Quality of service support for event detection in wireless sensor network

Yanjun Li

Nancy-Université
Université Henri Poincaré

Département des Sciences et Technologies
UFR STMIA

École doctorale IAEM Lorraine

Support de la Qualité de Service dans les Réseaux de Capteurs sans Fil pour la Détection D'événements (Quality of Service Support in Wireless Sensor Networks for Event Detection)

THÈSE

présentée et soutenue publiquement le 10 novembre 2010

pour l'obtention du

Doctorat de l'université Henri Poincaré – Nancy 1

(spécialité informatique)

par

LI Yanjun

Composition du jury

Rapporteurs :	GUYENNET Hervé, Professor, Université de Franche-Comté, France
	LOU Wei, Assistant Professor, The Hong Kong Polytechnic University, China
Examinateurs :	WANG Zhi, Associate Professor, Zhejiang University, China
	HU Fei, Professor, Shanghai Jiaotong Univerisity, China
	SCHOTT René, Professor, UHP Nancy 1, France
	SONG Ye-Qiong, Professor, INPL, France

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Loria

Mis en page avec la classe thloria.

Remerciements

It takes a tremendous amount of dedication and perseverance to complete a doctoral dissertation. At least that is what I have been told. I am not sure I possess either of those qualities in great enough quantity to complete a dissertation, but I am fortunate in that I have been constantly surrounded by many great and motivated people whose encouragement and assistance throughout this process has made up for any attributes I lack. I would like to sincerely extend my gratitude to those individuals here.

First and foremost I would like to thank my supervisors in France, Prof. Ye-Qiong Song and Prof. René Schott, for everything I have learned from them and for their infinite kindness and patience during my Ph.D. studies. I also thank my supervisors in China Prof. Zhi Wang and Prof. Youxian Sun, for the helpful discussions we had about my thesis and many other wireless network topics. Special thank to Dr. Chung Shue (Calvin) Chen, he has been working with me on quite a few research topics and I appreciate very much for all his help in shaping my thoughts and ideas. I am grateful to all the LORIA faculty, fellow graduate students, and staff for creating a great research environment and a fun place to work. Specially, I would like to thank the administrative assistant of TRIO group Ms. Laurence Benini, for her kindly help on all the tedious administrative things.

I would like to thank my jury, consisting of Prof. Hervé Guyennet, Prof. Wei Lou, Prof. Zhi Wang, Prof. Fei Hu, Prof. Ye-Qiong Song and Prof. René Schott, for their encouragement and discussions about my thesis.

It is my fortune being a member in TRIO group. I am grateful to all the group mates for being the surrogate family during the year I stayed there and for their continued moral support there after. From the staff, Boughami Najet, Maha Idrissi Aouad, Liliana Cucu, Bilel Nefzi, Shahram Nourizadeh, Dr. Jian Li, Dr. Ning Jia are especially thanked for their care and attention.

My friends in Nancy have helped a lot at various stage of my graduate career. I would like to thank Tingting Ding, Yuting Shi, Ping Lan, Xiaobai Zhou, Xianqing Mao, Dr. Wan Zhang, Dr. Dong Cheng, Dr. Weiming Dong, Dr. Jinsan Chen and Dr. Yun Wang, for helping me to solve many problems for my study or for my life.

I am immensely grateful to my parents and my husband for everything they have always been doing for me, and especially for all their help in the last few months. They took care of my life when I was preparing my thesis in Hangzhou, without them, it would not have been finished on time.

To my parents LI Xiaobai and GU Qinhua, and my husband YANG Hong.

