key *k* (associated with the name of the data)
  - Get(*k*) retrieves stored value associated with key *k*
- Hash key *k* into geo coordinates
  - Store and retrieve events for that key at the sensor closest to the *location*
  - Spreads key space storage load evenly across network

GHT

- Uses GPSR (Greedy Perimeter Stateless Routing) as its underlying routing algorithm
  - Geographic routing protocol
  - Packets are addressed to a particular location
  - Greedy forwarding
    - Forward packets to nodes that are always progressively closer to the destination

GHT



Shortcomings

- GHT
  - Hot-spots
    - A large percentage of events is mapped to few sensor nodes

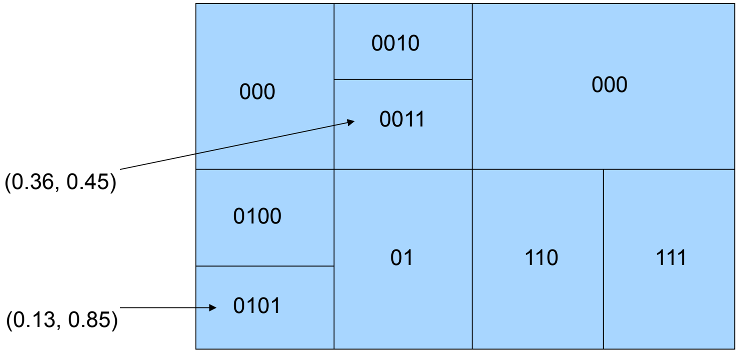
STDCS

- Goal of STDCS
  - Load-Balancing of storage load among sensors
- Features
  - Temporally evolving spatial indexing scheme to balance query load among sensors
  - Dynamic query hotspot detection and decomposition

STDCS

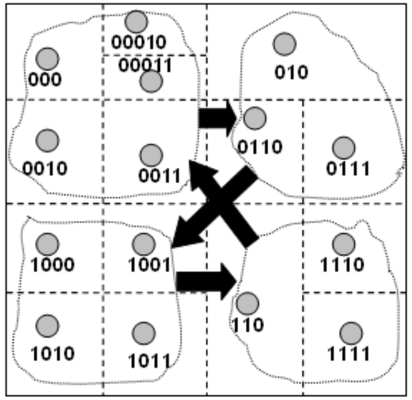
- Network is divided into zones
  - Each node is mapped to one zone
- Multi-dimensional ranges are hashed to a unique binary code
  - Binary codes are mapped to unique geographic zones for storage and retrieval

STDCS



STDCS

- Allows for switching the zones based on the formation of hot-spots
  - Continuously keeping track of hotspots using the Average Querying Frequency (AQF) metric



Shortcomings

- STDCS
  - Overkill, if there does not seem any reason for a hot-spot

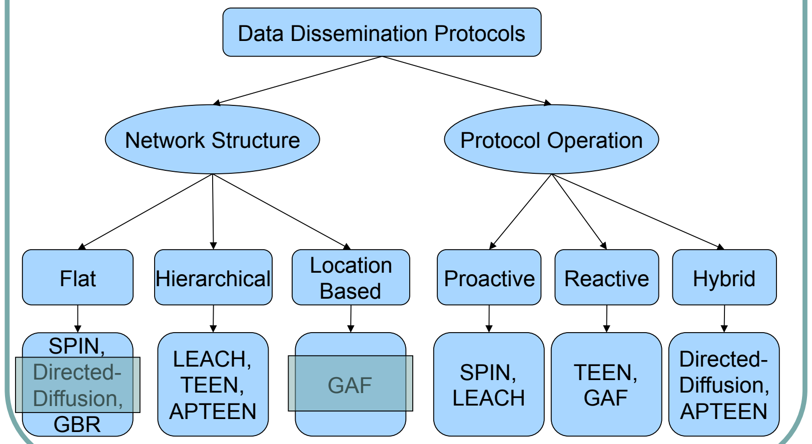
Research Challenges

- Research Challenges
  - Data Storage
  - Data Dissemination
  - Power Management
- Conclusion

Data Dissemination

- Requirements
  - Application specific
  - Data centric
  - Capable of aggregating data
  - Energy efficient

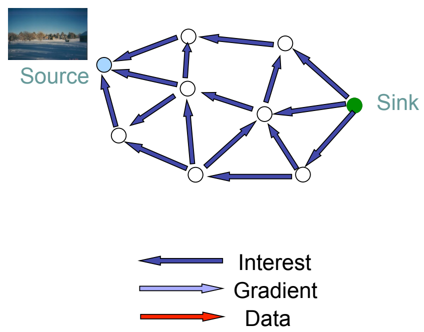
Data dissemination - Classification



Directed Diffusion

- Data is *named* using attribute-value pairs
- *Interests* are disseminated throughout the WSN
  - Sets up gradients to “draw” events from sources to sinks along multiple paths

Directed Diffusion



Directed Diffusion

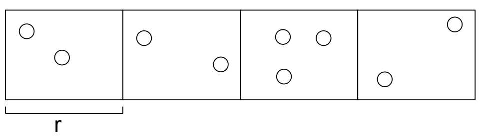
- Sink may *reinforce* one particular neighbour
  - For higher data rate
  - Shown as the data path in the previous slide
- *Negative* reinforcement to "repair" degraded links
  - Re-sending interest with lower data rate

Geographical Adaptive Fidelity (GAF)

- Motivation
  - Idle energy dominated energy consumption
- Solution
  - Put redundant nodes in sleep mode by using a *virtual grid*
    - Divide network into small virtual grids using location information
    - At any time only one device per grid is active
    - All nodes in adjacent grids can communicate with each other

GAF

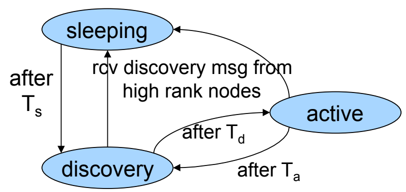
<u>Example of a virtual Grid</u>



r = size of virtual grid
R = transmission range
$r \leq R/\sqrt{5}$

GAF

- Three states in GAF
  - *sleeping*
  - *discovery*
    - find nodes within same grid (initial state)
  - *active*
    - Participates in routing



1. Start in *discovery* state,
2. After T_d,
   1. Broadcast discovery msg
   2. Enter *active* state
   3. Set timer T_a
   4. Periodically re-broadcast discovery msg (with a random number (R) between 0 and n)
3. After T_a, return to *discovery* state
4. Active node can change to *sleep* state when a higher-ranked node (> R) handles routing

Data Dissemination Drawbacks

- Directed Diffusion
  - High communication cost
  - Reinforcement may lead to many high quality paths (not needed)
  - High cost of set-up if interests change frequently
    - Real-time sensornets
- GAF
  - Duty cycle is based on application and system-level information
  - GAF decision to turn radio on/off is independent of routing protocols
    - Packet loss
    - GAF can inform routing protocol of impending suspension
  - What if the only node in a grid dies ?

Conclusion

- Focussed on Research Challenges
  - Data Storage
  - Data Dissemination
- Focus of next lecture
  - Power management
  - Further discussion about current research and a look at future direction and challenges

☺ Questions ☺