IEEE Int. Conf. on Networking

A New Adaptive Clustering Technique for Large-Scale Sensor Networks

Prof. Ming Yu

Dept. of Electrical & Computer Engineering
State University of New York
Binghamton, NY 13902
Email: *mingyu@binghamton.edu*

Outline of The Presentation

- Introduction
- System Architecture
- Adaptive Clustering
  - Real-Time Estimation of Number of Nodes
  - The Adaptive Clustering Algorithm
- Simulation Results
  - Example Network 1
  - Example Network 2
- Conclusion

Introduction

Problems in Networking Sensor Systems:

- A large numbers of small, relatively inexpensive and low-power sensors are connected via a wireless network.
- Through the network, the data extracted from these sensors is sent to a nearby base station (BS), which forwards the data to a remote data center for further processing.
- Challenging Problem: how to dynamically organize these sensors into a wireless communication network and route detected event information between field sensors and BS, i.e., the design of a networking protocol.

Networking Requirements

- Primary Requirements: energy efficiency, scalability, adaptivity, and security.
- Secondary Requirements: channel efficiency and network performance, which includes packet delay, packet loss ratio, throughput, and fairness.
- The focuses of this paper: energy efficiency, scalability, and adaptivity.
  - Energy-efficient: the energy consumption in delivering packets from source to destination is minimized.
  - Power-Aware: a route with nodes having higher residual (battery) power should be selected, although it may not be the shortest route.

Current Work

- Clustering is an energy-efficient architecture, wherein the individual nodes forward the information to their respective cluster heads.
  - The information is aggregated at the cluster head and then sent to the BS by the cluster head.
  - The cluster heads and BS usually form a multi-hop network, which must have a multi-hop routing protocol.

Current Work: Clustering

- The near term digital radio (NTDR): based on a two-tier clustering without using low-energy routing or MAC.
- d-hop clusters: directly aimed at minimizing the energy spent in the system.
- LEACH: the nodes were organized into cluster hierarchies and TDMA was applied within each cluster. Assumed the number of nodes and the optimal number of clusters to be formed are given.
- HEED: an improvement based on LEACH. Clustering uses node degree cost, which needs to know the number of active nodes. Not guarantee optimal selection of cluster heads in terms of energy.

Sensor and Data Networks

- The significant differences between the two networks:
  - Most traffic in sensor network is triggered by sensed events;
  - In reporting phase, the traffic goes from a hot-spot area, which consists of a few of the sensors, to a BS;
  - In polling phase, traffic goes from a BS to many sensors;
  - Many sensors coordinate for a common task and not all of them must be active.
- Therefore, a sensor network is more like the fault management system (FMS) of a data network rather than the data network itself.

Sensor and Data Networks

- Many of the existing sensor networking protocols have often applied data network concepts and produced inefficient algorithms.
- In a typical application of a sensor network, the routed event can be either an anomaly event autonomously sent out by sensors, or a polling command issued by the remote collection stations.
- We propose the ACT and port many FMS concepts and protocols to sensor networks, such as event routing, reporting, processing, correlation, by using SNMP traps and alarms.

System Architecture

Two-tier hierarchical architecture, includes an adaptive clustering technique and multi-hop routing protocol.

- ACT: adaptive clustering technique.
- In tier-1, cluster members communicate to each other over a hybrid MAC protocol consisting of both contention and schduling.
- In tier-2, cluster heads communicate over a multi-hop routing protocol.
- Focus: an adaptive clustering technique for large-scale sensor networks.

System Architecture

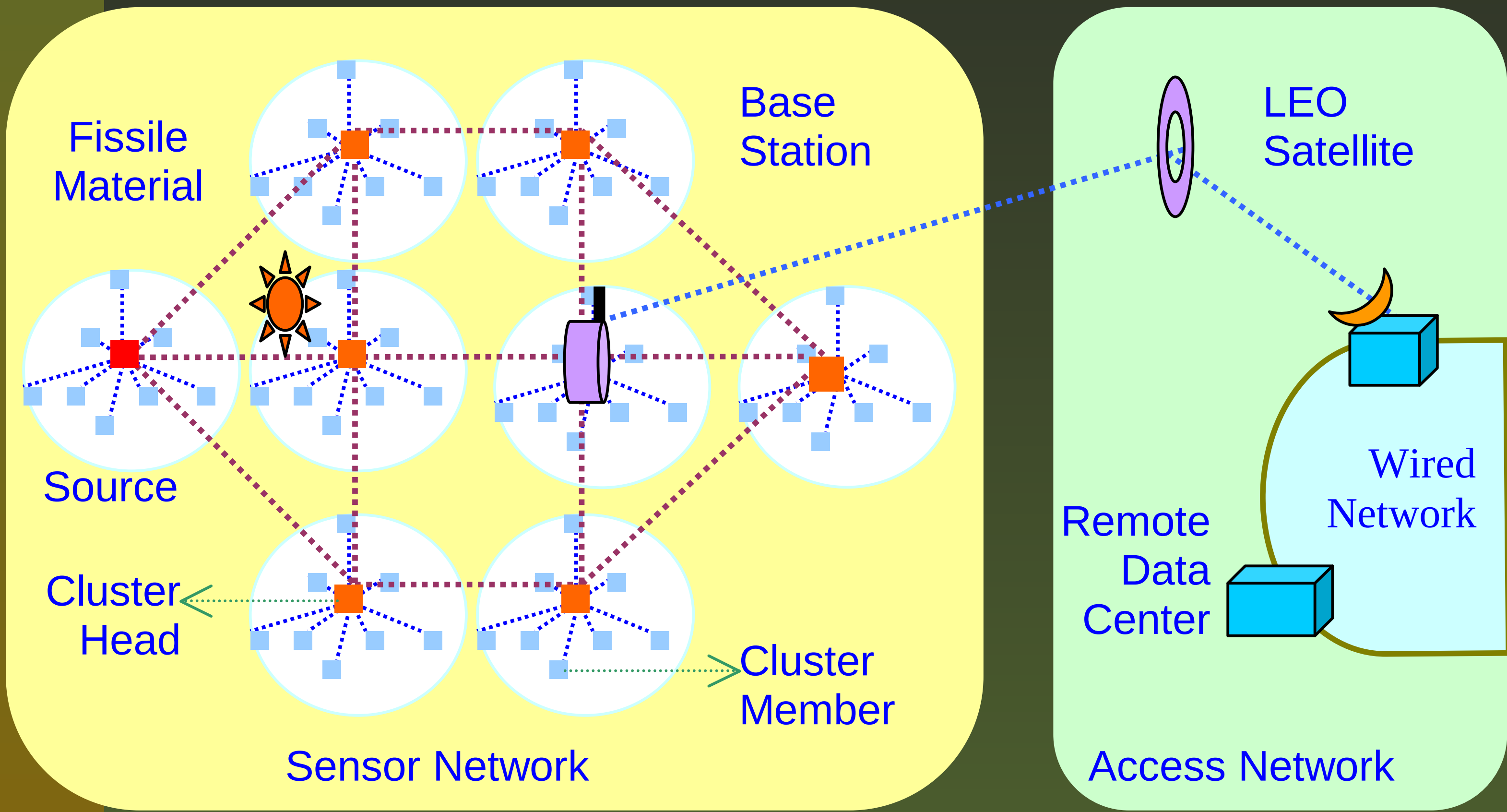

Figure 1: Architecture of A Sensor Network

Assumptions

- Network area: a square area of side $2a$, the area $A = 4a^2$.
- The sensors inside the area are distributed according to a homogeneous spatial Poisson process, with intensity of $\lambda\ sensors/m^2$, hence, the mean value of n is λA.
- The base station (BS) is at the center of the square area.

