Cost - Effective Lifetime - Oriented Network Planning in Wireless Sensor Networks

by

QUANHONG WANG

A thesis submitted to

the Department of Electrical and Computer Engineering

in conformity with the requirements for

the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
September, 2006

Your file Votre référence
ISBN: 978-0-494-18561-2
Our file Notre référence
ISBN: 978-0-494-18561-2

To my precious son - Xintai (LeLe)

给我亲爱的儿子 – 新泰（乐乐）

ABSTRACT

The limited energy supply and constrained capacities of individual devices raise many technical challenges in the efforts to implement an effective and efficient Wireless Sensor Network (WSN). In this thesis, we conduct an in-depth study on one of the fundamental design issues, namely network planning, in which types, numbers, and locations of heterogeneous devices are determined so that the total system cost is minimized while the requirements of lifetime, coverage, and connectivity are satisfied. To tackle this intricate problem, we propose a comprehensive and modular network planning framework, which is decomposed into a sensing domain and a communication domain.

In the sensing domain, we introduce the novel concept of information-oriented coverage, which measures the sensing quality by the information utility obtained from the sensing devices. We propose a couple of heuristic schemes to deploy a minimum number of sensing devices to achieve a desirable sensing quality.

The communication domain, formed by relaying devices, must ensure accurate delivery of all sensed information to the base station over a desirable lifespan. Due to the complexity of the general relaying device placement problem, we split the design of the communication domain into two phases.

In the first phase, we use a minimum set covering model to characterize the problem of placing a minimum number of relaying devices to ensure connectivity and lifetime requirements of each sensing device, and propose a recursive algorithm to obtain the optimal solution.

In the second phase, we propose Far-Near and Max-Min principles to make locally optimal decisions to place a minimum number of extra relaying devices, such that all data collected at the first phase relaying devices can be forwarded to the BS for a given duration. We envision two scenarios – either a relaying device has fixed or variable transmission range. Several schemes are proposed for the placement of the relaying devices in each scenario. Furthermore, a lower bound on the minimum number of second phase relaying devices is derived for each case above. Extensive simulations over different scenarios verify the effectiveness of our proposed schemes.

The network planning problems introduced in this work have not been considered previously and their formulations are novel. The devised techniques can serve as guidelines for WSN designers, solution providers, and system integrators of WSN applications.

ACKNOWLEDGEMENTS

I become myself today in completion of this thesis because of the people who have lent me a hand in the past years. I owe tremendous thanks to them for offering me hope, cheer, courage, and strength during this unforgettable journey.

First to Dr. Hossam Hassanein and Dr. Glen Takahara, my wonderful supervisors and mentors over the past four years. Working with them has been a pleasant experience. Their constant encouragement made me more self-confident; their wisdom led me to search new knowledge; their vision for the future greatly inspired me. I could always find them for guidance whenever I was lost or stuck. I am grateful to Dr. Hassanein and Dr. Takahara for the countless times they have spent with me: instructing my research, correcting my papers, and discussing every aspect of being a researcher, from career development to balancing school and family life. Their advice and comments have significantly enhanced my work. Their invaluable support during my time at Queen's University has enriched my life both academically and personally.

To Dr. Selim Akl whose breadth of knowledge and enthusiasm for research taught me what it means to be an educator and researcher. My work has been strengthened greatly from the discussions with him in the Special Interest Group for sensor networks.

I am thankful to Dr. Hussein Mouftah, Dr. Mohamed Ibnkahla and Dr. Chi-Hsiang Yeh. There are many excellent educators at Queen's, but I am fortunate enough to study with them. I appreciate them for showing me the value of a great teacher.

To Dr. Peter Marbach, who taught me my first course in computer networks and communications. His teaching has had an indelible impact on my study and career – it has inspired me to pursue my research in the related area.

To Dr. Robert Benkoczi, Dr. Baoxian Zhang, and Dr. Jun Zheng, I thank them for their advice, encouragement, and patience.

Thanks go to Dr. Steven Blostein, Ms. Bernice Ison, and Ms. Debra Fraser, for all their support and help with administrative issues in the Department of Electrical and Computer Engineering at Queen's University.

The talented and bright people and alumni of the Telecommunication Research Lab (TRL) at Queen's University have taught me much in the area of wireless communication. Their comments, feedback, and friendship have made my Ph.D process more fruitful and enjoyable. Thanks due to Dr. Nidal Nasser, Dr. Tiantong You, Hung Tam, Sylvia Tai, Hongyan Du, Jerry Yang, Christina Luo, Ayman Radwan, Waleed Al-Salih, Ahmed Hasswa, Abdulrahman Hijazi, and Bader Al-Manthari. My special gratitude goes to Afzal Mawji, for his editorial help of the thesis; and to Abd-Elhamid M. Taha, for his constant encouragement and help. I wish especially to thank Kenan Xu. He has always been available to share his wisdom and experience. My work has highly benefited from his suggestions, encouragement, and collaboration. Thank you Kenan for making my journey of study at Queen's University more joyful.

I am particularly blessed with a collection of friends. Thanks to Yumei Cao, Guihong Cui, Yinghong Fan, Quanhong Li, Dongqun Wu, Amy Y. Zhang, Hong Zhang, and Jianping Zhao (Mao Mao), for being my soul sisters, for always being ready to listen to me, to share my success, and to coach me with numerous phone calls and MSN chatting.

I would also like to thank the following people, whose friendship and care have been a blessing: Tao Wang and Pei Huang, Jeff Z. Huang and Tao Xu, Yan Zeng, Xin Kang and Yi Gao, Mei Sun and Liang Yao, Hongwu Zhao and Ning Hua, Xiaomei Zhou and Kun Hao, Sheng Wang, Qian Wu, Ying Qian, Xinzheng Li, Jie Zheng and Jinsong Wu, Tiger Li and Angela Zhu, Hong Xu and Dong Wang, Raymond Zhou and Carol Yu, Nancy, Stephen Monty, and Richard Rusheleau.

From the bottom of my heart, I express my deep sense of gratitude to my family. To my brother Guoqiang Wang, with whom I share a name and precious memories. I love you, and I am proud of you. My mother-in-law, Bichun Huang, for your help with educating my child, and for loving me as your daughter. To my sister-in-law, Xiaosong Li and Dan Lee, and their husbands Wim, and Weiqing Huang. Without your encouragement, support, and love, this thesis would not have been possible.

I am indebted to my parents Yuzhi Chen and Qinghui Wang, for giving me life; for teaching me the value of family and love; for sacrificing their time and efforts to help me in all aspects; for teaching me to be optimistic even in hard times. They have always been by my side, smiled for me, and provided me with endless power to achieve any goal I set. My life has been lightened by their love and guided by their knowledge.

This work could not have been done without the love and support by my precious son, Xintai Li. Your courage and bravery raises me; your patience and understanding inspires me; your love and care warms me; your faith in me helps me achieve goals. Thank you so much for choosing me as your mother, for bringing me joys, happiness, and laughter.

Finally, I am fortunate to have Shi Li, my wonderful husband. We have shared countless happinesses, excitements, surprises, even tears in the past 15 years. Shi has always been providing me support and strength; holding me whenever I encounter frustration, stress, or fatigue; taking care of the family when I was away. I appreciate his encouragement, patience, and love. I believe there is nothing impossible with him backing me up.

STATEMENT OF ORIGINALITY

I hereby certify that this Ph.D thesis is original and that all ideas and inventions attributed to others have been properly referenced.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES AND ILLUSTRATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

AGG	Aggregating
APTEEN	Adaptive Periodic Threshold-sensitive Energy – Efficient sensor Network
BECR	Biased Energy Consumption Rate
BER	Best Effort Relay
BS	Base Station
BT	Backward Channel
CDMA	Code Division Multiple Access
CEN	Candidate Elimination Node
CH	Cluster Head
CMLDA	Clustering – based heuristic for Maximum Lifetime Data Aggregation
CPLOAD	Collaborative Placement with Locally Optimal Allocation Decision
CO	Coordinating
CSIP	Collaborative Signal and Information Processing
CT	Communication Tier
DEFR	Densest Energy – Feasible Region
DS	Data Sink
DVS	Dynamic Voltage Scaling

EAD	Energy – Aware Data – centric routing
EAR	Energy – Aware Routing
EFR	Energy – Feasible Region
EIL	Expected Information Loss
EIU	Expected Information Utility
FC	Forward Channel
FPRN	First Phase Relay Node
GAF	Geographic Adaptive Fidelity
GBR	Gradient – Based Routing
GRS	Global Random Schedule
HC-EAR	Hop-Constrained Energy Aware Routing
IPDA	Independent Placement with Direct Allocation
IPT	Information Processing Tier
LAS	Localized Asynchronous Schedule
LEACH	Low Energy Adaptive Clustering Hierarchy
LHSPRN	Last Hop Second Phase Relay Node
LP	Linear Programming
MAC	Medium Access Control
MCU	Micro-Controller Unit
MCFA	Minimum Cost Forwarding Algorithm
MEMS	Micro-Electro-Mechanical System
MECN	Minimum Energy Communication Network
MINLP	Mixed – Integer Non- Linear Programming

MLDA	Maximum Lifetime Data Aggregation
MLR	Maximum Lifetime Routing
MN	Member Node
MRCF	Maximum – Residual – Capacity – First
MVPF	Max – Valuable – Point – First
NTBF	Nearest – To – BS – First
PEGASIS	Power-Efficient Gathering in Sensor Information Systems
RAND	Random
RF	Radio Frequency
RFD	Reduced Function Device
RN	Relay Node
RR	Routing/Relaying
SAR	Sequential Assignment Routing
SMECN	Small Minimum Energy Communication Network
SN	Sensor Node
SPIN	Sensor Protocols for Information via Negotiation
SPINDS	Smart Pairing and Intelligent Disc Search
SPRN	Second Phase Relay Node
ST	Sensing Tier
S/T	Sensing – and - Transmitting
SWATCH	Step-Wise AdapTive Hierarchical Clustering
TDMA	Time Division Multiple Access
TEEN	Threshold-sensitive Energy-Efficient sensor Network

VFA	Virtual Force Algorithm
WSN	Wireless Sensor Network

1. INTRODUCTION

1.1 Introduction to Wireless Sensor Networks

Rapid progress in micro-electro-mechanical systems (MEMS) and radio frequency (RF) technologies has fostered the development of low-power, inexpensive, and wireless communication capable miniature sensors. These sensor devices are capable of capturing various physical properties such as temperature, pressure, motion of an object etc., converting physical characteristics of the environment to quantitative measurements, and transmitting data packets containing measurements via wireless media. A Wireless Sensor Network (WSN) is composed of a group of such small-size sensors, which execute monitoring tasks in a collaborative and autonomous manner. To provision a WSN application over a large space, sensor nodes (SNs) may deliver sensed information across a number of intermediate relay nodes (RNs) to the base station (BS) in a multi-hop manner. Figure 1.1 illustrates the composition of a typical WSN.

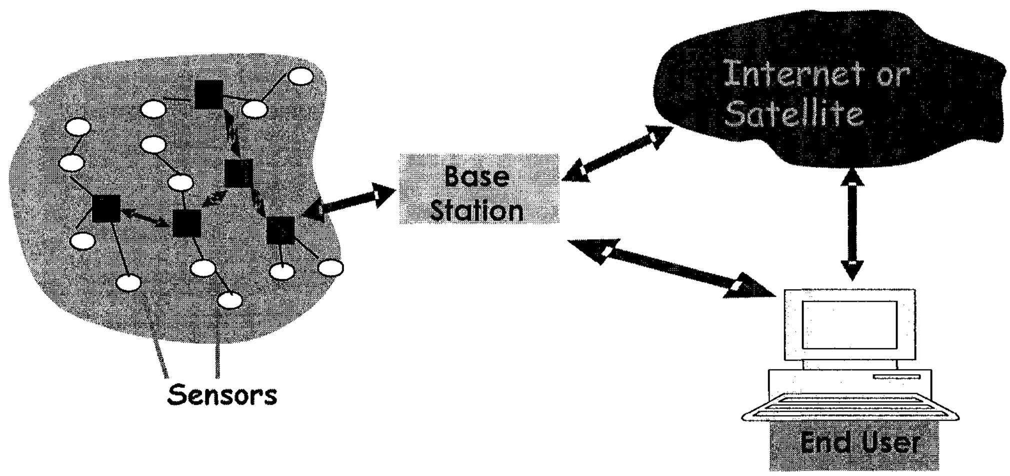


Figure 1.1 An Illustration of a Typical WSN

Widely recognized as a powerful and versatile platform for ubiquitous monitoring in the future, a WSN integrates the functions of sensing, data collection and storage, computation and processing, communication (e.g., over radio medium), and/or actuating.

From the functionality perspective, a typical WSN can be layered with three tiers: the Sensing Tier (ST), which connects the WSN with the physical world; the Communication Tier (CT), which transmits data from one end to the other via effective data collection and information dissemination; and the Information Processing Tier (IPT), which is responsible for processing collected data based on various purposes. The end user can then make appropriate decisions and reactions. Details of the layered structure are shown in Figure 1.2.

Furthermore, based on the direction of information transmission, a WSN consists of a Forward Channel (FC) and a Backward Channel (BC).

The forward channel begins from the physical world, goes through the ST, CT, and IPT, and terminates at the end-user. In the sensing tier, the SNs sense the physical characteristics, and generate data packets after A/D conversion. The communication tier in the forward channel mainly serves to collect data from individual SNs, and deliver them to the information processing tier. Data communication here is usually in a many-to-one fashion. As such, a tree-based routing topology has been adopted in a number of applications, such as in TinyDB [1]. The information processing tier is responsible for processing the collected data based on the end user's requirements. If the WSN is used for real-time environmental monitoring, some schemes for event detection, object identification, and localization may be applied. When the WSN is used for long-term statistics, data will be analyzed and stored; some functions for event prediction may also be executed. Based on the results provided by the information processing tier, the end user will make decisions. The appropriate feedback information will reach the sensing tier through the communication tier in the backward channel.

In the backward channel, the communication tier is responsible for disseminating real-time queries and control messages to the SNs. Typically, data communication in the backward channel is sent in a broadcast or multicast fashion. Based on the received information, the SNs may adjust their duty cycles, synchronize with one another, or conduct necessary actuation to the physical world.

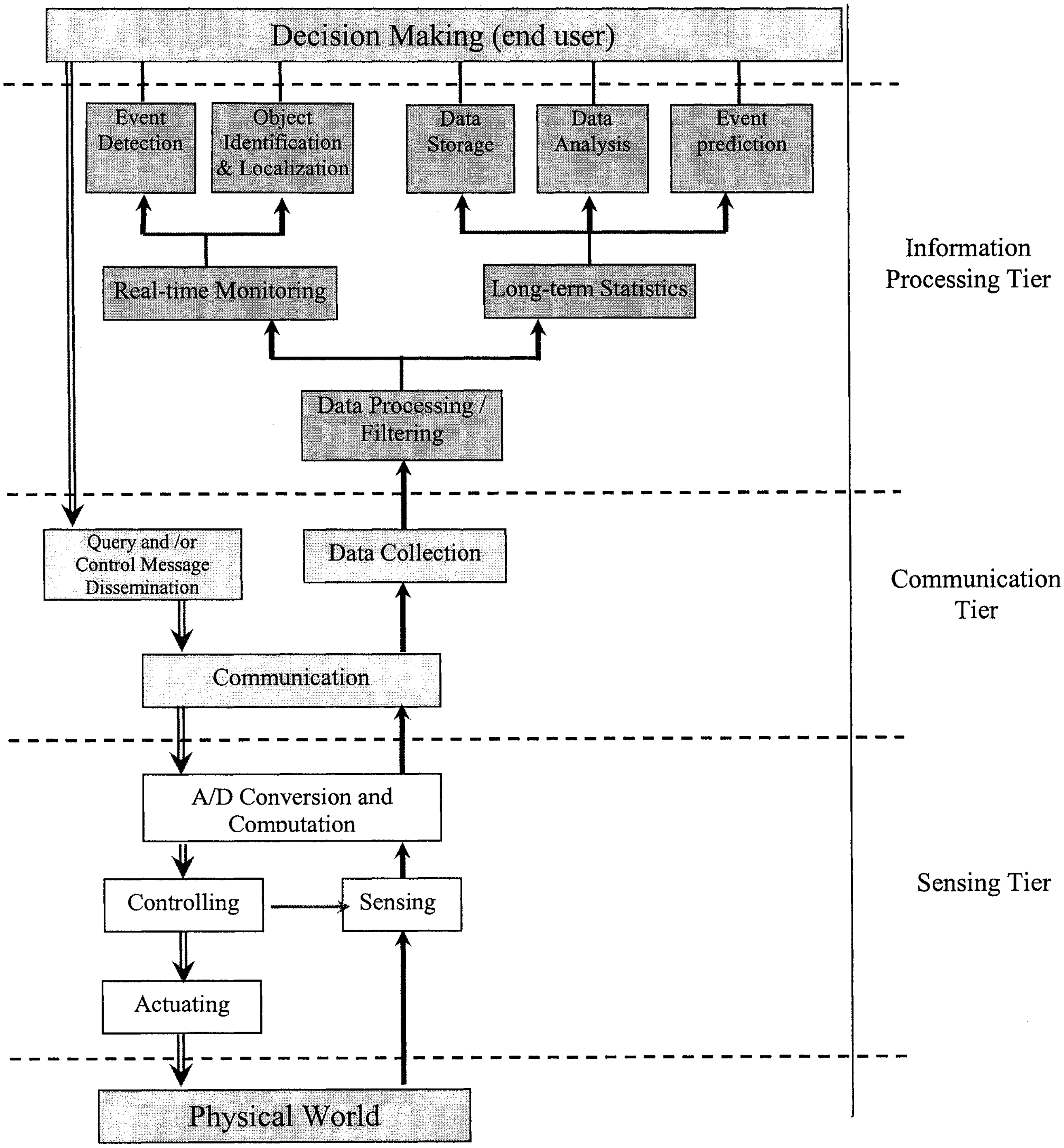


Figure 1.2 Functionality Perspective of a WSN

WSNs provide a revolutionary approach to reliable and ubiquitous environmental monitoring, event detection and reporting, and target localization and tracking. They are studied intensively and used by researchers to keep tabs on weather, crops, animals, mudslides, earthquakes, volcanos, and even glaciers [2]. A WSN is superior to a wired sensing system, such as a data acquisition system or a data logger, for two major reasons. Firstly, a WSN is highly cost effective. As the majority of SNs are powered by battery and communicate via radio, the wiring cost, including engineering labor cost and wiring material cost, is significantly reduced. Furthermore, if manufactured in mass volume, individual wireless SNs are expected to be very cheap. Secondly, WSNs are very flexible and can be deployed in almost any environment, even where conventional wired sensor systems are impossible, unavailable or inaccessible, such as inhospitable terrain, dangerous battlefields, outer space, deep oceans, etc. WSNs can be easily scaled from small to large according to the application settings.

A number of application-specific WSN systems have been developed and used in recent years. For example, in [3, 4], an experimental WSN is developed to detect and localize damage in buildings and bridges by monitoring and collecting the structural response to the ambient and force excitation. In this research, 10 SNs are deployed over a test structure. This system would allow engineers to find deficiencies in a building and prevent a potential disaster from happening in an effective manner. As another example, one experimental WSN is deployed for preventive device maintenance by Intel Corp. and the University of California at Berkeley [5]. In this study, more than one hundred SNs are deployed in a utility supply center to collect the vibration signatures of machinery. At the user side, data are imported into analysis software to determine the mechanical condition of the machinery. The probabilities of failure of machines can then be inferred. A similar deployment is also conducted in a shipboard environment. Another experimental WSN is deployed to collect environmental data, including temperature, relative humidity, and solar radiation, around a 70-meter tall coastal redwood tree [6]. The data provides a detailed picture of the complex spatial variation and temporal dynamics of the microclimate surrounding the tree. This network is composed of tens of sensor nodes.

In summary, the existing and potential applications of WSNs span a wide spectrum in various domains, such as [7, 8]:

- military command, control, communications, computing, intelligence, surveillance, reconnaissance and targeting (C^4ISRT)
- environmental detection and monitoring
- disaster prevention and relief
- medical care
- home intelligence
- industrial automation
- scientific exploration
- interactive surroundings
- smart surveillance
- ...

It is expected that WSNs can be integrated in many aspects of our lives to make them safer, healthier, and more convenient. The only limitation of their applications is our imagination.

1.2 Characteristics and Design Challenges

By gathering certain useful information from the physical world and conveying it to the end user, WSNs bridge the gap between the physical and computational worlds. People can be fully aware of the properties of an interesting object, such as its temperature and moisture dynamics across a vineyard, and interact with the environment. As such, it is essential that WSNs obtain sufficient data and transfer them in a reliable and timely fashion for a required time period. To meet this objective, a collection of unique characteristics have to be taken into consideration in designing the system. The salient features of WSNs and their impact on design issues are highlighted as follows:

a) Miniature devices and limited resources

To monitor the environment in a noninvasive fashion, the majority of wireless SNs are expected to be small in size. The tiny physical size imposes stringent constraints on the capacities of all hardware components onboard and results in an inherent problem of severe resource limitation in several aspects, including computing capability, memory, bandwidth, energy supply, etc. Due to the constraints on computing capability and memory, devices can only run light-weight operating systems and data processing algorithms. In addition, the small memory limits the volume of cached data packets and impairs the data recovery capability. The low bandwidth imposes bounds on the overall traffic rate through the network. To ensure no excess traffic overwhelms the system, the sampling rate, measurement resolution, and the scale of the network have to be confined accordingly. Localized data processing and compression algorithms are useful to eliminate redundancy and decrease traffic volume. In addition, most devices in a WSN operate with a non-replenishable battery. With limited power supply, a device can only operate continuously for a limited time period. Thus, efficiently utilizing the energy and extending the system lifetime are of utmost importance to many commercial applications.

b) Wireless ad-hoc networking and multi-hop routing

Ad-hoc networking is the default choice of WSNs in the majority of current research activities for a number of reasons. An ad-hoc architecture overcomes the difficulties of the predetermined infrastructure settings of other families of wireless networks. It avoids the single point of failure problem and eliminates significant networking cost.

With a low-power transceiver, the distance between two communicating nodes is relatively small. To extend the coverage of a WSN, traffic can be relayed in a multi-hop manner over a larger physical distance. To keep the cost of deployment and maintenance low, WSNs are expected to be rapidly deployed and reconfigured on demand without much engineering effort. New nodes can join the network and existing nodes can withdraw from the system without affecting the functionality of other nodes.

c) Error-prone devices and time-varying links

Individual sensor nodes may fail to function because of exhaustion of power at any time without notification to other nodes in advance. Also, nodes may be randomly damaged by natural forces such as rain, snow, lightning, wind, sun or animals. Moreover, radio communication encounters more errors than its wired counterpart because of background noise, signal attenuation and interference. The possible movement of devices is another cause of unstable wireless communication. In order to deliver data reliably over the unreliable network, usually in multiple hops, data recovery could be conducted hop-by-hop whenever possible or end-to-end when necessary [9]. In addition, fault tolerance can be achieved by the redundant deployment of devices [7], mesh networking [10], and implementing watchdogs on individual nodes [5].

d) Diverse applications

WSNs will provide a generic platform for a wide range of applications in the future. In different contexts, various applications have very diversified requirements, including the network scale, sampling rate, resolution, data volume, reliability, timeliness, robustness, etc. Some applications, such as monitoring and data collection in scientific research incur a large volume of sampled data. As data could be processed offline, the design should focus on communication reliability and energy efficiency. If a WSN is used in a mission-critical application, for example disaster rescue, the design would be more concerned with timely data delivery. As multiple performance aspects may be contradictory to each other, one has to make design choices to meet a performance tradeoff specific to each application.

1.3 Motivations

As discussed previously, energy supply is one intrinsic constraint in WSNs. Typically, energy consumption in a WSN occurs in sensing, computing and data processing, and communication. Tremendous efforts have been invested to achieve optimal utilization of energy supply and prolong system lifetime.

Some research work attempts to reduce all energy consumption sources at one time. For example, topology management algorithms are proposed, in which a subset of devices is strategically selected to be in an active state to provide the required coverage and connectivity, while other devices are put into inactive status [11]. Some research efforts focus on decreasing computing energy, such as using the proposed dynamic voltage scaling (DVS) [12] or system partitioning [13] techniques. While considering the fact that communication energy dominates the total amount of energy consumption in a WSN [14], a number of energy-aware schemes have been proposed. For example, signal and data processing algorithms are used to reduce the total traffic amount by eliminating redundant information reported by nearby devices [14, 15]; energy efficient routing protocols are proposed to either reduce the traffic amount or decrease long transmission distances at the cost of more localized communication among neighboring devices [16]; and low duty cycle scheduling techniques for MAC operation have been developed to avoid any energy wastage due to idle listening, collision, and overhearing [17].

However, if devices are not chosen and provisioned appropriately beforehand, even the best suite of energy-efficient protocols will not be able to provide the required system lifetime. For example, if the number of installed devices is insufficient or there are topological deficiencies due to ineffective device placement, the system lifetime will be degraded or, in the worst case, the system will be non-operational.

In this thesis, we investigate network planning techniques, which is the fundamental issue to ensure a successful WSN implementation. Specifically, we study the impact of device placement on the system lifetime and cost, which are among the greatest concerns of a WSN end user. Generally, the more devices used in a system, the longer the system lifetime will be. However, a user may not be able to afford a network with an arbitrary number of devices. To tackle the trade-offs between cost and lifetime, a WSN design could either aim at maximizing the system lifetime under a certain device cost budget, or aim at minimizing the device cost while ensuring a guaranteed lifetime. In this thesis, we consider the problem of the second form. To the best of our knowledge, little work has been reported in this area previously.

1.4 Networking Planning Techniques for WSNs

Network planning is a fundamental design issue for a WSN application. It refers to the procedure of determining types, numbers, and locations of different wireless devices in order to satisfy various design objectives, such as cost, required lifetime, sensing fidelity, connectivity, and reliability, etc. Figure 1.3 summarizes the general network planning methodology.

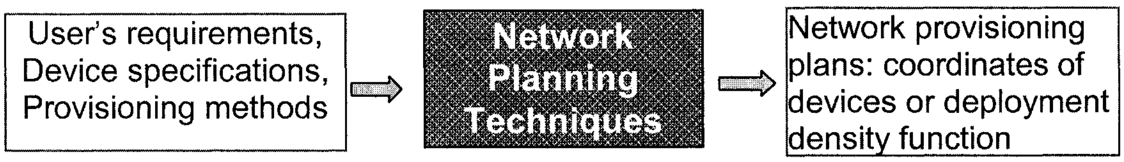


Figure 1.3 A General Networking Planning Methodology in WSNs

In this thesis, we consider the situation in which the sensing field is human accessible, thus devices can be installed at deliberately chosen locations. In such circumstances, a well designed planning strategy can provide optimal or near optimal solutions in terms of either lifetime maximization for a given network cost budget, or cost minimization for a guaranteed lifetime. Typical application scenarios include but are not limited to: industrial automation, agricultural or environmental monitoring, building security and surveillance, traffic monitoring, etc.

In the literature, the research efforts most relevant to our work in this thesis include SN deployment approaches and RN placement schemes. Much of the previous research of SN deployment is based on a binary disc sensing model (to be explained in detail in Chapter 3) and attempts to provide area coverage by deploying a minimum number of SNs [18]. Recently, researchers have started to consider the effects of RN provisioning on WSN lifetime [19, 20].

However, no solution has been presented to provide a comprehensive network planning strategy to determine the number and positions of different types of devices. In this thesis, we propose an integrated WSN planning framework to achieve guaranteed system lifetime under minimum device cost. We also present several device provisioning approaches in various scenarios.

1.5 Overview of Major Results and Contributions

A few schemes have been proposed to tackle the device placement problem with constraints in one or more aspects of lifetime, cost, coverage, connectivity, etc., in some specific scenarios. However, little research has taken all aspects into consideration together. It is the goal of this thesis to generalize, formulate, and solve the network planning problem in a WSN under various circumstances.

The major results and contributions of this thesis are highlighted as follows:

- We identify the generalized network planning problem, and provide a modular network planning framework of WSNs with different types of devices. Logically, we decompose a heterogeneous WSN into a sensing domain and a communication domain. In each domain we formulate the device provisioning problems and propose several schemes and algorithms to solve these problems. A brief structure of our proposed framework is shown in Figure 1.4.

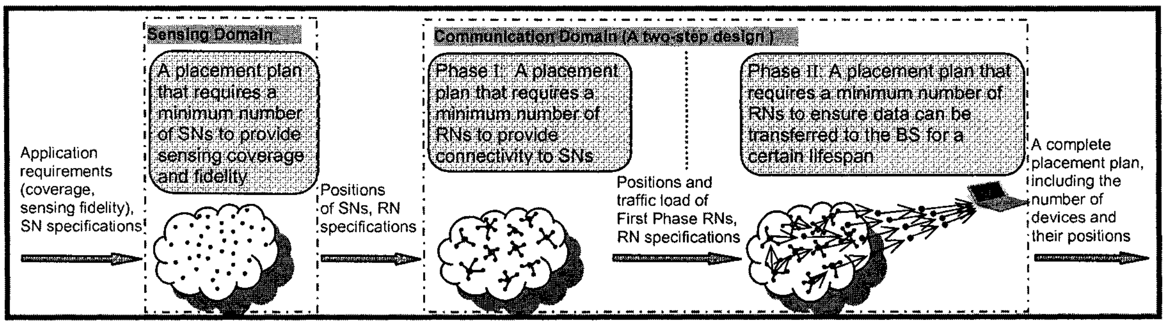


Figure 1.4 Modular Networking Planning Framework for WSN Design

- The sensing domain, formed by sensing devices, is dedicated to providing the required sensing coverage, which is typically application specific. In contrast to the conventional area sensing coverage, by which the whole sensing field is treated equally, we identify the novel concept of information-oriented sensing coverage. Based on the observation that location-related information significance over the sensing field can be differentiated, we propose to deploy a minimum number of SNs to achieve the desired expected information utility. This work paves a new direction for sensor deployment.

- The communication domain, composed of devices with data relay functionality, is dedicated to ensuring accurate delivery of all sensed information to a fixed BS over a desired lifespan. Because of the complexity of the general RN placement problem, we adopt a two-phase design procedure. In the first phase, we aim at placing a minimum number of RNs to guarantee that each SN connects to at least one RN. In the second phase, we aim at placing a minimum number of extra RNs, such that all data collected at the first phase RNs can be forwarded accurately to the BS.

- We develop a set of novel network planning strategies to accommodate various scenarios. These strategies provide a favorable trade-off between cost and lifetime with consideration of other system performance measurements, such as sensing coverage, connectivity, etc.

- Being the first integrated solution in the WSN community, the network planning techniques developed in this thesis will provide an unparalleled competitive advantage to direct users, such as solution providers and system integrators of WSN applications.

1.6 Thesis Organization

The remainder of this thesis is organized as follows:

In Chapter 2, we give an extensive review of state-of-the-art techniques in WSN design from the literature. As a huge amount of research efforts have been conducted to improve energy efficiency and make WSN implementation economically sound, we highlight those that are closely related to our research, including: hardware platform design, provisioning techniques of various devices, topology control and management, multi-hop communication protocols, and signal and data processing.

In Chapter 3, we provide background knowledge for the research in this thesis. Specifically, we present the functional view of a typical heterogeneous WSN; we introduce sensing models, with an emphasis on the signal strength based probability sensing model; and we describe the hierarchical communication model. Moreover, we explain the relationship between energy supply, device lifetime and transmission range. We also identify the generalized network planning problem and its variations in different scenarios. Finally, we present the modular framework, describing its components and the procedure to fulfill a complete network planning strategy.

In Chapter 4, we discuss the sensor deployment issue for information-oriented sensing coverage. Based on the observation of the differentiated importance of information over the sensing field, we describe a sensor deployment methodology to achieve a required expected information utility threshold with a minimum number of SNs. We present a couple of heuristic schemes to solve the problem. Performance of the proposed schemes is evaluated by comparison with random deployment and a simple information-oriented SN deployment strategy.

In Chapter 5, we study the placement of first phase RNs (FPRNs) to ensure connectivity of the SNs. Based on the device properties, we envision two scenarios: either RNs are equipped with ample energy supply or are energy constrained. In the former case, we formulate the corresponding RN placement problem as a minimum set covering problem, and develop a divide-and-conquer based optimal approach. For the latter case, we enhance the minimum set covering model.

In Chapter 6, we tackle the problem of provisioning second phase RNs (SPRNs) to guarantee that each device has a workable path to deliver its traffic to the BS. We identify two components of this problem: to determine the exact position of each SPRN and to determine a traffic allocation scheme such that each RN can select its next hop neighbors as well as the portion of traffic volume to be delivered to the BS through this particular neighbor. Due to the computational complexity, a globally optimal solution is not available for the combined SPRN positioning and traffic allocation problem. We propose several heuristic schemes to solve the problem in two scenarios: either the RN has fixed or adjustable transmission range. We also derive lower bounds on the minimum number of SPRNs in each scenario for comparison in performance analysis.

Finally, in Chapter 7, we conclude the thesis and discuss some directions for future work.

2. LITERATURE REVIEW

With the promise of being a revolutionary innovation for ubiquitous surveillance and monitoring, wireless sensor networks have attracted increasing interest from both academia and industry. As stated in the previous chapter, one of the critical challenges in a successful WSN implementation is limited lifetime, which is imposed by the constrained power supply of non-replenishable batteries on devices. Much of the recent research has attempted to improve energy efficiency and extend the overall system lifetime by devising potent schemes from all possible aspects, such as hardware platform design, network planning schemes, network protocols, network services, software design, etc. Another key issue that affects the implementation of WSNs is cost effectiveness. To provide the best benefit to the end user, a WSN is expected to be economically sound. Figure 2.1 illustrates a brief classification of various approaches. In the remainder of this chapter, we present an overview of state-of-the-art techniques for energy-efficient and/or cost effective WSNs, which are closely related to the research of this thesis.

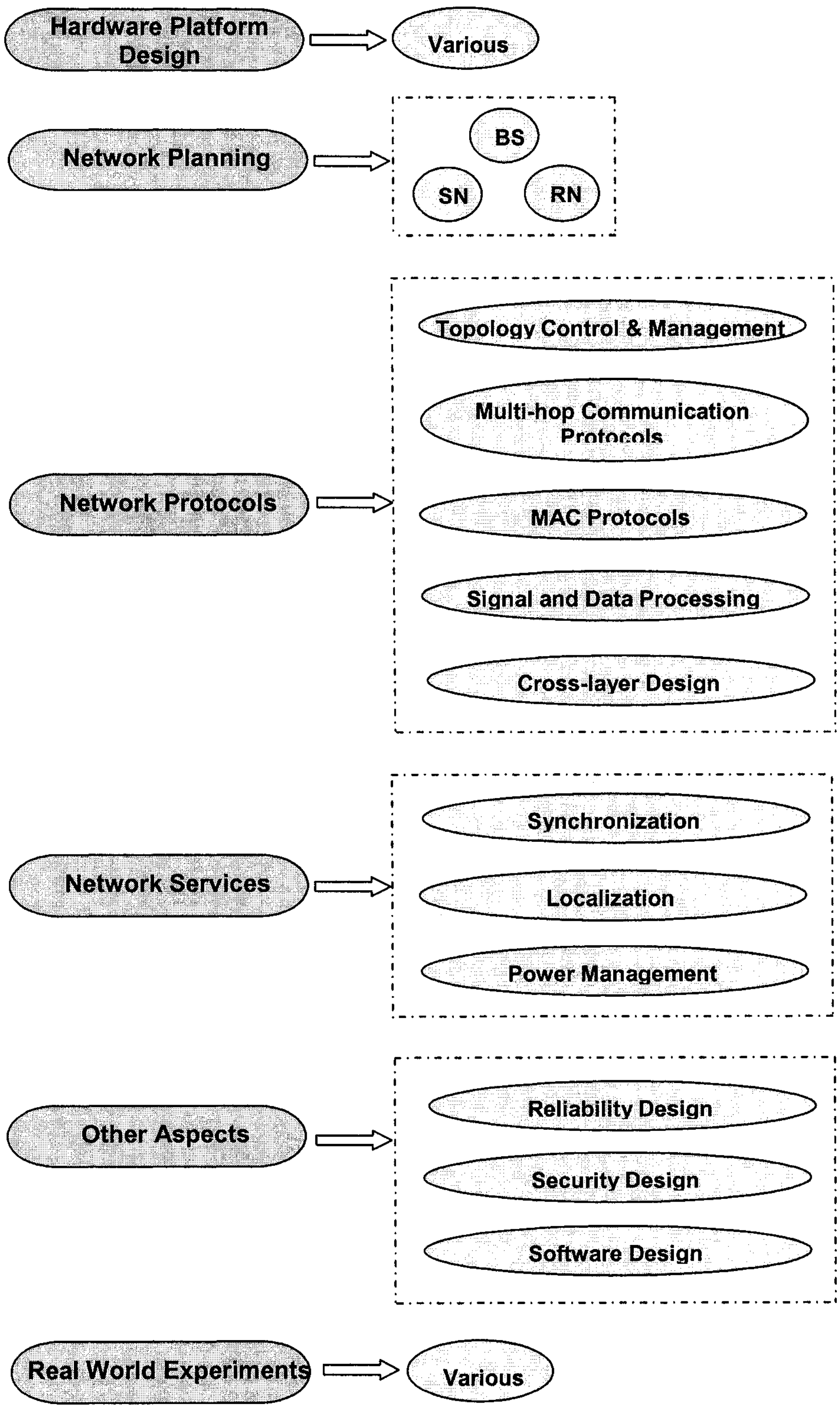


Figure 2.1 Energy-Efficient and Cost Effective Design Directions in WSNs

2.1 Hardware Platform Design

A wireless sensor node is an integration of sensing, signal processing, data collection and storage, and computation, as well as wireless communications along with an attached power supply on a single device. Although there are diverse hardware platforms, each node is composed of four essential components: 1) Power supply unit, which is usually an attached battery. 2) Sensing unit, which consists of the embedded sensor, actuator, and the A/D converter. It links the SN to the physical world. 3) Computing/processing unit, which is a micro-controller unit (MCU) or microprocessor with memory, and it provides intelligence to the SN. 4) Communication unit, which consists of a short range RF circuit and performs data transmission and reception. Some devices are also equipped with clock and positioning elements. The system architecture of a typical micro sensor node is shown in Figure 2.2 [21-23].

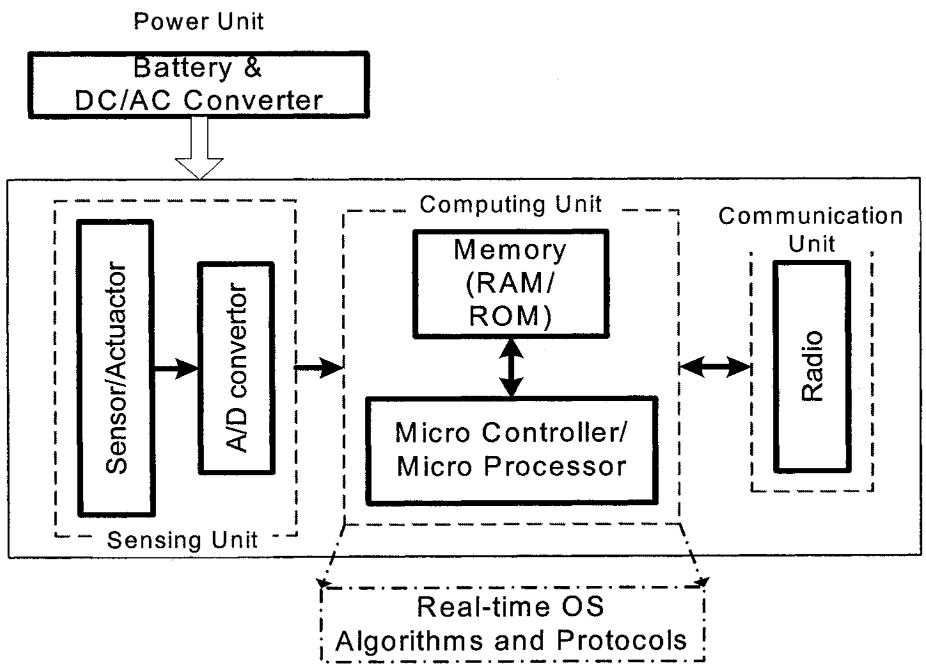


Figure 2.2 System Architecture of a Typical Micro-Sensor Node

To reduce cost, a SN is compact in size, and is equipped with a limited power supply. The tiny size of the device imposes constraints on the capabilities of the micro controller, memory, and radio range. To improve energy efficiency, hardware platforms utilize highly efficient micro-controllers, such as Atmel Atmegal 163, TI MSP430, or Intel PXA271. Most recent platforms use low power, low data rate, small range IEEE802.15.4 radios [24], such as the ChipCon CC2420, or a Bluetooth radio, such as the ZV4002 [25].

