Efficient Data Transport in Wireless Sensor Networks

Haibo Zhang

The School of Computer Science

The University of Adelaide

July 31, 2008

Contents

List of Figures

List of Tables

Abstract

Providing efficient data transport is one of the uppermost objectives in the design of wire-less sensor networks (WSNs) since the primary role for each sensor is to report the sensed data to the data sink(s). This thesis focuses on designing efficient data transport schemes for WSNs in the dimensions of energy consumption and time respectively. The developed schemes can be directly applied in a number of applications such as intrusion detection, target tracking, environment monitoring, etc., and can be further extended to underwa-ter acoustic sensor networks and unmanned aerial vehicles (UAVs) networks. With the development of WSN technologies, new challenging research problems such as real-time streaming data gathering and intelligent data communication are emerging. This thesis pro-vides useful foundation for designing next-generation data transport schemes for WSNs.

Energy is the most important resource in WSNs because sensor nodes are commonly powered by small batteries, and energy is directly related to the lifetime of nodes and the network. In this thesis, energy-efficient data transport schemes are designed for two major types of WSNs: event-driven sensor networks and time-driven sensor networks. A novel on-line routing scheme called EBGR (Energy-efficient Beaconless Geographic Routing) is designed for event-driven sensor networks characterized by dynamic network topology. The main advantage of EBGR is that it can provide energy-efficient sensor-to-sink routing without any prior neighborhood knowledge. Moreover, the total energy consumption for sensor-to-sink data delivery under EBGR has an upper bound. Time-driven sensor net-works, in which all sensors periodically report the sensed data to the sink(s), have been widely used for environment monitoring applications. Unbalanced energy consumption is an inherent problem in time-driven sensor networks. An efficient data gathering scheme, called EBDG (Energy-Balanced Data Gathering), is designed to balance energy consump-tion for the goal of maximizing network lifetime. Combing all advantages of corona-based network division, mixed-routing and data aggregation, EBDG can prolong network lifetime by an order of magnitude compared with conventional schemes.

Time-efficient data transport is another critical issue in WSNs since the data generated by the sensor nodes may become outdated after a certain time interval. This thesis focuses on the problem of providing real-time data gathering in time-driven sensor networks. A

novel data gathering scheme based on random access is proposed with the objective to minimize the average duration for completing one round of data gathering. Fully localized solutions have been designed for both linear networks and tree networks. A simple data gathering protocol called RADG (Random Access Data Gathering) is designed. Simulation results show that RADG outperforms CSMA based schemes when the size of the data packets is small.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying.

Singed:

Haibo Zhang

Acknowledgments

I first thank my principal supervisor Prof. Hong Shen for guiding my research. He has been part of my research in all the stages, starting from problem definition to publication of our solutions. His breadth of knowledge and his enthusiasm for research amaze and inspire me.

I am grateful to Dr. Cheryl Pope, my Ph.D co-supervisor. Cheryl and I have been working during the final days before the thesis submission deadline. Her advice and feedback about my research have greatly enhanced and strengthened the work. I thank her for the countless hours she has spent with me, discussing everything from research to career choices.

I also would like to give thanks to Prof. Yasuo Tan in JAIST for acting as my temporary supervisor before I transferred to the University of Adelaide.

I am deeply indebted to school of computer science at the University of Adelaide, and the 21th COE program in Japan Advanced Institute of Science and Technology (JAIST) for granting me scholarships, which enable me to pursue Ph.D study. Thanks them also for supporting me to present our work at some international conferences.

I would like to thank my friends and colleagues, especially Zonghua Zhang, Yingpeng Sang, Hui Tian, Shihong Xu and Yidong Li, for their insightful discussions and warm friendship.

I owe a special debt of gratitude to my parents and family. They have, more than anyone else, been the reason I have been able to get this far. Words cannot express my gratitude to my parents, who give me their support and love from across the seas. My wife, Yawen Chen, gives me her selfless support and love that make me want to excel.

Finally, I would like to thank all the people who either directly or indirectly provide me knowledge, experience, and support.

Published Papers

1. Haibo Zhang and Hong Shen, Balancing Energy Consumption to Maximize Network Lifetime in Data Gathering Sensor Networks, accepted for publication at IEEE Transactions on Parallel and Distributed Systems.

2. Haibo Zhang and Hong Shen, Brief Announcement: Balancing Energy Consumption for Data Gathering in Sensor Networks, the 27th Annual ACM Symposium on Principles of Distributed Computing (PODC), 2008.

3. Haibo Zhang and Hong Shen, EEGR: Energy Efficient Geographic Routing for Wireless Sensor Networks, the 36th International Conference on Parallel Processing (ICPP), 2007, China, #67.

4. Haibo Zhang, Hong Shen and Yasuo Tan, Optimal Energy Balanced Data Gathering in Wireless Sensor Networks, the 21st IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2007, USA, pp. 1-10.

5. Haibo Zhang, Hong Shen and Hui Tian, Reliable and Real-time Data Gathering for Multi-hop Linear Wireless Sensor Networks, the 1st International Conference on Wireless Algorithms, Systems and Applications (WASA), 2006, pp. 151-162.

6. Haibo Zhang and Hong Shen, Distributed Tuning Attempt Probability for Data Gathering in Random Access Wireless Sensor Networks, IEEE 20th International Conference on Advanced Information Networking and Applications (AINA), 2006, Vienna, Austria, pp. 643-648.

7. Haibo Zhang and Hong Shen, Haibin Kan, Reliability-Latency Tradeoffs for Data Gathering in Random-Access Wireless Sensor Networks, the 4th International Conference on Grid and Cooperative Computing (GCC), 2005, pp. 701-712.

Papers Under Review

1. Haibo Zhang and Hong Shen, Energy-efficient Beaconless Geographic Routing in Wireless Sensor Networks, *under second-round review*, IEEE Transactions on Paral-

lel and Distributed Systems.

2. Haibo Zhang and Hong Shen, Exploring Random Access Techniques for Data Gathering in Wireless Sensor Networks, Submitted to Wireless Communications and Mobile Computing in April 2008 .

Chapter 1

Introduction

Recent advances in sensor technology and wireless communication have motivated significant research interest in wireless sensor networks because of their promising potential to support a broad range of applications. Generally speaking, a wireless sensor network is composed of a large number of low-cost sensor nodes which are small in size and powered by inexpensive batteries. Each sensor node is composed of at least four components: sensing, computation, wireless communication and power unit. The sensing and computation components provide the functions for data acquisition and local data processing, and the power unit provides energy for running the other three components, whereas the wireless communication component enables all sensor nodes to self-organize into a wireless network and collaborate with each other to perform a given task.

WSNs have many advantages over conventional systems designed for monitoring or data collection applications. Ease of deployment is one of these attractive advantages. In conventional systems, expensive sensors often need to be placed in exact locations and wired to the end-users, which makes the deployment and configuration of sensors and wire extremely costly and awkward. However, in WSNs, the sensor nodes are cheap and small in size. Moreover, wireless communication eliminates the need for a fixed infrastructure. These characteristics enable rapid deployment of a large number of sensor nodes even in inaccessible terrains. For example, sensor nodes can be dropped from a plane flying over a remote or dangerous area. Another attractive feature of WSNs is self-organization. Instead of manual configuration, sensor nodes can self-organize into a wireless network

after deployment. This feature can support incremental node deployment in which new nodes can easily be assembled into the network. Moreover, this feature also provides the capabilities for sensor networks to dynamically adapt to device failures and changes in task.

The above attractive features ensure a wide range of applications for WSNs. For example, characteristics such as rapid deployment, self-organization, and fault tolerance make wireless sensor networks attractive for environment monitoring [6][9][7], target tracking [1], seismic detection, military surveillance, nuclear reactor control, fire detection, inventory tracking, smart spaces etc. In health [17], sensor nodes can be deployed to monitor patients and assist disabled patients. Some other commercial applications include indoor climate control, home automation, monitoring product quality, intelligent alarms, etc. In fact, because of the pervasive nature of micro-sensors, sensor networks have the potential to revolutionize the way we understand and construct complex physical systems.

1.1 Characteristics of WSNs

Wireless sensor networks have several special characteristics that make them different from other wireless networks like mobile ad hoc networks and cellular networks. These characteristics have posed great challenges for the design and implementation of WSNs.

First, because wireless sensor networks are generally composed of at least hundreds of sensor nodes, traditional IP-based protocols are not well suited for WSNs due to the high cost of assigning and maintaining unique IDs for such a large number of nodes. Second, sensor nodes are constrained in terms of energy, computation and storage capabilities. For example, Mica motes [4] are powered by AA batteries and equipped with Atmel AtMegal128L 8-bit microcontroller running at 8MHZ. Moreover, Mica motes have only 4KBytes of SRAM and 128 KBytes of Flash ROM. These constraints require that resources in WSNs should be carefully managed. Third, each sensor node has limited transmission range. For example, the maximum transmission range for Mica motes is only about 100 feet [4]. Therefore, data transport in WSNs is usually operated in a multi-hop manner. Fourth, unlike typical communication networks, wireless sensor networks are characterized by many-to-one data communication pattern in which multiple sources send their data to a particular sink. This characteristic may lead to some serious problems such as un-

balanced energy consumption, aggregated channel contention, etc. Fifth, sensor networks are usually densely deployed, and neighboring nodes may have overlapped sensing range. Thus data generated in wireless sensor networks is strongly correlated. Sixth, sensor networks are application-specific, i.e., design requirements of a sensor network change with applications. For example, the challenging problem of low-latency data delivery in tactical surveillance applications is different from that required for a periodic weather-monitoring task. Finally, wireless communications in WSNs are prone to failures. Data transport may experience serious data loss due to packet collision, environmental interference, device failure, etc.

1.2 Challenges for Data Transport in WSNs

Despite the innumerable applications of WSNs, there are some critical issues that need to be solved before WSNs can be widely used in practice. Achieving efficient data transport is one of the uppermost objectives in designing and implementing sensor networks since the primary role for each sensor is to deliver the sensed data to the sink(s). Moreover, WSNs are application-specific. Different applications may have different requirements on data transport such as energy consumption, transmission delay, transmission reliability, etc. However, due to the inherent characteristics, providing efficient data transport in WSNs is not as easy as in wired networks. The main challenges and issues in designing data transport schemes for WSNs are summarized as follows:

- *Energy Efficiency:* Energy is one of the most critical resources in WSNs, and energy consumption should be carefully managed so that network lifetime can be maximized. Wireless communication is the major source of energy consumption in wireless sensor networks. Thus designing energy efficient data transport schemes plays an important role in terms of improving energy efficiency and prolonging network lifetime.

- *Balanced Energy Consumption:* Unbalanced energy consumption is an inherent problem for WSNs characterized by multi-hop routing and many-to-one data communication pattern. Sensor nodes close to the sink experience much more traffic than

nodes far away from the sink since they should not only transmit the data generated by themselves but also need to relay the data they receive. These sensor nodes may be overused and die out earlier, resulting in network collapse although there may be still significant amount of energy in other sensors. Therefore, balancing energy consumption should be taken into account in designing data transport schemes for WSNs.

- *Scalability:* Wireless sensor networks are usually densely deployed to achieve full coverage and high fault-tolerance. The number of nodes in a wireless sensor network may be several orders of magnitude larger than that in wireless ad hoc networks [13]. This requires that data transport schemes designed for sensor networks must be scalable enough so that their performance is not affected by the number of sensor nodes. Therefore, localized data delivery algorithms, in which each node makes decisions on forwarding messages based solely on its neighborhood information, are particularly attractive to wireless sensor networks.

- *Network Dynamics:* Wireless sensor networks may be highly dynamic due to the following reasons. Firstly, sleep management schemes are commonly employed to save energy due to the low duty-cycles of WSNs. By letting each node periodically switch between *active* and *sleep* states, the energy wasted by idle listening can be significantly reduced. However, periodic sleeping can lead to frequently changing network topology. Secondly, in some applications, mobile sensor nodes are employed to improve energy efficiency or to balance energy consumption. Node mobility can also result in unpredictable and frequent topological changes. Thirdly, sensor nodes are prone to failures, and wireless communication quality can vary dramatically over time. These factors also make the network topology change dynamically.

- *Communication overhead:* Due to the limited battery power and the limited bandwidth, the communication overhead in WSNs must be minimized. Data delivery schemes that either use flooding to detect routes or try to maintain routing information are not well suited for sensor networks since both flooding and route maintenance can bring too much communication overhead. Moreover, the communication

overhead brought by route maintenance may exacerbate the extent of channel contention, resulting in too much degradation in terms of network performance.

- *Data Aggregation:* The data generated by sensor nodes may contain a lot of redundant data since sensor nodes are commonly densely deployed. Data aggregation is the combination of data from different sources according to a certain aggregation function, eliminating redundancy, minimizing the number of transmissions and thus saving energy.

- *Quality of Service:* Wireless sensor networks are application-specific. In some applications such as fire detection, the sensed data should be reported to the sink reliably and as quickly as possible so that some measures can be taken in time to deal with such events. However, in other applications, conservation of energy, which is directly related to network lifetime, is more important than data delivery reliability and latency. This is because that the network will lose coverage when some nodes run out of energy. Therefore, data transport schemes should provide versatile QoS to cater for different applications.

1.3   Issues Addressed in This Thesis

This thesis focuses on designing efficient data transport schemes for WSNs in the dimensions of energy consumption and time respectively.

The first part of this thesis focuses on designing energy-efficient data transport schemes for two major types of WSNs: event-driven sensor networks and time-driven sensor networks.

Event-driven sensor networks have been widely used for event-detection applications such as fire detection, intrusion detection [35], target tracking [1], etc. In these applications, a sensor node generates data and reports data to the sink(s) only when an event has happened within its sensing range. Thus the major communication paradigm in event-driven sensor networks is sensor-to-sink event reporting. Another special characteristic of event-driven sensor networks is dynamic network topology caused by node mobility, node activeness, etc. The characteristic of dynamic network topology, coupled with the unique

features of WSNs, makes routing in highly dynamic WSNs extremely challenging. Most existing routing schemes are based on maintaining routing table or neighborhood information. These routing schemes are not suitable for WSNs with dynamic network topology due to the high maintenance cost. The first issue addressed in this thesis is to design fully stateless, energy-efficient on-line geographic routing scheme, in which each node forwards packets without the help of any neighborhood knowledge, for WSNs with dynamic network topology.

Time-driven sensor networks have been widely used for data gathering applications such as habitat monitoring [26], volcanic-monitoring [6], etc. In these applications, all sensor nodes periodically sense the monitored environment and report the sensed data to the sink(s). Thus many-to-one data communication pattern is the major communication paradigm in time-driven data gathering sensor networks. Moreover, data transport in WSNs is commonly operated in a multi-hop manner due to the limited transmission range of each sensor node. The characteristics of multi-hop routing and many-to-one data communication pattern lead to a serious problem, called unbalanced energy consumption, in time-driven data gathering sensor networks. The sensors close to the sink have a high load of packets relay and may run out of energy quickly, resulting in significantly shortened network lifetime. The second issue addressed in this thesis is to design efficient energy-balanced data gathering scheme with the objective to maximize network lifetime.

Time-efficient data transport is another critical issue in WSNs. The second part of this thesis focuses on designing time-efficient data transport schemes for WSNs. In the past few decades, extensive work has been done on providing real-time data delivery in event-driven sensor networks [16][114][89][49][44][36]. Compared with data transport in even-driven sensor networks, providing real-time data gathering in time-driven data gathering sensor networks is much more challenging. This is because all sensor nodes must collaborate with each other to guarantee that the sensed data can be delivered to the sink(s) in time since the generated data may become outdated after a certain time interval. Moreover, a large amount of data may be generated in a very short period and transmitted through a shared radio channel during data gathering. The large burst of data packets may lead to high degree of channel contention, high probability of packet collision, and long delay of packet delivery, which makes the data gathering operation extremely challenging. The third issue

addressed in this thesis is to design real-time data gathering scheme with the objective to minimize the duration for completing each round of data gathering.

1.4   Main Contributions

The main contributions of this thesis are summarized as follows:

- An Energy-efficient beaconless Geographic Routing scheme, called EBGR, is designed for WSNs in which network topology changes frequently due to node mobility, node sleeping, node failures, etc. EBGR has several advantages over existing routing schemes. First, EBGR is fully stateless since all nodes are not required to maintain any neighborhood information, neither about their positions nor even about their existence. Second, EBGR can provide fully localized, loop-free, on-line sensor-to-sink routing at a low communication overhead. Third, EBGR is energy-efficient. The energy consumption for sensor-to-sink data transport can be theoretically bounded under EBGR. Moreover, it is demonstrated that the expected energy consumption approaches to the lower bound with the increase of node deployment density. To deal with the unreliable data communications, EBGR is also extended to lossy sensor networks. Extensive simulations have been performed to evaluate the proposed schemes, and simulation results show that the proposed schemes significantly outperform existing protocols in wireless sensor networks with highly dynamic network topology.

- An energy-balanced data gathering scheme is proposed to balance energy consumption and maximize network lifetime for data gathering sensor networks with uniform node deployment. The energy consumption balancing problem is formulated as an optimal data distribution problem by combining the ideas of corona-based network division and mixed-routing strategy together with data aggregation. A simple localized zone-based routing scheme is designed to balance energy consumption among nodes within each corona. An algorithm with time complexity $O(n)$ (n is the number of coronas) is designed to solve the transmitting data distribution problem aimed at balancing energy depletion among nodes in different coronas. The approach for

computing the optimal number of coronas in terms of maximizing network lifetime is also presented. Based on the mathematical model, an energy-balanced data gathering protocol (EBDG) is designed and the solution for extending EBDG to large-scale data gathering sensor networks is also presented. Simulation results demonstrate that EBDG can improve system lifetime by an order of magnitude compared with conventional multi-hop transmission schemes, direct transmission schemes and cluster-heads rotation schemes.

- A time-efficient data gathering scheme based on random channel access techniques is proposed for data gathering in wireless sensor networks with light data load. The data gathering problem is formulated as an optimization problem with the objective to minimize the expected duration for one round of data gathering on condition that each link can provide guaranteed transmission reliability. Fully localized solutions have been proposed for both linear networks and tree networks. A simple and scalable protocol, called RADG (Random Access Data Gathering) is designed for data gathering in WSNs. The performance of RADG is evaluated through extensive simulations by comparing with CSMA-based protocols. Simulation results show that RADG can significantly outperform handshaking (i.e. RTS/CTS) schemes when data packets are small.

1.5 Thesis Outline

The rest of this thesis is organized as follows:

Chapter 2 reviews related work in the areas of energy efficient routing, geographic routing, energy consumption balancing, time-efficient data transport and random channel access techniques in wireless sensor networks.

Chapter 3 focuses on addressing the problem of providing energy-efficient beaconless geographic routing for wireless sensor networks in which the network topology frequently changes over time. A novel energy-efficient beaconless geographic routing protocol called EBGR is presented first. Extensive theoretical analysis has been conducted to evaluate the performance of EBGR. First, it is proven that EBGR is loop-free in greedy forwarding

mode, and demonstrated that EBGR can provide guaranteed packet delivery as long as the network is connected. Second, the lower bound and upper bound on hop-count for sensor-to-sink packet delivery under EBGR are established, assuming no fails in greedy forwarding. To demonstrate the energy efficiency of EBGR, the upper bound on energy consumption as well as the expected energy consumption for sensor-to-sink data delivery is presented, assuming no packet loss and no failures in greedy forwarding. To deal with the unreliable data communication, EBGR is also extended to lossy sensor networks. Finally, extensive simulations have been conducted to evaluate the performance of the proposed scheme in three types of scenarios: mobility scenarios, random sleeping scenarios and high-variant link quality scenarios.

Chapter 4 concentrates on solving the problems of balancing energy consumption and maximizing network lifetime for data gathering WSNs. Based on coronas-based network division and mixed-routing techniques, the balancing energy consumption problem is divided into two sub-problems: Intra-Corona Energy Consumption Balancing (Intra-CECB) and Inter-Corona Energy Consumption Balancing (Inter-CECB). First, a localized zone-based routing scheme is proposed to solve the Intra-CECB problem. Then an off-line centralized algorithm is proposed to solve the Inter-CECB. Based on this model, an energy-balanced data gathering protocol (EBDG) is designed and the solution for extending EBDG to large-scale data gathering sensor networks is also presented. Finally, extensive simulations have been done to evaluate the performance of EBDG by comparing it with conventional multi-hop transmission schemes, direct transmission schemes and cluster-heads rotation schemes.

Chapter 5 addresses the problem of designing time-efficient data gathering scheme for WSNs with light data load. First, a simple example is presented to illustrate the advantage of random access techniques over handshaking (i.e. RTS/CTS) schemes for data gathering in sensor networks with small data packets. Then a novel scheme based on random access is proposed for data gathering in WSNs. In this scheme, it is assumed that time is slotted. In each time slot, each node v_i works in *transmitting* state with probability α_i, and works in *receiving* state with probability $1 - \alpha_i$, where α_i is referred to as the attempt probability for node i. The objective is to compute the optimal attempt probability for each node so that the expected data gathering duration can be minimized on condition that each link in

the data gathering structure can provide guaranteed packet delivery reliability. Localized solutions for both linear networks and tree networks have been designed. Based on this scheme, a data gathering protocol called RADG is designed. Finally, extensive simulations have been done to evaluate the performance of RADG by comparing with CSMA-based protocols.

Chapter 6 concludes this thesis and discusses the directions for future work.

Chapter 2

Literature Review

This chapter gives a detailed summary of previous work, which includes energy-aware routing, geographic routing, energy consumption balancing techniques, time-efficient data transport, and random channel access techniques.

2.1   Energy-aware Routing

Energy is one of the most important resources in wireless ad-hoc/sensor networks since the devices in such networks are commonly powered by batteries. In wireless networks, wireless communication is a major source of energy consumption. Thus it is essential to design energy-aware routing schemes to improve the energy efficiency of data transportation. In the past few decades, energy-aware routing has received much attention in wireless ad-hoc/sensor networks. Singh *et al.* in [105] introduced the concept of energy-aware routing and proposed five metrics, i.e., *minimize energy consumed/packet*, *maximize time to network partition*, *minimize variance in node power levels*, *minimize cost/packet*, and *minimize maximum node cost*, for routing in mobile ad hoc networks. Experimental results show that using these metrics in a shortest-cost routing algorithm reduces the cost of routing packets by $5 - 30\%$ over shortest-hop routing. In [76], three distributed power-aware routing algorithms (i.e. *distributed minimal power algorithm*, *distributed Max-Min algorithm* and *distributed Max-Min zP_{min}*) were designed to achieve energy-efficient packet delivery in WSNs. The *distributed minimal power algorithm* is a distributed version of

Dijkstra's algorithm that is guaranteed to be a minimal-power routing path algorithm by giving messages variable propagation delays. The *distributed Max-Min algorithm* is an enhanced version of the distributed Bellman-Ford algorithm. In this algorithm, each packet is delivered along the max-min path which is defined as the route from a node to the sink on which the minimal residual power of the nodes is maximized among all the routes. The *distributed Max-Min zP_{min}* algorithm is designed to obtain a tradeoff between minimizing the total power consumption and maximizing the minimal residual power of the network. Simulation results show that the *distributed Max-Min zP_{min}* algorithm can significantly extend network lifetime when sensor nodes are sparsely deployed. In [55], Li *et al* investigated the problem of designing a routing scheme to minimize the maximum energy utilization of a multi-hop wireless network with weak assumption of the traffic pattern and without ongoing collection of network information. Polynomial size linear programming models were developed to design such a routing scheme. An overview of energy-aware routing protocols can be found in [14].

The objective of energy-aware routing is to prolong network lifetime. Thus many routing schemes have been proposed aiming at maximizing network lifetime [29][66][110][56]. Chang and Tassiulas in [29] considered the problem of choosing routes between a set of source nodes and a set of sink nodes of an ad-hoc network so that the time until the first node uses up its energy is maximized. Two heuristic distributed algorithms were proposed to solve this problem. An identical energy efficient lifetime maximization problem has been studied in [66]. However the authors did not develop any distributed algorithms to solve the lifetime maximization routing problem. In [56], network lifetime maximization was formulated as a rate allocation optimization problem. A solution which uses the so-called *Lexicographic Max-Min (LMM) criterion* was proposed to maximize the bit rates for all the nodes until one or more nodes reach their energy limit. In [57], the authors addressed the problem of maximizing network lifetime of wireless sensor networks through optimal single-session flow routing. In [93], the authors studied the problem of lifetime maximization in interference-limited wireless sensor networks through cross-layer design techniques. In [104], several distributed algorithms were designed to maximize the lifetime of wireless sensor networks. In [19], the authors derived the upper bound on the lifetime of wireless sensor networks through optimal role assignment.

A wireless sensor network may be composed of hundreds or even thousands of nodes, and this feature requires that routing protocols for sensor networks must be highly scalable. Localized routing schemes, in which each node makes routing decisions only based on 1-hop or 2-hop neighbor information, are attractive to large-scale wireless sensor networks due to their simplicity and scalability. Stojmenovic *et al.* in [110] discussed the importance of designing localized power-aware routing protocols and proposed three fully localized routing algorithms to minimize total energy consumption and maximize network lifetime. A framework based on optimizing a cost over progress ratio was further proposed in [108] for designing energy-aware routing schemes in wireless networks. In [47], a localized probabilistic algorithm based on dominating set techniques was proposed to distribute information in wireless sensor networks. In [84], a novel analytical framework was introduced to analyze the relationship between energy efficiency and range of topology knowledge. A new forwarding scheme, called Partial Topology Knowledge Forwarding, was introduced, and the problem of determining the optimal knowledge range for each node to make energy efficient geographical routing decisions was tackled by Integer Linear Programming. It was demonstrated that the problem is intrinsically localized, i.e., a limited knowledge of the topology is sufficient to make energy efficient forwarding decisions. In [32], a localized algorithm (called AFA) was designed based on a novel aggregate fairness model. AFA is designed to work with any routing protocol. In particular, it allows the packets from a data source to follow an arbitrary set of forwarding paths to the base stations.

2.2   Geographic Routing

Geographic routing, in which each node forwards packets only based on the locations of itself, its directed neighbors and the destination, is particularly attractive to power-constrained sensor networks. The localized nature of geographic routing eliminates the overhead brought by route establishment and maintenance, giving the advantages of modest memory requirement at each node and high scalability in large distributed applications. For example, the MFR protocol proposed in [111] is one of the earliest geographic routing algorithms. In MFR, each node forwards its packets to the neighbor that has the maximum progress. In 'greedy' routing [39], each node forwards the packets to its neighbor closest

to the destination. An overview of geographic routing can be found in [102].

In geographic routing schemes, each node is required to have the knowledge of its geographic location. The location information can be obtained by equipping each sensor node with a GPS-capable antenna. However, providing each sensor with localization hardware (e.g., GPS) is expensive in terms of both cost and energy consumption. A reasonable solution to the localization problem is to allow some nodes (called seeds) to have their location information at all times, and allow other nodes to infer their locations by exchanging information with seeds. In [48], a range-free localization algorithm called APIT is proposed. In APIT, all possible triangles of the seeds are formed, and the location of a node is the center of intersection region of all triangles. In [99], a localization algorithm based on range measurements between pairs of nodes and the priori coordinates of sparsely located anchor nodes was proposed. Recently, Ananth *et al.* in [94] addressed the problem of retaining the benefits of geographic routing in the absence of location information by assigning virtual coordinates to each node. In [70], geographic routing in the presence of location errors was studied. Location accuracy is an important issue that affects the performance of geographic routing. Many approaches aiming at improving location accuracy have been proposed [53][95][87]. In [95], a localization service component was designed for reliably localizing wireless sensor networks in environments conducive to ranging errors by using a custom hardware-software solution. The ranging solution improves previous work, extending the practical measurement range threefold (20-30m) while maintaining a distance-invariant median measurement error of about 1% of maximum range (33cm).

One of the key challenges in geographic routing is how to deal with dead-ends, where greedy routing fails because a node has no neighbor closer to the destination. A variety of methods have been proposed to deal with this problem. The idea of employing routing in planar graphs to provide guaranteed packet delivery was first proposed in [92]. A well-known protocol called GPSR [24] implements the same idea to deal with routing dead-ends by combining greedy forwarding with perimeter routing. To guarantee delivery, many existing geographic routing algorithms [18][23][40][73] switch between the greedy forwarding mode and the recovery mode depending on the network topology. In GOAFR+[67], the authors proposed a method that is both asymptotically worst case optimal and average case efficient. Recently, Xing *et al.* in [120] demonstrated that greedy geographic

routing is good enough for sensing-covered sensor networks, and sophisticated techniques specially designed to handle routing dead-ends are not necessary in sensing-covered sensor networks.

Recently, some attention has been paid to improve energy-efficiency for geographic routing in sensor networks. In [118], a protocol called GPER was proposed to provide power-efficient geographic routing in WSNs. In GPER, each node first establishes a sub-destination within its maximum radio range and then employs a shortest-path algorithm to compute the minimum-energy path from itself to the sub-destination. Routing metrics based on normalized advance have been designed in [69][72]. Based on a realistic physical layer model, the $PRR \times DIST$ routing metric was first introduced in [69] to deal with the unreliable communication in wireless ad hoc networks. Seada *et al.* [103] addressed the *weak-link* problem and studied the energy and reliability trade-offs pertaining to geographic forwarding in lossy sensor networks under a realistic packet loss model.

In most existing routing schemes, nodes are required to maintain neighborhood information for making decisions on packet forwarding. However, in many applications, network topology changes frequently due to node periodic sleeping [27][31][34][122] and node mobility [107][112][81]. To maintain accurate neighborhood information, each node must broadcast a hello message to notify its neighbors whenever its state changes, leading to additional communication overhead and energy wastage. Recently, some work has been done to design beaconless routing algorithms to avoid exchanging topology and routing information. Beaconless geographic routing algorithms are fully reactive, with nodes forwarding packets without prior knowledge of their neighbors. The basic principle of beaconless forwarding can be described as follows: The *forwarder* i.e. the node that currently holds the packet, broadcasts it to its neighbors. The nodes within the forwarder's transmission range receive the packet, but only the nodes in the forwarding area are eligible for forwarding. The most suitable candidate is determined by contention. In [52], a beacon-less routing algorithm called BLR is proposed for mobile ad-hoc networks. BLR selects a forwarding node in a distributed manner among all its neighboring nodes without having any neighbor information, and optimized forwarding is achieved by applying a concept of Dynamic Forwarding Delay (DFD). Recently, Kalosha *et al* in [60] addressed the problem of designing beaconless geographic routing schemes to provide guaranteed deliv-

ery in wireless sensor networks. However, most existing beaconless geographic routing schemes are focused on solving the routing dead-ends problems, and the energy efficiency of beaconless routing is not well studied.

2.3 Energy-balancing Strategies

Wireless sensor networks are characterized by a many-to-one communication paradigm, i.e., most of the nodes in the network send their data to a few sink nodes. Moreover, packet delivery in sensor network is usually operated in a multi-hop manner since sensor nodes have limited transmission range. The above characteristics make unbalanced energy consumption an inherent problem for wireless sensor networks. The sensors close to the sink tend to deplete their energy budget faster than other sensors. This uneven energy consumption is apt to drastically reduce the useful lifespan of sensor networks and should be prevented to the largest extent possible. The existing approaches for balancing energy consumption in wireless sensor networks can be categorized into four groups: cluster heads rotation schemes, nonuniform node deployment schemes, data aggregation schemes and variable-power transmission schemes.

Clustering is particularly attractive to large-scale sensor networks [50][124]. By rotating cluster-heads within clusters, energy consumption can be distributed fairly evenly among nodes in each cluster. LEACH [50] is a clustering-based routing protocol, in which randomized cluster-heads rotation is employed to balance energy consumption among nodes within clusters. In HEED [124], cluster-heads are periodically selected based on node residual energy and other parameters such as node proximity to its neighbors or node degrees. To balance the energy consumption among cluster-heads, schemes such as EECS [74] and UCS [106] were proposed by partitioning the network into clusters with unequal size. Clusters closer to the base station have smaller sizes than those farther away from the base station. However, to achieve a desirable balance of energy consumption, cluster-head rotation must be performed frequently, which may add excessive communication overhead to the network, resulting in much energy wastage.

Nonuniform node distribution schemes, in which additional nodes are deployed to the area with heavy traffic, have been extensively studied to deal with the energy hole problem

[77][79][119]. In [77], a nonuniform node distribution strategy was proposed to increase the network data capacity. Liu *et al.* in [79] proposed a power-aware nonuniform node distribution scheme to deal with the so-called sink-routing hole problem for a long-termed connectivity in wireless sensor networks. Wu *et al.* in [119] proved that suboptimal balanced energy consumption is attainable only if the number of nodes grows with geometric proportion from the outer coronas to the inner ones except the outermost one. However, the nonuniform node distribution schemes may greatly increase the cost for deploying such networks since the increase of the number of sensor nodes makes the network more expensive than predicted.

Data aggregation has emerged as a useful paradigm in sensor networks [75][46]. The key idea is to combine data from different sensors to eliminate redundant transmissions. In [46], balancing energy consumption to increase network lifetime through data aggregation was discussed. In [75], the authors studied the problem of mitigating energy holes by traffic compression and aggregation. However, these works do not explore the possibility of avoiding energy holes in data gathering sensor networks.

Variable-power transmission is another attractive scheme for mitigating energy holes in wireless sensor networks [37] [45][58] [59] [90]. In [58], an energy balancing scheme called EBC(Energy Balanced Chain) was proposed to efficiently prolong network lifetime by actively controlling the nodes' hop distances. In EBC, nodes with higher traffic have a shorter hop distance than nodes with low traffic. Numerical results show that EBC performs significantly better than traditional hop-by-hop transmissions. In [90], the authors investigated the problems of avoiding energy holes and maximizing lifetime in sensor networks with uniform distribution and uniform reporting. They divided the network into coronas and presented an iterative process for computing the optimal size of each corona so that energy consumption is balanced among all coronas. Mixed-routing scheme, in which each node alternates between direct transmission and multi-hop transmission to transmit data, was first proposed in [45]. In [37], a slice-model based on mixed-routing was proposed and a probabilistic data propagation algorithm was designed to balance energy consumption among different slices. In [126], the problem of balancing energy consumption on a linear data gathering sensor network by considering energy consumption for both data transmission and data reception was studied. However, most the previous work focused

on balancing energy consumption across slices or coronas, and the problem of balancing energy consumption among nodes within the same slice or corona has not been well addressed. For example, in [90] and [37], the authors simply assume that the nodes in the same corona (slice) have the same probability of receiving packets from outer coronas (slices). Moreover, few work employs the combination of the above solutions to balance energy consumption. Unlike the existing work, the work done in this thesis is focused on both balancing energy consumption and maximizing network lifetime by combining the ideas of corona-based network division and mixed-routing strategy together with data aggregation. Solutions for balancing energy consumption among nodes both within the same corona and within different coronas have been designed.

2.4   Time-efficient Data Transport

Energy conservation in wireless sensor networks has been a primary design objective. However, this constraint has not overshadowed other key objectives like scalability, latency, reliability, etc. Many applications of wireless sensor networks are envisioned to handle critical scenarios where data retrieval time is crucial, i.e., delivering sensed information of each individual sensor node back to a central base station as fast as possible becomes most important. For example, in fire detection applications, it is desirable that each sensor can report its sensed event to the sink as quickly as possible and as reliably as possible. In health-monitoring applications, a sensor network may consist of many sensor instruments, which are attached to patients to monitor their physical conditions. Medical data sensed at each sensor is transmitted back to a central platform for diagnosis. Since the data is usually time varying in sensor networks, it is essential to guarantee that data can be successfully received by the sink the first time instead of being retransmitted due to collisions.

Most event-driven applications in WSNs are interactive, delay intolerant (real-time), mission critical, and non-end-to-end applications. The problem of providing real-time data transport in event-driven sensor networks has been extensively studied in the past few decades. In [16], a new distributed MAC protocol called DB-MAC was designed for delay-bounded applications in sensor networks. DB-MAC implements two mechanisms: *access with priority* and *path aggregation* to minimize the delay for packet delivery. In [114], the

authors considered the problem of optimizing delay in sequential change detection sensor networks. A censoring scheme, in which lagging sensors drop their delayed observations in order to mitigate network delay, was designed, and a lower bound on the network delay was also established. In [89], Nga *et al.* proposed a general reliability-centric framework for event reporting in Wireless sensor-actuator networks. This framework seamlessly integrates three key modules that process the event data, namely, an efficient and fault-tolerant event data aggregation algorithm, a delay-aware data transmission protocol, and an adaptive actuator allocation algorithm for unevenly distributed events. In [44], a delay-aware reliable transport (DART) protocol was designed. The objective of the DART protocol is to timely and reliably transport event features from the sensor field to the sink with minimum energy consumption. It incorporates the time critical event first (TCEF) scheduling mechanism to meet the application-specific delay bounds at the sink node. Furthermore, the DART protocol can accommodate multiple concurrent event occurrences in a wireless sensor field. Dousse *et al.* in [36] studied the problem of minimizing the time delay for a mobile intruder to be detected by a sensor with a connected path to the sink. It was shown that the probability the intruder proceeds undetected exhibits non-memoryless behavior over shorter distances and an exponentially decreasing tail.

Data gathering is a very important operation in time-driven sensor networks. Most of current data gathering researches have been emphasized on issues such as energy efficiency and network lifetime maximization. However, there are many emerging sensor network applications that require real-time data gathering since the data generated by the sensor nodes may become outdated after a certain time interval. In[80], a protocol named DMAC was designed to reduce packet delivery latency. It allows continuous packet forwarding by giving the sleep schedule of a node an offset that depends upon its depth on the tree. In [85], zhang *et al.* addressed the challenges of bursty convergecast in multi-hop wireless sensor networks, where a large burst of packets from different locations needs to be transported to a base station reliably with short delay. A protocols called RBC (for Reliable Bursty Convergecast) was designed. RBC employs a window-less block acknowledgment scheme that guarantees continuous packet forwarding and replicates the acknowledgment for a packet to improve channel utilization and to reduce the loss rate of ACK packets. RBC also employs differentiated contention control techniques to alleviate retransmission-incurred channel

contention. Some work has also been conducted on exploring energy-latency tradeoffs for data gathering in WSNs. In [78], schemes based on $energy \times delay$ metric were designed to balance the energy and delay cost for data gathering in wireless sensor networks. In [125], the authors explored the the energy-latency tradeoffs in wireless communication. They proposed algorithms to minimize the overall energy dissipation of the sensor nodes in the aggregation tree subject to the latency constraint.

2.5   Random Access Techniques

In most wireless sensor networks, all sensor nodes share the same communication channel. Contention occurs when two nearby sensor nodes both attempt to access the communication channel at the same time. Contention causes message collisions, which lead to too much energy wastage and decrease the lifetime of a sensor network. A Medium Access Control (MAC) protocol specifies how nodes share the channel, and hence plays a central role in the performance of a sensor network. MAC protocols are either contention-free or contention-based. In contention-free MAC protocols, collisions are completely avoided. Contention-free protocols include time-division multiple access (TDMA) [82], frequency division multiple access (FDMA) and code division multiple access (CDMA) [115]. Contention-based MAC protocols are also known as random access protocols, requiring no coordination among the nodes accessing the channel. Contention-based protocols include Pure ALOHA [11], slotted ALOHA [96], CSMA [65], etc.

