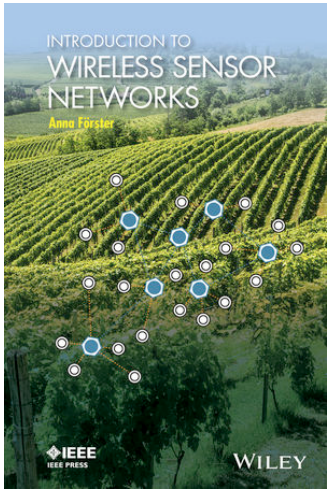


INTRODUCTION TO
WIRELESS SENSOR NETWORKS

CHAPTER 6:
DATA AGGREGATION AND CLUSTERING

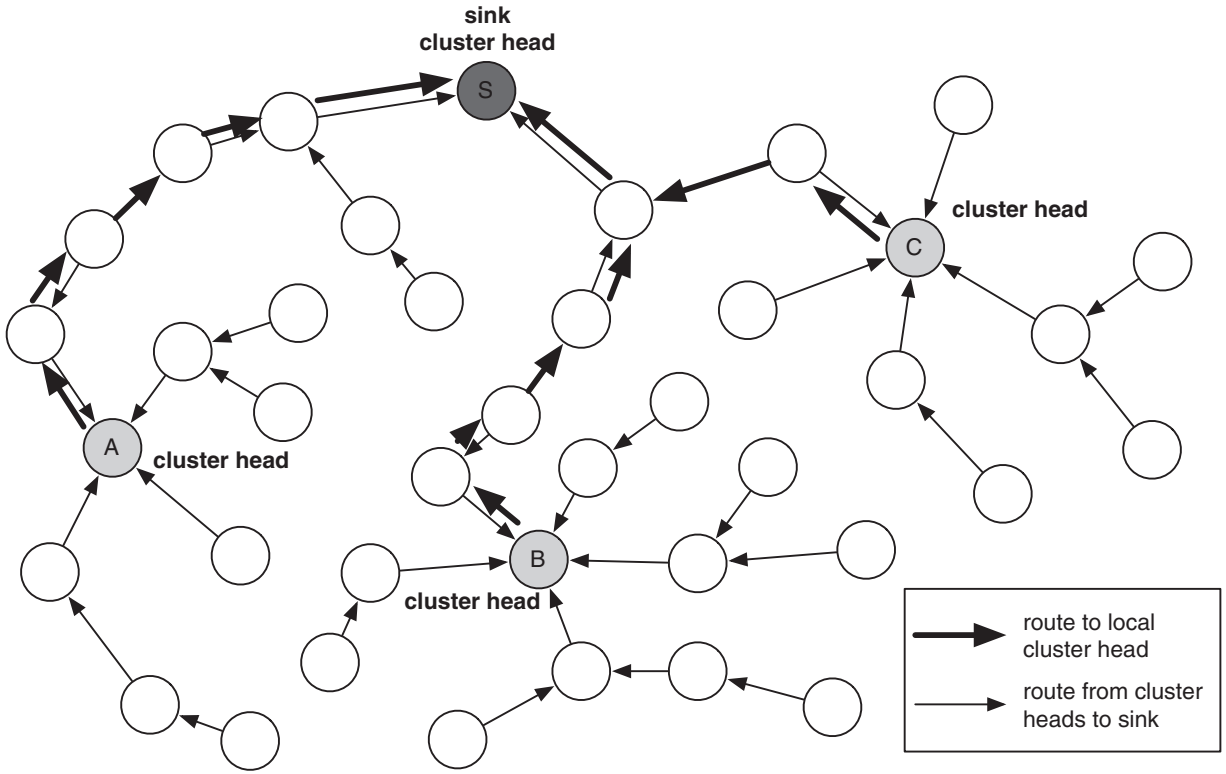
Anna Förster

OVERVIEW

1. Clustering Techniques
    1. Random Clustering
    2. Nearest Sink
    3. Geographic Clustering
2. In-Network Processing and Data Aggregation
    1. Compression
    2. Statistical Techniques
3. Compressive Sampling