Advances in Micro-Electro-Mechanical Systems (MEMS) and the continuous development in wireless communications are spurring more intelligent, less expensive, much smaller SNs. For example, *Piconodes* in the PicoRadio project, is a promising "system-on-chip" implementation to provide ubiquitous distribution of computation and communications for sensor/monitor networks. Each PicoRadio node has a small size, less than 0.10-0.15 cubic inches, consumes less than 10mW, and costs less than $1 [26, 27]. Another system, called *WINS* (Wireless Integrated Network Sensors), integrates multiple functions including sensing, signal processing, decision making and wireless networking capability in a compact, low power device. These intelligent sensors are tiny and powerful in establishing low cost and robust self-organizing networks for continuous sensing, event detection and identification [28, 29]. A project called μAMPS (Micro-Adaptive Multi-domain Power-Aware Sensors) [30] has the ultimate objective of implementing a micro-sensor system on a chip of 1 cubic cm, with the integration of MEMS sensors, A/D, data and protocol processing, and a radio transceiver on a single die. Moreover, the *Smart Dust* project aims to explore the limits on size and power consumption of self-organizing sensor nodes that are not more than a few cubic millimeters in size, i.e., small enough to float in the air, detecting and communicating for hours or days [31, 32].

In recent years, several sensor platforms, which run over the TinyOS operating system [25, 33], have been developed. For instance, a typical TinyOS based MOTE platform developed by Crossbow Technology Inc. [34] is flexible and in the form of modules. The first module is a processor/radio board. The battery-powered motes provide the computation capability and facilitate the wireless communication. A mote can work alone as a RN to deliver data for other nodes. The second module is a sensor/data acquisition board. They can be plugged in the motes to form SNs. The third module is the mote interface board, which works as the interface between motes and PCs, PDAs, etc. As the design of the MOTE is based on an open architecture, it allows researchers to easily redevelop and implement novel communications protocols into the system. The TinyNode platform introduced in [22] is a complete TinyOS supported, compact and low power SN. Its communication range exceeds the existing platform by a factor of 3-5 at a similar power consumption level.

For more information on the TinyOS-based hardware platforms, readers can refer to [25].

2.2 Network Planning

Network planning is a fundamental design issue in a WSN application. It determines the types, numbers, and positions of different devices in order to satisfy various objectives such as coverage, connectivity, cost and lifetime. Previously, people considered homogeneous

WSNs, in which all devices (except the BS) are identical in terms of functionality and capability. Much of the research on network planning has focused on the deployment of SNs to fulfill coverage requirements in various circumstances. Lately, people have advocated WSNs evolved to diverse forms containing a growing number of heterogeneous devices with enriched functions. Researchers have investigated how to place RNs dedicated to deliver data from SNs to the BS. The effects of RN placement on WSN lifetime and cost have been studied. Meanwhile, some efforts have been invested in positioning the base station(s) (BS). In this section, we will provide an overview of recent accomplishments in regard to these aspects.

2.2.1 Base Station Positioning

A BS in a WSN refers to the data sink, i.e., the final destination of data transmission. Considering the distance-related and communication dominant energy consumption property, the energy consumption rate among SNs (and/or RNs) is usually not the same because of the unequal distance from individual devices to the BS and imbalanced traffic loads on individual devices [35, 36]. As a result, some SNs (or RNs) run out of energy much earlier than others. The phenomena of biased energy consumption rate aggravates overall system lifetime. Therefore, some recent research studies how to position single or multiple, stationary or mobile BSs, such that the energy consumption rate among different devices is low and similar, so that the overall system lifetime is prolonged. Figure 2.3 categorizes different types of BS positioning schemes in the literature.

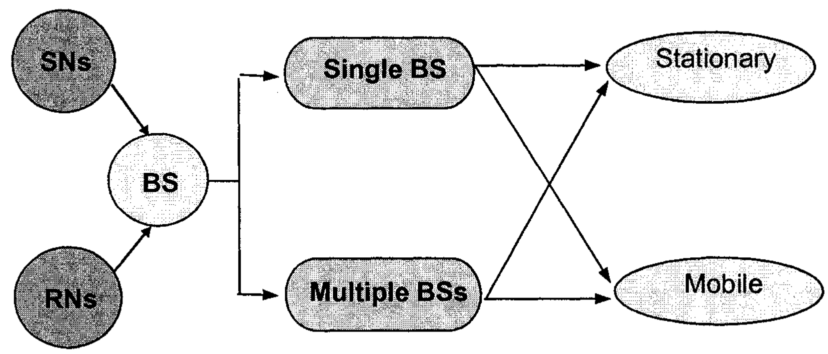


Figure 2.3 Positioning Strategies of Base Station (Gateway, Data Sink, etc)

In [37], in order to improve system performance in terms of lifetime, energy consumption, throughput, and transmission delay, Younis et al. present a solution to relocating the BS when necessary. As the problem of BS repositioning is NP-hard in nature, a method of heuristic searching is proposed. To find the best potential BS location, the product of traffic density and transmission power is used as the metric to decide when and where to relocate the BS. The effectiveness of the proposed approach is validated by simulation.

In [38], Gandham et al. propose to apply multiple and mobile BSs to prolong the system lifetime. The corresponding BS location problem is solved by integer linear programming.

In [39], Pan et al. discuss how to locate multiple BSs optimally in a WSN with a hierarchical network topology. Assuming the positions of the underlying SNs and application nodes (ANs, equivalent to RNs in some other work) are known, the authors present algorithms to determine optimal BS locations in different scenarios to maximize topological lifetime.

In [40], Maulin, et al. propose the optimal placement of SNs, RNs and BSs to ensure coverage, connectivity, and bandwidth in various circumstances.

However, applying multiple or mobile BSs is not feasible in many cases. First of all, using multiple BSs will raise the system cost. A BS has more complicated hardware than a SN or a RN and usually runs on wall power. Thus, its cost is much higher than the other devices. Secondly, it is not always feasible or convenient to move an established BS. Thirdly, according to the above-mentioned existing work, to determine the position of one or multiple BSs usually needs knowledge of accurate positions of all SNs and RNs. However, such knowledge is not available in many cases. Therefore, in a tremendous number of existing research efforts, an assumption is held that one or several BSs are located at fixed and pre-planned positions. In the rest of this thesis, we will focus on the work that assumes a single and stationary BS sitting at a known position.

2.2.2 Sensor Node Deployment

The objective of SN deployment is to realize desirable coverage requirements. In WSNs, coverage has a twofold meaning: range and spatial localization. Range refers to the geographic area of a designated sensing mission, while spatial localization emphasizes the relative spatial positions of SNs and targets, so as to extract accurate measurements. With the great diversity of deployment objectives, application scenarios, and SN properties, SN deployment strategies and mechanisms vary significantly from case to case, and numerous sensor deployment strategies have been proposed addressing various aspects of the sensing coverage problem.

According to the coverage mission, the majority of the existing research efforts assume SNs are reliable, and the related SN deployment strategies can be grouped into two categories: to achieve uniform coverage, or to achieve differentiated coverage. Some work deals with the situation with reliability consideration. Figure 2.4 shows a classification of SN deployment strategies.

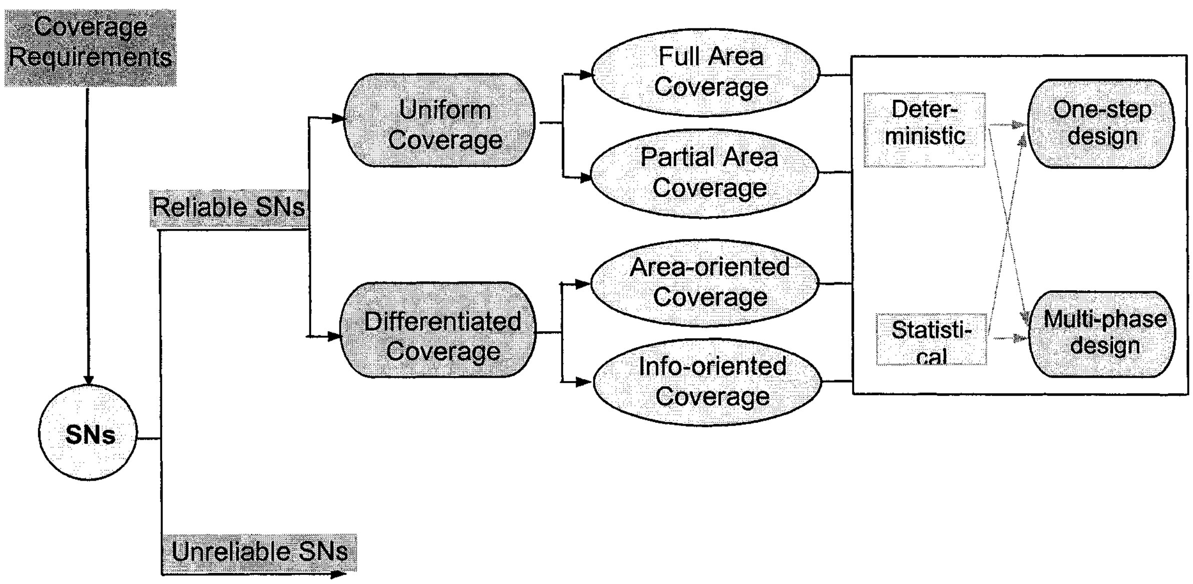


Figure 2.4 Classification of Sensor Node Placement Strategies

2.2.2.1 SN Deployment for Uniform Coverage

With uniform coverage, the entire sensing field is treated equally.

The first type of work in this category aims at achieving full area coverage at a minimum device cost. That is, any point in the sensing field is watched by at least one SN. Sensor deployment can be conducted deterministically or statistically, and by using stationary SNs or movement-capable SNs.

The deterministic strategy applies to two situations: (1) there is sufficient knowledge about the environment or the possible targets, as described in [41]; (2) sensor nodes can be placed in some pattern at deliberately chosen spots. For example, the sensing site is spatially modeled as a grid-based distribution of points, i.e., the two or three-dimensional space is represented by point coordinates. The granularity of the grid (the distance between adjacent grid points) is determined by the desired accuracy [42].

In [43, 44], several examples are illustrated to place SNs in some planned geometric topologies for medical care purposes. In [45], the sensing field is represented by a grid of points. The problem of achieving complete coverage at a minimum sensor cost is solved by integer linear programming. The proposed coverage schemes can be used for effective surveillance and target location. In [46], three grid-based deployment topologies, i.e., triangular, square, and hexagonal grids are discussed and compared. In [47], energy efficient organization schemes are proposed. Taking sensing coverage as the only constraint, sensors are partitioned into a maximum number of disjoint dominating sets, each of which can fully cover the sensing field and work alternately. In [48], the concept of a connected sensor cover is introduced. Both centralized and distributed self-organizing algorithms are proposed to tackle the problem of connectivity and full coverage for spatial queries. In [49], the connected and full area coverage problem is generalized into the connected k-coverage problem, i.e., any point in the sensor field should be covered by at least k sensors at any time. In [50], Bai et al. consider the full coverage and k-connected (k>1) sensor deployment problem. An optimal deployment pattern to achieve both full coverage and 2-connectivity is proposed, and its optimality is proven as well. The above two research efforts have great importance for fault tolerance and reliability.

In general, a deterministic strategy can provide an optimal solution for the desirable coverage and obtain high QoS and cost-efficiency at the same time. However, knowledge of the environment and targets may not be available *a priori*. Moreover, computational complexity makes the schemes not scalable to large networks.

Therefore, a random strategy is more realistic for large-scale WSN application or applications where the terrain is not readily accessible, such as battlefields or remote terrain [51-53]. However, coverage holes may occur due to the unevenness and uncertainty in sensor placements inherent in random deployment [54, 55]. In order to repair the coverage holes, movement-capable SNs are adopted in some research efforts. In [56], three protocols, called VECtor-based (VEC), VORonoi-based (VOR), and Minimax are proposed to move SNs from densely deployed regions to sparsely deployed areas after they have been randomly dropped on the sensing field. In [57], a heterogeneous mix of SNs is explored. The high-level nodes are deployed deterministically, while the low-level nodes are deployed randomly.

More work on using movement-capable sensors to achieve desired coverage can be found in [54]. Networks with movement-capable nodes are more flexible in achieving optimal coverage. However, mobility is a costly function. Compared with its stationary counterpart, a movement-capable SN is more complicated and consumes more energy, thus it is more expensive. Therefore, in this thesis, we will focus on deployment of stationary sensors.

The second type of research effort for uniform coverage aims at providing partial sensing coverage under a certain device cost. Although it is ideal to provide full area coverage, people argue that it is not always necessary, feasible, or reliable in practice. Therefore, in recent years, several papers advocate that partial coverage is sufficient, in contrast to full area coverage, in some wireless sensor applications. In [58], partial coverage properties are discussed to prolong the system lifetime. A heuristic algorithm is proposed to provide a given coverage guarantee while maintaining connectivity. A bound on the distance from an uncovered point or area to its closest sensor is derived as well. In [59], the partial coverage problem is studied from another angle: scheduling the duty cycle of sensors, such that alternative groups of active sensors form sensing coverage dominant sets, and any point in a sensing field is sensed with a certain probability at any time. Three algorithms, called Global Random Schedule (GRS), Localized Asynchronous Schedule (LAS) and Power-Aware Asynchronous Schedule (PAAS) are proposed in [59]. In [60], Howard et al. propose a potential-field based method to deploy sensor nodes automatically in an unknown environment. Since the sensing field is established in such a manner that each SN is repelled by both obstacles and other nodes, the entire network is self-spread throughout the environment and can reach the maximum coverage. In [61], based on an initial random distribution, a practical and distributed Virtual Force Algorithm (VFA) is introduced to reposition the sensors in order to enlarge the coverage to the desired optimal results. Each node autonomously decides its new position according to the attractive and repelling forces imposed by its neighbors. Simulation results showed that VFA can enlarge coverage significantly to a desired optimal result compared with the original random deployment.

2.2.2.2 SN Deployment for Differentiated Coverage

The deployment of SNs for uniform coverage cannot always satisfy the design requirements. For example, in some cases, the sensing targets may be distributed unevenly over the sensing area, or the value of information of individual targets may not be equal. In such cases, sensing coverage is differentiated. In [27], Willig et al. illustrate an example of biased placement of sensors in a large-scale office, in which the density of SNs close to windows is much higher than that in the middle of the room. In [53], Tilak et al. present some comparisons of different deployment strategies by means of simulations. In [62], Clouqueur et al. describe a scheme to sequentially deploy sensor nodes in steps by introducing path exposure as a metric of goodness. With the strategy of properly choosing the number of sensors in each step, the cost of deployment can be minimized to achieve the desired detection performance.

All of the above research efforts focus on area coverage, which measures the sensing quality in terms of the size of covered or uncovered areas. In contrast, the concept of information-oriented sensing coverage is introduced in [63, 64] , which is concerned with the actual benefits as a result of information obtained by sensors. The difference becomes very important when full coverage is not achievable with a limited number of SNs. As information-oriented sensing coverage is a new concept, many open problems remain to be solved. In Chapter 4, we describe the idea of information-oriented sensing coverage in detail and present a couple of schemes to solve the corresponding SN deployment problem.

2.2.2.3 Deployment of SNs with Reliability Considerations

Most of the research discussed above has an implicit assumption: every sensor node operates in a reliable manner. However, this is not always true in reality. Some proposals have been introduced to handle unreliable conditions. Considering the uncertainty of sensor detection, a statistical optimization framework is presented in [42]. Assuming a given set of detection probabilities in a sensor field, it optimizes the number of sensors and determines their positions so as to achieve sufficient grid coverage. In [65], Guibas et al. discuss the coverage and connectivity for WSNs with unreliable SNs. They derive necessary and sufficient conditions to cover a unit square region by a random grid network and maintain connectivity. They also formulate sufficient conditions for connectivity between active nodes. The framework described in [66] allows the sensor coverage areas to overlap such that each resolvable position is covered by a unique set of sensors. Using novel identification codes and based on a polynomial-time algorithm, it not only requires fewer sensors than existing proximity-based schemes to achieve the demanded coverage, but it is also robust against sensor failure or physical damage to the system. An alternate approach to achieve desirable and reliable coverage is by means of hardware redundancy, i.e., to deploy a greater density of SNs in a sensing region and exploit redundancy to extend the overall system lifetime by operating distinct subsets in turn based on local density and local demand [67]. This is effective when the cost of deploying a node during the initial placement is much smaller than the cost of adding a new node at a later time.

2.2.3 Relay Node Placement

As discussed in the previous subsection, much research on sensor deployment has focused on fulfilling coverage requirements in various circumstances. However, to make a network fully functional, the requirements on connectivity should also be guaranteed. Recently, researchers have started to consider the effects of RN provisioning on WSN lifetime. According to application circumstances, we classify the approaches into two categories: random deployment and deterministic placement.

2.2.3.1 Random Deployment of Relay Nodes

Random deployment is feasible in application scenarios such as large scale environmental monitoring, intrusion detection in battlefields, or relief and rescue from natural disasters. In such situations, either the sensing field is hostile or inaccessible to humans, or the sensing field is so large that a great number of devices are deployed. Hence, a random deployment of RNs is controlled by the deployment (probability) density function. The most popular density function in this regard is the uniform density function, by which the devices are spread randomly over the sensing field in an even manner. In [68], with the objective of minimizing the cost, optimal nodal densities (in terms of the ratio between sensor nodes and relay nodes) and nodal initial energy are derived to guarantee a lifetime threshold in a randomly deployed WSN.

While the random uniform placement of SNs provides even coverage of a sensing field, the random uniform placement of RNs is not energy efficient as shown in [35, 36]. In [35, 36], Xu et al. explain the biased energy consumption rate (BECR) phenomenon. Assuming a WSN composed of random, uniformly deployed SNs and RNs, and traffic being generated at the SNs uniformly across the network, the energy of RNs at different locations will be consumed at different rates in both the single-hop and multi-hop communication cases. In a single-hop transmission scenario, where a RN adapts its transmission energy and transmits data to the BS in one hop, the RNs which are farther away from the BS will drain energy faster than the RNs closer to the BS due to the longer transmission distance. As such, those RNs farther away from the BS will become unusable while a considerable amount of energy is still left in those closer to the BS. This phenomenon becomes worse when the network scale is large.

A weighted random deployment of RNs is proposed to overcome the BECR problem and maximize the usability of the network in [35, 36]. The work in [36] addresses the BECR phenomenon in the single-hop communication case, and the work in [35] deals with the BECR problem in the multiple-hop communication case. Given a random uniform deployment of N_{SN} SNs and the traffic pattern of each SN, these works determine the deployment density function for N_{RN} RNs, so that the lifetime of the WSN is maximized.

The essential idea of the solutions in [35, 36] is that the number of RNs deployed at a position should be proportional to the expected energy consumption rate at that location. To achieve this in a statistical manner, density functions which are proportional to the expected energy consumption rate at different locations are derived. In addition, both efforts consider the impacts of random deployment on the network connectivity.

In [69], another differentiated random RN deployment method is discussed based on the uniform random distribution of SNs in a two-dimensional field. The authors consider an application in a hostile environment where devices may fail due to physical attacks. Given a random uniform deployment of a known number of SNs, and a certain number of RNs as well as their initial energy supply, an upper bound on system lifetime is derived. Moreover, with a defined random physical attack model, a RN placement strategy is proposed, which expresses the density of RNs as a function of their distance to the BS. Such a weighted random RN placement approach aims at maximizing the network lifetime under a throughput requirement.

2.2.3.2 Deterministic Placement of Relay Nodes

Random deployment may not provide an optimal solution, and it may be difficult to realize the desired density function in many cases. Deterministic placement is applicable in circumstances where the number of devices involved is small to medium, and the sensing field is human accessible. As such, the devices can be installed at deliberately chosen locations. Typical application scenarios of deterministic placement include industrial automation, agricultural monitoring, building security and surveillance, traffic monitoring, etc. A well designed deterministic placement can provide optimal or near optimal solutions in terms of lifetime maximization with given resources or cost minimization for a guaranteed lifetime.

In [19], Hou et al. formulate a combined RN placement and energy provisioning problem as follows. Given the placement, traffic load and initial energy of some RNs (which guarantee the connectivity of all SNs), Hou et al. allocate an additional and limited amount of energy E at M locations (which can be either at existing FPRNs or newly added SPRNs) such that the system lifetime is maximized. Implicitly assuming RNs are able to adapt their transmission energy, in [19], Hou et al. first model this problem as a mixed-integer non-linear programming (MINLP) problem, which is known to be NP-hard. For such a MINLP problem, a heuristic algorithm called Smart Pairing and Intelligent Disc Search (SPINDS) is proposed, which can produce a satisfactory solution in polynomial time.

The above work makes an explicit assumption: the connectivity of SNs has been satisfied. In this dissertatoin, we consider the RN placement problem from another angle: given the placement of all SNs, determine the number and positions of RNs so that the WSN functions for a guaranteed lifetime at a minimum cost.

Due to different design constraints, the generalized design problem can be refined into different design scenarios. One of the constraints is the energy supply of a RN, which can be either unlimited (for practical purposes) or limited. When a RN has an unconstrained energy supply (e.g., rechargeable or simply ample enough relative to the projected lifetime of the SNs), the placement of RNs must provide connectivity to each SN with the constraint of the limited communication range of the SNs. When the energy supply of RNs is limited, the placement of RNs should not only guarantee the connectivity of SNs but also ensure that the paths from RNs to the BS are established without violating the energy limitation. Another factor is power control. RNs either transmit with a fixed transmission power, so that they can only reach next hop relays within a fixed transmission range, or they can adjust their transmission power, giving them a variable transmission range. In both cases, the traffic rates need to be explicitly constrained for every RN. In the remainder of this thesis, we tackle the RN placement problem in different scenarios.

2.3 Network Protocols

The network layer in WSNs is responsible for data delivery from source to destination via well-selected routes [70]. The ultimate goals of network protocols are: reliable information dissemination, collaborative data collection, and directed routing. Due to the unique characteristics of WSNs, energy efficiency is always a dominant consideration and protocol design is most likely application-specific.

In this section, we summarize various network protocols in several categories: topology control and architecture management, multi-hop routing protocols, and signal and data processing.

2.3.1 Topology Control and Architecture Management

Communication among SNs has been recognized as the major source of energy dissipation in WSNs [14, 21]. Experiments show that the energy ratio of communicating 1 bit over the wireless medium to that of processing the same bit could be in the range of 1,000-10,000 [14]. Topology control and architecture management are effective techniques to conserve communication energy.

2.3.1.1 Topology Control

A dense deployment of sensors ensures the required coverage and sufficient precision of detection. Meanwhile, the redundant data generated by densely deployed nodes can be treated as backups of one another, so as to ensure the reliable function of the network. In the process of system operation, some nodes may operate in low duty cycles by transitioning the hardware to sleep or off states to conserve energy. Ideally, for the purposes of energy conservation, a node is expected to be in inactive states as long as possible. However, in these states, the devices are unable to communicate and forward packets. The nodes would then be woken up in certain situations, such as when neighboring nodes becoming exhausted. Therefore, the active topology of the network changes over time. This leads to the critical issue of how to arrange sleep state transitions while ensuring robust, un-degraded information collection [21].

A typical approach is to periodically rotate the node functionality to achieve balanced energy consumption among nodes. The protocol SPAN, proposed in [71], is one example of the approach for wireless ad hoc networks. A limited number of random nodes are self-selected as coordinators to construct the backbone in a peer-to-peer fashion within the network for traffic forwarding, while others can make local decisions to transition to a sleep state or keep active. The Geographical Adaptive Fidelity (GAF) algorithm proposed in [72] is another way to rotate the active nodes within the network. By identifying equivalent nodes based on geographic locations on a virtual grid, they can substitute for each other directly and transparently without affecting the routing topology. Considering the fact that a WSN is only sensing its environment or waiting for interesting events to happen, a new technique – Sparse Topology and Energy Management (STEM), described in [73], claims to improve upon SPAN and GAF in terms of obtaining higher energy savings so as to prolong the system lifetime by trading off an increased latency to establish a multi-hop path.

2.3.1.2 Clustering and Hierarchical Architectures

Typically, energy conservation for communication can be achieved by a) compressing the traffic volume e.g. by processing data locally [74, 75], b) reducing the communication power, e.g., through a multi-hop data relay [68, 76], and c) decreasing wasteful energy consumption through, for example, reducing the energy expenditure due to idle sensing of the wireless channel, retransmission due to collision, overhearing, and/or overhead for exchanging control packets [76, 77].

Clustering is an effective energy saving framework, under which the above techniques can be adopted. Moreover, on top of a clustering architecture, different schemes can be employed at different layers, for example, MAC protocols, routing protocols, resource allocation schemes, duty cycle assignment and scheduling schemes, data aggregation and signal processing algorithms, etc.

Numerous clustering approaches for wireless ad hoc and/or sensor networks have been proposed in the literature. The essential problems faced by a clustering scheme are: a) how many clusters are to be formed, b) how to select cluster heads (CHs), and c) how to associate member nodes (MNs) to their corresponding CH. Since WSNs represent a variety of applications in which environment and technical requirements may differ greatly, the design of a clustering scheme is usually application oriented, and may vary case by case. Various clustering schemes can generally be classified into different categories based on several important criteria.

Depending on whether a central controller participates in the clustering formation, a clustering scheme could be either centralized or distributed. In a centralized scheme, such as LEACH-C [51, 78] and APTEEN [79], the central controller usually has powerful computational capability, is not resource constrained, and has global information of the system, including the position and residual energy of each individual node. Hence, it may produce the optimal number of clusters which are well spread in the network. However, a WSN often contains a very large number of nodes with limited computational and communication capabilities, so a centralized scheme is not always feasible, scalable, or robust. On the other hand, distributed clustering schemes only need local information exchange among the neighboring nodes [76, 80, 81], or are fully independent [51, 76, 82, 83]. Therefore, distributed clustering schemes are usually more attractive and practical for WSNs.

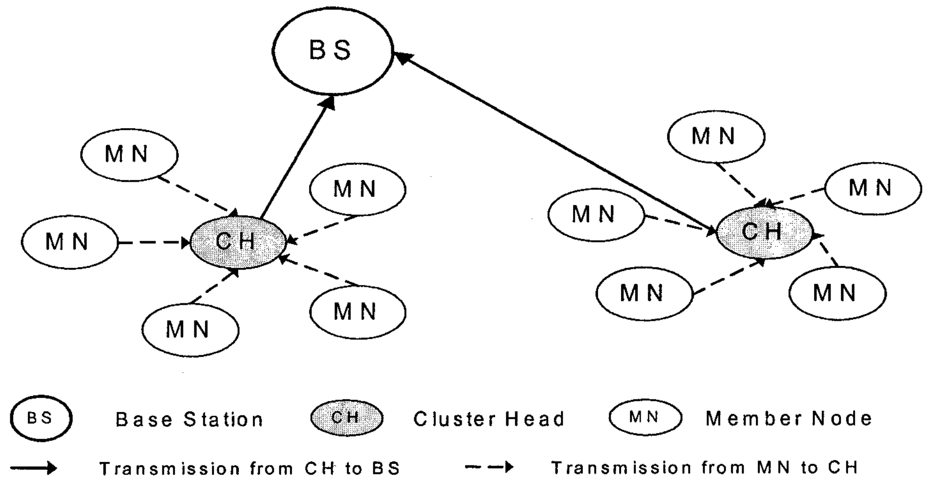
According to the types of nodes in a system, a clustering scheme can be classified as being for homogeneous or heterogeneous networks. Many of the existing clustering schemes make the assumption of a homogeneous network (only one type of node). Due to the fact that a CH has much heavier work load than its MNs and dissipates energy more quickly than the MNs, fairness is a critical issue in such systems. In [68, 83], Mhatre et al. propose a clustering scheme based on heterogeneous networks in which two types of nodes are deployed. The simpler nodes with less energy act as MNs and report their sensing data to their CHs. Nodes with higher intelligence and more initial energy will take the responsibility as CHs. The challenges in such schemes include the calculation of the ratio among different types of nodes and the initial energy level.

Based on the roles of devices in the system, a clustering scheme can be categorized as a static or dynamic one. In a static scheme, the CHs and their corresponding MNs will not change. However, in a static clustering scheme, it is difficult to realize fairness among the nodes. On one side, there could be an uneven number of MNs in different clusters, which results in the CHs with more MNs being overburdened. To solve such a problem, some load balancing schemes have been proposed [80, 84] with the objective that each CH serves the same number of MNs. For example, in [84], Chiasserini et al. attempt to solve the problem of optimal balanced k-clustering, where k denotes the specific number of CHs in a system.

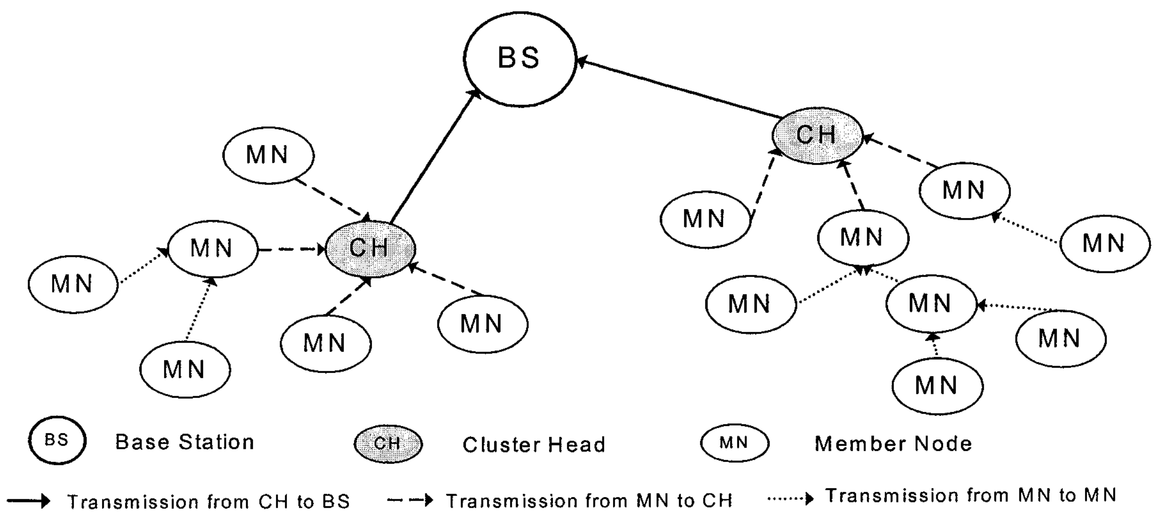
Based on minimum weight matching, the algorithm attempts to realize load balancing among different clusters by partitioning the nodes into groups such that each cluster has a similar number of nodes. It achieves minimum energy consumption by optimizing the total spatial distance between the MNs and the CHs. However, due to the non-optimized distances between a CH and its MNs, this type of scheme fails to minimize energy consumption. On the other side, the MNs are unable to dissipate their energy at the same pace due to unequal distance to their CHs. For example, by using single hop communication, the MNs which are far from the CHs will dissipate their energy more quickly than those closer to the CHs, while, when using multi-hop communication, the MNs closer to CH are overused to relay data for those that are far from the CH. To overcome this problem, in [68, 83], Mhatre et al. develop a scheme that has the MNs alternatively switch between single hop communication and multi-hop communication, but this requires a more complicated control scheme, which is energy consuming. Using a dynamic mechanism [51, 76, 82, 85] is a possible solution to provide better fairness while retaining simplicity. In these schemes, clusters are periodically regenerated, and nodes take responsibility as CHs in rotation. However, the regeneration of clusters is energy consuming. Therefore, it is better to have a hybrid scheme combining the advantages of both static and dynamic schemes.

Based on the CH selection criteria, a clustering scheme can be either weight-associated or weight-independent (or randomized). In the weight-associated scheme, the CHs may be selected based on node ID [86, 87], nodal degree [88], residual energy [79-81, 89], or a combination of several parameters [76, 90, 91]. In [91], Aoun et al. present a polynomial time algorithm by recursively computing minimum dominating sets to determine the position of a minimum number of gateways (CHs), while satisfying upper bounds on the maximum latency, the energy consumption on intermediate nodes and cluster sizes. On the other hand, CHs can be selected randomly in a weight-independent fashion [51, 82, 83, 92]. In general, a weight-associated scheme can produce more optimal clusters, but it has higher complexity for local/global information exchange and coordination, which sacrifices more energy and results in longer latency. Also, such a scheme is sensitive to topology changes in the system, and nodes are expected to be more intelligent. In contrast, the randomized scheme is not only simpler to implement, but also more robust to topology changes in the network. Such a scheme is more attractive in large scale WSNs with tiny, compact sensor nodes.

Depending on the distance from MNs to CHs, a clustering scheme can be either single hop [51] or multi-hop [82, 91-93], as shown in Figures 2.5 (A) and (B) respectively [94]. According to the hierarchical structure, a clustering scheme either has a single tier [51] or multiple tiers [79, 82, 93], as illustrated in Figure 2.6. Based on the CH selection procedure, the cluster can be formed in a single step [51, 82] or iteratively [76, 85, 95].



(A) Single-Hop Clustering Architecture



(B) Multi-Hop Clustering Architecture

Figure 2.5 Illustration of Single hop and Multi-hop Clustering Architecture

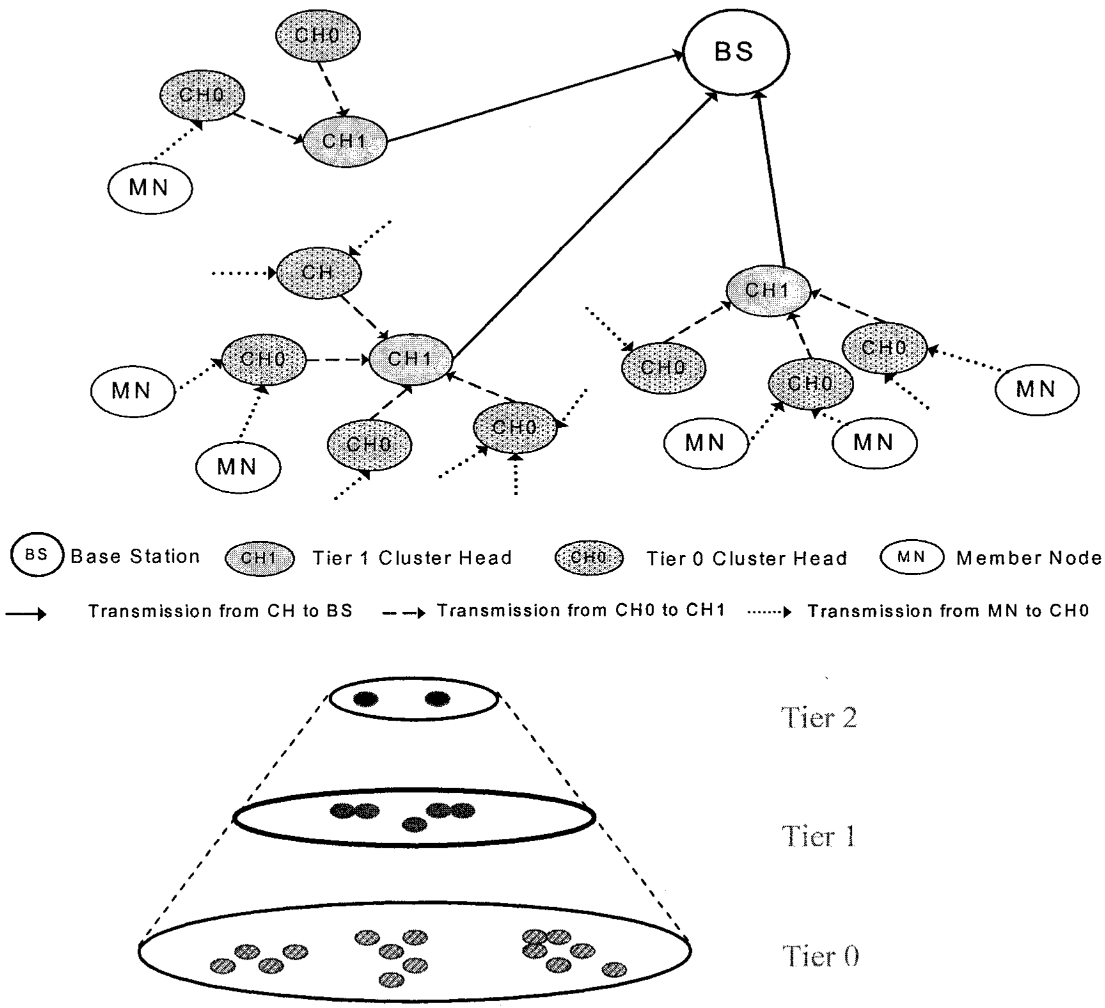


Figure 2.6 Multi-level Clustering Hierarchy Architecture

In a large-scale network, multi-hop clusters and multiple-level clustering hierarchies may be more appropriate to further decrease the communication distance. In [92], the Max-Min *d*-cluster scheme is proposed to generate *d*-hop clusters, which can achieve better load balancing among the clusters with fewer clusters than the single-hop clustering schemes [86, 87]. Max-Min *d*-cluster has no requirement of synchronization among the nodes.

Keeping the features of randomized creation and rotation of CHs as proposed in LEACH [51], as well as the advantages of a multi-hop clustering algorithm, in [82], Bandypadhyay et al. introduce a new energy efficient single level multi-hop clustering algorithm. The authors also provide the formulation for finding optimal parameter values to minimize the energy consumption. Moreover, based on the results of [96, 97] (also in [82]), Bandyopadhyay et al. present a novel energy-efficient hierarchical clustering algorithm with a total of h levels, (i.e., some of the cluster heads in level k-1 select themselves as k^{th} level cluster heads, and the remaining level k-1 cluster heads are cluster members in level k). They derive optimal parameters to achieve minimum energy consumption within the whole system. Experimental results for up to 10,000 nodes have been reported.

Due to the simplicity offered by the single step distributed clustering scheme of the LEACH scheme ([51]), in [98], Wang et al. developed a multi-dimensional Markov Chain model to evaluate the effects of the uncertainty associated with the randomized CH creation. Based on this theoretical analysis as well as the results on the optimal number of CHs derived in [51, 82], in [85], Wang et al. developed a step-wise adaptive clustering hierarchical protocol (SWATCH), in which clusters are generated in multiple steps. Both mathematical analysis and simulations prove that SWATCH outperforms other randomized distributed clustering scheme in terms of higher certainty in cluster formation.

