Data dissemination and gathering in sensor networks

Dmitrij Lagutin, dlagutin@cc.hut.fi

T-79.5401 Special Course in Mobility Management: Ad hoc networks, 18.4.2007

Contents

- Introduction
- Data dissemination methods
  - Flooding and gossiping
  - SPIN
  - Cost-field approach
- Data gathering methods
  - Direct transmission
  - PEGASIS
  - Binary scheme

Introduction

- Sensor networks consist of large amount of small sensor nodes that collect some kind of data from the environment (e.g. temperature or air pressure)
- Sensor networks differ from ad hoc networks in many ways
  - Number of nodes in sensor networks is significantly higher
  - Sensor nodes have very limited battery capacity, often the battery is not replaceable or rechargeable
  - Sensor nodes may not have an unique identifier like an IP address

Data dissemination

- Data dissemination is a process by which data and queries for data are routed in the sensor network
- Some concepts
  - Source is the node that generates data
  - Event is the information to be reported
  - Sink is the node that is interested in data
  - Interest is a descriptor for the event that node is interested in

Data dissemination

- Data dissemination is a two-step process
  - Node that is interested in some events, broadcasts its interests (like temperature or air humidity) to its neighbors periodically. Interests are then propagated to the whole network
  - Nodes that have the requested data, send back data after receiving the request
- Intermediate nodes in the sensor networks keep a cache of received interests and data

Data dissemination: Flooding

- In flooding method each node that receives the packet broadcasts it forward (as long as maximum hop count is not reached). Flooding is a simple method for data dissemination but it has many disadvantages:
  - Implosion, duplicate messages can be sent to the same node
  - Overlap, same event maybe sensed by more than one node because of overlapping coverage thus neighbors can receive duplicate reports of the same event
  - With flooding method many redundant transmissions occur and this decreases the lifetime of the network

Data dissemination: Gossiping

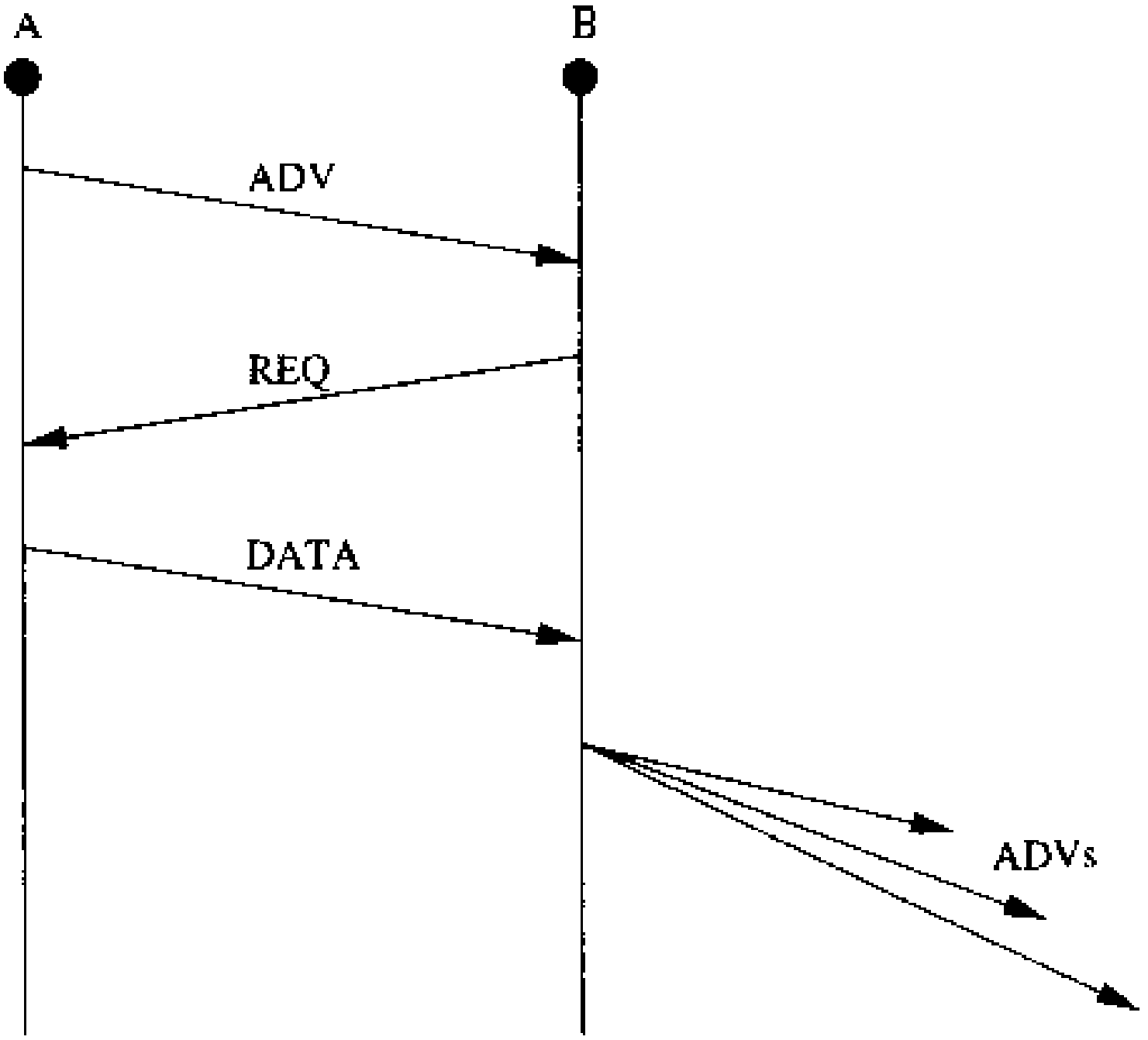
- Gossiping uses the basic idea of flooding, but the packet is sent to only one randomly selected neighbor
- Gossiping avoids the problem of implosion and it does not waste network resources as much as flooding
- However, since the neighbor is selected randomly, some nodes may not receive the message at all. Thus, gossiping is not very reliable method