Notations

- n: the number of sensors in the network, a random variable. For a particular realization, there are N sensors.
- p: the probability of a node becoming cluster head. Thus, on average, there are Np nodes that will become cluster heads, the rest $N(1-p)$ nodes will be cluster members in total.
- k: the average number of cluster heads. Therefore, there are $k = Np$ clusters, each on average has $m = N/k - 1$ member nodes; m is the average number of CMs within a cluster;

Adaptive Clustering

- The critical problem is how to estimate the parameters in real-time, N, the number of active nodes, and the optimal number of clusters, k_{opt}, or equivalently p_{opt}.
- Assume that each node
  - is capable of measuring its received signal power.
  - keeps a record of the minimum power of the signals it has received within its radio range during previous cluster updating cycle.

Radio Propagation Model

- Notations:
  - s_t and s_r are random variables that describe the powers of a signal a sensor node has transmitted and received at distance r, respectively;
  - κ is a dimensionless constant which depends on the antenna characteristics and average attenuation from blockage, while r_0 is a reference distance fro the antenna far-field;
  - $\gamma = 2 \sim 4$ is the path-loss exponent; for simplicity and without lossing generality, we choose $\gamma = 2$.
  - ψ_{dB} is a Gaussian-distributed random variable with mean zero and variance σ_ψ, which can be also measured.

Radio Propagation Model

- The widely used measurement-based radio propagation model is the path-loss model with log-normal shadowing:

$$\frac{s_r}{s_t}(dB) = 10\log_{10}\kappa - 10\gamma\log_{10}\frac{r}{r_0} + \psi_{dB}, \qquad (1)$$

- By taking mean values on the random variables in above Eqn.,

$$\frac{S_r}{S_t}(dB) = 10\log_{10}\kappa - 10\gamma\log_{10}\frac{r}{r_0}, \qquad (2)$$

where S_r and S_t are the mean values of s_r and s_t, respectively.

Radio Propagation Model

- Notations:
  - S: the signal power a node has received.
  - r: the distance of a node to its cluster head.
- Within a cluster, each node communicates to its cluster head by one hop distance; we have the further simplified model:

$$S(r) = \epsilon_0 / r^\gamma, \quad r \leq r_{max}, \tag{3}$$

where

  - ϵ_0: a constant related to the radio trasmission device and propagation environments;
  - r_{max}: the maximum allowable distance between a node and its cluster head.

Voronoi Cell

- Definition: a cell that covers most of the nodes in a cluster.
- ρ_m: the radius of the minimal ball that covers the Voronoi cell. The ball is centered at the nucleus of the cell.
- The probability that ρ_m is greater than a certain value ρ has an upper bound:

$$Prob\{\rho_m > \rho\} \leq 1 - [1 - \exp(-\mu p \lambda \rho^2)]^7, \quad (4)$$

where
  - $\mu = 2(\frac{\pi}{7} + \sin\frac{\pi}{14} + \cos\frac{5\pi}{14})$,
  - $p\lambda$: the equivalent intensity for the point process that describes the cluster head nodes.

Voronoi Cell

- The above Eq. can be simplified:

$$Prob\{\rho_m > \rho\} \leq 7\exp(-\mu p \lambda \rho^2). \tag{5}$$

- Objective: to ensure that the probability of all the member nodes are beyond the minimal ball is very low,
- Define a parameter (degree of clustering) α that:

$$Prob\{\rho_m > \rho\} \leq \alpha, \tag{6}$$

where α is a very small value specified from cluster design requirement.

Maximum Hop Distance

- By combining the two Eqns., we can choose

$$\alpha = 7\exp(-\mu p\lambda\rho^2). \tag{7}$$

- We can find such a ρ that must meet

$$\rho = \sqrt{-\ln(\alpha/7)/(\mu p\lambda)}. \tag{8}$$

In this way, we can make sure that $\rho_m \leq \rho$ in most cases, i. e., the minimal ball covers most of the nodes in the Voronoi cell.

Maximum Hop Distance

- Equivalently, the maximum distance of a node to a cluster head must be bounded by the above Eqn. in order to fall inside of the ball, that is,

$$r_{max} \leq \sqrt{-\ln(\alpha/7)/(\mu p \lambda)}. \tag{9}$$

- Almost all the nodes are within one hop distance from a cluster head. Therefore, we can let r_{max} take the maximum value.

Minimum Signal Power

- Using the equal sign in above Eqn.,

$$S_{min} = -\frac{\epsilon_0 \mu p \lambda}{\ln(\alpha/7)}, \tag{10}$$

where S_{min} is the minimum signal power a node has received.

- Therefore, we can approximately estimate the value of the number of active nodes, by using $N = \lambda A$, if the probability of a node becoming cluster head, i.e., p value, is given.

Energy Efficient Clustering

- In the network level, the requirement on clustering is to have as less number of clusters as possible.
- Within a cluster, the CMs must be as close to their CH as possible.
- To measure the cost incurred by the two requirements:

$$C(q) = e_1 k|D_0| + e_2 \sum_{j=1}^{k} m|D_1|, \tag{11}$$

where D_0 is the average distance from a CH to the BS; and D_1 is the average distance of a CM to its CH.

Energy Efficient Clustering

- After some manipulations, we have

$$C(p) = e_1 d_0 anp + e_2 np(\frac{1}{p} - 1)\frac{2}{3} r_{max}. \qquad (12)$$

- Now the energy efficient clustering becomes an optimization problem: how to choose an optimal p to minimize the total energy consumptions spent in intra- and inter-cluster communications?

How to Choose Energy Coefficient?

- Assume that the average energy consumption in network layer for a CH to receieve and transmit a unit of data (e. g., one bit, or one packet) over a hop distance of H are $\overline{E}_r$ and $\overline{E}_t$, respectively.
- The total energy spent in network layer:

$$e_1 = (\overline{E}_r + \overline{E}_t)/H, \tag{13}$$

where $\overline{E}_r = E_{elec}$, E_{elec} is the energy spent on electrical device for receiving a unit of data; and $\overline{E}_t = E_{elec} + \epsilon_0 H^\gamma$, ϵ_0 and γ can take different values, depending on H, as defined in Table **??**.

How to Choose Energy Coefficient?

- Assume that the average energy consumption in MAC layer for a CM is $\overline{E}_r$ for the receiving of the data bit per time unit (e. g., a time slot); $\overline{E}_t$ for the successful transmission of the bit per time unit; $\overline{E}_c$ for a collision that lasts for one time unit; and $\overline{E}_i$ for an idle that lasts for one time unit.
- The total energy spent in MAC layer:

$$e_2 = (w_s(\overline{E}_r + \overline{E}_t) + w_i\overline{E}_i + w_c\overline{E}_c)/\overline{D}_1, \qquad (14)$$

Estimating N Value

- It can be found that

$$p_{opt} = 1/(\sqrt{c_1 S_{min}} - 1), \tag{15}$$

where $c_1 = -4c_0^2 A \ln(\alpha/7)/(\epsilon_0 \mu)$.

- Assume that during the jth cluster updating cycle, the measurement of S_{min} is denoted as $\tilde{S}_{min}(j)$, the corresponding value of p_{opt} as $\tilde{p}_{opt}(j)$.

$$\tilde{N}(j) = -\frac{\ln(\alpha/7) A \tilde{S}_{min}(j)}{\epsilon_0 \mu \tilde{p}_{opt}(j)}, \tag{16}$$

where $\tilde{N}(j)$ is the calculated value of N in in the jth cluster updating cycle.

Smoothing Estimations

- By using a moving averaging model:

$$\hat{N}(j+1) = \beta\hat{N}(j) + (1-\beta)\tilde{N}(j+1), \tag{17}$$

where $0 < \beta < 1$ is a smoothing factor used to adjust the estimation speed and accuracy.