Table des matières

1

Introduction

1.1 Background

1.1.1 Wireless Sensor Networks

Wireless sensor networks (WSNs) have been one of the main research focus in wireless network area over the last decade, thanks to the evolution of the MEMS technology and the availability of low-power communication and computation hardware. WSNs are comprised of a large number of sensor devices that can sense the physical parameters (temperature, light, humidity, acceleration etc.) and communicate with each other via wireless channels, with limited energy and computing capability. Local sensed data are finally transmitted either to a base station (BS) which people may access via Internet, or directly to actuators which conduct actions in response. The latter is also referred to as wireless sensor and actuator networks (WSANs), which is an important extension to WSNs. WSNs are a fundamental part of the networking infrastructure for pervasive computing proposed by Mark Weiser [1] : "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it". This computing paradigm is expected to enable people to compute and communicate anytime and anywhere, meanwhile be gracefully integrated with human users. In WSN-based applications, sensor nodes can be elaborately manufactured and embedded in walls, clothes, human bodies or stick on machines, etc., to gather user-interested information, such as structural health, heart-beat rate or blood-pressure or body gesture of a person, stress or movement of a rotating machinery. These data provide the fundamental information of users' context and the underlying execution environment, which makes pervasive computing available. In a sense, WSNs have effectively bridged the gap between the physical world and the information world, profoundly transformed the way people interact with the nature.

1.1.2 Types of Applications

WSNs are being employed or have potential in a variety of applications, e.g. habitat and environment monitoring [2], medical care [3], military surveillance [4], industrial process control [5], asset tracking [6], home automation [7], as well as smart and interactive places [8]. Some of these applications share some basic characteristics. According to different interaction patterns, the type of applications can summarized as follows [9] :

– **Periodic measurements** :
Sensors periodically sense the surroundings and report the measured values to the BS. The reporting period can be decided by requirement of the application. Periodic measurements will trigger a event detection mode once a specific event is observed. An example is to attach wearable sensor nodes to the patients [3]. The physiological parameters like temperature and blood pressure are periodically measured for a long-term care.

– **Event detection** :
Sensor nodes report to the BS once they have detected the occurrence of a specified event. The simplest events can be detected locally by a single sensor node in isolation, e.g. a temperature threshold is exceeded ; more complicated types of events require the collaboration of nearby or even remote sensors to decide whether a event has occurred, e.g. a temperature gradient becomes too steep. If several different events can occur, event classification might be an additional issue.

– **Function approximation and border detection** :
The way a physical value like temperature changes from one place to another can be regarded as a function of location. A WSN can be used to approximate this unknown function, using a limited number of samples taken at each individual sensor node [10]. Similarly, a relevant problem can be to find areas or points of the same given value. An example is to find the isothermal points in a forest fire application to detect the border of the actual fire. This can be generalized to finding “edges” in such functions or to sending messages along the boundaries of patterns in both space and/or time.

– **Tracking** :
The source of an event can be mobile, e.g. an intruder in surveillance scenarios. The WSNs can be used to report updates on the event source’s position to the BS, potentially with estimates about speed and direction as well. To do so, typically sensor nodes have to cooperate before updates can be reported to the BS.

– **Control** :
In addition to the sensing capability, WSNs are also in some applications responsible for decision making and action taking. In such networks, the sensor nodes are not only a part of the passive identification process but also part of the active response procedure as they are closely integrated with the control system. These kinds of applications are referred to as control applications as they represent WSNs (or WSANs) in which sensing and actuating terms are blended together.

1.2 Motivation of Research on Quality of Service (QoS) in WSNs

Major research efforts in WSN area have resulted in many deployed testbeds and implementations. However, most of the current solutions and systems can only provide best effort service. Minimization of energy consumption or achieving "high efficiency" has been the objective of many communication protocols designed for WSNs. Their performance strictly depends on the configuration of the network and the load it carries, which cannot always be guaranteed to a user-satisfied level. While the boosting efforts of existing solutions on the research community should be acknowledged, these solutions fall short of addressing the specific requirements of the applications and cannot provide different levels of service. More specifically, real-time applications suffer from unpredictable performance levels when best-effort protocols and strategies are used.