Data dissemination: SPIN

- Sensor Protocols for Information via Negotiation
- SPIN uses negotiation and resource adaptation, thus it is more efficient than flooding
- SPIN is based on data-centric routing, available data is advertised first and then request for the data are sent
- In SPIN, sensor node that has collected the data sends ADV message with the description of the data to its neighbors

Data dissemination: SPIN

- If the neighbor is interested in the data, it sends REQ message back
  - Then the sensor node sends the data to the neighbor
- Neighbor also sends ADV messages to its neighbors, thus ADV messages propagate through the network
- SPIN-2 adds energy or resource threshold for participation in relaying SPIN messages
  - Nodes participate in the ADV-REQ-DATA handshake only if they have sufficient resources. This prolongs the lifetime of the network

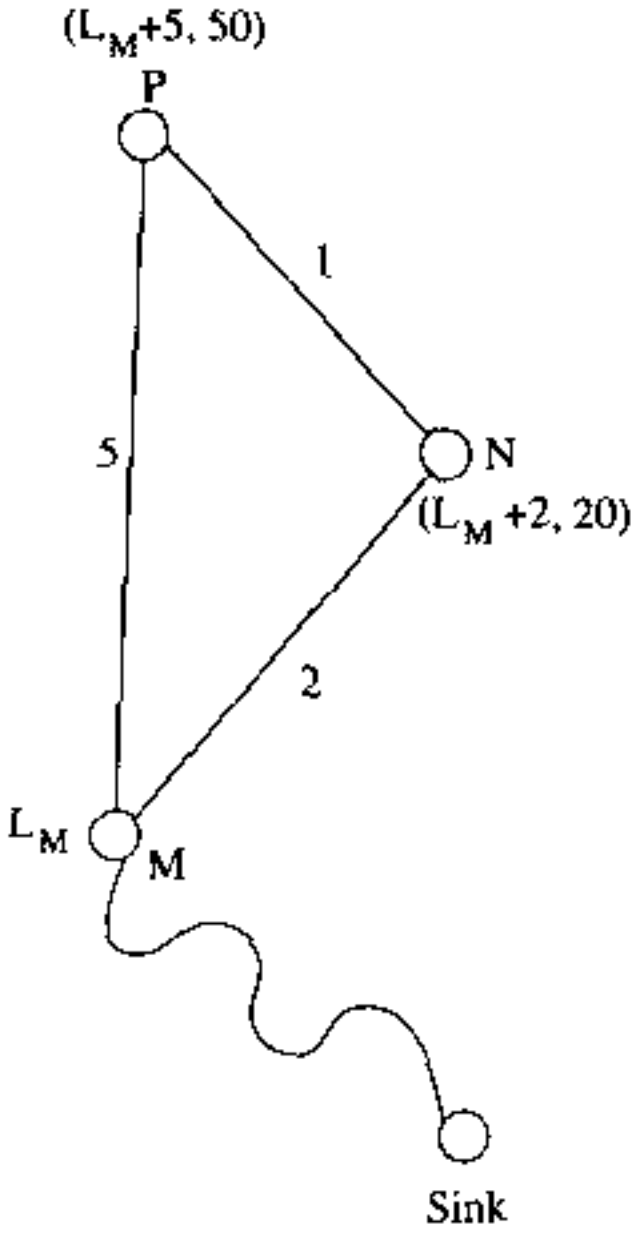

A
B
ADV
REQ
DATA
ADVs
Sensor node

Data dissemination: Cost-field approach

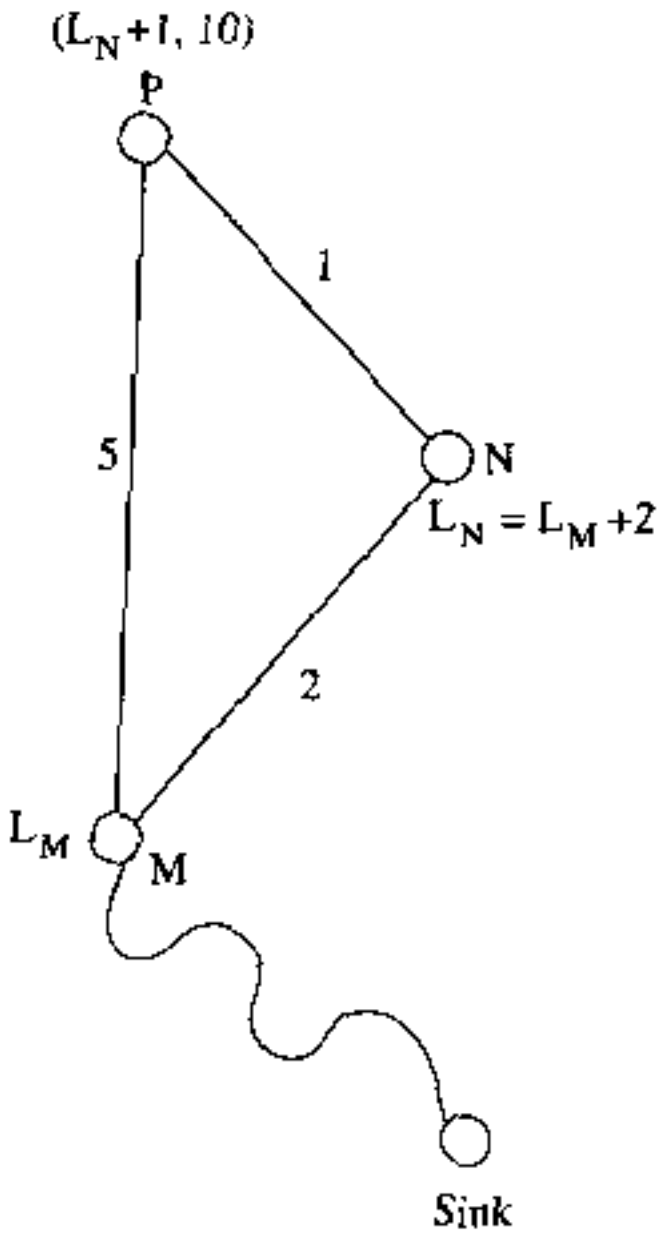
- Cost-field approach is a two-phase process
  - Cost field is set up based on metrics such as delay at all sensor nodes
  - The data is disseminated using the cost field
- Cost-field is set up starting from the sink node, which broadcasts an ADV packet with cost 0
- When node *N* hears an ADV packet from node *M*, it sets its own path cost to $min(L_N, L_M+C_{NM})$
  - L_N is the current total path cost from sink to node *N*
  - L_M is the cost from node *M* to sink
  - C_{NM} is the cost from node *N* to *M*