- The optimal number of clusters can be easily obtained

$$\hat{k}_{opt}(j+1) = \hat{N}(j+1)\tilde{p}_{opt}(j+1). \tag{18}$$

- Note that we can also have

$$\tilde{p}_{opt}(j+1) = 1/(\sqrt{c_1\tilde{S}_{min}(j)} - 1). \tag{19}$$

The Adaptive Clustering Algorithm

1. Initially, specify the value of α, such as $\alpha = 0.001$.
2. Each node measures and records the minimum value $S_{min}(j)$ in the current cycle.
3. Each node computes its p_{opt} value.
4. Each node computes its N and k_{opt} values.
5. By checking the inequality:

$$|\hat{N}(j) - \hat{N}(j-1)| \leq \delta. \tag{20}$$

where δ is a predefined QoS parameter. If yes, don't activate a cluster updating process. Go to step 2.

6. The node adopts the $p_{opt}(j)$ value and tries to become a cluster head with this new probability.
7. Let $j = j + 1$, go to step 2.

Remarks

- Each node actively monitors its received signal power levels during the cluster updating cycles.
- The updating process is triggered by some serious changes in network status, by trading off energy consumption and the adaptivity.
- Each node attempts to find and stay at its optimal cluster locally. The algorithm is carried out by all nodes simultaneously and independently.
- If no serious changes, the algorithm is aborted silently. Otherwise, the node will automatically adjust its cluster to adapt to the changes in real-time.

Conclusion

- We propose
  - a networking architecture for large-scale sensor networks
  - an adaptive clustering technique
- The simulation results have demonstrated that the two-tier hierarchical architecture
  - adapts to changes in network topology,
  - scales well to large network sizes,
  - is power-aware and energy-efficient.
- Future work would be further investigating the applicability of the proposed technique and algorithm to a larger number of other sensor networks.