Real-time and mission-critical applications require performance guarantee from the system on which they are implemented. As an example, a real-time intrusion detection application running on WSNs may require the event be captured within a given delay bound. A fire monitoring application may require the fire be detected with certain accuracy and delay bound. Other applications may require the network deliver the data packets to the sink with a given success probability. Similarly, the overall energy consumption of all communication events may be subject to system's energy budget. Applications running on multimedia WSNs, which form a new emerging class of WSNs, inherently require service guarantees due to real-time nature of its multimedia source content. These requirements are classified as QoS in wired and wireless networks, which we also adopt for WSN environment. We classify a protocol as QoS-based protocol if it can guarantee one or more performance metrics to upper layer or the application. Under this classification, solutions that simply minimize or maximize a performance metric (delay, energy consumption, packet loss probability, network lifetime etc.) without certain performance guarantees are considered non-QoS-based solution. We should note that most of the existing proposals for WSNs fall into non-QoS-based category.

1.3 Problems and Challenges

QoS has been the target of many communication protocols for numerous network types. In its broadest form, QoS refers to contract between the service provider and the customer. In wired networks, one of the main motivations for QoS solutions is the real-time multimedia applications that need bandwidth, delay, and jitter guarantees. ATM networks were proposed to support such requirements from ground up. Although ATM networks are not as widely in use as originally imagined, QoS support mechanisms proposed for ATM networks still inspire new solutions. In IP networks, QoS support of individual flows and differentiation of flows have been proposed to be handled through IntServ and DiffServ mechanism. In cellular networks, the motivation for QoS support is also inherent to the primary application of voice and video calls. In these networks, QoS support are provided through resource reservation mechanisms.

Providing QoS in WSNs differs dramatically from the traditional wired network systems due to its wireless nature. Furthermore, considering very limited resources in sensor nodes and the large scale of WSNs, per flow resource reservation-based approaches are especially ill-suited for WSNs. The following unique features and shortcomings of WSNs suggest new directions in QoS provisioning :

– **Resource constraints** : A typical WSN node is small in physical size and battery powered. Thus the components, such as Micro-Controller Unit (MCU), data and code memories, transceiver, and sensing unit, are selected to minimize their area and power consumption. This implies that computation, communication, and memory resources in nodes are very limited. Further, batteries are typically small-sized ; thus energy resources are also scarce. All these constraints require the QoS mechanism implemented in WSNs be simple and efficient.

– **Unbalanced traffic and energy consumption** : In most applications, traffic mainly flows from a large number of source nodes to a single or a small subset of sink nodes. The traffic and energy consumption is uneven throughout the network. This requires the designed QoS mechanisms to be able to handle unbalanced traffic and energy consumption, and prevent highly loaded nodes from dying prematurely.

– **Dynamic topology** : WSNs are prone to frequent topology changes. This is due to several reasons, such as node failure or addition, depleted batteries, intermittent radio interference, environmental factors. Applications require a degree of inherent fault tolerance and the ability to reconfigure themselves as the network topology evolves over time.

– **Scalability** : WSNs consisting of tens of thousands of nodes are envisioned. Further, the density of nodes can be high. These aspects depend on the requirements of the application for sensing coverage and robustness (redundancy). Thus, algorithm scalability is an important design criterion for WSN applications, i.e., QoS support should not degrade quickly when the number of nodes or the node density increases.

– **Distributed data processing** : Proper in-network data aggregation and fusion strategy can be a good complement to QoS protocols in reducing data redundancy and alleviating network congestion.

– **Service differentiation** : The content of data or high-level description reflects the criticality of the real physical phenomena and is thereby of different criticality or priority with respect to the quality of the applications. QoS mechanisms may be required to differentiate packet importance and set up a priority structure.

– **Probabilistic QoS guarantee** : Since the wireless channel is random and time-varying, it is difficult to provide deterministic guarantee in WSNs as it does in wired networks. Conventional deterministic QoS measures should be replaced by probabilistic ones.

Therefore, considering the problems and challenges of WSNs, while we design QoS-enabled algorithms, protocols and systems, at least a few of the above issues should be taken into account.