The categories described here may overlap, i.e. a specific scheme may have the properties of different domains. For instance, the cluster formation scheme of LEACH [51] is a fully distributed, dynamic, randomized (weight independent), single hop, single tier, single step clustering scheme used for homogeneous WSNs.

2.3.2 Multi-hop Communication Protocols

Providing efficient routing support in WSNs is very challenging due to several characteristics of WSNs that distinguish them from their counterparts in wired networks and in mobile ad-hoc networks.

- *Data-centric routing*: Assigning a globally unique ID for each and every sensor is neither necessary nor practical for a WSN consisting of a large number of SNs. Therefore, conventional address-centric routing protocols are not applicable to WSNs. A data-centric concept has been proposed for WSNs, in which a datum itself is named according to attributes such as event-type or geographic location of the SN [99]. The examples depicted in Figure 2.7 (A) and (B) demonstrate the difference between these two approaches. In address-centric routing, the intermediate node, M, has to forward all the packets received from different source nodes (e.g. S1, S2) to the destination D. However, in a data-centric routing model, node M first fuses the data from S1 and S2 by eliminating the redundant information, then it relays the processed data to D. This leads to higher efficiency and more energy savings.

- *Data redundancy*: Since multiple SNs may sense and therefore report similar data within the vicinity of a phenomenon, great redundancy may occur and needs to be reduced by communication protocols to improve energy efficiency and bandwidth utilization.

- *Resource-constrained*: SNs are tightly constrained in terms of battery energy, processing capacity and storage and thus require careful resource management.

- *Low data rate*: The data generating rate at a sensor is in general very low.

- *Non-mobility*: sensor nodes are usually non-mobile.

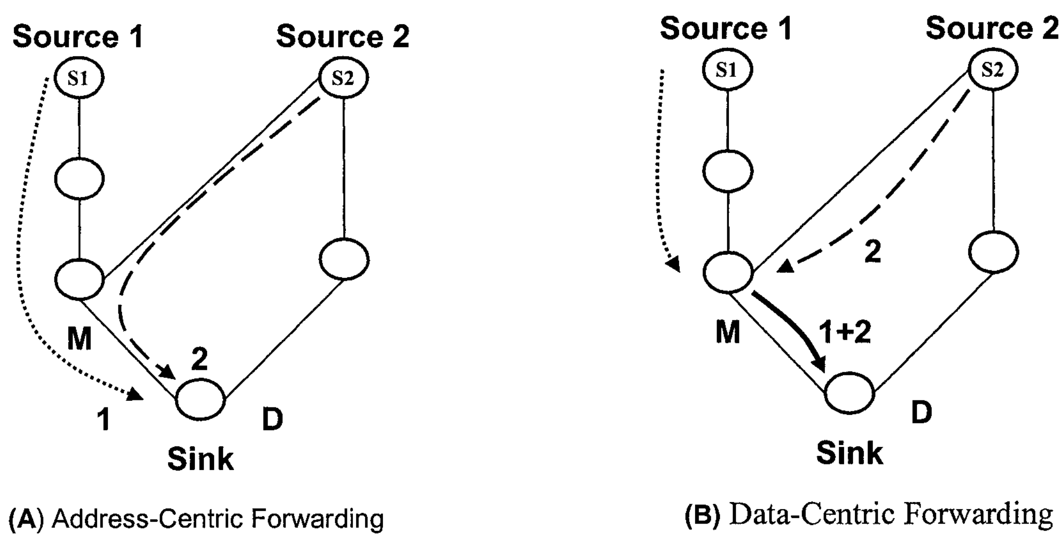


Figure 2.7 The Address-Centric Forwarding vs. the Data-Centric Forwarding

Tremendous research efforts have been invested in energy efficient data communication. Various routing protocols can be classified into three groups: a) collecting data from SNs to the BS collaboratively; b) disseminating information from the BS to SNs reliably; c) retrieving query-based information in a timely manner and accurately.

2.3.2.1. Energy-Efficient Information Collection Protocols

Information gathering is one essential task for all WSNs. One or multiple data-collecting and processing centers (BSs) are usually assumed to be placed at fixed locations, thus data are collected in a unidirectional manner. From the perspective of the system architecture, information collection protocols can be flat, hierarchical, or localized (geographical-based).

Flat Information Collection Protocols

The objective of an energy-efficient flat communication scheme is to find an optimal path connecting each source SN and the BS for energy-use optimization in different network environments. The metrics for energy optimization could be different, such as minimum hop, minimum total energy, maximum–minimum residual energy, minimum reluctance, etc., or a combination of two or more of the above [100]. For example, in [100], given the position and traffic generation rate at each node, the maximum lifetime routing (MLR) problem is defined as a linear programming (LP) problem. By using the link costs which are determined by both the energy consumption rates and residual energy levels at the source-destination pairs, a shortest cost routing algorithm is presented. In [101], the energy-aware routing problem is also formulated by LP, and the hop-constrained energy-aware routing (HC-EAR) is proposed. It guarantees that any packet from a source SN can reach the BS with a bound on the maximum number of hops, thus the maximum energy spent by any node is bounded. In [102, 103], several optimization problems for data flow management with delay constraints in WSNs have been considered. Based on a hierarchical architecture, the models based on LP and Integer LP formulations for finding the flow allocation strategies for either splittable or nonsplittable traffic flow are presented, respectively.

However, the implementation of these solutions depends on certain kinds of global state information of the entire network. This is greatly hindered by the limited capabilities of SNs and thus is not always practical. Designing a distributed heuristic solution for routing support to achieve energy-efficiency in WSNs is highly suitable. Therefore, many distributed routing protocols have been proposed.

In [104], Sohrabi et al. propose the Sequential Assignment Routing (SAR) scheme. Multiple information delivery trees are built such that the root of each tree is a one-hop neighbor of the BS. According to the energy supply, QoS metrics associated with the paths, and the priority level of each packet, a source SN selects a path on one of the trees for its data to be routed back to the BS. In [105], Boukerche et al. develop the Energy-Aware Data-centric routing heuristic (EAD) to alleviate the potentially large amount of storage overhead by storing two-hop neighborhood knowledge at SNs. EAD eliminates the necessity of knowing the positions of SNs. In [106], Chen et al. describe the Minimum Cost Forwarding Algorithm (MCFA). In MCFA, each SN maintains the minimum cost estimate from itself to the BS. Each SN broadcasts every packet to be delivered to its neighbors, who will rebroadcast the packet if it is on the minimum cost path between the source node and the BS. Thus, MCFA eliminates the cost of maintaining a unique ID or a routing table at each SN. In [107], Servetto et al. propose routing protocols with random walks to achieve load balance in a statistical manner through multipath routing. Each SN is assumed to have a unique ID and can alternately switch on and off. When a SN sends a packet towards the BS, its next hop node which is closer to the BS is selected according to a computed probability.

Hierarchical Multi-hop Information Collection Protocols

Depending on how the hierarchical structure is formed, hierarchical information collection protocols can be grouped as reserved tree based, chain based, and clustering based. Among these, the clustering-based approaches have received increased attention due to their effectiveness, lower complexity, and flexibility.

Low Energy Adaptive Clustering Hierarchy (LEACH) [51] is a well developed clustering-based protocol dedicated to continuous energy efficient information collection in WSNs. Each node selects itself as a CH with a calculated probability; other nodes make the decision to join a specific cluster that requires minimum communication energy. A periodic rotation of cluster formation is invoked to ensure balanced energy dissipation among all nodes. However, LEACH is used for proactive application scenarios, and it does not take into account the energy consumption for idle sensing of the channel, and the formation of clusters is not energy aware. Therefore, some efforts have been made to further improve its performance. Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [52] is a chain-based protocol. Instead of sending data packets directly to CHs as done in the LEACH protocol, each node forwards its packets to the destination through its closest neighbors. PEGASIS is reported to outperform LEACH by 100 to 300 percent. Threshold-sensitive Energy-Efficient Sensor Network (TEEN) [108] adopts the cluster formation method of LEACH, but uses thresholds to achieve enhanced control on SNs. This hierarchical routing protocol is responsive to sudden changes (or interest) in the sensing field. It also avoids the energy wastage due to idle sensing. Adaptive Periodic TEEN (APTEEN) [79] fits in the requirements of hybrid application scenarios using enhanced query management and a modified TDMA MAC protocol. In both TEEN and APTEEN, the concept of multi-level clustering is used. In Scalable Source Routing (SSR) [109], a virtual ring is formed that connects all nodes through predecessor/successor source routes. Similar to distributed hash tables in overlay networks [110], SSR maps the whole address space to the set of active nodes. SSR is a completely self-organizing routing protocol, which is suitable for random WSNs.

Localized Information Collection Protocols

In this type of protocol, SNs are distinguished by their locations, which can be identified by either using a GPS, or some localization protocols [111]. Based on the location information, energy conservation can be achieved by switching off a SN when there is another SN active within its vicinity.

In Minimum Energy Communication Network (MECN) [112], with the assistance of a low-power GPS, the relay region for each SN is computed. Then, a global minimum power path between any source – destination pair is found by using a localized search without considering all nodes in the network. Small MECN (SMECN) [113] enhances the MECN protocol by considering potential obstacles between any pair of SNs. In the Geographic Adaptive Fidelity (GAF) algorithm [72], each node uses its location information to associate itself with a "virtual grid" such that all nodes in a particular grid square are equivalent in the sense of forwarding packets. According to application and system information, nodes in the same grid can coordinate with each other to determine which one should sleep and the length of sleep. Nodes take turns to be in the active state and similar energy draining pace is pursued among nodes. In SPAN [71] , power saving is achieved by building an efficient backbone for information relaying without causing significant negative effects on the capacity or on the connectivity of the original network. With SPAN, each node makes its decision to be active as a coordinator based on its residual energy as well as an estimate of the number of its neighbors that would benefit from its being awake.

2.3.2.2. Energy-Efficient Information Dissemination Protocols

Information dissemination plays a critical role in WSNs. This is particularly the case in reactive and hybrid application scenarios, in which time sensitive information should reach other nodes as soon as a serious phenomenon (e.g., early warning of a fire) is detected by some nodes or when a command with certain attribute values should be spread in the network in a timely manner.

In general, information dissemination is conducted similar to flooding, but conventional flooding schemes will cause problems of redundancy and overlap, which lead to significant energy waste.

Pure flooding [114] is the simplest but most costly approach for disseminating information across the entire WSN. An advantage of pure flooding is simplicity and robustness, as its implementation does not depend on any kind of global network state information or neighbor-discovery mechanism. Deficiencies associated with pure flooding include explosion, overlapping, collision, and resource blindness [7, 115, 116]. These make it undesirable in resource-scarce WSNs. Gossiping [7] avoids the problem of explosion in pure flooding and works in such a way that a sender randomly selects one of its neighbors as the message receiver rather than broadcasting it to all neighboring nodes. Gossip routing is originally designed for those wired networks whose network nodes are connected via point-to-point links. To implement in WSNs, one possible solution is to employ TDMA- or CDMA-like MAC protocols to ensure the intended receiver is the only recipient in the transmission range of the sender. The implementation of gossiping requires a SN to maintain its one-hop neighborhood knowledge. Sensor Protocols for Information via Negotiation (SPIN) [115, 116] have been proposed to reduce energy wastage caused by the classic flooding scheme. A new type of control message – metadata is employed to allow negotiation between neighboring nodes, so that a node only forwards a packet to its neighbor who wants to receive the data. However, overhead for control messages is created for negotiation, which will lead to longer latency. Moreover, each individual node has to constantly maintain a neighbor list and update it periodically. This not only requires more memory space, but also costs extra energy. Therefore, generating and controlling metadata are critical to the success of the SPIN.

2.3.2.3. Query-based Information Retrieval Protocols

In many applications, communication protocols are query-driven, that is, the end-user initiates a query about a property of an interesting phenomenon, and the related SNs respond on demand. For example, the end-user may ask for "the temperature in room 717" or "the areas where the temperature is over 50°C". With such an attribute-based query, effective information retrieval protocols are required to have high accuracy and low delay, and consume as little energy as possible.

Directed Diffusion [117, 118] is designed to coordinate SNs to carry out distributed environment monitoring tasks collaboratively. It incorporates data centric distribution, in-network data aggregation, and data caching while enforcing adaptation to the empirically best path. It aims to establish efficient *n*-way communication from single or multiple sources to sinks. Much following work is proposed based on the architecture of directed diffusion and attempts to improve it.

In [119], Heidemann et al. present a physical implementation of directed diffusion with a wireless sensor network test-bed and show that the traffic can be reduced by up to 42% when deploying a duplicate suppression data aggregation method.

The Gradient-Based Routing (GBR) protocol [120] aims to enhance the Directed Diffusion protocol by adaptively adjusting the height of each node, which is initially set to the minimum-hop distance between the node and the BS. An adjustment on the height value associated with a node can be based on the change in either the amount of its remaining energy or the amount of load that the node carries. The gradient of a direct link is defined to be the difference between the height of the particular link's tail node and its head node. Each node forwards data packets onto its outgoing link resulting in the largest gradient.

The Energy-Aware Routing (EAR) protocol [121] combines the benefits of multiple-path routing and the availability of location information of nodes to improve the performance of Directed Diffusion. Each node is identified through a class-based address including its geographical location and type. Multiple paths are used based on a probability derived from the path energy quality to prolong the lifetime of the network.

Voronoi Scoping for Multi-sink WSNs [122] employs multiple sinks to achieve efficient query dissemination and tree construction for data gathering. It limits dissemination of query messages from each sink within its Voronoi cluster, i.e., only a subset of the SNs, which are as small as possible, while guaranteeing that each node in the network receives the message from the sink closest to it.

A brief summary of various communication protocols is shown in Figure 2.8. Readers interested in more energy efficient communication techniques may refer to [16, 123].

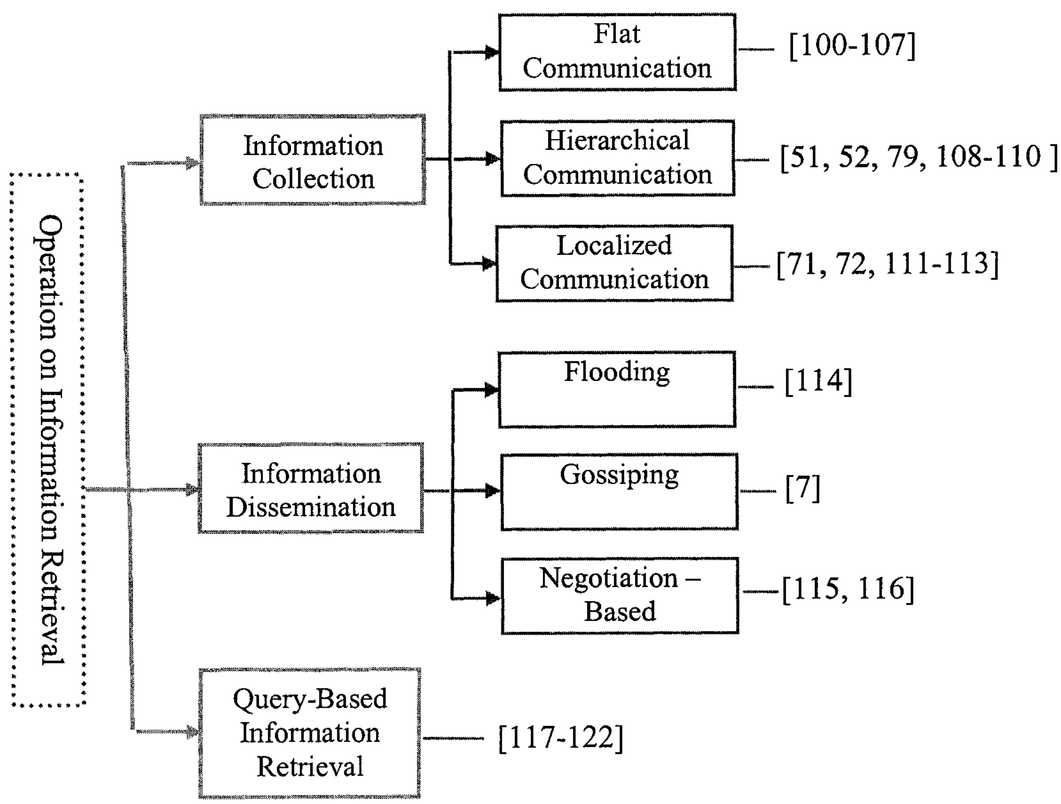


Figure 2.8 Classification of Data-centric Routing in Wireless Sensor Networks

2.3.3 Signal and Data Processing

As we mentioned in previous subsections, most energy efficient communication protocols incorporate localized signal and data processing to reduce the amount of traffic in the network. Two main categories of signal and data processing applied in WSNs are Data Aggregation and Collaborative Signal and Information Processing (CSIP).

2.3.3.1 Data Aggregation

The principle of data aggregation or fusion is to minimize traffic load (in terms of number and/or length of packets) through eliminating redundancy. It applies a novel data-centric approach to replace the traditional address-centric approach in data forwarding [99]. Specifically, when an intermediate node receives data from multiple source nodes, instead of forwarding all of them directly, it checks the contents of incoming data, and combines them by eliminating redundant information under the constraints of acceptable accuracy.

Several data aggregation algorithms have been reported in the literature. The most straightforward is duplicate suppression, i.e., if multiple sources send the same data, the intermediate node will only forward one of them. Using a maximum or minimum function is also possible. The greedy aggregation approach [124] can improve path sharing and attain significant energy savings when the network has higher node densities compared with the opportunistic approach. In [99], Krishnamachari et al. described the impact of source-destination placement on the energy costs and delay associated with data aggregation. They also investigate the complexity of optimal data aggregation. In [125], a polynomial-time algorithm for near-optimal Maximum Lifetime Data Aggregation (MLDA) is described for data collection in WSNs. The scheme is superior to others in terms of system lifetime, but has a high computational expense for large sensor networks. In [15], a simple and efficient clustering-based heuristic for maximum lifetime data aggregation (CMLDA) is proposed for both small and large-scale sensor networks. In [126], the methods for using sketches to produce accurate aggregation results are presented for a WSN database. By generalizing the duplicate intensive sketches, the computation and communication overhead remains at a low level. In [127], the trade-off between energy conservation and accuracy by using data aggregation is discussed. In contrast to the conventional snapshot aggregation problems, the periodic aggregation problem is considered. A distributed estimation algorithm is presented as well.

2.3.3.2 Collaborative Signal and Information Processing (CSIP)

CSIP schemes are also able to reduce the amount of traffic transmitted and thus result in energy efficiency in WSNs. With the combination of interdisciplinary techniques, such as low-power communication and computation, space-time signal processing, distributed and fault-tolerant algorithms, adaptive systems, and sensor fusion and decision theory, CSIP is expected to provide solutions to many challenges, including dense spatial sampling of interesting events, distributed asynchronous processing, progressive accuracy, optimized processing and communication, data fusion, querying and routing tasks [128]. CSIP can be implemented either through coherent signal processing on a small number of nodes in a cluster or through non-coherent processing across a larger number of nodes when synchronization is not a strict requirement [67]. CSIP algorithms can be classified [75] as information-driven schemes [14], mobile agent based schemes [129], which attempt to reduce the system traffic by employing an agent – transmitting the integration process (code) to the data sites instead of moving original data directly, and relation-based schemes [65], which use a top-down approach to select the sensor nodes to sense and communicate based on a high-level description of the task.

Besides the work discussed in this chapter, energy efficiency can be achieved in many other aspects. For example, energy efficient MAC protocols have been developed, including protocols for channel detection and allocation, collision avoidance, scheduling, power management, etc. [17, 130]. The concept of cross layer design has been proposed and applied [131, 132]. Moreover, network services such as synchronization [133] and localization [134] are also helpful to improve energy efficiency. Due to space limitations, we cannot describe approaches for energy efficiency and/or cost reduction in all aspects.

2.4 Discussion

Having thoroughly studied various aspects of the WSN design, we observe that the successful implementation of a WSN cannot be realized without a well constructed networking planning strategy. Being the first and essential step in building a WSN, all other functional protocols can only operate on top of a planning strategy.

Therefore, it is our focus in this thesis to explore the network planning issue and provide a comprehensive and feasible solution. As far as we know, no similar work has been reported in the literature. Our work shall fill the blank in this field. The research in this thesis is rooted from some existing work, but differs significantly from them.

Firstly, all existing research work concerning device provisioning only tackles the placement of a single type of device with an assumption that other devices have been deployed properly. However, a heterogeneous WSN may contain multiple types of devices. We consider a network planning strategy from the ground up for a WSN with two or more kinds of devices. Moreover, we take coverage, connectivity, lifetime, and cost into consideration at one time. Our research focuses on the applications for data collection from all SNs to the BS.

Due to the high complexity of the generic network planning problem, we propose a modular framework, consisting of a sensing domain and a communication domain. Such a framework is based on the functionality separation of devices, and it effectively and significantly simplifies the network planning efforts. This is partially inspired by the well accepted hierarchical system architecture (e.g., reviewed in Section 2.3.1). This makes it easy and practical to collaborate with other efficient schemes, such as multi-hop communication protocols (e.g., described in Section 2.3.2) and signal and information processing schemes (e.g. , introduced in Section 2.3.3).

Secondly, in the sensing domain, most of the previous work considers the uniform area coverage, in which the whole area over the entire sensing field is treated equally. We find that the points of a sensing field may be differentiated based on location-related importance. Therefore, we propose the concept of information-oriented sensing coverage, which is novel in the area of SN deployment. Moreover, taking into account that the simplified 0-1 binary disc sensing model cannot represent the nature of strength attenuation of a collection of signals, we utilize the signal strength based sensing model. Detailed description of such a sensing model is presented in Chapter 3.

Thirdly, in the communication domain, the design objective is to use a minimum number of RNs to ensure each and every device has at least one workable path to deliver all its traffic amount to the BS with a lifetime guarantee. As the connectivity properties of SNs and RNs differ from one another, we split the design of the communication domain into two phases. The first phase is to ensure the connectivity for all SNs while the second phase is to ensure the connectivity for all RNs.

For the placement of first phase RNs (FPRNs), we formulate the problem as a minimum set covering model. No similar work has been reported in the literature. For the placement of second phase RNs (SPRNs), we identify two components: the RN position determination and traffic allocation. Our formulated SPRN placement problem is a complimentary problem to the one described in [19]. To determine the positions of SPRNs, we propose the Far-Near and Max-Min principles. Traffic allocation and RN positioning couple tightly and affect one another greatly. Although many traffic allocation and routing protocols have been developed as reviewed in Section 2.3.2, all of them are implemented based on a deployed WSN. In the procedure of SPRN placement, we use linear programming to provide a locally optimal solution for traffic allocation.

A detailed description of our proposed network planning framework will be illustrated in Chapter 3. In the same chapter, we also introduce the system architecture, the sensing model and the communication model, and introduce the generalized network planning design problem and its variations. In Chapters 4-6, we propose several schemes to solve the corresponding problems defined in Chapter 3.

3. SYSTEM MODELS AND PROBLEM STATEMENT

In this chapter, we will first present the system architecture. Next, we will introduce the sensing model and the communication model. Then, we will describe the energy model and lifetime constraints. Finally, we will describe the generic problem of network planning in WSNs and its variations under different scenarios.

3.1 System Architecture

When sensing and communication tasks become more complicated, a WSN tends to contain more devices and/or operate in a larger area. It is an inevitable tendency that WSNs evolve from homogeneous networks to heterogeneous networks. In the former case, all nodes are identical in terms of initial energy, computation and communication capability, etc.; in the latter case, multiple types of nodes with different functionalities and capacities co-exist in a system.

Compared with homogeneity, heterogeneity is a more cost-effective approach to a complex system. Devices with different functionalities and capacities can take different responsibilities, such as sensing various physical quantities; routing and relaying; data processing and aggregating; clustering and coordinating, etc. Using a limited number of highly capable but more expensive nodes can lead to a reduction in device cost for the rest of the network. Moreover, overall system lifetime can be extended because of better fairness of energy dissipation among the different devices.

Moreover, the standardization of WSN protocols, such as IEEE802.15.4 [24] and ZigBee [135], enables devices of various types to communicate and collaborate together. Thus, manufacturers with different expertise can provide products with different functions at different prices. For example, of the more than 100 members in the ZigBee Alliance nowadays, Honeywell and Omron specialize in designing and producing sensing and control devices, Motorola and Samsung are known for their networking technology [135], and Intel has invested a lot of effort in system design and experimental implementation in the real world.

We now describe how the basic functionalities of a device can be abstracted. Figure 3.1 demonstrates a heterogeneous WSN composed of nodes with different functionalities and capacities:

- Sensing-and-transmitting (*ST*): refers to the functionalities of sampling the physical attributes of the environment, conducting A/D conversion, assembling data packets, and transmitting packets to neighboring nodes.

- Routing/relaying (*RR*): refers to the functionalities of establishing and maintaining the routing scheme, and forwarding received data packets to next-hop devices toward the destination.

- Aggregating (*AGG*): refers to the functionality of combining two or more pieces of raw data into a single stream. By doing so, the amount of traffic to be relayed may be reduced.

- Coordinating (*CO*): refers to the capability of managing the operation of the other nodes. With this functionality, a device can conduct synchronization and scheduling of medium access among the competing nodes within its proximity, and instruct other nodes to switch to power-saving modes. Hence the power wastage due to collision and idle listening can be effectively reduced.

- Data Sink (*DS*): represents the destination of data transmissions. This function typically resides in the Base Station (BS).

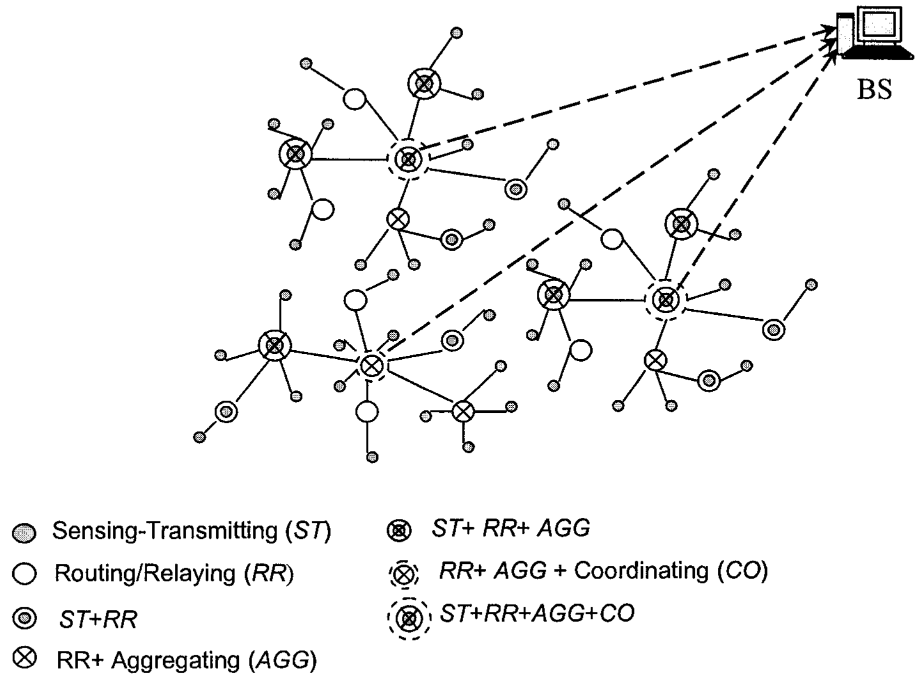


Figure 3.1 The Functional View of a Heterogeneous WSN

As sensing and communication are the primary functionalities of a WSN, in the following sections, we will introduce the sensing model and the communication model in detail.

3.2 Sensing Model

3.2.1 Binary disc sensing model

Most previous work in the literature adopted a binary disc sensing model, i.e., a point is sensed with probability 1 if it is within a fixed sensing radius of at least one sensor, or with probability 0 otherwise.

The 0-1 model is simple, thus it is easy to utilize in modeling the area sensing coverage problem. However, such a simplified sensing model is not always realistic. Therefore, several recent efforts have studied more generalized sensing models [136, 137], which are signal strength based sensing models.

3.2.2 Signal strength based probabilistic sensing model

The signal strength based sensing model attempts to capture the property that signal strength decays as it travels farther away from a source. As such, the probability that a SN detects a signal decreases as the distance between the source and the SN increases. Our research of SN deployment is independent of any particular sensing model. However, to facilitate the discussion, we use the signal strength based probabilistic sensing model, which is described in [136]. The model is valid for a variety of signals, such as radio, acoustics, vibration, etc.

According to [136], the received signal strength $P_r(d)$, at a particular location with separation d between the source and the sensor, is the difference between transmission signal strength at the source O, $P_t(o)$, and the path loss $PL(d)$[1].

$$P_r(d) = P_t(o) - PL(d) \tag{3-1}$$

$$PL(d) = \overline{PL}(d_0) + 10n\log\left(\frac{d}{d_0}\right) + X_\sigma \tag{3-2}$$

where X_σ is a zero mean Gaussian distributed random variable (in dB) with standard deviation σ, which describes the phenomenon of log-normal shadowing; d_0 is a close-in

[1] All signal strengths are in dB.

radial distance; n is the path loss exponent; and $\overline{PL}(d_0)$ is the average path loss at d_0, which can be obtained based on either experimental measurements or on a free space assumption from transmission to d_0 [138].

Let $p_{s\to o}$ denote the sensing probability provided by a sensor S, to the point of interest O. Assuming a SN can only successfully detect a signal with strength greater than a threshold γ, according to [138], we have :

$$p_{s\to o} = \Pr ob[P_r(d) > \gamma] = Q\left(\frac{\gamma - \overline{P}_r(d)}{\sigma}\right) \tag{3-3}$$

where

$$\overline{P}_r(d) = P_t(o) - \overline{PL}(d) = P_t(o) - \left[\overline{PL}(d_0) + 10n\log\left(\frac{d}{d_0}\right)\right] = \overline{P}_r(d_0) - 10n\log\left(\frac{d}{d_0}\right) \tag{3-4}$$

and the Q-function is defined as:

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^{\infty}\exp(-\frac{z^2}{2})dz = \frac{1}{2}\left[1 - erf\left(\frac{x}{\sqrt{2}}\right)\right] \tag{3-5}$$

$$erf(x) \triangleq \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}dz \tag{3-6}$$

Figure 3.2 illustrates an example diagram of the above signal strength based sensing model.

Moreover, in a sensing field, one point of interest could be covered by multiple sensors. The joint sensing probability is given by [136]:

$$p^{(m)} = 1 - \prod_{i=1}^{m}(1 - p_{s_i \to o}) \tag{3-7}$$

where a total of m sensors cover a single point of interest and $p_{s_i \to o}$ can be calculated according to Eqs. (3-1) to (3-6).

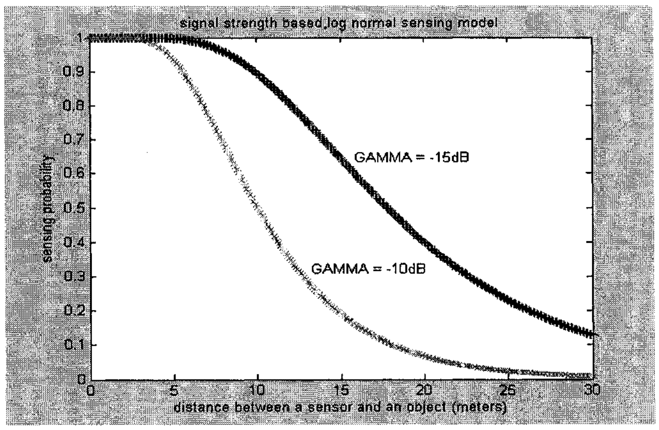


Figure 3.2 Illustration of Signal Strength Based Sensing Model

3.3 Communication Model

As communication energy is the major component of total energy consumption in a WSN, it is always a primary objective to provide energy efficient communication by either reducing traffic volume or by separating the single hop long transmission distance into multiple short ones.

It has been well accepted that clustering based hierarchical communication is an effective platform to achieve these objectives. Nodes with higher capacities (in terms of more complicated hardware, processing/computational ability, extra initial energy), take the responsibility as Cluster Heads (CHs). They have a much heavier workload, and thus consume more energy than the normal Member Nodes (MNs).

In this thesis, we adopt such a hierarchical communication model. Although the devices can have various functionalities, from the perspective of communication capability, we classify them into three types: the Sensor Nodes (SNs), the Relay Nodes (RNs) and the Base Station (BS).

Individual SNs may differ in a few aspects including the physical properties they sense, the rate at which they generate data, their transmission ranges, etc. However, all SNs are assumed to be small and equipped with a constrained and un-replenished power supply. The functionalities of a SN are limited to sensing the environment, generating data, and transmitting data to a neighboring RN within its proximity. For cost and energy efficiency, SNs do not relay traffic for other nodes. One example of such a SN is a Reduced Function Device (RFD) defined in the IEEE802.15.4 standard [24]. Moreover, the traffic pattern of an individual SN is assumed to be predictable.

A RN is devoted to relaying data which are generated at SNs to the BS in one or multiple hops. It may have limited or (practically) unlimited energy supply. It may have fixed transmission power or, alternatively, it may be able to adjust its transmission power to avoid energy wastage. A RN is assumed to have greater computation capacity and carry a greater power supply than the SNs. A RN is capable of routing/relaying, and acts as a Cluster Head (CH) when active to group the SNs in its vicinity into a cluster. A RN also coordinates MAC layer operation within the cluster and fulfills data processing (e.g., aggregating data) so that energy is efficiently utilized.

The BS is generally located away from the sensing field. It acts as the Data Sink (DS) for all data packets.

The hierarchical communication model is shown in Figure 3.3.

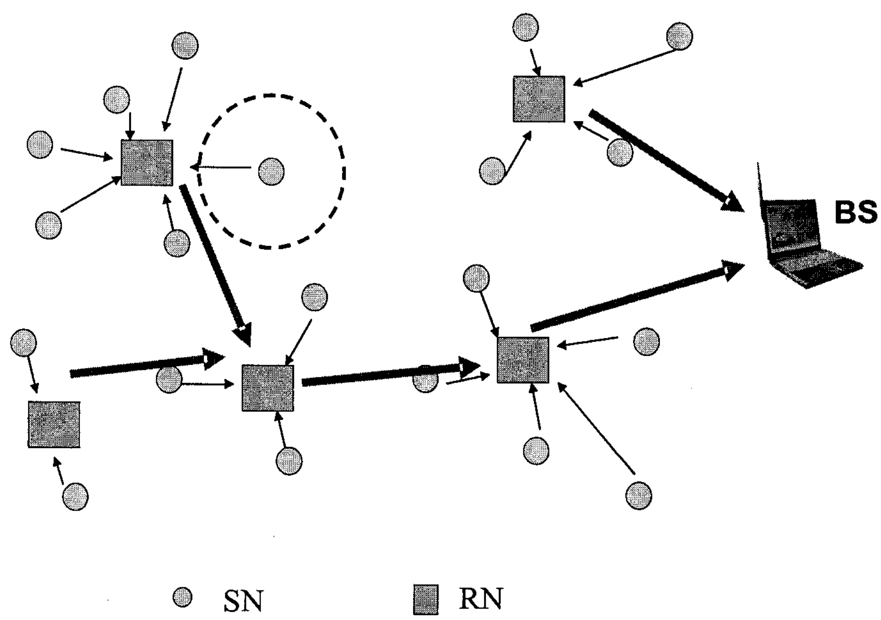


Figure 3.3 A Hierarchical Communication Model of a Heterogeneous WSN

3.4 Energy, Lifetime and Transmission Range

In this thesis, we consider wireless data communication as the dominant energy consumption source in a WSN. The energy consumption of a device is a function of its radio transmission power and the amount of traffic it carries. In order for an energy-constrained node to function for a desired lifetime T, its transmission radius and/or its traffic amount should be constrained.

The traffic pattern of a SN is considered predictable and obtainable through the product specification or experimental measurement. Therefore, given the average data rate of a SN, the required calendar lifetime T can be translated into the cumulative traffic amount to be generated by the SN. Using the well recognized radio model [78], in order to send the desirable amount of traffic, the transmission distance of a given SN should be confined according to:

$$d_{SN} \le \sqrt[n]{\frac{J_{SN}/(R_{SN} \cdot T) - \alpha_{1_{SN}}}{\alpha_{2_{SN}}}} \tag{3-8}$$

where J_{SN} is the portion of its initial energy allocated for the successful transmission of data packets, R_{SN} is the average traffic rate during its lifetime, $\alpha_{1_{SN}}$ and $\alpha_{2_{SN}}$ are transmitter circuit parameters, and n is the path loss exponent.

Once the data is sent to RNs from SNs, RNs further convey the data to the BS. In order to achieve the projected lifetime T, the traffic rate at which an energy-limited RN receives and sends should be carefully planned. Injecting too much traffic into the RN will violate its energy constraint. For any RN, say RN_i, the incoming traffic rate should be equal to the outgoing traffic rate. Thus, we will not explicitly distinguish between the two in the following unless necessary. Let R_i denote the traffic rate of RN_i. In order not to violate the energy constraint, the following condition should be satisfied.

$$\sum_{j:RN_j \text{ is } RN_i\text{'s next hop } RN} r_{ij} \cdot T \cdot (\alpha_{1_{RN}} + \alpha_{2_{RN}} \cdot d_{ij}^{\ n} + \beta_{RN}) \leq J_i \tag{3-9}$$

$$\sum_{j:\, RN_j \text{ is } RN_i\text{'s next hop relay}} r_{ij} = R_i \ , \tag{3-10}$$

where $\alpha_{1_{RN}}$, $\alpha_{2_{RN}}$ and β_{RN} are circuit specific parameters and r_{ij} is the traffic rate at which RN_i transmits to RN_j, d_{ij} is the distance between RN_i and RN_j, and J_i is the initial energy of RN_i.

The relationships between energy supply, traffic rate, and transmission distance are clearly shown in inequality (3-9). Depending on its power control capability, a RN may either transmit with fixed transmission power, in which case it is only able to reach other RNs within its fixed transmission range, or a RN may be capable of adjusting its transmission power to reach the intended next hop relay or the BS with minimal required transmission power. In the former case, given a RN's initial energy J_{RN} and transmission range d_{RN}, the maximum traffic rate in order to achieve the lifetime T is given by:

$$R_{RN} \leq \frac{J_{RN}}{\left(\alpha_{1_{Rn}} + \alpha_{2_{RN}} \cdot d_{RN}^{n} + \beta_{RN}\right) \cdot T} \tag{3-11}$$

This is defined as the capacity of the RN in the remainder of this thesis. If the traffic rate injected into a RN does not exceed its capacity, the lifetime constraint is satisfied. When a RN has variable transmission range, it only transmits data over a small distance when handling a high traffic rate, or over a long distance if its traffic rate is low.

3.5 Problem Statement

In this research, we consider the generic device placement planning problem as follows.

Given a specific sensing task, determine the number and placements of heterogeneous devices, so that the total network cost is minimized while the constraints of lifetime, coverage, and connectivity are satisfied.

According to the energy consumption model discussed in Section 3.4, lifetime constraints for SNs can be transformed to the constraints on their transmission range. Hence, to guarantee the lifetime requirement of a SN, at least one RN must be located within the range of a SN's transmission range. If so, the connectivity requirement of a SN is satisfied automatically. Hence, the lifetime and connectivity constraints for SNs can be considered together.

The constraints on RNs are more complicated compared with those of SNs. First, the possible locations of the RNs are restricted by the SN connectivity requirements. Next, the initial energy limitation (if any) of a RN imposes the constraints on its lifetime, which depends on its transmission range as well as the total traffic amount of SNs or other RNs it serves. Last but not least, for a RN with limited transmission range, it usually cannot reach the BS directly. Thus, each RN must have at least one (multi-hop) path to the BS.