Pure ALOHA and slotted ALOHA protocols are representative random access protocols. In Pure ALOHA, each node is allowed to access the channel whenever it has data to transmit. If the packet successfully reaches the destination (receiver), the next packet is sent. If the packet fails to be received at the destination, it is sent again. Since ALOHA uses a shared medium for transmission, data packets may frequently collide. The heavier the communications volume, the worse the collision problem becomes. The result is the degradation of system efficiency, because when two frames collide, the data contained in both packets is lost. To minimize the number of collisions, thereby optimizing network efficiency and increasing throughput, an enhanced version of Pure ALOHA called Slotted Aloha was developed. Slotted Aloha employs signals called beacons that are sent at precise

intervals and tell each source when the channel is clear to send a packet.

To further improve throughput and reduce data collision, a more sophisticated protocol called Carrier Sense Multiple Access with Collision Avoidance(CSMA/CA) was designed. CSMA works as listen-before-talk, i.e., before transmitting a frame, the station senses the medium (carrier sensing). If the medium is found idle at least for DIFS (DCF Inter-Frame Space) time period, the station starts transmission, and other stations wait until medium becomes idle again at least for DIFS time period. CSMA has been employed as the basic medium access mechanism of IEEE 802.11. To reduce the probability of data collision, a binary slotted exponential random backoff procedure is employed. Whenever a back-off occurs, the backoff time is randomly set from a uniform distribution over the interval $[0, contention\_window]$, while the contention window(CW) will be doubled for a retry and reset for a new packet. Since the probability that two packets at different nodes are set to the same backoff time is small, this mechanism can efficiently reduce data collision.

An additional mechanism, RTS/CTS(Request To Send/Clear To Send), is employed to solve the hidden terminal problem found in wireless networks that use CSMA. With RTS/CTS, the sender and receiver perform a handshake mechanism by exchanging RTS and CTS control frames. Based on this handshaking, collisions can only occur at the beginning when the RTS frame is transmitted. However, RTS/CTS mechanism can improve performance only when the data payload is relatively large. Using it for small data frames may result in significant overhead causing inefficient capacity utilization and higher delays.

Random channel access schemes have also been employed to solve other issues such as link assessment and neighbor discovery in wireless networks. In [64], the authors exploited the random scheme to perform link assessment task. In [113], directed and gossip-based neighbor discovery algorithms were designed based on random access. Some work has also been done on tuning the attempt probability for the sensor nodes in random access wireless networks. In [62], a generalized gradient ascent algorithm was proposed to self-learn the optimal attempt probability to maximize the throughput for each node. Two iterative, decentralized algorithms were presented in [116] to computing the global optimal rates for the problem of max-min fair rate allocation in ALOHA networks. In [15], the authors viewed the problem of optimizing the node attempt probability as a non-cooperative game and showed the existence of a Nash equilibrium giving an optimal channel access rate for

each node. However, all these approaches only consider the performance for each single node, such as maximizing the throughput or obtaining fairness for each node. In most applications of WSNs, each sensor node is not independent, and all the sensors must work in a cooperative manner to optimize network performance.

Chapter 3

Energy-efficient Beaconless Geographic Routing

3.1 Introduction

Energy is one of the most important resources in wireless sensor networks (WSNs) since sensor nodes are commonly powered by small batteries and are expected to work functionally for a long period. In many applications of WSNs, replacing or recharging batteries is either too expensive or impractical both because of the large number of sensor nodes and because of the hazardous environment in which the sensor nodes are deployed. Thus energy efficiency is a paramount concern in the design of WSNs. In addition to energy constraint, sensor nodes are also subject to further restrictions in computing power, bandwidth and storage capability, which indicates that sensor nodes can not deal with complex operations. These unique features have posed great challenges for designing routing schemes for WSNs. As pointed out by Stojmenovic *et al.* in [110], routing protocols for wireless sensor networks should be simple, localized, scalable and energy-efficient with low communication overhead.

Geographic routing, in which each node forwards packets only based on the locations of itself, its directed neighbors and the destination, is particularly attractive to resource-constrained sensor networks. The localized nature of geographic routing eliminates the overhead brought by route establishment and maintenance, indicating the advantages of

modest memory requirement at each node and high scalability in large distributed applications. In conventional geographic routing schemes, each node is required to maintain more or less accurate position information of all its direct neighbors, and the position of a node is made available to its direct neighbors by periodically broadcasting beacons. In WSNs with invariant or slowly changing network topology, maintaining neighbor information can greatly improve network performance because of the reusability of the maintained information and the low maintenance cost. However, in many application scenarios, WSNs are highly dynamic and the network topology may frequently change due to node mobility, node sleeping [38][68], node or link faults, etc. In highly dynamic scenarios, routing protocols based on maintaining neighbor information suffer from at least three drawbacks. First, the maintenance of neighbor information causes too much communication overhead and results in significant energy expenditure due to the frequent transmission of beacon messages. Secondly, the collected neighbor information can quickly get outdated, which in turn leads to frequent packet drops. Thirdly, the maintenance of neighbor information consumes memory which is also a scarce resource in WSNs.

To overcome the drawbacks of conventional geographic routing schemes in scenarios with dynamic network topology, several beaconless geographic routing protocols [52] [41] [22] [30] [60] have been proposed. Beaconless routing schemes, in which each node forwards packets without the help of beacons and without the maintenance of neighbor information, are fully reactive. When a node has a packet to transmit, it broadcasts the packet to its neighbors. The most suitable neighbor for further relaying the packet is determined based on a contention mechanism, in which each neighbor determines a proper delay for further forwarding the packet based on how well it is suited as the next-hop relay. Therefore, beaconless routing schemes are robust to topology changes since the forwarding decision is based on the actual topology at the time a packet is forwarded. However, in most existing beaconless routing schemes such as BLR [52], CBF [41] and GDBF [30], each node forwards packets based on hop-count routing metrics (e.g., each node selects its neighbor closest to the destination as its next-hop relay). These routing metrics are simple in implementation, but they can not guarantee energy efficiency which is a major concern in WSNs.

In this chapter, we address the problem of providing energy-efficient beaconless ge-

ographic routing for dynamic wireless sensor networks in which the network topology frequently changes over time. A novel routing protocol called Energy-efficient Beaconless Geographic Routing (EBGR) is designed. Without any prior knowledge of neighbors, EBGR tries to minimize the total energy consumed by delivering each packet to the sink. EBGR works as follows: each sensor node first calculates its ideal next-hop relay position based on the optimal forwarding distance in terms of minimizing the total energy consumption for delivering a packet to the sink. When a node has a packet to transmit, it first broadcasts a RTS message to detect its best next-hop relay. All suitable neighbors in the relay search region participate in the next-hop relay selection process using a timer-based contention mechanism. Each candidate that receives the RTS message sets a delay for broadcasting a corresponding CTS message based on a discrete delay function, which guarantees that the neighbor closest to the optimal relay position has the shortest delay. The neighbor has the minimum delay broadcasts its CTS message first, and the other candidates snooping the CTS message notice that another node has responded the request and quit the contention process. Finally, the packet is unicasted to the established next-hop relay. If there is no node in the relay search region, the forwarding node enters into a beaconless recovery mode to recover from the local minimum. Extensive theoretical analysis has been conducted to evaluate the performance of EBGR. It is proven that the total energy consumption for delivering each packet to the sink under EBGR has an upper bound, assuming no data loss and no failures in greedy forwarding. Moreover, it is demonstrated that the expected energy consumption derived based on uniform node distribution approaches to the lower bound with the increase of node deployment density. EBGR is also extended to handle the unreliable communication links which are common in realistic applications of WSNs. Simulation results show that the proposed scheme significantly outperforms existing schemes in terms of energy efficiency in WSNs with highly dynamic network topology.

The remainder of this chapter is structured as follows. The related work on power-aware routing and geographic routing is discussed in Section 4.2. The system models are described in Section 4.3. The energy-efficient beaconless geographic routing protocol EBGR is described in Section 3.4. In Section 3.5, we give extensive theoretical analysis for EBGR. In Section 3.6, we extend EBGR to lossy wireless sensor networks. In Section 3.7, we evaluate our schemes through extensive simulations and present the comparisons with

other protocols. Finally, we conclude the chapter and discuss future extensions in Section 3.8. For convenience, a list of the parameters used in this chapter is given in Appendix B.

3.2 Related Work

Traditional geographic routing schemes commonly employ maximum-distance greedy forwarding techniques. The MFR protocol proposed in [111] is one of the earliest geographic routing algorithms in which each node forwards its packets to the neighbor that has the maximum progress [1]. In 'greedy' routing [39], each node forwards the packets to its neighbor closest to the destination. However, these schemes can not guarantee that packets are delivered in an energy-efficient manner since energy consumption is not taken into account for making forwarding decisions. Recently, some work has been done on improving energy efficiency for geographic routing. In[118], a protocol called GPER was proposed to provide power-efficient geographic routing in WSNs. In GPER, each node first establishes a sub-destination within its maximum radio range and then employs the shortest-path algorithm to compute the minimum-energy path from itself to the sub-destination. In [103], Seada *et al.* addressed the *weak-link* problem and studied the energy and reliability trade-offs pertaining to geographic forwarding in lossy sensor networks using the $PRR \times DIST$ metric.

To deal with the dynamic network topology, beaconless routing schemes, in which each node forwards packets without any neighbor knowledge, have been proposed. Heissenbuttel and Braun [52] proposed the Beacon-Less Routing (BLR) algorithm. BLR selects a forwarding node in a distributed manner among its neighboring nodes whereas each node is not required to have information about neighboring nodes, neither about their positions nor even about their existence. In BLR, optimized forwarding is achieved by applying a concept of Dynamic Forwarding Delay(DFD). In contention-based forwarding (CBF) [41], Fuessler *et al.* proposed a technique called *active selection method*, in which a forwarding node selects its next-hop relay through broadcasting control message before transmitting the data packet. The implicit geographic forwarding (IGF) proposed by Blum *et al* [22]

---

[1]Given node u, the progress that node u obtains by transmitting a packet to its neighbor a is defined as the projection of $|ua|$ onto the line connecting u and the final destination of the transmitted packet.

and the geographic random forwarding (GeRaF) proposed by Zorai [128] implement the same ideas but focus on the integration of beaconless routing with IEEE 802.11 MAC layer. However, most of the proposed beaconless schemes employ hop-count based routing metric, which is not efficient in terms of energy consumption.

Most geographic routing protocols use greedy forwarding as its basic mode of operation. However, greedy forwarding may fail when a node can not find a better neighbor than itself to forward the packet. To recover from a local minimum, GFG [92], GPSR[24] and GOAFR+ [67] route a packet around the faces of a planar subgraph (e.g., *Relative Neighborhood Graph* (RNG) and *Gabriel Graph* (GG)) when a local minimum is encountered. The planar subgraphs are constructed based on neighborhood information. However, in beaconless routing neighborhood information is not a priori knowledge. In BLR [52], a *Request-response* approach was proposed for recovering from local minima. The forwarder broadcasts a request message and all neighboring nodes respond. If a node is closer to the destination, it becomes the next hop relay. In [60], algorithms for constructing different proximity graphs in beaconless routing were designed. Two schemes, Beaconless Forwarder Planarization (BFP) and Angular Relaying, were proposed to recover from local minima for beaconless routing. To provide guaranteed delivery in WSNs, most existing geographic routing algorithms [18][23][40][73] [52] [60] switch between the greedy forwarding mode and recovery mode depending on the network topology.

3.3   Preliminaries

3.3.1   Network Model

We consider a general network model in which a large number of sensors are densely deployed in a two-dimensional terrain. Without loss of generality, it is assumed that no two nodes locate at the same position. All sensor nodes are equipped with the same radio transceiver that enables a maximum transmission range R. Each node knows its own location as well as the location of the sink. We use the *Unit Disk Graph (UDG)* communication model in the first stage of analysis. In this model, any two nodes u and v can communicate with each other reliably if and only if $|uv| \leq R$, where $|uv|$ is the Euclidean

distance between u and v. In Section 3.6, based on a realistic communication model in which data loss is estimated by packets reception rate, we extend our scheme to achieve localized energy-efficient beaconless routing in the presence of unreliable communication links.

3.3.2 Energy Model

The *First Order Radio Model* proposed in [51] has been widely used for measuring energy consumption in wireless communications [20][76][84][110][118]. In this model, the energy for transmitting one bit data over distance x is $\epsilon_t(x) = a_{11} + a_2 x^k$, where a_{11} is the energy spent by transmitter electronics, a_2 is the transmitting amplifier and k $(k \geq 2)$ is the propagation loss exponent. The energy for receiving one bit data is $\epsilon_r = a_{12}$, where a_{12} is the energy spent by receiver electronics. Therefore, the energy consumed by relaying one bit data (i.e., receiving one bit data and then transmitting it over distance x) is

$$\varepsilon_{relay}(x) = a_{11} + a_2 x^k + a_{12} \equiv a_1 + a_2 x^k \tag{3.1}$$

where $a_1 = a_{11} + a_{12}$.

3.3.3 Characteristics of Power-adjusted Transmission

In [109], the characteristics of energy consumption for power adjusted transmission were investigated using a generalized form of the *First Order Radio Model*. Given a source node u and a destination node v, d denotes the distance between u and v, and $\xi(d)$ represents the total energy consumed by delivering one bit data from u to v. The following lemmas hold according to the analysis presented in [109].

Lemma 1. *[109] If* $d \leq \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}}$, *direct transmission is the most energy-efficient way to deliver packets from u to v.*

Lemma 2. *[109] If* $d > \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}}$, $\xi(d)$ *is minimized when all hop distances are equal to $\frac{d}{N}$, and the optimal number of hops is $\lfloor \frac{d}{d_o} \rfloor$ or $\lceil \frac{d}{d_o} \rceil$ where $d_o = \sqrt[k]{\frac{a_1}{a_2(k-1)}}$.*

Lemma 3. *The total energy consumption for delivering one bit data over distance d satisfies $\xi(d) \geq a_1 \cdot \frac{k}{k-1} \cdot \frac{d}{d_o}$.*

From Lemma 2, it can be observed that d_o is the optimal forwarding distance in terms of minimizing $\xi(d)$ when d is an integral multiple of d_o. Even if d can not be divided exactly by d_o, d_o is also a good approximation of the optimal forwarding distance. Moreover, d_o remains constant for given sensor device and application environment since d_o only depends on a_1, a_2 and k. Thus d_o can act as an effective metric to guide localized packet forwarding to provide energy-efficient routing. Based on this observation, in our study, the ideal next-hop relay position for any node u in terms of minimizing the total energy consumption for delivering a packet from node u the sink is defined as follows:

Definition 1. *Given any node u, the ideal position of its next-hop relay, denoted by f_u, is defined as the point on the straight line from u to the sink s where $|uf_u| = d_o$.*

Since each node u has the knowledge of its own location as well as the location of the sink, node u can easily compute its ideal next-hop relay position f_u. In our scheme, each node makes fully localized and stateless forwarding decisions based on the location of its ideal next-hop relay position.

3.4 Energy-efficient Beaconless Geographic Routing (EBGR)

The general idea of EBGR is to employ the optimal forwarding distance d_o to make energy-efficient localized forwarding decisions. Each forwarding node chooses the neighbor closest to its ideal relay position as its next-hop relay. In this way, each packet is expected to be delivered along the minimum energy route from the source to the sink.

The main mechanism of EBGR is depicted as follows: first, the forwarding node broadcasts a RTS message to its neighbors to detect its best next-hop relay, but only the nodes in the *relay search region* (see Fig. 3.1) are candidates for further forwarding the packet. Each candidate that receives the RTS message schedules a CTS transmission to the forwarder with some delay determined by a discrete delay function which guarantees that the

neighbor closest to the ideal next-hop relay position of the forwarding node has the shortest delay. The node that has the shortest delay broadcasts its CTS message first and establishes itself as the next-hop relay. Finally, the forwarding node unicasts its packet to the established relay node. If there is no node in the relay search region, beaconless angular relaying is employed to recover from the local minimum.
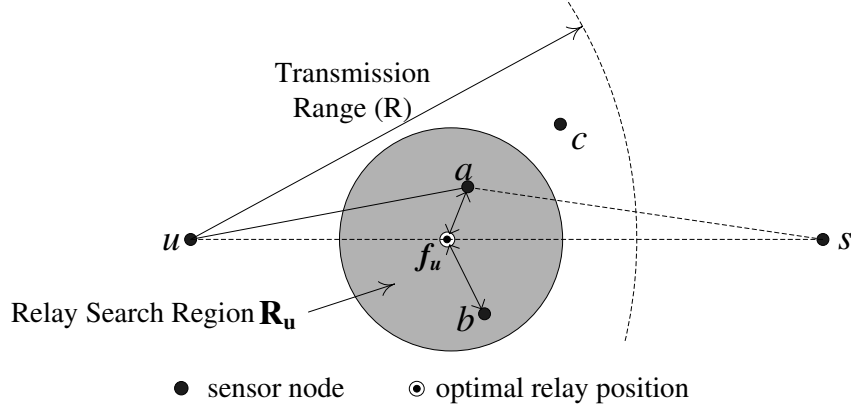
Figure 3.1: Greedy forwarding in EBGR in which u is the forwarder, and s is the sink. Only a and b are eligible candidates since they are in the relay search region.

In the following, we describe the details of EBGR including the relay search region, the greedy forwarding procedure, the discrete delay function, and the beaconless recovery procedure.

3.4.1   Relay Search Region

Since the best next-hop relay for any node u is the neighbor closest to its ideal relay position f_u, there is no need for all neighbors of node u to participate in the contention for acting as the next-hop relay. In EBGR, each node has a relay search region which is defined as follows:

Definition 2. *Given any node u, its next-hop relay search region, denoted by $\mathbf{R}_u$, is defined as the disk centered at $u's$ ideal next-hop relay position f_u with radius $r_s(u)$ where $r_s(u) \leq |uf_u| = d_o$.*

For any node u, only the neighbors in its relay search region $\mathbf{R}_u$ are candidates for further forwarding the packets transmitted from node u. The concept of relay search region

is introduced to prohibit the unsuitable neighbors from participating the relay contention procedure.

3.4.2 Greedy Forwarding

Given any node u, let $|us|$ be the distance from node u to the sink s. Node u calculates $|us|$ since it has the knowledge of its own position as well as the position of the sink. If the sink is in $u's$ transmission range and $|us| \leq \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}}$, node u transmits its packets directly to the sink because relaying the packets by some intermediate nodes is no more energy-efficient than direct transmission (see Lemma 1). Otherwise, node u detects it best next-hop relay based on the procedure given as follows.

Let (x_u, y_u) and (x_s, y_s) be the coordinates of node u and the sink s respectively. By Definition 1, the location of f_u, denoted by (x_{uo}, y_{uo}), can be computed as follows:

$$
\begin{cases}
x_{uo} = x_u - \frac{d_o}{|us|}(x_u - x_s); \\
y_{uo} = y_u - \frac{d_o}{|us|}(y_u - y_s).
\end{cases}
$$

When node u has a packet to transmit, it broadcasts a RTS message, which also contains the location of its ideal next-hop relay position as well as the radius of its relay search region, to detect its best next-hop relay. For any neighbor w that receives the RTS message from node u, it first checks if it falls in $\mathbf{R}_u$. If $w \notin \mathbf{R}_u$, the RTS message is simply discarded. Otherwise, node w generates a CTS message which also contains its own location and sets a proper delay, denoted by $\delta_{w \to u}$, for broadcasting the CTS message based on a discrete delay function given in next subsection. If node w overhears a CTS message broadcasted by another candidate before $\delta_{w \to u}$ is due, node w cancels broadcasting its own CTS message; otherwise node w broadcasts its CTS message when $\delta_{w \to u}$ is due. When node u receives the CTS message from its neighbor w, the next hop relay for u, denoted by $n(u)$, is updated if $n(u)$ is null or $|wf_u| < |n(u)f_u|$. Finally, node w unicasts its packet to its next-hop relay $n(u)$.

3.4.3 Discrete Delay Function

In EBGR, the selection of the next-hop relay is performed by means of contention through RTS/CTS handshaking. The forwarding node selects its best next-hop relay based on the CTS messages received from the candidates in its relay search region. To reduce the communication overhead incurred by relay selection, a discrete delay function is designed to promote the best relay and to suppress the broadcasting of CTS message by other unsuitable neighbors.
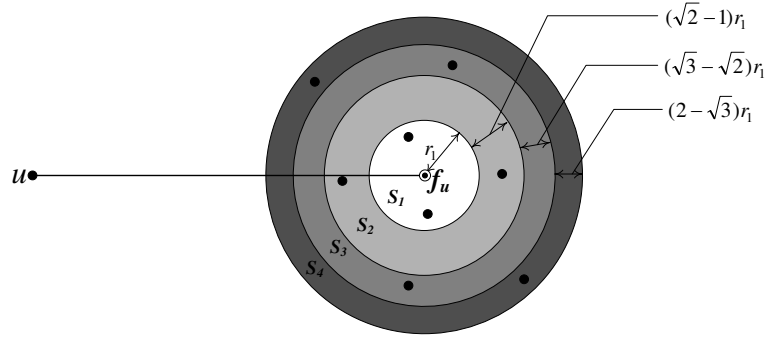
Figure 3.2: The relay search region $\mathbf{R}_u$ is divided into 4 coronas S_1, S_2, S_3, S_4 where S_i has width $(\sqrt{i} - \sqrt{i-1})r_1$ since all coronas have the same area size.

For any node u, its relay search region $\mathbf{R}_u$ is divided into n concentric coronas $S_1, S_2...S_n$ where all coronas have the same area size (see Fig. 3.2). Thus the width of the ith corona is $(\sqrt{i} - \sqrt{i-1})r_1$ where r_1 is the radius of S_1 and $r_1 = \frac{r_s(u)}{\sqrt{n}}$. If $v \in S_i$, the distance between node v and f_u satisfies that $\frac{\sqrt{i-1} \cdot r_s(u)}{\sqrt{n}} \leq |vf_u| < \frac{\sqrt{i} \cdot r_s(u)}{\sqrt{n}}$. Therefore, given any node $v \in \mathbf{R}_u$, v must locate in S_m where $m = \left\lfloor \left(\frac{\sqrt{n} \cdot |vf_u|}{r_s(u)}\right)^2 \right\rfloor + 1$.

For any node $v \in \mathbf{R}_u$, instead of broadcasting the CTS message immediately after receiving a RTS message from node u, node v broadcasts its CTS message with delay $\delta_{v \to u}$. Let γ be the delay for transmitting a packet over a unit distance. $\delta_{v \to u}$ is defined as follows:

$$\delta_{v \to u} = \frac{\gamma \cdot r_s(u)}{\sqrt{n}} \cdot (2 \sum_{i=1}^{m} \sqrt{i} - \sqrt{m} - 1) + \gamma \cdot (|vf_u| - \sqrt{\frac{m-1}{n}} r_s(u)), \qquad (3.2)$$

$$\text{where} \quad m = \left\lfloor \left(\frac{\sqrt{n} \cdot |vf_u|}{r_s(u)}\right)^2 \right\rfloor + 1.$$

The delay computed by Equation (3.2) guarantees:

1. Nodes in S_i broadcast their CTS messages earlier than nodes in S_j where $j > i$.

2. Given node $v_i \in S_i$ and node $v_j \in S_j$ $(1 \le i < j \le n)$, it must satisfy that $\delta_{v_j \to u} - \delta_{v_i \to u} > \gamma \cdot |v_i v_j|$, which means that v_j can overhear the CTS message broadcast by v_i and cancels broadcasting its own CTS message before $\delta_{v_j \to u}$ is due.

3. For all nodes located in the same corona S_i, the node closest to f_u broadcasts its CTS message first because the second term $\gamma \cdot (|v f_u| - \frac{(m-1) r_s(u)}{\sqrt{n}})$ in Equation (3.2) guarantees that the node closer to f_u has a shorter delay.

If there is only one node in the innermost non-empty corona, the above delay function guarantees that the number of CTS/RTS messages broadcasted for detecting the best relay for u is minimized (only 2 messages, one CTS message broadcasted by u and one RTS message broadcasted by the neighbor closest to f_u). If there are multiple nodes in the innermost non-empty corona, there may be more than one CTS broadcasted. For example, there are two nodes v_i and v_j in the innermost non-empty corona where $\delta_{v_j \to u} - \delta_{v_i \to u} < \gamma \cdot |v_i v_j|$. Both v_i and v_j broadcast CTS messages since the CTS message broadcasted by node v_i can not reach node v_j before $\delta_{v_j \to u}$ is due. However, the above delay function can still significantly reduce the number of CTS messages broadcasted because only the nodes in the most inner nonempty corona have the chance to broadcast CTS messages. Obviously, the performance of the discrete delay function depends on the number of coronas the relay search region is divided into. A large n can significantly reduce the number of CTS messages broadcasted but may also incur additional delay when the first several coronas are empty. In practice, the proper number of coronas can be determined based on the application requirements.

3.4.4   Beaconless Recovery

When node u broadcasts a RTS message to detect its best next-hop relay, it sets its timer to t_{max} and starts the timer. t_{max} is large enough to guarantee that node u can receive the CTS message from the furthest neighbor in $\mathbf{R}_u$ before the timer is expired. If node u receives no CTS message till the timer is expired, it assumes that there is no neighbor in its relay search region. To recover from the local minimum, the beaconless angular

relaying proposed in [60] is employed in EBGR. The angular relaying algorithm works in two phases: *selection phase* and *protest phase*. In the selection phase, the forwarder u broadcasts a RTS message to its neighbors, and the neighbors answer with CTS messages in counter-clockwise order according to an angular-based delay function. After the first candidate w answers with a valid CTS, the protest phase begins. First, only the nodes in $N_{GG}(v, w)$ (i.e., the Gabriel circle having uw as diameter) are allowed to protest. If a node x protests, it automatically becomes the next hop relay. After that, only nodes in $N_{GG}(v, x)$ are allowed to protest. Finally, the forwarder sends the packet to the selected (first valid or last protesting) candidate.

The detailed operations for relay selection is give in Algorithm **RS**.

Algorithm: RS /*Relay Selection performed at node u*/

1:     If $(|us| \leq \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}})$ and $(|us| \leq R)$ then

2:         $n(u) = s$;

3:     Else Calculate (x_{uo}, y_{uo});

4:   **On broadcasting the RTS message:**

5:     Broadcast RTS message $(r_s(u), (x_u, y_u), (x_{uo}, y_{uo}))$;

6:   **On Receiving a message *M* from node *w*:**

7:     If *M is a RTS message* $(r_s(w), (x_w, y_w), (x_{wo}, y_{wo}))$ then

8:       If $u \notin \mathbf{R}_w$ then

9:         Discard message *M*;

10:       Else

11:         Set $\delta_{u \to w}$ according to Equation (3.2);

12:      If *M is a CTS message* $((x_w, y_w), (x_v, y_v))$ then

13:        If $(x_v = x_u)$ and $(y_v = y_u)$ then

14:          If $(n(u) = Null)$ or $(|wf_u| < |n(u)f_u|)$ then

15:            $n(u) = w$;

16:        Else If u *received a RTS message from* v and $\delta_{u \to v}$ *is not due* then

17:            Cancel broadcasting CTS message to v;

18:   **On broadcasting *reply* message:**

19:       For each delay label $\delta_{u \to w}$ do

20:       If $\delta_{u \to w}$ *is due* then

21:         Generate CTS message $((x_u, y_u), (x_w, y_w))$;

22:         Broadcast $((x_u, y_u), (x_w, y_w))$;

23:   **On beaconless recovery:**

24:       If t_{max} *is due* then

25:         Employing Beaconless angular relaying to recover from local minimum;

3.5   Theoretical Analysis

In this section, we present extensive theoretical analysis for EBGR. We first prove that EBGR is loop-free in greedy forwarding mode, and demonstrate that EBGR can provide guaranteed data delivery as long as the network is connected. Based on a simplified MAC model without packet loss, the unit disk graph model without failures in greedy forwarding and uniform node deployment, we establish the bounds on hop-count, the upper bound on energy consumption and the expected energy consumption for sensor-to-sink data delivery under EBGR.

3.5.1   Notations and Definitions

In [84], two terms, *progress* and *advance*, were introduced to distinguish different forwarding rules in geographic routing. Suppose that node u forwards its packets to its neighbor v for relay to the sink s. The *progress*, denoted by $P(u, v)$, is defined as the projected distance of $|uv|$ on the straight line passing through u and s, and the *advance*, denoted by $A(u, v)$, is defined as the difference between $|us|$ and $|vs|$. Therefore,

$$P(u, v) = |uv| cos\angle vus; \tag{3.3}$$

$$A(u, v) = |us| - |vs|. \tag{3.4}$$

We define two metrics, called *energy over progress ratio* and *energy over advance ratio* to analyze the characteristics of energy consumption in EBGR. Let $\gamma_P(u, v)$ and $\gamma_A(u, v)$ be the *energy over progress ratio* and the *energy over advance ratio* for relaying one bit data from u to v respectively. $\gamma_P(u, v)$ and $\gamma_A(u, v)$ are defined as follows:

$$\gamma_P(u, v) = \frac{\varepsilon_{relay}(|uv|)}{P(u, v)} = \frac{a_1 + a_2 |uv|^k}{|uv| cos\angle vus}; \tag{3.5}$$

$$\gamma_A(u, v) = \frac{\varepsilon_{relay}(|uv|)}{A(u, v)} = \frac{a_1 + a_2 |uv|^k}{|us| - |vs|}. \tag{3.6}$$

Without loss of generality, we assume that the maximum transmission range (R) is no less than $2d_o$ since the analysis approach for $R < 2d_o$ is the same.

3.5.2 Guaranteed Delivery
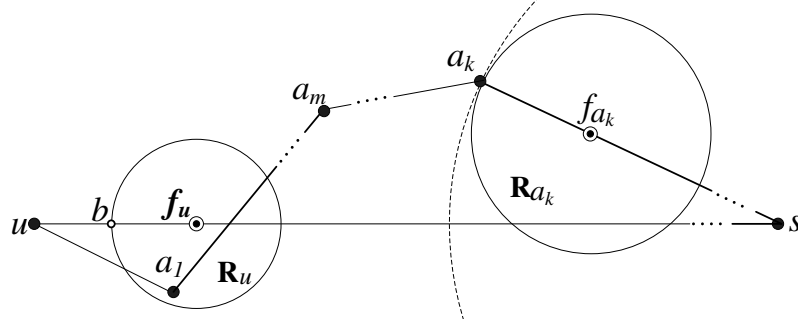
Figure 3.3: Illustration of loop-free packet forwarding in greedy forwarding mode

Theorem 1. *EBGR is loop-free in greedy forwarding mode.*

Proof. Let u be a source node and s be the sink. If s locates in the transmission range of node u and $|us| \leq \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}}$, node u sends its packets directly to the sink without any relay. The theorem holds in this case.

If $|us| > \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}}$, the packets generated by node u may be relayed by some intermediate nodes before arriving at the sink. As shown in Fig. 3.3, the maximum distance from s to any point in $\mathbf{R}_u$ is $|sb|$. $\forall a_1 \in \mathbf{R}_u$, $|a_1s| \leq |bs| < |us|$ because $r_s(u) \leq d_o$ and no two nodes locate at the same position. Therefore, $A(u, a_1) = |us| - |a_1s| > 0$, which means that each forwarding must obtain a positive advance.

Let a_k be a node that relays the packets generated by node u, and $\overrightarrow{a_0a_1...a_m...a_{k-1}a_k}$ represents the routing path from u to a_k in EBGR where $a_0 = u$. For any relay node a_m prior to a_k in this routing path, $A(a_k, a_m) = |a_ks| - |a_ms| < 0$, which means that a_k can not forward its packets to a_m. Hence, the theorem holds. $\square$

EBGR works in two modes: *greedy forwarding* and *beaconless angular relaying*. In the *greedy forwarding* mode, EBGR is loop-free according to Theorem 1. When greedy forwarding fails, the beaconless angular relaying scheme is employed to recover from local minima. The beaconless angular relaying uses the *Select* and *Protest* methods to avoid crossing edges which might cause a routing loop. In [60], it is proven that the angular relaying algorithm can always select the first edge of the Gabriel subgraph in counter-clockwise order. Thus there are no routing loops in angular relaying. Therefore, EBGR can provide guaranteed delivery as long as the network is connected.

3.5.3 Bounds on Hop Count

For any node u, let $C(u)$ be the minimum relay search region that covers only one neighbor, and let r be the radius of $C(u)$. Since nodes are uniformly deployed, the minimum relay search regions for all nodes in the network have the same size. Then we have the following theorem.

Theorem 2. *If there are no failures in greedy forwarding, the number of hops, denoted by N, for delivering a packet from u to the sink s under EBGR where $|us| = d$ satisfies*

$$\frac{d}{d_o + r} - 1 < N < \frac{d}{d_o - r} + 1. \tag{3.7}$$

Proof. Let v be the node in $C(u)$. As shown in Fig. 3.4, $|vs|$ is maximized when v locates at point a and minimized when v locates at point b. By Equation (3.4), $A(u, v) = |us| - |vs|$. Therefore, $d_o - r \le A(u, v) \le d_o + r$.
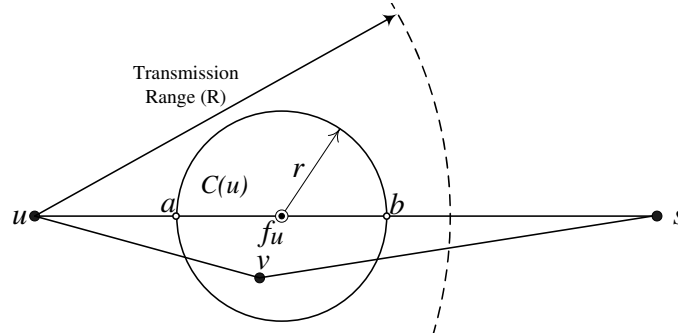
Figure 3.4: $C(u)$ is the minimum relay search region that covers only one node, and r is the radius of $C(u)$.

Let $\overrightarrow{ua_1 a_2 ... a_{N-1} s}$ denote the routing path from u to the sink s. Based on Equation (3.4),

$$d = |us| = A(u, a_1) + \sum_{i=1}^{N-2} A(a_i, a_{i+1}) + A(a_{N-1}, s).$$

For the prior $N - 1$ hops, the best relay is chosen based on the same metric (i.e., the neighbor closest to the ideal next-hop relay position). Therefore,

$$(N - 1)(d_o - r) \le d - A(a_{N-1}, s) \le (N - 1)(d_o + r).$$

That is,

$$\frac{d - A(a_{N-1}, s) + (d_o + r)}{d_o + r} \leq N \leq \frac{d - A(a_{N-1}, s) + (d_o - r)}{d_o - r}.$$

For the last hop, the packet is directly transmitted to the sink. Based on Lemma 1, $0 < A(a_{N-1}, s) \leq \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}}$. By Lemma 19 (see appendix), $0 < A(a_{N-1}, s) < 2d_o < 2(d_o + r)$. Thus

$$N \geq \frac{d - A(a_{N-1}, s) + (d_o + r)}{d_o + r} > \frac{d - 2(d_o + r) + (d_o + r)}{d_o + r} = \frac{d}{d_o + r} - 1;$$

and,

$$N \leq \frac{d - A(a_{N-1}, s) + (d_o - r)}{d_o - r} < \frac{d + (d_o - r)}{d_o - r} = \frac{d}{d_o - r} + 1.$$

Therefore,

$$\frac{d}{d_o + r} - 1 < N < \frac{d}{d_o - r} + 1.$$

$\square$

3.5.4   Upper Bound on Energy Consumption

In this subsection, we establish the upper bound on energy consumption for sensor-to-sink data delivery under EBGR, assuming no packet loss and no failures in greedy forwarding. In EBGR, the best next-hop relay for each node is detected through RTS/CTS handshaking. Since the discrete delay function can effectively suppress unsuitable candidates for broadcasting CTS messages, the number of RTS/CTS messages broadcasted is proportional to the number of hops for delivering the data to the sink. Therefore, the energy consumed by broadcasting and receiving RTS/CTS can be viewed as a part of energy spent by data delivery. In this analysis, the energy consumption for sensor-to-sink data delivery is referred to as the sum of energy consumed by the nodes in the routing path for delivering one bit data from the source to the sink.

In the following, we first prove that the position that maximizes the *energy over advance ratio* $\gamma_A(u, v)$ must be located on the border of the minimum relay search region in Lemma 4. Then in Lemma 6, we demonstrate that the *energy over advance ratio* at each hop is upper bounded. Finally, the upper bound on energy consumption for sensor-to-sink data

delivery is given in Theorem 3.

Lemma 4. *Given $v \in C(u)$, v must locate on the border of $C(u)$ when $\gamma_A(u,v)$ is maximized.*

Proof. Suppose v is located inside $C(u)$. Let v' be a point on the border of $C(u)$ and v' has the same *x*-coordinate with v. By Equation (3.6),

$$
\begin{aligned}
\gamma_A(u,v') - \gamma_A(u,v) &= \frac{\varepsilon_{relay}(|uv'|)}{A(u,v')} - \frac{\varepsilon_{relay}(|uv|)}{A(u,v)} \\
&= \frac{a_1 + a_2(|uv'|^k)}{|us| - |v's|} - \frac{a_1 + a_2(|uv|^k)}{|us| - |vs|}.
\end{aligned}
$$

Since $|v's| > |vs|$ and $|uv'| > |uv|$, $\gamma_A(u,v') - \gamma_A(u,v) > 0$. Hence, this lemma holds. $\square$
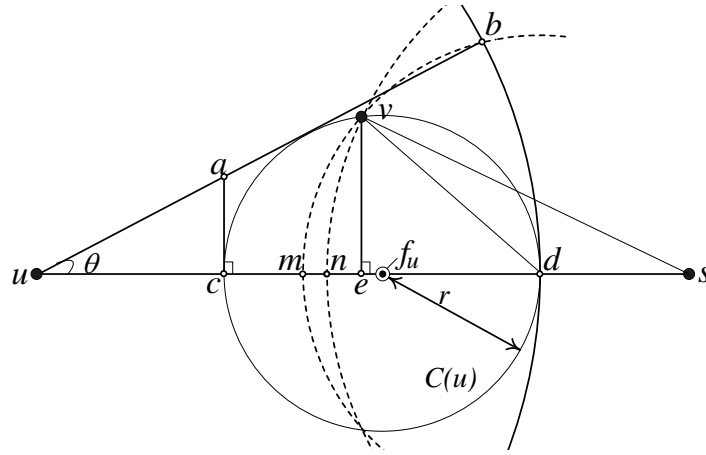
Figure 3.5: $C(u)$ is the minimum relay search region, and v is a node located on the border of $C(u)$ where $|vs| = |ns|$, $|vd| = |md|$, $|ud| = |ub|$.