Simulation Results

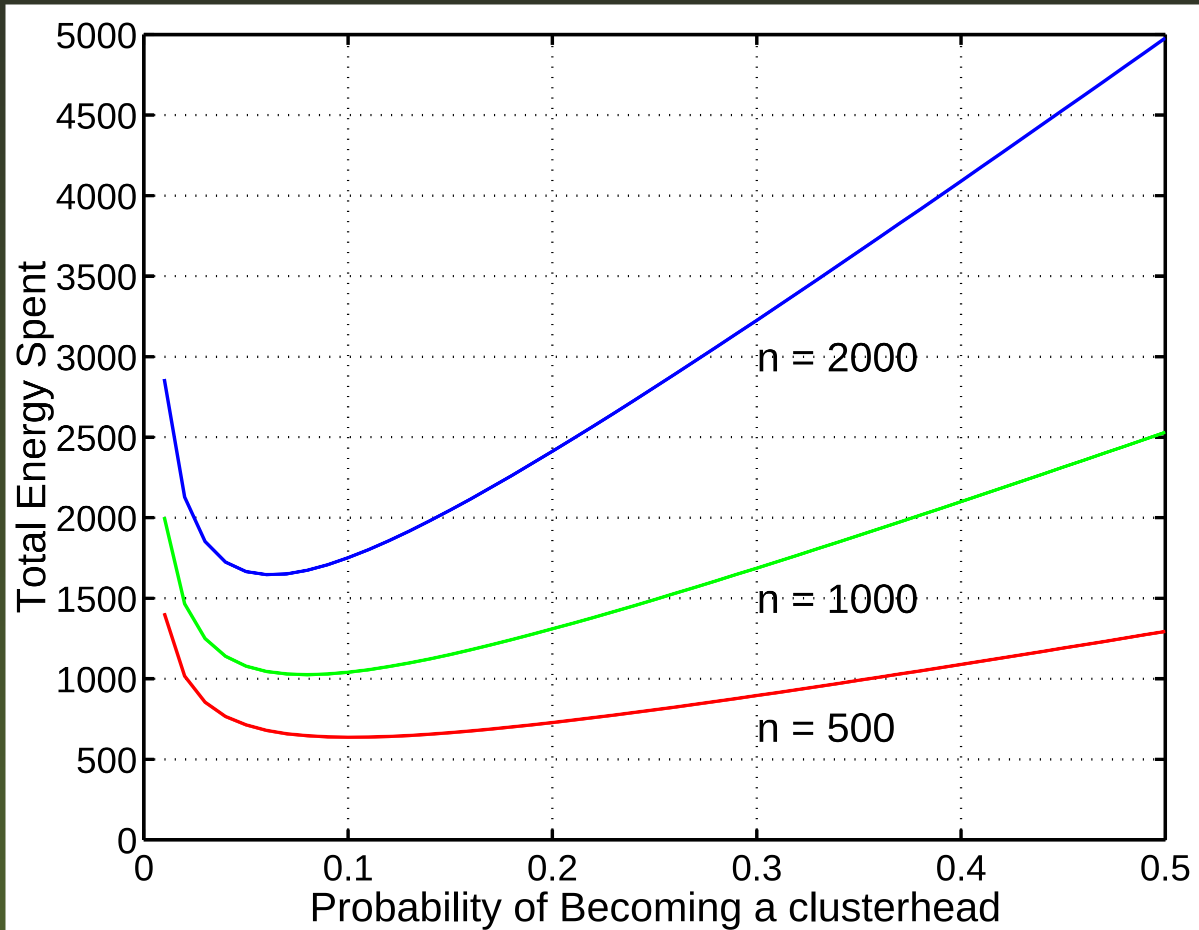

Figure 2: Optimal Probability and Total Energy Spent

Simulation Results

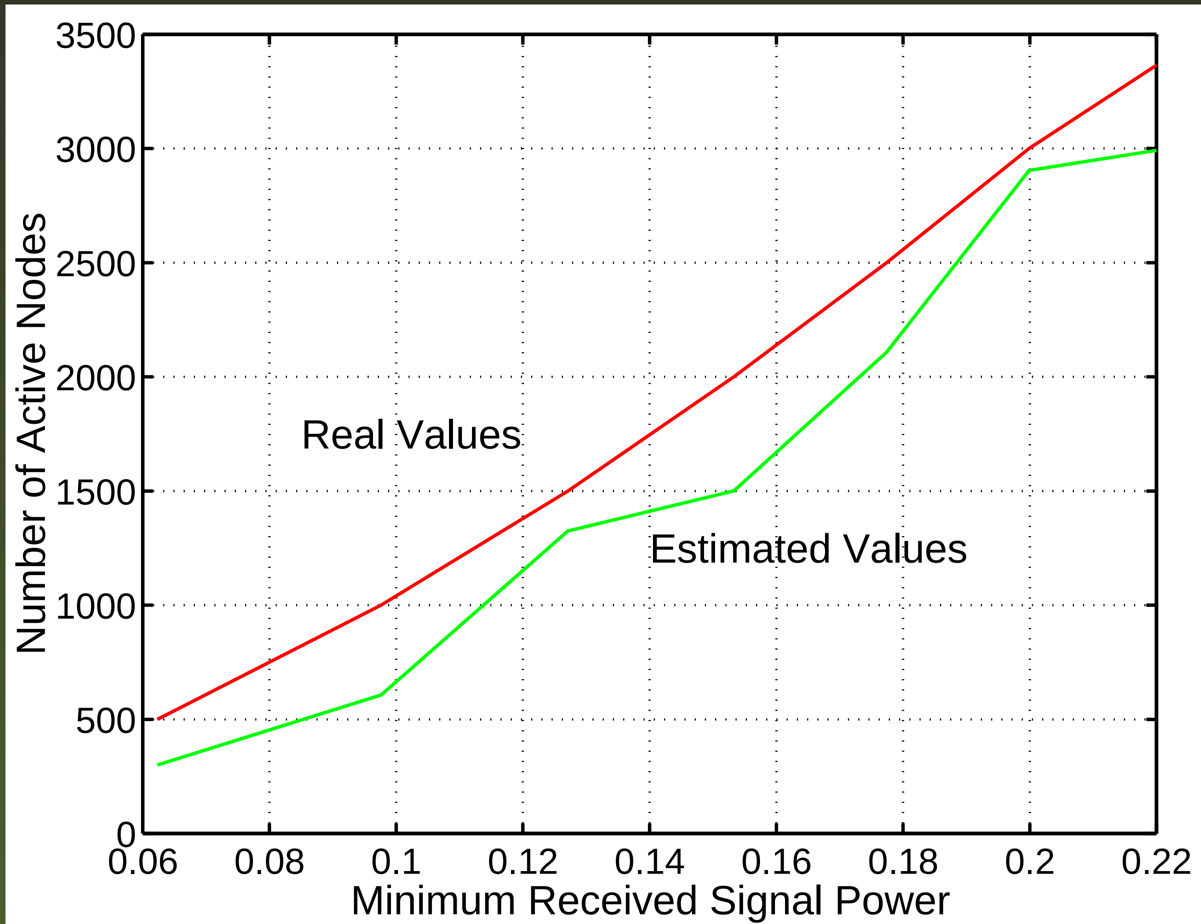

Figure 3: $\tilde{S}_{min}$ and $\hat{N}$ Relationship

Simulation Results