Therefore, the general form of the network planning problem is shown in Figure 3.4.

minimize: f_{cost}(system cost) *subject to:* g_{SN}(SNs lifetime requirements and connectivity constraints) $g_{RN\text{-}L}$(RNs lifetime requirements) $g_{RN\text{-}C}$(RNs connectivity constraints)

Figure 3.4 The General Form of the Network Planning Problem

The cost of a device is determined by its functionality and capability, as well as the manufacturer. With the different combinations of the functions and power supply methods, there may be different types of nodes in a single network. Each of these functionalities and power supply methods is associated with a certain amount of cost. The more functions a device has, the more expensive it is likely to be. In addition, the capacities, such as power supply and memory size, also have an impact on the device cost. For convenience and without losing generality, in this thesis, we assume that devices of the same type have the same cost.

Finding a globally optimal solution to the above general problem is highly non-trivial. Therefore, we propose a modular framework. Specifically, differentiated by basic functionalities of sensing and communication, the entire network can be separated into two domains: a sensing domain and a communication domain. The sensing domain aims to meet the coverage requirement under mission criticality; the communication domain aims to provide connectivity support. The resultant positions of SNs from the sensing domain design, as well as other properties of SNs, such as their transmission range and traffic amount, are input parameters to the communication domain design problem. Moreover, due to the diversity of RN specifications, the communication domain design is split into two phases. The design of phase one is dedicated to satisfying SN connectivity with a minimum cost under lifetime constraints while phase two is to ensure that each device has at least one path to deliver traffic to the BS with a minimum extra cost while satisfying lifetime requirements. Figure 3.5 illustrates our proposed framework as well as its components.

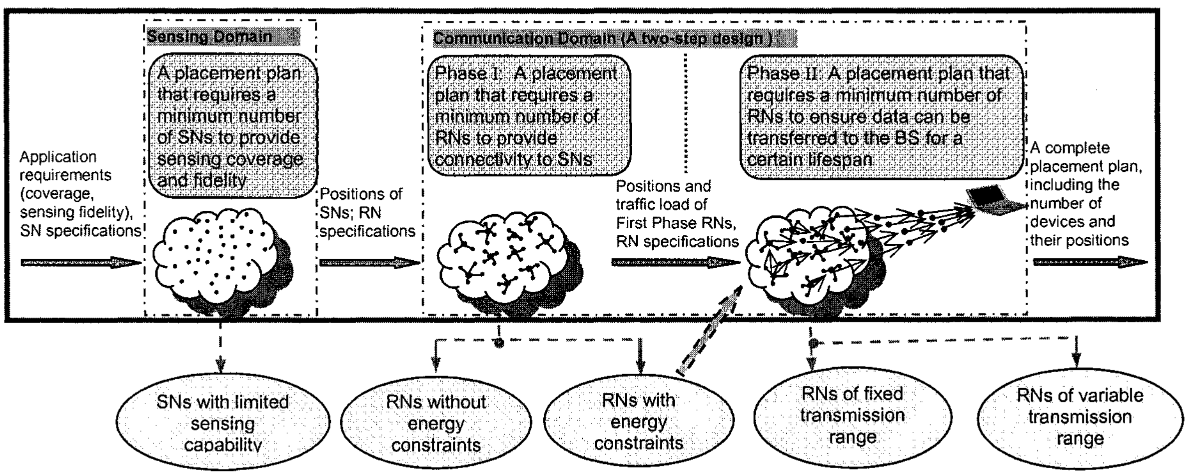


Figure 3.5 Networking Planning Framework and its Components

By using such a modular framework, we obtain (but are not limited to) the following advantages:

- It is easy and practical for standardization and commercialization. Coupled with an open communication standard, the network components from different manufacturers can be integrated in an economical manner.

- The overall network lifetime will be prolonged. The SNs can unload the heavy burden for transmission, thus, the simple and non-expensive nodes can satisfy the lifetime constraints. Moreover, since the SNs are treated individually based on their initial energy and traffic pattern, the bottleneck nodes will be deleted.

- The network can be easily upgraded and has low maintenance cost.

A. Design of Sensing Domain

Because sensing is the basic task of a WSN, the framework begins with the sensing domain specification, which is determined by the application and consists of three major components: problem identification, sensor selection, and sensor deployment strategy. First, the required physical measurements and critical sensing spots should be recognized. Then, the suitable types of SNs will be determined. For the sake of cost conservation, it is desirable to use products which can be found on the market. Lastly, a sensor deployment strategy will be developed with the objective of minimum cost under a coverage constraint. A general sensing domain design problem is:

Given a specific sensing task and SN specifications, determine the numbers and positions of SNs, so that the desirable sensing coverage is satisfied with a minimum device cost.

As sensing coverage requirements differ greatly from one case to another, the design of the sensing domain can vary considerably. In Chapter 4, we will present a sensor deployment proposal that determines the positions of a minimum number of SNs to achieve information-oriented sensing coverage.

B. Design of Communication Domain

Since SNs have a limited power supply, instead of affording long distance transmission, they can only support local communication within a short distance. Moreover, the energy consumption for communication is much higher than that for sensing and computation. To conserve energy, SNs do not relay information for one another. To ensure that all SNs can report their information to the destination, a communication domain should operate on top of the sensing domain to provide connectivity under lifetime constraints. The challenges include deciding how many RNs are sufficient and where they should be located. Then the generic communication domain design problem becomes:

> *Given a specific sensor deployment, determine the number and positions of RNs, so that the total network cost is minimized while the constraints of lifetime and connectivity are satisfied.*

Because device properties can greatly vary, the design of the communication domain appears different from one case to another. Among all the factors, though, the energy supply and transmission range of a RN are critical to distinguish the design scenarios.

On one hand, the initial energy of a RN can be unconstrained, fixed, or planned. The meaning of unconstrained is in the sense that a RN has ample energy supply (e.g., wall power, solar power, or high capacity battery) relative to the lifetime requirements, hence, the energy constraint is not a critical issue compared to that for the SNs. In such a case, the energy/lifetime constraints on RNs are relaxed. The meaning of fixed is that a RN is equipped with an un-replenished and limited power supply. Planned means, a RN is strategically equipped with individualized initial energy, which is based on several factors, such as its location, the total traffic volume it has to handle, etc.

On the other hand, the RNs can either have fixed or variable transmission range. Fixed refers to the case in which a RN always transmits with a fixed power, thus the transmission from a source RN can only reach the destinations that fall within its fixed transmission range. When a RN is able to adjust its transmission power to the surrounding conditions (e.g., distance to its destination, SNR level), its transmission range is adjustable, too.

In the following, we envision three typical scenarios, which correspond to the above variations. Solutions to these problems will be explored in detail in Chapters 5 and 6. Figure 3.6 illustrate a brief summary.

SCENARIO ONE: This is a basic version of the generalized communication domain design problem. SNs are energy constrained and transmit data over a small distance. RNs have ample energy supply (wall power, solar power, or high capacity battery) so that their energy constraints are not a critical issue compared to that of the SNs. In such a case, the energy/lifetime constraints on RNs are relaxed. RNs can send data to the BS in one hop, or via multi-hops by a communication mechanism. Therefore, the optimization problem is:

given a deployment of SNs, find a minimum number of RNs and their positions so that each SN can reach at least one RN in a single hop, and the lifetime constraints of the SNs are satisfied.

In Chapter 5, we model this problem as a minimum set covering problem and propose an algorithm to solve it optimally.

SCENARIO TWO: In this scenario, the power constraints on RNs are explicitly considered. Thus, each RN can only relay a limited amount of traffic for the SNs within a restricted distance. Furthermore, each RN is assumed to have fixed transmission range. RNs have to relay data in a multi-hop fashion to the far-away BS. In this scenario, the problem is the same as that in scenario one but with the added constraint that each RN must have at least one (multi-hop) path to the BS, while the lifetime constraint is satisfied.

As a result, this becomes a compound problem of RN placement and traffic allocation. We should not only determine the positions of RNs that collect information from SNs, but also decide the traffic relaying routes for each RN to the BS. As these two components are coupled with one another tightly, it is difficult to provide a globally optimal solution at one time. Therefore, we propose a two phase RN placement strategy. The first phase RN (FPRN) placement is dedicated to satisfying connectivity and lifetime constraints of each and every SN. This problem is further modeled as an enhanced minimum set covering problem, and its solution is discussed in Chapter 5. In addition, the second phase RN (SPRN) placement is dedicated to ensuring the connectivity and lifetime constraints of all RNs. The detailed solution to this problem is presented in the first part of Chapter 6.

SCENARIO THREE: Similar to the situation in scenario two, RNs are energy constrained. However here each RN can adjust its transmission power adaptively. Thus, its transmission range is adjustable, too. This is also a compound problem of RN placement and traffic allocation. We can also adopt the two phase planning strategy. The first phase planning is similar to that of scenario two. The second phase planning is more flexible. The techniques are discussed in the second part of Chapter 6.

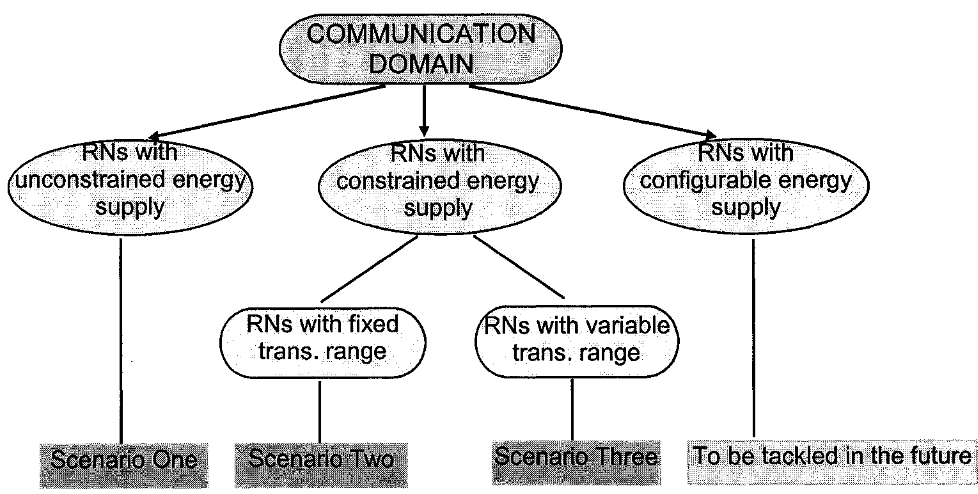


Figure 3.6 Scenarios in the Design of Communication Domain

3.6 Remarks

In terms of the literature, research on network planning for WSNs is quite small and ours is the first work we know of to consider a unified approach to network planning. In this thesis, we do not explicitly consider any specific MAC and routing protocols. Our aim is to provide optimal (or near optimal) network planning strategies that are nonetheless practical in the sense that they can be supported by existing MAC and routing protocols, rather than tailoring our solutions to every imaginable set of constraints that MAC and/or routing protocols might impose. Our view is that although the latter approach may be important and possibly necessary in specific cases, tackling in a comprehensive way the resulting network design problems is much too large a problem for a single thesis. Moreover, optimizing the network planning for just one or a few MAC or routing schemes may narrow the applicability of the "optimal" strategies to just these schemes or others like them, depending on how stringent the constraints are. Thus, our approach is to consider what is possible in terms of optimizing network planning, and in doing so we would like to limit the possibilities only by the most necessary constraints, which we take to be those imposed by the hardware limitations of the devices themselves.

Moreover, the problem of optimally placing WSN devices is to a large extent orthogonal to the MAC protocol used. This is because a WSN is energy limited as opposed to bandwidth limited. Implicitly, we are only assuming that the MAC protocol is working efficiently in the sense of keeping the number of packet collisions small so that we can regard the energy wastage due to retransmissions as negligible. Since we assume that the traffic patterns of SNs is predictable (see Section 3.4), our implicit assumption is quite reasonable and weak. To further support this assumption, when specifying the initial energy of devices (J_{SN}, J_i or J_{RN} in Eqs. (3-8), (3-9) and (3-11), respectively), we can make the assumption that these values specify only the energy that will be used for successful data packet transmission (this includes the energy spent for the transmission of any control packets in the MAC layer used in the successful data packet transmissions). In other words, by conservatively estimating the energy wasted due to retransmission (which can be tailored to the specific MAC protocol used and the specific traffic patterns expected in a given application), we can appropriately adjust J_{SN} and J_{RN} , which are inputs to our formulations. The MAC protocol is essentially orthogonal to the formulations themselves and impose a non-significant impact on the network planning strategies.

Network planning is the first step of a WSN design and is done off-line. Part of the solution to the design problems in scenarios two and three of Section 3.5 are data paths and traffic allocations that, if used, are energy efficient enough to satisfy the given lifetime constraint. By using some centralized routing and traffic management protocols, the desirable traffic allocation objective can be achieved. More generally, we are implicitly assuming that the routing algorithm used, whether it be centralized or distributed, is energy efficient. As reviewed in Chapter 2, there are numerous energy efficient routing protocols that have been proposed in the literature.

4. SENSOR DEPLOYMENT FOR INFORMATION-ORIENTED SENSING COVERAGE

As discussed in previous chapters, a comprehensive and integrated network planning strategy begins from the deployment of sensor nodes. Sensing coverage is an essential measurement of the quality of service provided by a WSN. In this chapter, we discuss the issue of information-oriented sensing coverage and propose a couple of schemes to fulfill the required sensing coverage.

4.1 Information-Oriented Sensing Coverage

Numerous sensor deployment strategies have been proposed in the literature. Many of them are dedicated to ensuring full area coverage, that is, any point in a physical region should be covered by at least one sensing node [45, 47, 61, 139, 140]. However, in many realistic scenarios, the full area coverage requirement cannot always be satisfied or may be unduly difficult to satisfy [58, 59]. This has led to claims of the inevitable necessity of partial area coverage for some real world applications [58]. For such applications, full coverage is either impossible or unnecessary and partial coverage is more realistic. Some reasons why full area coverage may be difficult or infeasible are highlighted below.

- Infeasibility of full area coverage may arise due to a combination of several factors such as signal attenuation with increase of distance, constrained sensing capability, and large sensing field. For example, the magnetic sensor HMC1002 described in [133] can only detect a small magnetic signal at a distance of about 1ft from a stationary target, and about 8-10 ft from slowly moving vehicles. It is difficult to achieve full area coverage if such sensors are used to detect objects and track their motion on a large sensing field.

- Full area coverage may not be worth the cost. For example, if sensors are deployed randomly, a huge number of sensors should be used to ensure complete area coverage in a large field [55]. On the other hand, if grid-based or other deterministic deployment strategies are used, a moderate number of sensors can be carefully deployed to achieve full coverage initially. However, to maintain full coverage as the system operates for a period of time, a large number of redundant sensors are needed [55].

- Full area coverage may be infeasible due to stealth requirements. Even though a sufficient budget is provided, an excess deployment of sensors would severely deteriorate the stealthiness of the WSN. Therefore, in such applications, only a limited number of sensors can be used. This conflicts with the full area coverage requirement.

In addition, absolute full area coverage is not always necessary. For example, in some applications, the objects to be monitored or the possible or crucial occurrences of an event are distributed unevenly over the sensing field. In addition, the detection priorities of individual objects or events may vary with location. For example, on a battlefield, certain objects, such as the general command centers, may be more critical than other objects. As such, intrusions at these key objects invoke higher security concerns. To protect the objects efficiently, engineers should choose to boost the sensing quality around a command center instead of maximizing the sensed area overall. In other words, it is more appropriate to differentiate sensing coverage instead of assigning equal importance to all points of the entire physical area as conventional area-oriented deployment strategies do.

The importance of location-dependent differentiation of sensing priority exists in a wide range of applications, including early warning systems for wildland fires, environmental monitoring, and intrusion detection. The focus of interest should be on the desired objects, areas, or events with greater importance rather than the full sensing field. More abstractly, we say that the objects being monitored on the sensing field or the occurrence of events at different places have different *information* values. Failure to detect an interesting object or event would result in different information loss or penalty, or equivalently different utility or gain, depending on its location. For such applications, a deployment scheme which minimizes the expected information loss (EIL) or maximizes the expected information utility (EIU) for a given number of sensors, or minimizes the number of sensors required to achieve a given information threshold, is more appropriate.

Therefore, in this chapter, we introduce the concept of information-oriented sensing coverage for sensor deployment. In contrast to the conventional area-oriented coverage, information-oriented sensing coverage is more directly concerned with the actual, or at least perceived, benefits resulting from information obtained by sensors. The difference becomes particularly important when full coverage is not achievable with a limited number of sensors.

The property of information differentiation is overlooked in previous work on sensor deployment. Though they do not consider sensor deployment, in [141], Yan et al. do propose an adaptive sensing coverage protocol to provide differential surveillance. The differential sensing coverage is achieved by adjusting the duty cycle of individual sensors in a densely deployed WSN. However, uneven workloads on sensors result in unbalanced energy dissipation among the SNs. Those SNs, which stay active longer than others, deplete their energy faster. Thus, the WSN can no longer be functional if a certain percentage of sensors are out of energy. In our concept of information-oriented deployment, in addition to the fundamental property of providing differential sensing coverage, we have the goal of keeping the SNs consuming their energy at an even pace.

In brief, the significant benefits of sensor deployment for information-oriented sensing coverage are as follows:

- Firstly, it improves a WSN's sensing quality in a cost-effective manner. Taking into consideration the distribution of differential information utility over the whole sensing field, sensors are deployed deliberately at the positions with greater importance. As a result, a differential sensor deployment strategy needs fewer sensors to provide a guaranteed quality of sensing coverage than those proposals which treat all points in the sensing field equally.

- Secondly, for differential sensing the implementation is easier and more practical than the scheme that individualizes the duty cycles of SNs, as this requires SNs to be location-aware and synchronized. Both these two functionalities are costly in terms of computational complexity and energy consumption. Therefore, it is more difficult to implement in reality. In contrast, differential sensor deployment avoids the problems of sensor location-awareness and synchronization.

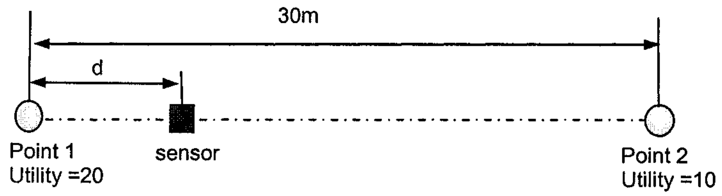
For concreteness, in this chapter we use the *information utility* of a point of interest to quantify its importance. Such utility is application-determined and can be obtained based on previous experience or knowledge of a user's requirements. Given a set of points of interest on the sensing field with known information utility, an information-oriented sensor deployment solution either determines the positions of a given number of sensors such that the EIU is maximized, or it determines the positions of a minimum number of sensors such that the EIU reaches a desirable threshold. In this chapter we focus on the latter problem.

To our best knowledge, the concept of information-oriented sensing coverage and its use for sensor deployment is novel. The work in this chapter is a new direction for sensor deployment research. In the remainder of this chapter, we will formulate the information-oriented sensor deployment problem, and present several heuristic algorithms to solve it. In addition, discussions on performance of the proposed schemes will be based on some experimental results.

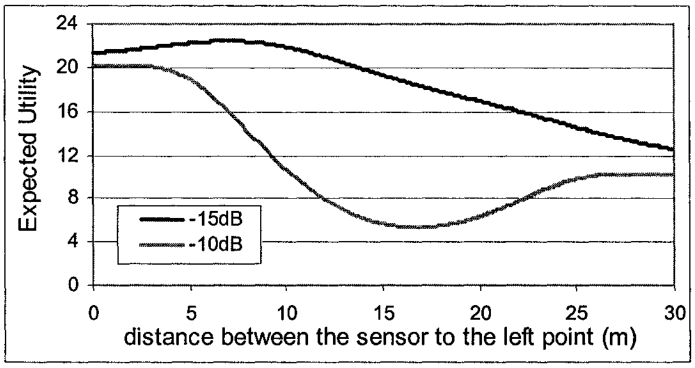
4.2 Problem Formulation

Before formulating the generic sensor deployment problem, we begin with a brief example to illustrate the impact of sensor position on the sensing quality. We assume that two points of interest are located 30 meters away from one another with information utilities[2] of 20 and 10, respectively. We also assume that one sensor is used to monitor both points. We assume that the signal-strength-based sensing model is used (see Section 3.2.2). Figure 4.1 (A) shows the network structure. Figure 4.1 (B) illustrates the obtained EIU by placing the sensor at different positions along the line between these two points, when the received signal strength threshold γ, is set at -10dB and -15dB respectively. The example clearly indicates that the position of the sensor does affect the obtained EIU significantly.

[2] Information utility here has a generic definition, which reflects the importance of information at the point of interest and is application determined.



(A) Positions of Points of Interest and the SN



(B) Expected Utility vs. Position of the Sensor at different. γ.

Figure 4.1 Illustration of the Impact of Sensor Position on Expected Information Utility

Next, we will formulate a generic information oriented SN deployment problem. For convenience, we first list the notations to be used in Table 4-1.

Table 4-1 Notation for Information Oriented SN Deployment Formulation

Notation	Description
N	Total number of points of interest
$O = \{o_1, o_2, \ldots, o_N\}$	A set of N points of interest
$o_i = (x_i, y_i)$	An element of set O, which is one point of interest with coordinates (x_i, y_i) on the two-dimensional field
iu_i	Value of information utility at point of interest o_i
n_s	Total number of sensors to be deployed
$S = \{s_1, s_2, \ldots, s_{n_s}\}$	A set of n_s sensors
$s_j = \{x_{s_j}, y_{s_j}\}$	An element of set S, which is a sensor located at (x_{s_j}, y_{s_j}) on the two-dimensional field
$n_s^{(\min)}$	Minimum value of n_s
$p_i^{(k)}$	Joint sensing probability provided by a total of k sensors to the point of interest o_i
$eiu^{(k)}$	Expected information utility obtained by using k sensors
EIU_{th}	Threshold of expected information utility

The general sensor deployment problem for information-oriented sensing coverage can be described as follows:

Given:

- A total of N points of interest on the two-dimensional sensing field;
- The value of *information utility* at each point of interest, iu_i;
- The sensing probabilities $p_i^{(k)}$ (e.g., as described in Section 3.2.2);
- The threshold of sensing quality, EIU_{th}.

Find:

The minimum number of sensors, $n_s^{(\min)}$, as well as their optimal positions, such that the obtained EIU can reach a pre-fixed threshold EIU_{th}.

To sum up, the formulation of the above problem is shown in Figure 4.2.

$$
\begin{array}{ll}
\text{min.} & n_s \\
s.t. & n_s > 0,\ n_s \text{ is a positive integer} \\
 & eiu^{(n_s)} = \sum_{i=1}^{N} p_i^{(n_s)} \cdot iu_i \geq EIU_{th} \\
 & \min(x_1,\ldots,x_N) \leq x_{S_j} \leq \max(x_1,\ldots,x_N), j = 1,2,\ldots,n_s \\
 & \min(y_1,\ldots,y_N) \leq y_{S_j} \leq \max(y_1,\ldots,y_N), j = 1,2,\ldots,n_s
\end{array}
$$

Figure 4.2 Formulation of Sensor Deployment for Information-Oriented Sensing Coverage

Due to the computational complexity of the formulated problem, it is not trivial to find a globally optimal solution using existing methods or a software package within reasonable time. Therefore, we propose a couple of heuristic schemes to provide locally optimal solutions.

4.3 Sensor Deployment Schemes

In order to find the minimum number of sensors as well as their optimal positions that satisfy all constraints, we propose a GREEDY strategy and a RECURSIVE strategy.

The principle of these two strategies is as follows. We begin by using one sensor. Then, we increment the number of sensors by one at each stage. In each stage, we solve the complementary problem of the primal one described in Section 4.2. That is, we aim at achieving maximum *EIU* by determining the optimal positions of a given number of SNs.

The procedure terminates when the obtained *EIU* reaches or exceeds the demanded threshold. In addition, based on how to determine the positions of sensors in each stage, we have a GREEDY algorithm and a RECURSIVE algorithm. Figure 4.3 shows the flow chart of the procedure.

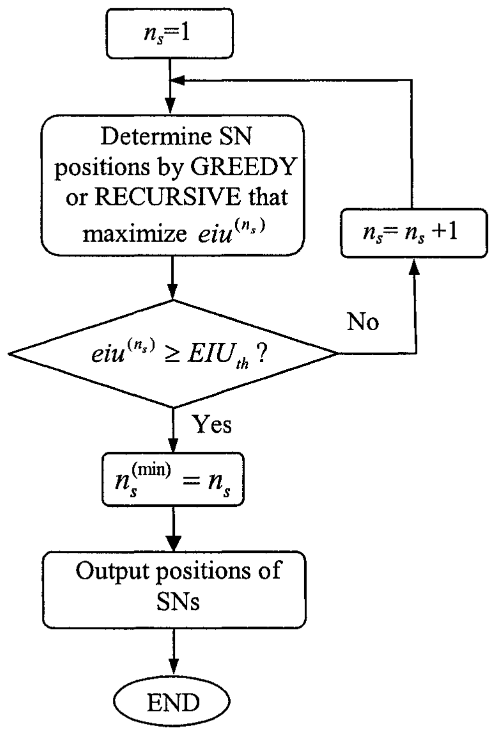


Figure 4.3 Flow Chart of Sensor Deployment Procedure

4.3.1 The GREEDY Algorithm

The basic idea of the GREEDY algorithm (called GREEDY hereafter) is to split the overall optimization problem into stages. At each stage, one SN is added. Based on the positions of the SNs added in previous stages, the newly added one will maximize the increment of obtained *EIU*. Let $eiu^{(m)}$ denote the *EIU* at the end of the m^{th} stage. i.e., $eiu^{(m)} = \sum_{i=1}^{N} p_i^{(m)} \cdot iu_i$.

Let $\Delta^{(m)}$ denote the increment of the *EIU* from the *(m-1)*st stage to the m^{th} stage, i.e., $\Delta^{(m)} = eiu^{(m)} - eiu^{(m-1)}$, m =1, 2, ... (where $eiu^{(0)} = 0$). In each stage m (m = 1, 2, ...), GREEDY aims to maximize $\Delta^{(m)}$.

The full procedure of GREEDY is as follows:

In stage 1: Determine the position of the first sensor.

$$\Delta^{(1)} = eiu^{(1)} = \sum_{i=1}^{N} p_i^{(1)} \cdot iu_i = \sum_{i=1}^{N} p_{s_1 \to o_i} \cdot iu_i \tag{4-1}$$

Assuming all sensors are identical and plugging Eqs. (3-3) and (3-4) in Eq. (4-1), we have:

$$\Delta^{(1)} = \sum_{i=1}^{N} iu_i \cdot Q\left\{\frac{\gamma - \overline{p_r}(d_0) - 10n\log d_0 + 10n\log d_{i,s_1}}{\sigma}\right\} \tag{4-2}$$

where d_{i,s_1} is the Euclidian distance between sensor s_1 (x_{s_1}, y_{s_1}) and the i^{th} point of interest.

Letting $a = \dfrac{\gamma - \overline{p_r}(d_0) - 10n\log d_0}{\sigma}$ and $b = \dfrac{10n}{\sigma}$, we rewrite Eq.(4-2) as:

$$\Delta^{(1)} = \sum_{i=1}^{N} iu_i \cdot Q(a + b \cdot \log d_{i,s_1}) \tag{4-3}$$

Using the definition of the Q-function as shown in Eq. (3-5), we have:

$$\begin{aligned}\Delta^{(1)} &= \sum_{i=1}^{N} iu_i \cdot \frac{1}{2}\left[1 - erf\left(\frac{a + b\log d_{i,s_1}}{\sqrt{2}}\right)\right] \\ &= \frac{1}{2}\sum_{i=1}^{N} iu_i - \frac{1}{2}\sum_{i=1}^{N} iu_i \cdot erf\left(\frac{a + b\log d_{i,s_1}}{\sqrt{2}}\right)\end{aligned} \tag{4-4}$$

In order that $\Delta^{(1)}$ is maximized, we need to find the position of sensor $s_1(x_{s_1}, y_{s_1})$, such that

$e^{(1)} = \sum_{i=1}^{N} iu_i \cdot erf(\frac{a + b\log d_{i,s_1}}{\sqrt{2}})$ is minimized.

Since $e^{(1)}$ is a continuous function with respect to (x_{s_1}, y_{s_1}), to obtain the numerically optimal results of (x_{s_1}, y_{s_1}), we need to find those coordinates that satisfy:

$$\begin{cases} \dfrac{\partial e^{(1)}}{\partial x_{s_1}} = 0 \\ \dfrac{\partial e^{(1)}}{\partial y_{s_1}} = 0 \end{cases} \tag{4-5}$$

Using the chain rule for partial derivatives, we have:

$$\begin{cases} \dfrac{\partial e^{(1)}}{\partial x_{s_1}} = \dfrac{\sqrt{2} \cdot b}{\sqrt{\pi} \cdot \ln 10} \cdot \sum_{i=1}^{N} iu_i \cdot \exp\left\{ -\dfrac{(a + b\log_{10} d_{i,s_1})^2}{2} \right\} \cdot (d_{i,s_1})^{-3/2} \cdot x_{s_1} = 0 \\ \dfrac{\partial e^{(1)}}{\partial y_{s_1}} = \dfrac{\sqrt{2} \cdot b}{\sqrt{\pi} \cdot \ln 10} \cdot \sum_{i=1}^{N} iu_i \cdot \exp\left\{ -\dfrac{(a + b\log_{10} d_{i,s_1})^2}{2} \right\} \cdot (d_{i,s_1})^{-3/2} \cdot y_{s_1} = 0 \end{cases} \tag{4-6}$$

where $d_{i,s_1} = \sqrt{(x_{s_1} - x_i)^2 + (y_{s_1} - y_i)^2}$

Among these candidate coordinates, the one that leads to the minimum $e^{(1)}$ should be found.

It is not trivial to obtain an optimal numerical solution for the above problem. Alternatively, optimal results can be obtained by exhaustive enumeration. However, the solution space is infinitely large when searching exhaustively on the continuous two-dimensional space. As such, we partition the entire sensing space into a grid, and search for a solution on the grid vertices. It is obvious that the finer the grid, the more accurate the result. As the tile edge length of the grid approaches zero, the result obtained from searching on the grid vertices approaches the optimal one. Moreover, when the grid is fine enough, the search results will not rely on the shape of the grid. For convenience, we use a grid with vertices on a square in our experiments, which will be discussed later in this chapter.

In stage m (m>1): Determine the position of the m^{th} sensor based on deployment of the first (m-1) sensors obtained in stages 1 to (m-1), such that the incremental *EIU* is maximized, i.e.,

$$\text{Max. } \Delta^{(m)} = eiu^{(m)} - eiu^{(m-1)} = \sum_{i=1}^{N} iu_i \cdot \left(\prod_{j=1}^{m-1} q_{s_j \to o_i} \right) \cdot p_{s_m \to o_i}$$

$$= \frac{1}{2} \sum_{i=1}^{N} iu_i \cdot \left(\prod_{j=1}^{m-1} q_{s_j \to o_i} \right) \cdot \left[1 - erf\left(\frac{a + b \log d_{i,s_m}}{\sqrt{2}} \right) \right] \tag{4-7}$$

where, $q_{s_j \to o_i} = 1 - p_{s_j \to o_i}$, j=*1, 2, ..., m-1*

Similar to stage 1, the optimal value of (x_{S_m}, y_{S_m}) can be approximated by exhaustive search on a grid.

4.3.2 The RECURSIVE Algorithm

Although GREEDY aims at finding the optimal positions of the SNs, the position of a newly added SN depends on those added previously. Therefore, GREEDY can only provide a local optimal solution. To improve over GREEDY, we present another heuristic algorithm, namely a RECURSIVE algorithm (called RECURSIVE hereafter). In every stage, RECURSIVE will operate in cycles. In each cycle, starting from an initial deployment of SNs, we attempt to reposition each and every SN step by step. In each step, one SN is repositioned to a new place if such a repositioning can lead to an increase in the EIU. The RECURSIVE procedure will terminate until the movement of any sensor in a cycle cannot lead to an increase in EIU. The pseudo-code for RECURSIVE in stage k is illustrated in Figure 4.4.

Initialization:

Deploy k sensors initially by some strategy, such as random deployment, max-valuable - object- first (MVOF), or using the greedy algorithm presented previously;
Set increasedEIU to a positive value;
Calculate EIU based on current sensor deployment:

$$eiu^{(k)} = \sum_{i=1}^{N} iu_i \cdot (1 - \prod_{j=1}^{k} q_{s_j \to o_i});$$

Reposition the sensors iteratively:

```
while (increasedEIU > 0)
    set increasedEIU = 0;
    for m = 1: 1: k
        Calculate new position for sensor s_m :
            Determine the new position of s_m , using the same method as described in
            stage k of GREEDY. Specifically, assume the other k-1 SNs have been
            deployed, and s_m is the kth one to be deployed.
        Calculate eiu_new^(k) , the value of EIU based on new position of s_m^new .
        if ( eiu_new^(k) > eiu^(k) )
            reposition s_m to the new position s_m^new ;
            update increasedEIU = increasedEIU + eiu_new^(k) - eiu^(k) ;
            update eiu^(k) = eiu_new^(k) ;
        end if
    end for
end while
```

Figure 4.4 Pseudo-Code of the RECURSIVE Algorithm

4.4 Experiments and Results

To gain insight into the behavior of the proposed algorithm, we evaluate its performance under various scenarios. As the number of potential design scenarios is immense, the algorithm is executed on randomly generated networks.

The network is on a square field of dimension 200 × 200 m^2. A total of 20 points of interest are distributed randomly in the field. The information utilities at the points of interest are independent and identically distributed random variables from a uniform distribution on [min_utility, max_utiltiy]. In each scenario, a total of k sensors are deployed. In each group of experiments, we generate 30 independent random networks. The results presented in this thesis are the averages over 30 runs. The simulation program was written in MATLAB.

We use the expected information utility as the evaluation metric. In all experiments, the results are compared with a naive information-oriented sensor deployment strategy, Max-Valuable-Point-First (MVPF) deployment, by which sensors are located at the k points of interest with the highest utility values, and with random (RAND) deployment, by which sensors are spread uniformly randomly in the field. All results are normalized with respect to the total information utility on the entire system.

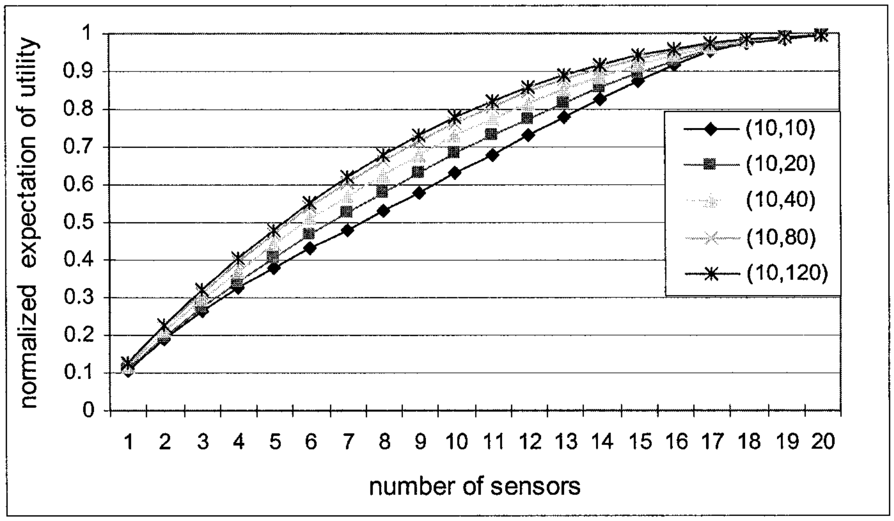
We conduct three groups of experiments under different scenarios. In the first group, we vary the information utility within different ranges. In the second group, we vary the sensing capability of sensors by adjusting the received signal threshold γ. In the third group, we vary total number of points of interest. The parameters used in our experiments are listed in Table 4-2. If not mentioned specifically, the information utility range in the experiments is (10, 20), $\gamma = -10dB$, and the tile edge length of the grid is 2. Experimental results are shown in Figures 4.5 – 4.8.

Table 4-2 Experimental Parameters

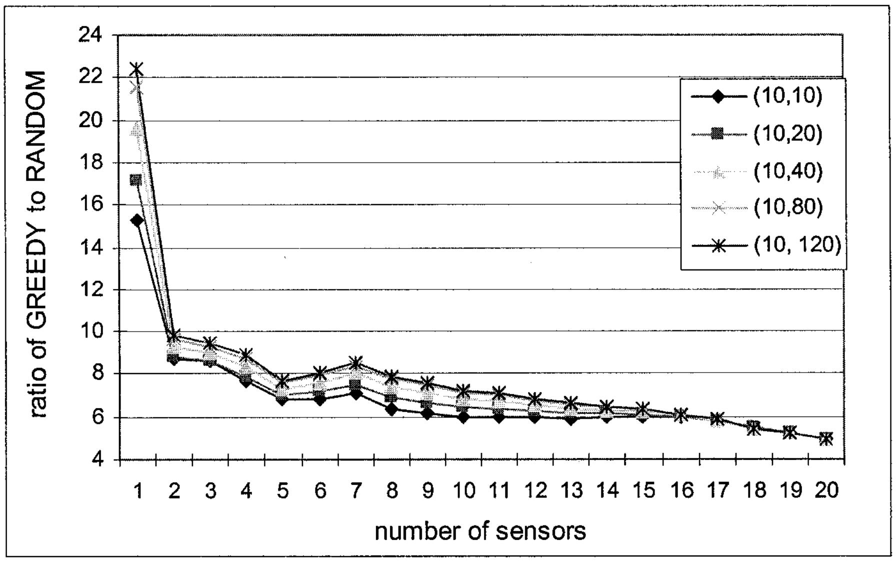
Parameters	Value
N	20, 40, 80, 160
γ	-5dB, -10dB, -15dB, -20dB, -25dB
$\overline{P}_r(d_0)$	10dB
d_0	1 meter
σ	4dB
n	2
k	1-N
utility range	(10, 10), (10, 20), (10, 40), (10, 80) (10, 120)
Tile edge length	1,2, 8,16

Experimental results indicate that the proposed GREEDY algorithm and RECURSIVE algorithm outperform MVPF and RAND in all scenarios. Regardless of other parameters, the more sensors that are deployed, the higher the expected information utility. When the number of sensors equals the number of points of interest, the expected utility is 1, which implies that the detection probability at any point of interest is approaching 1. Moreover, the more SNs deployed, the smaller the difference between the proposed algorithms and the simple ones. In all scenarios, the ratio of the proposed algorithms to RAND is much higher than that to MVPF. This implies that the planned deployment is better than the unplanned one, and the more carefully the deployment is planned, the better the results.