1.4 Architecture of QoS Support in WSNs

1.4.1 Protocol Layer Based QoS Support

Although there is no general consensus within the WSN research community about the layer structure in WSNs, a layered communication stack like TCP/IP model is commonly used. From a layered view, each protocol layer builds on the foundation provided by the lower layers, as show in Figure 1.1. Upper protocol layers define the requirements that ultimately come from the application layer. Based on the available capabilities, the upper layer controls the operation of the lower layer by selecting the level of service. The lower layer provides feedback of its operation, such as link failures, which allows the upper layer to adjust its operation accordingly. In the following, we discuss QoS metrics in each layer and review existing QoS mechanisms and protocols.

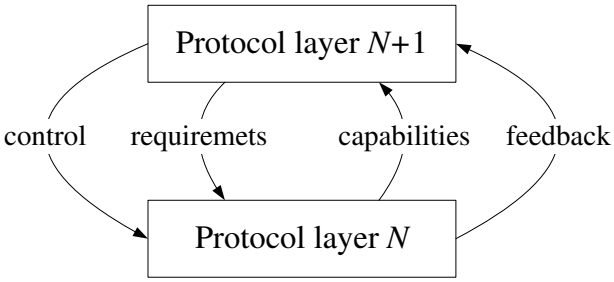


FIGURE 1.1: Protocol layer based QoS architecture.

Application layer :

The application layer has the best knowledge about the QoS requirements. Therefore, an application passes the QoS requirement to lower layers by associating a generated packet with the QoS requirements. The network aims to fulfill these requirements and take actions in each layer respectively. The application layer is also responsible for making the sensor measurements and controlling sensors. As such, an application may configure sensors based on measurement accuracy versus time interval, thus making a tradeoff between the data accuracy and energy usage.

Transport layer :

The two main tasks for the transport layer are congestion control and reliable transmission. Typically, congestion control limitations the sending of traffic to reduce the utilized bandwidth. As the congestion is reduced, the overall reliability in the network is increased because the data link layer does not have to drop frames. However, throughput limitations may increase delays as the source node must hold onto the generated packets that much longer. Therefore, QoS awareness is required to make a decision regarding which traffic is being more limited than other traffic. The energy efficiency of a transport protocol depends on the number of transmissions required to deliver a message. This is greatly affected by the acknowledgment scheme being utilized. With the positive acknowledgment scheme, the receiver acknowledges each packet. If

an acknowledgment is not received within a certain time interval, the packet is resent. With the negative acknowledgment scheme, missing packets are requested by the receiver node. This reduces messaging as acknowledgments are sent only when required. However, the drawback of the scheme is that the sender does not know when a packet is successfully transmitted and must therefore buffer sent packets longer than with positive acknowledgments.

Due to resource constraint, WSNs rarely implement the transport layer. Instead, the transport layer functionality is often combined with the other layers. Pump-Slowly, Fetch-Quickly (PSFQ) [11] is a WSN communication protocol that allows reliable data transfer with low communication cost. To allow operation on resource-constrained sensor nodes, PSFQ combines the functionality of the transport and network layers. The Reliable Multisegment Transport (RMST) protocol [12] is designed toward guaranteed delivery of large blocks of data from sensors to sinks. It is tightly integrated with directed diffusion. The Event-to-Sink Reliable Transport (ESRT) protocol [13] considers the situation where the sink node does not care which nodes send packets ; it is only interested in receiving a sufficient number of packets to present reliable and credible information to the user. This is called event-to-sink reliability. ESRT works by carefully adjusting the sensor's reporting rate to make a sufficient number of packets be received, and not many more packets than needed be received to avoid congestion and save energy.

Network layer :

The network layer controls QoS with traffic shaping and routing protocol. The traffic shaping performs congestion control by classifying packets and providing queuing disciplines that provide per class QoS and fairness. For example, a node may drop low-priority traffic to ensure enough resources for higher priority data. The routing protocol is responsible for selecting an end-to-end routing path fulfilling the desired QoS characteristics. As a route that satisfy one QoS metric may not be optimal to others, the route selection has to make a trade-off between different QoS metrics. The shortest path route is not always the most energy efficient. Because the transmission power requirement is proportional to the square of the distance, it might be more energy efficient to forward data through two short hops than through one long hop. As longer routes usually have higher latency, the route selection therefore has to make a tradeoff between energy and delay. The routing protocol design can make the selection between maintenance and routing energy consumption, thus determining how dynamic the protocol is. Generally, the geographic routing techniques provide the best adaptability to network dynamics, and is favorable to real-time design, while the table-based routing techniques are more energy efficient in static networks.