Lemma 5. *If v locates on the border of $C(u)$ and $r > 0$,*

$$
\frac{A(u,v)}{P(u,v)} \geq \frac{r}{d_o + r - \sqrt{d_o^2 - r^2}}.
$$

Proof. As shown in Fig. 3.5, $|vs| = |ns|$ and $|vd| = |md|$. Clearly, $|un| \geq |um|$ because $\angle vns > \angle vms$. By Equation (3.3) and Equation (3.4),

$$
\frac{A(u,v)}{P(u,v)} = \frac{|us| - |vs|}{|ue|} = \frac{|us| - |ns|}{|ue|} \geq \frac{|ud| - |md|}{|ue|} = \frac{|ud| - |vd|}{|ue|}.
$$

Assume u is the origin and the straight line from u to v is the x-axis. $(0,0)$ and (x,y) represent the coordinates of u and v respectively. Since v locates on the border of $C(u)$, $(x - d_o)^2 + y^2 = r^2$. Thus

$$\frac{A(u,v)}{P(u,v)} \geq \frac{d_o + r - \sqrt{2r(r + d_o - x)}}{x} \geq \frac{r}{d_o + r - \sqrt{d_o^2 - r^2}}.$$

$\square$

Lemma 6. *Given* $v \in C(u)$, $\gamma_A(u,v) < \max\{\frac{2a_2 d_o^k[k-1+(\frac{d_o-r}{d_o+r})]}{(d_o-r)(1+\sqrt{\frac{d_o-r}{d_o+r}})}, \frac{2a_2 d_o[(k-1)d_o^k+(d_o+r)^k]}{(d_o+r)(\sqrt{d_o^2-r^2}+d_o-r)}\}$.

Proof. From Lemma 4, v must locate on the border of $C(u)$ when $\gamma_A(u,v)$ is maximized. By Equation (3.6) and Lemma 5,

$$\gamma_A(u,v) = \frac{a_1 + a_2|uv|^k}{A(u,v)} = \frac{a_1 + a_2|uv|^k}{P(u,v)\frac{A(u,v)}{P(u,v)}} = \gamma_P(u,v)\frac{P(u,v)}{A(u,v)} \leq \gamma_P(u,v)\frac{d_o + r - \sqrt{d_o^2 - r^2}}{r}.$$

$$(3.8)$$

Let $\overline{ub}$ be the tangent of the circle centered at f_u with radius r (see Fig. 3.5). a and b are two points on the line segment from a to b where $|ua| = (d_o - r)\sec\theta$ and $|ub| = d_o + r$. Let ψ denote the set of points located on the line segment between a and b. $\forall w \in \psi$,

$$\gamma_P(u,w) = \frac{a_1 + a_2|uw|^k}{P(u,w)} = \frac{1}{\cos\theta} \cdot \frac{a_1 + a_2|uw|^k}{|uw|}, \qquad (3.9)$$

where $(d_o - r)\sec\theta \leq |uw| \leq d_o + r$. Clearly, $\gamma_P(u,w)$ given in Equation (3.9) is strictly concave with respect to $|uw|$. Therefore,

$$\max_{w \in \psi} \gamma_P(u,w) = \max\{\gamma_P(u,a), \gamma_P(u,b)\}. \qquad (3.10)$$

Similar to the proof of Lemma 4, it is easy to prove that $\gamma_P(u,v) \leq \max_{w \in \psi} \gamma_P(u,w)$. By Equation (3.8) and Equation (3.10),

$$\gamma_A(u,v) < \max\{\gamma_P(u,a), \gamma_P(u,b)\} \cdot \frac{d_o + r - \sqrt{d_o^2 - r^2}}{r}, \qquad (3.11)$$

where $\gamma_P(u,a) = \frac{a_1 + a_2 d_o^k(\frac{d_o-r}{d_o+r})^{\frac{k}{2}}}{d_o - r}$ and $\gamma_P(u,b) = \frac{d_o(a_1 + a_2(d_o+r)^k)}{(d_o+r)\sqrt{d_o^2-r^2}}$. By Lemma 2, $d_o =$

$\sqrt[k]{\frac{a_1}{a_2(k-1)}}$. Replacing $\gamma_P(u, a)$ and $\gamma_P(u, b)$ in Equation (3.11),

$$\gamma_A(u, v) < \max\{\frac{2a_2 d_o^k[k - 1 + (\frac{d_o-r}{d_o+r})]}{(d_o - r)(1 + \sqrt{\frac{d_o-r}{d_o+r}})}, \frac{2a_2 d_o[(k - 1)d_o^k + (d_o + r)^k]}{(d_o + r)(\sqrt{d_o^2 - r^2} + d_o - r)}\}.$$

$\square$

Theorem 3. *If there are no failures in greedy forwarding and no packet loss in EBGR, the total energy consumption, denoted by $\xi(d)$, for delivering one bit data from source u to the sink s under EBGR where $|us| = d$ satisfies*

$$\xi(d) < \max\{\frac{2a_2 d_o^k[k - 1 + (\frac{d_o-r}{d_o+r})]}{(d_o - r)(1 + \sqrt{\frac{d_o-r}{d_o+r}})}, \frac{2a_2 d_o[(k - 1)d_o^k + (d_o + r)^k]}{(d_o + r)(\sqrt{d_o^2 - r^2} + d_o - r)}\} \cdot d. \quad (3.12)$$

Proof. Let $\xi'(d)$ and N' be the total energy consumption and the number of hops respectively for delivering one bit data for u to the sink s when all hops including the last one use the same metric (*i.e.*, the neighbor closest to the ideal next-hop relay position) to choose relay. Let d_i and $\gamma(i)$ denote the distance and the energy over advance ratio for the ith hop respectively.

$$\xi'(d) = \underbrace{a_{11} + a_2 d_1^k}_{u} + \underbrace{\sum_{i=2}^{N'} \varepsilon_{relay}(d_i)}_{relay} + \underbrace{a_{12}}_{s} = \sum_{i=1}^{N'} \varepsilon_{relay}(d_i) = \sum_{i=1}^{N'} \gamma(i) \cdot d_i.$$

Since all hops are independent and use the same routing metric, $\xi'(d)$ is maximized when the packet is forwarded to the neighbor with the maximum energy over advance ratio at each forwarding. By Lemma 1, $\xi(d) < \xi'(d)$ because the metric used for the last forwarding in EBGR saves energy. Therefore,

$$\xi(d) < \xi'(d) < \left(\max_{i=1}^{N'-1} \gamma(i)\right) \cdot \sum_{i=1}^{N'} d_i = d \cdot \max_{i=1}^{N'-1} \gamma(i).$$

By Lemma 6, the upper bound on energy over advance ratio only depends on a_2, d_o, k

and r. So, all hops have the same upper bound on energy over advance ratio. Therefore,

$$\xi(d) < \max\{\frac{2a_2d_o^k[k-1+(\frac{d_o-r}{d_o+r})]}{(d_o-r)(1+\sqrt{\frac{d_o-r}{d_o+r}})}, \frac{2a_2d_o[(k-1)d_o^k+(d_o+r)^k]}{(d_o+r)(\sqrt{d_o^2-r^2}+d_o-r)}\} \cdot d.$$

□

Let r_{ul} denote the ratio of the upper bound to the lower bound on energy consumption for delivering one bit data over distance d in EBGR. We have the following corollary.

Corollary 1.

$$r_{ul} < \max\{\frac{k-1+(\frac{d_o-r}{d_o+r})^{\frac{k}{2}}}{k(1-\frac{r}{d_o})\sqrt{\frac{d_o-r}{d_o+r}}}, \frac{k-1+(1+\frac{r}{d_o})^k}{k(1-\frac{r^2}{d_o^2})}\}.$$

When $a_{11} = a_{12} = 50nJ/bit$, $a_2 = 100pJ/bit/m^2$ and $k = 2$, $r_{ul} < \frac{1000}{(\sqrt{1000}-r)(\sqrt{1000}-r^2)}$.

Proof. By Lemma 3 and Theorem 3, we have

$$r_{ul} = \frac{\xi_{upperbound}(d)}{\xi_{lowerbound}(d)};$$

$$< \max\{\frac{\frac{2a_2d_o^k[k-1+(\frac{d_o-r}{d_o+r})]}{(d_o-r)(1+\sqrt{\frac{d_o-r}{d_o+r}})} \cdot d}{a_1 \cdot \frac{k}{k-1} \cdot \frac{d}{d_o}}, \frac{\frac{2a_2d_o[(k-1)d_o^k+(d_o+r)^k]}{(d_o+r)(\sqrt{d_o^2-r^2}+d_o-r)}}{a_1 \cdot \frac{k}{k-1} \cdot \frac{d}{d_o}} \cdot d\}$$

$$< \max\{\frac{2a_2d_o^k[k-1+(\frac{d_o-r}{d_o+r})] \cdot d}{a_1 \cdot \frac{k}{k-1} \cdot \frac{d}{d_o} \cdot (d_o-r)(1+\sqrt{\frac{d_o-r}{d_o+r}})}, \frac{2a_2d_o[(k-1)d_o^k+(d_o+r)^k] \cdot d}{a_1 \cdot \frac{k}{k-1} \cdot \frac{d}{d_o} \cdot (d_o+r)(\sqrt{d_o^2-r^2}+d_o-r)}\}$$

By Lemma 2, $d_o = \sqrt[k]{\frac{a_1}{a_2(k-1)}}$. Replacing d_o with $\sqrt[k]{\frac{a_1}{a_2(k-1)}}$ in Equation 3.13,

$$r_{ul} < \max\{\frac{k-1+(\frac{d_o-r}{d_o+r})^{\frac{k}{2}}}{\frac{k}{2}\sqrt{\frac{d_o-r}{d_o+r}}(1-\frac{r}{d_o}+\sqrt{1-\frac{r^2}{d_o^2}})}, \frac{k-1+(1+\frac{r}{d_o})^k}{\frac{k}{2}(1+\frac{r}{d_o})(1-\frac{r}{d_o}+\sqrt{1-\frac{r^2}{d_o^2}})}\}.$$

Since $1-\frac{r}{d_o} < \sqrt{1-\frac{r^2}{d_o^2}}$,

$$r_{ul} < \max\{\frac{k-1+(\frac{d_o-r}{d_o+r})^{\frac{k}{2}}}{k(1-\frac{r}{d_o})\sqrt{\frac{d_o-r}{d_o+r}}}, \frac{k-1+(1+\frac{r}{d_o})^k}{k(1-\frac{r^2}{d_o^2})}\}.$$

When $a_{11} = a_{12} = 50nJ/bit$, $a_2 = 100pJ/bit/m^2$ and $k = 2$, by Lemma 2, $d_o =$

$\sqrt{\frac{a_1}{a_2}} = \sqrt{1000}$. Thus

$$r_{ul} < \frac{1000}{(\sqrt{1000} - r)(\sqrt{1000 - r^2})}$$

□

From Corollary 1, it is worth noting that $r_{ul} \to 1$ when $r \to 0$. Therefore, the upper bound on energy consumption is close to the lower bound when sensor nodes are densely deployed. However, r_{ul} becomes infinite when r approaches d_o. This phenomenon can be explained as follows: when r approaches to d_o, as can been seen in Fig. 3.5, the advance obtained by forwarding the data to the next-hop relay is very small if the next-hop relay is located at the position closest to node u. Thus $\gamma_A(u, v)$ must be very large since the energy spent by electronic circuit (i.e, a_{11} and a_{12}) becomes the dominant part in the total energy consumption. Obviously, the worst case for delivering a data packet to the sink is that the next-hop relay for each forwarding node is located at the position that maximizes the energy over advance ratio. Let $P_w(d)$ denote the probability that the worst case happens for delivering packets from u to the sink s where $|us| = d$. The following theorem shows that $P_w(d)$ approaches to 0 when r approaches to d_o.

Theorem 4. $P_w(d)$ *monotonically decreases with the increase of d and r. $P_w(d) \to 0$ when $r \to d_o$.*

Proof. Let N be the number of hops for delivering a packet from u to s where $|us| = d$. Let $p(i)$ denote the probability that the packet is forwarded in the way that the energy over advance ratio at hop i is maximized. For the prior $N - 1$ hops, the forwarding at each hop is the same and independent. Thus $p(i) = p(j) \equiv p$ $(1 \le i, j \le N - 1)$. We have

$$P_w(d) = p(N) \cdot \prod_{i=1}^{N-1} p(i) < p^{N-1}.$$

Since nodes are deployed with uniform distribution, $p \propto \frac{1}{\pi r^2}$. For the hop count N, it increases with the increase of d. Therefore, $P_w(d)$ monotonically decreases with the increase of d and r.

From Equation (3.6), the location that maximizes $\gamma_A(u, v)$ approaches to u when $r \to d_o$ because the advance obtained for each forwarding is very small and a_1 becomes the

dominant part in energy consumption. When $r \to d_o$, $N \to \infty$. Therefore, $P_w(d) \to 0$ when $r \to d_o$. □

3.5.5 Expected Energy Consumption

Let $E[\gamma_A(u,v)]$ denote the expected energy over advance ratio for one hop forwarding in EBGR. We have the following lemma.
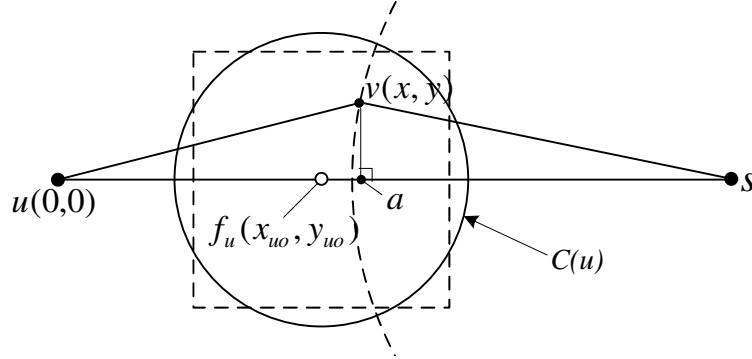
Figure 3.6: Approximation of advance where v is the only node in $C(u)$ and $|vs| = |bs|$. The dashed square is the approximation of $C(u)$, and it has the same area size with $C(u)$.

Lemma 7.

$$E[\gamma_A(u,v)] \approx \rho \iint_{C(u)} \frac{a_1 + a_2(x^2 + y^2)^{\frac{k}{2}}}{x} dxdy, \tag{3.13}$$

where ρ is the node deployment density.

Proof. Let v be the node in $C(u)$. Since sensor nodes are uniformly distributed with density ρ, by Equation (3.6),

$$
\begin{aligned}
E[\gamma_A(u,v)] &= \iint_{C(u)} \frac{\varepsilon_{relay}(|uv|)}{A(u,v)} \rho dxdy \\
&= \rho \iint_{C(u)} \frac{a_1 + a_2|uv|^k}{|us| - |vs|} dxdy.
\end{aligned}
$$

When nodes are densely deployed, $C(u)$ is small. As shown in Fig. 3.6, the advance obtained by forwarding the packet to v is close to the progress, that is, $|us| - |vs| \approx |ua|$.

Thus

$$
\begin{aligned}
E[\gamma_A(u, v)] &\approx \rho \iint_{C(u)} \frac{a_1 + a_2 |uv|^k}{|ua|} dx dy \\
&\approx \rho \iint_{C(u)} \frac{a_1 + a_2 (x^2 + y^2)^{k/2}}{x} dx dy.
\end{aligned}
\tag{3.14}
$$

$\square$

Theorem 5. *If there are no failures in greedy forwarding and no packet loss, the expected energy consumption, denoted by $E[\xi(d)]$, for delivering one bit data from source u to the sink s where $|us| = d$ satisfies*

$$
E[\xi(d)] \approx \rho d \iint_{C(u)} \frac{a_1 + a_2 (x^2 + y^2)^{\frac{k}{2}}}{x} dx dy.
$$

Proof. Let N be the number of hops to deliver one packet from u to s. By Lemma 7, it is easy to see that the approximation of $E[\gamma_A(u, v)]$ only depends on ρ. Therefore, the prior $N - 1$ hops have the same approximated energy over advance ratio, denoted by $E[\gamma]$, due to the same forwarding procedure. Let $E[\gamma(N)]$ be the energy over advance ratio for the last hop. As shown in Fig. 3.6,

$$
E[\gamma(N)] = \rho \iint_{C(u)} \frac{a_1 + a_2 (|sv|)^k}{|sv|} dx dy
$$

When nodes are densely deployed, $C(u)$ is small and $E[\gamma(N)] \approx E[\gamma]$. Furthermore, the effect of the last hop on the total energy consumption is small for large d. Therefore,

$$
\begin{aligned}
E[\xi(d)] &= E[\sum_{i=1}^{N} \gamma(i) d_i] \approx E[\gamma] d \\
&\approx \rho d \iint_{C(u)} \frac{a_1 + a_2 (x^2 + y^2)^{k/2}}{x} dx dy.
\end{aligned}
$$

$\square$

Let r_{el} be the ratio of the approximated expected energy consumption to the lower

bound on energy consumption for delivering one bit data from a node to the sink. We have the following corollary.

Corollary 2.

$$r_{el} = \frac{1}{2} + \frac{12d_o^2 + \pi r^2}{24\sqrt{\pi}rd_o} \ln \frac{2d_o + \sqrt{\pi}r}{2d_o - \sqrt{\pi}r} \quad when \quad k = 2$$

and $r_{el} < 1.5$ when $a_{11} = a_{12} = 50nJ/bit$, $a_2 = 100pJ/bit/m^2$.

Proof. As shown in Figure 3.6, a square centered at f_u with side l is used to approximate $C(u)$ where $l^2 = \pi r^2$. Since $C(u)$ is the minimum relay search region, $\rho \pi r^2 = 1$, that is, $\rho = \frac{1}{\pi r^2}$. When $k = 2$,

$$
\begin{aligned}
E[\xi(d)] &\approx \frac{d}{\pi r^2} \int_{d_o-l}^{d_o+l} \int_{-l}^{l} \frac{a_1 + a_2(x^2 + y^2)}{x} dx dy \\
&\approx \frac{a_2(12d_o^2 + \pi r^2)}{12\sqrt{\pi}r} \ln \frac{2d_o + \sqrt{\pi}r}{2d_o - \sqrt{\pi}r} + a_2 d_o.
\end{aligned}
\tag{3.15}
$$

By Equation (3.15) and Lemma 3,

$$r_{el} = \frac{1}{2} + \frac{12d_o^2 + \pi r^2}{24\sqrt{\pi}rd_o} \ln \frac{2d_o + \sqrt{\pi}r}{2d_o - \sqrt{\pi}r}, \quad where \quad 0 < r \leq d_o.$$

When $a_{11} = a_{12} = 50nJ/bit$, $a_2 = 100pJ/bit/m^2$ and $k = 2$, by Lemma 2, $d_o = \sqrt{\frac{a_1}{a_2}} = \sqrt{1000}$. When $r \to 0$, $r_{el} \to 1$. When $r = \sqrt{1000}$, $r_{el} < 1.5$. $\qquad \square$

3.5.6   Summary of Analysis on Energy Consumption

To demonstrate the energy efficiency of EBGR, we give the comparison between r_{ul} and r_{el}. Similar with [19] [51], the system parameters are set as follows: $k = 2$, $a_{11} = a_{12} = 50nJ/bit$ and $a_2 = 100pJ/bit/m^2$. Under this setting, the optimal hop distance is $\sqrt{1000}m$ according to Lemma 2. The maximum transmission range R for all nodes is set to $80m$ which is larger than $2d_o$. Fig. 3.7 plots r_{ul} and r_{el} under different size of the relay search region. It can be seen that r_{ul} is close to 1 when r is small because a small r means that the best relay for each node keeps very close to its ideal next-hop relay position. With the increase of r, r_{ul} first increases slightly and reaches around 2 when $r = 15m$. After that,
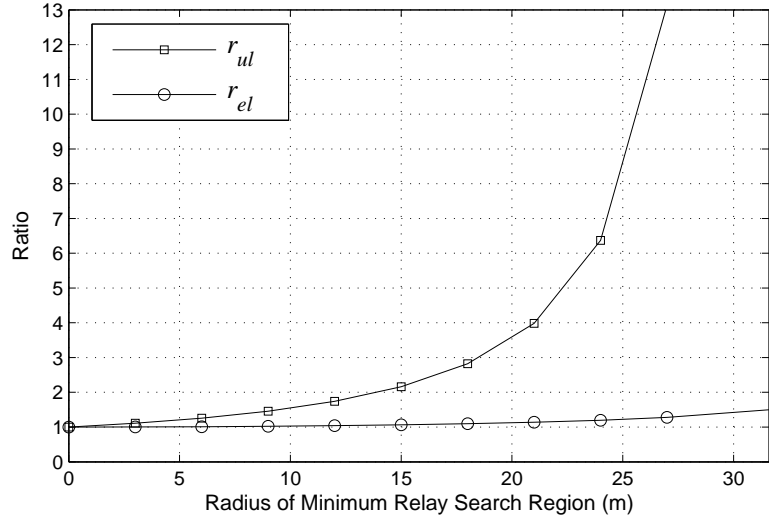
Figure 3.7: Comparison of r_{ul} and r_{el} with the variation of the radius of the minimum relay search region

r_{ul} increases quickly with the increase of r and approaches to infinity when r comes near to $\sqrt{1000}m$. However, r_{el} increases slowly with the increase of r. Even if $r = \sqrt{1000}m$, r_{el} is less than 1.5. This is because the probability that packets are delivered along the worst path decreases with the increase of r. By Theorem 4, the probability that a packet is delivered along the worst path approaches to 0 when $r \to d_o$. Therefore, the expected energy consumption does not increase dramatically when r approaches to d_o.

3.6 Extension to Lossy Wireless Sensor Networks

The protocol design and theoretical analysis discussed in the previous sections are based on the UDG model in which transmissions between any two nodes within the communication range are assumed to be reliable. In this section, we extend EBGR to lossy sensor networks to provide energy-efficient sensor-to-sink routing in the presence of unreliable communication links.

3.6.1 Routing Metric

To capture the characteristics of data loss, *packet reception rate (PRR)*, which is the ratio of successful transmissions to the total number of transmissions, is used to measure the

quality of communication links. Let $PRR(u,v)$ be the packet reception rate for link (u,v). The expected number of transmissions that guarantees one successful transmission from u to v is $\frac{1}{PRR(u,v)}$.

Packets may be lost due to many reasons such as data corruption, collision or the attenuation of signal strength. In the case where a packet is lost before reaching the receiver, nearly the same amount of energy is dissipated by listening [121]. Therefore, the total expected energy consumption for successfully relaying one bit data from u to v, denoted by $E[\varepsilon(u,v)]$, can be approximately modeled as

$$E[\varepsilon(u,v)] \simeq \frac{\varepsilon_{relay}(|uv|)}{PRR(u,v)}.$$

By Equation (3.6), the expected energy over advance ratio, denoted by $E[\gamma_A(u,v)]$, for successfully relaying one bit data from u to v satisfies

$$
\begin{aligned}
E[\gamma_A(u,v)] &\simeq \frac{\varepsilon_{relay}(|uv|)}{PRR(u,v)A(u,v)} \\
&\simeq \frac{1}{PRR(u,v)} \cdot \frac{\varepsilon_{relay}(|uv|)}{A(u,v)}.
\end{aligned}
\tag{3.16}
$$

Note that the second part ($\frac{\varepsilon_{relay}(|uv|)}{A(u,v)}$) in Equation (3.16) is the energy over advance ratio when transmission from u to v is reliable. Motivated by this observation, we propose a new metric for providing energy-efficient routing in lossy sensor networks. Instead of choosing the neighbor closest to the ideal next-hop relay position among all candidates in the relay search region, node u chooses the neighbor that minimizes $\frac{|vf_u|}{PRR(u,v)}$ as its next hop relay. We refer to the extended version of EBGR using this routing metric as EBGR-2.

3.6.2 Blacklisting and Discrete Delay Function

For any node $v \in \mathbf{R}_u$, node v is blacklisted from participating in the contention for acting as packet relay for node u if $PRR(u,v) < \delta$. Let $B(u)$ be the set of remaining nodes in $\mathbf{R}_u$ after blacklisting. Clearly, $\max_{v \in B(u)} \frac{|vf_u|}{PRR(u,v)} \leq \frac{r_s(u)}{\delta}$. A *Discrete Delay Function* similar to the one in EBGR is then used to reduce the number of CTS messages broadcasted. The principle of this discrete delay function is the same with that in EBGR. The nodes in $B(u)$

are divided into n sets S_1, S_2,...S_n based on the following rule: If $v \in S_i$, $\frac{(i-1) \cdot r_s(u)}{\delta} \leq \frac{|vf_u|}{PRR(u,v)} < \frac{i \cdot r_s(u)}{\delta}$. The delay for node v to broadcast its CTS message after receiving a RTS message from u, denoted by $\delta_{v \to u}$, is defined as follows:

$$\delta_{v \to u} = 2(m-1) \cdot \gamma \cdot r_s(u) + \gamma \cdot \frac{|vf_u|}{PRR(u,v)}, \quad \text{where} \quad m = \left\lfloor \frac{n \cdot \delta \cdot |vf_u|}{r_s(u) \cdot PRR(u,v)} \right\rfloor + 1.$$

(3.17)

The delay setting ensures that the neighbors in S_1 broadcast the CTS messages first. Within each set, the neighbor that has a smaller $\frac{|vf_u|}{PRR(u,v)}$ is assigned with a shorter delay. Furthermore, the CTS message broadcasted by a node in one set can be snooped by the nodes in other sets before they broadcast their own CTS messages. It is easy to see that the number of messages needed to be broadcasted is minimized when there is only one node in innermost non-empty set.

3.7 Simulation Results and Analysis

In this section, we evaluate the performance of EBGR and EBGR-2 through simulations. We first demonstrate the energy efficiency of EBGR in dynamic scenarios where the network topology changes frequently due to node mobility and node sleeping. Then we present the performance of EBGR-2 under a realistic radio model and compare EBGR-2 with the $PRR \times DIST$ metric specially designed for routing in lossy sensor networks.

3.7.1 Simulation Settings

To study the behavior of EBGR and EBGR-2, we have implemented a simulation package based on OMNeT++ version 3.3 [5]. In all simulations, 200 sensor nodes are randomly deployed in a $500m \times 400m$ region. There is only one sink which is placed at the center of the region. Three scenarios are designed to evaluate the performance of the proposed schemes.

- **Mobility Scenario**: In this scenario, all sensor nodes move according to the Random Walk Mobility Model [97]. A sensor node moves from its current location to a new location by randomly choosing a speed from range $[minspeed, maxspeed]$ and a

direction from range $[0, 2\pi]$. Each movement continues for an interval of 10 seconds. New speed and direction are chosen at the end of each interval. If a node reaches a simulation boundary, it bounces off the border with an angle determined by the incoming direction and continues the movement along the new path.

- **Random Sleeping Scenario**: In this scenario, all nodes remain static after deployment, and the Random Independent Sleeping (RIS) scheme [68] is employed to extend network lifetime. The simulation time is divided into intervals with length of $\mathcal{T}_{sleep}$. At the beginning of each interval, each node decides to work in *active* state with probability p and to enter into *sleep* state with probability $1 - p$. With this sleeping scheme, the expected network lifetime can be increased by a factor close to $1/p$.

- **High-variant Link Quality Scenario**: In this scenario, all nodes remain static and no sleeping scheme is employed. However, the link quality changes dynamically over time. The behavior of each link is modeled according to a realistic channel model proposed in [25]. The simulation time is divided into link quality estimation intervals with length of $\mathcal{T}_{esti}$. At the end of each interval, each node estimates the packet reception rate (PRR) of a link from a neighbor to itself by counting the number of beacon or RTS/CTS messages received from that neighbor.

For performance analysis, in addition to EBGR and EBGR-2, we have implemented another three routing schemes: GPER[118], BLR[52] and $PRR \times DIST$[103]. GPER is a beacon-based geographic routing scheme in which each node maintains neighbor information by periodically broadcasting beacons. Based on the maintained neighbor information, each node first chooses its neighbor closest to the sink as a sub-destination and then uses a shortest-path algorithm to compute the energy optimal next-hop relay. BLR is a beaconless geographic routing scheme based on hop-count routing metric. In all simulations done in this section, the *closest-to-destination* routing metric, in which each node chooses its neighbor closest to the sink as its next-hop relay, is employed in BLR. The $PRR \times DIST$ routing metric is introduced to provide energy efficient geographic routing in the presence of unreliable transmissions. In this metric, both packet reception rate (PRR) and hop forwarding

distance are taken into account to make routing decisions aiming at obtaining a good trade-off between shorter, high-quality links and longer lossy links. For each neighbor that is closer to the destination, the product of the reception rate and the distance improvement achieved by forwarding to this neighbor is computed, and the neighbor with the highest value is chosen as the next-hop relay. In [103], simulations and experiments show that $PRR \times DIST$ is the best forwarding metric for making localized geographic forwarding decisions in lossy wireless networks with ARQ mechanisms.

The underlying MAC protocol is IEEE 802.11, and the configuration of the MAC protocol is described as follows: For beacon-based schemes (i.e.,GPER and $PRR \times DIST$), the RTS/CTS exchange function is turned off to reduce communication overhead since RTS/CTS handshaking is not necessary for these two schemes. For fair comparison, both BLR and EBGR use the RTS/CTS handshaking for selecting the next-hop relay, avoiding packet duplication and reducing packet collisions. In all simulations, the maximum transmission range for each node is set to $80m$. The beacon message is set to 20 bytes. The RTS message is 25 bytes and the CTS message is 20 bytes. For the parameter settings in the delay function of EGBR and EBGR-2, the number of coronas/sets (i.e., n) is set to 20, and the transmission delay (i.e., λ) is $10^{-6}s/m$. The recovery timers for both EBGR, EGBR-2 and BLR are set to $40ms$. For the energy model, the energy spent by transmitter electronics on transmitting or receiving one bit data (i.e. a_{11} and a_{12}) is set to $50nJ/bit$, and the transmitting amplifier (a_2) is set to $10pJ/bit/m^2$, and the propagation loss exponent (k) is set to 2.

In each simulation run, 20 nodes are selected as sources and each source generates 40 data packets with a payload of 128 bytes. The simulation is terminated until the sink receives all the data packets generated in the network, and the simulation results are the average of 10 independent runs.

3.7.2   Performance of EBGR in Mobility Scenarios

In this set of simulations, we evaluate the performance of EBGR in mobile scenarios in which the network topology changes frequently due to node mobility. The parameters of the Random Walk Mobility Model are set as follows: $minspeed$ is set to $0m/s$, and

$maxspeed$ is varied from $0m/s$ to $50m/s$ to provide different levels of mobility. The simulated beacon intervals for GPER are 0.5, 1.0 and 2.0 seconds.
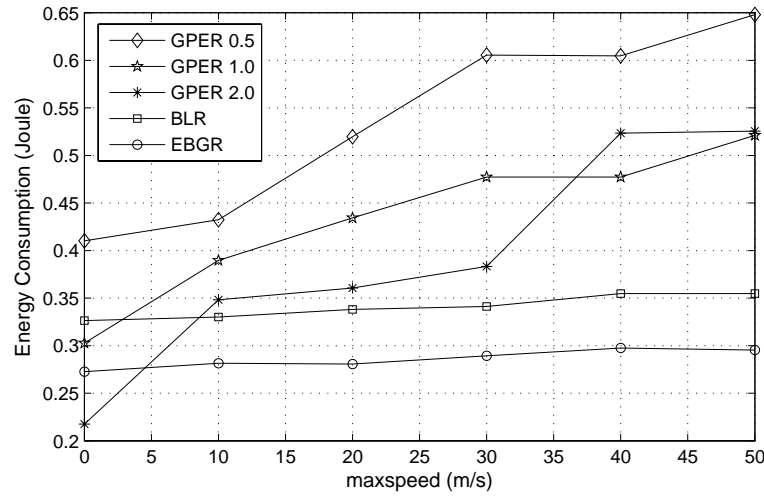
Figure 3.8: The total energy consumption under EBGR, BLR and GPER with different mobility levels
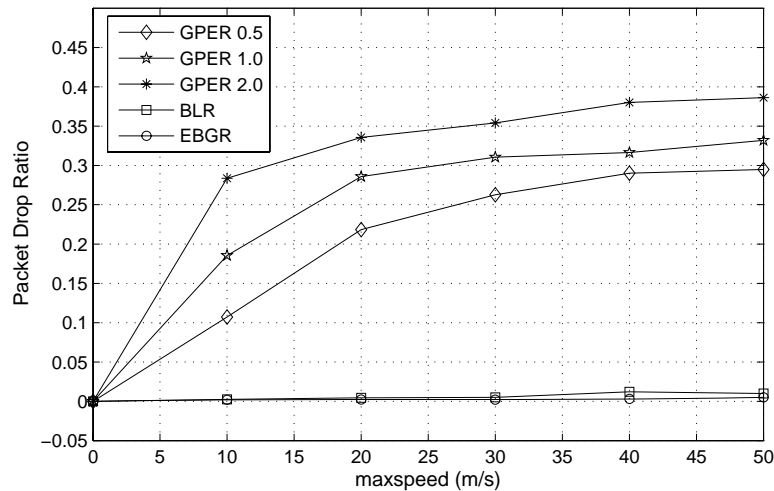
Figure 3.9: The Packet drop ratio under EBGR, BLR and GPER with different mobility levels

Fig. 3.8 shows the total energy consumption of GPER, BLR and EBGR under different mobility levels. The total energy consumption is referred to as the sum of the energy spent by each node in the network during the simulation time. As can be seen from Fig. 3.8, EBGR and BLR are much more robust to network topology changes than GPER since each forwarding decision in EBGR and BLR is made based on the actual topology at the time
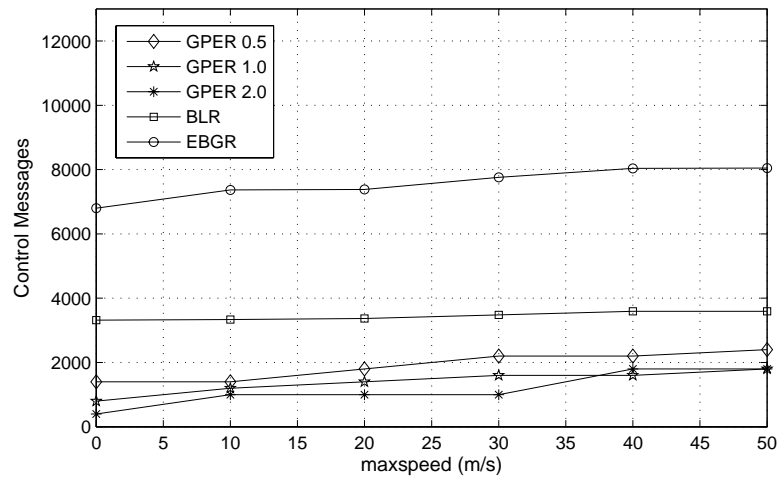
Figure 3.10: The number of control messages under EBGR, BLR and GPER with different mobility levels
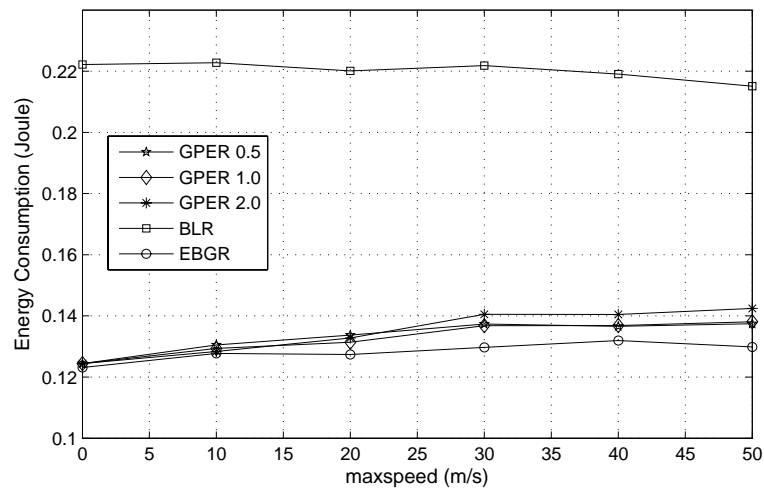
Figure 3.11: The sum of energy along routing path under EBGR, BLR and GPER with different mobility levels

the packet is transmitted. With the increase of $maxspeed$, the energy consumption under EBGR and BLR only increase slightly due to slight packet drops (see Fig. 3.9) and the suboptimal power routes (see Fig. 3.11) caused by node movement. From Fig. 3.9, we can observe that the data packet drop ratios in EBGR and BLR are close to 0. As the maximum movement speed increases, the data packet drop ratio in BLR is a little larger than that in EBGR because each node in BLR tends to choose the neighbor close to the border of the transmission range as its next-hop relay, and it is possible that the established next-relay moves out of the transmission range before receiving the data packet.

In contrast, node movement has a great impact on the performance GPER. When $maxspeed$ is lower than $5m/s$, GPER with a beacon interval of 2.0 seconds consumes less energy than EBGR because of the low packet drop rate (as shown in Fig. 3.9) and the low beacon message overhead (as shown in Fig. 3.10). As the maximum movement speed increases, the total energy consumption under GPER increases abruptly. With a maximum movement speed of $50m/s$, GPER with a beacon interval of 0.5s and 1s consumes 109% and 65% more energy than EBGR. This phenomenon can be explained as follows: in scenarios with high node movement speed, the maintained information in GPER becomes outdated quickly, resulting in frequent packet drops. As shown in Fig. 3.9, with a maximum movement speed of $50m/s$, the packet drop ratio can be as high as 0.38 for GPER with a beacon interval of 2 seconds. To provide guaranteed packet delivery, dropped packets must be retransmitted, resulting in much energy wastage. Moreover, frequent packet drops lead to a long sensor-to-sink data delivery delay. Since the number of beacon messages broadcasted in GPER scales proportionally to the running time, much additional energy is consumed by exchanging beacon messages.

It is worth noting that EBGR consumes much less energy than BLR. This result can be explained as follows. First, in BLR, each node chooses its neighbor closest to the sink as its next-hop relay. The hop distance is much larger than the optimal forwarding distance d_o when each node has a large maximum transmission range, resulting in significant energy dissipation since energy consumption on data transmission is proportional to a square of the transmission distance. Fig. 3.11 shows the sum of energy spent on successful data packet transmission(the energy wasted by unsuccessful packet transmission is not taken into account in order to demonstrate the quality of routing path.). It can be seen that BLR

consumes at least 80% more energy than EBGR, which means that the routing paths in BLR are not energy efficient. Second, in BLR, the power level for broadcasting RTS message must be large enough so that all nodes in the transmission range of the transmitter can receive the RTS message, whereas in EBGR, the power level for broadcasting RTS message only need to guarantee that all neighbors in the relay search region can receive the RTS message. The power level for broadcasting RTS message is much smaller than that in BLR with a large maximum transmission range. Fig. 3.10 shows that the number of RTS/CTS messages broadcasted in BLR is 50% less than that in EBGR since the number of RTS/CTS messages is proportional to the number of hops. However, it doesn't mean that the energy spend by broadcasting RTS/CTS messages in BLR must be smaller than EBGR because each node spends more energy to broadcast RTS/CTS messages in BLR.