Data dissemination: Cost-field approach

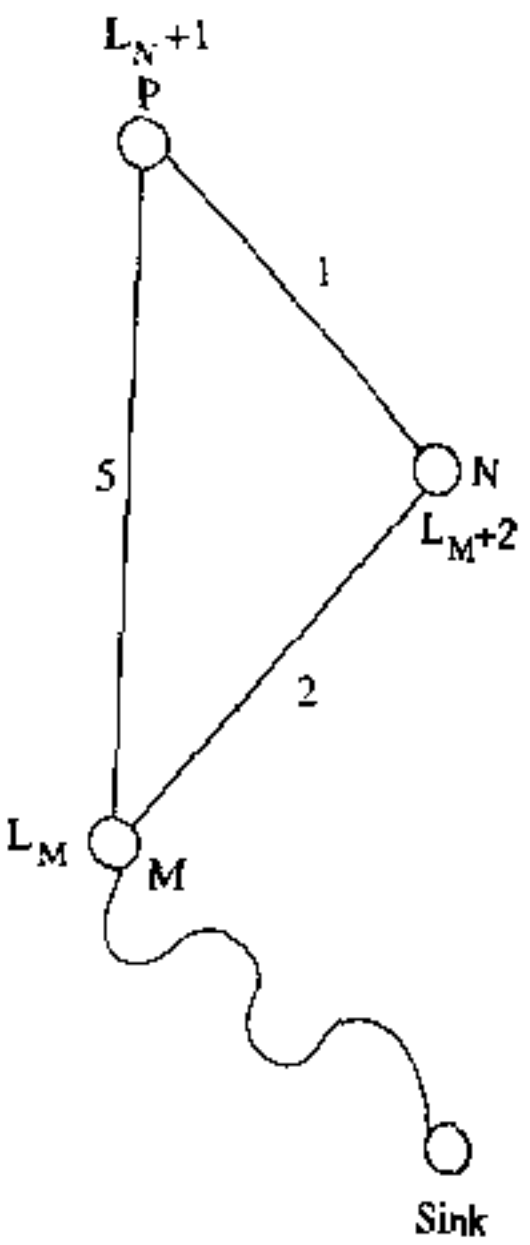
- Cost-field approach does not need to maintain explicit information about the optimal path
- Back-off timers are used to avoid transmissions of non-optimal costs. Otherwise, the approach would work like flooding which is very ineffective
  - Node *N* will broadcast received ADV message forward only after time $\gamma * C_{NM}$ has passed, γ is the parameter of the algorithm



(a) Time T, after M's ADV



(b) Time T + 20, after N's ADV



(c) Time T + 30, after P's ADV

Data dissemination: Cost-field approach

- Second phase of the cost-field approach is data dissemination
- A source sends its message to sink S with cost C_S
  - The message also contains a cost-so-far field that is initially set to 0
- Each node will forward the packet if the cost-so-far field plus its own cost equals the original source-to-sink cost
  - This ensures that the optimal path is chosen, intermediate nodes also update the cost-so-far field while forwarding

Data gathering

- In data gathering, all sensor nodes send their collected data to the base station periodically or on demand
- The goal of data gathering algorithms is to minimize the transmission delay and power consumption of data gathering
- Data gathering vs. Data dissemination
  - In data dissemination other nodes can request the data, not just the base station
  - In data dissemination data is transmitted on demand, while in data gathering data may be transmitted periodically