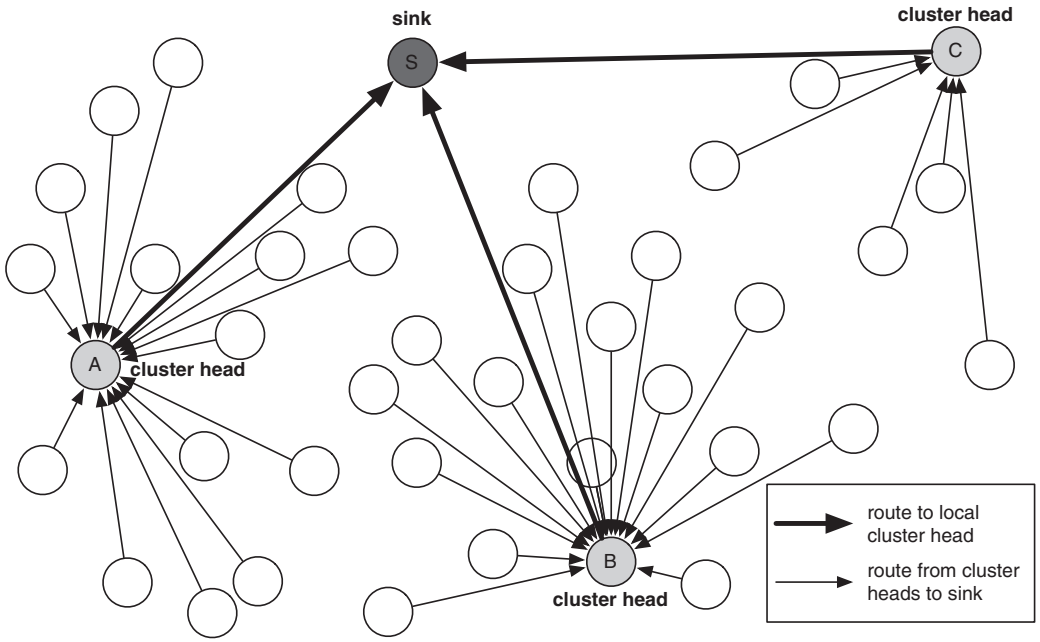
Clustering Techniques

- Developed to handle very large networks



Random Clustering

- The cluster heads are selected randomly among all nodes
- Cluster head roles are updated from time to time to balance their energy consumption
- Example: LEACH