3.7.3   Performance of EBGR in Random Sleeping Scenarios

Sleeping schemes are attractive to applications such as fire detection and intrusion detection since sensor networks in these applications commonly work on a low duty cycle and each node generates data only when it detects an event. Thus by letting nodes periodically switch between *active* and *sleep* states, a large amount of energy wasted in an idle state can be saved. However, periodic node sleeping can result in frequent network topology change. In this section, we evaluate the performance of EBGR in scenarios where random independent sleeping (RIS) is employed to extend network lifetime.

The simulated protocols are slightly modified in order to integrate with the RIS sleeping scheme. For GPER, instead of using a constant beacon interval, each node broadcasts a beacon message only when it switches between *active* and *sleep* states since there is no need to broadcast beacon message if its work status does not change. For EBGR and BLR, a node that has data to transmit broadcasts a RTS message only when it works in *active* state and its remaining active time is large enough to finish forwarding one data packet. A neighbor node can join in the relay contention process only when its remaining active time is large enough to receive the data if it is selected as the next hop relay.

Fig. 3.12 shows the energy consumption under EBGR, BLR and GPER with different sleeping probability where the length of the time interval in RIS is set to 4 seconds. When
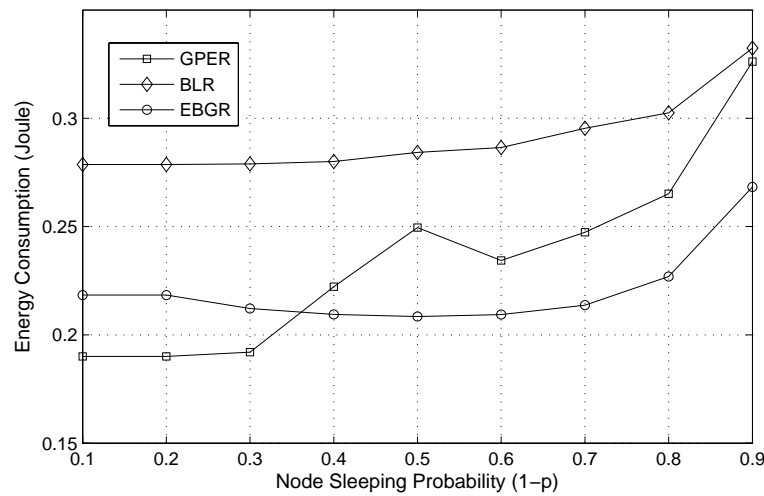
Figure 3.12: The total energy consumption under EBGR, BLR and GPER in random sleeping scenarios with different sleeping probability

node sleeping probability is smaller than 0.36, GPER outperforms EBGR and BLR because most of nodes work in active state and the amount of beacon messages broadcasted is small due to the low frequency of states switching. With the increase of node sleeping probability, the energy consumption under GPER increases. There is a peak at $p = 0.5$ due to the high frequency for state switching. When node sleeping probability is larger than 0.6, the energy consumption for all three protocols increases with the increase of node sleeping probability. The higher the node sleeping probability is, the smaller the number of active nodes is. Thus, nodes with data to transmit switch to recovery mode frequently, resulting in much energy consumption. Moreover, GPER can not avoid temporary loops although it can prevent infinite loops, which also leads to energy wastage since it is possible that a given packet reaches the same node more than once in GPER.

The size of time interval (i.e., T_{sleep}) in RIS has a great impact on the performance of GPER. Fig. 3.13 shows the energy consumption under different intervals where node sleeping probability is set to 0.4, and the simulated packet generation rates are $1 packet/5s$ and $1 packet/10s$. It can be seen that EBGR and BLR are independent of the size of T_{sleep} since the next-hop relay for each forwarder is detected in an on-line manner, and the number of data packets and RTS/CTS messages broadcasted for sensor-to-sink data delivery is proportional to the number of hops. For GPER, packet forwarding is based on the maintained the neighbor information. The longer the time interval is, the smaller the amount of
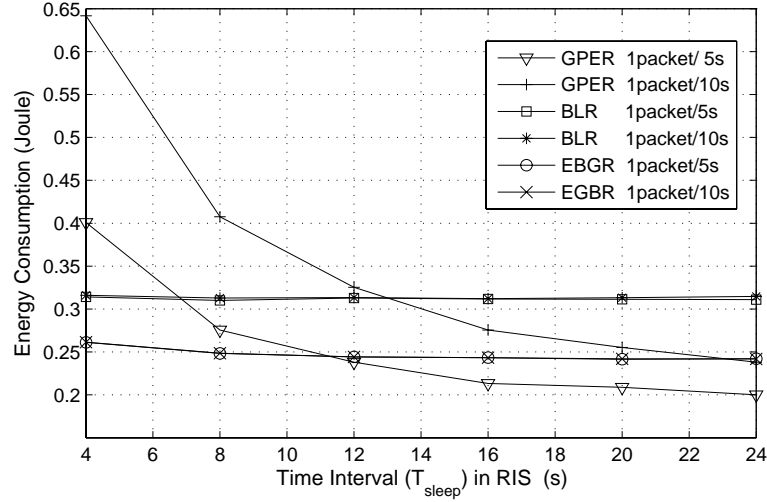
Figure 3.13: The total energy consumption under EBGR, BLR and GPER in random sleeping scenarios with different time interval

beacon messages broadcasted, and the larger the amount of data packets transmitted in one interval. As shown in Fig. 3.13, the total energy consumption under GPER decreases with the increase of T_{sleep}. When data generation rate is $1packet/5s$, EBGR outperforms GPER when T_{sleep} is less than $11s$. With the further increase of time interval, GPER consumes less energy than EBGR since less energy is spent on exchanging beacon messages. It is worth noting that GPER with a low data generation rate consumes much more energy than GPER with a high data generation rate because the running time with a low data generation rate is long and much more beacon messages need to be broadcasted. Thus, in contrast to GPER, EBGR is more suitable for event-detection applications in which data generation rate is very low.

3.7.4 Performance of EBGR-2 in High-variant Link Quality Scenarios

In this section, we evaluate the performance of EBGR-2 in scenarios where data transmission experiences frequent loss. The loss behavior for each link is modeled based on a realistic channel model proposed in [25]. In this model, two nodes exhibit full connectivity when the distance between them is below a distance D_1. Nodes are disconnected if they are at least distance D_2 away from each other. In the transitional region between D_1 and D_2, the expected reception rate decreases smoothly with some variation. The packet reception

rate for a link that has a distance d is computed as follows:

$$PRR(d) = \begin{cases} 1, & \text{if} \quad d < D_1; \\ [\frac{D_2-d}{D_2-D_1} + X]_0^1, & \text{if} \quad D_1 \leq d < D_2; \\ 0, & \text{if} \quad d \geq D_2. \end{cases} \quad (3.18)$$

where $[.]_a^b = \max\{a, \min\{b, .\}\}$ and $X \sim N(0, \sigma)$ being a Gaussian variable with variance σ^2. In our simulations, the parameters of this model are set as follows: $D_1 = 20$, $D_2 = 60$ and $\sigma = 0.3$. To demonstrate the energy efficiency of EBGR-2, we compare EBGR-2 with the $PRR \times DIST$ metric designed for routing in lossy sensor networks.
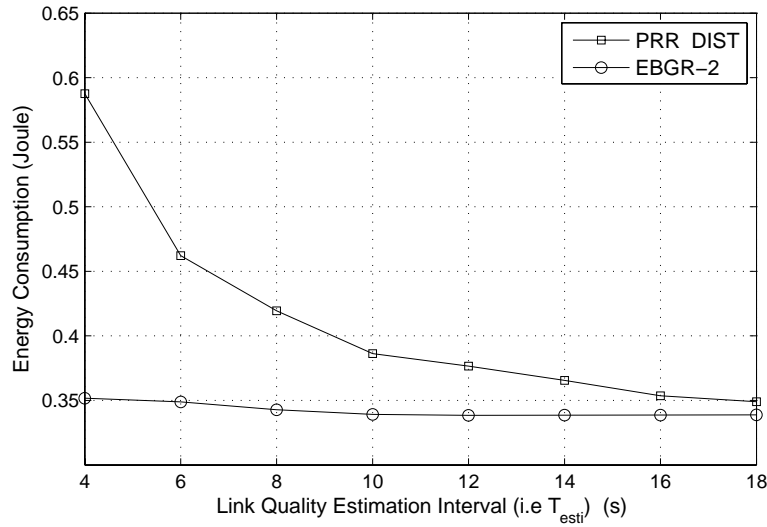
Figure 3.14: The total energy consumption under EBGR-2 and $PRR \times DIST$ with different link quality estimation interval

For the routing scheme based on $PRR \times DIST$ metric, each node broadcast beacon message to update the packet reception ratio for each link at the end of each link quality assessment interval. Fig. 3.14 shows the total energy consumption under EBGR-2 and $PRR \times DIST$ with the variation of different link quality assessment interval T_{esti}. With the increase of T_{esti}, the energy consumption under EBGR-2 keeps roughly constant, whereas the energy consumption under $PRR \times DIST$ decreases because the number of beacon messages broadcasted decreases with the increase of T_{esti}. Fig. 3.15 shows the energy spent on transmitting data packets with different link quality assessment interval. In contrast, $PRR \times DIST$ consumes much more energy than EBGR-2. In $PRR \times DIST$, each node
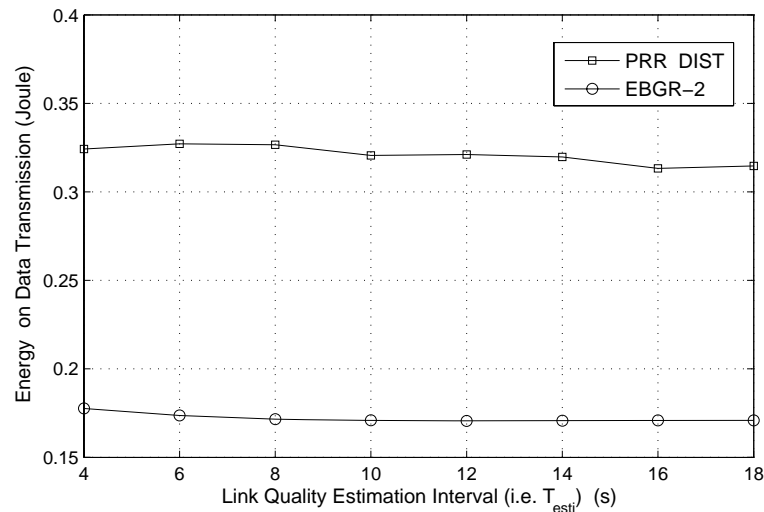
Figure 3.15: The energy consumption on data transmission under EBGR-2 and $PRR \times DIST$ with different link quality estimation interval

chooses its neighbor that maximizes the product of PRR and $DIST$ as its next hop relay. Thus each node tends to choose the neighbor with a large hop distance as its next hop relay, whereas a large hop distance means a large amount of energy consumption for each packet forwarding. As shown in Fig. 3.15, $PRR \times DIST$ consumes around 88% more energy on data transmission than EBGR-2.

3.8   Summary and Future Work

Providing energy-efficient routing is an important issue in the design of WSNs. In this chapter, we present a novel energy-efficient beaconless geographic routing protocol EBGR which takes advantages of both geographic routing and power-aware routing to provide loop-free, stateless and energy-efficient sensor-to-sink routing in dynamic WSNs. The performance of EBGR is evaluated through both theoretical analysis and simulations. We establish the bounds on hop count and the upper bound on energy consumption under EBGR for sensor-to-sink data delivery, assuming no packet loss and no failures in greedy forwarding. Furthermore, we demonstrate that the approximated expected energy consumption under EBGR keeps close to the lower bound when sensor nodes are densely deployed. To deal with the unreliable communication links in WSNs, we extend EBGR to provide

energy-efficient routing in lossy sensor networks. Extensive simulations have been performed to compare our schemes with the existing protocols. Simulation results show that our protocols save significantly more energy than routing protocols based on neighborhood maintenance in highly dynamic scenarios.

There are some interesting future research directions regarding the concept of energy-efficient geographical routing in WSNs. By taking the residual energy into account for making forwarding decision, our schemes can be extended to alleviate the unbalanced energy consumption in the network while still guaranteeing that the total energy consumption for sensor-to-sink data delivery is bounded. Another extension is to integrate other energy conserving schemes such as data aggregation to further reduce energy consumption and maximize network lifetime.

Chapter 4

Energy-Balanced Data Gathering

4.1 Introduction

In many applications of WSNs such as remote habitat monitoring, battlefield monitoring and environmental data (e.g., temperature, light, humidity, and vibration, etc.) collection, hundreds or even thousands of low-cost sensor nodes may be dispersed over the monitoring area. The sensors self-organize into a wireless network, termed data gathering sensor network, in which each sensor node must periodically report its sensed data to the sink(s). Sensor nodes in such large-scale data gathering sensor networks are generally powered by small inexpensive batteries in expectation of surviving for a long period. Therefore, energy is of utmost importance in power-constrained data gathering sensor networks, and energy consumption should be well managed to maximize the post-deployment network lifetime.

Sensor-to-sink direct transmission is the most easiest way for each sensor node to report its sensed data to the data sink(s) if the transmission range of the sensor nodes is large enough to reach a data sink[1]. However, in wireless communications, the transmission power is proportional to the square or quadruple of the transmission distance if the transmission power can be adjusted according to the transmission distance [112] (e.g., for Mica2 motes, the transmission power can be tuned from -20dBm to +10 dBm [2]). If each node employs power-adjusted sensor-to-sink direct transmission to report the sensed data, the nodes farther away from the sink run out of energy quickly due to the long transmission

---

[1]Mica2 motes (MPR410CB and MPR420CB) have a maximum outdoor range of $1000 ft$ [2]. SmartMesh-XT (M2135) motes have a maximum outdoor range of $400m$ [3].

distance. To save energy, multi-hop routing is more preferable than sensor-to-sink direct transmission. However, multi-hop routing schemes tend to overuse the nodes close to the sink and make them run out of energy quickly, leading to the existence of *energy holes* [75] around the sink(s). Experimental results in [77] show that, by the time the sensors close to the sink exhaust their energy budget, up to 90% of the initial budget may still be available in the nodes farthest away from the sink. Therefore, unbalanced energy consumption is an inherent problem for both direct transmission schemes and multi-hop routing schemes, and this unbalanced energy depletion can make the network collapse early due to the death of some critical nodes, resulting in significant network lifetime reduction.

Mixed-routing schemes, in which each node alternates between hop-by-hop transmission mode and direct transmission mode to transmit data, is an attractive scheme for balancing energy consumption due to their simplicity and effectiveness. In direct transmission mode, each node sends its data directly to the sink without any relay, and this mode helps to alleviate the relay burden for the nodes close to the sink. In hop-by-hop transmission mode, each node forwards the data to its next-hop neighbors, and this mode helps to relieve the burden of long distance transmission for the nodes far away from the sink. Therefore, it is possible to obtain fairly even energy consumption among all nodes by properly allocating the amount of data transmitted in the two modes. In [37], a slice model based on mixed-routing was proposed for balancing energy consumption in sensor network with uniform node deployment and uniform event generation rate. In this model, the network is divided into slices, and all nodes in each slice use the same probability to transmit data in direct transmission mode. The solution for computing the optimal probability of direct transmission for each slice in terms of balancing energy consumption among different slices was presented. However, this model only considers the energy consumption for transmitting data, and the energy consumption for receiving data was not taken into account. Moreover, the problem of balancing energy consumption among nodes within a slice and the problem of maximizing network lifetime through balancing energy consumption were not addressed in [37].

This chapter is focused on addressing the problems of balancing energy consumption and maximizing network lifetime for uniform data gathering WSNs. Uniform node deployment, which is also the basic assumption in [37] [90], seems to be a strong assump-

tion. However, in many applications, near-uniform node deployment is one of the easiest and most practical approach to provide full sensing coverage and connectivity. For example, in environment monitoring applications [6] [9], thousands of sensors can be deployed near-uniformly by dropping the sensors from an airplane to the monitoring area. In building monitoring applications [21], grid deployment schemes have been widely used. In [8], wireless sensor networks were deployed in a grid topology to a farm at Belmont for monitoring animal behaviors. Moreover, a lot of work [71] [83] [38] has been done on designing random uniform node deployment schemes to provide full coverage and connectivity in WSNs.

In this chapter, efficient schemes have been designed for solving the problems of balancing energy consumption and maximizing network lifetime by taking the advantages of corona-based network division, mixed-routing and data aggregation. Similar to the slice model in [37], the network is divided into coronas centered at the data sink. Based on corona-based network division, the energy-balanced data gathering problem is divided into two sub-problems: intra-corona energy consumption balancing and inter-corona energy consumption balancing. For the former sub-problem, a fully localized zone-based routing scheme is designed to balance energy consumption among nodes within each corona. For the latter sub-problem, an algorithm with time complexity $O(n)$ (n is the number of coronas) is designed to compute the optimal data distribution ratios for different coronas to ensure that the energy consumption among nodes in different coronas can be balanced. Different from the work in [37], this study considers energy consumption for both data transmission and reception. Moreover, data aggregation techniques, which have been regarded as efficient solutions for reducing energy consumption and alleviating energy unbalance, is employed in this study. The purpose of balancing energy consumption is to prolong network lifetime. The solution for computing the optimal number of coronas in terms of maximizing network lifetime is also presented. Based on the schemes proposed in this chapter, an energy-balanced data gathering protocol, called EBDG, is designed, and the solution for extending EBDG to large scale sensor networks is also presented.

The rest of this chapter is organized as follows. The related work on energy balancing in WSNs is discussed in Section 4.2. The system models and the statement of the problems to be solved in this chapter are described in Section 4.3. Section 4.4 presents the zone-based

routing scheme for balancing energy consumption among nodes in each corona. Section 4.5 describes the solution for balancing energy consumption among nodes in different coronas. In section 4.6, the problem of maximizing network lifetime is studied. Section 4.7 presents the design of EBDG and Section 4.8 gives the solution to extend EBDG to large-scale data gathering sensor networks. In Section 4.9, EBDG is evaluated through extensive simulations by comparing with several popular routing protocols. Finally, Section 4.10 gives the conclusion and future extensions of this work. For convenience, a list of the parameters used in this chapter is given in Appendix C.

4.2 Related Work

Cluster-heads rotation schemes have been proposed for distributing energy consumption fairly evenly in clustered sensor networks. LEACH [50] is a clustering-based routing protocol, in which randomized cluster-heads rotation is employed to balance energy depletion among nodes within clusters. In HEED [124], cluster-heads are periodically selected based on node residual energy and other parameters such as node proximity to its neighbors or node degrees. To balance the energy depletion among cluster-heads, schemes such as EECS [74] and UCS [106] were proposed. In these schemes, the network is divided into clusters with unequal size, and clusters closer to the base station have smaller sizes than those farther away from the base station. However, to achieve a desirable balance of energy depletion, cluster-head rotation must be performed frequently, which may add excessive communication overhead to the network, resulting in much energy wastage.

Mixed-routing schemes, in which each node alternates between direct transmission and multi-hop transmission to transmit data, was first proposed in [45]. In [37], a slice-model based on mixed-routing was proposed and a probabilistic data propagation algorithm was designed to balance energy depletion among different slices. However, energy consumption is only balanced among slices, and the problem of balancing energy consumption among nodes within the same slice was not addressed. For example, in [90] and [37], the authors simply assume that the nodes in the same corona (slice) have the same probability of receiving packets from outer coronas (slices). In [59], the authors proved that an energy-balanced mixed-routing scheme beats every other possible routing strategy in terms of net-

work lifetime maximization[2]. In [91], the authors gave a formal definition of an optimal data propagation algorithm with the objective to maximize network lifetime, and found a simple necessary and sufficient condition for the data propagation algorithm to be optimal. In [91], a spreading technique was also employed to balance energy consumption among sensors of a same slice. In [101], spreading schemes based on the energy histogram were designed for energy efficient routing in wireless sensor network. However, these spreading schemes can not guarantee balanced energy consumption among all nodes in the network. Different from existing work, in this study, balancing energy consumption and maximizing network lifetime are achieved by combining the advantages of corona-based network division, mixed-routing and data aggregation. Energy consumption is balanced among nodes both within a corona and among different corona.

4.3   System Models and Problems Statement

4.3.1   Network Model

Similar to the models in [37], [90] and [119], all sensor nodes are uniformly distributed in a circular monitoring area $\mathcal{A}$ of radius R with node distribution density ρ. There is only one sink which is located at the center of $\mathcal{A}$. Every node has the same maximum transmission range r_{max} and the same amount of initial energy budget. Each node has the knowledge of its location which can be easily obtained by equipping each sensor node with GPS-capable antenna. Even when GPS coordinates are not available, some localization algorithms [48] [100] can be pre-performed at the network setup phase since sensor nodes usually remain static in data gathering applications and such localization operation only needs to be performed once.

It is assumed that $r_{max} \geq R$ which ensures that each node can directly communicate with the sink. This assumption puts constraints on the size of the networks since sensor nodes usually have limited transmission range. In Section 4.8, a solution based on cluster-

---

[2]An natural extension of mixed-routing schemes is to allow each node sending its data to other intermediate nodes (e.g., 2-hop neighbors, 3-hop neighbors, etc.). In [59], the authors demonstrated that such complicated transmission scheme can not achieve more lifetime extension since forwarding the data to other intermediate nodes does not relive the traffic burden of the nodes close to the sink.

ing techniques is designed to remove this constraint so that the schemes designed in this chapter can be used in large-scale wireless sensor networks.

Similar to [42], the operation of data gathering is divided into *rounds*. In each round, all nodes wake up, generate the data, perform data aggregation and send the data to the sink. Between two adjacent rounds, all nodes turn off their radios to save energy. It is assumed that all nodes have the same data generation rate and the amount of data generated by every node in each round is l bits.

4.3.2 Aggregation Model

A general aggregation model proposed in [85] is employed. In this model, the amount of data outputted, denoted by $\varphi(x)$, is a linear function of the amount of input data x.

$$\varphi(x) = mx + c. \tag{4.1}$$

This model captures the following aggregation scenarios depending on the values of m and c.

- If $m = 0$, $c > 0$, the model corresponds to the scenario in which each node can aggregate the data it generates and receives into a single packet of fixed size. This case is suitable to applications with aggregation operations such as *min* or *max* (e.g. temperature and humidity), *sum* (e.g. event count), and *yes-no* (e.g. intrusion detection).

- If $0 < m < 1$, $c = 0$, the model corresponds to the scenario in which all sensor nodes can compress the data they generate and receive by a factor of m.

- If $m = 1$, $c = 0$, the model corresponds to the scenario in which all nodes do not perform any data aggregation.

4.3.3 Energy Model

This work is based on the *First Order Radio Model* which has been described in Section 3.3.2 of Chapter 3. Each sensor node can adjust its transmission power according to the

transmission distance. The energy for transmitting one bit data over distance x is given by $\epsilon_t(x) = a_{11} + a_2 x^k$, where a_{11} is the energy spent by transmitter electronics, a_2 is the transmitting amplifier and k $(k \geq 2)$ is the propagation loss exponent. The energy for receiving one bit data is given by $\epsilon_r = a_{12}$, where a_{12} is the energy spent by receiver electronics. In this study, it is assumed that $a_{11} = a_{12}$. The energy consumption for data sensing is not taken into account since all nodes have uniform data generation rate and the energy spent by sensing data has been balanced among all nodes. Compared with data communication, the energy spent by data aggregation is much smaller, and is not taken into account in this study.

4.3.4   Problems Statement
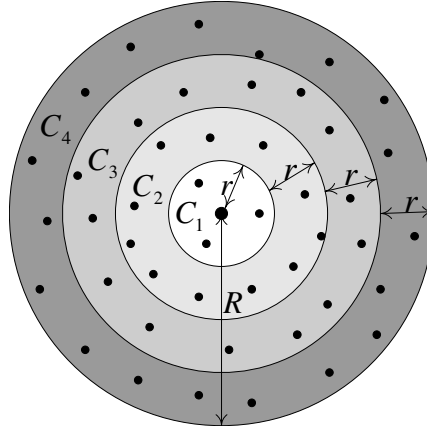
Figure 4.1: Illustration of network division (n=4)

The circular monitoring area $\mathcal{A}$ is divided in to n concentric coronas C_1, C_2, ..., C_n centered at the sink with the same width r $(r = \frac{R}{n})$, as shown in Figure 4.1. To balance energy consumption, each node alternates between two transmission modes: hop-by-hop transmission and direct transmission. For any node $u \in C_i$, it forwards its data to a neighbor in C_{i-1} when hop-by-hop transmission mode is used, and transmits its packets directly to the sink when direct transmission mode is used. Since direct transmission consumes much more energy than hop-by-hop transmission mode, balanced energy consumption might be achieved by optimally distributing the amount of data for hop-by-hop and direct transmissions at each node.

Let T be the total number of data gathering rounds performed during the entire network lifetime. For any node u, $F(u)$ and $D(u)$ denote the amount of data **Forwarded** in hop-by-hop transmission mode and transmitted in **Direct** transmission mode in T rounds of data gathering respectively.

Definition 3. *The data distribution ratio for node u, denoted by $p(u)$, is defined as the ratio of the amount of data transmitted by u in direct transmission mode to the total amount of data transmitted by u in two transmission modes, i.e.,*

$$p(u) = \frac{D(u)}{F(u) + D(u)}. \tag{4.2}$$

Similar to the slice model in [37] and [45], all nodes in corona C_i have the same data distribution ratio, denoted by p_i. Since the maximum distance from a node in C_i to the data sink is ir, it is assumed that all nodes in C_i spend the same energy $\varepsilon_t(ir)/bit$ for direct transmission. It is also assumed that all nodes in the network use the same energy $\varepsilon_t(r')/bit$ for hop-by-hop transmission (r' is larger than r in order to balance energy consumption among nodes in the same corona, which is explained in detail in Section 4.4.3). Based on this model, this chapter is focused on solving the following problems.

1. How to balance the energy consumption among nodes in the same corona? This problem is referred to as Intra-Corona Energy Consumption Balancing (Intra-CECB) and is solved in Section 4.4.

2. How to compute the optimal data distribution ratio p_i for each corona C_i so that energy consumption can be balanced among nodes in different coronas? This problem is referred to as Inter-Corona Energy Consumption Balancing (Inter-CECB) and is solved in Section 4.5.

3. What is the optimal number of coronas n in terms of maximizing the network lifetime? This problem is referred to as Network Lifetime Maximization (NLM) and is solved in Section 4.6.

4.4   Intra-Corona Energy Consumption Balancing

This section is focused on solving the Intra-CECB problem. The sufficient and necessary condition for Intra-CECB is presented first. It is proven that the energy consumption among nodes in each corona can be balanced if and only if the amount of data received by nodes in this corona is balanced. Based on this observation, a localized zone-based routing scheme is designed to guarantee balanced energy consumption among nodes in each corona.

4.4.1   Sufficient and Necessary Condition for Intra-CECB

For every node $u \in C_i$, let $S(u)$ and $E(u)$ denote the total amount of data received and the total amount of energy spent by node u in T rounds of data gathering respectively. Using the energy model given in Section 4.3.3,

$$E(u) = F(u) \cdot \varepsilon_t(r') + D(u) \cdot \varepsilon_t(ir) + S(u) \cdot a_{11}, \qquad (4.3)$$

where the first term $F(u) \cdot \varepsilon_t(r')$ represents the energy spent by hop-by-hop transmissions, the second term $D(u) \cdot \varepsilon_t(ir)$ denotes the energy spent by direct transmissions and the last term $S(u) \cdot a_{11}$ is the energy consumption for receiving data.

Obviously, energy consumption is balanced among all nodes in C_i only when

$$E(u) = E(v), \qquad \forall u \in C_i \quad \forall v \in C_i, \quad 1 \leq i \leq n.$$

Lemma 8. $\forall u \in C_i, \forall v \in C_i$, $F(u) = F(v)$ and $D(u) = D(v)$ if and only if $S(u) = S(v)$.

Proof. Since all nodes in the network have the same data generation rate ($l\ bits/round$), the total amount of data generated by each node in T rounds of data gathering is Tl. By the aggregation model given in Section 4.3.2, the total amount of data transmitted by node u after aggregation is $\varphi(Tl + S(u))$. By Definition 3,

$$D(u) = p_u \cdot (F(u) + D(u)) = p_u \cdot \varphi(Tl + S(u)); \qquad (4.4)$$

$$F(u) = (1 - p_u) \cdot (F(u) + D(u)) = (1 - p_u) \cdot \varphi(Tl + S(u)). \qquad (4.5)$$

Since all nodes in C_i have the same data distribution ratio p_i, therefore, $F(u) = F(v)$ and $D(u) = D(v)$ if and only if $S(u) = S(v)$. $\quad\square$

Theorem 6. *Energy consumption is balanced among nodes in C_i if and only if the amount of data received by nodes in C_i is balanced.*

Proof. By substituting $D(u)$ and $F(u)$ in Equation (4.3),

$$E(u) = \varphi(Tl + S(u)) \cdot [(1 - p_i) \cdot \varepsilon_t(r') + p_i \cdot \varepsilon_t(ir)] + S(u) \cdot a_{11}.$$

Clearly, $E(u)$ only depends on $S(u)$ because all the other parameters are the same for all nodes in C_i. Thus energy consumption is balanced among nodes in C_i if and only if all the nodes in C_i receive the same amount of data. $\quad\square$

By Theorem 6, it can be observed that the Intra-CECB problem can be simplified as the problem of balancing the amount of data received by nodes within each corona.

Suppose that all nodes in C_i receive the same amount of data in T rounds of data gathering. From Lemma 8, all nodes transmit the same amount of data in hop-by-hop transmission mode, and also the same amount of data in direct transmission mode. For simplicity of notation, let F_i and D_i denote the amount of data transmitted by each node in C_i through hop-by-hop transmission mode and direct transmission mode respectively. S_i represents the amount of data received by each node in C_i and E_i denotes the energy depleted by each node in C_i. The following lemma holds.

Lemma 9.
$$\frac{S_i}{F_{i+1}} = \frac{2i + 1}{2i - 1}, \qquad 1 \le i < n.$$

Proof. Let N_{C_i} and $N_{C_{i+1}}$ be the number of nodes in C_i and C_{i+1} respectively. Then

$$N_{C_i} = \rho[\pi(ir)^2 - \pi((i-1)r)^2] = \rho\pi r^2(2i - 1);$$
$$N_{C_{i+1}} = \rho[\pi((i+1)r)^2 - \pi(ir)^2] = \rho\pi r^2(2i + 1).$$

Since $F_{i+1} \cdot N_{C_{i+1}} = S_i \cdot N_{C_i}$,

$$\frac{S_i}{F_{i+1}} = \frac{N_{C_{i+1}}}{N_{C_i}} = \frac{2i + 1}{2i - 1}.$$

□

4.4.2   Zone-based Routing Scheme

The basic idea of the zone-based routing scheme is described as follows: each corona is divided into sub-coronas and each sub-corona is further divided into zones, as shown in Figure 4.2. There is a one-to-one mapping between the zones in two adjacent corona. When hop-by-hop mode is used, data communication is performed between nodes in two corresponding zones. The objective is to design the optimal network division scheme so that the amount of data received by nodes in each corona can be balanced. By Theorem 6, the energy consumption among nodes in each corona is balanced if the amount of data received by nodes in this corona is balanced.
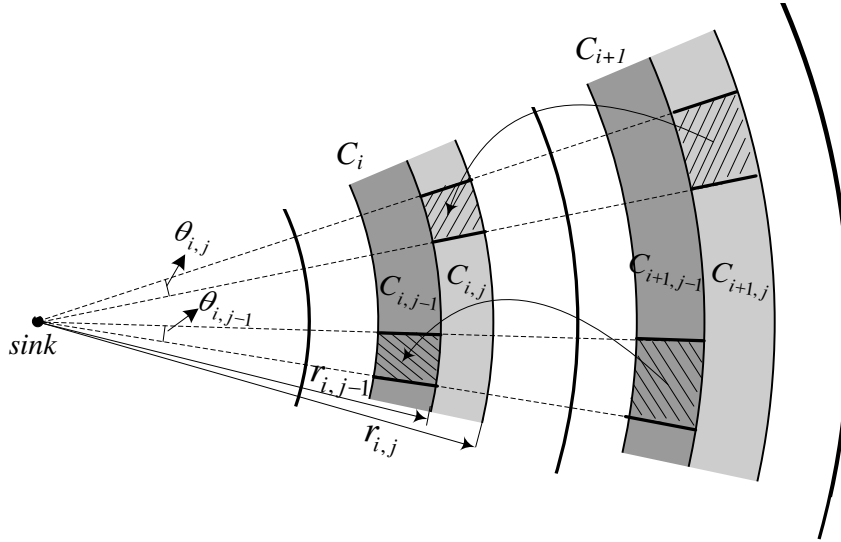
Figure 4.2: The partition of coronas C_i and C_{i+1}

Consider any two adjacent coronas C_i and C_{i+1} where nodes in C_{i+1} forward their data to nodes in C_i when hop-by-hop transmission mode is used. C_{i+1} is termed as *source* corona and C_i is termed as *destination* corona. As shown in Figure 4.2, the *source* corona C_{i+1} is divided into w sub-coronas $C_{i+1,1}$, $C_{i+1,2}$,..., $C_{i+1,w}$ with equal width $\frac{r}{w}$, and the *destination* corona C_i is divided into w sub-coronas $C_{i,1}$, $C_{i,2}$,..., $C_{i,w}$ with width $\Delta_{i,1}$, $\Delta_{i,2}$,..., $\Delta_{i,w}$ respectively. Each sub-corona $C_{i,j}$ is further divided into h_j equal-size zones $Z_{i,j,1}$, $Z_{i,j,2}$,..., Z_{i,j,h_j} in such a way that $Z_{i,j,k}$ and $Z_{i+1,j,k}$ are the regions $C_{i,j}$ and $C_{i+1,j}$

overlap with a circular sector centered at the sink with radius R and central angle $\theta_{i,j} = (\theta_2 - \theta_1)$, where $\theta_1 = \frac{2\pi(k-1)}{h_j}$ and $\theta_2 = \frac{2\pi k}{h_j}$ (θ_1 and θ_2 are the angles of the two radii of the sector in polar coordinates). For each destination zone, it covers λ nodes where λ is a user-defined parameter. Obviously, different sub-coronas are divided into different number of zones (i.e., h_j for sub-corona $C_{i,j}$ is unequal to h_k for sub-corona $C_{i,k}$). In this network division scheme, all destination zones have the same size and all source zones also have the same size.

In zone-based routing scheme, $Z_{i+1,j,k}$ is referred to as a *source* zone, and $Z_{i,j,k}$ is referred to as the corresponding *destination* zone for $Z_{i+1,j,k}$. When hop-by-hop transmission mode is used, each node u in $Z_{i+1,j,k}$ forwards its data to all nodes in $Z_{i,j,k}$ with equal probability. To achieve this, each node maintains the location information of the nodes in its destination zone. Since the number of nodes in each destination zone is small, this scheme does not add too much complexity to the zone-based routing scheme.

It is easy to see that the above approach can evenly distribute the data forwarded by nodes in $Z_{i+1,j,k}$ among nodes in $Z_{i,j,k}$. In the following lemma, it is proven that the amount of data received by all nodes in each sub-corona $C_{i,j}$ is balanced in the zone-based routing scheme.

Lemma 10. *For $\forall u \in C_{i,j}$, $\forall v \in C_{i,j}$, $S(u) = S(v)$, $\quad 1 \leq i \leq n$, $1 \leq j \leq m$.*

Proof. Let $A_{Z_{i,j,k}}$ and $A_{C_{i,j}}$ be the area size of $Z_{i,j,k}$ and $C_{i,j}$ respectively. By Equation (4.5), the amount of data forwarded by each node in C_n to nodes in C_{n-1} is the same and equal to $(1 - p_n)\varphi(Tl)$ because nodes in C_n do not receive any data. For any $u \in Z_{n-1,j,k}$, by the definition of $Z_{n-1,j,k}$,

$$S(u) = \frac{A_{Z_{n,j,k}}}{A_{Z_{n-1,j,k}}} \cdot (1 - p_n)\varphi(Tl) = \frac{A_{C_{n,j}}}{A_{C_{n-1,j}}} \cdot (1 - p_n)\varphi(Tl).$$

Therefore, the amount of data received by each node in $C_{n-1,j}$ is balanced.

Suppose that the amount of data received by all nodes in $C_{i,j}$ is balanced and equal to $S_{i,j}$. For any $u \in Z_{i-1,j,k}$, by Equation (4.5) and the definition of $Z_{i-1,j,k}$,

$$S(u) = \frac{A_{Z_{i,j,k}}}{A_{Z_{i-1,j,k}}} \cdot (1 - p_i)\varphi(Tl + S_{i,j}) = \frac{A_{C_{i,j}}}{A_{C_{i-1,j}}} \cdot (1 - p_i)\varphi(Tl + S_{i,j}).$$

Obviously, all the nodes in $C_{i-1,j}$ also receive the same amount of data. Hence, the lemma is proved. □

As shown in Fig. 4.2, let $r_{i,j}$ be the distance between the sink and the outer border of $C_{i,j}$, i.e., $r_{i,j} = (i-1)r + \sum_{k=1}^{j} \Delta_{i,k}$ ($0 \leq j \leq m$). Clearly, $r_{i,0} = (i-1)r$ and $r_{i,m} = ir$. The following theorem gives the condition that guarantees balanced data receiving among all nodes in corona C_i.

Theorem 7. *The amount of data received by nodes in C_i is balanced in the zone-based routing scheme if*

$$r_{i,j} = r \cdot \sqrt{(i-1)^2 + \frac{2i-1}{2i+1} \cdot (\frac{j^2}{m^2} + \frac{2ij}{m})}. \tag{4.6}$$

Proof. From Lemma 10, the amount of data received by nodes in each sub-corona is balanced in zone-based routing scheme. By Lemma 9, the amount of data received by all nodes in each corona C_i can be balanced if

$$\frac{A_{C_{i+1,1}}}{A_{C_{i,1}}} = \frac{A_{C_{i+1,2}}}{A_{C_{i,2}}} = ... = \frac{A_{C_{i+1,j}}}{A_{C_{i,j}}} = ... = \frac{A_{C_{i+1,m}}}{A_{C_{i,m}}} = \frac{2i+1}{2i-1}.$$

For $1 \leq j \leq m$,

$$\frac{\sum_{k=1}^{j} A_{i+1,k}}{\sum_{k=1}^{j} A_{i,k}} = \frac{\pi(ir + \frac{jr}{m})^2 - \pi(ir)^2}{\pi r_{i,j}^2 - \pi(i-1)^2 r^2} = \frac{2i+1}{2i-1}.$$

Therefore,

$$r_{i,j} = r \cdot \sqrt{(i-1)^2 + \frac{2i-1}{2i+1} \cdot (\frac{j^2}{m^2} + \frac{2ij}{m})}.$$

□

Let $\phi_{i,j-1}$ and $\phi_{i+1,j}$ be the circles with radii $r_{i,j-1}$ and $ir + \frac{jr}{m}$ respectively. Obviously, $\phi_{i,j-1}$ is the inner boundary of $C_{i,j}$ and $\phi_{i+1,j}$ is the outer boundary of $C_{i+1,j}$. Let $\beta_{i+1,i,j}$ be the width of the corona delimited by circles $\phi_{i,j-1}$ and $\phi_{i+1,j}$. The following corollary holds.