SPEED [14] uses stateless nondeterministic geographic forwarding and guarantees a certain delivery speed for a packet. In SPEED, the next hop is selected randomly among the neighbors with the probability that is proportional to the link speed. Only the nodes that advance towards the target and meet the delivery time are included in the selection. The link speed is calculated by dividing the distance between nodes by measured link delay. The next-hop selection is combined with feedback received from neighbors. A backpressure packet re-routing around large-delay links is used to reduce or divert the traffic injected to a congested area. SPEED is especially suitable for real-time data delivery, but it does not include any other QoS metrics in its routing decisions. MMSPEED [15] is an extension of SPEED which supports service dif-

ferentiation and probabilistic QoS guarantee. For delivery timeliness, multiple network-wide packet delivery speed options are provided for different traffic types according to their end-to-end deadlines. In supporting service reliability, probabilistic multi-path forwarding is used to control the number of delivery paths based on the required end-to-end reaching probability. These methods are implemented in a localized way with dynamic compensation for the inaccuracies of local decisions. Like SPEED, since all mechanisms in MMSPEED work locally without global network state information and end-to-end path setup, it is scalable and adaptive to network dynamics. However, both SPEED and MMSPEED have a common deficiency : energy consumption metric has not been taken into account. A real-time power-aware routing (RPAR) protocol [16] is proposed to achieve application specified communication delay at low energy cost by dynamically adjusting transmission power and routing decisions. It allows the application to control the tradeoff between energy consumption and communication delay by specifying packet deadlines. Important practical issues like lossy links, memory and bandwidth constraints and scalability are considered. Since RPAR adjusts the transmission power from time to time, the network topology may change frequently. Thus it employs a novel neighborhood management mechanism which is more efficient than the periodic beacon scheme adopted by SPEED and MMSPEED.

Besides, security guarantee is getting more and more important in WSN. Several types of attacks can be executed on the network layer, such as black hole attack, misdirections, neglect and greed, homing etc. In [17], attacks in the network layer and secure routings as countermeasures are discussed in detail.

MAC layer :

The MAC layer is responsible for dividing bandwidth to the traffic based on their priority and QoS requirements. TDMA-based MAC protocols can naturally implement per flow throughput guarantees, because the slot usage of each node is negotiated separately. On the other hand, CSMA can provide per class bandwidth differentiation by assigning each priority with different backoff times. The flexibility of the bandwidth usage is another factor. In this respect, CSMA is better as nodes may use bandwidth as required. CSMA has the further benefit of having smaller delays in lightly loaded networks as a node does not have to wait for its slot to transmit. However, when the network congests, delay and throughput get worse due to the increased number of collisions. The MAC mainly controls reliability with the used retransmission scheme, but also by avoiding the collisions and hidden node problems. In addition, the adjustment of transmission power affects reliability. High transmission power enables more reliable transmission, but on the other hand, might cause additional interference within a network.

Some commonly used MAC protocols, for example, S-MAC [18], T-MAC [19] and B-MAC [20], which reduce delay in a best effort approach. Q-MAC [21] is a CSMA-based sensor MAC that priorizes frames so that higher priority packets experience lower latencies. The distributed scheduling between nodes is performed by adjusting the contention window according to the traffic priority, thus giving higher priority frames faster channel access. The priority is calculated from an application-defined importance rating, the number of transmitted hops, residual energy and the proportional load of the queue. In simulations, Q-MAC is able to differentiate latency between priorities while using the same amount of energy and having comparable throughput as S-MAC. An implicit prioritized access protocol (I-EDF) [22] is designed espe-