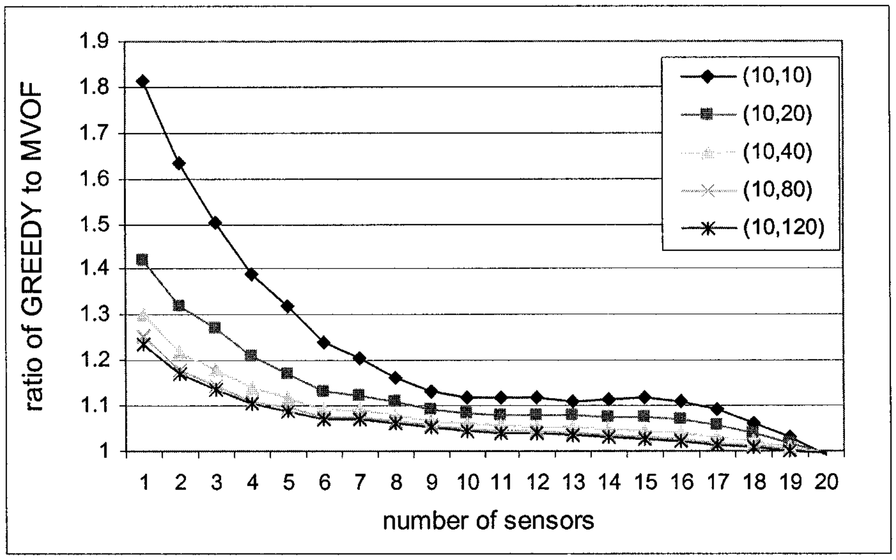
From Figures 4.5 (B) and (C), we find that the smaller the information utility range, the smaller the difference between GREEDY and RAND, and the higher the difference between GREEDY and MVPF. A similar trend applies to the RECURSIVE algorithm. This is because when deploying SNs, MVPF only consider the utility of each point of interest, but ignores the sensing probabilities. Because MVPF always gives higher priority to the points of interest with higher information utility, when the information utility varies in a large range, those points with high information utility dominate the expected utility. On the other hand, when using RAND, the placement of SNs ignores the differentiation of the information utility. So when the utility range is high, the impact of unplanned deployment becomes more significant. When deploying the sensors, the proposed GREEDY and RECURSIVE algorithms take both the differentiation of information utility and the geographic distribution of the points of interest into consideration.



(A) Expected Utility vs. Utility Range for GREEDY



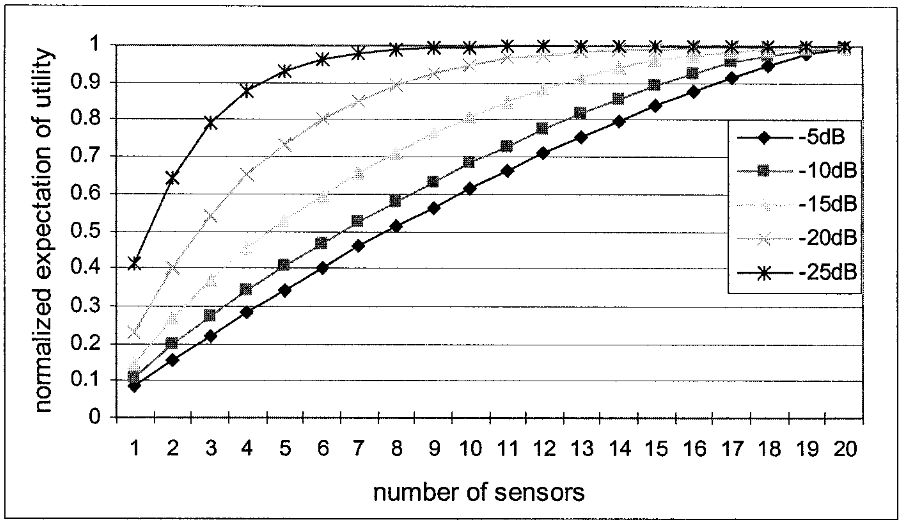
(B) Comparison of GREEDY and RAND



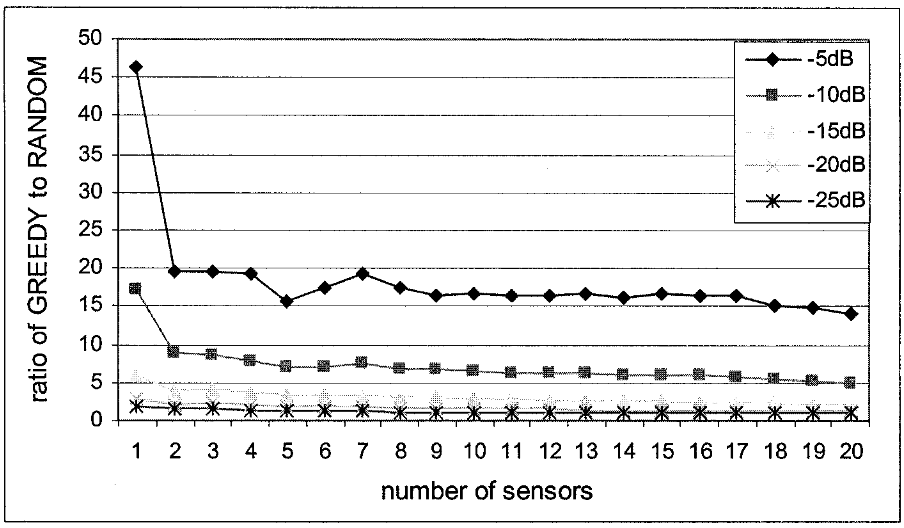
(C) Comparison of GREEDY and MVPF

Figure 4.5 Performance of GREEDY vs. Utility Range

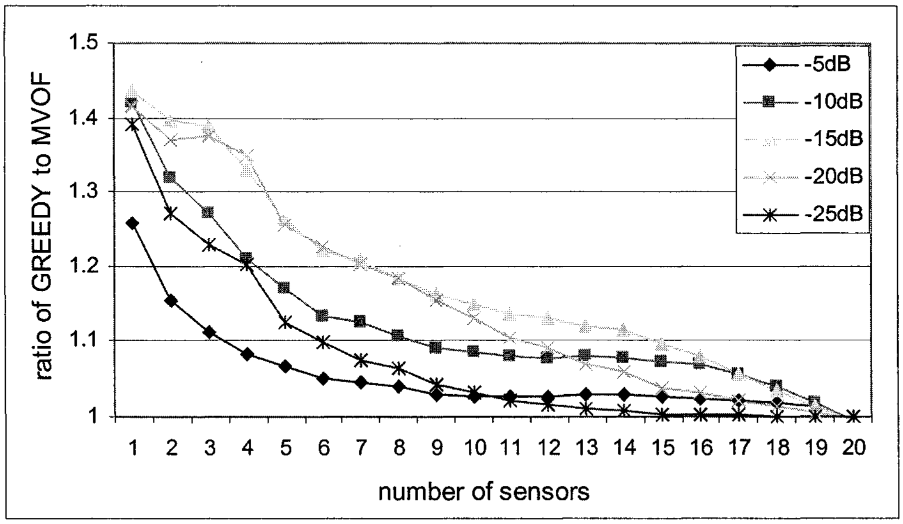
From Figure 4.6 (B), we observe that the difference between GREEDY and RAND decreases when the sensing capability improves. This is because SNs work collaboratively to provide a joint sensing capability for a specific point of interest. When the sensing capability is modest, the SNs should be placed more carefully, so that they can sit at more appropriate positions to collaborate with other SNs. From Figure 4.6 (C), we find that the difference between GREEDY and MVPF increases when the sensing capability increases ($\gamma = -5dB$ to $\gamma = -15dB$), while the difference will decrease if the sensing capability increases further ($\gamma = -15dB$ to $\gamma = -25dB$).



(A) Expected Utility vs. Sensing Capacity for GREEDY



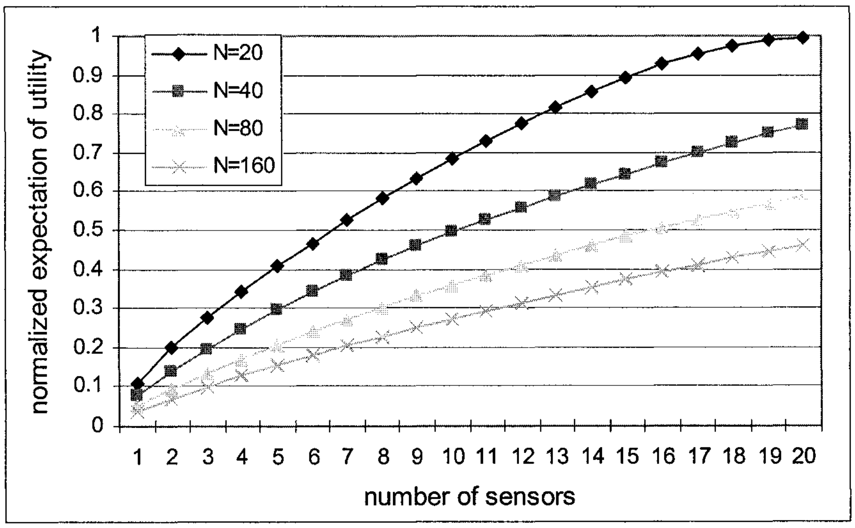
(B) Comparison of GREEDY and RAND



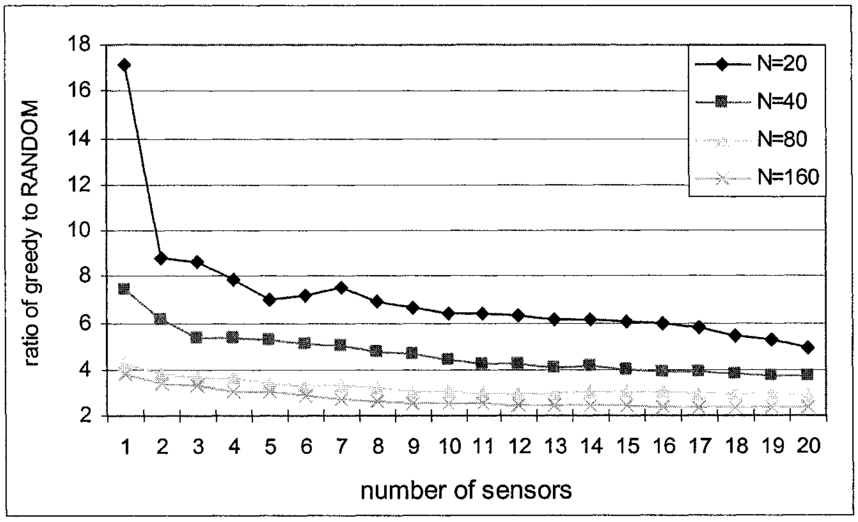
(C) Comparison GREEDY and MVPF

Figure 4.6 Performance of GREEDY vs. Sensing Capacity

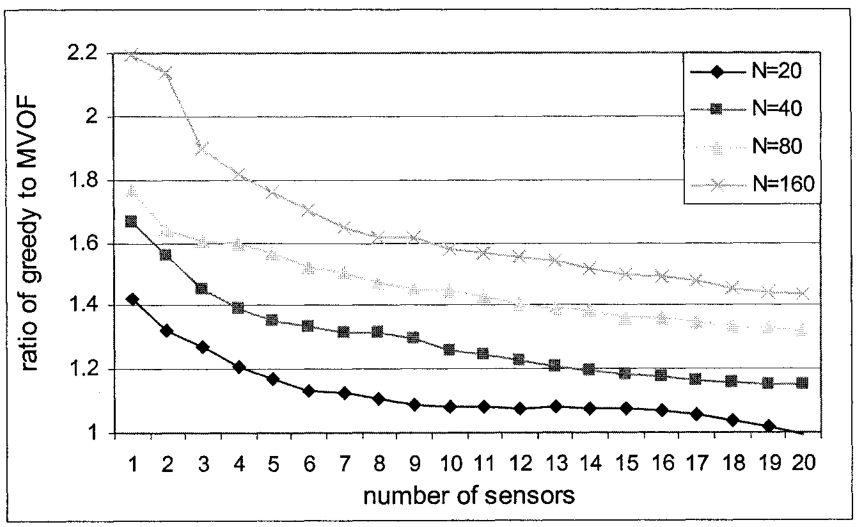
From Figure 4.7 (B) and (C) we observe that the higher the number of points of interest, the smaller the difference between GREEDY and RAND, while the greater the difference between GREEDY and MVPF.



(A) Expected Utility vs. Number of Points of Interest for GREEDY



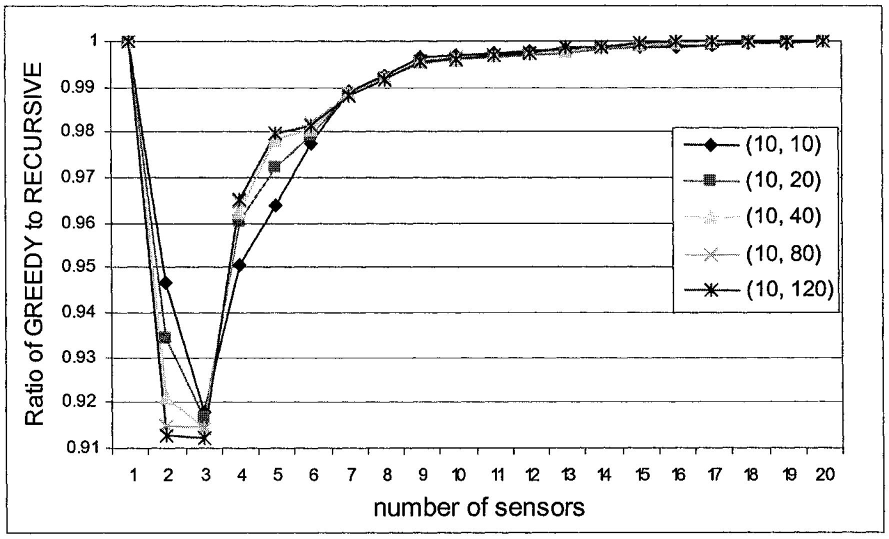
(B) Comparison of GREEDY and RAND



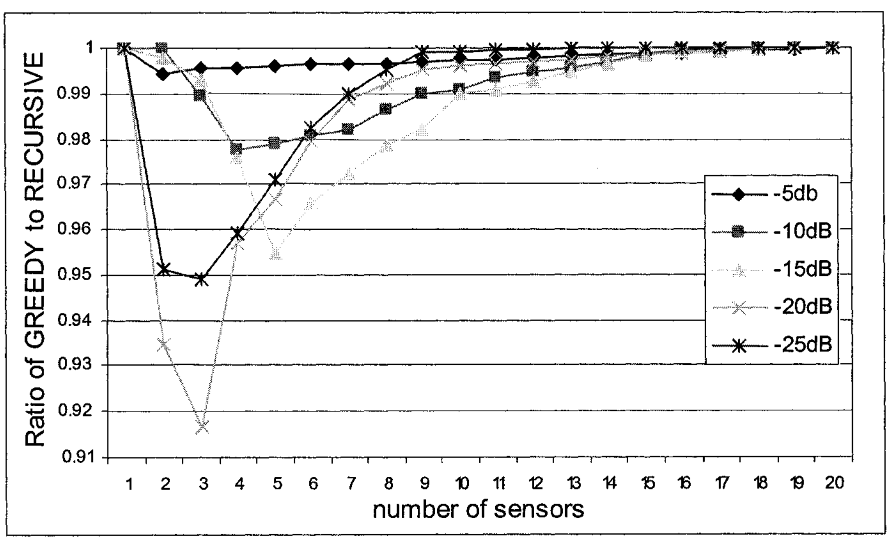
(C) Comparison of GREEDY and MVPF

Figure 4.7 Performance of GREEDY vs. Number of Points of Interest

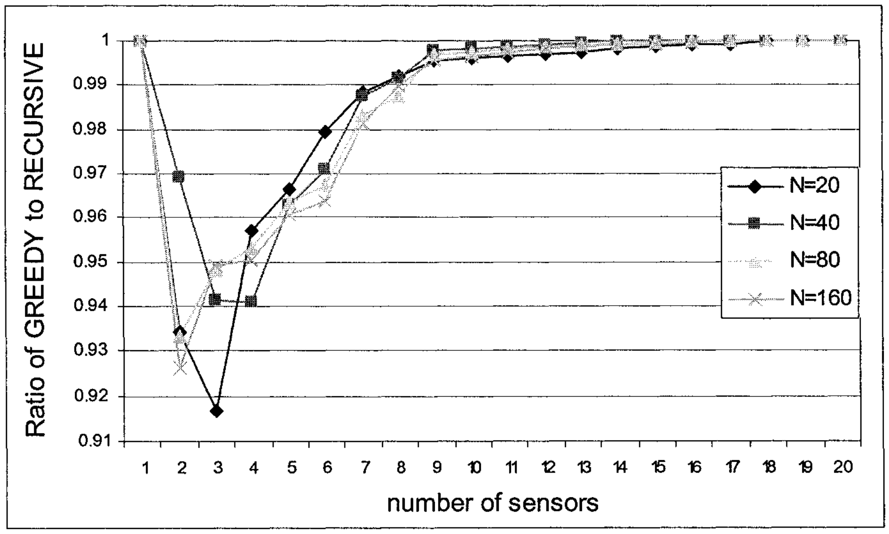
Figure 4.8 (A)–(C) illustrates the comparison using GREEDY and RECURSIVE. We observe that RECURSIVE provides equal or better SN positions than GREEDY. Thus the minimum number of SNs by using RECURSIVE is no larger than using GREEDY. However, RECURSIVE has much higher computational complexity than GREEDY. When a large number of SNs are needed, or the difference between GREEDY and RECURSIVE is within a moderate range, using GREEDY is more appropriate.



(A) Performance of Proposed Algorithms vs. Utility Range ($\gamma = -20dB$)



(B) Performance of Proposed Algorithms vs. Sensing Capacity



(C) Performance of Proposed Algorithms vs. Number of Points of Interest ($\gamma = -20dB$)

Figure 4.8 Comparison of GREEDY and RECURSIVE

Furthermore, we examine the impact of grid length on the accuracy of the result. As indicated previously, when using the proposed algorithm, a loss of accuracy may occur due to searching on the discrete grid vertices instead of on the continuous space. Figure 4.9 illustrates the results of running GREEDY using different grid lengths. Not surprisingly, the performance degrades with an increase in grid length. However, it also indicates that the degradation is less than 2% when the grid length is no more than 4, and it is about 6% when the grid length is 8. Since the complexity of the grid search is inversely proportional to the square of the grid length, it is promising to run the GREEDY algorithm at the larger grid length while keep the performance at a good level.

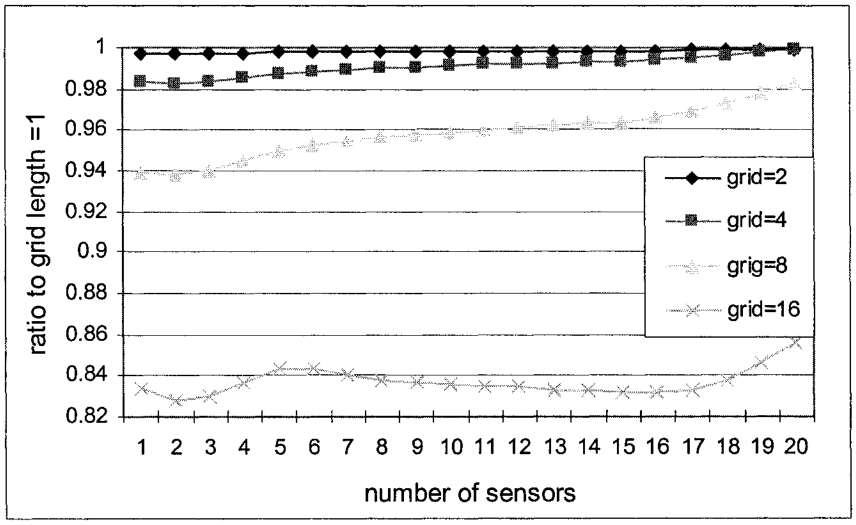


Figure 4.9 Performance of GREEDY vs. Tile Edge Length of the Grid

4.5 Summary

In this chapter, we explore the information-oriented sensor deployment problem in wireless sensor networks. Based on the observation of the differentiated significance of the information over the sensing field, we present a sensor deployment methodology to minimize the required number of sensors for a given expected information utility threshold. To the best of our knowledge, the problem of sensor deployment for information-oriented sensing coverage is first identified in this thesis. The work in this thesis paves a new direction for sensor deployment.

Due to computational complexity, a global optimal solution is not available now. We propose a couple of algorithms which provide locally optimal solutions and outperform the naive information-oriented sensor deployment strategy, Max-Valuable-Point-First (MVPF) deployment, and random deployment. Moreover, the computational complexity of the proposed GREEDY algorithm is $O(n_s \cdot N_{grid} \cdot \log(N_{grid}))$, and the complexity of the proposed RECURSIVE algorithm is bounded by $O\left((n_s \cdot N_{grid} \cdot \log(N_{grid}))^m\right)$, where n_s is the number of deployed sensors, N_{grid} is the total number of vertices of the grid for enumeration search, and m is a finite integer, which is the total number of cycles of sensor re-positioning. Although both are polynomial-time algorithms, GREEDY is computationally less expensive.

5. DEVICE PROVISIONING IN THE COMMUNICATION DOMAIN: PHASE ONE

5.1 Background

In Chapter 4, we discuss the problem of sensing domain design, in particular, the sensor deployment to achieve required coverage. However, due to the capability limitation, SNs can only transmit over a small distance. For the sake of energy conservation, the SNs do not relay traffic for other nodes. In order to ensure that a WSN is fully connected, i.e., all nodes can deliver their data to the BS via a workable route, we need to add RNs, which are responsible for collecting data from the SNs and establishing the data delivery routes to the BS. This will be the focus of the communication domain design.

A generalized deterministic RN placement problem is described as follows.

> *Given a specific SN deployment and properties of SNs, determine the number and placements of RNs, so that the total network cost is minimized while the constraints of lifetime and connectivity are satisfied.*

As discussed in Chapter 3, the generalized design problem can be refined into different design scenarios according to different device properties. One of the constraints is the energy supply of a RN, which can be either unlimited or limited. When a RN has an unconstrained energy supply (rechargeable or simply ample enough relative to the projected lifetime of the SNs), the placement of RNs must provide connectivity to each SN given the constraint of the limited communication range of the SNs. When the energy supply of RNs is limited, the placement of RNs should not only guarantee the connectivity of the SNs, but also ensure that the paths from RNs to the BS are established without violating the energy limitation. In [142], a two-phase RN placement procedure is proposed. The placement of the first phase RNs (FPRNs) provide connectivity support to the SNs. The second phase RNs (SPRNs) are further added to establish the routes from FPRNs to the BS.

In this chapter, we discuss the placement of FPRNs, which are dedicated to providing connectivity to the SNs under lifetime constraints. The connectivity constraints for RNs will be considered in Chapter 6.

We begin with the scenario that RNs have no energy constraints. We show that the optimization design is equivalent to the *minimum set covering problem* algorithm [143]. We propose a dynamic programming algorithm to solve the minimum set cover problem. Next, we extend the problem to the scenario in which RNs have limited energy supply. We provide an enhanced minimum set covering model to solve it. To the best of our knowledge, this is the first effort to use the minimum set covering model in the design of WSNs.

5.2 Placement of FPRNs without Energy Constraints

Recall SCENARIO ONE defined in Chapter 3, where the RN has no energy constraints. We also assume that they can adjust their transmission power adaptively. Thus, the lifetime and connectivity constraints of RNs are relaxed. Furthermore, we assume that RNs have identical cost. In such cases, the FPRN placement problem can be refined as:

Given a specific SN deployment and properties of SNs, determine the minimum number of RNs and their optimal positions, such that the constraints of lifetime and connectivity of SNs are satisfied.

Define $S = \{s_1, s_2, ..., s_{Ns}\}$, the set of given SNs at known positions in a two dimensional space, (e.g., the result obtained from Chapter 4 or obtained from other sensor deployment strategies). As stated in Chapter 3, the lifetime constraint of an SN, s_i, can be transferred to its confined transmission range, r_i^{SN}. Moreover, let $V = \{v_1, v_2, ..., v_m\}$ denote the set of RNs to be placed; let d_{iu}^{sr} denote the Euclidean distance from a SN s_i to a RN v_u. Then, the above RN placement problem becomes:

Given the set of SNs, S, find the minimum value of m, $m^{(\min)}$, and the optimal positions of RNs, such that for any $s_i \in S$, there exists at least one RN, v_u, and $d_{iu}^{sr} \leq r_i^{SN}$.

5.2.1. Modeling of Minimum Set Covering Problem

Considering that a SN can be represented by a disc centered at the position of the SN with a radius equal to its confined transmission range, Figure 5.1 demonstrates a set of SNs and their restricted transmission range based on the lifetime constraint. As individual SNs can have different properties, such as the amount of energy supply they are carrying, the physical characteristics they are measuring, the sampling rate, the traffic volume they are generating, etc., they do not necessarily have the same transmission range. Thus, these SN discs do not necessarily have equal radii.

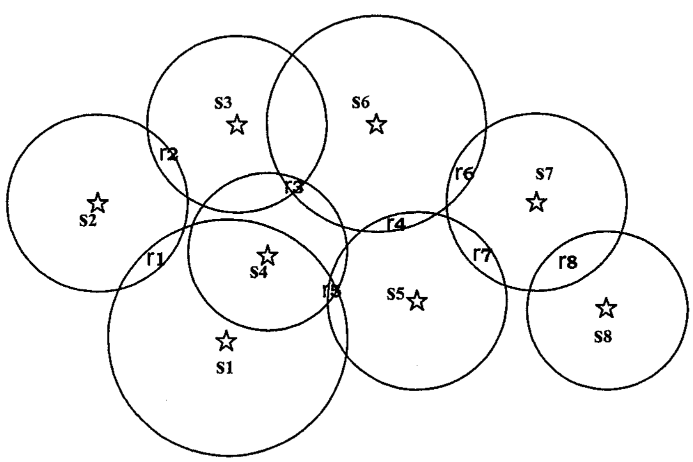


Figure 5.1 A Set of SNs and their Transmission Ranges

Since a SN cannot reach the BS directly, at least one RN has to be deployed within its transmission disc to relay its data to the BS. A trivial way is to deploy one RN for each SN.

However, if a RN is placed at a position where multiple discs overlap, the corresponding SNs will all benefit from this RN. Thus, the number of the RNs is decreased.

Given a set of *Ns* disks, denoted by $S = \{s_1, s_2, ..., s_{Ns}\}$, let a subset sr of S denote the region which is covered/overlapped by the corresponding subset of discs. For example, the subset $sr = \{s_1, s_2\}$ denotes the region where the disks of s_1 and s_2 overlap. A RN placed at sr_i can serve more SNs than a RN placed at sr_j if $sr_j \subset sr_i$, and vice versa.

A region sr_i is said to be a *densest region* if there is no region sr_j, satisfying $sr_i \subset sr_j$. Let r denote an individual densest region and R_O denote the set of all densest regions of the disc set S. In Figure 5.1, $R_O = \{r_1, r_2, ..., r_8\}$. For any subset sr_k of the set S, there must be an $r_l \in R_O$, so that $sr_k \subseteq r_l$. That is, if one RN is placed at a region sr_k, there must be a densest region r_l in which a RN will serve the same or more SNs. Therefore, when designing the optimal placement of RNs, the non-densest regions can be ignored. Hereafter, we can restrict our consideration to the densest regions as the candidate positions of RNs.

As such, the problem of finding the optimal placement of RNs for a set of SNs becomes equivalent to finding a *minimum set cover* for the disc set S using its corresponding set of densest regions R_O.

The generalized minimum set cover problem can be stated as follows:

Given a finite set X, let F be a family of subsets of X, such that for any $x \in X$, there is at least one $f \in F$, so that $x \in f$. A set cover is a subset $C \subseteq F$ whose members cover all of X, i.e., $X = \bigcup_{f \in C} f$. A set cover which has the smallest size is called a minimum set cover from F for X. X and F are the parameters of a minimum set cover problem.

Specifically, we can model the design problem illustrated in Figure 5.1 by a minimum set covering problem with the parameters summarized in Figure 5.2.

$X = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8\}$
$F = \{f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8\}$

$f_1 = \{s_1, s_2\}$	$f_2 = \{s_2, s_3\}$
$f_3 = \{s_3, s_4, s_6\}$	$f_4 = \{s_5, s_6\}$
$f_5 = \{s_1, s_4, s_5\}$	$f_6 = \{s_6, s_7\}$
$f_7 = \{s_5, s_7\}$	$f_8 = \{s_7, s_8\}$

Figure 5.2 The Minimum Set Cover Model Corresponding to the Scenario in Figure 5.1

5.2.2. An Optimization Algorithm to the Minimum Set Cover Problem

The minimum set covering problem is introduced in [143] and the decision version of the problem is NP-complete, i.e., it is unlikely we can find a polynomial-time optimization algorithm for it. In [143], a polynomial-time approximation algorithm, called GREEDY-SET-COVER is presented and it is proven to be a $(\ln|X|+1)$ approximation, where $|X|$ is the total number of elements to be covered. That is, the size of the set cover found by the GREEDY-SET-COVER is at most $(\ln|X|+1)$ times the size of the minimum set cover. The problem addressed here typically has a set of tens to hundreds of elements (densest regions).

Therefore, in the worst case, the size of the solution found using the approximation algorithm can be a few times the size of the optimal results, and hence not acceptable.

Next, we propose an optimization algorithm using the dynamic programming technique. The gist of the proposed algorithm is ***divide-and-conquer***. Specifically, by executing three major functions recursively, the algorithm splits the overall minimum set cover problem into a series of minimum set covering problems of smaller size (in terms of the number of uncovered elements and the candidate subsets) iteratively [144].

For example, Figure 5.3 demonstrates how the minimum set cover problem in Figure.5.2 can be split into minimum set cover problems of smaller size. In part (A), since the element s_8 is only covered by the subset f_8, f_8 has to be in the minimum set cover. Taking the subset f_8 as a component of the minimum set cover, the elements s_7 and s_8 are covered. Therefore, we only need to search a minimum set cover for the uncovered elements $\{s_1, s_2, s_3, s_4, s_5, s_6\}$ using F=$\{f_1, f_2, f_3, f_4, f_5\}$($f_6$ and f_7 are taken away from F by the compress function, which will be detailed later). Part (B) gives a graphic demonstration of this smaller problem. In part (C), every element is covered by more than one subset. For element s_1, it is covered by f_1 and f_5. At least one of them has to be in the minimum set cover. If we choose f_1 to cover s_1 (it also covers s_2) and we can find a minimum set cover for the other elements $\{s_3, s_4, s_5, s_6\}$ using $F=\{f_2, f_3, f_4, f_5\}$, the union of two parts are a candidate minimum set cover. If we choose f_5 to cover s_1 (it also covers s_4 and s_5) and find a minimum set cover for the other elements $\{s_2, s_3, s_6\}$ using $F=\{f_1, f_2, f_3, f_4\}$, the union of two parts is another candidate minimum set cover. By comparing all the candidate solutions, the one of smallest size is the final solution.

$X = \{s_1,s_2,s_3,s_4,s_5,s_6,s_7,s_8\}$
$F = \{f_1,f_2,f_3,f_4,f_5,f_6,f_7,f_8\}$

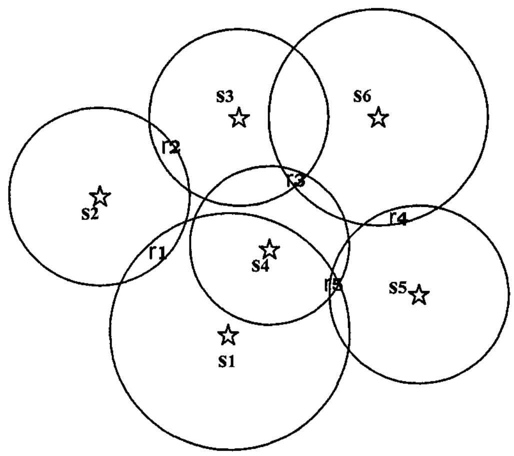
$f_1 = \{s_1,s_2\}$ $f_2 = \{s_2,s_3\}$
$f_3 = \{s_3,s_4,s_6\}$ $f_4 = \{s_5,s_6\}$
$f_5 = \{s_1,s_4,s_5\}$ $f_6 = \{s_6,s_7\}$
$f_7 = \{s_5,s_7\}$ $f_8 = \{s_7,s_8\}$

$\rightarrow$ f_8 covers s_7,s_8 +

$X = \{s_1,s_2,s_3,s_4,s_5,s_6\}$
$F = \{f_1,f_2,f_3,f_4,f_5\}$

$f_1 = \{s_1,s_2\}$ $f_2 = \{s_2,s_3\}$
$f_3 = \{s_3,s_4,s_6\}$ $f_4 = \{s_5,s_6\}$
$f_5 = \{s_1,s_4,s_5\}$

(A) Simplex Element Case



(B) A Graphical Representation of the Partial Minimum Set Cover Problem of (A)

$X = \{s_1,s_2,s_3,s_4,s_5,s_6\}$
$F = \{f_1,f_2,f_3,f_4,f_5\}$

$f_1 = \{s_1,s_2\}$ $f_2 = \{s_2,s_3\}$
$f_3 = \{s_3,s_4,s_6\}$ $f_4 = \{s_5,s_6\}$
$f_5 = \{s_1,s_4,s_5\}$

f_1 covers s_1,s_2 +

$X = \{s_3,s_4,s_5,s_6\}$
$F = \{f_2,f_3,f_4,f_5\}$

$f_2 = \{s_2,s_3\}$ $f_3 = \{s_3,s_4,s_6\}$
$f_4 = \{s_5,s_6\}$ $f_5 = \{s_1,s_4,s_5\}$

f_5 covers s_1,s_4,s_5 +

$X = \{s_2,s_3,s_6\}$
$F = \{f_1,f_2,f_3,f_4\}$

$f_1 = \{s_1,s_2\}$ $f_2 = \{s_2,s_3\}$
$f_3 = \{s_3,s_4,s_6\}$ $f_4 = \{s_5,s_6\}$

(C) Multip ase

Figure 5.3 The Divide-and-Conquer Strategy

For a clear description of the algorithm, the following concepts are defined.

Definition 1: In a set cover problem with parameters (X , F), let F_i denote the set of subset $f \in F$ which covers element i , i.e., $F_i = \{f_k : f_k \in F, i \in f_k\}$. For instance, in the example shown in Figure 5.1, $F_1 = \{f_1, f_5\}$, $F_2 = \{f_1, f_2\}$,..., $F_8 = \{f_8\}$. For the sake of convenience, in the algorithm introduced next, we use the index number of the subset to represent a subset, hence $F_i = \{k : f_k \in F, i \in f_k\}$. The degree of an element $i \in X$, denoted by $Deg(i)$, is the size of the set F_i . Furthermore, we call the element with degree one the simplex element; otherwise, it is a multiplex element.

Let G denote the minimum set cover (a group of indexes of the subsets in F), the pseudo code for the proposed minimum set cover algorithm is illustrated in Figure 5.4. As illustrated in Figure 5.4, three major functions have been recursively invoked in the algorithm.

***Decompose()**:* This function takes three parameters, X, F, i . X is the set of elements to be covered, F is a set of subsets of X , i is the index of the subset which has been chosen to be a component of the set cover. The function generates the minimum cover for X from F .

The function first adds the index i into the solution set G. As such, the elements in $F[i]$ are covered. We can restrict the cover search for the remaining uncovered elements using unused subsets. The *compress()* function (to be detailed later) is invoked to clear out the covered elements in X and F , respectively. After the compression, the function recursively finds the simplex element and its associated subset (as discussed in the divide-and-conquer strategy), adds the subset index to the solution set, and compresses the parameters, until all elements are covered, or there exist only complex elements. In the latter case, the *multiplex_search()* function is invoked to find the minimum set cover for the residual uncovered elements.

multiplex_search()*:* This function takes two parameters X , F , whose meanings are the same as in the *decompose()* function.

For an element $j \in X$, one of its $Deg(j)$ associated subsets has to be in the solution. By arbitrarily picking a subset $f \in F_j$ to cover j , we can find a candidate solution for the minimum set cover problem by invoking *decompose(*X,F,i*)*, where i is the index of f . Comparing the results of multiple candidate solutions, the one of minimum size is the final solution.

compress()*:* This function takes three parameter X,F,i , and is invoked once after the decompose function is invoked. It gets rid of the elements and subsets that do not make an impact on the search for the minimum set cover of smaller size.

```
Minimum_Set_Cover( X,F ) {
   j =next_simplex_element( X,F );
   i =the index of f ∈ F which covers j ;
   if i ≠ 0
        G =decompose( X,F,i )
   else
        G =multiplex_search( X,F )
}
G =decompose( X,F,i ) {
   Do {
      G = G ∪ {i}
      [ X,F ]=Compress( X,F,i );
      j =next_simplex_element( X,F );
      i =the index of f ∈ F which covers j ;
   } while i ≠ 0
   If X ≠ φ  G = G ∪ multiplex_search(X,F);
   Return G ;
}
G =multiplex_search( X,F ) {
    Pick an element i ∈ X ;
    Find the set F_i of the subsets which covers i ;
    For k=1: | F_i | {
         CG[k] =decompose( X,F, F_i[k] );
    }
    // the index of the candidate solution of minimum size;
    l =min_candidate( CG );
    Return CG[l];
}
[ X,F ]=compress( X,F,i ) {
    X = X − F[i] ;
    For m=1: | F |  F[m] = F[m] − F[i] ;
    For m=1: | F | {
         for n=m+1: | F |+1 {
           if F[n] ⊆ F[m]  F[n] = φ ;
           if F[m] ⊂ F[n]  F[m] = φ ;
         }
     }
 }
```

Figure 5.4 An Optimal Minimum Set Cover Algorithm using the Divide-and-Conquer Strategy

Figure 5.5 gives an example of the *compress()* function. It assumes that f_1 is picked as a component of the minimum set cover. In the step one, the elements in f_1={1,2} are cleaned out from X and $f \in F$. In the second step, f_2 is reset to empty since f2 is included in f_3. In other words, f_3 can cover elements covered by f_2. The benefit of the *compress()* function is reconstructing the set cover problem in an algorithm-efficient manner. For example, element 3 becomes simplex (only covered by f_3) after the second step in Figure 5.5. Hence, the decomposition can continue on by taking f_3 as the next component of the minimum set cover.

Note that the *compress()* function does not change the index number of the subset in F; instead it cuts off the elements and subsets which do not further impact the search from X and F. The resultant set G can therefore unambiguously include the minimum set cover.

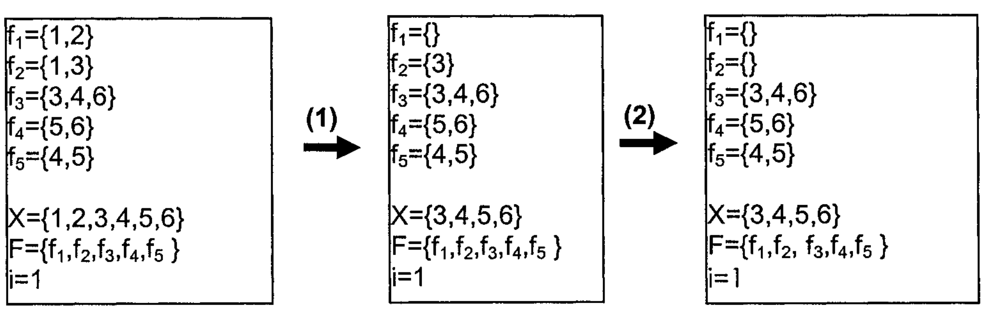


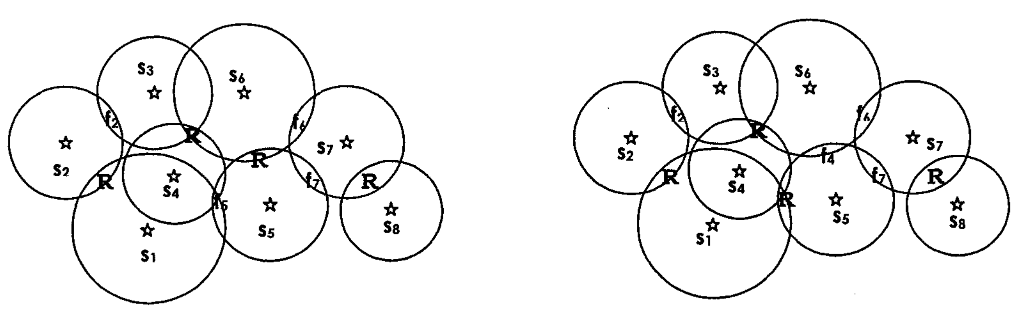
Figure 5.5 An Illustration of the *compress()* function

The main function, *minimum_set_cover()*, will first find a simplex element i in X, F, and invoke the decomposition thereafter. If no simplex element is found, the *exhaustive_search()*

is invoked. When X becomes empty in the end, G will include the indices of the subsets that comprise a minimum set cover.