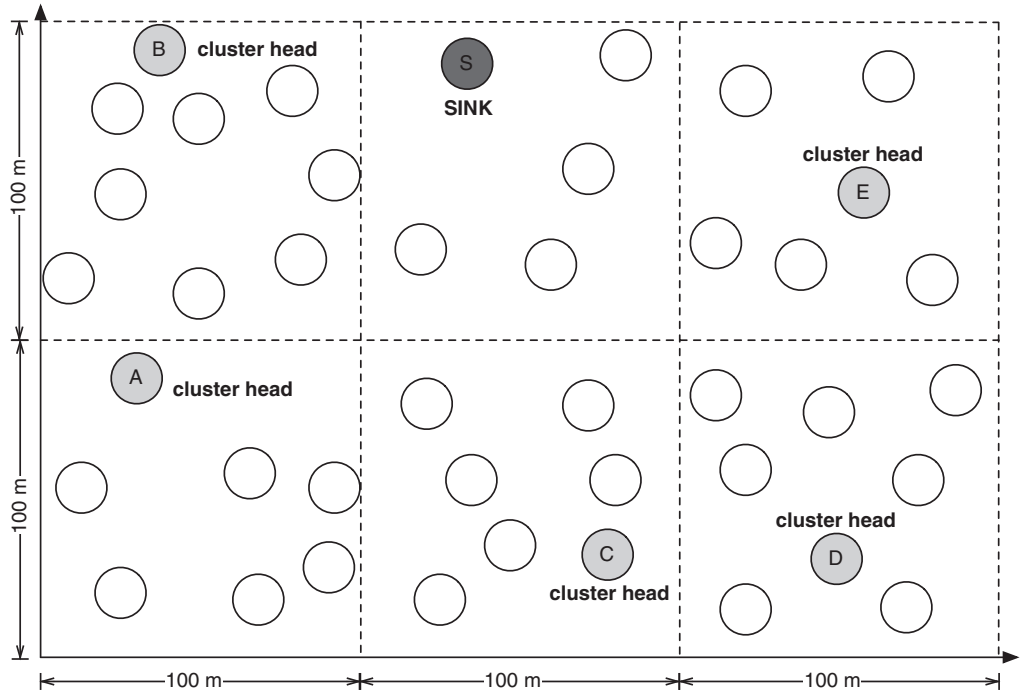


Nearest Sink

- Developed to balance better the sizes of the clusters
- More than one sink deployed in a network (e.g. with more resources than other nodes)
- Every sink announces itself via full-network broadcast
- Every node selects the “nearest” sink – in terms of hops, ETX, delay, etc.

Geographic Clustering

- Divides the network into geographic areas
- One sink per area
- All nodes belong to that single sink
- Balances very well the size of clusters

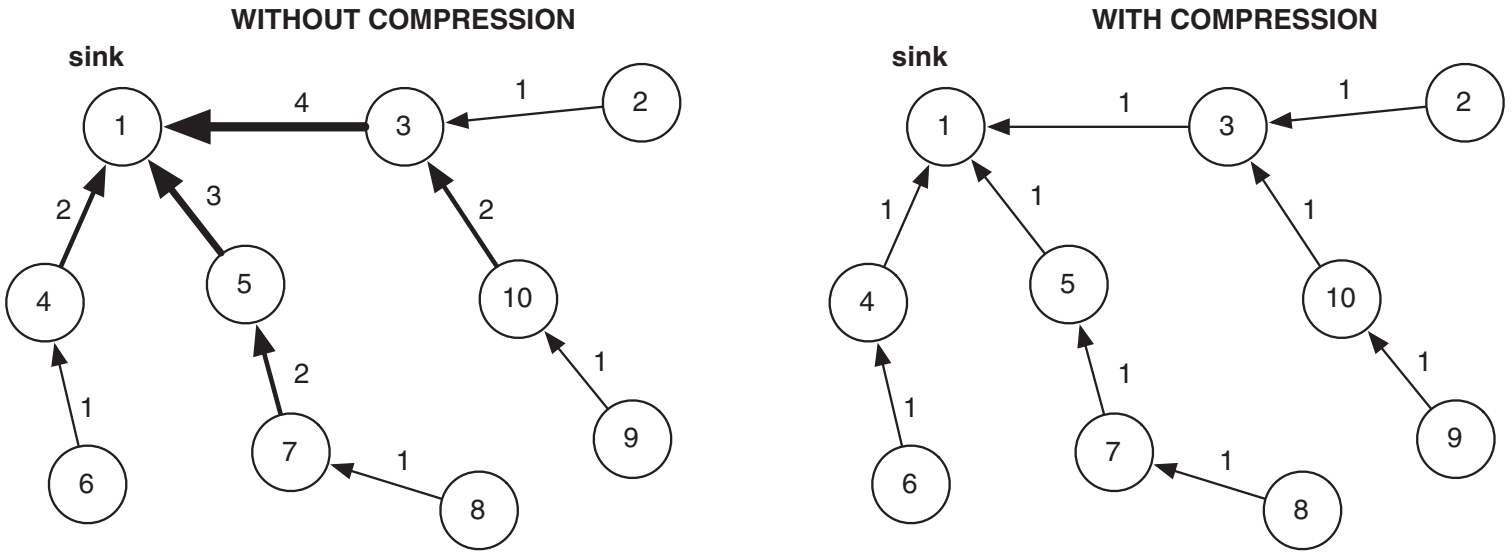


In-Network Processing and Data Aggregation

- Used usually together with clustering
- Saves additional energy by not transmitting all data to the sink
- Variants:
  - Compression of data
  - Statistical Techniques
  - Compressive Sampling