cially for HRT-WSNs. A cellular backbone network is adopted and different frequency channels are assigned. In a cell, time is divided into frames and all nodes are frame synchronized and follow earliest deadline first (EDF) schedule for packet transmission to guarantee bounded delay. A capable router node is required at the center of each cell and equipped with two transceivers for separate transmission and reception. Inter-cell communication is supported by a globally synchronized TDMA scheme and the messages are ordered by their earliest deadlines too. The mixed FDMA-TDMA scheme offers a collision-free solution. Simulations show that I-EDF can provide high throughput and low latency even in heavy loads. However, the system architecture and requirements appear impractical for WSNs. Nodes are assumed synchronized. Routers need to be deployed specifically following the cellular structure and topology knowledge is required. Z-MAC [23] is a hybrid MAC protocol which dynamically switches between CSMA and TDMA depending on the level of contention. It uses the knowledge of topology and a loosely synchronized clock as hints to improve MAC performance under high contention. When the hints are not obvious, it just behaves like CSMA. Generally, Z-MAC outperforms B-MAC under medium to high contention while it is a little worse under low contention. Although Z-MAC is not specifically designed for RT service, the idea of switching between TDMA and CSMA based protocols is inspiring. IEEE 802.15.4 standard specifies the physical layer (PHY) and MAC sublayer for low-data-rate low-cost wireless personal area networks (WPANs) of fixed, portable, and moving devices with no battery or very limited energy consumption requirements [24]. It supports star as well as peer-to-peer topologies. These features make it promising for WSNs. Basically, the medium access employs CSMA-CA mechanism. However, by the optional superframe structure, time slots can be reserved for devices with time critical data upon their allocation request message. In the beacon-enabled synchronized mode, the PAN coordinator may allocate portions of the active superframe to form guaranteed time slots (GTSs) and provide contention-free period (CFP), which starts after the slotted CSMA-CA based contention access period (CAP).

Physical layer :

The physical layer comprises not only the transceiver, but also MCU, sensors, and the energy source. Therefore, the physical layer determines limits for the other layers. While the transceiver causes most of the energy usage, it also imposes several other limitations to the communication protocols. The data rate limits maximum achievable throughput, whereas the used coding scheme affects reliability. As the communication range is limited, the transceiver determines the minimum network density that is needed to route data. MCU puts limits to computational capabilities and available memory, therefore preventing complex protocols and applications. Also, its energy efficiency in sleep and active modes have a significant impact on energy usage. Physical sensors have certain accuracy and acquisition time-limiting sampling intervals. To overcome these limitations, the network may need to sample data in several nodes on the same region and combine this data to get more detailed values, thus consuming more energy. Still, if a sensor supports selecting sensing accuracy, the accuracy may be purposefully reduced to make a tradeoff against energy.

Summary :

The operation on separate layers can be complementary. For example, a radio on the physical layer affects the reliability of a single transmission ; the MAC layer is responsible for the

reliability of a link ; and the routing and transport layers are responsible for true end-to- end reliability. Also, for a real QoS, cooperation between layers is required. Otherwise, each layer may try to maximize different QoS metrics, which will have unpredictable and possibly undesirable results.

1.4.2 Data Centric QoS Support

Due to the data centric nature of WSNs, it is natural to build the QoS architecture based on the procedure data sensing, data transmission and data processing. Figure 1.2 shows the related QoS metrics in each component of the data centric QoS architecture.