The proposed algorithm uses the divide-and-conquer strategy to decompose a problem of large size into a series of problems smaller size. It is indeed an application of the dynamic programming technique, which has been proven to yield optimal solutions, while incurring less computation than exhaustive enumeration [143].

However, the complexity of the proposed algorithm in the worst case is $O\left(K^{|X|}\right)$, where $K = \max_{i \in X} Deg(i)$. For a complex problem, the GREEDY-SET-COVER algorithm described in [143] might be a useful tool which has only linear complexity and is proven to be a $(\ln|X|+1)$ approximation. For example, a solution to the problem in Figure 5.1 is obtained by the recursive algorithm, and is illustrated in Figure 5.6(A). The result from the GREEDY-SET-COVER algorithm is compared in Figure 5.6(B). In this case, they are both optimal.



R marks the RN positions

(A) Placement Solution by Optimal Algorithm **(B)** Placement Solution by Greedy Algorithm

Figure 5.6 An Example using both the Optimal and Greedy Algorithms

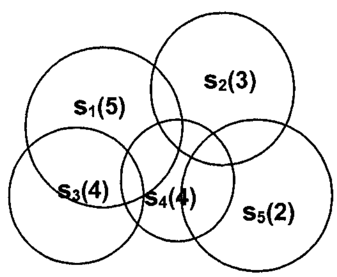
5.3 Placement of FPRNs with Energy Constraints

In the previous sections, we assume that the RNs have unlimited energy supply. Hence, each of them can serve all SNs that can reach it. In contrast, in some scenarios, a RN is only equipped with limited energy supply. In such cases, a RN can only handle a constrained amount of traffic. As a result, the minimum set covering model described in Section 5.2 should be modified.

Due to the power limitation on RNs, when a RN has a fixed transmission range, the amount of traffic that a RN can handle is limited, as shown in Eq.(3-11). As such, some densest regions are not energy feasible (see Figure 5.7). We enhance the criteria of candidate locations with an energy constraint. After finding the eligible candidate locations by applying the enhanced criteria, the minimum set covering model is applicable to the first phase RNs problem. The new candidate locations are therefore called *densest energy-feasible regions*.

Definition 2: *Energy-Feasible Region (EFR) and Densest Energy-Feasible Region* (*DEFR*). For a set of SNs, $S = \{s_1, s_2, ..., s_N\}$, a region is a subset, $R \subseteq S$, for which the intersection of the transmission discs of all $s_i \in R$ is nonempty. The term region will also be used to refer to the associated intersection area. Physically, if a RN is deployed in a region, it is possible that all associated SNs can transmit data to the RN in one hop. A region is energy feasible if a RN deployed in this region can relay all traffic from the associated SNs while meeting the lifetime constraint. Such a region is called an Energy-Feasible Region (EFR). An EFR R^* is a DEFR if there is no other EFR R, so that $R^* \subset R$.

The concepts of region, EFR, and DEFR are illustrated in Figure 5.7. As shown, the number of EFRs is finite for a finite SN set S and it is bounded by $O(|S|)^2$ [145]. By finding a minimum set cover for S among EFRs using the algorithm in Section 5.2, the first phase placement problem is solved.



A set of SNs={s_1, s_2, s_3, s_4, s_5} and their traffic, denoted by the numbers in brackets. Assume a RN can relay a maximum traffic volume of 10 in order to meet the lifetime constraint. Subsets, such as {s_1, s_3, s_4}, {s_1, s_2, s_4} are regions. Each region, such as {s_1, s_3, s_4}, has an associated intersection of the transmission disks. {s_1, s_3}, {s_1, s_4}, {s_3, s_4} are also regions. However, {s_1, s_3, s_4} is not energy feasible, as the sum of the traffic of its associated nodes is greater than 10. {s_1, s_3}, {s_1, s_4}, {s_3, s_4} are EFRs, as in each of them the total traffic of their component nodes does not exceed 10. However, the EFR {s_2, s_5} is not DEFR since it belongs to the EFR {s_2, s_4, s_5}.

Figure 5.7 An Illustration of Regions, EFRs, and DEFRs

Following the same line of logic as in Section 5.2, we argue that the size of a minimum set cover from EFRs is the same as the size of a minimum cover from DEFRs. Therefore, we constrain the candidate placement positions to the set of DEFRs.

The placement of FPRNs is locally optimal in the sense that the number of RNs required to connect SNs directly is minimized.

Furthermore, if a RN has a variable transmission power, we can tune it to the level that corresponds to the optimal transmission range [146], which leads to the best energy efficiency.[3] The enhanced minimum set cover model described above is still applicable.

[3] Details of the optimal transmission range will be discussed in Chapter 6.

5.4 Summary

In this chapter, we discuss the first phase of the design of the communication domain, which ensures the connectivity of SNs. Depending on whether the RNs are energy constrained or not, we formulate the refined problems as a minimum set covering problem and enhanced minimum set covering problem respectively. We also propose corresponding optimal algorithms to solve them. As far as we know, the work in this thesis is the first to use a minimum set covering model to formulate the deterministic RN placement problem.

Furthermore, in the case that RNs do not have energy constraints, the procedure of network planning completes after RNs are placed in this first phase. However, in the case that RNs do have energy constraints, either the transmission distance and/or the total amount of traffic that each RN can handle are limited as well. As a result, it could be infeasible to ensure that each RN can find a workable route to relay traffic to the BS. Then, extra RNs must be placed to guarantee the connectivity for each relay node. We define those RNs as second phase relay nodes (SPRNs). In Chapter 6, we will formulate the problem of SPRN placement in different scenarios, and will present the corresponding proposals to solve the problem.

6. DEVICE PROVISIONING IN THE COMMUNICATION DOMAIN: PHASE TWO

In Chapter 5, we discussed the RN placement in the first phase, by the end of which, every SN has found a RN to forward its traffic. However, when the RNs have limited initial energy supply, we need to place more RNs in the second phase, so that every RN will be able to find its neighbor(s) via which it relays traffic to the BS, with an eye to minimizing the number of second phase relay nodes (SPRNs).

Actually, the placement of SPRNs has two components: to determine the exact positions of SPRNs; and to decide the traffic allocation scheme, i.e., for each RN (either FPRN or SPRN), select its next hop neighbor and the portion of traffic volume to be forwarded to the BS through this particular neighbor. In general, the optimal solution to the combined problem of SPRN positioning and traffic allocation is not yet available. Therefore, we propose several heuristic schemes to solve it in different scenarios. In addition, in order to evaluate the performance of our proposed schemes, we derive the lower bounds of the number of SPRNs for comparison.

Moreover, as RNs may have either fixed or variable transmission range, the solutions to SPRN placement differ, too. In the following sections, we will present proposals corresponding to these scenarios.

6.1 Relay Nodes with Fixed Transmission Range

The network of RNs is abstracted as a directed graph *G(V, A)*. Initially, $V=\{v_0, v_1, \ldots, v_N\}$ is the set of FPRNs and A is the edge set. The vertex v_0 stands for the BS. Each RN has an associated weight which indicates the traffic load it has. Let w_i denote the traffic load of v_i, d_{ij} denote the Euclidean distance from RN v_i to RN v_j, and r_i denote the fixed transmission radius of RN v_i. There would be a directed edge from v_i to v_j if $d_{ij} \le r_i$ and $d_{j0} < d_{i0}$, i.e., if the transmission by v_i can reach v_j and v_j is closer to the BS than v_i. Also, let C_i denote the capacity of v_i, which can be calculated according to Eq. (3-11).

In the remainder of this section, we first identify two essential design principles. Next, we present three heuristic algorithms to implement the principles. In order to analyze the performance of the proposed algorithms, we derive a theoretical lower bound of the number of SPRNs. Then, the performance evaluation is conducted by simulation.

6.1.1 Far-Near and Max-Min Principles in the Placement of SPRNs

A. Far-Near Principle

This refers to the principle that the placement decision in the second phase should consider the RNs which are farthest from the BS and evolves step-by-step to the RNs that are closest to the BS. The rationale is that data are to be forwarded towards the BS. Hence RNs that are closer to the BS should relay the traffic for other farther nodes. This principle helps to avoid energy wastage incurred due to unnecessary detours in traffic relaying.

B. Max-Min Principle

This refers to the principle of maximally utilizing the capacity of existing RNs, while introducing a minimum number of new RNs. Specifically, from far to near to the BS, each node will distribute its workload to other existing neighboring nodes first. Only when the existing neighboring nodes of a given node cannot support the traffic load from it will a new RN be added.

6.1.2 Localized Heuristic Algorithms

Following the principles above, three heuristic algorithms are proposed, which are differentiated by how the traffic load of a RN is forwarded and distributed to neighboring RNs.

A. Nearest-To-BS-First (NTBF) Algorithm

This algorithm applies the above described Far-Near, Max-Min principles directly. Starting from the farthest RN, say RN_{far}, we calculate the total residual capacity of its adjacent neighbors. The capacity of each RN can be calculated as per Eq.(3-11), and the workload of each RN is the sum of its relayed traffic load. Thus the residual capacity at a given node is the difference between its capacity and its workload. If the total residual capacity is bigger than the workload of RN_{far}, then its workload is distributed to its adjacent neighbors; otherwise, a new RN will be introduced as its next hop relay. The algorithm is described in Figure 6.1.

Step 0: (Initialization) let $U = V - \{v_0\}$.

Step 1: Pick a node v_i, which is farthest from the BS among the nodes in the set U.

If $d_{i0} \le r_i$, Delete v_i from U, go to Step 3. Otherwise, go to Step 2.

Step 2: (Traffic Distribution with Nearest-To-BS-First)

Calculate the total residual capacity (RC) of the neighboring nodes of node v_i as follows:

$$RC = \sum_k (C_k - w_k),\ where\ d_{ik} \le r_i,\ and\ d_{k0} < d_{i0}.$$

If $RC > w_i$, i.e., the residual capacity is equal to or larger than the traffic load of node v_i, then v_i distributes its traffic load to the neighboring nodes, by first filling up the capacity of the node nearest to the BS, then to the node next nearest to the BS, and so on. If there are two or more neighboring nodes closest to BS, then the one with higher residual capacity will be chosen first. Delete v_i from U, go to Step 3.

Else a new RN should be deployed along the line of node v_i and the BS, at the point that is at distance r_i away from v_i. Add the new RN into U and delete v_i from U, go to Step 3.

Step 3: If $U = \varnothing$, EXIT. Otherwise, go to Step 1.

Figure 6.1 Nearest-To-BS-First Algorithm

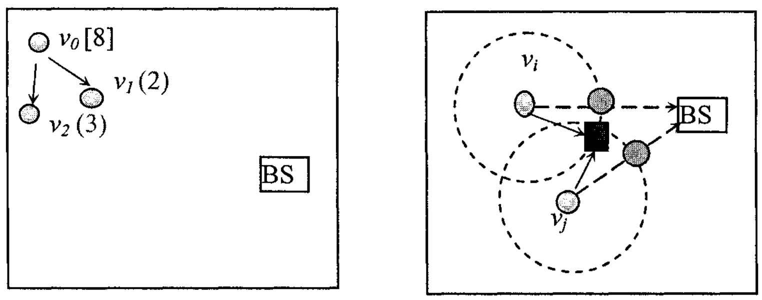
B. Max-Residual-Capacity-First (MRCF) Algorithm

We observe that by using the NTBF algorithm, the workloads among the nodes could become unbalanced, since the traffic distribution is sensitive to the distance between a node and the BS. A tiny difference in distance to the BS could lead to significant variance in the workloads of two nodes. Thus, some potential traffic path segments may become jammed, causing more new RNs to be needed to setup new paths. As such, we introduce the Max-Residual-Capacity-First (MRCF) algorithm to maintain better load balance among the RNs. The performance improvement by doing so will be considered in the next subsection.

The MRCF algorithm is similar to the NTBF algorithm, except for Step 2. When distributing the workload of a given RN to its neighbors, it first fills up the capacity of the neighboring RN with maximum residual capacity, then the neighboring RN with second to maximum residual capacity, and so on. If two or more neighboring nodes have the same highest residual capacity, then the one nearer to the BS is chosen first.

C. Best-Effort Relaying (BER) Algorithm

In the Nearest-To-BS-First algorithm, a new RN is added if the farthest RN's neighbors cannot relay its total traffic. However, the capacity of its neighbors is potentially wasted as its workload will be passed to the new RN without bothering the existing neighboring RNs. Therefore, a new FNMM based heuristic algorithm is proposed in this thesis. Specifically, before adding a new RN, we attempt to utilize the existing RNs to relay traffic in a best-effort manner. That is, the traffic relay will be arranged even if a RN's neighbors cannot serve it all. In addition, when placing a new RN, the location is picked to not only make the RN be as close to the BS as possible, but also to be strategically placed so as to serve as many existing RNs as possible. For example, in Figure 6.2 (A), v_1 and v_2 are assigned a portion of the traffic load from v_0, even though they cannot handle all of the traffic load of v_0. In Figure 6.2 (B), one SPRN is needed by the Best-Effort-Relaying algorithm, two are required by Nearest-To-BS-First algorithm.



[a] – traffic load

(b) – residual capacity

(A)

SPRN placed by NTBF

SPRN placed by BER

(B)

Figure 6.2 Illustration of Best-Effort-Relaying Algorithm

To describe the algorithm, we use the same notation defined at the beginning of this Section. In addition, we define the residual load of a node to refer to the difference between its total traffic load and the traffic load for which the next step RN has been assigned. Furthermore, we denote by U the set of RNs under consideration. Representing a RN v_i by a disc centered at v_i with radius r, let $INS(v_i)$ denote the intersecting neighbor set of a RN v_i, i.e., $INS(v_i) = \{v_j : d_{ij} \le 2r, v_j \in V\}$. Let $GIR(W)$ denote the geometrical intersecting region covered by transmission discs of all RNs in the RN set W . The Best-Effort-Relaying algorithm is described in Figure 6.3.

Step 0 – *Initialization:* let $U = V$;
Step 1 – For each RN $v_i \in U$, doing the following:
If v_i has zero residual capacity and can reach the BS in one hop, remove v_i from U.
Step 2 – From far to near, each $v_i \in U$ relays as much traffic load as possible to its neighboring RNs which are closer to the BS;
Update residual capacity and workload for the RN;
If v_i is fully loaded and has zero residual load, remove v_i from U.
Step 3 – Up to this step, we have a set of RNs which cannot distribute their workload under the existing RNs, therefore, new RNs are necessary to be added
Pick the farthest RN $v_i \in U$;
If v_i has zero residual traffic load, remove v_i from U and start over from Step 5.
Else if residual load of v_i equals to its capacity, go to Step 4;
Else
Find $INS(v_i)$
If ($INS(v_i) = \varnothing$), go to Step 4;
Else
set $W = \{v_i\}$;
pick the farthest RN $v_j \in INS(v_i)$, if $GIR(\{v_j\}) \cap GIR(W) \neq \varnothing$, and the total residual load of RNs in $\{v_j\} \cup W$ is no more than the capacity of a new node, then update $W = W \cup \{v_j\}$ and $GIR(W)$; else do nothing.
The operation proceeds until the total load is equal to the capacity of a new node, or all $v_j \in INS(v_i)$ have been checked. Place the new node in $GIR(W)$ as close to the BS as possible, relay the load from RNs in the set W to the new RN.
Update node residual capacities and workloads of RNs in U
Remove all nodes in W from U, add the new node to U
Go to Step 5;
Step 4 – Place a new RN along the line joining v_i to the BS, as close to the BS as possible, and assign the load from v_i to the new RN; remove v_i from U; add the new node to U.
Step 5 – repeat Steps 1, 2 & 3 until all nodes in U can reach the BS in one hop.

Figure 6.3 Best-Effort-Relaying Algorithm

6.1.3 Theoretical Lower Bound

In the following, we derive a lower bound on the number of SPRNs. The notation from Section 6.1.2 is used and we define $d_{\max} = \max_{v_j \in V} d_j$. Given a BS and a RN set V, draw M concentric circles of radii $1r$, $2r$, ..., Mr centered at the BS, where $M = \lceil d_{\max} / r \rceil$. Thus, all FPRNs are within the circles.

Define *shell m*, denoted by S_m, to be the region between the m^{th} and $(m+1)^{st}$ circles. A FPRN v_j is said to be *covered* by the m^{th} shell if $m*r < d_j \le (m+1)*r$, where ($m = 0, 1, ... M-1$). The definitions are illustrated in Figure 6.4.

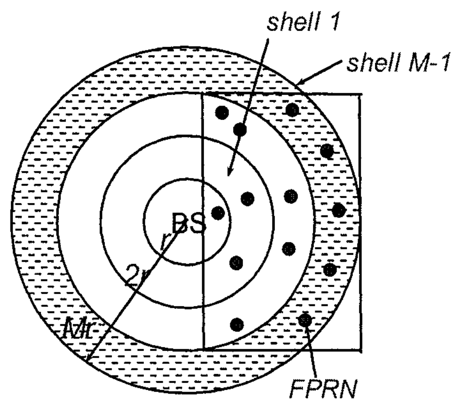


Figure 6.4 Definition of Shells

Lemma. Let w_j and rc_j be the workload and residual capacity of v_j. The total workload from all FPRNs in S_m is $W_m = \sum_{v_j \text{ belongs to } S_m} w_j$ and the total residual capacity is $RC_m = \sum_{v_j \text{ belongs to } S_m} rc_j$.

Denote the capacity of each RN by *C*. A lower bound on the optimal number of SPRNs is:

$$n_{\min} = \sum_{m=0}^{M-2} \left(\max\left(0, \left\lceil \left(\sum_{k=m+1}^{M-1} W_k - RC_m \right) / C \right\rceil \right) \right) \tag{6-1}$$

Proof: As all traffic transmission is towards and terminates at the BS, traffic residing in S_m must be relayed across each of the inner *shells* $S_{m-1}, \ldots, S_1, S_0$. Thus, starting from the farthest shell S_{M-1}, W_{M-1} amount of workload will pass through S_{M-2}. As the capacity of existing FPRNs is limited, some SPRNs may necessarily be added in S_{M-2} to carry W_{M-1}. Also, if the existing RNs are utilized completely before adding new nodes, the minimum number of SPRNs to be added in S_{M-2} is:

$$n_{\min}^{(M-2)} = \begin{cases} 0 & \text{if } RC_{M-2} \geq W_{M-1} \\ \lceil (W_{M-1} - RC_{M-2})/C \rceil & \text{otherwise} \end{cases} \tag{6-2}$$

Next the total workload relayed from S_{M-2} towards the BS will be $(W_{M-1} + W_{M-2})$. In general, following the same line of logic, the total workload from S_{m+1} towards the BS will be the accumulated workload of all the nodes in the shells $S_{M-1}, S_{M-2}, \ldots, S_{m+1}$; i.e., $\sum_{k=m+1}^{M-1} W_k$

Therefore, the minimum number of SPRNs to be added in the m^{th} shell is:

$$n_{\min}^{(m)} = \begin{cases} 0 & \text{if } RC_m \geq \sum_{k=m+1}^{M-1} W_k \\ \left\lceil \left(\sum_{k=m+1}^{M-1} W_k - RC_m \right)/C \right\rceil & \text{otherwise} \end{cases} \tag{6-3}$$

Then $\sum_{m=0}^{M-2} n_{\min}^{(m)}$ is a lower bound to the number of new nodes. ■

In the next subsection, we will compare the results obtained from our proposed schemes with the lower bound.

6.1.4 Performance Evaluation

To gain insight into the proposed algorithms, we compare their performance against the lower bound in some representative scenarios. We first assume that the FPRNs are placed in grid networks. We then assume the FPRNs are placed in random networks; the performance is evaluated in a statistical manner.

6.1.4.1 Performance Evaluation in Grid Networks

The grid network is set up as follows. Sixteen FPRNs are placed at the vertices of a 4x4 grid with the BS to the left of the grid. As shown in Figure 6.5, a is the grid length and the BS is located at [-b, ε], where $\varepsilon = 0.001a$ is a small perturbation value so that distances from FPRNs to the BS are distinct. We use $b = 0.55a$. The nodes are numbered in increasing order by their distance to the BS, i.e, node 16 is the farthest, while node 1 is the closest. All RNs have the same capacity of 8, meaning they can relay 8 units of traffic.

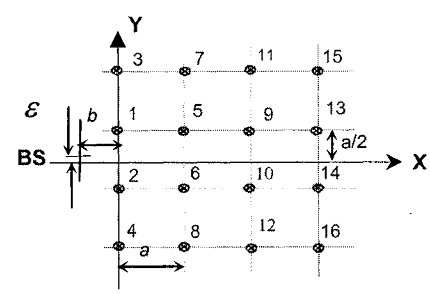


Figure 6.5 A Grid Network and its Parameters

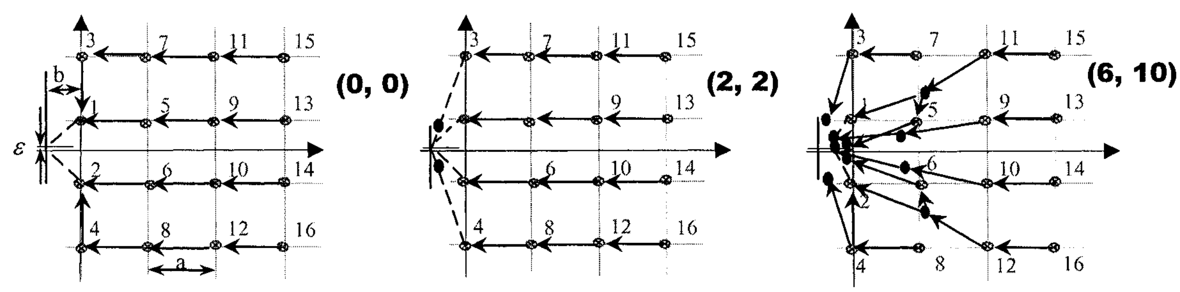
A. NTBF vs. MRCF

We first conduct two groups of experiments with both the NTBF and MRCF algorithms. In the first group of experiments, the transmission range of the RNs is $r = 1.2a$, so that each RN reaches its adjacent neighbors but not the diagonal neighbors. In the second group of experiments, the transmission range of RNs is enlarged to $r = 1.5a$. Thus, each RN has both adjacent neighbors and diagonal neighbors. Each group consists of three experiments, and the initial workload on FPRNs (directly from SNs) is 1, 2 and 4 respectively in all three experiments. Accordingly, the lower bounds on the optimal number of RNs in the three experiments of two groups are 0, 2, and 6 respectively.

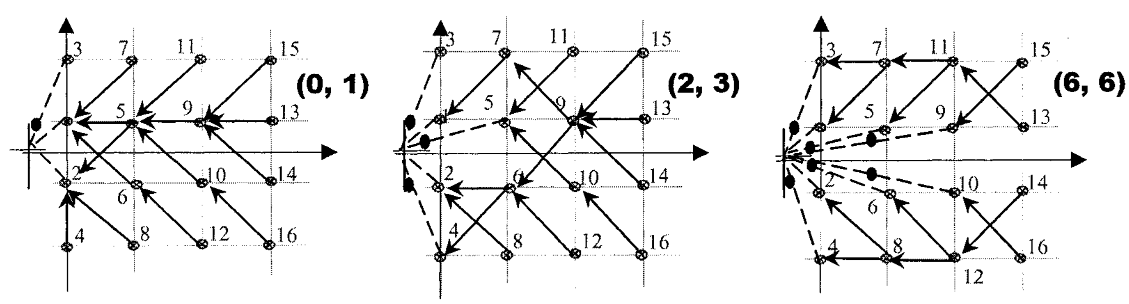
Figure 6.6 illustrates the placement of SPRNs by the NTBF algorithm. The number of the SPRNs is 0, 2 and 10 respectively when the transmission range is small. Contrary to intuition, the number increases as the transmission range increases in scenarios where the traffic load is 1 and 2 (comparing (A) and (D), (B) and (E)). This undesired increase is due to an uneven workload distribution among the RNs, i.e., the RNs closer to the BS will take a heavier load than the farther RNs. For example, in Figure 6.6 (D), node 9 is only a little closer to the BS than node 10 is, but its final workload is twice that of node 10. This will further create an unbalanced traffic distribution for nodes closer to the BS. This problem is due to the fact that the NTBF scheme prefers to fill up the capacity of the RNs closer to the BS. As a result, some traffic paths are unnecessarily congested.

We examine the MRCF algorithm using the same two groups of experiments as above. The results in the first group of experiments (short transmission range) are the same as those for the MRCF algorithm (thus omitted). In the second group, when the traffic load is 1 or 2, the MRCF algorithm does not exhibit the load unbalance problem and the numbers of RNs are optimal (the numbers of SPRNs placed by the algorithm are equal to the lower bounds), as shown in Figure 6.7. When the traffic load is 4, the MRCF algorithm has the same results as the NTBF algorithm.

However, the MRCF algorithm does not always have better placement than NTBF provides. We set up an experiment in which the FPRNs do not have equal initial workloads. Specifically, we set FPRNs 1-4 with initial workload 2, 5-8 with initial workload 3, and 9-16 with initial workload 4. In the case $r = 1.2a$, the results for both NTBF and MRCF are shown in Figure 6.8, where the MRCF algorithm requires one more RN than the NTBF algorithm.



(A) load =1, tx_range=1.2a **(B)** load =2, tx_range=1.2a **(C)** load =4, tx_range=1.2a

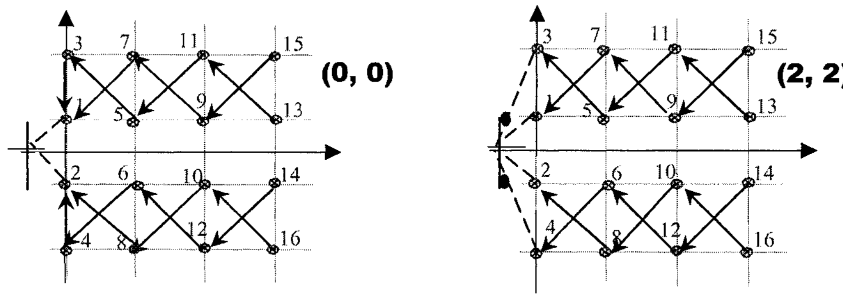


(D) load =1, tx_range=1.5a **(E)** load =2, tx_range=1.5a **(F)** load =4, tx_range=1.5a

⊗ First Phase Relaying Node ● Second Phase Relaying Node

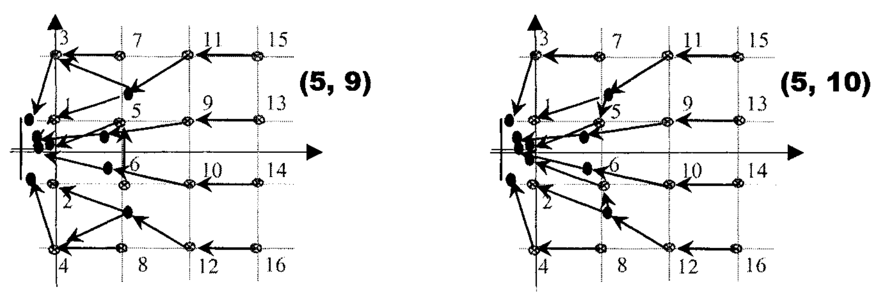
(a, b): **a** indicates the lower bound and **b** represents the number from the algorithms.

Figure 6.6 SPRN Placement and Traffic Distribution by Nearest-To-BS-First Algorithm



(A) load =1, tx_range=1.5a **(B)** load =2, tx_range=1.5a

Figure 6.7 SPRN Placement and Traffic Distribution by Max-Res-Cap-First Algorithm



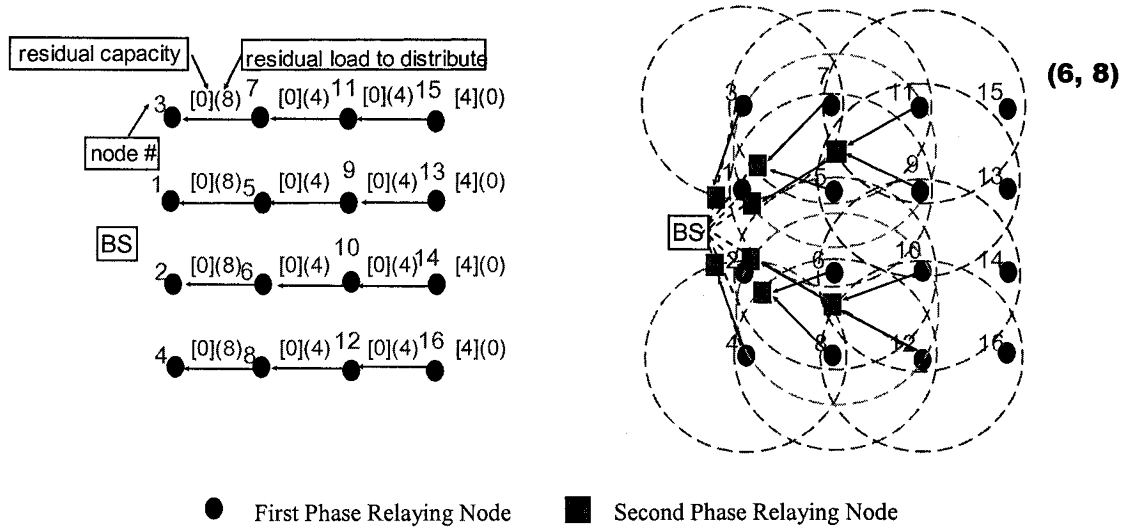
(A) Nearest-To-BS-First algorithm **(B)** Max-Res-Cap-First algorithm

Figure 6.8 SPRN Placement and Traffic Distribution, FPRNs with Unequal Initial Workload (1-4: 2; 5-8: 3; 9-16: 4). The Transmission Range is r = 1.2a

B. BER Algorithm

Here we evaluate the BER algorithm on the experiments above. In most cases, the BER algorithm achieves optimal results. Limited by space, we illustrate only two experimental results which are better than both the NTBF and MRCF algorithms.

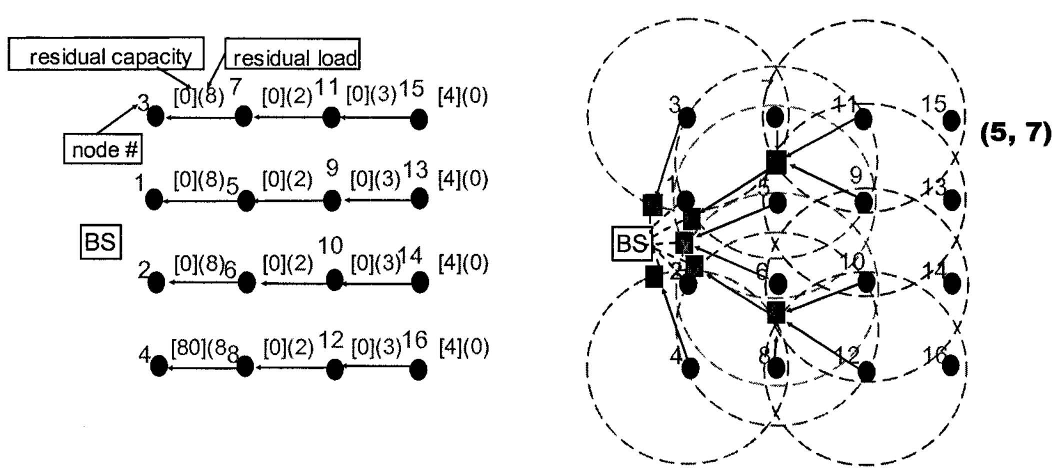
Figure 6.9 is the result of the third experiment of the first group, in which the traffic load on FPRNs is identically 4 and the transmission range is $r = 1.2a$. In this case, the BER algorithm requires 8 SPRNs (compared to 10 for both NTBF and MRCF). Figure 6.11 is the placement for the unequal traffic load experiment. The BER algorithm requires 7 SPRNs (compared to 9 by NTBF and 10 by MRCF).



(A) After Best Effort Relay

(B) Final RN placement

Figure 6.9 SPRN Placement of Best-Effort-Relaying Algorithm (load=4, rx_range=1.2a)



(A) After Best Effort Relay

(B) Final RN placement

Figure 6.10 SPRN Placement of Best Effort Relay Algorithm (tx_range=1.2a, unequal traffic load)

In brief, the results of all experiments and lower bounds are listed in Table 6-1.

Table 6-1 Results of Experiments in Grid Networks

Tx. range	Initial load	lower bound	BER	NTBF	MRCF
1.2a	1	0	0	0	2
	2	2	2	2	2
	4	6	8	10	10
	2,3,4	5	7	9	10
1.5a	1	0	0	1	0
	2	2	2	3	2
	4	6	6	6	6
	2,3,4	5	5	6	6

It can be seen that the BER algorithm provides at least as good or better results than the other two algorithms. In fact, in the 8 experiments, the BER algorithm produces optimal or near optimal solutions as its results are the same as or close to the lower bounds.

We remark that the BER algorithm outperforms the NTBF algorithm and the MRCF algorithm for two reasons. First, a RN forwards its traffic load as much as possible to RNs closer to the BS. Thus, the capacity of existing RNs is maximally utilized before any new RN is added. Second, when adding a new SPRN at any time, not only the farthest RN is considered, but also its intersected nodes are taken into account. Then, two or more RNs can share one common SPRN as their next hop relay. In such a way, the RNs can achieve better utilization.

Even though the experimental scenario is far from exhaustive, we argue that the Best-Effort-Relaying is better than the Nearest-To-BS-First and Max-Res-Cap-First algorithms in general. This argument will be further supported by an evaluation on random networks.

6.1.4.2 Performance Evaluation on Random Networks

As the number of potential design scenarios is immense, and an optimal placement is not available in general, in this section, the three algorithms are executed on randomly generated networks. Their performance is evaluated and compared in a statistical manner.

The random network is a square field of dimension 300x300, wherein the BS is located at the fixed position (0, 0). A total of 100 FPRNs are distributed randomly in the field. The coordinates of the four corners of the square field are (20, -150), (320, -150), (20, 150), and (320, 150), the transmission range of a RN is 15, and the initial workload of the FPRNs has a uniform distribution on [25, 75]. Four experiments are executed, in which the capacity of RNs is set to 75, 100, 125, and 150 respectively. In each of the experiments, 500 independent random networks are generated, and the results shown in this Section are the averages over 500 runs. The simulation is written in C++. Four metrics are used to evaluate the performance of the three heuristic algorithms.

(1) Number of SPRNs: This criterion directly reflects the total device cost incurred in a system given that the cost of an individual device is determined.
(2) Composite energy cost. This is the sum of capacities of all RNs in the network. It reflects the system cost in a unified manner and makes the placement designs comparable across the various scenarios. Regardless of other factors, the overall system cost is proportional to this metric.

(3) Energy utilization. This is the ratio of the workload on the RNs to their capacity. The higher the utilization of the nodes, the fewer new SPRNs are added. In this research, we calculate the average utilization of FPRNs and the average utilization of SPRNs separately.

(4) Coefficient of Variance (C.O.V.) of the energy utilization. It is a measurement of load balance among the RNs, defined for a given network as the standard deviation of node utilization divided by the average utilization over all nodes. The C.O.V. is computed separately for FPRNs and SPRNs.

Figure 6.11 demonstrates the results from applying the three algorithms to the four experiments. From these figures, we observe the following.

(1) All three algorithms require fewer SPRNs as the RN's capacity increases. This is a desirable property, as nodes with higher capacity should relay more traffic than those with lower capacity.

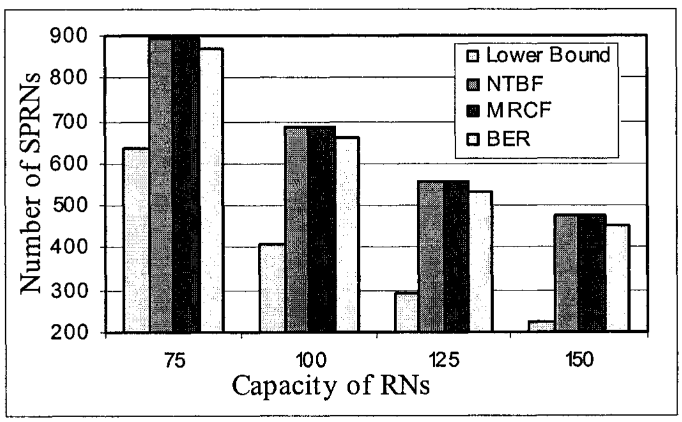
(2) Using all the three schemes, the ratio to the theoretical lower bound is 1.4 – 2.1, which is quite acceptable.

(3) The number of SPRNs used by the NTBF and the MRCF algorithms are similar to one another with the latter having smaller C.O.V. of the energy utilization, which supports the assertion that the MRCF algorithm achieves better load balancing among the nodes.

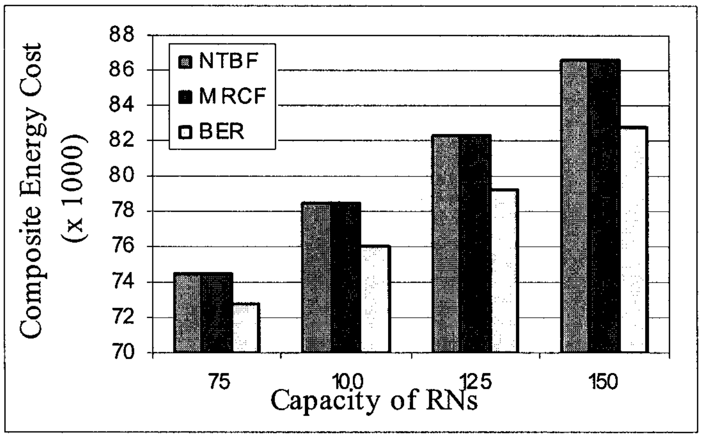
(4) The BER algorithm outperforms the other two algorithms in terms of smaller number of SPRNs, higher utilization, and smaller C.O.V., though the improvement is moderate.

(5) From Figure 6.11 (B) and (C), we notice that with the increase in individual node capacity, the comprehensive energy cost increases, while the resource utilization decreases. This suggests that an energy economical system may prefer to use a large number of devices with less energy rather than using a smaller number of devices with higher energy. This could be helpful for the network designer when choosing the devices.

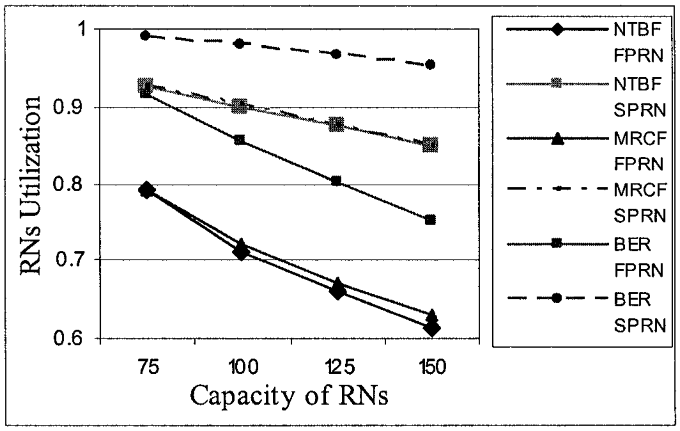
Furthermore, we checked the standard deviations of the average utilization in every case (running each of the three algorithms in the 4 experiments). Results show that in all experiments, the standard deviations of average utilization fall in [0.0037, 0.0234], and standard deviation of average COV is within [0.013, 0.024]. These results indicate that our proposed schemes perform stably in this scenario.