Data gathering: Direct transmission

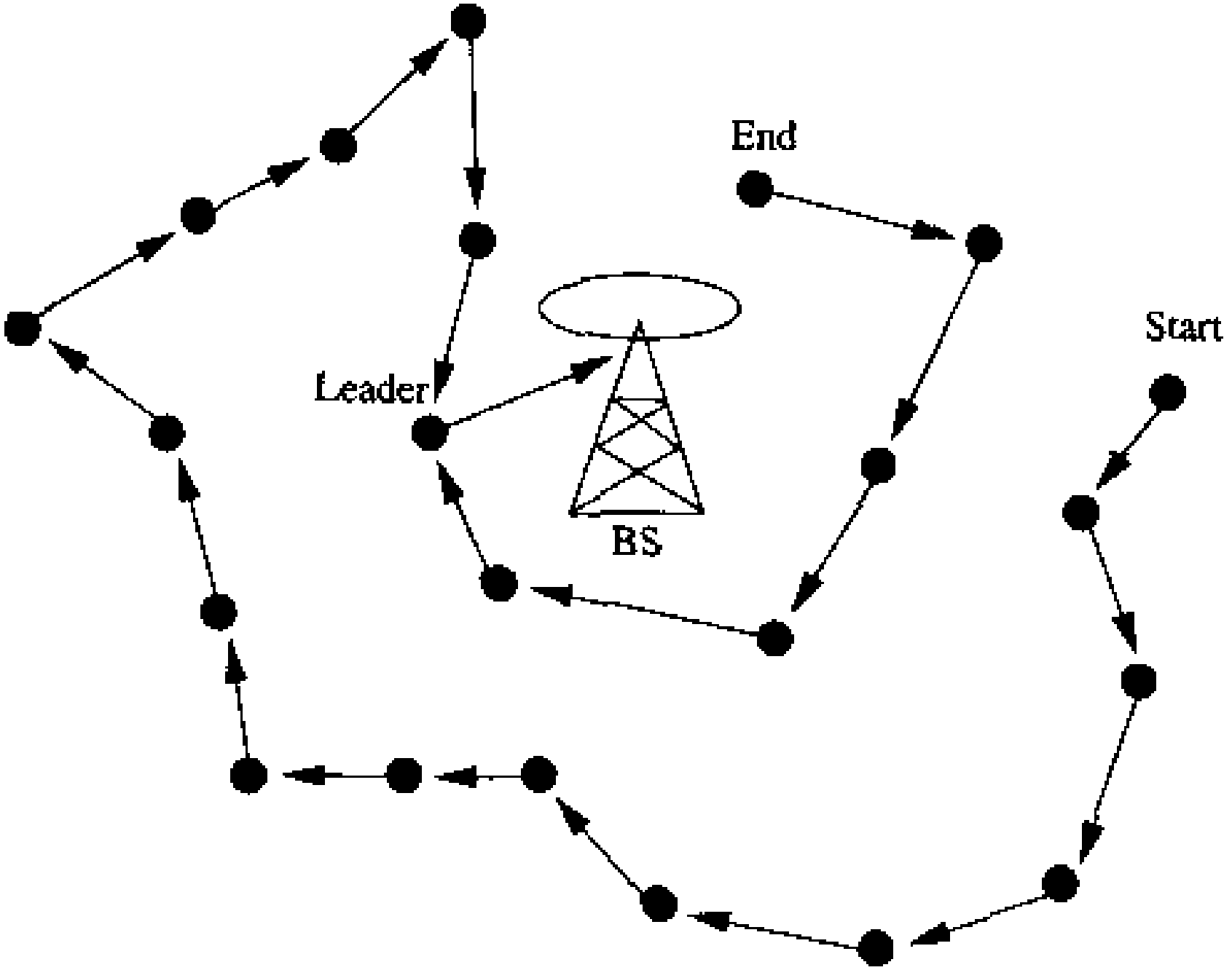
- All nodes send their data directly to the base station
- Very simple but inefficient method
  - Some nodes may be very far away from the base station, thus transmissions will use high amount of energy
  - In order to avoid collision, nodes must take turns while transmitting to the base station, thus the delay in direct transmission is also high