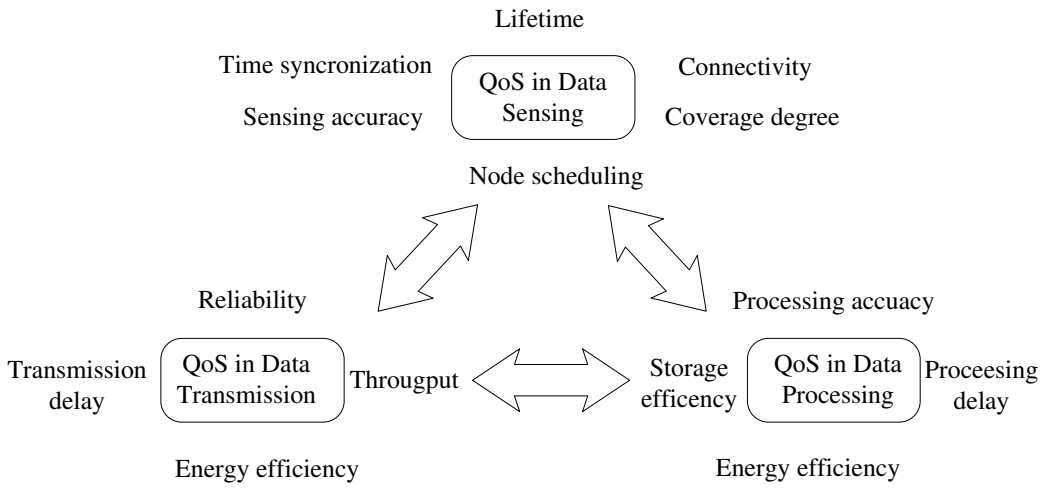


FIGURE 1.2: Data centric QoS architecture.

Data sensing :

In WSNs, the state of the environment or a target is assessed by the sensed parameters like temperature, humidity, sound and accelerations etc. Therefore, sensing related service should be guaranteed to ensure an accurate reflection of the physical world. The most important QoS metrics in this part are the quality of coverage and connectivity. The former indicates the possibility of event or target being captured, while the latter the possibility of the information being transmitted to the base station. Since energy efficiency is always important throughout the design of the system, network lifetime should be a concern in the sensing tasks, e.g., a well-designed sleeping scheduling is desired. It is worth noting that there should be tradeoffs in coverage and sleep scheduling design as the effects of the two actions are contradictive.

Data transmission :

There are different QoS metrics for different types of data in the transmission procedure. For real-time data, the timeliness and reliability is the prior concern, while for non-real-time data, throughput is more important. The task of transmission can be mapped to MAC layer and network layer in the protocol layer based architecture, namely, the MAC provides channel access

(single-hop) guarantee and in the network layer the routing design ensures end-to-end (multi-hop) transmission.

Data processing :

Data processing strategies such as data compressing, aggregation, and decision fusion based on the sensed data, play an important role in the quality of surveillance, data timeliness and reducing communication overheads. The QoS metrics in data processing include processing accuracy, delay, energy efficiency and storage efficiency, etc.

1.5 Contributions

The structure of the thesis is illustrated in Figure 1.3. Considering the event detection application, the user gives a set of QoS requirements, e.g. the quality of coverage and connectivity of the deployed network, the deadline of critical packets, the accuracy of the detection information, the energy efficiency etc. These QoS requirements can be mapped to different function layers and be satisfied by executing respective QoS mechanisms. To meet the required quality of coverage and connectivity, the most effective way is to customize a deployment strategy for the area of interest (AoI). After deployment, as critical data packets are transmitted through multi-hops to the BS, a real-time QoS based routing is desired to guarantee the end-to-end delay and reliability as well as energy efficiency and load balance. Finally to ensure required quality of detection or information accuracy, one of the solutions is to design a detection fusion rule with guaranteed false alarm rate, and robust in dynamic WSN environment.

The main contributions of the thesis are as follows :

1. To better understand how various parameters impact on the quality of connectivity, we analyze the connectivity of WSN in a non-isotropic log-normal shadowing environment. An explicit expression of node non-isolation probability is derived as the upper bound of 1-connectivity, based on an analytical link model which incorporates important parameters such as path loss exponent, shadowing variance of channel, modulation, encoding method etc. A tight lower bound for the minimum node density that is necessary to obtain an almost surely connected network is also given. Besides, we find giant component size a good relaxed measure of connectivity in some applications that do not require full connectivity.
2. To meet the quality of coverage and connectivity together, a fine deployment strategy is developed. Assuming each point in the area is associated with a coverage threshold, which must be satisfied after nodes are deployed, the resulting deployment problem is formulated as a multi-objective optimization problem, which seeks to minimize both the gap between the generated coverage probability and the required threshold and the number of deployed nodes with the constraint of maintaining the network connectivity. Heuristic methods based on tabu search (TS) and generic algorithm (GA) are proposed. Simulations show that GA and TS deployment outperforms random and regular lattice deployment. Furthermore, GA provides diverse solutions with better tradeoffs between two objectives.