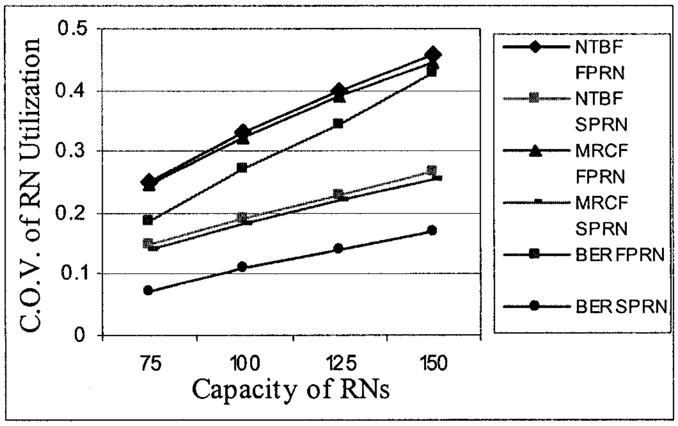
(A) Number of SPRNs vs. capacity



(B) Composite energy cost vs. capacity



(C) RNs utilization vs. capacity



(D) C.O.V of RNs utilization vs. capacity

Figure 6.11 Experiment Results in Random Networks

6.2 Relay Nodes with Variable Transmission Range

In contrast to the work in Section 6.1, in this Section, we assume RNs can adjust their transmission power according to the distance to the intended destination. To make the best use of the power adaptivity of the RN, we propose two heuristic solutions, namely, Independent Placement with Direct Allocation (IPDA) and Collaborative Placement with Locally Optimal Allocation Decision (CPLOAD). Furthermore, a lower bound on the minimum number of SPRNs is provided. The effectiveness of our proposals is investigated through simulation using numerical examples.

6.2.1 Optimal Transmission Range

It is a common belief that multi-hop transmission is more energy efficient than single-hop direct transmission when the distance between the source and the destination is long [146]. However, should the number of hops increase with no bound?

Figure 6.12 plots the energy consumption for delivering one data bit over a distance $D=1000$ *meters* against the number of hops m. In the figure, when m is small, the energy consumption decreases as m increases. However, after the value of m exceeds a certain value (14 in this example), the energy consumption increases with m. The reason is as follows. As m increases, the number of receiving operations increases. Thus, the energy saving from transmitting is gradually surpassed by the energy cost on the extra reception when m becomes larger. From Figure 6.12, it is safe to conclude that both too short and too long transmission distances are not energy efficient. Next, we derive the most energy efficient multi-hop relay arrangement.

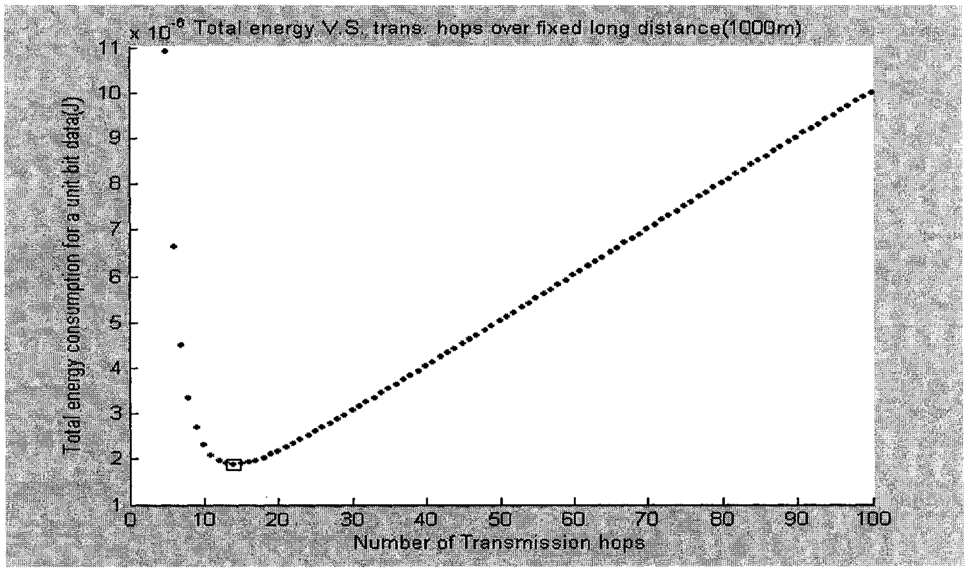


Figure 6.12 Total Energy vs. Hops Over Fixed Long Distance (delivering 1 data bit)

$\alpha_1 = \beta = 50nJ/bit$, $\alpha_2 = 0.0013pJ/bit/m^4$, n=4, as in [78]

In [146], the optimal hop distance in terms of minimizing energy consumption from a source to a destination with given distance was derived independently as:

$$d_{opt} = \sqrt[n]{\frac{\alpha_1 + \beta}{(n-1) \cdot \alpha_2}} \tag{6-4}$$

The optimal distance d_{opt} only depends on the circuit parameters and propagation environment. When the distance between the source and the destination is smaller than or equal to d_{opt}, one-hop transmission is most energy efficient. When the distance is larger than d_{opt}, it would be desired to adopt multi-hop transmission with the distance of each hop being as close to d_{opt} as possible. In reality, the distance D may not be an integer multiple of d_{opt}. It is easy to show that the optimal multi-hop relay arrangement will be one of two ways as follows. The first option is a $\left\lceil D/d_{opt} \right\rceil$ hop relay with evenly separated distances. In this case, the per-hop distance is smaller than d_{opt}. The second option is a $\left\lfloor D/d_{opt} \right\rfloor$ hop relay with evenly separately distances. In this case, the per-hop distance is greater than d_{opt}. The two multi-hop relay arrangement methods are illustrated in Figure 6.13.

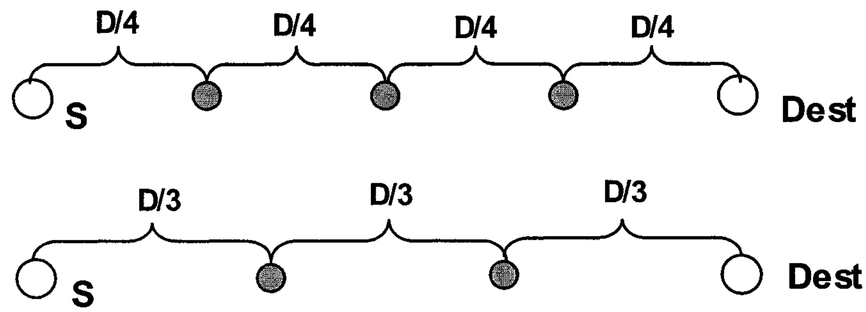


Figure 6.13 An Illustration of Multi-hop Relay Arrangement when(D/d_{opt}) =3.5

Furthermore, the capacity corresponding to the optimal transmission range is

$$C_i^{opt} = \frac{E_i}{(\alpha_1 + \beta)\cdot T} \cdot (1 - \frac{1}{n}) \tag{6-5}$$

The optimal transmission distance as well as the corresponding traffic capacity will be used as guideline parameters when designing device placement in the next section. The effectiveness of using the optimal transmission range will be verified by numerical results shown in Section 6.2.4.

6.2.2 Heuristic SPRN Placement and Traffic Allocation Algorithms

In this subsection, we describe our proposed schemes to solve the combined RN placement and traffic allocation problem. To describe the algorithms, we first abstract the network of

RNs as a directed graph $G(V, A)$. Initially, $V=V_F=\{v_1, v_2, ..., v_N\}$ is the set of FPRNs and A is the edge set. Each RN has an associated weight, which represents the total traffic load it has. Let w_i denote the traffic load of v_i, d_i denote the Euclidean distance from v_i to the BS, and d_{ij} denote the Euclidean distance between v_i and v_j.

6.2.2.1 Independent Placement with Direct Allocation (IPDA)

This is a straightforward and easy to implement method. Based on the traffic load and energy supply on a FPRN v_i, if all its traffic is forwarded to a single next hop relay, its maximum transmission distance (also called critical transmission range in this thesis) can be calculated as:

$$d_i^{(critical)} = \sqrt[n]{\frac{E_i - w_i \cdot T \cdot (\alpha_1 + \beta)}{w_i \cdot T \cdot \alpha_2}} \tag{6-6}$$

Then an independent path for each FPRN can be established by placing the SPRNs along the line joining the FPRN and the BS at the interval of its critical transmission distance. An example of using the IPDA algorithm is shown in Figure 6.14. Since the traffic load at each FPRN may be different, the per-hop distance on each path may be different too.

By using the IPDA algorithm, all RNs except those that can reach the BS directly will use up their energy. This is helpful to reduce the number of newly added SPRNs. However, two important aspects are ignored. First, the traffic allocation for an existing node only depends on one newly added RN. It does not take into account the capability of the existing RNs. Second, the transmission distance of the RNs depends on its initial traffic load, which is not likely to be the optimal transmission range in many cases. Both may result in inefficient energy utilization.

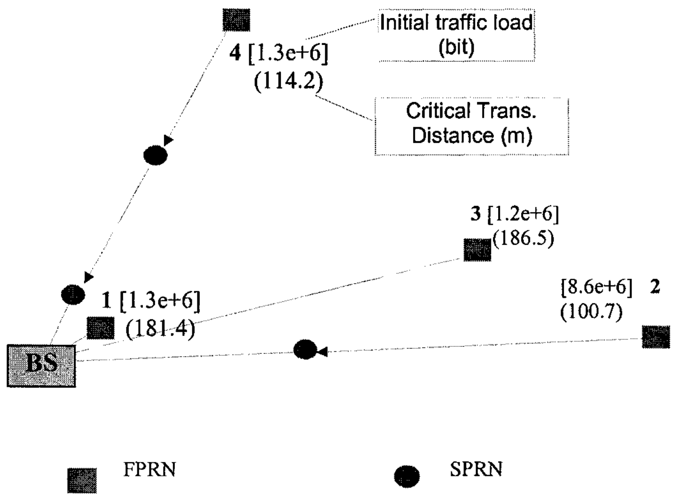


Figure 6.14 An Example Using IPDA Algorithm

6.2.2.2. Collaborative Placement with Locally Optimal Allocation Decision (CPLOAD)

To overcome the problems in the IPDA algorithm in minimizing the total number of newly added RNs, we propose the CPLOAD algorithm. The CPLOAD algorithm is based on the following principles:

(1) Making the best use of the resources of the existing RNs before adding any new RN. In general, it is desired to arrange traffic allocation for each RN among the existing RNs. However, due to energy constraints, the existing nodes might not be capable of relaying all the traffic of a given FPRN to the BS, and so new RNs are needed.

(2) Avoiding detouring traffic as much as possible. Since the BS is the ultimate destination, all traffic should be transmitted towards the BS. Any unnecessary traffic detour will lead to energy wastage.

(3) Adopting optimal transmission distance whenever possible. By doing so, the traffic relaying will be highly energy efficient.

Using the same notations defined earlier, we further define w_{ij} and p_{ij} to be the traffic load and the fraction of v_i's traffic load allocated to v_j, i.e., $p_{ij} = \frac{w_{ij}}{w_i}$. Define G_c to be the candidate set of RN under consideration (initially, $G_c = \{v_0\} \cup V_F$, where v_0 refers to the BS). Define G_i to be the set of all nodes that can be directly reached by node v_i within a certain transmit power level in its dynamic transmission range, with $d_j < d_i$ if $v_j \in G_i$. The CPLOAD scheme operates as follows:

(a) Locally Optimal Allocation Decision

Pick the node v_i from G_c, where v_i is the farthest from the BS in G_c. By doing so, the traffic can always be allocated to nodes that are closer to the BS, which can avoid energy wastage due to unnecessary traffic detouring. We follow the principle that v_i should maximally use its energy to forward its traffic load as close to the BS as possible. We show that such traffic allocation is a linear programming (LP) problem with the following objective function and constraints:

$$\min \sum_{j \in G_i} w_i \cdot p_{ij} \cdot (\alpha_1 + \alpha_2 \cdot d_j^n) \tag{6-7}$$

subject to:

$$\sum_{j \in G_i} p_{ij} = 1 \tag{6-8}$$

$$\left(\alpha_1 + \beta + \alpha_2 \cdot \sum_{j \in G_i} p_{ij} \cdot d_{ij}^n \right) \cdot w_i \cdot T \leq E_i \tag{6-9}$$

$$w_i \cdot p_{ij} + w_j \leq C_j, \; j, s.t. v_j \in G_i \tag{6-10}$$

$$p_{ij} \geq 0; \; for\ all\ j, s.t.\, v_j \in G_i \tag{6-11}$$

With the objective function given in (6-7), the traffic load w_i will be delivered as close to the BS as possible, so that the least energy consumption is required for further traffic relaying.

Constraint (6-8) is due to the local traffic balance condition, i.e., all incoming traffic at a RN v_i should be transmitted to next hop relay(s).

Constraint (6-9) is due to the energy limitation at node v_i, i.e., the total energy consumed at v_i cannot exceed its energy supply.

Constraint (6-10) is due to the energy limitation on other RNs, and

$$C_j = \max\left(w_j, C_j^{(opt)}, \frac{E_j}{(\alpha_1 + \alpha_2 \cdot d_j^n + \beta) \cdot T} \right) \tag{6-12}$$

where $C_j^{(opt)}$ is calculated as per Eq.(6-5).

The capacity of v_j is based on several considerations:

1) The initial workload cannot exceed a node's capacity; this is the first term on the right hand side of Eq. (6-12).
2) When a RN is far from the BS, its traffic load is expected to be delivered to the BS over multiple hops, with each hop distance at the optimal transmission range. This will lead to minimum overall energy consumption. Thus, we have the second term on the right hand side of Eq. (6-12).
3) When a RN is close to the BS, the total traffic amount it can afford depends on its actual distance to the BS. This is the third term in Eq. (6-12).

The non-negative constraints given in (6-11) are obvious.

If the above LP problem has an optimal solution, the traffic load of all v_i's next hop relay(s) is updated according to $w_j^{(new)} = w_j + p_{ij} \cdot w_i$. The i^{th} row of the next hop relay table and traffic allocation table is updated accordingly. Then v_i is deleted from the candidate set G_c.

If the above LP problem has no feasible solution (e.g., some of the constraints are conflicting), it implies that v_i cannot allocate all its traffic to the existing RNs, and a new RN, $v_i^{(new)}$, should be added. Go to step (b).

(b) New RN Placement

(c) Repeat steps (a) and (b) until $G_c = \{v_o\}$.

(d) Redundancy Elimination

By executing steps (a)-(c), a RN placement design and the accompanying traffic allocation decisions have been formed for a given WSN. In the process, when a new RN is added, its position only depends on the RN under examination without considering the benefit to other RNs. Such local optimality results in resource duplication in some cases. For example, in Figure 6.15 (A), RNs v_1^{new} and v_2^{new} have been added to relay traffic load for v_1 and v_2 to the BS respectively. We illustrate replacing these two RNs in Figure 6.15 (B), wherein a strategically placed RN, say v_{NEW} can relay traffic for both v_1 and v_2 to the BS. The redundancy elimination procedure is a systematic approach to remove resource duplication and decrease the number of RNs further.

We advocate confining the redundancy elimination to last hop SPRNs (LHSPRNs), namely, RNs that connect to the BS in one hop, for three reasons. Firstly, LHSPRNs are close to the BS. Compared with other nodes that are further from the BS, they are more likely to have residual energy. Secondly, SPRNs fan in the vicinity of the BS. Geometrically, they are closer to each other. Thirdly, it reduces the complexity dramatically. Thus, the redundancy elimination procedure is designed as follows:

(i) Initialize the candidate elimination nodes (CENs), G_L as all LHSPRNs.

(ii) Eliminate redundant pairs. For an arbitrary un-checked pair of CENs, v_i and v_j, find the predecessors of v_i and v_j. Based on the positions of those predecessors as well as the traffic load the predecessors relayed to v_i and v_j, we can check if a new node, v_{new} can be relocated at a spot such that it can replace v_i and v_j without violating energy constraints. If so, add v_{new} in G_L, reallocate the traffic load of the predecessors to v_{new}, and delete v_i and

v_j from G_L. Then exhaustively check all pairs of updated CENs until no pair can be merged.

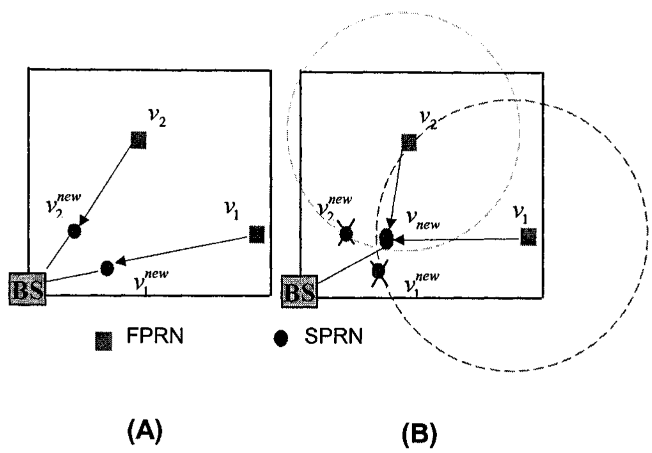


Figure 6.15 An Example of Redundancy Elimination

The pseudo-code for redundancy elimination is shown in Figure 6.16.

initialize G_L;
if two or more nodes in G_L
    pair nodes in G_L;
    do
    {
        pick an unchecked pair, v_i and v_j;
        find predecessors of v_i and v_j;
        draw circles centered at the predecessors with radii of the corresponding distances from the predecessors to v_i and v_j respectively;
        if a common intersection of these circles exists
            if a v_{new} (placed at the point in the intersection nearest to the BS) is capable of delivering all traffic load originally from the predecessors to v_i and v_j to the BS directly
                add v_{new} to G_L;
                reallocate traffic from v_i and v_j 's predecessors to v_{new};
                delete v_i and v_j from G_L;
                pair v_{new} with other nodes in G_L;
            end if
        end if
    }while (not all pairs in G_L have been checked)
end if

Figure 6.16 Pseudo Code for Redundancy Elimination

The flowchart of the CPLOAD algorithm is given in Figure. 6.17.

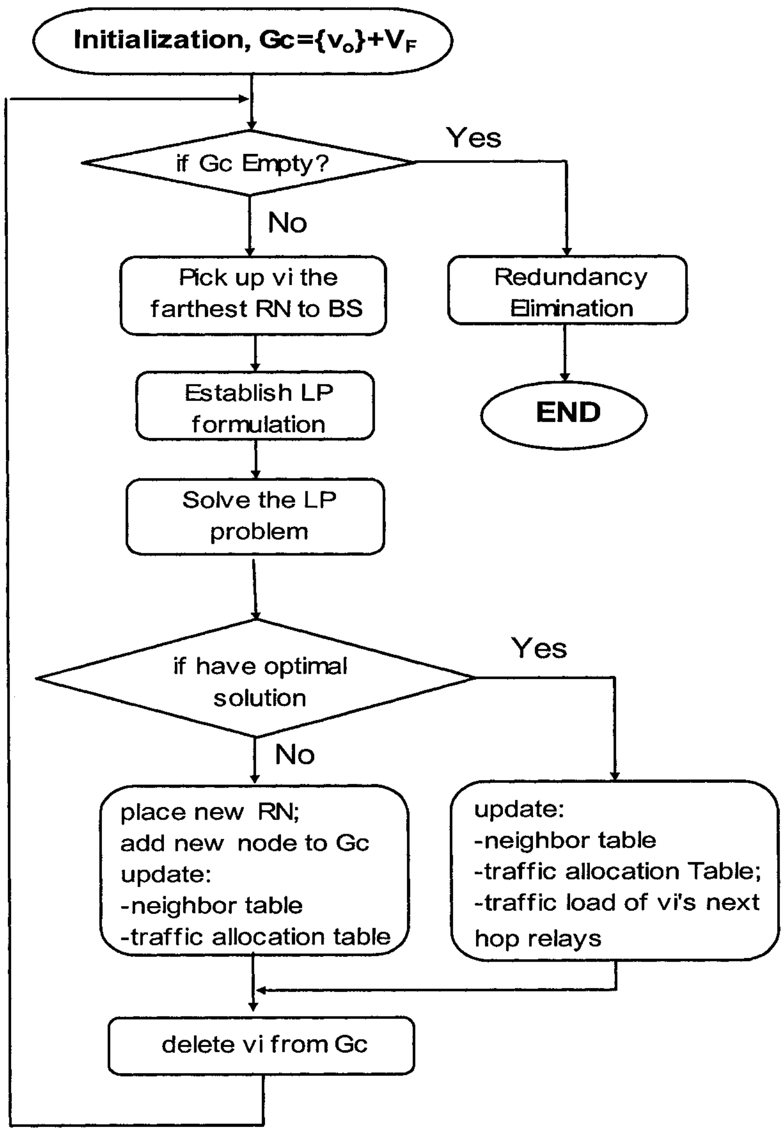


Figure 6.17 Flowchart of CPLOAD Algorithm

6.2.3 Lower Bound on the Number of SPRNs

The ultimate goal of this research is to place a minimum number of SPRNs in order to relay all traffic on FPRNs to the BS. The number of SPRNs is affected by many factors, such as the topology of the given network composed of the BS and FPRNs, and the traffic loads at the FPRNs. As shown above, finding the optimal placement of SPRNs is non-trivial. In this section, we derive a lower bound on the optimal number of SPRNs for a given network.

To obtain this bound, we first derive the minimum total energy consumption (excluding the energy cost on the BS) for relaying all traffic on FPRNs to the BS. The optimal number of SPRNs should be lower bounded by the total required energy and the initial energy of individual RNs.

In Section 6.2.1, we found that the most energy efficient multi-hop relay arrangement can be obtained by comparing the energy consumption of two methods. For a FPRN v_i with traffic load w_i, let $m_i = \frac{d_i}{d_{opt}}$, the energy consumption for transmitting traffic of amount w_i from v_i to the BS by using the two methods would be:

if $d_i > d_{opt}$

$$e_i^{(1)} = w_i \cdot T \cdot \lceil m_i \rceil \cdot (\alpha_1 + \alpha_2 \cdot (d_i / \lceil m_i \rceil)^n + \beta) \tag{6-13}$$

$$e_i^{(2)} = w_i \cdot T \cdot \lfloor m_i \rfloor \cdot (\alpha_1 + \alpha_2 \cdot (d_i / \lfloor m_i \rfloor)^n + \beta) \tag{6-14}$$

if $d_i \le d_{opt}$

$$e_i^{(1)} = e_i^{(2)} = w_i \cdot T \cdot (\alpha_1 + \alpha_2 \cdot d_i^n + \beta) \tag{6-15}$$

Define $e_i^{(\min)}$ to be the minimum energy consumption for relaying the traffic load w_i at v_i to the BS. We have:

$$e_i^{(\min)} = \min(e_i^{(1)}, e_i^{(2)}) \tag{6-16}$$

Define $e_{total}^{\min}$ to be the minimum total energy consumption for relaying all traffic load at all FPRNs to the BS. We have:

$$e_{total}^{\min} = \sum_{i:v_i \in V_F} e_i^{(\min)} \tag{6-17}$$

The lower bound of the number of SPRNs $n_s^{(opt)}$ is:

$$n_s^{(opt)} = \max\left(0, \left\lceil \frac{e_{\min} - \sum_{i:FPRN} E_i}{E_S} \right\rceil\right) \tag{6-18}$$

where E_i is the initial energy at FPRN i, and E_s is the initial energy at each individual SPRN. We note that this lower bound only depends on the initial energy of each individual RN and the traffic loads and distances to the BS of FPRNs. Further improvement to the bound is possible with more detailed knowledge about the physical locations of the FPRNs, though we do not pursue this here.

6.2.4 Performance Evaluation

In this section, we evaluate the proposed algorithms by comparing the number of SPRNs with the lower bound in randomly generated networks in various scenarios.

Initially, a total of N FPRNs are distributed randomly in a two dimensional square field. The initial workload of the FPRNs has a uniform distribution within [a, b]*C^{max}, where C^{max} is the maximum traffic load at each RN, which can be calculated according to Eq. (3-11). We set the same initial energy for all RNs. The parameters used in the experiments are listed in Table 6-2.

Table 6-2 Parameters Used in Simulations

Para	Values	Para	Values
α_1	50nJ/bit*	E	2J
α_2	0.0013pJ/bit/m^4	T	10000s
β	50nJ/bit	C^{max}	2000 bits/s
n	4	C_{opt}	1500 bits/s

We use the same values of α_1, α_2, β and n as set in [51, 78].

In each scenario, 50 independent random networks are generated. The results presented in this thesis are the averages over the 50 runs. The simulation program is written in MATLAB.

6.2.4.1. An Example Using the CPLOAD Algorithm

Here we demonstrate how the CPLOAD algorithm can distribute the load and reduce the number of SPRNs. Let there initially be 20 FPRNs, randomly distributed in a two dimensional square field with four corners positioned at [-200, -200], [-200, 200], [200,-200] and [200, 200]. The BS is located at [0, 0] (i.e., at the center of the field). The initial traffic load at each FPRN is uniformly distributed within [0.3-0.5] C^{max}.

The resulting relay node deployment after executing the CPLOAD algorithm is shown in Table 6-3 and Figures 6.18-20, in which RNs numbered 1-20 are FPRNs and 21-29 are the SPRNs required.

Table 6-3 lists the traffic allocation decision for each RN, including its next hop relay(s) and the portion of its traffic load relayed to each of them. Figure 6.18 illustrates the initial network set-up with FPRNs and the resultant SPRNs deployment. Figure 6.19 shows the initial traffic load of FPRNs and the final traffic load of all the FPRNs and SPRNs. As expected, the average traffic load of SPRNs is higher than that of FPRNs. Figure 6.20 shows the energy consumption on each RN. In general, SPRNs use their energy more efficiently than FPRNs.

Table 6-3 Example of Using the CPLOAD Algorithm

RN	xi	yi	relays	RN	xi	yi	relays	RN	xi	yi	relays
1	105.0	81.2	25(1)	11	12.6	-110.2	28(1)	21	66.4182	-108.4683	26(1)
2	10.0	-31.2	BS(1)	12	-85.0	-31.5	BS(1)	22	-49.6392	91.833	29(1)
3	-172.8	-61.6	23(1)	13	146.5	90.3	24(1)	23	-88.9403	-31.7279	BS(0.63) 5(0.37)
4	120.0	-195.9	21(1)	14	25.8	136.8	22(0.27) 24(0.73)	24	64.1205	39.5121	BS(0.78) 20(0.22)
5	-16.9	-18.5	BS(1)	15	157.1	-60.1	16(0.62) 21(0.38)	25	48.6932	37.6725	BS(1)
6	-145.0	61.9	9(1)	16	60.3	10.6	BS(1)	26	27.0028	-44.0985	BS(1)
7	-95.1	-67.3	27(1)	17	69.1	-160.0	11(0.63) 21(0.37)	27	-36.99	-26.1793	BS(1)
8	-174.5	-135.6	3(0.36) 7(0.64)	18	-94.3	174.4	22(1)	28	4.5059	-39.5299	BS(1)
9	-60.8	34.6	BS(1)	19	180.4	91.3	1(0.85) 13(0.15)	29	-9.4394	17.463	BS(1)
10	-4.3	-50.6	BS(1)	20	50.5	5.0	BS(1)				

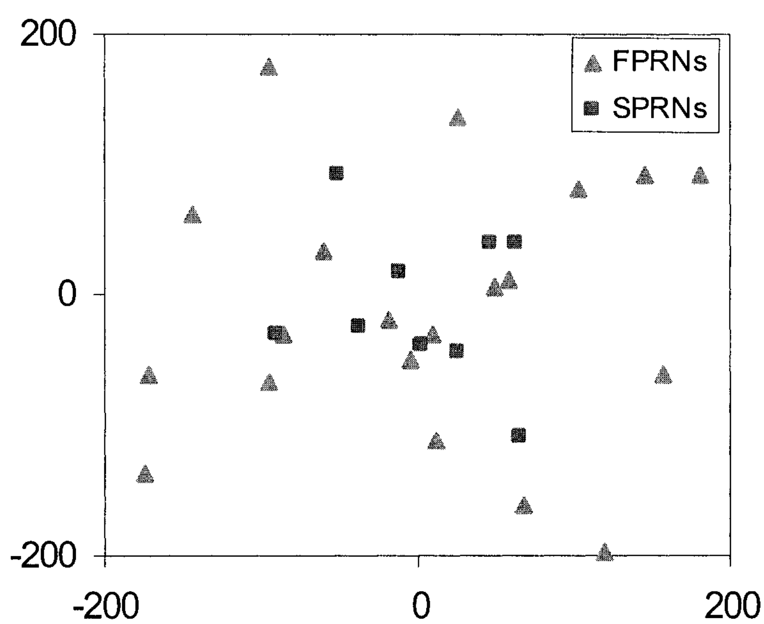


Figure 6.18 Positions of FPRNs and SPRNs

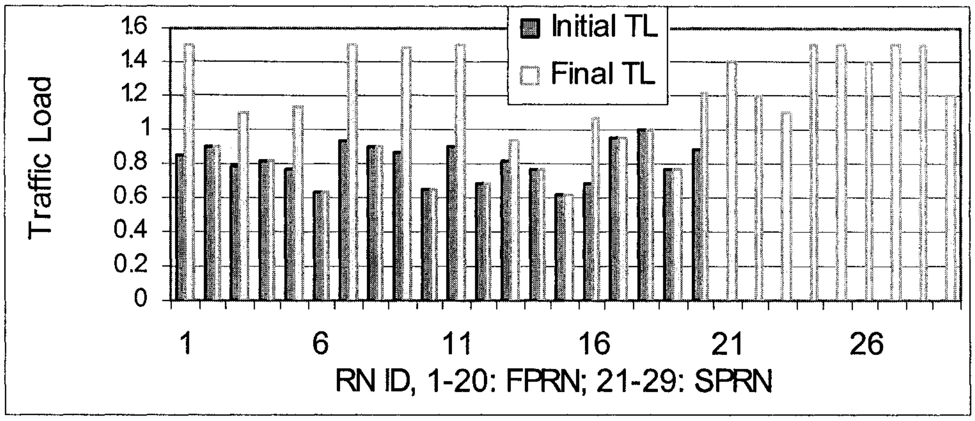


Figure 6.19 Traffic Load on the RNs (X1e7bits) (RN #1-20: FPRNs, 21-29: SPRNs)

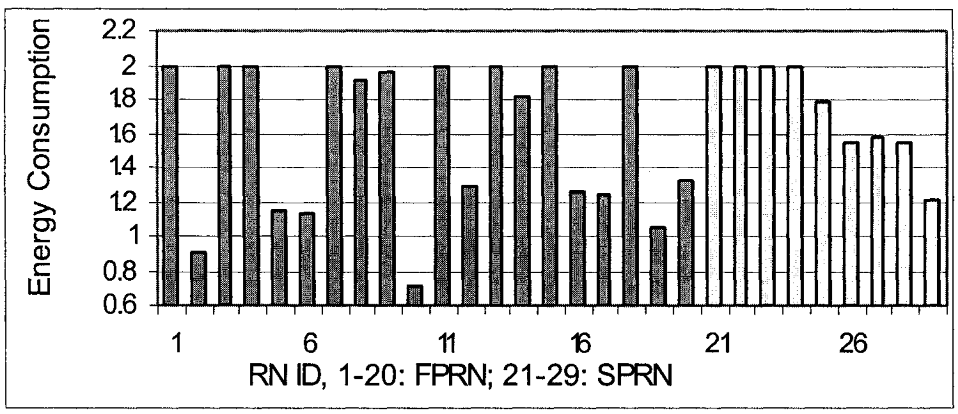


Figure 6.20 Energy Consumption on the RNs (J) (RN #1-20: FPRNs, 21-29: SPRNs)

6.2.4.2. Performance Comparison

In the following, we evaluate the performance of the proposed algorithms under different scenarios. The number of SPRNs is represented by a normalized value, which is the ratio of the actual number of SPRNs to the lower bound. The traffic load is also represented by a normalized value, which is the ratio of the actual traffic volume to C^{max}.

A. Effect of initial traffic load

In the first group of experiments, the network is generated as follows: a total of N =100 FPRNs are randomly scattered in a square field with the four corners positioned at [600, -200], [600, 200], [1000, -200], [1000, 200]. The BS is fixed at [0, 0], (which is called the far BS). In different scenarios, the normalized initial traffic loads of the FPRNs are set uniformly within [0.1-0.3], [0.3-0.5], [0.5-0.65] and [0.65-0.85] respectively. The results are shown in Figure 6.21.

In all scenarios, CPLOAD outperforms IPDA in terms of the number of SPRNs and energy consumption. Moreover, we also observe that:

(1) As shown in Figure 6.21, when the BS is far from the sensing field, the number of SPRNs resulting from CPLOAD is close to the lower bound for all traffic conditions. This indicates that, in this setting, the locally optimal traffic allocation works well and that the derived lower bound is relatively tight.

(2) The lighter the initial traffic load, the better the CPLOAD algorithm outperforms IPDA. This verifies the effectiveness of the locally optimal traffic allocation scheme. When the traffic load is light, a node can transmit over a longer distance, so it may have more candidate next hop relays. Moreover, the existing RNs are more likely to have high residual capacity. The CPLOAD algorithm takes advantage of this and assigns traffic to the existing RNs whenever possible before adding a new node.

(3) The higher the initial traffic load, the closer the two algorithms perform, and the closer the number of SPRNs is to the lower bound. In particular, when the initial traffic load is [0.65 - 0.85], i.e., with average at 0.75, which is the same as the capacity corresponding to the derived optimal transmission distance, the number of SPRNs using both CPLOAD and IPDA is very close to the derived lower bound. This validates the effectiveness of the optimal transmission range. In such a case, most of the nodes do not have residual capacity to afford traffic relaying for others, and the traffic allocation to existing RNs is likely infeasible. So CPLOAD and IPDA operate similarly. Moreover, when the initial traffic load is close to the optimal capacity, a RN will transmit at a distance close to the optimal transmission distance. This leads to high energy efficiency.

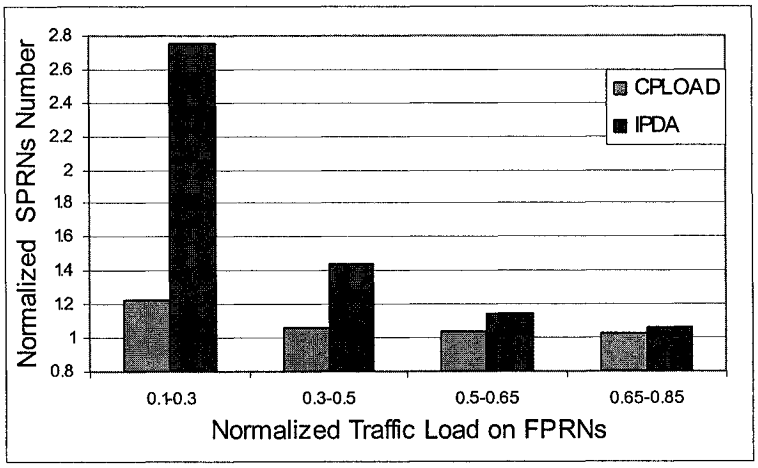


Figure 6.21 Number of SPRNs vs. Initial Traffic Load

B. Effect of BS location

In this set of experiments, the impact of the BS positioning on performance is investigated. The FPRN field set-up is similar to that in the first group of experiments. We move the BS towards the field of FPRNs to [200,0] (called the close BS), [400, 0] (called the closer BS) and [600,0] (called the edge BS) and conduct another 3 groups of experiments. The results for executing CPLOAD are shown in Figure 6.22. Note that under the same traffic load, the closer the BS is to the FPRN field, the higher the number of normalized SPRNs. This can also be attributed to the effectiveness of the optimal transmission distance. When the BS is closer to the RN field, a greater percentage of nodes transmit at a distance much different from the optimal transmission distance. This generally degrades energy efficiency compared to the case when the BS is away from the RN field.

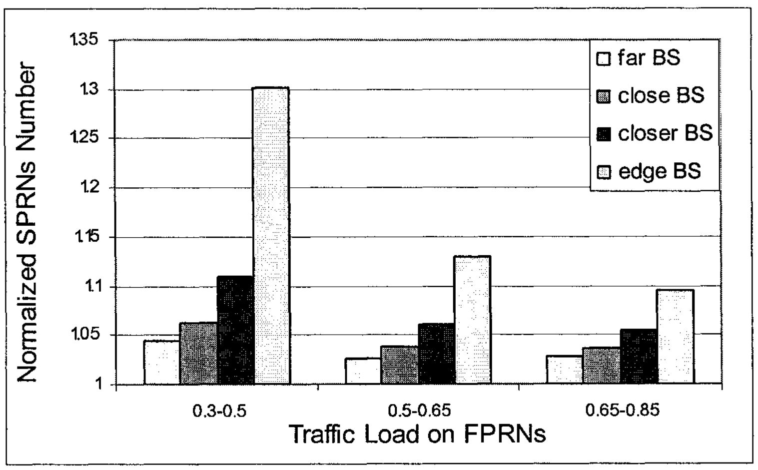


Figure 6.22 Impact of BS Location on Number of SPRNs (Number of FPRNs = 100)

C. Effect of number of FPRNs

This group of experiments investigates the impact of the number of FPRNs on performance. The network setup is similar to that in the previous case A, except that the number of FPRNs is set at 30, 50, 100 and 150, respectively in each scenario. The results of using CPLOAD are shown in Figure 6.23. As expected, at a specific traffic load, the normalized number of SPRNs is insensitive to the number of FPRNs.

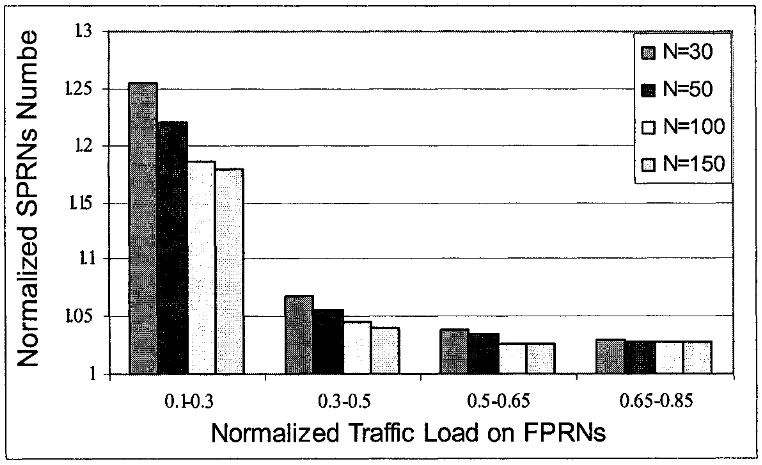


Figure 6.23 Number of SPRNs vs. Number of FPRNs at Different Initial Traffic Load (far BS)

From the experiments above, we observe that CPLOAD generally outperforms IPDA. When the initial traffic load on devices is light and/or the BS is far from the sensing field, the advantage of CPLOAD is more significant. Moreover, the effectiveness of optimal transmission distance is also validated by the simulation results.

6.3 Summary

Based on the work in Chapters 4 and 5, in this chapter, we further explore the problem of device provisioning in the second phase of the communication domain. . The objective is to use a minimum number of SPRNs so that all the given traffic at existing nodes can be relayed to the BS. We identify two components of the problem, i.e., to determine the exact positions of the devices, and to decide the traffic allocation scheme for each RN to its neighboring nodes. As the two components are coupled tightly to one another, a globally optimal solution is not available now. Therefore, we propose the Far-Near, Max-Min principles as guidance to solve the compound problem.