Corollary 3. $\beta_{i+1,i,j} \leq (1 + \frac{1}{m})r, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m.$

Proof. Let $\alpha = \frac{j}{m}$ where $0 \le j \le m$. Therefore, $0 \le \alpha \le 1$. For $\forall i > 0$, it can be proved that $\frac{2i-1}{2i+1}(\alpha^2 + 2i\alpha) > 2(i-1)\alpha + \alpha^2$. By Equation (4.6),

$$
\begin{aligned}
r_{i,j} &= r \cdot \sqrt{(i-1)^2 + \frac{2i-1}{2i+1} \cdot (\alpha^2 + 2i\alpha)} \\
&\ge r \cdot \sqrt{(i-1)^2 + 2(i-1)\alpha + \alpha^2} = (i-1+\alpha)r,
\end{aligned}
$$

where the equality is held only when $j = 0$ and $j = m$.

$$
\begin{aligned}
\beta_{i+1,i,j} &= ir + \frac{jr}{m} - r_{i,j-1} \\
&\le ir + \frac{jr}{m} - (i-1)r + \frac{(j-1)r}{m} = (1 + \frac{1}{m})r,
\end{aligned}
$$

where the equality is held when $j = 1$. $\qquad\qquad\square$

4.4.3 Hop-by-hop Transmission Range

In the zone-based routing scheme, the zones are created based on the following rule: each *destination* zone covers λ nodes. As shown in Figure 4.3, the optimal hop-by-hop transmission range, denoted by r'_w, should be the minimum distance that guarantees any node in any *source* zone can transmit its data to any node in its corresponding *destination* zone. Therefore, r'_w must be larger than r. The following lemma gives the range of r'_w, and the optimal r'_w in terms of maximizing network lifetime is derived in Section 4.6.
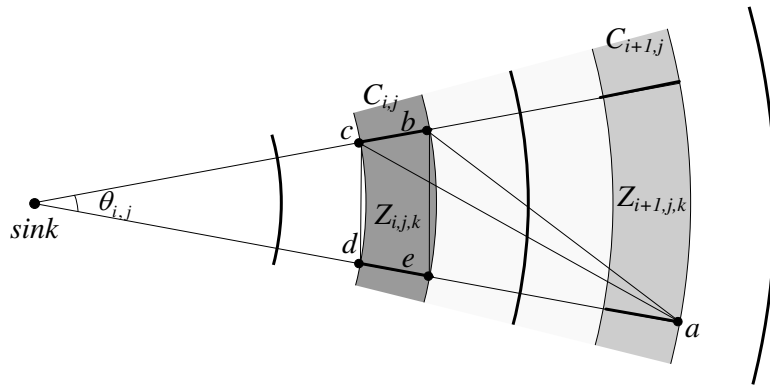
Figure 4.3: Illustration of the range of r'_w

Lemma 11. $(1 + \frac{1}{w})r < r'_w < (1 + \frac{1}{w})r + \frac{3\lambda w}{\rho r}$.

Proof. As shown in Figure 4.3, the minimum transmission distance that guarantees all nodes in $Z_{i+1,j,k}$ can communicate with any node in $Z_{i,j,k}$ is either $|ac|$ or $|ab|$ ($|ab| > |ac|$ when $\theta_{i,j}$ is relatively large). By Corollary 3,

$$r'_w \geq max\{|ab|, |ac|\} > max|ad| = (1 + \frac{1}{w})r.$$

Since

$$|ac| < |ad| + |cd| < |ad| + |be|;$$
$$|ab| < |ae| + |be| < |ad| + |be|,$$

therefore, $r'_w < max\{|ad| + |be|\}$. Let $\widehat{be}$ be the length of the small arc from b to e. By Lemma 9, Since $Z_{i+1,j,k}$ covers $\frac{\lambda(2i+1)}{2i-1}$ nodes because $Z_{i,j,k}$ covers λ nodes. Therefore, $\rho A_{Z_{i+1,j,k}} = \frac{\lambda(2i+1)}{2i-1}$, i.e., $\theta_{i,j} = \frac{2\lambda w^2(2i+1)}{\rho r^2(2iw+2j-1)(2i-1)}$.

$$
\begin{aligned}
|be| \quad < \quad \widehat{be} &= r_{i,j} \cdot \frac{2\lambda w^2(2i+1)}{\rho r^2(2iw+2j-1)(2i-1)} < ir \cdot \frac{2\lambda w^2(2i+1)}{\rho r^2(2iw+1)(2i-1)} \\
&< \quad \frac{\lambda w}{\rho r} \cdot \frac{2iw(2i+1)}{(2iw+1)(2i-1)} < \frac{3\lambda w}{\rho r}.
\end{aligned}
$$

Since $|ad| \leq (1 + \frac{1}{w})r$, $r'_w < max\{|ad| + |be|\} < (1 + \frac{1}{w})r + \frac{3\lambda w}{\rho r}$. $\qquad\square$

4.5 Inter-Corona Energy Consumption Balancing

This section concentrates on solving the Inter-CECB problem. By Lemma 8 and Theorem 6, all nodes in C_i transmit the same amount of data F_i through hop-by-hop transmission mode, transmit the same amount of data D_i through direct transmission mode, and receive the same amount of data S_i from nodes in C_{i+1} when energy consumption is balanced among nodes in C_i. Let $f_i = \frac{F_i}{T}$, $d_i = \frac{D_i}{T}$, and $s_i = \frac{S_i}{T}$. The network can be mapped onto a linear network model, as shown in Figure 4.4.

Based on Equation (4.1),

$$f_i + d_i = m(s_i + l) + c. \tag{4.7}$$
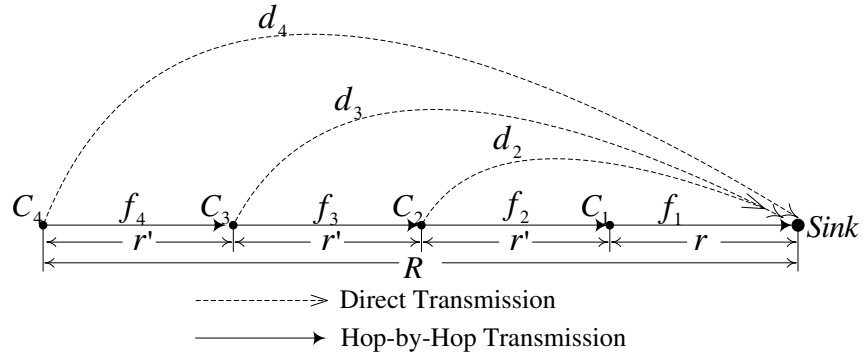
Figure 4.4: Mapping the network onto linear model

By Definition 3,

$$p_i = \frac{D_i}{F_i + D_i} = \frac{d_i \cdot T}{f_i \cdot T + d_i \cdot T} = \frac{d_i}{f_i + d_i} = \frac{d_i}{m(s_i + l) + c}. \tag{4.8}$$

Let $e_i = \frac{E_i}{T}$. By Equation (4.3),

$$
\begin{aligned}
e_i &= \frac{F_i}{T} \cdot \varepsilon_t(r') + \frac{D_i}{T} \cdot \varepsilon_t(ir) + \frac{S_i}{T} \cdot a_{11} \\
&= f_i \cdot \varepsilon_t(r') + d_i \cdot \varepsilon_t(ir) + s_i \cdot a_{11}.
\end{aligned} \tag{4.9}
$$

Obviously, the Inter-CECB problem can be formulated as the following transmitting data distribution problem.

$$
\begin{cases}
\text{Compute} \quad p_i, \quad 1 \le i \le n; \\
s.t. \quad e_1 = e_2 = ... = e_i = ... = e_{n-1} = e_n.
\end{cases} \tag{4.10}
$$

4.5.1 Optimal Data Distribution Ratio Allocation

Since balanced energy consumption might be achieved by optimally distributing the amount of data for hop-by-hop and direct transmissions at each node, this section is focused on computing the optimal data distribution ratio for each corona.

Lemma 12.

$$
s_i =
\begin{cases}
0, & i = n; \\
(m(s_{i+1} + l) + c - d_{i+1}) \cdot \frac{2i+1}{2i-1}, & 1 \le i < n.
\end{cases}
$$

Proof. If $i = n$, C_n is the outermost corona. Since all nodes in C_n do not receive any packets during each round, thus $s_n = 0$.

If $1 \leq i < n$, the nodes in C_i must relay the data received from nodes in C_{i+1}. By Lemma 9,

$$s_i = \frac{F_{i+1}}{T} \cdot \frac{2i+1}{2i-1} = f_{i+1} \cdot \frac{2i+1}{2i-1}.$$

By Equation (4.7), $f_i = m(s_i + l) + c - d_i$. Therefore, $s_i = (m(s_{i+1} + l) + c - d_{i+1}) \cdot \frac{2i+1}{2i-1}$. □

Lemma 13.

$$d_i = \frac{d_{i+1}[\varepsilon_t((i+1)r) - \varepsilon_t(r')] - (s_i - s_{i+1})(m\varepsilon_t(r') + a_{11})}{\varepsilon_t(ir) - \varepsilon_t(r')}, \quad where \quad 2 \leq i < n.$$

Proof. When $i > 1$, all nodes use the same transmission range r' for hop-by-hop transmission, and energy consumption is balanced among these coronas only when $e_{i+1} = e_i$. Thus

$$f_{i+1} \cdot \varepsilon_t(r') + d_{i+1} \cdot \varepsilon_t((i+1)r) + s_{i+1} \cdot a_{11} = f_i \cdot \varepsilon_t(r') + d_i \cdot \varepsilon_t(ir) + s_i \cdot a_{11}.$$

i.e.,

$$(f_i - f_{i+1}) \cdot \varepsilon_t(r') + (s_i - s_{i+1}) \cdot a_{11} = d_{i+1} \cdot \varepsilon_t((i+1)r) - d_i \cdot \varepsilon_t(ir). \qquad (4.11)$$

From Equation (4.7),

$$f_i - f_{i+1} = m(s_i - s_{i+1}) - (d_i - d_{i+1}).$$

By substituting $f_i - f_{i+1}$ in Equation (4.11),

$$d_i = \frac{d_{i+1}[\varepsilon_t((i+1)r) - \varepsilon_t(r')] - (s_i - s_{i+1})(m\varepsilon_t(r') + a_{11})}{\varepsilon_t(ir) - \varepsilon_t(r')}. \qquad (4.12)$$

□

For all nodes in C_1, only the direct transmission mode is used since all packets are directly transmitted to the sink. By Definition 3, $p_1 = 1$.

Lemma 14.

$$d_2 = \frac{(s_1 - s_2) \cdot a_{11} + m(s_1 \varepsilon_t(r) - s_2 \varepsilon_t(r')) + (ml + c)(\varepsilon_t(r) - \varepsilon_t(r'))}{\varepsilon_t(2r) - \varepsilon_t(r')}. \quad (4.13)$$

Proof. Since all nodes in C_1 transmit the data directly to the sink, therefore, $e_1 = [m(s_1 + l) + c] \cdot \varepsilon_t(r)$. By $e_1 = e_2$,

$$f_2 \cdot \varepsilon_t(r') + d_2 \cdot \varepsilon_t(2r) + s_2 \cdot a_{11} = [m(s_1 + l) + c] \cdot \varepsilon_t(r) + s_1 \cdot a_{11}.$$

By Equation (4.7), $f_2 = m(s_2 + l) + c - d_2$. Therefore,

$$d_2 = \frac{(s_1 - s_2) \cdot a_{11} + m(s_1 \varepsilon_t(r) - s_2 \varepsilon_t(r')) + (ml + c)(\varepsilon_t(r) - \varepsilon_t(r'))}{\varepsilon_t(2r) - \varepsilon_t(r')}.$$

$\square$
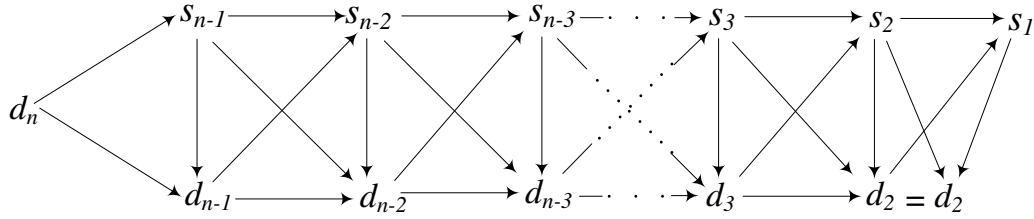
Figure 4.5: Illustration of the iterative process for computing p_i

From Lemma 12, s_i can be expressed by s_{i+1} and d_{i+1}. From Lemma 13, it can be seen that d_i $(2 \le i < n)$ can be expressed by s_i, s_{i+1} and d_{i+1}. As illustrated in Figure 4.5, both s_i and d_i $(2 \le i < n)$ can be expressed by d_n after a few iterations. By Lemma 14, d_2 can also be expressed by d_n because both s_1 and s_2 are functions of d_n. Therefore, d_n can be computed by making the d_2 computed by Equation (14) equal to the d_2 computed by Equation (4.12). Once d_n is computed, s_i and d_i can be obtained by backward computation.

Based on Equation (4.8), p_i can be finally obtained. The detailed algorithm for computing the data distribution ratios for all coronas is given as follows.

Algorithm: DDRC (Data Distribution Ratio Computation)

1:   For i:=n-1 downto 2 do
2:       Express s_i by d_n;   //according to Lemma 12
3:       Express d_i by d_n;   //according to Lemma 13
4:   Express d_2 by d_n;   // according to Lemma 14
5:   Compute d_n by making d_2 in Equation (4.12) equal to d_2 in Equation (4.13);
6:   For i:=2 to n do
7:       Compute d_i and s_i; //according to Lemma 12 and Lemma 13
8:       $p_i = \frac{d_i}{m(s_i+l)+b}$;

Obviously, the time complexity of Algorithm DDRC is $O(n)$ where n is the number of coronas.

4.5.2   Numerical Results and Analysis

To understand the relationship among data distribution ratios of different coronas and the relationship between data distribution ratios and data aggregation, some numerical results are given. The network parameters are set as follows: the radius of the circular monitored area is set to $500m$, and the area is divided into 50 coronas with equal width of $10m$. The hop-by-hop transmission range is set to $15m$. The parameters in the energy model are set as follows: $a_{11} = 50nJ/bit$, $a_2 = 10pJ/bit$ and $k = 2$.

Figure 4.6 plots the data distribution ratios for all coronas with different aggregation parameters. For all cases, with the increase of corona ID, the data distribution ratio first decreases rapidly, then remains more or less constant in the middle coronas and finally increases at the last few coronas. This behavior can be explained based on the fact that energy depleted by a transmission is proportional to the square of the transmission distance. For the nodes close to the sink, direct transmission mode is frequently used due to the small direct transmission distance. For the nodes in the last several coronas, although the direct transmission distance is large, the data distribution ratio increases with the increase of corona ID because nodes in such coronas have lighter data transmission burden than nodes in the middle coronas and nodes in the coronas close to the sink.

From Figure 4.6, it is also easy to see that data aggregation has a great impact on the
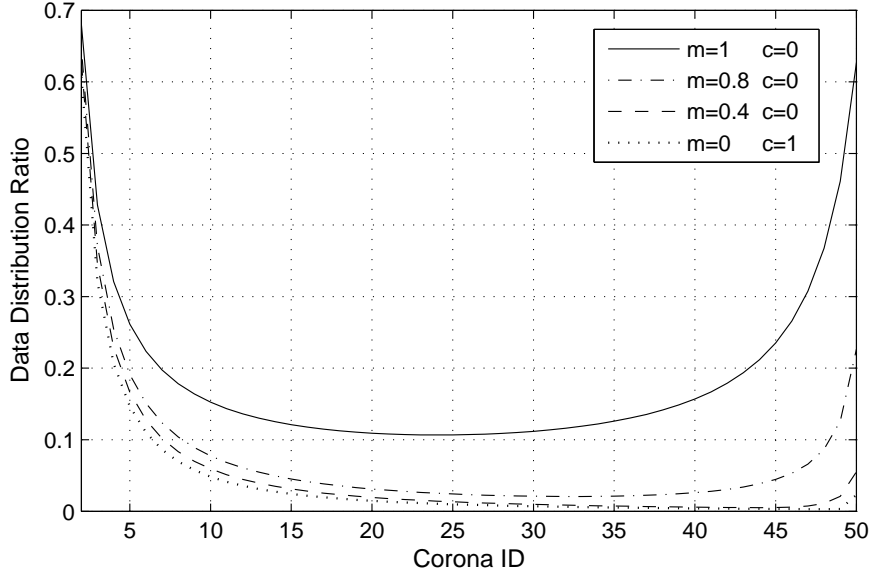
Figure 4.6: The data distribution ratios for different coronas with different aggregation parameters

distribution ratios. When m is small, the data sent after compression is small and most data are transmitted in hop-by-hop transmission mode which is more energy-efficient than direct transmission. Therefore, data aggregation is an effective approach for mitigating unbalanced energy consumption in wireless sensor networks.

4.6 Network Lifetime Maximization

This section is focused on solving the problem of maximizing network lifetime by balancing energy consumption among all nodes in the network. For nodes in outermost corona C_n,

$$
\begin{aligned}
e_n &= f_n \cdot \varepsilon_t(r') + d_n \cdot \varepsilon_t(nr) \\
&= (ml + c - d_n) \cdot (a_{11} + a_2 \cdot (r')^k) + d_n \cdot (a_{11} + a_2 \cdot R^k).
\end{aligned} \tag{4.14}
$$

When energy consumption is balanced, all nodes have the same energy consumption equal to e_n. Thus the NLM problem can be formulated as the following optimization problem.

$$
\begin{cases}
Min & e_n; \\
s.t. & e_i = e_j, \quad 1 \leq i, j \leq n; \\
& nr = R; \\
& r \geq \sqrt{\frac{\lambda}{\rho\pi}}.
\end{cases}
$$

where the first constraint ensures that energy is balanced among all nodes, and the second constraint ensures that the network is divided into n coronas with equal width; while the last constraint ensures that the innermost corona C_1 covers at least λ nodes since each destination zone in zone-based routing scheme must contain λ nodes.

4.6.1 Optimal Number of Sub-Coronas

Given n coronas, p_i and d_i can be computed by Algorithm DDRC. By Equation (4.14), e_n is minimized only when r' is minimized because all the other parameters in Equation (4.14) except r' are fixed for given n. Based on Lemma 11, the optimal hop-by-hop transmission range r'_w depends on the number of sub-coronas w. As shown in Fig. 4.7, each zone can be approximated as a rectangular. If w is too small, r'_w becomes large since $r'_w > (1 + \frac{1}{w})r$. On the other hand, if w is too large, the width of each sub-corona is very small, and each zone becomes a long and narrow strip, which also results in a larger r'_w. Therefore, there is an optimal number of sub-coronas in terms of minimizing r'.

Let w_o be the optimal sub-corona number and r'_o be the optimal hop-by-hop transmission distance in terms of minimizing r'.

Theorem 8. $w_o \approx w^*$, $r'_o \approx \sqrt{\frac{(w^*+1)^2 r^2}{(w^*)^2} + (\frac{3\lambda(3w^*-1)}{4\rho r})^2}$ *where* $\frac{(w^*)^3(3w^*-1)}{w^*+1} = \frac{16\rho^2 r^4}{27\lambda^2}$.

Proof. As shown in Figure 4.7,

$$
|ac| = \sqrt{|ad|^2 + |cd|^2} = \sqrt{|ad|^2 + (\frac{|ab| + |ec|}{2})^2}.
$$

When each corona is divided into w^* sub-coronas, each zone should not be long and narrow.
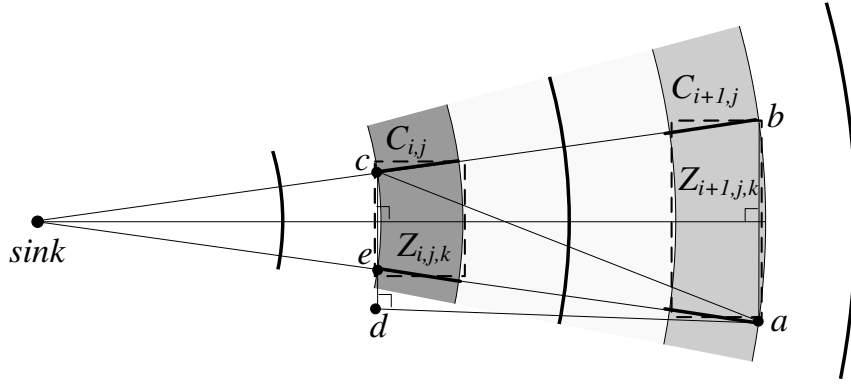
Figure 4.7: Approximated hop-by-hop transmission range

By approximating $Z_{i+1,j,k}$ as a rectangular,

$$\rho \cdot |ab| \cdot \frac{r}{w} \approx \frac{\lambda(2i+1)}{2i-1}.$$

Thus $|ab| \approx \frac{\lambda w(2i+1)}{\rho r(2i-1)}$. Since

$$\frac{|ce|}{|ab|} = \frac{|se|}{|sa|} = \frac{r_{i,j-1}}{ir + \frac{jr}{w}} \approx \frac{(i-1)r + \frac{(j-1)r}{w}}{ir + \frac{jr}{w}} = \frac{(i-1)w + j - 1}{iw + j},$$

therefore,

$$
\begin{aligned}
|ce| + |ab| &\approx \frac{2iw + 2j - w - 1}{iw + j} \cdot \frac{\lambda w(2i+1)}{\rho r(2i-1)} \\
&< \frac{\lambda w}{\rho r} \cdot \frac{(2iw + w - 1)(2i+1)}{(i+1)(2i-1)} \\
&< \frac{3\lambda(3w-1)}{2\rho r}.
\end{aligned}
$$

Since $|ad| \approx (1 + \frac{1}{w})r$, thus

$$r'_o = \max |ac| \approx \sqrt{\frac{(w+1)^2 r^2}{w^2} + \left(\frac{3\lambda(3w-1)}{4\rho r}\right)^2}.$$

Let $f(w) = \frac{(w+1)^2 r^2}{w^2} + \left(\frac{3\lambda(3w-1)}{4\rho r}\right)^2$. Obviously, $f(w)$ is a concave function of w. Let w^* be the value that satisfies $\frac{\partial f(w)}{\partial w}|_{w^*} = 0$, i.e., $\frac{(w^*)^3(3w^*-1)}{w^*+1} = \frac{16\rho^2 r^4}{27\lambda^2}$. Therefore, $w_o \approx w^*$ and $r'_o \approx \sqrt{\frac{(w^*+1)^2 r^2}{(w^*)^2} + \left(\frac{3\lambda(3w^*-1)}{4\rho r}\right)^2}$. $\qquad \square$

4.6.2 Optimal Number of Coronas

As shown in Figure 4.6, most data is transmitted through hop-by-hop transmission mode which is more energy efficient than long-distance direct transmission. However, compared with direct transmission mode, additional energy is spent by receiving data in hop-by-hop transmission mode. If the network is divided into only a few coronas, it is likely to incur significant energy dissipation due to the long hop-by-hop transmission distance. On the other hand, if the corona number is too large, a large amount of energy may be wasted by receiving data. The objective of the NLM problem is to compute the optimal number of coronas in terms of maximizing the network lifetime.

By Equation (4.12), d_i is a non-linear function of r'. Based on Theorem 8, the optimal hop-by-hop transmission distance can be approximated as a non-linear function of n because $r = \frac{R}{n}$. Therefore, the NLM problem is a non-linear integer programming problem. An important result has been reported in [117] that each exact algorithm for solving non-linear integer programming problem has an exponential computational complexity, which implies that it belongs to the class of NP-hard problem. In practice, such problems are usually solved by using heuristic algorithms such as Tabu search [43] and Simulated annealing [88]. In our scheme, the optimal number of coronas is computed off-line using Simulated annealing algorithm and then distributed to all nodes in network set-up phase.

4.7 The EBDG Protocol

In this section, the energy-balanced data gathering protocol EBDG is designed. The operation of EBDG is divided into two phases: network set-up phase and data gathering phase.

4.7.1 Network Set-up Phase

As discussed in Section 4.5 and Section 4.6, network parameters such as the optimal number of coronas n, the optimal number of sub-coronas w and the optimal data distribution ratio for each corona can be computed in an off-line manner. In network set-up phase, the sink distributes these global parameters to all nodes through broadcasting, and then each node establishes its corona, sub-corona and zone identifications based on these parameters.

In hop-by-hop transmission mode, EBDG uses the zone-based routing scheme presented in Section 4.4.2 to balance energy consumption among nodes within the same corona. Consider three consecutive coronas C_{i+1}, C_i and C_{i-1}. C_i acts as the destination corona when nodes in C_{i+1} forward the data to nodes in C_i, and C_i acts as the source corona when nodes in C_i forward the data to nodes in C_{i-1}. Since the sub-corona dividing approaches for source and destination coronas are different, each node in the network may belong to different zones when transmitting and receiving. $Z_{i,sj,sz}$ represents the source zone that u locates in when u forwards its data to nodes in C_{i-1}, and $Z_{i,dj,dz}$ denotes the destination zone that u locates in when u receives data from nodes in C_{i+1}. The following gives the approach to compute the corona ID, sub-corona ID and Zone ID for any node u.

For any node u, let (r_u, θ_u) be the polar coordinates of u. Since the network is divided into coronas with the same width, the corona ID (i.e. i) of node u can be computed as follows:

$$i = \lceil \frac{r_u}{R/n} \rceil = \lceil \frac{nr_u}{R} \rceil.$$

When nodes in C_{i+1} forward the data to nodes in C_i, C_i acts as destination corona, and C_i is divided into sub-coronas based on Theorem 7. Thus the destination sub-corona ID (i.e. dj) should be j' where $r_{i,j'-1} \leq r_u < r_{i,j'}$. In the zone-based routing scheme, each destination zone should contain λ nodes. Thus $\theta_{i,j}$ (as shown in Figure 4.3) should be $\frac{2\lambda}{\rho(r_{i,dj}^2 - r_{i,dj-1}^2)}$ where ρ is the node deployment density. The destination zone ID (i.e. dz) of node u can be computed as follows:

$$dz = \lceil \frac{\theta_u \cdot \lfloor \frac{2\pi}{\theta_d} \rfloor}{2\pi} \rceil, \quad \text{where} \quad \theta_d = \frac{2\lambda}{\rho(r_{i,dj}^2 - r_{i,dj-1}^2)}.$$

When nodes in C_i forward the data to nodes in C_{i-1}, C_i acts as source corona and C_i is divided into w sub-coronas with the same width $\frac{r}{w}$. Thus the source sub-corona ID (i.e. sj) of node u can be computed as follows:

$$sj = \lceil \frac{r_u - (i-1)\frac{R}{n}}{\frac{R}{wn}} \rceil = \lceil \frac{wnr_u - w(i-1)R}{R} \rceil.$$

By Lemma 9, each source zone in C_i should contain $\frac{\lambda(2i+1)}{2i-1}$ nodes since its correspond-

ing destination zone contains λ nodes. Thus the source zone ID (i.e. sz) of node u can be computed as follows:

$$sz = \lceil \frac{\theta_u \cdot \lfloor \frac{2\pi}{\theta_s} \rfloor}{2\pi} \rceil, \quad \text{where} \quad \theta_s = \frac{2\lambda w^2(2i+1)}{r^2(2i-1)(2iw-2w+2j-1)}.$$

4.7.2 Data Gathering Stage

In EBDG, all sensor nodes work in two states: *active* and *sleep*. In *active* state, each node can transmit data, receive data and perform data aggregation. In *sleep* state, each node turns off its radio to save energy.

Let $\mathcal{T}_{round}$ be the time duration for completing one round of data gathering. $\mathcal{T}_{round}$ is divided into wn time slots τ_0, τ_1,...,τ_i,..., τ_{wn-1}. In time slot τ_0, nodes in the outermost sub-corona (i.e., $C_{n,w}$) and in the corresponding destination sub-corona $C_{n-1,w}$ wake up. For any node $u \in C_{n,w}$, node u transmits its data directly to the sink if direct transmission mode is used; otherwise, node u forwards the data to nodes in its corresponding destination zone located in $C_{n-1,w}$. Similarly, in time slot τ_i, nodes in $C_{n-\lfloor \frac{i}{w} \rfloor,w-(i-\lfloor \frac{i}{w} \rfloor w)}$ and $C_{n-\lfloor \frac{i}{w} \rfloor-1,w-(i-\lfloor \frac{i}{w} \rfloor w)}$ wake up, and nodes in $C_{n-\lfloor \frac{i}{w} \rfloor,w-(i-\lfloor \frac{i}{w} \rfloor w)}$ finish transmitting data. Based on this scheduling, all nodes in the network finish reporting data to the sink in an order from the outermost sub-corona to the innermost sub-corona. This scheduling scheme provides the opportunity for each node to perform data aggregation. Moreover, it can effectively reduce the probability of data collusion.

In zone-based routing scheme, each source corona is divided into sub-coronas with equal size. Thus different sub-coronas have different number of nodes. For each time slot, it should be large enough so that all nodes in its corresponding source sub-corona can finish data transmission. In EBDG, the length of time slot τ_i, denoted by l_i, is set as follows:

$$l_i = \frac{A_{C_{n-\lfloor \frac{i}{w} \rfloor,m-(i-\lfloor \frac{i}{w} \rfloor w)}}}{\pi R^2} \cdot \mathcal{T}_{round}, \tag{4.15}$$

where $A_{C_{n-\lfloor \frac{i}{w} \rfloor,w-(i-\lfloor \frac{i}{w} \rfloor w)}}$ is the area size of $C_{n-\lfloor \frac{i}{w} \rfloor,w-(i-\lfloor \frac{i}{w} \rfloor w)}$. As can be seen from Equation (4.15), the time allocated time slot i is proportional to the number of nodes in $C_{n-\lfloor \frac{i}{w} \rfloor,w-(i-\lfloor \frac{i}{w} \rfloor w)}$.

To implement the above transmission scheme, each node u maintains two timers: $T_s(u)$ and $T_d(u)$. $T_s(u)$ is introduced for controlling u to enter into active state to transmit data, whereas $T_d(u)$ is introduced for controlling u to enter into active state to receive data. Initially, for any node $u \in C_{i,j}$ ($1 \le i \le n, 1 \le j \le w$), $T_s(u) = \sum_{k=i}^{n} \sum_{v=j}^{w} l_{(k-n)w+v-w} - l_0$ and $T_d(u) = \sum_{k=i+1}^{n} \sum_{v=j+1}^{w} l_{(k-n)w+v-w} - l_0$. This initialization guarantee that all nodes in $C_{i,j}$ can automatically enter into active state to start transmission at the beginning of the time slot allocated to $C_{i,j}$ for data transmission, and all nodes in $C_{i,j}$ can automatically enter into active state to receive data forwarded from nodes in $C_{i+1,j}$ at the beginning of the time slot allocated to $C_{i,j}$ for receiving data. The data gathering phase is triggered by having the sink broadcasting a control message, and all nodes start the two timers after receiving the control message.

If $T_s(u)$ expires, node u enters into active state, generates data, performs data aggregation and then transmits its data. Let $\mathcal{T}_{inter\_rounds}$ be the interval between two adjacent data gathering rounds. After u finishes transmitting, $T_s(u)$ is set to $\sum_{k=i}^{n} \sum_{v=j}^{w} l_{(k-n)w+v-w} - l_0 + \mathcal{T}_{inter\_rounds}$ which is the time node u should wake up to transmission data in next data gathering round. Node u then starts $T_s(u)$ and enters into sleep state to wait for the next data gathering round.

If $T_d(u)$ expires, u enters into active state to receive the data destined to it. At the same time, $T_d(u)$ is set to $l_{(n-i-1)w+w-j}$ (i.e., the length of the time slot in which node u should finish receiving data) and started. When $T_d(u)$ expires again, which means all nodes in the source corona have finished transmitting in this round, $T_s(d)$ is set to $T_d(u) = \sum_{k=i+1}^{n} \sum_{v=j+1}^{w} l_{(k-n)w+v-w} - l_0 + \mathcal{T}_{inter\_rounds}$ (i.e., the time node u should wake up to receive data in next data gathering round) and started.

In EBDG, each node alternates between hop-by-hop and direct transmission modes to balance energy consumption. To do this, each node $u \in C_i$ keeps records of the total amount of data transmitted by hop-by-hop mode (i.e., $F(u)$) and the total amount of data transmitted by direct transmission mode (i.e., $D(u)$) since the start of the data gathering phase. At any round, if $D(u) < p_i \cdot (F(u) + D(u))$, node u transmits the data in direct transmission mode; otherwise, the data is transmitted in hop-by-hop transmission mode in this round.

It is worthwhile noting that, at any time slot, there are only two sub-coronas whose

nodes stay in active state. Moreover, the nodes in one corona only transmit data, whereas the nodes in the other corona only receive data. Therefore, this transmission scheme can significantly reduce the probability of data collision. For nodes in each source zone, the transmissions are scheduled in EBDG to avoid data collision since the number of nodes in each source zone is relatively small. In EBDG, the source zones in each sub-corona are assigned with different frequency band (e.g. FDMA) so that the possibility of interfering with nearby zones is reduced. The detailed algorithm performed at each node for data gathering is given as follows:

Algorithm: DG /*Data Gathering Operation at $u \in C_{i,j}$*/

Initialization:

1:   $T_s(u) = \sum_{k=i}^{n} \sum_{v=j}^{w} l_{(k-n)w+v-w} - l_0$;

2:   $T_d(u) = \sum_{k=i+1}^{n} \sum_{v=j+1}^{w} l_{(k-n)w+v-w} - l_0$;

3:   Start $T_s(u)$ and $T_d(u)$;

Transmitting Phase:

4:       If $T_s(u) = 0$ then

5:           Enter into *active* state;

6:           Generate data, and perform data aggregation;

6:           If $D(u) < p_i \cdot (F(u) + D(u))$ then;

7:               Transmit data in direct transmission mode;

8:           Else

9:               Transmit data in hop-by-hop transmission mode;

10:          $T_s(u) = \sum_{k=i}^{n} \sum_{v=j}^{w} l_{(k-n)w+v-w} - l_0 + \mathcal{T}_{inter\_rounds}$, start $T_s(u)$;

11:          Enter into *sleep* state;

Receiving Phase:

12:       If $T_d(u) = 0$ then

13:           Enter into *active* state;

14:           $T_d(u) = l_{(n-i-1)w+w-j}$, start $T_d(u)$;

15:           Receiving data packets;

16:           If $T_d(u) = 0$ then

17:               $T_d(u) = \sum_{k=i+1}^{n} \sum_{v=j+1}^{w} l_{(k-n)w+v-w} - l_0 + \mathcal{T}_{inter\_rounds}$;

18:           Enter into *sleep* state;

For $u \in Z_{i,j,k}$, u can forward each packet to nodes in $Z_{i-1,j,k}$ with same probability if all packets generated by nodes in $Z_{i,j,k}$ have the same size. This can be easily done by each node in $Z_{i,j,k}$ maintaining a list of the nodes in $Z_{i-1,j,k}$ and using a uniform random function to determine the destination for each packet.

4.8 Extension to Large-Scale Data Gathering Sensor Networks

In mixed-routing schemes, the basic requirement is that all nodes must have the capability to communicate with the sink directly. However, in practice, most realistic sensor motes usually have limited transmission range, which indicates that mixed-routing scheme is not suitable for large-scale sensor networks. In this section, by employing the advantages of clustering techniques, the solution for extending EBDG to large-scale sensor networks is presented.

Clustering is a promising technique for large-scale sensor networks because of its high scalability and efficiency. By dividing the whole network into small clusters, each node only needs to communicate with its cluster-head through single or multi-hop routing, thereby eliminating the requirement for large communication range. Similar to the work in [86], it is assumed that the network is composed of two kinds of nodes: regular nodes and cluster-head nodes. The regular nodes have battery energy E_0, and perform sensing, data aggregation as well as packet relaying. The cluster-head nodes are equipped with battery energy E_1 which is much lager than E_0, and these nodes are responsible for collecting data from nodes within each cluster and transmitting the data to the remote sink. The data gathering in the extended EBDG is performed as follows:

- Intra-Cluster: In each cluster, EBDG is employed to gather the data from all nodes to the cluster head. Therefore, energy consumption is balanced among nodes within each cluster.

- Inter-Cluster: All the cluster heads form into a super-cluster with the sink acting as the cluster head. Then EBDG can be used to balance energy consumption among cluster heads. Besides this approach, some schemes based on mobile sink [86][112] can also be used to achieve balanced energy consumption among cluster heads.

Schemes based on balancing energy consumption through cluster heads rotation have been proposed, and several protocols such as LEACH, HEED, EECS and UCS have been designed. Compared with cluster heads rotation schemes, EBDG has the following advantages:

1. In LEACH, the clusters are dynamically created. In each running of the clustering algorithm, a large amount of energy is consumed. In EBDG, the clustering operation is performed only once, and the clusters remain unchanged after the network setup phase.

2. All of the network parameters are computed and distributed only once at the network setup phase, and the overhead caused by frequent cluster-heads rotation has been avoided.

3. By heterogeneous network deployment, the regular nodes are relieved of the burden of receiving a large amount of data when acting as cluster heads, which may quickly drain the energy at each node especially when the packet receiving cost is large. Therefore, network lifetime can be significantly extended.

4.9   Simulation Results and Analysis

In this section, the performance of EBDG is evaluated through extensive simulations. To demonstrate the efficiency of EBDG in terms of balancing energy consumption and maximizing network lifetime, EBDG is compared with a conventional multi-hop routing scheme, a direct transmission scheme and a cluster-heads rotation scheme.

4.9.1   Simulation Setup

In all simulations, the sensor nodes are near-uniformly deployed in a circular area, and the radius of the circular area can vary from $50m$ to $1000m$. There is only one data sink which is located at the center of the circle. All sensor nodes have the same initial energy $1 Joule$. For the radio model, the parameters are set as follows: $a_{11} = a_{12} = 50nJ/bit$, $a_2 = 10pJ/bit/m^2$ and $k = 2$ (same as the settings in LEACH [50] and HEED [124]). All simulations are based on a perfect MAC layer without collision and data loss since the design of EBDG is based on collision-free MAC protocols. In each round of data gathering, every node generates $512\ bits$ raw data, and the data compression ratio can vary from 0 to 1. When each node can fuse the data it generates together with the data it receives into

Table 4.1: Simulation Parameters for Evaluating EBDG

Type	Parameter	Value
Network	Network Area Radius (R)	$50m \sim 1000m$
	Node distribution	Near-Uniform
	Node deployment density (ρ)	$0.04 nodes/m^2$
	Initial energy	$1 Joule/node$
Application	Data generation rate (l)	$512 bits/node/round$
	Compression Ratio (m)	$0 \sim 1$
	Packet Size when $m = 0$ (c)	50 bits
Radio Model	$a_{11} = a_{12}$	$50nJ/bit$
	a_2	$10pJ/bit/m^2$
	k	2

one packet with constant size, the size of packet after compression is set to 50 $bits$. In all simulations, the network lifetime is defined as the time elapsed until 1% of the nodes in the network run out of energy. Table 4.1 lists the system parameters in detail.