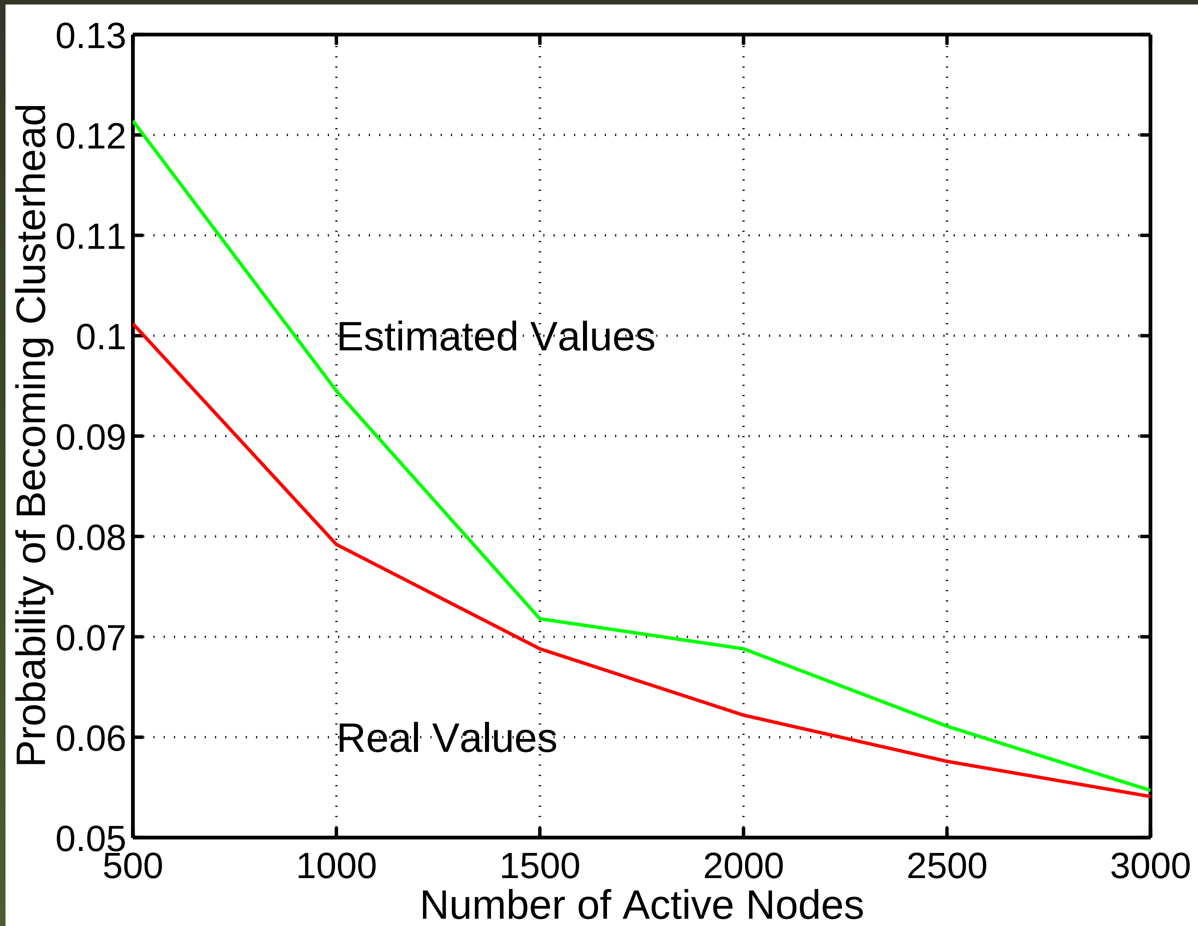

Figure 4: Optimal Probability and Number of Nodes

Simulation Results

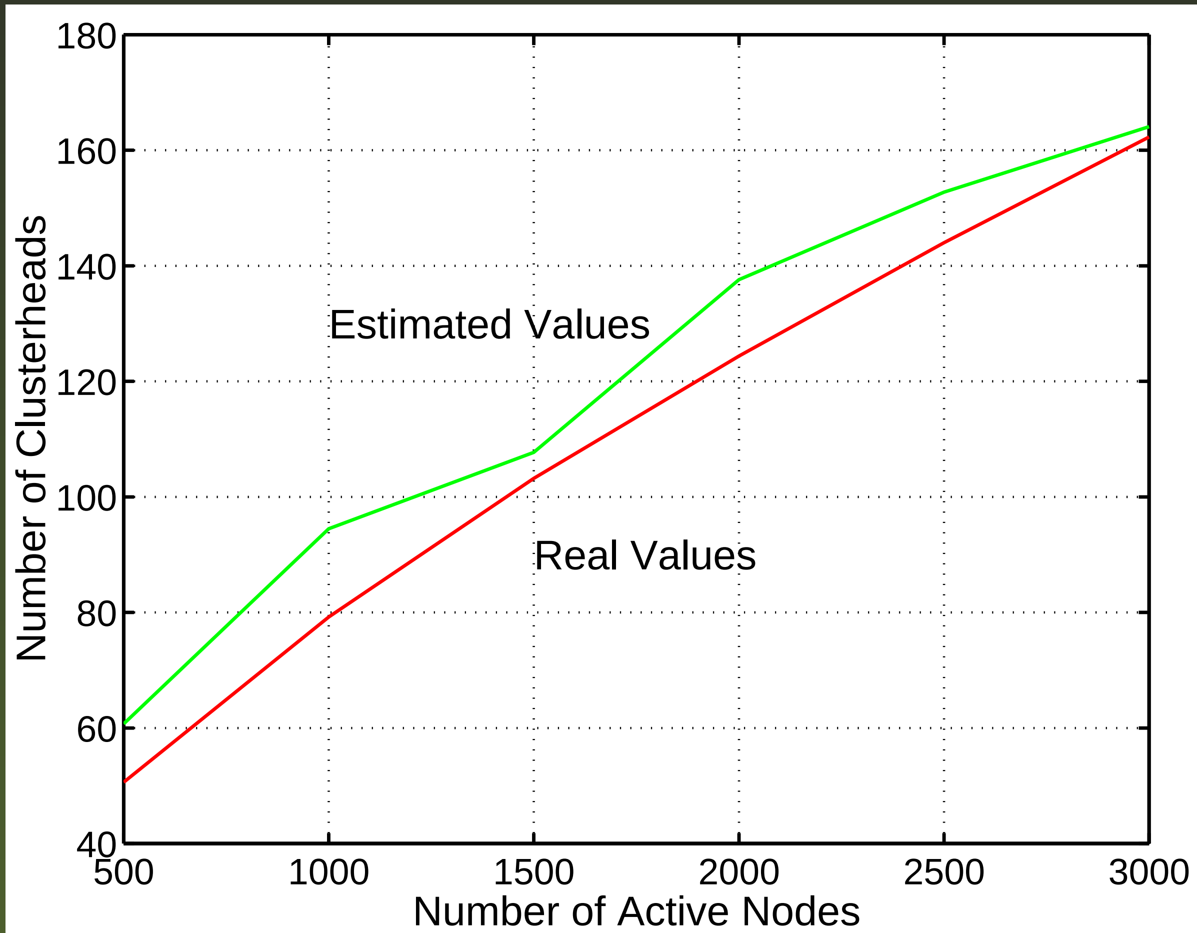

Figure 5: Numbers of Clusterheads and Nodes

Simulation Results

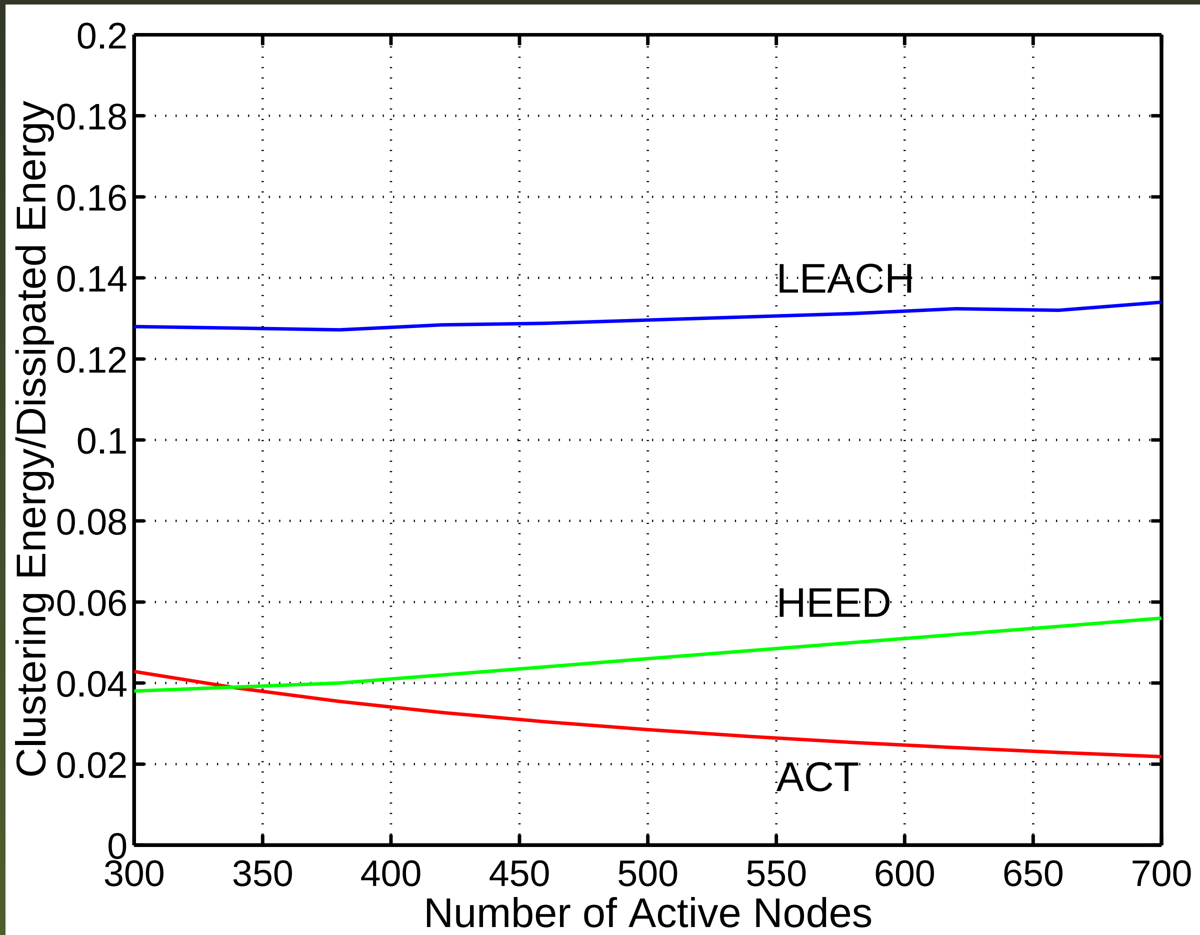