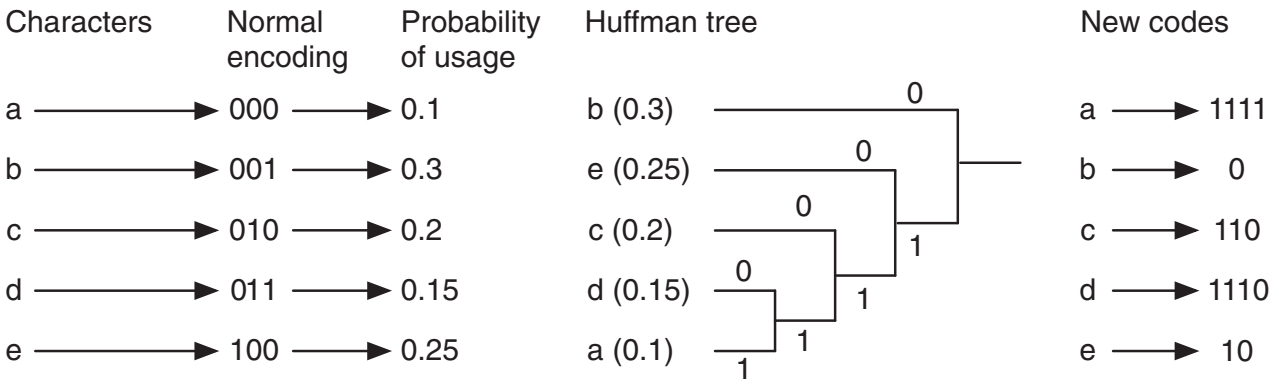
Compression

- Several packets get combined into a single one (recall that individual packets are costly, while larger packets are almost as costly as small ones)



Huffman Codes

- Compresses the contents of the data packets
- Based on observation: some letters/symbols are used more often than others
- Consequence: use shorter codes for often used symbols
- Compare: ASCII code uses the same code length for all characters



Example with normal encoding

abbcdbbbeee → 000 001 001 010 011 001 001 001 100 100 100

Example with Huffman decoding

abbcdbbbeee → 1111 0 0 110 1110 0 0 0 10 10 10

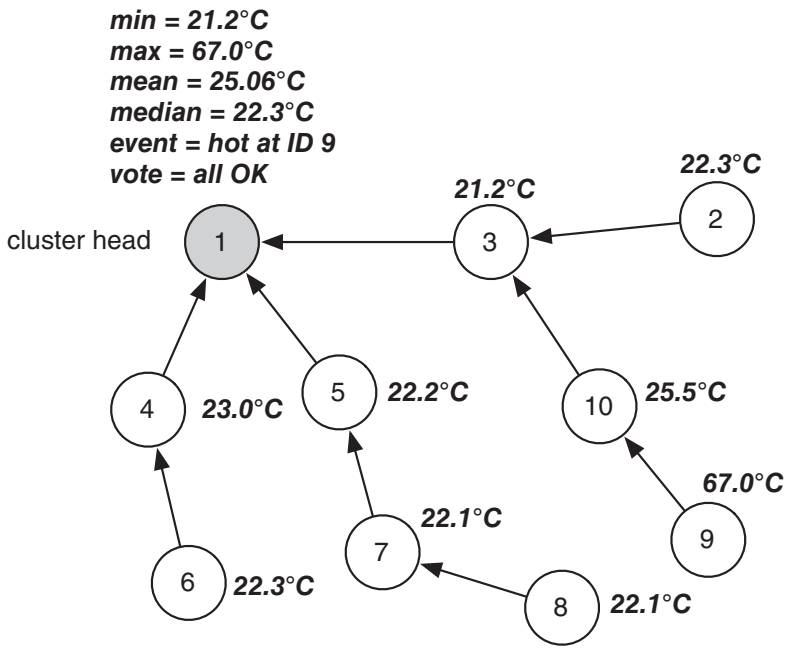
Model Prediction

- Develop a mathematical model of the data
- Example: Temperature in the room should be between 21° and 23°.
- If the data matches the model, do not transmit the data
- Very complex models can be developed (e.g. for soil temperature in a day-night rhythm)