Figure 4.8 gives an example of network division when the radius of circular area is $500m$. In this example, n (the number of coronas) is set to 2, and the boundary of coronas is marked with thick lines. w (the number of sub-coronas) is also set to 2, and the boundary of sub-coronas is marked with thin lines. λ (the number of nodes contained in each destination zone) is set to 6. Thus sub-coronas $C_{2,1}$ and its destination sub-coronas $C_{1,1}$ are divided into 4 zones. Sub-coronas $C_{2,2}$ and its destination sub-coronas $C_{1,2}$ are divided into 6 zones.
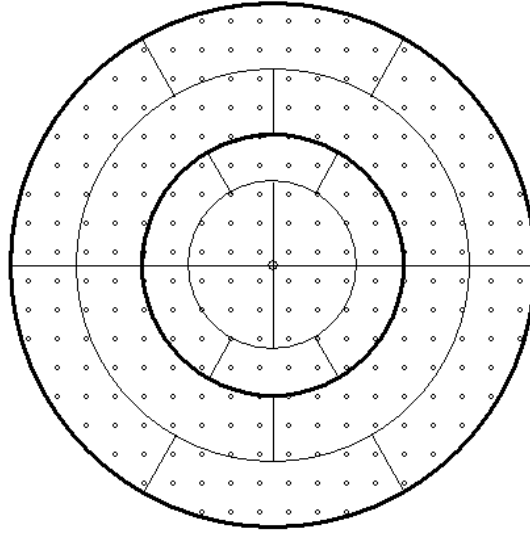
Figure 4.8: Network division and zones creation ($n = 2$, $w = 2$ and $\lambda = 3$)

4.9.2 Comparison with multi-hop routing and direct transmission schemes

In EBDG, energy consumption among all nodes in the network is balanced in expectation that all nodes should run out of energy at nearly the same time. This set of simulations are focused on evaluating the performance of EBDG in terms of energy consumption balancing and network lifetime extension by comparing with a multi-hop routing scheme and a direct transmission scheme. For multi-hop routing, a scheme based on minimum energy data gathering tree, denoted by MEDGT, is referred for comparison with EBDG. In MEDGT, each packet is delivered along the route that minimizes the total energy consumption for delivering the packet to the sink. For the direct transmission scheme (DT), in each round, every node generates its data, performs data compression and transmits the data directly to the sink without any relay.
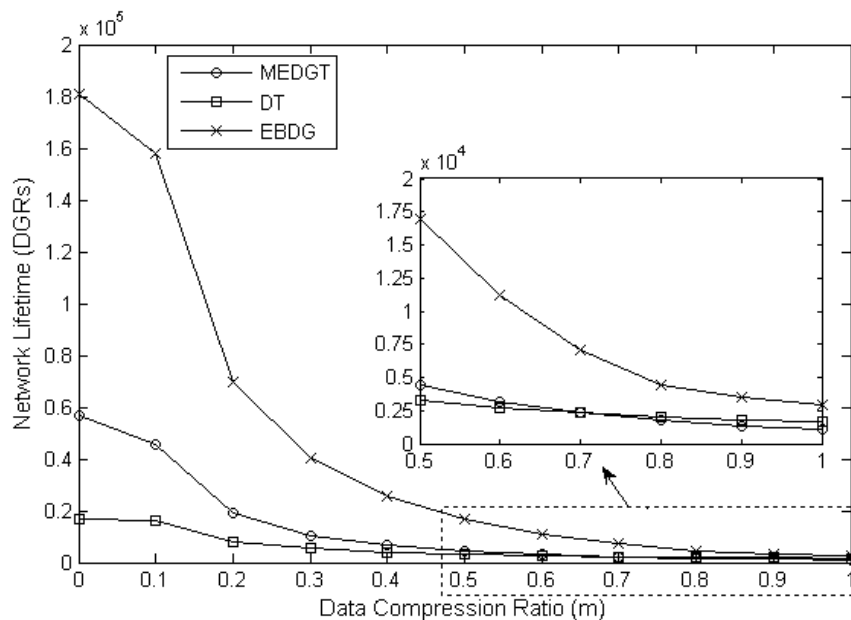
Figure 4.9: Network lifetime with different data compression ratio when $R = 340m$ and $\lambda = 5$

Figure 4.9 plots the lifetime obtained by the three schemes with different data aggregation parameters when $R = 340m$ and $\lambda = 5$. It can be seen that EBDG outperforms MEDGT and DT for all cases. When $m = 0$ and $c = 50bits$, the lifetime obtained by EBDG is 3 times of that obtained by MEDGT and more than 9 times of that obtained by

DT. Even when data is not aggregated, the lifetime obtained by EBDG has been doubled compared with MEDGT. From Figure 4.9, it is worth mentioning that data aggregation is an efficient way for prolonging network lifetime. With the decrease of m, the lifetime achieved by all three schemes has been significantly extended.
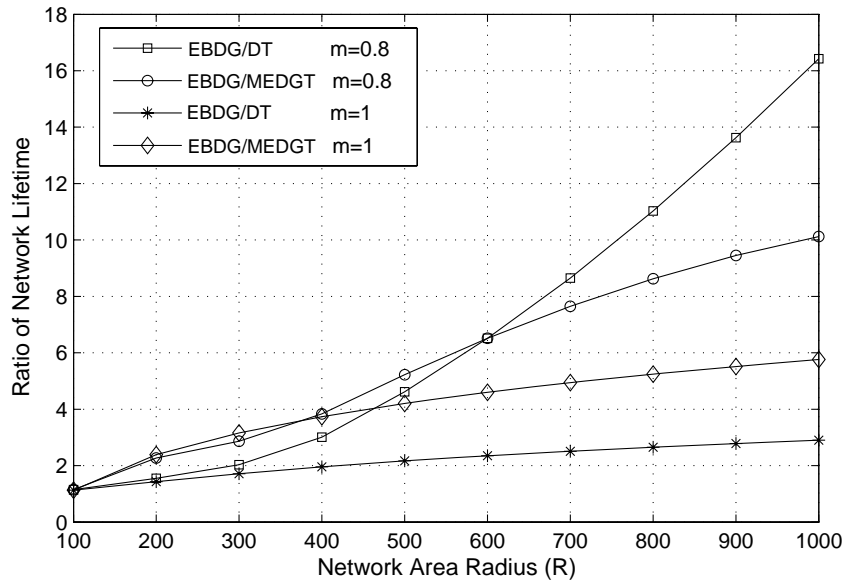
Figure 4.10: Network lifetime extension with different data compression ratio and different network size

Figure 4.10 shows the lifetime extension achieved by EBDG compared with MEDGT and DT under different data compression ration m and network size R. When $R = 100m$, the lifetime extension is small and the two ratios are the same because MEDGT and DT use single hop transmission for delivering packets to the sink. With the increase of R, for all cases, the two ratios increase. The lifetime achieved by EBDG is nearly 6 times of that obtained by MEDGT and nearly 3 times of that obtained by DT when $m = 1$ and $R = 1000m$. It is worth noting that the ratio of EBDG to MEDGT is larger than the ratio of EBDG to DT when $m = 1$, which means that the performance of direct transmission is better than multi-hop routing scheme when there is not data aggregation. The same result is also demonstrated in [85]. When $m = 0.8$, the performance of DT is better than that of MEDGT when $R < 600m$. With the further increase of R, DT performs poorly, and the lifetime achieved by EBDG is nearly 16 times of DT and 10 times of MEDGT when $R = 1000m$.
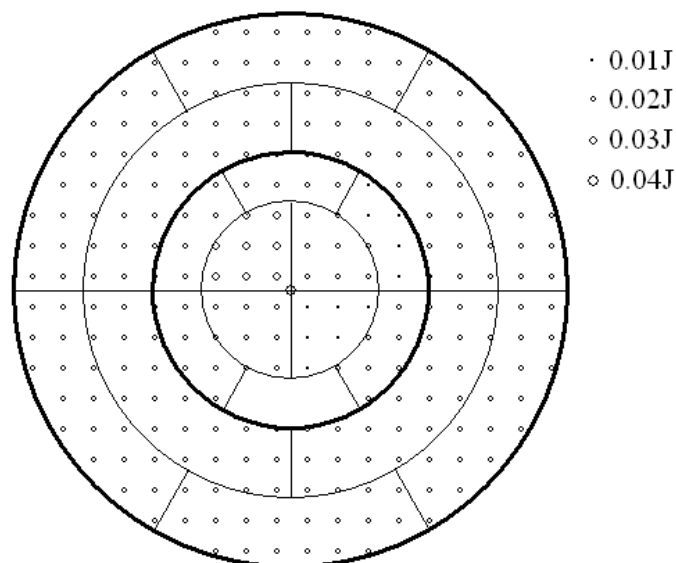
Figure 4.11: A snapshot of distribution of residual energy

Figure 4.11 gives a snapshot of the distribution of residual energy in the network after the nodes in one zone run out of energy. The residual energy for the nodes in the second corona have an uniform distribution since all nodes in this corona do not receive data during data gathering and all nodes have the same data distribution ratio. For the nodes in the first corona, the residual energy among different nodes varies slightly since node distribution is not absolutely uniform. However, the difference between the maximum residual energy and the minimum residual energy is less than 0.05J. Thus EBDG can efficiently balance energy consumption among nodes in the network.

4.9.3   Comparison with cluster-head rotation scheme

In this simulation, EBDG is compared with LEACH in which cluster-heads are frequently rotated to balance energy consumption among nodes within clusters. For simplicity, only one cluster is considered in this simulation. When EBDG is performed, the cluster-head is located at the center of the circle and remains unchanged during the entire network lifetime. The energy consumption at the cluster head is not taken into account since the cluster head in EBDG is equipped with much more power than regular nodes. When LEACH is performed, the cluster head is re-elected at the beginning of each round. To guarantee that there is only one cluster-head at each round, the node that first broadcasts the cluster-head

advertisement message is selected as the new cluster-head. In this simulation, the size of advertisement message is set to $10\ bits$. Figure 4.12 plots the lifetime extension obtained by
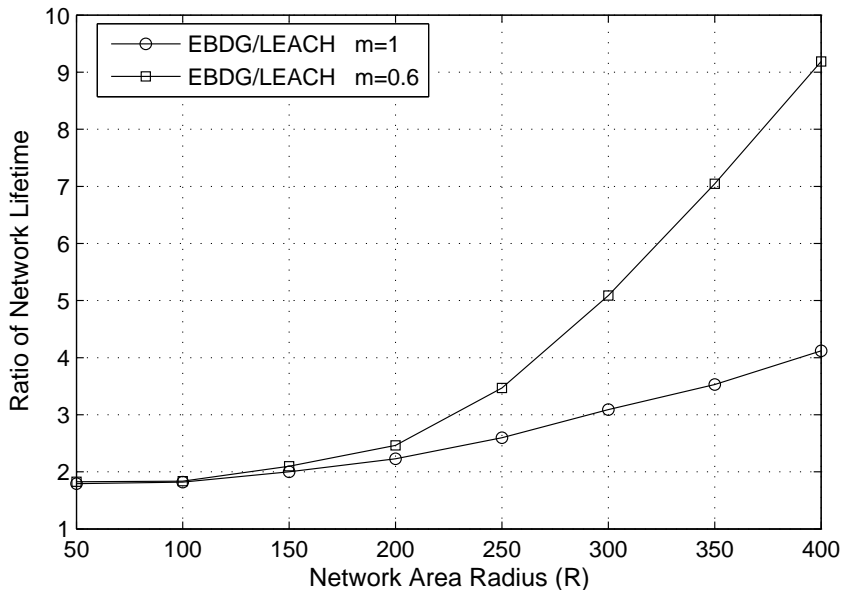
Figure 4.12: Network lifetime extension compared with LEACH

EBDG by comparing with LEACH under different data compression ratio m and network size R. It can been seen that the lifetime extension is small when R is small. With the increase of R, EBDG significantly outperforms LEACH because each node in LEACH uses direct transmission to deliver packets to the cluster-head and the energy consumption for transmitting each packet significantly increases with the increase of R. When $m = 0.6$, the lifetime obtained by EBDG is more than 9 times of that obtained by LEACH because the data can be aggregated in EBDG when hop-by-hop transmission mode is used.

4.9.4 Comparison with maximum lifetime data gathering scheme

In [61], Kalpakis *et al.* studied the problem of maximizing lifetime data gathering in wireless sensor networks for scenarios both with data aggregation (MLDA: Maximum Lifetime Data aggregation) and without data aggregation (MLDR: Maximum Lifetime Data Routing). A centralized algorithm which can generate a near-optimal data gathering schedule in terms of maximizing network lifetime was proposed. In MLDA/MLDR, the data gathering schedule is a collection of directed trees rooted at the sink that span all the sensors.

Each tree may be used for one or more rounds, and lifetime maximization is achieved by optimally allocating the number of rounds to each tree. Fig. 4.13 gives the comparison of network lifetime achieved by optimal schedule, MLDA/MLDR and EBDG for scenarios both with and without data aggregation. It can be seen that the lifetime achieved by EBDG keeps very close to that of MLDA/MLDR which is near-optimal. This phenomenon can be explained using the example given in Fig. 4.14. T_1, T_2 and T_3 are the three data gathering tree generated by MLDA/MLDR, and are allocated with 3076, 3076 and 10765 rounds respectively. In EBDG, the network is divided into 3 slices (i.e., node 1 and 2 are in one slice) as show in Fig. 4.14, and the data distribution ratio for node 3 (i.e., p_3) is 0.6363. It is easy to see that $p_3 = 0.6363 = \frac{10765}{10765+3076+3076}$ which means that the schedule generated by MLDA/MLDR is equivalent to that in EBDG.

Figure 4.13: Network lifetime: Optimal schedule, MLDA/MLDR and EBDG
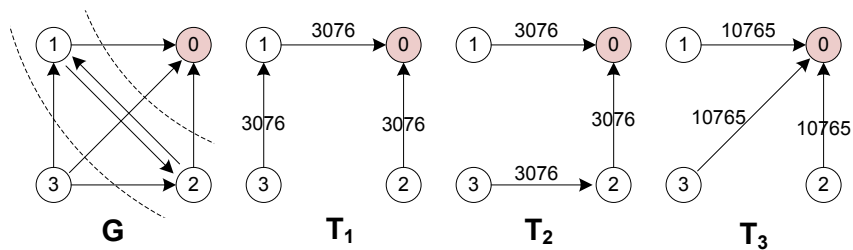
Figure 4.14: Network Graph and data gathering trees in MLDR

4.10   Summary and Future Work

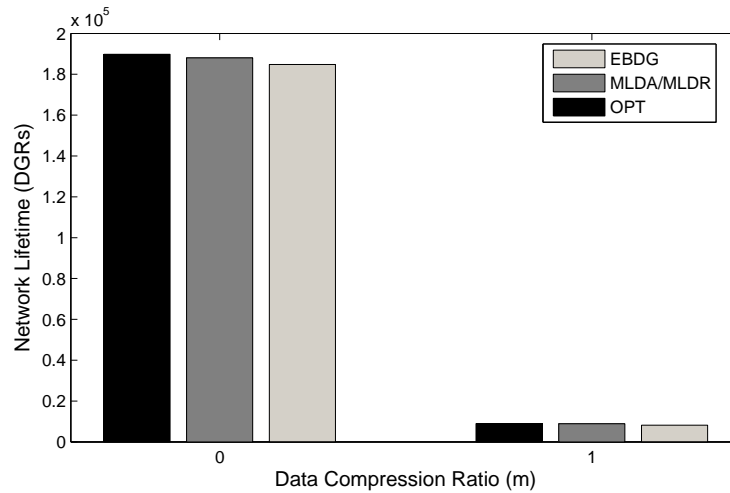Unbalanced energy consumption is an important problem in wireless sensor networks, which results in the network lifetime reduction dramatically. In this chapter, an efficient solution was designed to maximize network lifetime through energy consumption balancing techniques for uniformly deployed data gathering sensor networks. The energy balanced data gathering is formulated as an optimal data distribution problem by combining the ideas of corona-based network division and mixed-routing strategy together with data aggregation. First, a fully localized zone-based routing scheme was designed to balance energy consumption among nodes within each corona. Then an algorithm with time complexity $O(n)$ (n is the number of coronas) was designed to solve the data distribution problem aimed at balancing energy consumption across coronas. The approach for computing the optimal number of coronas in terms of maximizing network lifetime was also presented. Finally, an energy-balanced data gathering protocol (EBDG) was designed and the solution for extending EBDG to large-scale data gathering sensor networks was also presented. The performance of EBDG was evaluated through simulations, and simulation results demonstrate that EBDG can improve system lifetime by an order of magnitude compared with conventional multi-hop transmission schemes, direct transmission schemes and cluster-heads rotation schemes.

Future extensions of this work can be done in following directions. First, this work is focused on balancing energy consumption based on a collision-free MAC protocol. Future work can be done to extend it to sensor network with contention-based MAC protocols. However, different nodes may have different packet loss rate. If contention-based MAC protocols are employed, the energy balancing problem studied in this chapter is equivalent to the problem of balancing energy consumption in sensor networks with non-uniform nodes deployment and non-uniform data generation rate, which still remains as an open problem. Secondly, in this study, it is assumed that all nodes in the same corona have the same data distribution ratio since assigning each node with a different data distribution ratio significantly increases the complexity of the problems. In sparsely deployed sensor network, assigning each node with a different data distribution ratio may be practical. So the future work is to revise the scheme proposed in this chapter to sparse sensor networks

by letting different nodes have different data distribution ratios.

Chapter 5

Time-efficient Random Access Data Gathering

5.1 Introduction

Data gathering is a critical operation needed for collecting useful information from sensor nodes to a data sink which is responsible for further processing and end-user queries. Providing time-efficient data gathering is crucial to some applications such as traffic monitoring and health-care monitoring since the data generated by the sensors may become outdated after a certain time interval, and wrong decisions may be made based on the outdated information collected at the sink. Furthermore, a long data gathering duration[1] also means a large amount of energy wastage if no scheme is used to avoid idle listening.

Generally, the operation of data gathering is divided into rounds [42]. In each round, all sensor nodes sense the environment, generate the data, and report the data to the data sink. Thus a large amount of data may be generated in a very short period and transmitted through a shared radio channel during data gathering. The large burst of data packets may lead to high degree of channel contention and high probability of packet collision. The situation is further exacerbated when data packets are delivered over multi-hop routes. For example, in [127], the authors observed that around 50% packets are lost for most events in Lites with the default radio stack of TinyOS. To provide guaranteed data delivery, lost

---

[1]The data gathering duration is defined as the number of time slots needed for all nodes to finish one round data gathering.

packets must be retransmitted, leading to long data delivery delay.

To reduce packet collision incurred by hidden terminals, RTS/CTS handshaking mechanism has been employed in contention-based protocols such as carrier sense multiple access (CSMA). In mobile ad-hoc networks, the RTS/CTS handshaking can improve system performance since the data payload is much larger than the RTS/CTS messages. However, in most wireless sensor networks, data packets are very small. For example, the data payload in TinyOS is limited at 29 bytes [10]. In sensor networks deployed for gathering some environmental parameters such as temperature, humidity, pressure, etc, the data payload in the packets generated in such applications is only around several bytes. Thus RTS/CTS handshaking can add too much communication overhead to the network and lead to too much bandwidth wastage when data packets are small. In [33], the authors observed that the total percentage of bandwidth invested on RTS/CTS messages accounts for nearly 46%, which also implies that RTS/CTS handshaking has further deteriorated channel contention and results in more serious data loss. Moreover, the exchanging RTS/CTS messages also results in additional delay for data delivery.

Random access schemes have been seen as good solutions for bursty packet networks due to their simplicity and low delay (under light load) for bursty traffic [123]. It is well known that the random access channel throughput increases initially with increasing aggregate traffic generation rate [98]. A representative protocol based on random access techniques is the Aloha protocol [96]. However, in Aloha, a node immediately transmits whenever it has data to transmit irrespective of whether any other nodes are using the channel or not. The lack of cooperation of transmission among different nodes may result in too much packet collision. In data gathering schemes based on random access techniques, each sensor node should optimally attempt a transmission. Each sensor node should not be too aggressive to cause too much collisions, and at the same time, each sensor node should not be too conservative so as to miss the chance of a successful transmission.

This chapter focuses on designing time-efficient data gathering scheme for wireless sensor networks with small data packets. A novel scheme based on random access techniques is proposed for performing the data gathering operation in sensor networks. In this scheme, it is assumed that time is slotted. In each time slot, each node v_i works in *transmitting* state with probability α_i and works in *receiving* state with probability $1 - \alpha_i$. The data gathering

problem is formulated as an optimization problem, and the objective is to compute the optimal probability for each node to attempt a transmission so that the expected duration for performing one round of data gathering can be minimized. Fully localized solutions have been designed for both linear networks and tree networks. Based on this scheme, a simple protocol called RADG (Random Access Data Gathering) is designed. The performance of RADG is evaluated through simulations by comparing with CSMA-based protocols. Simulation results show that RADG can significantly outperform protocols both with and without RTS/CTS handshaking when data packets are small.

The rest of this chapter is organized as follows. Section 5.2 gives a simple example to illustrate the advantages of random access techniques over handshaking schemes for the data gathering operation. Section 5.3 reviews related work. System models and problem formulation are presented in Section 5.4. In Section 5.5, the solution for linear wireless sensor networks is presented. The solution for networks with random topology is presented in Section 5.6. In Section 5.7, a simple and scalable protocol called RADG is designed. The performance of RADG is evaluated through simulations in Section 5.8. Finally, Section 5.9 gives the summary and future work. For convenience, a detailed description of the variables used in this chapter is listed in Appendix D.

5.2   An Illustrating Example

In this section, a simple example is presented to demonstrate the advantages of random access techniques over handshaking schemes for data gathering in WSNs.

In this example, it is assumed that the size of a data packet is nearly the same as the size of control messages (i.e. RTS/CTS) used in handshaking schemes. This assumption is reasonable for sensor networks with small data packets. Consider a one-hop sensor network composed of n sensor nodes $(v_1, v_2, ..., v_n)$ and one data sink, as shown in Figure 5.1. In the handshaking scheme illustrated in Figure 5.2, the source node a must first send a RTS message to request for data transmission, and the data message can be transmitted only when node a receives a CTS message from the destination b. Therefore, the lower bound on the number of messages transmitted for successfully transmitting a data packet from a to b is 3, and this lower bound can be achieved only when there are no collisions for
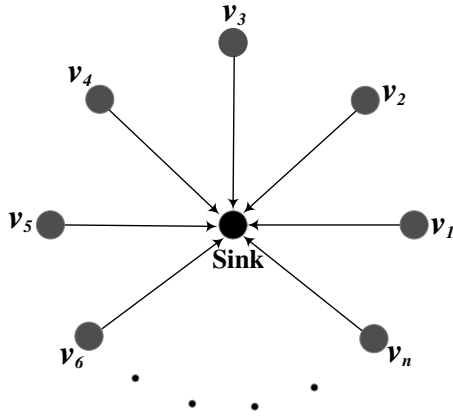
Figure 5.1: one-hop network
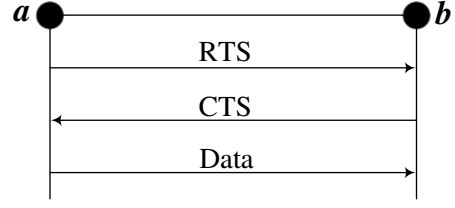
Figure 5.2: RTS/CTS example

RTS and CTS transmission. Obviously, for the one-hop network given in Figure 5.1, the lower bound on the total number of messages transmitted for performing one round of data gathering in handshaking-based schemes is $3n$, and the duration of performing one round of data gathering is proportional to the number of messages broadcasted.

In time-slotted random access schemes, at each time slot, each node works in one of the following two states: *transmitting* and *receiving*. For each node v_i, it works in *transmitting* state with probability α_i and works in *receiving* state with probability $1 - \alpha_i$. For the network given in Figure 5.1, to achieve fairness, all nodes should have the same transmitting probability denoted by α. At any time slot, only one of the following n scenarios may occur: *one node transmits, two nodes transmit, ... , n nodes transmit*. Let $p(k)$ be the probability that k nodes transmit simultaneously at a time slot. Then

$$p(k) = C_n^k \cdot \alpha^k (1 - \alpha)^{n-k}.$$

Let $E[N_{slot}]$ be the expected number of messages transmitted in one time slot. $E[N_{slot}]$ can be computed as follows:

$$E[N_{slot}] = \sum_{i=1}^{n} i \cdot p(i) = \sum_{i=1}^{n} i \cdot C_n^i \cdot \alpha^i (1 - \alpha)^{n-i}.$$

Give node v_i, the probability that it can successfully transmit a packet to the data sink at each time slot is $\alpha(1 - \alpha)^{n-1}$. Thus the expected number of time slots needed for v_i to

transmit its data packet to the data sink is $\frac{1}{\alpha(1-\alpha)^{n-1}}$. Let $E[N]$ be the expected number of messages transmitted in one data gathering round. Then

$$
\begin{aligned}
E[N] &= E[N_{slot}] \cdot \frac{1}{\alpha(1-\alpha)^{n-1}}; \\
&= \frac{\sum_{i=1}^{n} i \cdot C_n^i \cdot \alpha^i (1-\alpha)^{n-i}}{\alpha(1-\alpha)^{n-1}}.
\end{aligned}
$$
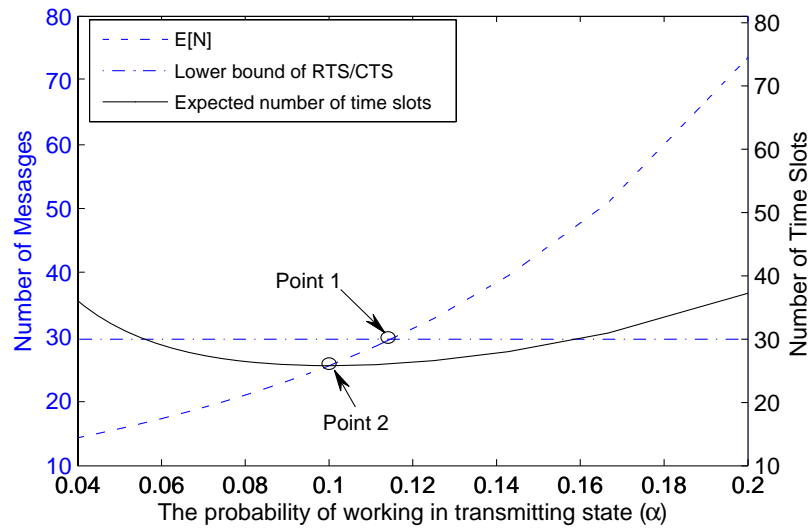
Figure 5.3: Number of message and number of time slots *vs.* α
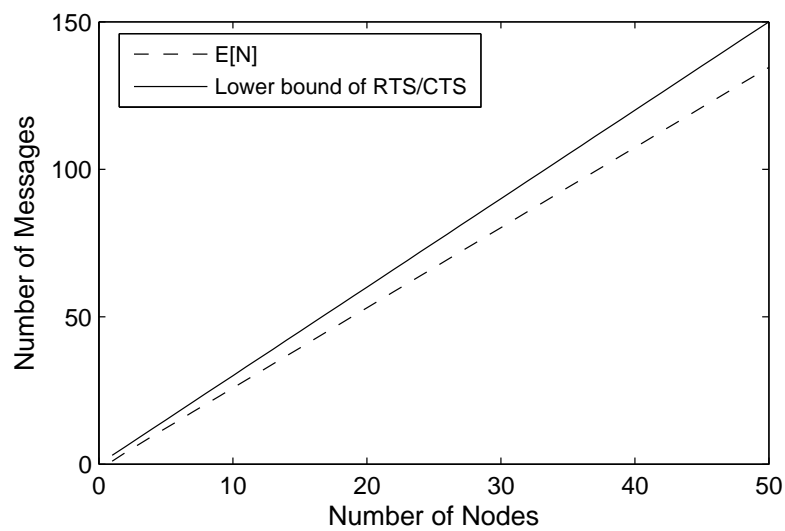
Figure 5.4: Message *vs.* number of nodes

Figure 5.3 gives the comparison of random access scheme and RTS/CTS based scheme when $n = 10$. It can be observed that the number of messages transmitted in the random channel access scheme decreases with the the decrease of α. Especially, when $\alpha < 0.116$ (point 1), the expected number of messages transmitted in random access scheme is even lower than the lower bound of messages transmitted in the RTS/CTS based scheme. It is also worth noting that there is an optimal $\alpha = 0.1$ (point 2) in terms of minimizing the expected number of time slots for completing one round of data gathering in the random access scheme. Figure 5.4 plots the expected number of messages for completing one round of data gathering with the variation of number of nodes in the network, where α is set to the value that minimizes the expected number of time slots. It can be seen that $E[N]$ is always less than the lower bound on number of messages in RTS/CTS based scheme for all cases. In real applications with bursty traffic patterns, the random access scheme may significantly outperform the RTS/CTS based scheme since the probability that RTS/CTS messages collide is very high.

From Figure 5.3, it can be seen that the setting of α has a great impact on the performance of the random access scheme. This chapter focuses on dealing with the problem of tuning the transmitting probability for each node to optimize the performance of the random access scheme.

5.3   Related Work

The Aloha protocol [11] and its slotted version [96] are two representative protocols based on random channel access techniques. Aloha is a time-division multiple access scheme. A node, whenever it has data to transmit, immediately transmits irrespective of whether any other nodes are using the channel or not. In Slotted-Aloha, time is divided into time slots. A node starts transmitting data only at the beginning of the next time slot. In both Aloha and Slotted-Aloha, nodes do not perform carrier sensing before transmission and the cooperation of data transmission among different nodes that share the same channel is also not taken into account. Thus data packets may easily collide in these two protocols.

Carrier sense multiple access (CSMA) has been widely used in wireless communication systems, and it is also the basic medium access mechanism in IEEE 802.11. In IEEE

802.11, DIFS (DCF Inter-frame Space) and SIFS (Short Inter-frame Space) are employed to control medium access. These Inter-frame Spaces can effectively reduce data collision. However, they also add additional delay for data transmission. To reduce packet collision, a random backoff procedure is employed in IEEE 802.11. Whenever an unsuccessful transmission occurs, the sender increases its contention window, from which it chooses its backoff time, exponentially. This procedure can also effectively reduce collision since the probability that the data packets at different nodes are set to the same backoff time is very small. However it may incur too much data delivery delay in a bursty traffic scenario since the contention window may become very large. In IEEE 802.11, RTS/CTS handshaking can further reduce the probability of data packet collisions particularly in the presence of hidden terminals. However, RTS/CTS handshaking can add too much communication overhead and packet delivery delay when data packets are small.

Some work has also been done on tuning the attempt probability for data transmission in random access wireless networks. In [64], the authors exploited the random scheme to perform link assessment task. In [113], directed and gossip-based neighbor discovery algorithms were designed based on random access. In [62], a generalized gradient ascent algorithm was proposed to self-learn the optimal attempt probability to maximize the throughput for each node. Two iterative, decentralized algorithms were presented in [116] to compute the global optimal rates for the problem of max-min fair rate allocation in ALOHA networks. In [15], the author views the problem of optimizing the node attempt probability as a non-cooperative game and shows the existence of a Nash equilibrium giving an optimal channel access rate for each node. However, most existing work focuses on maximizing throughput and does not consider other factors such as packet delivery reliability and delay. The work presented in this chapter is different from existing work. This study focuses on investigating the problem of exploring random access techniques for data gathering in sensor networks, taking into account both data transmission reliability and delay. In data gathering, each sensor node is not independent, and all the sensors must work in a cooperative manner since the quality of a link depends on the attempt probabilities of several nodes. The objective is to compute the optimal transmission probability for each sensor node so that the expected data gathering duration can be minimized.

5.4   System Models and Problem Formulation

5.4.1   System Models

The wireless sensor network is modeled as a directed graph $G = (V, L)$, where $V = (v_1, v_2, ..., v_n)$ is the set of sensor nodes, and L is the set of wireless links. Let $L(i, j)$ be the link from v_i to v_j. For each node v_i, $N_1(i)$ denotes the set of nodes that are within one hop of v_i, and $N_2(i)$ represents the set of nodes that are within two hops of v_i. The following models are employed in this study.

- **Antenna Model:** Each node can transmit and receive on a common carrier frequency using omnidirectional antennas.

- **Channel Access Model:** Similar to Slotted-Aloha, time is slotted and nodes are synchronized on time slots. At each time slot during data gathering, every node can work in two states: *transmitting* and *receiving*. For node v_i, it decides to work in *transmitting* state with probability α_i and decides to work in *receiving* state with probability $(1 - \alpha_i)$, where α_i is called the *attempt probability* for node v_i.

- **Collision Model:** Collision occurs when multiple nodes simultaneously transmit packets to the same destination. Although it is possible that the receiver may correctly decode the packet if the signal to interference ratio (SIR) exceeds a given threshold, in this study, it is assumed that all colliding packets are lost.

To avoid data queuing delay, it is also assumed that each node can fuse the data packets stored in its transmitting queue into a fixed-size transmitting packet whenever it attempts a transmission. This assumption is suitable for applications with aggregation operations such as *min* or *max* (e.g. temperature and humidity), *sum* (e.g. event count), and *yes-no* (e.g. intrusion detection). In applications where the data generated by different nodes can not be aggregated, network coding techniques [12][28][63] can be employed to meet this requirement. In random network coding [54], each node encodes the data it generates and receives with random coefficients into a fixed-size transmitting packet, and the decoding is only performed at the data sink. Network coding can achieve significant throughput improvement.

5.4.2 Problem Statement

In random access schemes, as demonstrated in Section 5.2, optimally setting the attempt probability for each node plays an important role in terms of minimizing data gathering duration. An improper attempt probability setting results in increased packet collisions, leading to long data delivery delay. Moreover, in WSNs, all sensor nodes may have different number of neighbors, and each sensor node also needs to adjust its attempt probability to cooperate with its neighboring nodes. The objective of this chapter is to design efficient solutions to compute the optimal attempt probability for each node in terms of minimizing the expected duration for one round of data gathering. To address the problem in detail, some definitions are first given as follows:

Definition 4. *At time slot t, $P_L(i,j)$ is defined as the probability that time slot t is collision-free for $L(i,j)$.*

$$P_L(i,j) = \alpha_i(1 - \alpha_j) \prod_{k \in (N_1(j) - \{v_i\})} \left(1 - \alpha_k\right). \tag{5.1}$$

It is obvious that time slot t is collision-free for $L(i,j)$ if and only if node v_i is in *transmitting* state; while node v_j and all its one hop neighbors except v_i are in *receiving* state at this time slot.

Let $\gamma(i,j)$ be the expected reliability for packet transmission over link $L(i,j)$. Since the states a node works in different time slots are independent, the expected number of time slots, denoted by $E[S_{i,j}]$, needed for transmitting a data packet from node v_i to v_j with reliability $\gamma(i,j)$ can be computed as follows:

$$E[S_{i,j}] = \frac{\gamma(i,j)}{P_L(i,j)} = \frac{\gamma(i,j)}{\alpha_i(1 - \alpha_j) \prod_{k \in (N_1(j) - \{v_i\})} \left(1 - \alpha_k\right)}. \tag{5.2}$$

Since each node can fuse the data in its transmitting packet queue into one packet whenever it attempts a transmission, there is no queuing delay for packet delivery. The following gives the definition for the expected delay for sensor-to-sink packet delivery.

Definition 5. *Let $\psi(i)$ be the set of links on the path from node v_i to the sink. $Delay(i)$ is the delay for delivering a packet from node v_i to the sink. The expected delay, denoted by*

$E[Delay(i)]$ *is defined as follows:*

$$E[Delay(i)] = \sum_{L(k,j)\in\psi(i)} E[S_{k,j}]. \tag{5.3}$$

Let T_{round} be the expected duration for one data gathering round. Obviously, $T = \max_{v_i \in V} E[Delay(i)]$. The objective is to compute the optimal attempt probability for each node so that T_{round} can be minimized on condition that each link in the data gathering structure can provide user-predefined reliability. The problem is formulated as follows:

$$\begin{aligned}
Minmize \quad & T_{round} = \max_{v_i \in V} E[Delay(i)]; \\
s.t. \quad & \gamma(k,j) \geq \tau(k,j) \quad L(k,j) \in \psi(i); \\
& 0 \leq \alpha_i \leq 1, \quad v_i \in V.
\end{aligned}$$

where $\tau(k,j)$ is the user-defined reliability for data transmission from v_k to v_j. This problem is referred to as the Attempt Probability Tuning (APT) problem.

5.5 Solution for Linear Wireless Sensor Networks

This section focuses on solving the APT problem for linear networks in which the sink is placed at one end of the network. The case that the sink is placed in an inner position may be seen as a particular case of a tree network rooted at the sink, which will be studied in next section. It is assumed that each node v_i except the sink and the last node v_n has only two one-hop neighbors: v_{i-1} and v_{i+1}, as shown in Figure 5.5. Hence, all the packets must be delivered along the unique path during data gathering.
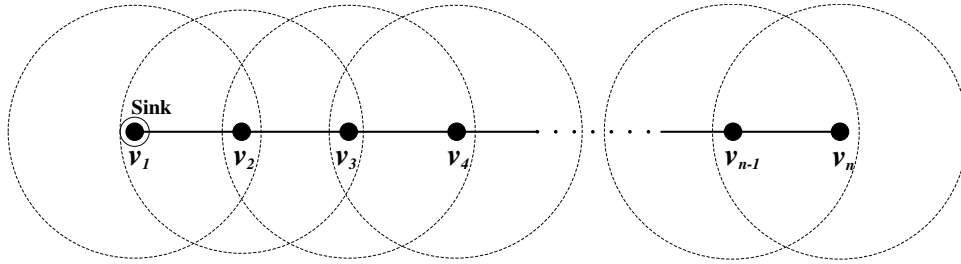
Figure 5.5: A linear network composed of n sensor nodes

Obviously,

$$T_{round} = E[Delay(n)] = \sum_{j=1}^{n-1} E[S_{j+1,j}], \qquad (5.4)$$

where node v_1 is the sink. Then the APT problem can be rewritten as

$$Minimize \quad E[Delay(n)] = \sum_{j=1}^{n-1} E[S_{j+1,j}];$$
$$s.t. \quad \gamma(k, k-1) \geq \tau(k, k-1), \; 1 < k \leq n;$$
$$0 \leq \alpha_i \leq 1, \quad v_i \in V.$$

5.5.1 Localized Algorithm for Attempt Probability Computation

Let $LS = \{L(i, i-1), 1 < i \leq n\}$. Obviously, only the links in *LS* are useful for packet delivery during data gathering.