Figure 6: Ratio of Clustering Energy to Total Energy

Simulation Results

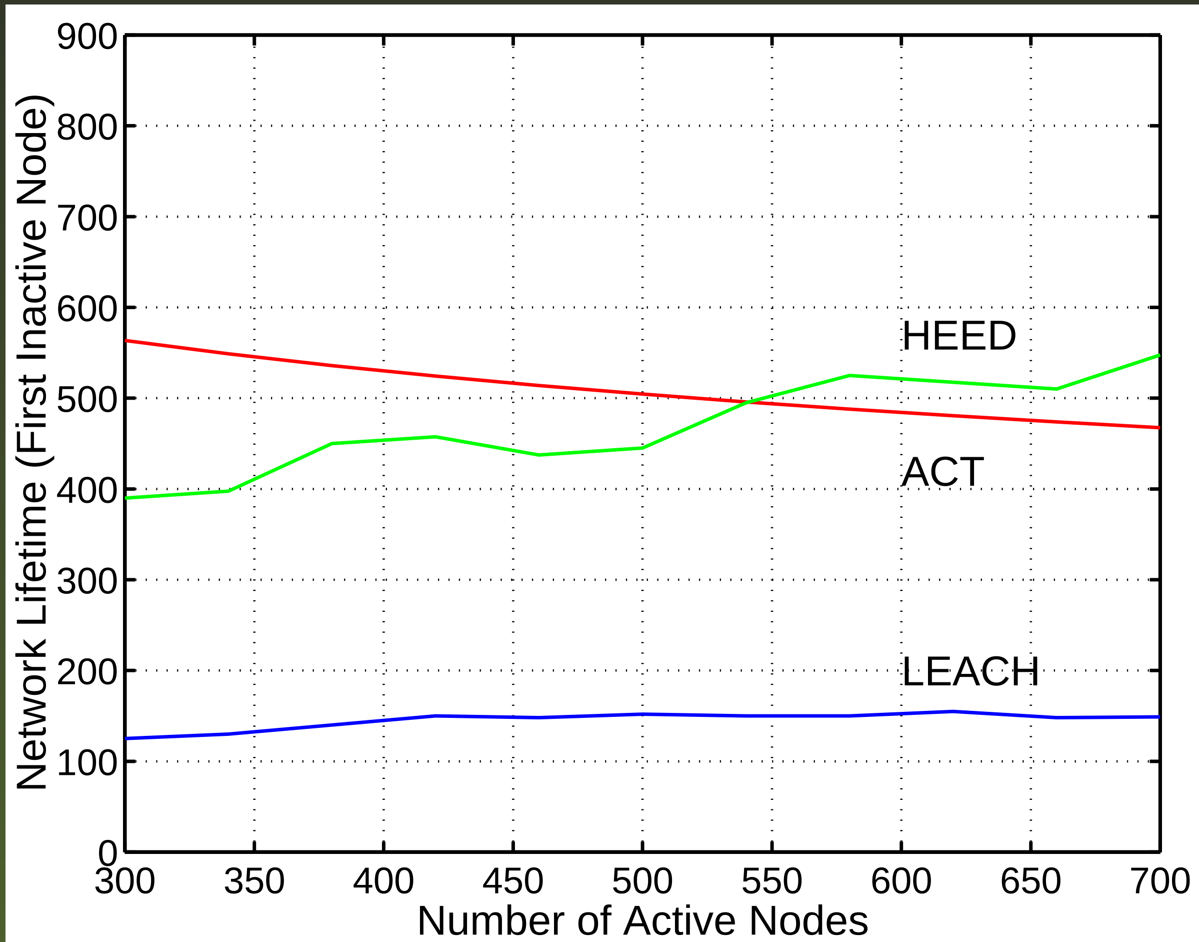

Figure 7: Network Lifetime Until First Node Inactive

Simulation Results

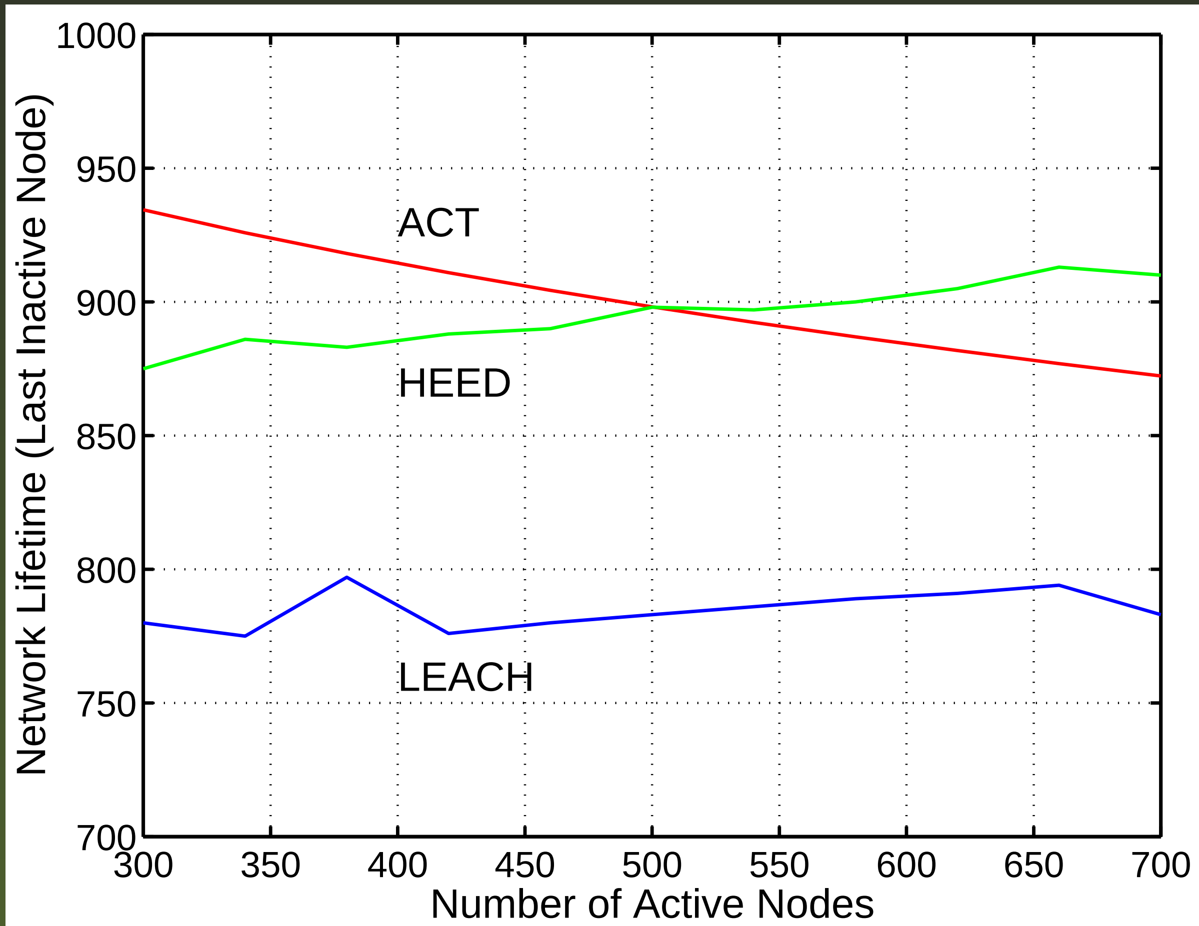

Figure 8: Network Lifetime Until Last Node Inactive