Considering that the RNs have fixed transmission range, we present several polynomial- time heuristic schemes, i.e., NTBF, MRCF, and BER, to solve the formulated problem. The computational complexity of these schemes is bounded by $O\left(N_{RN}^{2} \cdot \log N_{RN}\right)$, where N_{RN} is the total number of RNs. We derive a theoretical lower bound on the minimum number of SPRNs. Analysis and numerical results indicate that the proposed algorithms can provide optimal or near optimal solution in some scenarios.

Furthermore, we analyze the impact of transmission range on energy efficiency. Considering each RN can adjust its transmission power and using the optimal transmission distance, a polynomial-time Collaborative Placement with Locally Optimal Allocation Decision (CPLOAD) algorithm is proposed to solve the additional relay nodes placement. We also offer a theoretical lower bound on the minimum (optimal) number of SPRNs. Numerical results indicate that the proposed CPLOAD can provide a near optimal solution in some scenarios. Experimental results also advocate the argument that transmitting at the optimal distance can lead to high energy efficiency. The research in this chapter can be a vital guideline in WSN design.

Moreover, the RN provisioning problem tackled in this chapter is fundamental and critical in WSN design. Combining the work in Chapters 4 and 5, a comprehensive network planning solution is offered.

7. CONCLUDING REMARKS

Wireless Sensor Networks (WSNs) exhibit many advantages over their conventional wired counterparts, including ease of deployment, self-organization, reliability, versatility, scalability and flexibility. With the rapid advancement of technology, they will turn various challenging detection and monitoring tasks into reality in the near future.

7.1 Summary and Contributions

As the interest for WSNs from both academia and industry has increased rapidly in recent years, many technical issues have been identified and studied. Some were introduced in Chapter 2. In this thesis, we focus on the network planning issues of heterogeneous WSNs in a controlled environment. Network planning is a design process that determines the types, numbers, and locations of WSN devices. The goal of network planning is to achieve a favorable trade-off between system performance such as sensing coverage, connectivity, lifetime, reliability, etc., and the overall cost. In this thesis, we conduct a deep study on the network planning problem for a heterogeneous network. We investigate the network planning problems for two functional domains, namely, the sensing domain and the communication domain. For each domain, the network planning efforts are subject to different performance requirements. For the sensing domain, a minimum number of SNs are arranged so as to achieve a given network information coverage. For the communication domain, a minimum number of RNs are added to provide lifetime-guaranteed network connectivity.

Figure 7.1 recaps the WSN network planning process, the design framework, and its components. It also summarizes various scenarios under consideration, as well as the schemes developed accordingly.

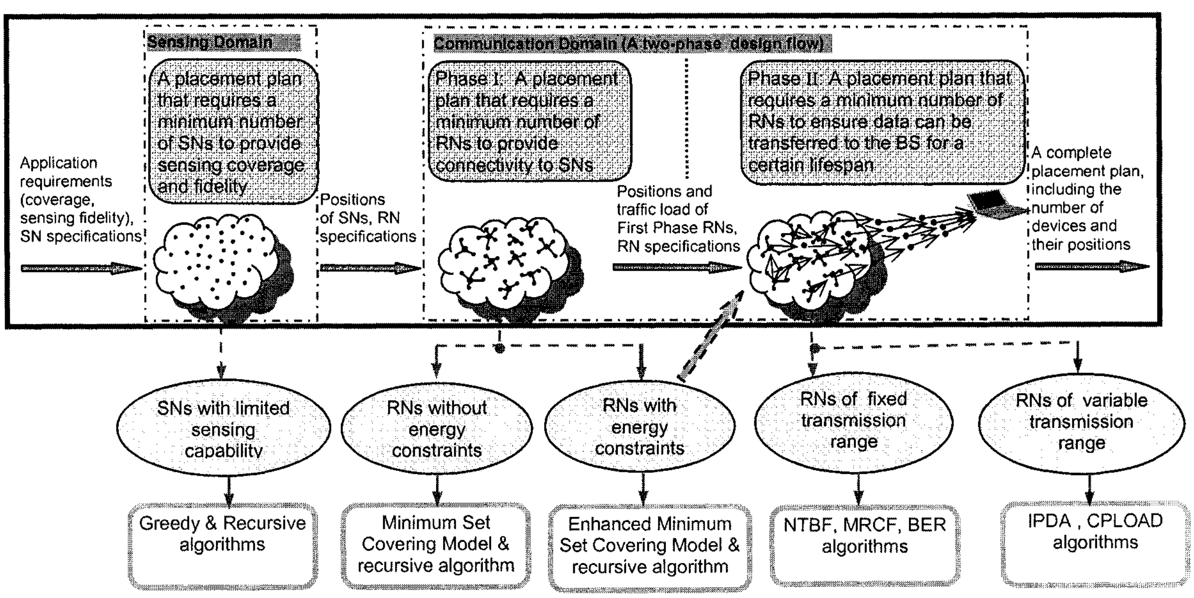


Figure 7.1 Networking Planning Framework, its Components, and Proposed Algorithms

The network planning starts from the sensing domain. The performance of the planning scheme for this domain is measured by the information obtained by the collection of SNs. Our research models the signal attenuation with distance that occurs for many physical signals in reality. In addition, the uneven distribution of information across the sensing field is explicitly considered. The planning schemes aim at maximizing the information gained by a given number of SNs which are deliberately placed according to the plan. Two heuristic schemes are proposed, namely, Greedy algorithm and Recursive algorithm. These schemes can apply to various signal attenuation models.

Taking the outcome from the sensing domain (the number and coordinates of SNs) as the input, the planning design for the communication domain aims at ensuring that the traffic generated at SNs will be relayed back to the BS for a given lifespan. Aiming at different objectives, the planning efforts are further divided into two phases.

The first phase aims at finding a placement plan that requires a minimum number of RNs to provide lifetime-guaranteed connectivity to SNs. Two design scenarios are addressed, namely, RNs have no energy constraints or RNs have energy constraints. For the former case, the planning problem is only concerned with connectivity of SNs. The problem is modeled by the minimum set covering model and a recursive algorithm is proposed to solve it. For the latter case, the planning problem also has to ensure RNs have enough capacity (in terms of energy) to last for a required lifespan. The problem is modeled by the enhanced minimum set covering model. The same recursive algorithm above can solve this problem.

The second phase aims at finding a placement plan that requires a minimum number of RNs to ensure data carried by RNs in the first phase can be transferred to the BS for a certain lifespan. Two design scenarios are considered, namely, RNs have fixed transmission range or RNs can adjust its transmission power. For the former case, an analytical lower bound on the number of RNs required is derived. Furthermore, three algorithms are proposed, namely, Nearest-To-BS-First (NTBF), Maximum-Residual-Capacity-First (MRCF), and Best-Effort-Relaying (BER). For the latter case, the design is more challenging as the transmission power optimization is involved. Two heuristic schemes are proposed, namely, Independent Placement with Direct Allocation (IPDA) and Collaborative Placement with Locally Optimal Allocation Decision (CPLOAD). A lower bound on the minimum number of RNs in this case is also derived.

Network planning is the first step of implementing a successful WSN application. Our research documented in this thesis represents some of the latest progress on this subject. The major contributions and significance of our work are as follows:

- We propose a comprehensive and modular network planning framework containing multiple phases in heterogeneous WSNs. This framework significantly and effectively simplifies the network planning effort and can be extended to a situation where more than two types of devices are involved. Thanks to the modularity of this framework, we can improve the algorithms for one phase with minimum effect on the other. In this way, design reuse is encouraged.

- We explore the differentiated information coverage problem while modeling the signal (emitted by the target) attenuation in the target detection. The model can better catch the physical characteristics of some physical signals, such as electromagnetic waves and acoustic waves, as compared to the commonly used binary detection model. We propose two novel information-oriented sensor placement algorithms, whose efficacy is proven by simulation. Additionally, these two algorithms are independent of the actual signal attenuation properties.

- We progressively design several RN placement schemes to provide lifetime-guaranteed connectivity, aiming at minimizing the overall system cost in two design phases. We explore the performance of these schemes by deriving bounds on the number of RNs required and simulating a large number of randomly generated networks.

7.2 Future Work

As network planning imposes significant impacts on the network performance, it is a fundamental issue of WSN engineering. Some future work that can build upon the accomplishments of this thesis is recommended as follows.

1) In future work, the performance of the schemes proposed in this thesis must be evaluated in a real sensor network and in various applications. A deep understanding of a network planning scheme in the context of a given networking protocol suite, for example, IEEE 802.15.4 [24] and ZigBee's specifications, should be gained from such evaluations.

2) Note that in the planning design of RN placements, the ideal routing paths and traffic allocation arrangements are also decided as part of the solutions. In order that the theoretically optimal results are achieved, routing schemes that decide the data paths consistent with the ideal data paths should be used. This is easily done if the routing paths are calculated in a centralized manner and broadcast to the RNs. However, existing distributed routing schemes which only make use of local information may not generate the ideal data paths or traffic distribution. Efforts should be invested in tackling the joint problem of device provisioning and design of distributed routing protocols, or of device provisioning under particular constraints imposed by a given distributed routing protocol.

3) Currently, our network planning design is based on two types of devices, and RNs in each phase are assumed identical in terms of energy supply and cost. However, devices with different power capacities (e.g., devices may be equipped with configurable initial energy supply), sensing and transmitting range, and computing/processing abilities may co-exist in a network in order to construct a more efficient and practical system. Therefore, the network planning problem with many types of devices and a more complicated cost function should be examined in the future.

4) Efficient numerical solutions for optimal information-based sensing and approximate solutions should be pursued, as well as studying how to adopt the proposed methodology to sensor scheduling in a densely deployed environment, which aims at selecting a group of

active sensors while putting all others to a sleep or off state to extend the lifetime of the whole system.

5) The device placement approaches presented in the literature focus on applications in two-dimensional fields. Optimal deterministic device placement in a spatially constrained three-dimensional environment such as a multi-storied building or a collection of tall trees is an interesting and challenging open problem.

6) Network planning affects other aspects of performance, such as reliability. Placement strategies that allow the network to keep functioning in spite of unexpected disruption from some devices should be studied too.

7) The research work described in this thesis represents some of the latest progress in the field of WSN planning. Network planning becomes more and more important as the technology advances and the market spreads. A commercially available network planning strategy, which implements the schemes presented here would greatly facilitate engineering efforts.

BIBLIOGRAPHY

[1] S. Madden, J. Hallerstein, and W. Hong, "TinyDB: In-Network Query Processing in TinyOS", http://telegraph.cs.berkeley.edu/tinydb/tinydb.pdf, 2003.

[2] E. Guizzo, "Into Deep Ice", *IEEE Spectrum*, vol. 42, pp. 28-35, 2005.

[3] N. Xu, S. Rangwala, K. K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin, "A Wireless Sensor Network for Structural Monitoring", *Proc. 2nd International Conference on Embedded Networked Sensor Systems (SenSys'04)*, pp. 13-24, 2004.

[4] J. Paek, K. Chintalapudi, R. Govindan, J. Caffrey, and S. Masri, "A Wireless Sensor Network for Structural Health Monitoring: Performance and Experience", *Proc. 2nd IEEE Workshop on Embedded Networked Sensors (EmNetS-II)* pp. 1-10, 2005.

[5] L. Krishnamurthy, R. Adler, P. Buonadonna, J. Chhabra, M. Flanigan, N. Kushalnagar, L. Nachman, and M. Yarvis, "Design and Deployment of Industrial Sensor Networks: Experiences from A Semiconductor Plant and the North Sea", *Proc. the 3rd International Conference on Embedded Networked Sensor Systems (ACM SenSys '05)*, pp. 64-75, 2005.

[6] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, and W. Hong, "A Macroscope in the Redwoods", *Proc. the 3rd international conference on Embedded networked sensor systems (ACM SenSys '05)*, pp. 51-64, 2005.

[7] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Network: a Survey", *Computer Networks (Elsevier) Journal*, vol. 38, pp. 393-422, 2002.

[8] Q. Wang, H. Hassanein, and K. Xu, "A Practical Perspective on Wireless Sensor Networks," in *chapter 9, Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems (Ilyas/Mahgoub, Eds)*: CRC Press, 2004.

[9] C.-Y. Wan, A. T. Campbell, and L. Krishnamurthy, "Pump-Slowly, Fetch-Quickly (PSFQ): A Reliable Transport Protocol for Sensor Networks", *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 862-872, 2005.

[10] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless Mesh Networks: a Survey", *Computer Networks*, vol. 47, pp. 445-487, 2005.

[11] A. Cerpa and D. Estrin, "ASCENT: Adaptive Self-configuring Sensor Networks Topologies", *Proc. IEEE INFOCOM'02*, pp. 1278-1287, 2002.

[12] A. Sinha and A. Chandrakasan, "Dymamic Power Management in Wireless Sensor Networks", *IEEE Design & Test of Computers*, vol. 18, pp. 62-74, 2001.

[13] A. Wang and A. Chandrakasan, "Energy Efficient System Partitioning for Distributed Wireless Sensor Networks", *Proc. IEEE ICASSP'01*, vol. 2, pp. 905-908, 2001.

[14] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative Signal and Information Processing: An Information Directed Approach", *Proceedings of IEEE*, vol. 91, pp. 1199-1209, 2003.

[15] K. Dasgupta, K. Kalpakis, and P. Namjoshi, "An Efficient Clustering-based Heuristic for Data Gathering and Aggregation in Sensor Networks", *Proc. IEEE WCNC'03*, vol. 3, pp. 1948-1953, 2003.

[16] J. N. Al-Karaki and A. E. Kamal, "Routing Techniques in Wireless Sensor Networks: A Survey", *IEEE Wireless Communications*, vol. 11, pp. 6-28, 2004.

[17] Y. Wei, J. Heidemann, and D. Estrin, "Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks", *IEEE/ACM Transaction on Networking*, vol. 12, pp. 493-506, 2004.

[18] 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.

[19] Y. T. Hou, Y. Shi, H. D. Sherali, and S. F. Midkiff, "On Energy Provisioning and Relay Node Placement for Wireless Sensor Networks", *IEEE Transactions On Wireless Communications*, vol. 4, pp. 2579-2590, 2005.

[20] Q. Wang, K. Xu, G. Takahara, and H. Hassanein, "On Lifetime-Oriented Device Provisioning in Heterogenous Wireless Sensor Networks: Approaches and Challenges", *IEEE Network*, vol. 20, pp. 26-33, 2006.

[21] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-Aware Wireless Microsensor Networks", *IEEE Signal Processing Magazine*, vol. 19, pp. 40-50, 2002.

[22] H. Dubois-Ferriere, R. Meier, L. Fabre, and P. Metrailler, "TinyNode: A Comprehensive Platform for Wireless Sensor Network Applications", *Proc. ACM IPSN'06*, pp. 358-365, 2006.

[23] V. Handziski, J. Polastre, J. H. Hauer, C. Sharp, A. Wolisz, and D. Culler, "Flexible Hardware Abstraction for Wireless Sensor Networks", *Proc. 2nd European Workshop on Wireless Sensor Networks (EWSN 2005)*, pp. 145-157, 2005.

[24] IEEE802.15.4, "IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks specific requirements part 15.4: wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs)", *IEEE Std 802.15.4-2003*, pp. 1-670, 2003.

[25] P. Levis, "TinyOS Programming", *http://csl.stanford.edu/~pal/pubs/tinyos-programming-1-0.pdf*, 2006.

[26] J. M. Rabaey, J. Ammer, J. L. d. Silva Jr, D. Patel, and R. Shad, "PicoRadio Supports Ad Hoc Ultra-Low Power Wireless Networking", *IEEE Computer Magazine*, vol. 33, pp. 42-48, 2000.

[27] A. Willig, R. Shah, J. Rabaey, and A. Wolisz, "Altruists in the PicoRadio Sensor Network", *Proc. 4th IEEE International Workshop on Factory Communication Systems*, pp. 175-184, 2002.

[28] G. Asada, M. Dong, T. S. Lin, F. Newberg, G. Pottie, W. J. Kaiser, and H. O. Marcy, "Wireless Integrated Network Sensors: Low Power Systems on a Chip", *Proc. the 24th IEEE European Solid-State Circuits Conference (ESSCEIC'98)*, pp. 9-16, 1998.

[29] G. J. Pottie and W. J. Kaiser, "Wireless Integrated Network Sensors", *Communications of the ACM*, vol. 43, pp. 51-58, 2000.

[30] R. Min, M. Bhardwaj, S.-H. Cho, E. Shih, A. Sinha, A. Wang, and A. Chandrakasan, "Low-Power Wireless Sensor Networks", *Proc. IEEE VLSID'01*, pp. 205-210, 2001.

[31] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next Century Challenges: Mobile Networking for "Smart Dust"", *Proc. ACM/IEEE MOBICOM'99*, 1999.

[32] K. S. J. Pister, "SMART DUST - Autonomous Sensing and Communication in a Cubic Millimeter", *Internet Article*, *http://www-bsac.eecs.berkeley.edu/~pister/SmartDust/.* 2003.

[33] P. Levis and e. al., "T2: A Second Generation OS for Embedded Sensor Networks", *Technical Report No. TKN-05-07, Telecommunication Networks Group, Technical University, Berlin*, 2005.

[34] Crossbow, *Crossbow Technology Inc, http://www.xbow.com/*, 2006.

[35] K. Xu, H. Hassanein, and G. Takahara, "Relay Node Deployment Strategies in Heterogeneous Wireless Sensor Networks: Multiple-Hop Communication Case", *Proc. IEEE SECON'05*, pp. 575-585, 2005.

[36] K. Xu, H. Hassanein, G. Takahara, and Q. Wang, "Relay Node Deployment Strategies in Heterogeneous Wireless Sensor Networks: Single Hop Communication Case", *Proc. IEEE GLOBECOM'05*, vol. 1, 2005.

[37] M. Younis, M. Bangad, and K. Akkaya, "Base-station Repositioning for Optimized Performance of Sensor Networks", *Proc. IEEE VTC2003*, vol. 5, pp. 2956-2960 Vol.5, 2003.

[38] S. R. Gandham, M. Dawande, R. Prakash, and S. Venkatesan, "Energy Efficient Schemes for Wireless Sensor Networks With Multiple Mobile Base Stations", *Proc. IEEE GLOBECOM'03*, vol. 1, pp. 377-381, 2003.

[39] J. Pan, L. Cai, Y. T. Hou, Y. Shi, and S. X. Shen, "Optimal Base-Station Locations in Two-tiered Wireless Sensor Networks", *IEEE Transactions on Mobile Computing,* , vol. 4, pp. 458-473, 2005.

[40] P. Maulin, R. Chandrasekaran, and S. Venkatesan, "Energy Efficient Sensor, Relay and Base Station Placements for Coverage, Connectivity and Routing", *Proc. IEEE IPCCC2005*, pp. 581-586, 2005.

[41] S. Musman, P. E. Lehner, and C. Elsaesser, "Sensor Planning for Elusive Targets ", *Journal of Computer and Mathematical Modeling*, vol. 25, pp. 103-115, 1997.

[42] S. S. Dhillon, K. Chakrabarty, and S. S. Lyengar, "Sensor Placement for Grid Coverage under Imprecise Detections", *Proc. International Conference on Information Fusion (FUSION'02)* vol. 2, pp. 1581-1587, 2002.

[43] A. Salhieh, "Power Efficient Topologies for Wireless Sensor Network", *Proc. International Conference on Parallel Processing*, pp. 156-168, 2001.

[44] L. Schwiebert, S. K. S. Gupta, and J. Weinamann, "Research Challenges in Wireless Networks of Biomedical Sensors", *Proc. ACM SIGMOBILE'01*, pp. 151-165, 2001.

[45] K. Chakrabarty, S. Iyengar, H. Qi, and E. Cho, "Grid Coverage or Surveillance and Target Location in Distributed Sensor Networks", *IEEE Transactions on Mobile Computers*, vol. 51, pp. 1448-1453, 2002.

[46] E. S. Biagioni and G. Sasaki, "Wireless Sensor Placement for Reliable and Efficient Data Collection", *Proc. IEEE HICSS'03*, pp. 127b, 2003.

[47] M. Cardei, D. MacCallum, X. Cheng, M. Min, X. Jia, D. Li, and D.-Z. Du, "Wireless Sensor Networks with Energy Efficient Organization", *Journal of Interconnection Networks*, vol. 3, pp. 213-229, 2002.

[48] H. Gupta, S. R. Das, and Q. Gu, "Connected Sensor Cover: Self-organization Sensor Networks for Efficient Query Execution", *Proc. ACM MOBICOM'03*, pp. 189-200, 2003.

[49] Z. Zhou, S. Das, and H. Gupta, "Connected k-Coverage Problem in Sensor Networks", *Proc. IEEE ICCCN'04*, pp. 373-378, 2004.

[50] X. Bai, S. Kumer, D. Xuan, Z. Yun, and T. Lai, "Deploying Wireless Sensors to Achieve Both Coverage and Connectivity", *Proc. ACM MobiHoc'06*, pp. 131-142, 2006.

[51] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks", *IEEE Tran. On Wireless Communications*, vol. 1, pp. 660-670, 2002.

[52] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data Gathering Algorithms in Sensor Networks Using Energy Metrics", *IEEE Tran. on Parallel and Distributed Systems*, vol. 13, pp. 924-935, 2002.

[53] S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, "Infrastructure Tradeoffs for Sensor Networks", *Proc. ACM WSNA'02*, pp. 49-58, 2002.

[54] N. Ahmed, S. S. Kanhere, and S. Jha, "The Holes Problem in Wireless Sensor Networks: A Survey", *ACM Mobile Computing and Communication Review*, vol. 9, pp. 4-18, 2005.

[55] K. Xu, G. Takahara, and H. Hassanein, "On the Robustness of Grid-based Deployment in Wireless Sensor Networks", *Proc. ACM IWCMC'06*, pp. 1183-1188, 2006.

[56] G. Wang, G. Cao, and T. L. Porta, "Movement-Assisted Sensor Deployment", *Proc. IEEE INFOCOM'04*, vol. 4, pp. 2469-2479, 2004.

[57] J.-J. Lee, B. Krishnamachari, and C.-C. J. Kuo, "Impact of Heterogeneous Deployment on Lifetime Sensing Coverage in Sensor Networks", *Proc. IEEE SECON'04*, pp. 367-376, 2004.

[58] Y. Liu and W. Liang, "Approximate Coverage in Wireless Sensor Networks", *Proc. IEEE LCN'05*, pp. 68-75, 2005.

[59] S. Ren, Q. Li, H. Wang, X. Chen, and X. Zhung, "A Study on Object Tracking Quality under Probabilistic Coverage in Sensor Networks", *Mobile Computing and Communication Review*, vol. 9, pp. 73-76, 2005.

[60] A. Howard, M. J. Mataric, and G. S. Sukhatme, "Mobile Sensor Network Deployment using Potential Fields: A Distributed, Scalable Solution to the Area Coverage Problem", *Proc. the 6th International Symposium on Distributed Autonomous Robotics Systems (DAR02)*, pp. 299-308, 2002.

[61] Y. Zou and K. Chakrabarty, "Sensor Deployment and Target Localization Based on Virtual Forces", *Proc. IEEE INFOCOM2003*, pp. 1293-1303, 2003.

[62] T. Clouqueur, V. Phipatanasuphorn, P. Ramanathan, and K. K. Saluja, "Sensor Deployment Strategy for Target Detection", *Proc. ACM WSNA'02*, pp. 42-48, 2002.

[63] Q. Wang, K. Xu, H. Hassanein, and G. Takahara, "Deployment for Information Oriented Sensing Coverage in Wireless Sensor Networks", *Proc. IEEE GLOBECOM'06, to appear*, 2006.

[64] K. Xu, G. Takahara, H. Hassanein, and Q. Wang, "A Generic Framework of Differential Deployment of Sensor Nodes in Wireless Sensor Networks", *Proc. IEEE GLOBECOM'06, to appear*, 2006.

[65] L. J. Guibas, "Sensing, Tracking, and Reasoning with Relations", *IEEE Signal Processing Magazine*, vol. 19, pp. 73-85, 2002.

[66] S. Ray and et al., "Robust Location Detection in Emergency Sensor Networks", *Proc. IEEE INFOCOM2003*, vol. 2, pp. 1044-1053, 2003.

[67] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the World with Wireless Sensor Networks", *Proc. IEEE ICASSP2001*, vol. 4, pp. 2033-2036, 2001.

[68] V. 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.

[69] X. Wang, W. Gu, S. Chellappan, K. Schosek, and D. Xuan, "Lifetime Optimization of Sensor Networks under Physical Attacks", *Proc. IEEE ICC'05*, pp. 3295-3301, 2005.

[70] J. F. Kurose and K. W. Ross, *Computer Networking, a Top-Down Approach Featuring the Internet*, 1st ed: Addison-Wesley Longman, 2000.

[71] 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.

[72] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed Energy Conservation for Ad Hoc Routing", *Proc. ACM MOBICOM'01*, pp. 70-84, 2001.

[73] C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. Srivastava, "Optimizing Sensor Networks in the Energy-Latency-Density Design Space", *IEEE Tran. on Mobile Computing*, vol. 1, pp. 70-80, 2002.

[74] F. Zhao, J. Shin, and J. Reich, "Information-driven Dynamic Sensor Collaboration for Tracking Applications," in *IEEE Signal Processing Magazine*, vol. 19, 2002, pp. 61-72.

[75] H. Qi, P. T. Kuruganti, and Y. Xu, "the Development of Localized Algorithm in Wireless Sensor Networks," in *Sensors Magazine*, vol. 2, 2002, pp. 286-293.

[76] O. Younis and S. Fahmy, "Distributed Clustering in Ad-hoc Sensor Networks: a Hybrid, Energy-efficient Approach", *Proc. IEEE INFOCOM'04*, pp. 629-640, 2004.

[77] W. Ye, J. Heidemann, and D. Estrin, "An Energy-Efficient MAC Protocol for Wireless Sensor Networks", *Proc. IEEE INFOCOM'02*, 2002.

[78] W. Heinzelman, "Application-specific Protocol Architecture for Wireless Networks," in *Electrical and Computer Engineering*, vol. Ph. D: Massachusetts Institute of Technology, 2000.

[79] A. Manjeshwar, Q. Zeng, and D. P. Agrawal, "An Analytical Model for Information Retrieval in Wireless Sensor Networks Using Enhanced APTEEN Protocol", *IEEE Tran. on Parallel and Distributed Systems*, vol. 13, pp. 1290-1302, 2002.

[80] S. Ghiasi, A. Srivastava, X. Yang, and M. Sarrafzadeh, "Optimal Energy Aware Clustering in Sensor Networks," in *Sensors Magazine*, vol. 19, 2002, pp. 258-269.

[81] S. Basagni, "Distributed Clustering Algorithm for Ad Hoc Networks", *Proc. International Symposium on Parallel Architectures, Algorithms and Networks (I-SPAN)*, pp. 310-315, 1999.

[82] S. Bandyopadhyay and E. J. Coyle, "An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks", *Proc. IEEE INFOCOM'03*, vol. 3, pp. 1713-1723, 2003.

[83] V. Mhatre and C. Rosenberg, "Design Guideline for Wireless Sensor Networks: Communication, Clustering and Aggregation", *Ad Hoc Networks Journal, Elsevier Science*, vol. 2, pp. 45-63, 2004.

[84] C. F. Chiasserini, I. Chlamtac, P. Monti, and A. Nucci, "Energy Efficient Design of Wireless Ad Hoc Networks", *Proc. Networking2002*, pp. 387-398, 2002.

[85] Q. Wang, K. Xu, H. Hassanein, and G. Takahara, "SWATCH: a StepWise AdapTive Clustering Hierarchy for Energy-Efficient Communication in Wireless Sensor Network (WSNs)", *Proc. IFIP Networking2005*, 2005.

[86] D. J. Baker and A. Ephremides, "The Architecture Organization of A Mobile Radio Network via A Distributed Algorithm", *IEEE Tran. on Communications (Legacy, pre 1988)*, vol. 29, pp. 1694-1701, 1981.

[87] A. Ephremides, J. E. Wieselthier, and D. J. Baker, "A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling", *Proceedings of IEEE*, vol. 75, pp. 56-73, 1987.

[88] M. Gerla and J. T. C. Tsai, "Multi Cluster, Mobile, Multimedia Radio Network", *Wireless Networks*, vol. 1, pp. 255-265, 1995.

[89] A. Safwat, H. Hassanein, and H. T. Mouftah, "Power-Aware Virtual Base Stations (PW-VBS) for Wireless Mobile Ad Hoc Communications", *the Journal of Computer Networks*, vol. 41, pp. 331-346, 2003.

[90] M. Chatterjee, S. K. Das, and D. Turgut, "WCA: A Weighted Clustering Algorithm for Mobile Ad Hoc Networks", *Journal of Clustering Computing, special issue on mobile ad hoc networking*, vol. 5, pp. 1448-1453, 2002.

[91] B. Aoun and R. Boutaba, "Clustering in WSN with Latency and Energy Consumption Constraints", *Journal of Network and Systems Management*, vol. 14, pp. 415-439, 2006.

[92] A. D. Amis, R. Prakash, T. H. P. Vuong, and D. T. Huynh, "Max-min D-cluster Formation in Wireless Ad Hoc Networks", *Proc. IEEE INFOCOM2000*, vol. 1, pp. 32-41, 2000.

[93] S. Banerjee and S. Khuller, "A Clustering Scheme for Hierarchical Control in Multi-hop Wireless Networks", *Proc. IEEE INFOCOM'01*, vol. 2, pp. 1028-1037, 2001.

[94] G. Gupta and M. Younis, "Load-balanced Clustering of Wireless Sensor Networks", *Proc. IEEE ICC'03*, vol. 3, pp. 1848-1852, 2001.

[95] H. Chan and A. Perrig, "ACE: An Emergent Algorithm for Highly Uniform Cluster Formation", *Proc. the First European Workshop on Sensor Networks (EWSN)*, 2004.

[96] S. G. Foss and S. A. Zuyev, "On a Voronoi Aggregative Process Related to a Bivariate Poisson Process", *Advances in Applied Probability*, vol. 28, pp. 965-981, 1996.

[97] F. Baccelli and S. Zuyev, "Poisson Voronoi Spanning Trees with Applications to the Optimation of Communication Networks", *Operations Research*, vol. 47, pp. 619-631, 1999.

[98] Q. Wang, H. Hassanein, and G. Takahara, "Stochastic Modeling of Distributed, Dynamic, Randomized Clustering Protocols for Wireless Sensor Networks", *Proc. ICPP2004 Workshops*, pp. 456-463, 2004.

[99] B. Krishnamachari, D. Estrin, and S. Wicker, "Impact of Data Aggregation in Wireless Sensor Networks", *Proc. International Workshop on Data Aggregation in Wireless Sensor Networks*, pp. 575-578, 2002.

[100] J.-H. Chang and L. Tassiulas, "Maximum Lifetime Routing in Wireless Sensor Networks", *IEEE/ACM Trans. on Networking*, vol. 12, pp. 609-619, 2004.

[101] S. R. Gandham, M. Dawande, and R. Prakash, "Hop-Constrained Energy-Aware Routing in Wireless Sensor Networks", *Proc. IEEE GLOBECOM'05*, vol. 5, pp. 3112-3116, 2005.

[102] R. R. Benkoczi, H. Hassanein, S. Akl, and S. and Tai, "Data Relaying with Constraints in Hierarchical Sensor Networks", *Journal of Combinatorial Optimization*, pp. 113-123, 2006.

[103] S. Tai, R. Benkoczi, H. Hassanein, and S. Akl, "An Energy Consumption Study of Wireless Sensor Networks with Delay-Constrained Traffic", *Proc. IEEE GLOBECOM'06, to appear*, 2006.

[104] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, "Protocols for Self-organization of A Wireless Sensor Network", *IEEE Personal Communications*, vol. 7, pp. 16-27, 2000.

[105] A. Boukerche, X. Cheng, and J. Linus, "Energy-Aware Data-Centric Routing in Microsensor Networks", *Proc. The 6th International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems*, 2003.

[106] F. Ye, A. Chen, S. Liu, and L. Zhang, "A Scalable Solution to Minimum Cost Forwarding in Large Sensor Networks", *Proc. IEEE ICCCN'01*, pp. 304-309, 2001.

[107] S. Servetto and G. Barreneschea, "Constrained Random Walks on Random Graphs: Routing Algorithms for Large Scale Wireless Sensor Networks", *Proc. 1st ACM International Workshop of Wireless Sensor Networks and Applications*, pp. 12-21, 2002.

[108] A. Manjeshwar and D. P. Agrawal, "TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks", *Proc. IEEE IPDPS'02 workshop*, pp. 2009-2015, 2002.

[109] T. Fuhrmann, "Scalable Routing for Networked Sensors and Actuators", *Proc. IEEE SECON'05*, pp. 240-251, 2005.

[110] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications", *Proc. ACM SIGCOMM'01*, pp. 149-160, 2001.

[111] A. Savvides, C.-C. Han, and M. Srivastava, "Dynamic Fine-Grained Localization in Ad-Hoc Networks of Sensors", *Proc. ACM MobiCom'01*, pp. 166-179, 2001.

[112] V. Rodoplu and T. H. Meng, "Minimum Energy Mobile Wireless Networks", *IEEE JSAC*, vol. 17, pp. 1333-1344, 1999.

[113] L. Li and J. Y. Halpern, "Minimum-Energy Mobile Wireless Networks Revisited ", *Proc. IEEE ICC'01*, pp. 278-283, 2001.

[114] A. S. Tanenbaum, *Computer Networks*: Prentice-Hall, 1996.

[115] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks", *Proc. IEEE/ACM MobiCom'99*, pp. 174-185, 1999.

[116] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-Based Protocols for Disseminating Information in Wireless Sensor Networks", *Wireless Networks*, vol. 8, pp. 169-185, 2002.

[117] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks", *Proc. ACM/IEEE MobiCom'00*, pp. 56-67, 2000.

[118] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion for Wireless Sensor Networking", *IEEE/ACM Trans. on Networking*, vol. 11, pp. 2-16, 2003.

[119] J. Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan, "Building Efficient Wireless Sensor Networks with Low-Level Naming", *Proc. 18th ACM Symposium on Operating Systems Principles*, pp. 146-159, 2001.

[120] C. Schurgers and M. B. Srivastava, "Energy Efficient Routing in Wireless Sensor Networks", *Proc. MILCOM on Communications for Network-Centric Operations: Creating the Information Force, 2001.*, 2001.

[121] R. Shah and J. Rabaey, "Energy Aware Routing for Low Energy Ad Hoc Sensor Networks", *Proc. IEEE WCNC'02*, pp. 350-355, 2002.

[122] H. Dubois-Ferriere and D. Estrin, "Efficient and Practical Query Scoping in Sensor Networks," Tech. Report # 39, CENS/UCLA 2004.

[123] Q. Wang and H. Hassanein, "A Comparative Study of Energy Efficient (E2) Protocols for Wireless Sensor Networks," in *chapter 18, Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems (Ilyas/Mahgoub, Eds)*: CRC Press, 2004.

[124] C. Intanagonwiwat, D. Estrin, and R. Govindan, "Impact of Network Density on Data Aggregation in Wireless Sensor Networks", *Technical Report 01-750, University of Southern California, November*, 2001.

[125] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Efficient Algorithm for Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks", *Technical Report UMBC-TR-02-13, Computer Science and Electrical Engineering Department, University of Maryland, Baltimore, County*, 2002.

[126] J. Considine, F. Li, G. Kollios, and J. Byers, "Approximate Aggregation Techniques for Sensor Databases", *Proc. 20th International Conference on Data Engineering (ICDE2004)*, pp. 449-460, 2004.

[127] A. Boulis, S. Ganeriwal, and M. B. Srivastava, "Aggregation in Sensor Networks: An Energy-Accuracy Trade-off", *Proc. IEEE workshop on Sensor Network Protocols and Applications*, pp. 128-138, 2003.

[128] S. Kumar, F. Zhao, and D. Shepherd, "Collaborative Signal and Information Processing in Microsensor Networks", *IEEE Signal Processing Magazine*, vol. 19, pp. 13-14, 2002.

[129] H. Qi, S. S. Iyengar, and K. Chakrabarty, "Multi-Resolution Data Integration Using Mobile Agents in Distributed Sensor Networks", *Multi-Resolution Data Integration Using Mobile Agents in Distributed Sensor Networks*, vol. 31, pp. 383-391, 2001.

[130] J. Gutiérrez, E. H. Callaway, and R. L. Barret, *Low-Rate Wireless Personal Area Networks: Enabling Wireless Sensors with IEEE 802.15.4*: Standards Information Network, IEEE Press, 2003.

[131] A. J. Goldsmith and S. B. Wicker, "Design Challenges for Energy-Constrained Ad Hoc Wireless Networks", *IEEE Wireless Communications*, vol. 9, pp. 8-27, 2002.

[132] A. Safwat, H. Hassanein, and H. Mouftah, "Optimal Cross-Layer Designs for Energy-Efficient Wireless Ad Hoc and Sensor Networks", *Proc. IEEE IPCCC'03* pp. 123-128, 2003.

[133] T. He, S. Krishnamurthy, J. A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh, "Energy-Efficient Surveillance System Using Wireless Sensor Networks", *Proc. ACM MobiSYS'04*, 2004.

[134] Y. Xu and W.-C. Lee, "On Localized Predication for Power Efficient Object Tracking in Sensor Networks", *Proc. 1st Int'l Workshop on Mobile Distribued Computing (MDC)* pp. 434-439, 2003.

[135] ZigBee, http://www.zigbee.org, 2006.

[136] N. Ahmed, S. S. Kanhere, and S. Jha, "Probabilistic Coverage in Wireless Sensor Networks", *Proc. IEEE LCN2005*, pp. 672-679, 2005.

[137] S.-P. Kuo, Y.-C. Tseng, F.-J. Wu, and C.-Y. Lin, "A Probabilistic Signal-Strength Based Evaluation Methodology for Sensor Network Deployment", *Int. J. Ad Hoc and Ubiquitous Computing*, vol. 1, pp. 3-12, 2005.

[138] T. S. Rappaport, *Wireless Communications: Principles and Practice*: Prentice Hall, Inc., 1996.

[139] S. Adlakha and M. Srivastava, "Critical Density Thresholds for Coverage in Wireless Sensor Networks", *Proc. IEEE WCNC'03*, pp. 1615-1620, 2003.

[140] J.-J. Lee, B. Krishnamachari, and C.-C. J. Kuo, "Impact of Heterogeneous Deployment on Lifetime Sensing Coverage in Sensor Networks", *Proc. IEEE SECON'04*, pp. 367-376, 2004.

[141] T. Yan, T. He, and J. Stankovic, "Differentiated Surveillance for Sensor Networks", *Proc. ACM SenSYS'03*, pp. 51-62, 2003.

[142] Q. Wang, K. Xu, H. Hassanein, and G. Takahara, "Minimum Cost Guaranteed Lifetime Design for Heterogeneous Wireless Sensor Networks", *Proc. IEEE IPCCC'05*, pp. 599-604, 2005.

[143] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed: the MIT Press, 2001.

[144] K. Xu, Q. Wang, H. Hassanein, and G. Takahara, "Optimal wireless sensor networks (WSNs) deployment: minimum cost with lifetime constraint", *Proc. IEEE WiMob2005*, vol. 3, pp. 454-461, 2005.

[145] C.-F. Huang and Y.-C. Tseng, "The Coverage Problem in a Wireless Sensor Network", *Mobile Networks and Applications*, vol. 10, pp. 519-528, 2005.

[146] I. Stojmenovic and X. Lin, "Power Aware Localized Routing in Wireless Networks", *IEEE Trans. on Parallel and Distributed Systems*, vol. 12, pp. 1122-1133, 2001.