Lemma 15. *Given any node v_i in a linear network, its attempt probability α_i only influences $E[S_{i,i-1}]$, $E[S_{i+1,i}]$ and $E[S_{i+2,i+1}]$.*

Proof. By Equation (5.2), $E[S_{i,j}]$ only depends on α_k where $k \in \{j\} \bigcup N_1(j)$. In linear networks illustrated in Figure 5.5, each intermediate node v_i has only two one-hop neighbors: v_{i-1} and v_{i+1}, i.e., $v_i \in N_1(i-1)$ and $v_i \in N_1(i+1)$. Therefore, α_i only influences $L(i, i-1)$, $L(i+1, i)$ and $L(i+2, i+1)$ for all the links in *LS*. Hence, the lemma is proved. □

Definition 6. *$f(\alpha_i)$ is defined as the expected latency for delivering a packet along the links that α_i influences.*

$$f(\alpha_i) = \begin{cases} E[S_{2,1}] + E[S_{3,2}], & i = 1; \\ E[S_{i,i-1}] + E[S_{i+1,i}] + E[S_{i+2,i+1}], & 1 < i < n-1; \\ E[S_{n,n-1}] + E[S_{n-1,n-2}], & i = n-1; \\ E[S_{n,n-1}], & i = n. \end{cases} \qquad (5.5)$$

Let α_i^* be the optimal attempt probability for v_i. Then the following two theorems hold.

Theorem 9. $\alpha_1^* = 0, \alpha_n^* = 1$.

Proof. By Lemma 15, α_1 only influences $E[S_{2,1}]$ and $E[S_{3,2}]$. By Equation (5.2), both $E[S_{2,1}]$ and $E[S_{3,2}]$ strictly decreases with the decrease of α_1. By Definition 5, $E[Delay(n)]$ strictly decreases with the decrease of α_1. Thus $\alpha_1^* = 0$. Since α_n only influences $E[S_{n,n-1}]$ which decreases with the increase of α_n, $E[Delay(n)]$ strictly decreases with the increase of α_n. Thus $\alpha_n^* = 1$. Consider the data gathering scenario, the theorem is correct because v_1 is the data sink and it only needs to receive packets; while v_n is the last node in the linear network and it only needs to send its own packet. $\square$

Theorem 10. For node $v_i \in V, 1 < i < n$, $\frac{\partial f(\alpha_i))}{\partial \alpha_i}|_{\alpha_i^*} = 0$.

Proof. For any intermediate node v_i, it should not only send its own packet but also relay the packets it receives from v_j where $j < i$. Thus $0 < \alpha_i^* < 1$. It is easy to prove that $Delay(n)$ is continuous and differential when $0 < \alpha_i < 1$. According to the extremum existence theorem for continuous and differential multi-variable functions, there must be a minimum for $Delay(n)$ when $\frac{\partial Delay(n))}{\partial \alpha_i} = 0$. Since α_i^* should minimize $Delay(n)$, $\frac{\partial Delay(n))}{\partial \alpha_i}|_{\alpha_i^*} = 0$. From Lemma 15 and Definition 6, $\frac{\partial Delay(\alpha))}{\partial \alpha_i} = \frac{\partial f(\alpha_i))}{\partial \alpha_i}, 1 < i < n$. Therefore, $\frac{\partial f(\alpha_i))}{\partial \alpha_i}|_{\alpha_i^*} = 0$. $\square$

The following gives the details of the procedure for computing the optimal attempt probability for each node. First, the local data structure maintained at each node, the message communication between the nodes and the attempt probability update procedure performed at each node are described respectively. Then a localized algorithm based on only two-hop neighbor information is presented.

Local Data Structure: By Lemma 15, all the links that α_i influences are within two-hop neighborhood of node v_i. Therefore, to compute $f(\alpha_i)$, each node only needs to store the attempt probability of the nodes in $N_2(i)$. To achieve this, each sensor node v_i maintains a simple local data structure called *<PNtab>* which records $\alpha_j, j \in \{i\} \bigcup N_2(i)$.

Communication: At each time slot, if node v_i is in *transmitting* state, it broadcasts *<PNtab>* to its one hop neighbors; otherwise, it listens to receive the *<PNtab>* message from its one hop neighbors. Once node v_i has successfully received a *<PNtab>* message, it updates its own *<PNtab>* based on the following two rules:

1. If v_i receives a *<PNtab>* message from v_{i-1}, v_i only needs to update α_{i-1} and α_{i-2} with the corresponding values in the received *<PNtab>* message.

2. If v_i receives a *\<PNtab\>* message from v_{i+1}, v_i only needs to update α_{i+1} and α_{i+2} with the corresponding value in the received *\<PNtab\>* message.

The *\<PNtab\>* update procedure is illustrated by an example. For instance, if v_3 receives a *\<PNtab\>* message from v_2, v_3 only needs to replace α_1 and α_2 with the corresponding values in the received *\<PNtab\>* message. Although α_4 also exists in the received *\<PNtab\>*, it may be outdated, and v_3 can get the latest α_4 from the *\<PNtab\>* broadcasted by v_4.

Computation: For each intermediate node v_i, an iterative scheme is designed to compute the best attempt probability for each node. The main idea of this iterative scheme is that each node v_i iteratively adjusts its attempt probability α_i so that $f(\alpha_i)$ is minimized. Whenever node v_i receives a *\<PNtab\>* message and updates its own *\<PNtab\>*, it updates its attempt probability α_i with value that minimizes $f(\alpha_i)$. Finally, this interactive local optimization is expected to lead to a global optimal solution.

The localized algorithm **APT-LN** designed for computing the attempt probability for each node in linear networks is given as follows.

Algorithm APT-LN: Tune α for Linear Network(node v_i)
1: Initialize α_i and $\bar{\alpha}_i$;
2: state=GenerateState($\bar{\alpha}_i$);
3: If state=*Transmitting* then
4:     broadcast(i,*\<PNtab\>*);
5: If state=*Receiving* then
6:     If SuccessReceive(i-1,*\<PNtab\>*);
7:         UpdatePNtab($\alpha_{i-2}, \alpha_{i-1}$);
8:     If SuccessReceive(i+1,*\<PNtab\>*);
9:         UpdatePNtab($\alpha_{i+1}, \alpha_{i+2}$);
10:     $\alpha_i = \alpha_i'$, where $\frac{\partial f(\alpha_i)}{\partial \alpha_i}\vert_{\alpha_i'} = 0$;

In **APT-LN**, $\bar{\alpha}_i$ is the attempt probability used during the process of attempt probability computation. The reason for introducing $\bar{\alpha}_i$ can be illustrated as follows: during the running of the algorithm, α_i changes dynamically, and α_i may become very small. If node v_i uses α_i which is very small to control the broadcasting of *\<PNtab\>* message. The neighbors of node v_i need a long time to receive the *\<PNtab\>* message broadcasted by node v_i, which may significantly affect the convergence speed of algorithm **APT-LN**. Obviously, the suitable setting for each $\bar{\alpha}_i$ should ensure that all nodes have the same probability to

receive $<PNtab>$ message from its neighbors. From the analysis in Section 5.2, a proper setting of $\bar{\alpha}_i$ is $\frac{1}{|N_1(i)|}$ where $|N_1(i)|$ is the cardinality of $N_1(i)$.

5.5.2 Numerical Results and Analysis

Figure 5.6 plots the convergence of algorithm **APT-LN** for three independent runs on a linear sensor network composed of 8 sensor nodes. The expected reliability for all links in LS is set to 1. It can be seen that algorithm **APT-LN** has very fast convergence rate. The expected data gathering duration drops quickly in the first several time slots, and it takes less than 50 time slots for the algorithm to converge to the optimal result.
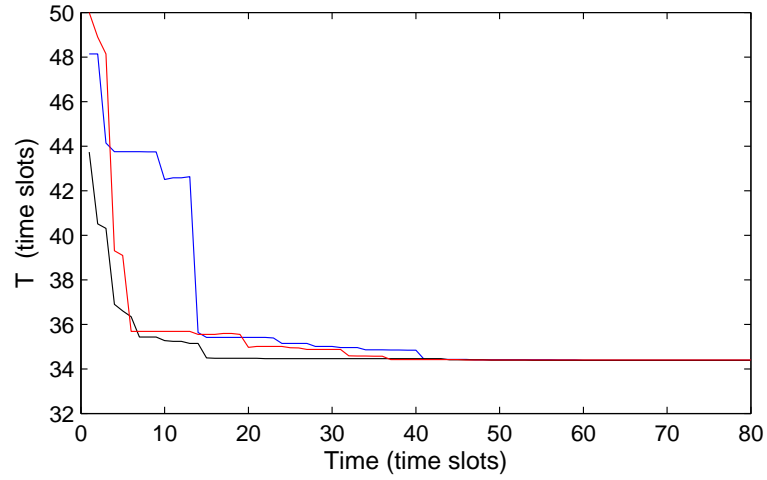
Figure 5.6: Convergence speed of algorithm **APT-LN** on a linear sensor network composed of 8 nodes

By Equation (5.2) and Definition 5, the optimal attempt probability for each node is independent of link reliability if all links in LS are set to the same reliability. Table 5.1 gives the attempt probability computed by Algorithm **APT-LN** when all links in LS are set to the same reliability.

Table 5.1: Attempt probability for each node

Node v_i	1	2	3	4	5	6	7	8
α_i	0	0.265	0.288	0.332	0.359	0.404	0.555	1

Table 5.2 gives the expected number of time slots for each link in LS and the the expected data gathering duration T_{round} with different link reliability τ. It can be easily

observed that there is a tradeoff between link reliability and the duration of data gathering. In practice, the required link reliability can be adjusted to cater for the requirement of different applications.

Table 5.2: $E[S_{i,j}]$ and T_{round} with different link reliability

$E[S_{i,j}]$	$2,1$	$3,2$	$4,3$	$5,4$	$6,5$	$7,6$	$8,7$	T_{round}
τ=0.85	3.2	4.0	4.9	5.0	4.9	4.0	3.2	29.2
τ=0.90	3.4	4.2	5.2	5.3	5.2	4.2	3.4	30.9
τ=0.95	3.6	4.5	5.5	5.6	5.5	4.5	3.6	32.8
τ=1	3.7	4.7	5.7	5.8	5.7	4.7	3.7	34

5.6   Solution for Networks with Random Network Topology

Having solved the linear network APT problem, this section focuses on solving the APT problem for sensor networks with random topology. In wireless sensor networks deployed for data gathering applications, many-to-one communication paradigm which involves multiple sources and one data sink is the main communication pattern. Thus the corresponding data flow generally resembles a reverse-multicast structure which is called the *data gathering tree*. In this study, it is assumed that the data gathering tree (e.g., minimum hop tree, minimum spanning tree, etc.) has been constructed. The objective is to solve the problem of tuning the attempt probability for each node in the data gathering tree to minimize the expected data gathering duration.

Let $T = (V, E)$ be the data gathering tree where V is the set of n sensor nodes $v_1, v_2, ..., v_n$, and E is the set of directed links in T. Let F_i be the father node of v_i, and C_i be the set of children of v_i. $Ts(i)$ denotes the maximum expected time slots for the descendants of v_i to deliver a packet to v_i. Thus $Ts(i)$ can be computed as follows:

$$Ts(i) = \max_{j \in C_i}\{Ts(j) + E[S_{j,i}]\}. \tag{5.6}$$

Let v_1 be the sink. Obviously, $Ts(1) = \max_{i \in V} E[Delay(i)]$. Thus the ATP Problem

can be rewritten as

$$Minmize \quad T_{round} = Ts(1);$$
$$s.t. \quad \gamma(k, j) \geq \tau(k, j) \quad L(k, j) \in E;$$
$$0 \leq \alpha_i \leq 1, \quad v_i \in V.$$

5.6.1 Localized Algorithm for Attempt Probability Computation

In tree networks, each node may have more than two neighbors. As can be seen from Equation (5.2), $E[S_{i,j}]$ depends on the attempt probabilities of all neighbors of node v_j, which means that the attempt probability for node v_i may influence all links $L(m, n)$ where $m \in N_i$ or $n \in N_i$. We design a heuristic localized algorithm to solve the APT problem for tree networks.

Local storage and communication: Although α_i influences all links initiated from v_i and destined to v_i, only the links belonging to the data gathering tree are useful for data gathering. So α_i should be adjusted to maximize the performance of the links that α_i influences in the data gathering tree.

Lemma 16. *For tree networks, all the links α_i influences are within two hops of v_i.*

Proof. From Equation (5.2), α_i only influences $E[S_{k,j}]$ where $v_i \in \{v_j\} \bigcup N_1(j)$. Given $L(k, j)$,

- if $v_i = v_j$, $L(k, j)$ must be within one hop from v_i;

- if $v_i \in N_1(j)$ and $v_i = v_k$, $L(k, j)$ must be within one hop from v_i;

- if $v_i \in N_1(j)$ and $v_i \neq v_k$, the number of hops from v_i to v_k is 2.

Hence, the lemma is proved. ☐

By Lemma 16, v_i can compute any $E[S_{k,j}]$ that α_i influences if v_i has the knowledge of the attempt probabilities of the nodes in $N_2(i)$. To achieve this, each node v_i maintains a table *<PNtab>* which records the attempt probability for v_j where $j \in \{i\} \bigcup N_2(i)$. At each time slot, if a node is in *transmitting* state, it broadcasts its *<PNtab>* to its neighbors.

Otherwise, if node v_i is in *receiving* state and successfully receives a *<PNtab>* message from its neighbor v_j, it only updates $\alpha_k, k \in \{j\} \bigcup N_1(j)\backslash\{i\}$ with the corresponding value in the received *<PNtab>*.

Computation: The attempt probability tuning is divided into two phases: *stabilization phase* and *latency balancing phase*.

Obviously, the one round data gathering duration is the maximum data delivery delay from the sensor nodes to the data sink. In a tree structure, each node has a unique path to the sink. Therefore, the path started from the sensor node which has the maximum delay and ended at the data sink is the **critical path** that has a great impact on the one round data gathering duration.
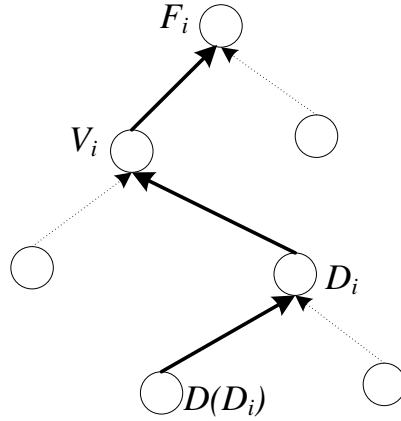
Figure 5.7: Critical path of node v_i

As shown in Figure 5.7, for node v_i where $i \in (V - \{v_1\})$, let $D_i = j$ where $j \in C_i$ and $Ts(j) + E[S_{j,i}] = \max_{k \in C_i}(Ts(k) + E[S_{k,i}])$. $D(D_i)$ is the child of D_i where $Ts(D_i) + E[S_{D_i,i}] = \max_{k \in C_i}(Ts(k) + E[S_{k,D_i}])$. By Lemma 16, all the links α_i influences are within two hops of v_i. In this study, the path from $D(D_i)$ to F_i is called the **critical path for node** v_i.

Definition 7. $f(\alpha_i)$ *is defined as the expected latency for delivering a packet along the critical path for node* v_i.

$$f(\alpha_i) = E[S_{D(D_i),D_i}] + E[S_{D_i,i}] + E[S_{i,F_i}]. \tag{5.7}$$

In the *stabilization phase*, each node adjusts its attempt probability with the objective to

minimize the expected latency for delivering a packet along its critical path. In this phase, the data gathering duration T_{round} can be significantly reduced since the critical paths have a great impact on the data gathering duration. By Equation (5.2) and Equation (5.7), $f(\alpha_i)$ is a strict concave function with respect to α_i. Therefore, there is an optimal α_i' that can minimize $f(\alpha_i)$ at each iteration, and α_i' satisfies $\frac{\partial f(\alpha_i))}{\partial \alpha_i}|_{\alpha_i'} = 0$. For any node v_i, it updates α_i with α_i' whenever it receives a <*PNtab*> message from its neighbors.

In tree networks, an intermediate node may have several branches. The *stabilization phase* can reduce the expected latency for packet delivery along the critical paths. However, it can not guarantee that the expected latency for different branches is balanced. In the *latency balancing phase*, the data gathering duration is further decreased by balancing the expected data delivery delay among different branches.

Lemma 17. *Given node v_i, $E[S_{i,F_i}]$ increases with the decrease of α_i. For any other link $L(m,n) \in E$ that α_i influences, $E[S_{m,n}]$ decrease with the decrease of α_i.*

Proof. By Equation (5.2),

$$E[S_{i,F(i)}] = \frac{\gamma(i, F_i)}{P_L(i, F_i)} = \frac{\gamma(i, F_i)}{\alpha_i(1 - \alpha_{F_i}) \prod_{k \in (N_1(F_i) - \{v_i\})} (1 - \alpha_k)}.$$

Obviously, $E[S_{i,F(i)}]$ increases with the decreases of α_i.

For any link $L(m,n) \in E$ except $L(i, F_i)$, v_i must be a neighbor of v_n if α_i influences $L(m,n)$. Thus

$$
\begin{aligned}
E[S_{m,n}] &= \frac{\gamma(m,n)}{\alpha_m(1 - \alpha_n) \prod_{k \in (N_1(n) - \{v_i\})} (1 - \alpha_k)}; \\
&= \frac{\gamma(m,n)}{\alpha_m(1 - \alpha_n)(1 - \alpha_i) \prod_{k \in (N_1(n) - \{v_i\} - \{v_m\})} (1 - \alpha_k)}.
\end{aligned}
$$

Obviously, $E[S_{m,n}]$ decreases with the decrease of α_i. $\qquad\square$

Lemma 18. *Let $Ts^*(i)$ be the value of $Ts(i)$ when each node v_i has obtained the optimal attempt probability α_i^*, then*

$$Ts^*(j) + E[S_{j,i}] = Ts^*(k) + E[S_{k,i}], \quad j \in C(i), k \in C(i). \tag{5.8}$$
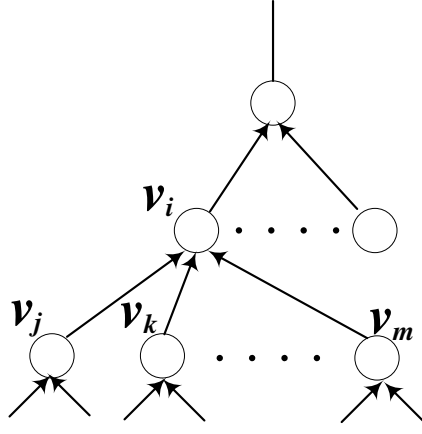
Figure 5.8: Illustration for Lemma 18

Proof. Suppose that there are two node $v_j \in C_i, v_k \in C_i$, as shown in Figure 5.8, where $Ts^*(j) + E[S_{j,i}] \neq Ts^*(k) + E[S_{k,i}]$ when each node has been assigned with the optimal attempt probability. Without loss of generality, suppose $Ts^*(j) + E[S_{j,i}] > Ts^*(k) + E[S_{k,i}]$. From Lemma 16, α_k only influences $\{E[S_{p,F_k}], v_p \in C_{F_k}\}$, $\{E[S_{q,k}], v_q \in C_k\}$ and $\{E[S_{l,n}], v_n \in C_k, v_l \in C_n\}$. By Lemma 17, $E[S_{k,i}]$ increases with the decrease of α_k. For any other link $L(m,n)$ that α_k influences, $E[S_{m,n}]$ decreases with the decrease of α_k. Therefore, there must be a α_k' where $\alpha_k' < \alpha_k^*$ and $Ts'(1) < Ts^*(1)$, which implies that α^* is not optimal. Hence, the lemma is proved. □

The basic mechanism of *latency balancing* is described as follows: the critical path for each node remains unchanged, and each node adjusts its attempt probability to balance latency among different branches. As shown in Figure 5.8, let $v_j = D(i)$. For any $v_m \in C_i$ and $v_m \neq v_j$, by Lemma 17, $Ts(i)$ decreases with the the decrease of α_m until $Ts(j) + E[S_{j,i}] = Ts(m) + E[S_{m,i}]$. If $Ts(m) + E[S_{m,i}] < Ts(j) + E[S_{j,i}]$, α_m decreases with Δ which is a user-defined parameter; otherwise, α_m increases with Δ. The parameter Δ is introduced to adaptively tune the attempt probability for each node with the objective to balance latency among different branches. The parameter Δ should be neither too small nor too large. A small Δ results in slow convergence speed. While a large Δ results in a large T_{round}. This phenomenon is demonstrated by the numerical results in Section 5.6.2.

To switch from *stabilization phase* to *latency balancing phase*, each node v_i maintains a counter denoted by θ_i. When the probability tuning procedure is triggered, θ_i is initialized to 0. Whenever node v_i update its attempt probability α_i with a different value, θ_i is reset

to 0; otherwise, θ_i is increased by 1. If θ_i exceeds a user-defined threshold, node v_i switchs from *stabilization phase* to *latency balancing phase*.

The detailed algorithm for tuning the attempt probability in tree network is given as follows:

Algorithm APT-TN: Tune α for Tree Network(node u)

1:    Initialize α_u, $\bar{\alpha}_u$ and Δ;
2:    state=GenerateState($\bar{\alpha}_u$);
3:    If state=*Transmitting* then
4:       broadcast(u, *Ts(u)*, <*PNtab*>);
5:    If state=*Receiving* then
6:       If SuccessReceive(v, *Ts(v)*, <*PNtab*>);
7:        Update $\alpha_m, m \in \{v\} \bigcup N_1(v) \backslash \{u\}$;
      Case Self-Stabilization:
8:        $\alpha_u = \alpha'_u$, where $\frac{\partial f(\alpha_i))}{\partial \alpha_i}|_{\alpha'_u} = 0$;
9:        Update $Ts(u)$ and D_u;
      Case Latency-Balance:
10:       If $u = D_{F_u}$ then
11:        $\alpha_u = \alpha'_u$, where $\frac{\partial f(\alpha_i))}{\partial \alpha_i}|_{\alpha'_u} = 0$
12:        else if $Ts(u) + S_{u,F_u} < Ts(D_{F_u}) + S_{D_{F_u},F_u}$ then
13:          $\alpha_u = \alpha_u - \Delta$
14:         else $\alpha_u = \alpha_u + \Delta$;
15:       Update *Ts(u)*;

5.6.2   Numerical Results and Analysis

The algorithm **APT-TN** is evaluated on a tree network composed of 60 nodes, as shown in Figure 5.9. For node v_i, the attempt probability $\bar{\alpha}_i$ used during attempt probability tuning is set to $\frac{1}{|N_1(i)|}$.

Figure 5.10 plots the convergence of the algorithm during the stabilization phase for three independent runs. For all runs, the expected data gathering duration drops quickly and finally converges to the minimum value, and it takes only around 60 time slots for the algorithm to achieve stability. This phenomenon can be explained as follows: at each time slot during the stabilization phase, the tree network is divided into linear networks since each node updates its attempt probability aiming at minimizing the expected latency for delivering packet along its critical path. It can be seen that the expected data gathering duration has been reduced to around 90 time slots from more than 1000 time slots during the stabilization phase.
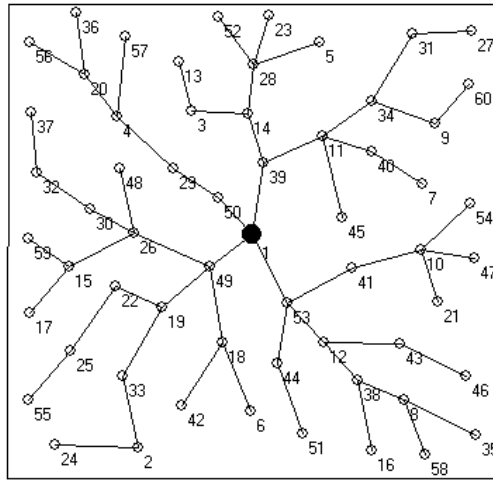
Figure 5.9: A tree data gathering network composed of 60 nodes
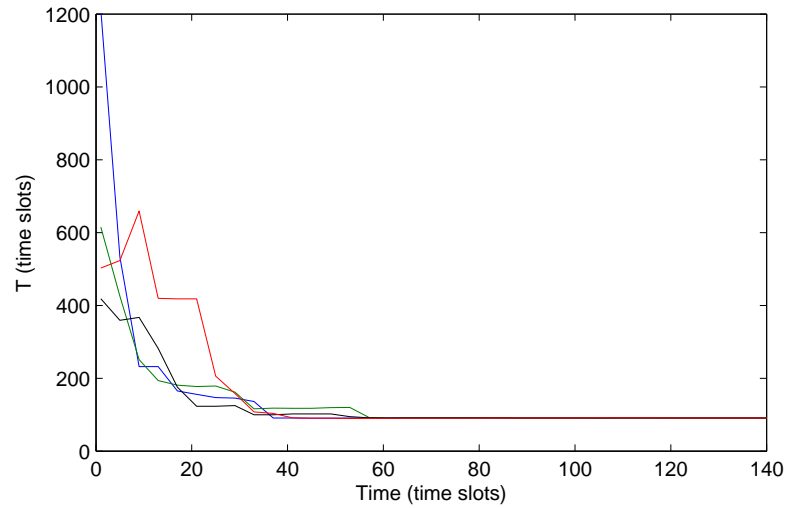
Figure 5.10: Convergence speed of self-stabilization phase

Figure 5.11 plots the convergence of the algorithm during the latency balancing phase with different setting of Δ. When $\Delta = 0.001$, the data gathering latency decreases stably but the convergence speed is slow. While $\Delta = 0.01$, the convergence speed is a little fast but there are many ups and downs after 20 time slots, and the final T_{round} computed when $\Delta = 0.01$ is a little larger than that when $\Delta = 0.001$. During the latency balancing phase, the data gathering duration T_{round} is further reduced, and it takes only around 58 time slots to perform one round of the data gathering operation.
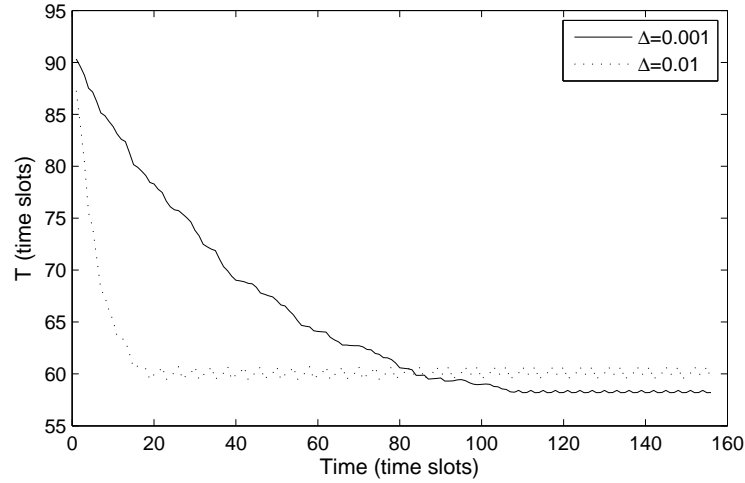
Figure 5.11: Convergence speed of latency balance phase

5.7   Protocol RADG

In this section, a simple and scalable data gathering protocol called RADG is designed based on the proposed random access scheme. In RADG, an aggregated acknowledgement scheme is designed to improve channel utilization and to reduce packet collision. Moreover, periodic node sleeping is employed to save energy by turning off the radio in the period between two consecutive data gathering rounds.

5.7.1   Transmission Stopping Rules

If link $L(i,j)$ is collision-free at time slot t, node v_i can successfully transmit a packet to node v_j. However, v_i does not know that the packet has been received by v_j, and it may continue to retransmit the same packet in the following available time slots. So how can each node know when to stop retransmitting the same packet? In traditional MAC protocols such as CSMA/CA, an ACK message is sent back to the transmitter to acknowledge the reception of the transmitted packet. However, explicit ACK does not work well in the model proposed in this study since link $L(j,i)$ may not be collision-free at time slot t, and the ACK message may be lost with high probability. In RADG, two schemes, *Time-to-Live (TTL)* and *Aggregated Acknowledge (AAck)*, are designed to solve this problem.

- *Time-to-Live (TTL)*: When a node v_i receives a data packet, it associates the data

packet with a timer called TTL which controls the number of time slots allocated to v_i for transmitting a packet. It initializes the TTL timer with value of S_{i,F_i} and starts the timer once the data packet is added into the data transmitting queue. TTL decreases by one at each time slot. If TTL reaches 0, the corresponding packet is discarded. By Equation (5.2), S_{i,F_i} can guarantee that the packet can be transmitted to v_{F_i} with expected reliability $\tau(i, F_i)$. In practice, TTL can be assigned with a larger threshold than S_{i,F_i} to improve reliability. By adjusting the threshold assigned to TTL, it is easy to get tradeoffs between reliability and latency.

- *Aggregated Acknowledge (AAck)*: Each node v_i maintains a set denoted by R_i which records the source IDs of the packets (the source ID of a packet is the ID of the node that generated this packet) in its transmitting queue. Whenever v_i transmits a packet, R_i is piggybacked in the packet. If v_i receives a packet from its children, it adds the source ID of the packet into R_i and adds the packet into the transmitting queue if the packet is not a duplicate. On the other hand, when v_i overhears a packet from its next hop relay v_{F_i} (i.e., the father of v_i for tree networks), it extracts the set R_{F_i} from the packet. For any source ID that appears both in R_i and R_{F_i}, it means that the data packet transmitted by v_i has been successfully received by its father node v_{F_i}. Then v_i removes the source ID from R_i and removes the packet with the same source ID from its transmitting queue. This scheme makes full use of the broadcasting nature of wireless communication and does not introduce additional ACK message. Thus it does not exacerbate the degree of channel congestion and packets collision.

5.7.2   Duplicate Packet Elimination

In RADG, duplicate packets may be generated when a transmitter does not receive the Aggregated ACK from its father until the next available time slot for transmission. Without duplicate packet elimination, the duplicate packets will be forwarded, possibly causing more retransmissions, more contention, and more energy wastage. In RADG, the duplicate packets can be easily detected. When node v_i receives a packet from its children, it first checks if the source ID of the packet appears in R_i. If the source ID already exists in R_i, the received packet is a duplicate and is discarded; otherwise, v_i adds the source ID into R_i

Table 5.3: Simulation Parameters for Evaluating RADG

Tpye	Parameter	Value
Network	network area	$500m \times 500m$
	number of nodes	60
	maximum transmission range	60m
	channel bit rate	$(100 \sim 1000)kbps$
	data packet size	$(10 \sim 40)$bytes
RADG	slot time	$\frac{data-packet-size}{channel-bit-rate}$
	Per-link Transmission Reliability	$(0.8 \sim 1)$
CSMA	slot time	$20\ \mu s$
	DIFS	$50\ \mu s$
	SIFS	$10\ \mu s$
	RTS	20 bytes
	CTS	14 bytes
	ACK	14 bytes
	(CW_{min}, CW_{max})	(31,1023)

and adds the packet to its transmitting packet queue.

5.7.3 Periodical Sleeping

In data gathering applications, the duration between two consecutive data gathering rounds may be several seconds, minutes and even hours. Periodic node sleeping schemes have been widely used in many protocols such as S-MAC [96], T-MAC [34] and ACENT [27] to conserve energy in idle state. In RADG, each node enters into sleep state during the period between any two consecutive data gathering rounds.

5.8 Performance Evaluation

In this section, the performance of RADG is evaluated through extensive simulations, and RADG is compared with a CSMA-based protocol in terms of data gathering duration and communication overhead.

5.8.1 Simulation Setup

RADG is implemented in OMNeT++ [5] network simulator. All the simulations are run on the network given in Figure 5.9 in Section 5.6.2. The wireless sensor network was generated by randomly deploying 60 sensors into a $500m \times 500m$ region with a data sink located at the center of the region. The maximum transmission range for each node is set to 60m. The maximum number of neighbors for the nodes in the tree is 6, and the average number of neighbors is 4.3. A minimum-hop tree was constructed as the data gathering tree.

RADG is compared with CSMA which has been employed in IEEE 802.11 MAC Standard. Both CSMA with RTS/CTS handshaking and without RTS/CTS handshaking have been considered. The parameters are set to the default values in IEEE 802.11b, and the details of the parameter settings are given in Table 5.3. For RADG, the size of a time slot should be proportional to the size of a data packet to guarantee that one data packet can be transmitted in each time slot. In all simulations, the size of time slot in RADG is set to $\frac{data\ packet\ size}{channel\ bit\ rate}$. Since both RADG and CSMA are contention-based protocols, all the results obtained are the average values of 100 runs. RADG has advantages over CSMA based protocols when the user desired packet transmission reliability is less than 1. However, in CSMA based protocols, the link reliability is not tunable. To make fair comparisons, in all simulations, the user desired packet transmission reliability for each link is set to 1.

5.8.2 Data Gathering Duration *vs.* Packet Size

This simulation is designed to investigate the effect of data packet size on the performance of RADG and CSMA. Figure 5.12 presents the comparison of data gathering duration (T_{round}) under RADG and CSMA with the variation of data packet size. Since data packets in sensor networks are much smaller than those in wireless LANs, the size of data packet in this simulation is varied from 10 bytes to 40 bytes. As can be seen from Figure 5.12, T_{round} in RADG is much smaller than CSMA based schemes when the data packet size is small. For instance, when data packet size is 15 bytes, T_{round} in RADG has been reduced by 16.4% and 40.9% compared to CSMA without RTS/CTS and CSMA with RTS/CTS respectively. This phenomena can be explained as follows: Firstly, in CSMA based schemes,
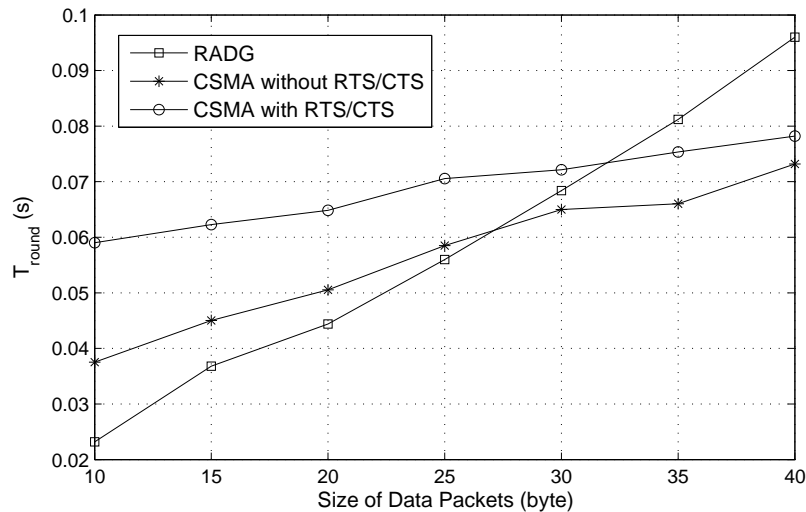
Figure 5.12: Effect of data packet size on the performance of RADG and CSMA

explicit ACK is employed to acknowledge the reception of data packets, and this acknowledgement scheme can result in much additional delay when data packet size is small. While in RADG, the reception of data packets is acknowledged using an aggregated-Ack scheme which is an implicit acknowledgement scheme, thus no additional delay is introduced. Secondly, in CSMA based schemes, DIFS and SIFS are employed to control the nodes' access to the medium. The delay incurred by these inter-frame spaces can not be ignored when the data packet is small. Thirdly, in CSMA with RTS/CTS, the exchanging of RTS/CTS message also adds significant delay for packet delivery. This is why RADG can significantly outperforms CSMA with RTS/CTS when the data packet is very small.

From Figure 5.12, it also can be seen that T_{round} in RADG increases rapidly with the increase in data packet size. When the data packet is larger than 27 bytes, the T_{round} achieved by CSMA without RTS/CTS is smaller than the T_{round} achieved by RADG. When data packet is larger than 32 bytes, the performance of CSMA with RTS/CTS is much better than RADG. This is because, in RADG, the size of the time slot is proportional to the size of the data packet so that one data packet can be transmitted in each time slot. When data collision occurs, a node has to wait for the next available time slot to retransmit the data packet. Therefore, data delivery may experience longer delay than CSMA based schemes when the data packet size is relatively large.

5.8.3 Amount of Data Transmitted *vs.* Data Packet size
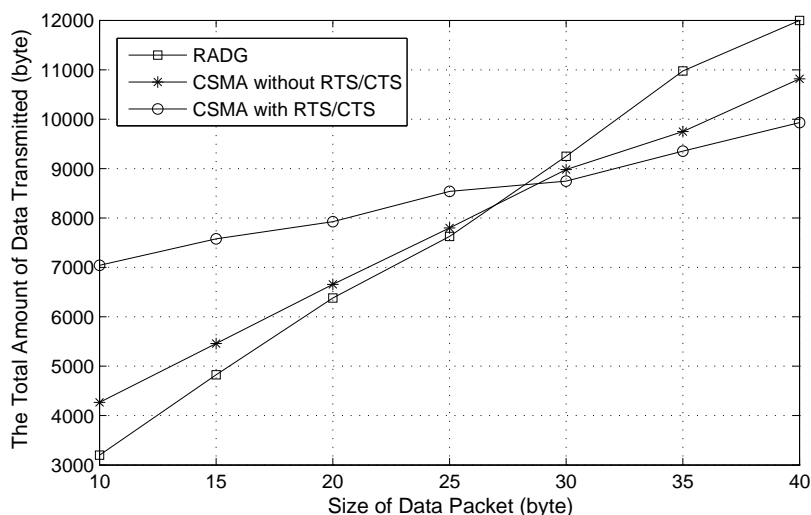
Figure 5.13: The amount of data transmitted under different data packet size

Figure 5.13 plots the average amount of data transmitted in one data gathering round against the size of the data packet. When the data packet size is small, the average amount of data broadcasted in RADG is less than that in CSMA without RTS/CTS. This is because CSMA employs explicit ACK, and the ACK message takes up a large portion of the total communication overhead when the data packet size is small. In RADG, since no explicit ACK messages are employed, the amount of data transmitted can be reduced. With the increase of data packet size, the average amount of data transmitted in RADG becomes larger than that in CSMA without RTS/CTS. This is because the ACK message has little effect on the total amount of data transmitted when the data packet size is relative large.