Simulation Results

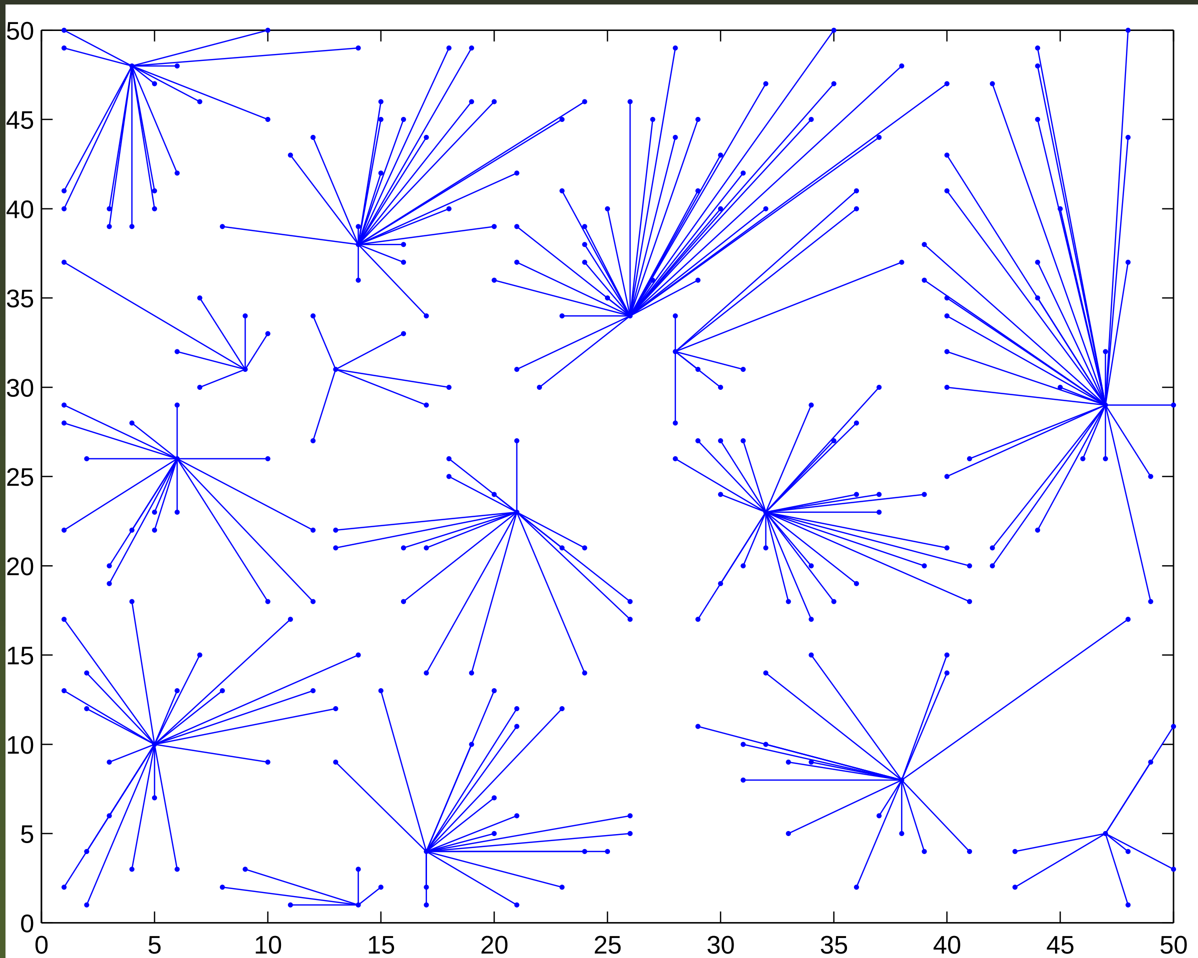

Figure 9: Adaptive Clustering

Simulation Results

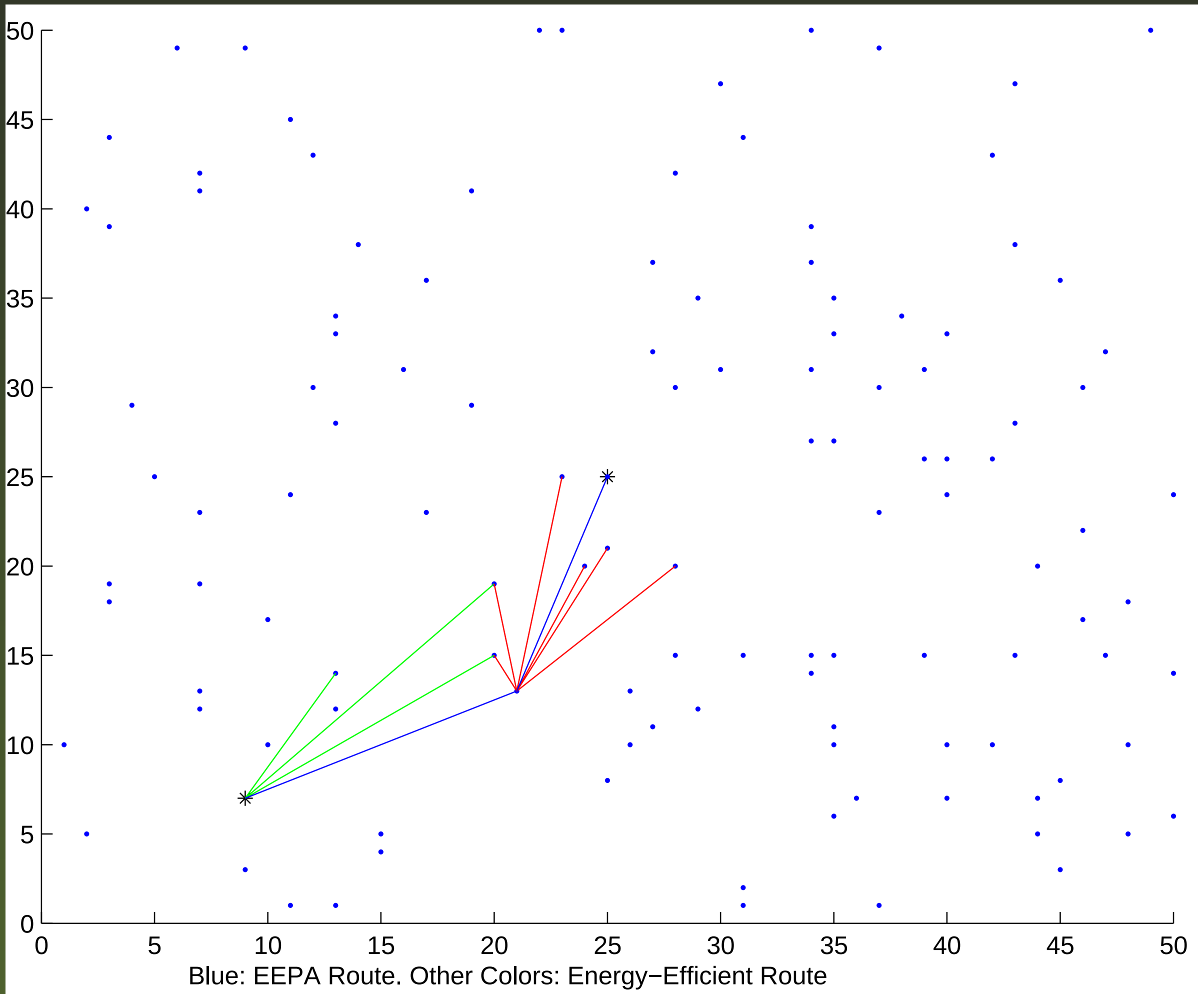

Figure 10: Energy-Efficient Routing Scenario 1

Simulation Results

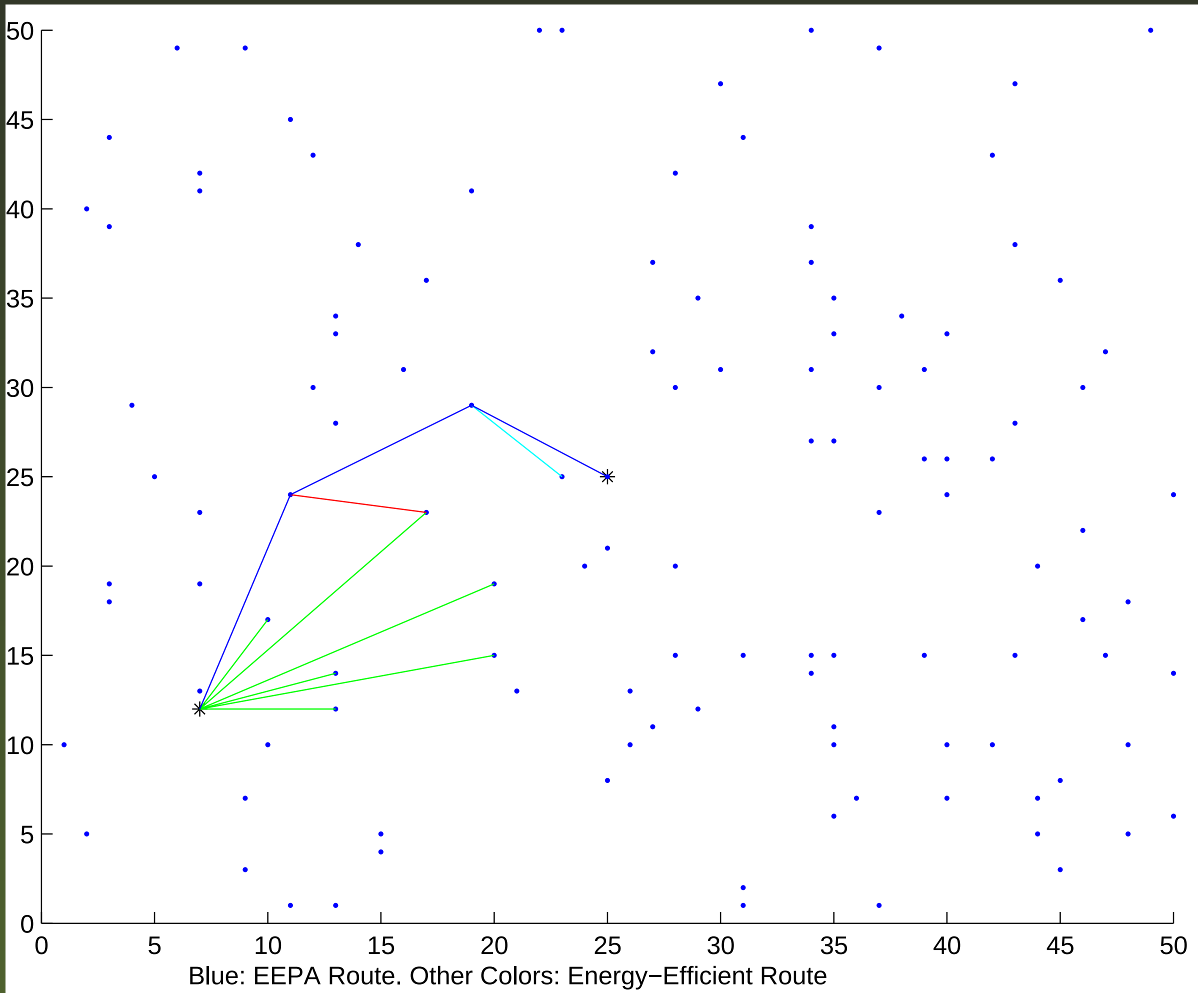