Data gathering: PEGASIS

- Power-Efficient Gathering for Sensor Information Systems
- PEGASIS aims to
  - Minimize the transmission distances over the whole network
  - Minimize the broadcast overhead
  - Minimize the amount of messages that are sent to the base station
  - Distribute the energy consumption equally between all nodes

Data gathering: PEGASIS

- PEGASIS assumes that all sensor nodes know the topology of the whole network
- A chain of sensor nodes is constructed before the data transmission begins
  - Construction of the chain starts from the node farthest from the base station and uses a greedy algorithm
  - At each step in the chain, data aggregation is carried out and only one message is passed to the next node
- The node that is selected as a leader transmits then all information to the base station
- The delay is *O(N)* where *N* is the amount of sensor nodes in the network



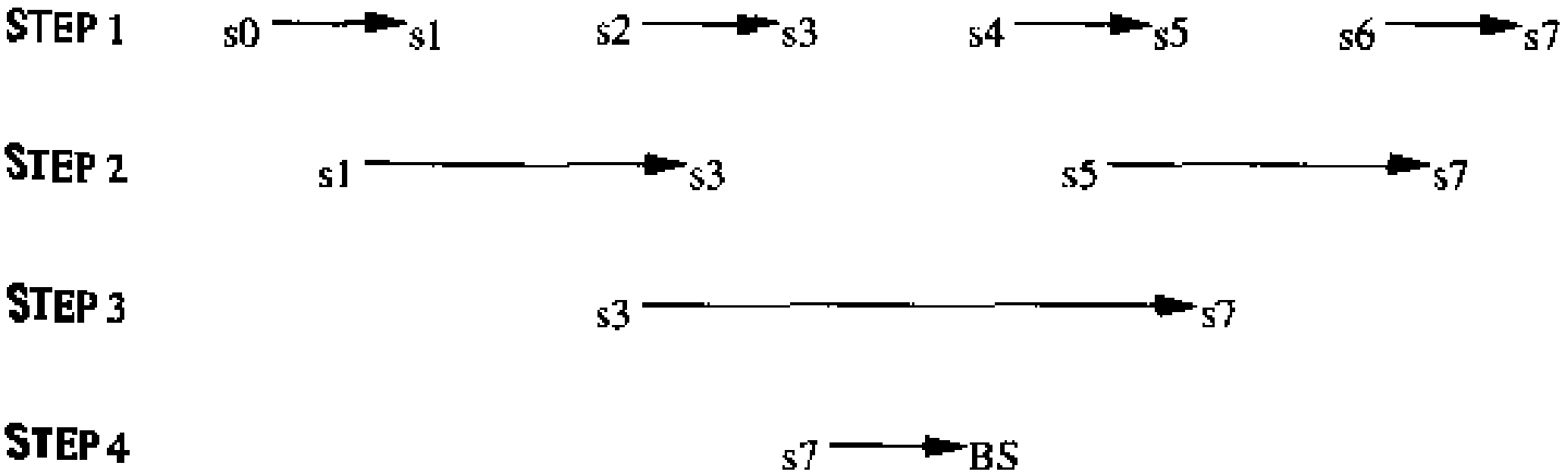
Sensor node

Data gathering: Binary scheme

- Binary scheme is also a chain-based scheme
- Nodes are divided into different levels. Nodes that receive messages at one level rise to the next level
- Has a low delay of $O(log_2 N)$
- Binary scheme requires support for simultaneous transmissions from the network, it works with e.g. CDMA. A similar Chain-based three-level scheme do not use simultaneous transmissions, thus it works with all networks.

Data gathering: Binary scheme

- Example: eight nodes s0...s7, latency for transmitting all data to the base station is four steps



Conclusions

- Several different approaches exist to solve the problem of data dissemination and gathering
  - Simple approaches like flooding and direct transmission are very ineffective
- Sensor networks are challenging because of their decentralized nature and since sensor nodes have very limited battery capacity