- Problem: if no data arrives, how to know that the node is still alive?
- Solution: Heart-beat protocol. Sends a small packet every extended period of time (e.g. once per day)

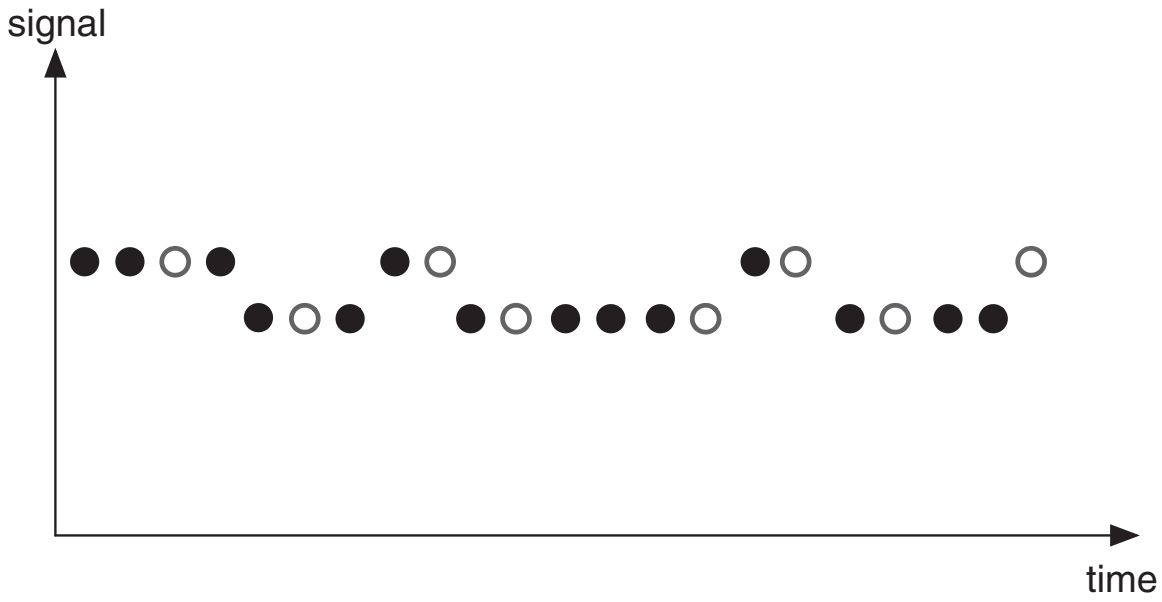
Statistical Techniques

- On the way to the sink, statistically evaluate the data.
- Once statistically evaluated, the original data cannot be recovered



Compressive Sensing

- Minimizes the number of required sensor data to fully depict the underlying process/signal.
- Mathematically quite complex (refer to literature for more details)
- Can save a lot of energy, if underlying process follows the mathematical model (not all natural processes do!)



Clustering and Data Aggregation: Summary

- **Clustering** organizes the network into sub-networks called clusters. Clusters can be identified randomly or geographically. Each cluster gathers all of its data on a single node, called the cluster head, and then sends the data in a compressed or processed way to the sink.
- **Compression** refers to minimizing the required communication overhead and storage for individually sampled data. For example, several data items can be unified in a single data packet or data compression techniques can be applied to further reduce their volume.
- **In-network processing** refers to calculating some statistical values or to allow voting between the nodes of a cluster. For example, only the mean values can be transferred to the sink.
- **Compressive sensing** refers to a concept from signal processing, which randomly samples a signal instead of continuously. This reduces the sampled data at the source directly and no further compression is needed.

Clustering and Data Aggregation: Summary

- Clustering is a communication organization measure and needs to be used together with other techniques to achieve communication savings.
- Compression and compressive sensing preserve the full properties of the originally sensed signal, and sometimes it can even be completely reconstructed. These two concepts are our preferred options.
- Statistical in-network processing does not allow for reconstruction of the original signal, nor of all of its properties, and has to be used very carefully.