Figure 11: Energy-Efficient Routing Scenario 2

Simulation Results

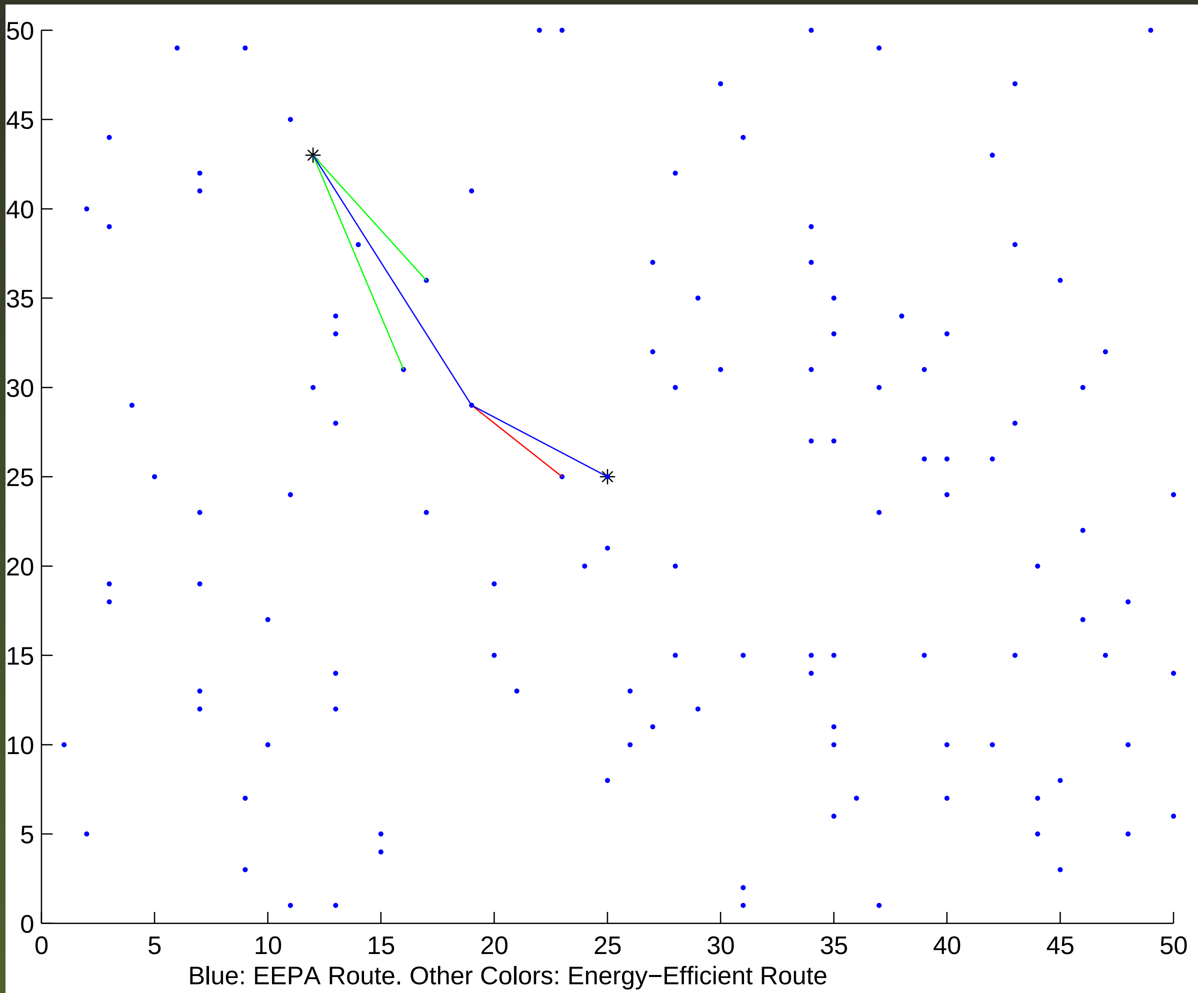

Figure 12: Energy-Efficient Routing Scenario 3

Simulation Results

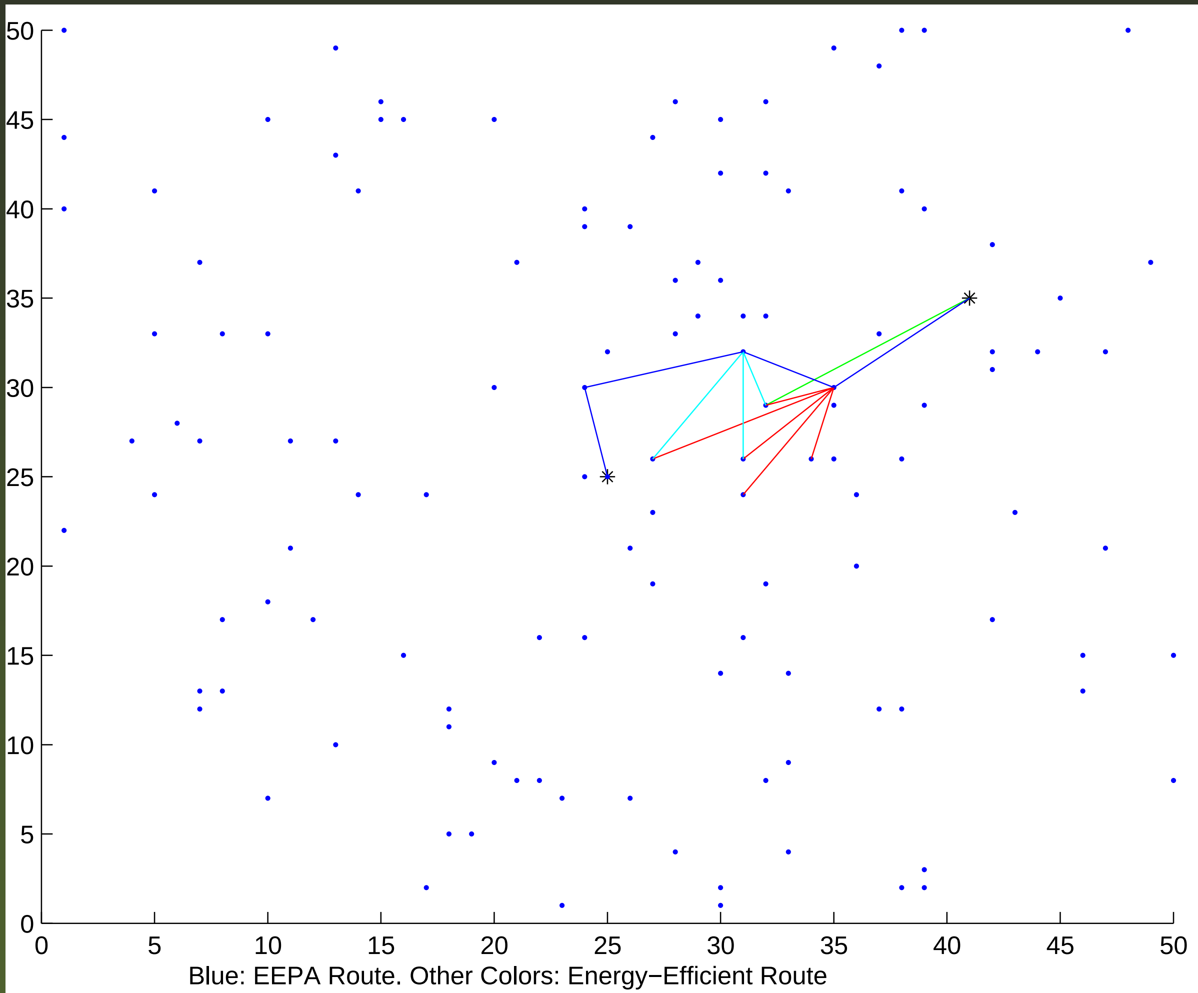

Figure 13: Energy-Efficient Routing Scenario 4