From Figure 5.13, it is worth noting that RADG significantly outperforms CSMA with RTS/CTS when the data packet size is small. The amount of data transmitted in CSMA with RTS/CTS is nearly 2 times of that in RADG when the data packet is 10 bytes. With the increase of the size of data packet, the amount of data transmitted in RADG increases quickly. The performance of CSMA without RTS/CTS is better than that of RADG when the data packet is larger than 20 bytes. This is because data transmission in RADG does not perform carrier sensing and frequent data collisions result in many data retransmissions. From Figure 5.13, it is also worth noting that CSMA with RTS/CTS outperforms CSMA without RTS/CTS when the data packet is larger than 30 bytes. Therefore, the RTS/CTS

handshaking can improve the performance of CSMA based schemes when the data payload is relatively large.
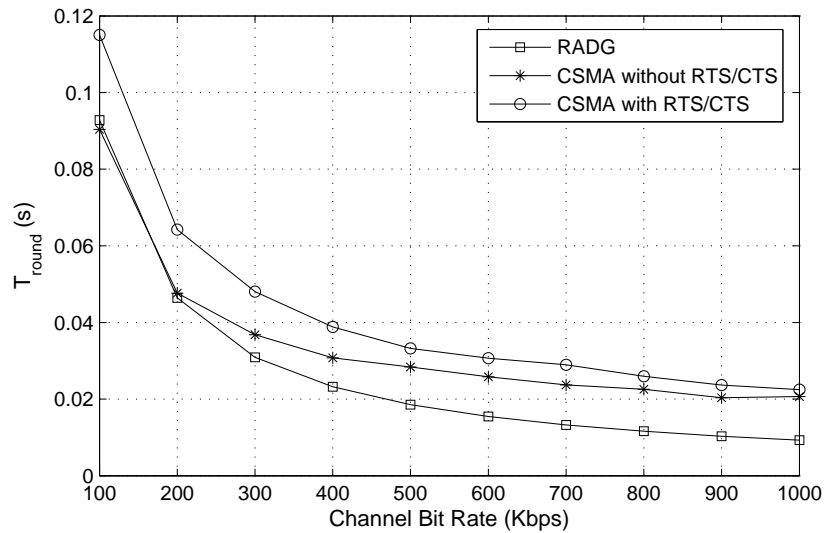
Figure 5.14: T_{round} under different channel bit rate

5.8.4 Impact of Channel Bit Rate

This simulation is designed to investigate the effect of channel bit rate on the performance of RADG and CSMA schemes. Figure 5.14 plots the data gathering duration T_{round} under RADG and CSMA with the channel bit rate when the data packet size is set to 20 bytes. It can be seen that T_{round} decreases with the increase of channel bit rate for all three schemes. When the channel bit rate is no larger than 200 $kpbs$, RADG has the same performance as CSMA without RTS/CTS. With the increase of channel bit rate, the performance of RADG is slightly better than the other two schemes. This is because that the time duration for transmitting one data packet decreases with the increase of channel bit rate. As discussed before, in CSMA based protocols, the inter-frame spaces and the explicit ACK scheme have a great impact on data delivery delay when the time duration for transmitting one data packet is small.

Figure 5.15 depicts the amount of data transmitted in the three schemes with the channel bit rate. As the channel bit rate increases, the duration of transmitting each data packet decreases. In CSMA based protocols, small data transmission duration can effectively reduce
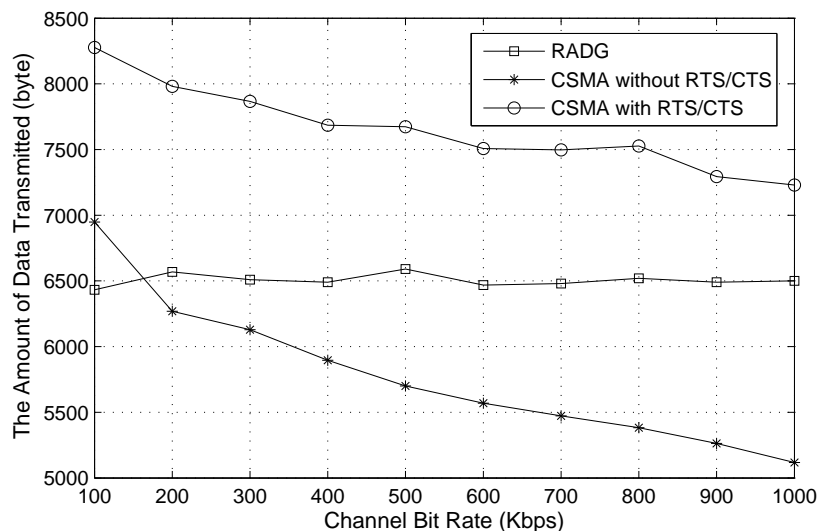
Figure 5.15: The amount of data transmitted under different channel bit rate

packet collisions. Thus the amount of data transmitted in CSMA based schemes decreases with the increase of channel bit rate. For RADG, the number of data packets transmitted only depends on the attempt probabilities of the nodes in the network. Therefore, channel bit rate has little impact on the amount of data transmitted in one data gathering round.

5.9   Summary and Future Work

This chapter investigated the problem of designing time-efficient random access scheme for data gathering in wireless sensor networks. First, a simple example was presented to illustrate the advantages of random access techniques over handshaking schemes for data gathering when data packets are small. Then a novel scheme based on random access was proposed for performing data gathering operation in WSNs. In this scheme, the data gathering problem is formulated as an optimization problem by considering both data transmission reliability and latency. The objective is to compute the optimal attempt probability for each node so that the expected data gathering duration can be minimized. Fully localized solutions, in which each node only needs to maintain two hop neighbor information, have been proposed for both linear and tree networks. Based on this model, a simple protocol called RADG which employs both data fusion and aggregated acknowledgement was designed for data gathering in sensor networks. Extensive simulations have been done to evaluate

the performance of RADG by comparing with CSMA based protocols. Simulations re-
sults show that RADG can significantly outperform RTS/CTS handshaking schemes when
data packets are small. With the increase of data packet size, the performance of RADG
degrades, and RTS/CTS handshaking can effectively improve system performance when
data payloads are relatively large. Thus RADG can be used for sensor networks deployed
for collecting some environmental parameters (such as temperature, humidity, etc) since
the data packets in such networks are very small. Moreover these applications also require
that the data can be collected within a time interval since the data collected exceed a time
interval may be invalid.

In this study, it is assumed that each node can fuse the data in its transmitting queue
into a fixed sized data packet based on data aggregation or network coding techniques
whenever it attempts a transmission. This approach can effectively reduce the data queuing
delay. However, it may also add additional complexity when a network coding scheme
is employed. So one future extension of this work is to solve the data gathering problem
based on random access technique by taking data queuing delay into account. Another
interesting area of work is to explore the tradeoff between communication overhead and
the data gathering latency based on random access techniques.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Wireless sensor networks have been identified as one of the most important technologies for the coming decades because of their unprecedented potential for a wide range of applications such as military sensing, traffic control, traffic surveillance, industrial and manufacturing automation, environment monitoring, building monitoring, etc. Different from conventional wireless networks, WSNs have special characteristics such as large number of sensor nodes and limited resources (e.g. energy, memory, computation power). These characteristics have posed great challenges for designing and implementing wireless sensor networks.

Providing efficient data transport is a critical issue in WSNs. This thesis focuses on designing efficient data transport schemes for WSNs. The main contents of this thesis are summarized as follows:

- Chapter 3 addresses the problem of providing energy-efficient sensor-to-sink routing in WSNs with dynamic network topology. An energy-efficient beaconless geographic routing scheme called EBGR is designed. EBGR can provide energy-efficient sensor-to-sink routing without the help of any neighborhood information. Extensive theoretical analysis and simulations have been conducted to evaluate the performance of EBGR. Simulation results show that EBGR significantly outperforms existing protocols in wireless sensor networks with highly dynamic network topol-

ogy.

- Chapter 4 addresses the problem of balancing energy consumption to maximize network lifetime for uniform data gathering sensor networks. An energy-balanced data gathering scheme called EBDG is designed. Combining the advantages of corona-based network division, mixed-routing and data aggregation, EBDG can provide balanced energy consumption among nodes both located in the same corona and located in different coronas. Simulation results show that EBDG can extend network lifetime by an order of magnitude compared with conventional schemes.

- Chapter 5 addresses the problem of providing time-efficient data gathering in WSNs with small data packets. A random access data gathering protocol called RADG is designed. As validated by simulations, RADG can outperform CSMA based schemes in WSNs with small data packets in terms of reducing the time duration for completing one round of data gathering.

In summary, this thesis provides efficient solutions for data transport in wireless sensor networks. The schemes presented in this thesis are relatively easy to implement, and can be directly used in a number of applications. The EBGR protocol can be directly applied in applications such as intrusion detection, target tracking, searching and rescuing, etc. With minor modifications, EBGR can be extended to underwater acoustic sensor networks and unmanned aerial vehicles (UAVs) networks. The EBDG protocol can be directly applied in applications such as environment monitoring, habitat monitoring, building monitoring, etc. The RADG protocol can be directly applied in applications such as building monitoring, health-care monitoring, home automation, industrial and manufacturing automation, etc. This thesis also provides useful foundations for addressing some open problems in designing data transport schemes for WSNs.

6.2 Future Work

There are still many areas to explore within this research topic. The energy-efficient beaconless geographic routing scheme proposed in Chapter 3 is a very flexible framework due

144

to its simplicity, and this scheme can be easily extended by taking other factors into account to further improve energy efficiency and extend network lifetime. For example, the factor of residual energy can also be utilized to guide forwarding in EBGR. When each node chooses its best next-hop relay, it selects a node with the largest residual energy in its relay search region as its next hop relay. This extension can alleviate the imbalance of energy consumption in the network, leading to significant network lifetime extension.

Most current research on balancing energy consumption is focused on uniform wireless sensor networks in which sensor nodes are uniformly deployed and all sensor nodes have uniform data generation rate. However, balancing energy consumption in non-uniform wireless sensor networks still remains an open problem. Since the data traffic pattern in non-uniform sensor networks is unpredictable, it is not easy to find optimal solutions to solve this problem. However, it is possible to design heuristic solutions to alleviate the unbalanced energy consumption. Future work will be carried out in this direction.

Providing flexible Quality of Service (QoS) is becoming more and more important for WSNs because sensor networks are application specific and different applications may have different QoS requirements. Due to the special characteristics of WSNs, an ideal protocol should take into account of factors such as available resources, channel quality, dynamically changing network topology to provide reliable and real-time data delivery. Since energy efficiency and real-time QoS support are both important to WSNs, it is desirable to design energy-efficient QoS-aware routing protocols for WSNs. Future work will be carried out in this direction.

Appendix A

Analysis of Characteristic Distance on Energy Consumption

In this appendix, the analysis of some characteristic distance on energy consumption. Lemma 19 gives the range of the characteristic distance in terms of minimizing the total energy consumption for delivering each packet to the sink.

Lemma 19. $d_o < \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}} < 2d_o, \quad when \quad k \geq 2.$

Proof. Let

$$f(k) = \frac{\sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}}}{\sqrt[k]{\frac{a_1}{a_2(k-1)}}} = \sqrt[k]{\frac{k-1}{1-2^{1-k}}} = \sqrt[k]{2^k\left(\frac{k-1}{2^k-2}\right)}.$$

When $k \geq 2$, $1 - 2^{1-k} < k - 1 < 2^k - 2$. Thus $1 < f(k) < \sqrt[k]{2^k} = 2$. By Lemma 1, $d_o = \sqrt[k]{\frac{a_1}{a_2(k-1)}}$. Therefore, $d_o < \sqrt[k]{\frac{a_1}{a_2(1-2^{1-k})}} < 2d_o$. $\qquad\square$

Appendix B

Description of Parameters Used in Chapter 3

Parameter	Description
R	Maximum Transmission Range
$\|uv\|$	Euclidean distance between node u and node v
a_{11}	energy spent by transmitter electronics for transmitting one bit data
a_{12}	energy spent by receiver electronics for receiving one bit data
a_1	$a_{11} + a_{12}$
a_2	transmitting amplifier
k	propagation loss exponent
$\varepsilon_{tx}(x)$	the energy spent by transmitting one bit data over distance x
ε_{rx}	the energy spent by receiving one bit data
$\varepsilon_{relay}(x)$	the energy spent by relaying one bit data
(x_u, y_u)	the geographic location of node u
$\xi(d)$	the total energy spent on delivering one bit data from u to v where $\|uv\| = d$
d_o	the optimal hop distance $d_o = \sqrt[k]{\frac{a_1}{a_2(k-1)}}$
f_u	the optimal relay position of node u
$\mathbf{R}_u$	the relay research region of node u
$r_s(u)$	the radius of $\mathbf{R}_u$
n_u	the next relay of node u
γ	the delay for transmitting a packet over a unit distance
$\delta_{v->u}$	the delay that node v broadcasts a reply message to node u
S_i	the ith concentric ring in the relay search region
$P(u, v)$	the progress that node u obtained by forwarding its packets to node v

$A(u, v)$	the advance that node u obtained by forwarding its packets to node v
$\gamma_{P(u,v)}$	the energy over progress ratio
$\gamma_{P(u,v)}$	the energy over advance ratio
$C(u)$	the minimum relay search region of node u that covers only one active node
r	the radius of $C(u)$
N	the number of hops in the routing path
r_{ul}	ratio of the upper bound to the lower bound on energy consumption
$P_w(d)$	the probability that the worst case happens for delivery packets over distance d
ρ	active nodes distribution density
r_{el}	ratio of the approximated expected energy consumption to the lower bound on energy consumption
$PRR(u, v)$	the packet reception rate for link (u, v)
δ	the threshold for blacklisting

Appendix C

Description of Parameters Used in Chapter 4

Parameter	Description
$\varphi(x)$	aggregation function where $\varphi(x) = mx + c$
C_i	the ith corona
r	the width of each corona
T	network lifetime
$F(u)$	the amount of data forwarded by node u in hop-hop-hop transmission mode
$D(u)$	the amount of data transmitted in direct transmission mode
$p(u)$	the data distribution ratio of node u
p_i	all nodes in C_i have the same data distribution ratio p_i
$\varepsilon_t(ir)$	energy spent by a node in C_i to transmit one bit data to sink
$\varepsilon_t(r')$	energy spent by fording one bit data to nodes in next corona
$S(u)$	the total amount of data received by node u
$E(u)$	the total energy consumption of node u
ρ	node deployment density
N_{C_i}	the number of nodes in C_i
$C_{i,j}$	the jth sub-corona in C_i
$\Delta_{i,j}$	the width of sub-corona $C_{i,j}$
$Z_{i,j,k}$	the kth zone in sub-corona $C_{i,j}$
$A_{Z_{i,j,k}}$	the area size of zone $Z_{i,j,k}$
$r_{i,j}$	the outer boundary of sub-corona $C_{i,j}$ $r_{i,j} = (i-1)r + \sum_{k=1}^{j} \Delta_{i,k}$
N	the number of hops in the routing path
r_{ul}	the ratio of the upper bound to the lower bound on energy consumption

$P_w(d)$	the probability that the worst case happens for delivery packets over distance d
ρ	active nodes distribution density
r_{el}	the ratio of the approximated expected energy consumption to the lower bound on energy consumption
$PRR(u,v)$	the packet reception rate for link (u,v)
δ	the threshold for blacklisting

Appendix D

Description of Parameters Used in Chapter 5

Parameter	Description
v_i	node i
α_i	attempt probability for node v_i
$p(k)$	probability that k nodes transmit simultaneously at a time slot
$E[N_{slot}]$	expected number of messages transmitted in one time slot
$E[N]$	expected number of messages transmitted in one data gathering round
$L(i,j)$	link from v_i to v_j
$N_1(i)$	set of nodes within one hop of node v_i
$N_2(i)$	set of nodes within two hops of node v_i
$P_L(i,j)$	probability that a time slot is collision-free for $L(i,j)$
$\gamma(i,j)$	expected reliability for packet transmission over link $L(i,j)$
$E[S_{i,j}]$	expected number of time slots needed for transmitting a data packet from v_i to v_j with reliability $\gamma(i,j)$
$\psi(i)$	set of links on the path from v_i to the sink
$E[Delay(i)]$	expected delay for delivering a packet from v_i to the sink
T_{round}	expected duration for one data gathering round
LS	set of links that are useful for data gathering in the linear network
$f(\alpha_i)$	expected latency for delivering a packet along the links that α_i influences

α_i^*	optimal α_i in terms of minimizing T_{round}
$\overline{\alpha_i}$	attempt probability v_i uses during the process of attempt probability computation
$Ts(i)$	maximum expected time slots for the descendants of v_i to deliver a packet to v_i
F_i	father of v_i in the data gathering tree
C_i	set of children of node v_i
D_i	child of v_i that $Ts(D_i) + E[S_{D_i,i}] = \max_{k \in C_i}(Ts(k) + E[S_{k,i}])$
TTL	Time-to-Live timer used to control the number of time slots allocated for a data transmission

Bibliography

[1] "Calibration and tracking in sensor networks," *http://people.csail.mit.edu/rahimi/projects/sensor-networks/*.

[2] "Crossbow, mica2 wireless measurement system datasheets," in *http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/6020-0042-04_A_MICA2.pdf*.

[3] "Dustnetwork, smartmesh-xt m2135, wireless measurement system datasheets," in *http://www.dustnetworks.com/docs/M2135.pdf*.

[4] "http://web.syr.edu/ gkamathh/topics/micamote.html."

[5] "http://www.omnetpp.org/index.php."

[6] "Monitoring volcanic eruptions with a wireless sensor network," in *http://www.eecs.harvard.edu/ mdw/proj/volcano/*.

[7] "New sensor technology advances australia's water management," in *http://www.csiro.au/news/ps1k9.html*.

[8] "Smart farm project in csiro ict, australia," in *http://www.sensornets.csiro.au/belmont.htm*.

[9] "Soil moisture monitoring with wireless sensor networks," in *http://www.csse.uwa.edu.au/adhocnets/WSNgroup/soil-water-proj/*.

[10] "Tinyos 2.0.2 documentation," in *http://www.tinyos.net/scoop/special/support*.

[11] N. Abramson, "The aloha system - another alternative for computer communication," in *AFIPS Conference Proceedings*, 1970, pp. 295 – 298.

[12] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, pp. 1204–1216, 2000.

[13] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, pp. 102 – 114, 2002.

[14] J. N. Al-karaki and A. E. Kamal, "Routing techniques in wireless sensor network: A survey," *IEEE Wireless Communications*, pp. 6 – 28, 2004.

[15] E. Altman, V. Borkar, and A. Kherani, "Optimal random access in networks with two-way traffic," in *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2004, pp. 609 – 613.

[16] G. D. Bacco, T. Melodia, and F. Cuomo, "A mac protocol for delay-bounded applications in wireless sensor networks," in *Proceedings of the Mediterranean Ad Hoc Networking Workshop (MedHocNet)*, 2004.

[17] H. Bai, M. Atiquzzaman, and D. Lilja, "Wireless sensor network for aircraft health monitoring," in *Proceedings of the First International Conference on Broadband Networks (BROADNETS)*, 2005, pp. 748–750.

[18] L. Barrire, P. Fraigniaud, L. Narayanan, and J. Opatrny, "Robust position-based routing in wireless ad hoc networks with irregular transmission ranges," *Wireless Communications and Mobile Computing*, vol. 3, pp. 141–153, 2003.

[19] M. Bhardwaj and A. P. Chandrakasan, "Bounding the lifetime of sensor networks via optimal role assignments," in *21th IEEE International Conference on Computer Communications (INFOCOM)*, 2002, pp. 1587–1596.

[20] M. Bhardwaj, T. Garnett, and A. P. Chandrakasan, "Upper bounds on the lifetime of sensor networks," in *Proceedings of IEEE International Conference on Communications (ICC)*, 2001, pp. 785–790.

[21] Y. Bi, L. Sun, H. Zhu, T. Yan, and Z. Luo, "A parking management system based on wireless sensor network," *ACTA AUTOMATICA SINICA*, vol. 32, pp. 968 – 977, 2006.

[22] B. Blum, T. He, S. Son, and J. Stankovic, "IGF: A state-free robust communication protocol for wireless sensor networks," in *University of Virinia, USA, Tech. Rep. CS-2003-11*, 2003.

[23] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks," *Wireless Networks*, vol. 7, pp. 609 – 616, 2001.

[24] H. T. K. Brad Karp, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th annual international conference on Mobile computing and networking(MobiCom)*, 2000, pp. 243–254.

[25] M. Busse, T. Haenselmann, and W. Effelsberg, "Energy-efficient forwarding schemes for wireless sensor networks," in *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2006, pp. 125 – 133.

[26] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao, "Habitat monitoring: Application driver for wireless communications technology," in *Proceedings of ACM SIGCOMM Workshop on Data Communications*, 2001, pp. 20–41.

[27] A. Cerpa and D. Estrin, "ASCENT: Adaptive self-configuring sensor networks topologies," *IEEE Transactions on Mobile Computing*, vol. 3, pp. 272–285, 2004.

[28] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Proceedings of the ACM SIGCOMM 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2007, pp. 169–180.

[29] J.-H. Chang and L. Tassiulas, "Routing for maximum system lifetime in wireless adhoc networks," in *Proceedings of 37th Annual Allerton Conference on Communication, Control and Computing*, 1999.

[30] M. Chawla, N. Goel, K. Kalaichelvan, A. Nayak, and I. Stojmenovic, "Beacon-less position based routing with guaranteed delivery for wireless ad-hoc and sensor networks," in *IFIP International Federation for Information Processing*, 2006, pp. 61–70.

[31] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "SPAN: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," *Wireless Networks*, vol. 8, pp. 481 – 494, 2002.

[32] S. Chen and Z. Zhang, "Localized algorithm for aggregate fairness in wireless sensor networks," in *Proceedings of the 12th annual international conference on Mobile computing and networking (MobiCom)*, 2006, pp. 274–285.

[33] R. R. Choudhury, A. Chakravarty, and T. Ueda, "Implicit mac acknowledgment: An improvement to 802.11," in *Proceedings of the 4th IEEE/ACM Wireless Telecommunications Symposium (WTS)*, 2005.

[34] V. Dam, Tijs, and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Proceedings of the First ACM Conference on Embedded Networked Sensor Systems*, 2003, pp. 171–180.

[35] I. Demirkol, F. Alagoz, H. Delic, and C. Ersoy, "Wireless sensor networks for intrusion detection: Packet traffic modeling," *IEEE Communications Letters*, vol. 10, pp. 22 – 24, 2006.

[36] O. Dousse, C. Tavoularis, and P. Thiran, "Delay of intrusion detection in wireless sensor networks," in *Proceeding of the 7th ACM international symposium on Mobile ad hoc networking and computing(MobiHoc)*, 2006, pp. 155 – 165.

[37] C. Efthymiou, S. Nikoletseas, and J. Rolim, "Energy balanced data propagation in wireless sensor networks," in *Proceedings of the 18th International Parallel and Distributed Processing Symposium (IPDPS)*, 2004, p. 225a.

[38] C. fan Hsin and M. Liu, "Network coverage using low duty-cycled sensors: random and coordinated sleep algorithms," in *Proceedings of the third international symposium on Information processing in sensor networks (IPSN)*, 2004, pp. 433 – 442.

[39] G. G. Finn, "Routing and addressing problems in large metropolitan-scale internetworks," Technical Report ISI/RR-87-180, Tech. Rep., 1987.

[40] H. Frey and I. Stojmenovic, "On delivery guarantees of face and combined greedy-face routing in ad hoc and sensor networks," in *14th IEEE International Conference on Network Protocols(MobiCom)*, 2006, pp. 390–401.

[41] H. Füler, J. Widmer, M. Käsemann, M. Mauve, and H. Hartenstein, "Contention-based forwarding for mobile ad hoc networks," *Ad Hoc Networks*, vol. 1, pp. 351–369.

[42] A. Giridhar and P. R. Kumar, "Maximizing the functional lifetime of sensor networks," in *Proceedings of The Fourth International Conference on Information Processing in Sensor Networks (IPSN)*, 2005, pp. 5– 12.

[43] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Computers and Operations Research*, vol. 13, pp. 533 – 549, 1986.

[44] V. C. Gungor and zgr B. Akan, "Delay aware reliable transport in wireless sensor networks," *International Journal of Communication Systems*, vol. 20, pp. 1155 – 1177, 2007.

[45] W. Guo, Z. Liu, and G. Wu, "An energy-balanced transmission scheme for sensor networks," in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys)*, 2003, pp. 300 – 301.

[46] Haenggi M., "Energy-balancing strategies for wireless sensor networks," in *Proceedings of the 2003 International Symposium on Circuits and Systems (ISCAS)*, 2003, pp. 828–831.

[47] Hauspie M and Panier A, "Localized probabilistic and dominating set based algorithm for efficient information dissemination in ad hoc networks," in *IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, 2004, pp. 11 – 20.

[48] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proceedings of the 9th*

annual international conference on Mobile computing and networking (MobiCom),
2003, pp. 81–95.

[49] T. He, P. Vicaire, T. Yan, L. Luo, L. Gu, G. Zhou, R. Stoleru, Q. C. J. A. Stankovic, and T. Abdelzaher, "Achieving real-time target tracking using wireless sensor networks," in *Proceedings of the 12th IEEE Real-Time and Embedded Technology and Applications Symposium*, 2006, pp. 37 – 48.

[50] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 660–670, 2002.

[51] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Hawaii International Conference on System Sciences*, 2000, pp. 4–7.

[52] M. Heissenbttel, T. Braun, T. Bernoulli, and M. Wlchli, "BLR: beacon-less routing algorithm for mobile ad hoc networks," *Computer Communications*, vol. 11, pp. 1076–1086, 2004.

[53] P. Hii and A. Zaslavsky, "Improving location accuracy by combining wlan positioning and sensor technology," in *Workshop on Real-World Wireless Sensor Networks*, 2005.

[54] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proceedings of IEEE International Symposium on Information Theory*, 2003, pp. 442 – 448.

[55] Y. L. J. H. R. Holte, "Optimal traffic-oblivious energy-aware routing for multihop wireless networks," in *25th IEEE International Conference on Computer Communications (INFOCOM)*, 2006, pp. 1–12.

[56] T. Hou, Y. Shi, and H. Sherali, "Rate allocation in wireless sensor networks with network lifetime requirement," in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2004, pp. 67–77.

[57] Y. T. Hou, Y. Shi, J. Pan, and S. F. Midkiff, "Maximizing the lifetime of wireless sensor networks through optimal single-session flow routing," *IEEE Transactions on Mobile Computing*, vol. 5, pp. 1255–1266, 2006.

[58] I. Howitt and J. Wang, "Energy balanced chain in distributed sensor networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, 2004, pp. 1721 – 1726.

[59] A. Jarry, P. Leone, O. Powell, and J. Rolim, "An optimal data propagation algorithm for maximizing the lifespan of sensor networks," in *Proceedings of Distributed Computing in Sensor Systems (DCOSS)*, 2006, pp. 405–421.

[60] H. Kalosha, A. Nayak, S. Rhrup, and I. Stojmenovi, "Select-and-protest-based beaconless georouting with guaranteed delivery in wireless sensor networks," in *Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM)*, 2008, pp. 346–350.

[61] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Maximum lifetime data gathering and aggregation in wireless sensor networks," in *Proceedings of the IEEE International Conference on Networking*, 2002, pp. 685–696.

[62] A. Karnik and A. Kumar, "Distributed optimal self-organisation in a class of wireless sensor networks," in *23th Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, 2004.

[63] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Mdard, and J. Crowcroft, "XORs in the air: practical wireless network coding," in *Proceedings of the ACM SIGCOMM 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2006, pp. 243 – 254.

[64] A. Keshavarzian, E. Uysal-Biyikoglu, F. Herrmann, and A. Manjeshwar, "Energy-efficient link assessment in wireless sensor networks," in *23th Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, 2004, pp. 1751–1761.

[65] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part I - carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Transactions on Communications*, vol. 23, pp. 1400–1416, 1975.

[66] K. D. Konstantinos Kalpakis and P. Namjoshi, "Maximum lifetime data gathering and aggregation in wireless sensor networks," in *Proceedings of IEEE International Conference on Networking (ICN)*, 2002, pp. 685–696.

[67] F. Kuhn, R. Wattenhofer, Y. Zhang, and A. Zollinger, "Geometric ad-hoc routing: of theory and practice," in *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, 2003, pp. 63–72.

[68] S. Kumar, T. H. Lai, and J. Balogh, "On k-coverage in a mostly sleeping sensor network," in *Proceedings of the 10th annual international conference on Mobile computing and networking (MobiCom)*, 2004, pp. 144 – 158.

[69] J. Kuruvila, A. Nayak, and I. Stojmenovic, "Hop count optimal position based packet routing algorithms for ad hoc wireless networks with a realistic physical layer," in *Proceedings of the 1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, 2004, pp. 398–405.

[70] S. Kwon and N. B. Shroff, "Geographic routing in the presence of location errors," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 50, pp. 2902 – 2917, 2006.

[71] L. Lazos and R. Poovendran, "Stochastic coverage in heterogeneous sensor networks," *ACM Transactions on Sensor Networks*, vol. 2, pp. 325 – 358, 2006.

[72] S. Lee, B. Bhattacharjee, and S. Banerjee, "Effcient geographic routing in multi-hop wireless networks," in *the 6th ACM international symposium on Mobile ad hoc networking and computing(MobiHoc)*, 2005, pp. 230 – 241.

[73] B. Leong, S. Mitra, and B. Liskov, "Path vector face routing: geographic routing with local face information," in *13th IEEE International Conference on Network Protocols(MobiCom)*, 2005, pp. 12–22.

[74] C. Li, M. Ye, G. Chen, and J. Wu, "EECS: an energy efficient clustering scheme in wireless sensor networks," in *Proceedings of 24th IEEE International Performance, Computing, and Communications Conference (IPCCC)*, 2005, pp. 535 – 540.

[75] J. Li and P. Mohapatra, "Analytical modeling and mitigation techniques for the energy hole problem in sensor networks," *Pervasive and Mobile Computing*, vol. 3, pp. 233–254, 2007.

[76] Q. Li, J. Aslam, and D. Rus, "Distributed energy-conserving routing protocols for sensor network," in *Proceedings of the IEEE the 36th Hawaii International Conference on System Science*, 2003.

[77] J. Lian, K. Naik, and G. B. Agnew, "Data capacity improvement of wireless sensor networks using non-uniform sensor distribution," *International Journal of Distributed Sensor Networks*, vol. 2, pp. 121 – 145, 2006.

[78] S. Lindsey, C. Raghavendra, and K. Sivalingam, "Data gathering in sensor networks using the energy*delay metric," in *Parallel and Distributed Processing Symposium., Proceedings of the 15th International Parallel and Distributed Processing Symposium (IPDPS)*, 2001, pp. 2001 – 2008.

[79] Y. Liu, H. Ngan, and L. M. Ni, "Power-aware node deployment in wireless sensor networks," in *Proceedings of IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC)*, 2006, pp. 128–135.

[80] G. Lu, B. Krishnamachari, and C. Raghavendra, "An adaptive energy-efficient and low-latency mac for data gathering in sensor networks," *Wireless Communications and Mobile Computing*, vol. 7, pp. 863 – 875, 2007.

[81] J. Luo and J.-P. Hubaux, "Joint mobility and routing for lifetime elongation in wireless sensor networks," in *24th Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, 2005, pp. 1735– 1746.

[82] J. Martin, "Communication satellite systems," *Prentice Hall, New Jersey*, 1978.

[83] S. Megerian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, "Worst and best-case coverage in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 4, pp. 84–92, 2005.

[84] T. Melodia, D. Pompili, and I. F. Akyildiz, "Optimal local topology knowledge for energy efficient geographical routing in sensor networks," in *Proceedings of the 24th Conference on Computer Communications (INFOCOM)*, 2004.

[85] V. Mhatre and C. Rosenberg, "Design guidelines for wireless sensor networks: communication, clustering and aggregation," *Ad Hoc Networks*, vol. 2, pp. 45–63, 2004.

[86] V. P. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, and N. Shroff, "A minimum cost heterogeneous sensor network with a lifetime constraint," *IEEE Transactions on Mobile Computing*, vol. 4, pp. 4–15, 2005.

[87] Y. Min, K. Kirill, M. Sameer, Sundresh, W. Kim, and G. Agha, "Resilient localization for sensor networks in outdoor environments," in *the 25th IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2005, pp. 643 – 652.

[88] T. Murata and H. Ishibuchi, "Performance evaluation of genetic algorithms for flow-shopscheduling problems," in *Proceedings of the First IEEE Conference on Evolutionary Computation*, 1994, pp. 812–817.

[89] E. Nga, Y. Zhou, M. Lyu, and J. Liu, "Reliable reporting of delay-sensitive events in wireless sensor-actuator networks," in *Proceeding of IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2006, pp. 101–108.

[90] S. Olariu and I. Stojmenovic, "Design guidelines for maximizing lifetime and avoiding energy holes in sensor networks with uniform distribution and uniform reporting," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, 2006, pp. 1 – 12.

[91] O. Powell, P. Leone, and J. Rolim, "Energy optimal data propagation in wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 67, pp. 302–317, 2007.

[92] I. S. Prosenjit Bose, Pat Morin and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks," in *Proceedings of the 3rd ACM International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, 1999, pp. 48–55.

[93] M. R, S. Cui, L. S., and A. Goldsmith, "Cross-layer design for lifetime maximization in interference-limited wireless sensor networks," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, 2005, pp. 1964– 1975.

[94] A. Rao, S. Ratnasamy, C. Papadimitriou, S. Shenker, and I. Stoica, "Geographic routing without location information," in *the 9th annual international conference on Mobile computing and networking(MobiCom)*, 2003, pp. 96 – 108.

[95] S. Ray, R. Ungrangsi, F. D. Pellegrini, A. Trachtenberg, and D. Starobinski, "Robust location detection in emergency sensor networks," in *IEEE Conference on Computer Communications (INFOCOM)*, 2003, pp. 1044–1053.

[96] L. G. Roberts, "Aloha packet system with and without slots and capture," *ACM SIGCOMM Computer Communication Review*, vol. 5, pp. 28 – 42, 1975.

[97] M. Sanchez and P. Manzoni, "A java-based ad hoc networks simulator," in *Proceedings of the SCS Western Multiconference Web-based Simulation Track*, 1999.

[98] J. Sarker and S. Halme, "Auto-controlled algorithm for slotted aloha," in *IEE Proceedings - Communications*, 2003, pp. 53–58.

[99] C. Savarese, J. Rabaey, and J. Beutel, "Locationing in distributed ad-hoc wireless sensor networks," in *Proceedings of the International Conference on Acoustics*, 2001, pp. 2037–2040.

[100] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proceedings of the 7th annual international conference on Mobile computing and networking(MobiCom)*, 2001, pp. 166 – 179.

[101] C. Schurgers and M. B. Srivastava, "Energy efficient routing in wireless sensor networks," in *Proceedings of Military Communications Conference (MILCOM)*, 2001, pp. 357– 361.

[102] K. Seada and A. Helmy, "Geographic protocols in sensor networks," Electrical Engineering Department, University of Southern California, Tech. Rep., 2004.

[103] K. Seada, M. Zuniga, A. Helmy, and B. Krishnamachari, "Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks," in *Proceedings of the 2nd international conference on Embedded networked sensor systems (SenSys)*, 2004, pp. 108–121.

[104] R. C. Shah, "Distributed algorithms to maximize the lifetime of wireless sensor networks," in *http://citeseer.ist.psu.edu/shah05distributed.html*, 2005.

[105] S. Singh, M. Woo, and C. S. Mghavendra, "Power-aware routing in mobile ad hoc networks," in *Proceedings of the 4th ACM/IEEE Interantional Conference on Mobile Computing and Networking (MobiCom)*, 1998, pp. 181–190.

[106] S. Soro and W. B. Heinzelman, "Prolonging the lifetime of wireless sensor networks via unequal clustering," in *Proceedings of IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2005, p. 236.2.

[107] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Performance analysis of mobility-assisted routing," in *7th ACM international symposium on Mobile ad hoc networking and computing(MobiHoc)*, 2006, pp. 49 – 60.

[108] I. Stojmenovic, "Localized network layer protocols in wireless sensor networks based on optimizing cost over progress ratio," *IEEE Network*, pp. 21–27, 2006.

[109] I. Stojmenovic and X. Lin, "Power-aware localized routing in wireless networks," in *Proceeding of the 14th International Parallel and Distributed Processing Symposium (IPDPS)*, 2000, p. 371.

[110] ——, "Power-aware localized routing in wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 12, pp. 1122–1133, 2001.

[111] H. Takagi and L. Kleinrock, "Optimal transmission ranges for randomly distributed packet radio terminals," *IEEE Transactions on Communications*, vol. 32, p. v, 1984.

[112] Y. Tirta, B. Lau, N. Malhotra, S. Bagchi, and Z. L. adn Yung-Hsiang Lu, "Controlled mobility for efficient data gathering in sensor networks with passively mobile nodes," *Sensor Network Operations*, vol. 3.2, pp. 92–113, 2006.

[113] S. Vasudevan, J. Kurose, and D. Towsley, "On neighbor discovery in wireless netorks with directional antennas," in *24th Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, 2005, pp. 2502– 2512.

[114] Venkata K., Prasanthi M., and A. Kumar, "Optimizing delay in sequential change detection on ad hoc wireless sensor networks," in *Proceeding of the 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON)*, 2006, pp. 306 – 315.

[115] A. Viterbi, "CDMA: Principles of spread spectrum communication," *Addison Wesley, Reading*, 1995.

[116] X. Wang and K. Kar, "Distribtuted algorithm for max-min fair rate allocation in aloha networks," in *41th Annual Allerton Conference*, 2004.

[117] L. A. Wolsey and G. L. Nemhauser, in *Integer and Combinatorial Optimization*, 1998.

[118] S. Wu and K. S. Candan, "GPER: Geographic power efficient routing in sensor networks," in *Proceedings of the IEEE International Conference on Network Protocols (ICNP)*, 2004, pp. 161–172.

[119] X. Wu, G. Chen, and S. K. Das, "On the energy hole problem of nonuniform node distribution in wireless sensor networks," in *Proceedings of IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2006, pp. 180–187.

[120] G. Xing, C. Lu, and Q. Huang, "Greedy geographic routing is good enough in sensing covered networks," Washington University, Tech. Rep., 2004.

[121] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in *Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, 2001, pp. 70–84.

[122] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated, adaptive sleeping for wireless sensor networks," *ACM/IEEE Transactions on Networking*, vol. 12, pp. 493–506, 2004.

[123] H. Yin and H. Liu, "Distributed rate adaptive packet access (DRAPA) for multicell wireless networks," *IEEE Transactions on Wireless Communications*, vol. vol.3, pp. 432–441, 2004.

[124] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, pp. 366–379, 2004.

[125] Y. Yu, B. Krishnamachari, and V. K. Prasanna, "Energy-latency tradeoffs for data gathering in wireless sensor networks," in *23th Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, 2004.

[126] H. Zhang, H. Shen, and Y. Tan, "Optimal energy balanced data gathering in wireless sensor networks," in *Proceedings of the 21th International Parallel and Distributed Processing Symposium (IPDPS)*, 2007, pp. 1 – 10.

[127] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *the 1st ACM Conference on Embedded Networked Sensor Systems (Sensys)*, 2003, pp. 1 – 13.

[128] M. Zorzi, "A new contention-based MAC protocol for geographic forwarding in ad hoc and sensor networks," in *Proceedings of IEEE International Conference on Communications (ICC)*, 2004, pp. 3481 – 3485.