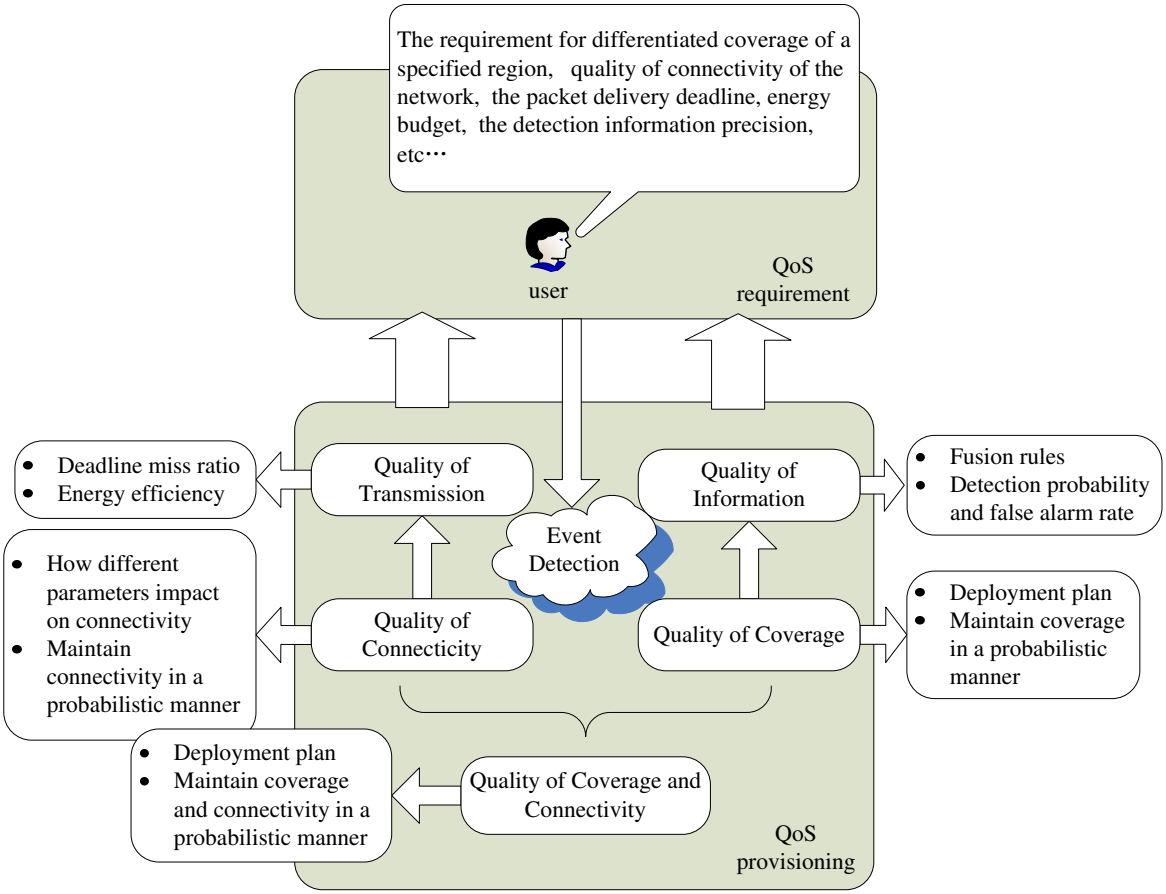


FIGURE 1.3: The structure of the thesis.

3. To satisfy real-time and reliable delivery requirement, a two-hop neighborhood information based real-time routing protocol is proposed. The packet deadline is mapped to a network-wide velocity while routing decision is made based on the novel two-hop velocity integrated with energy balancing mechanism. Initiative drop control is embedded to enhance energy utilization efficiency while reducing packet deadline miss ratio. The proposed routing has one-hop more prediction capability as using a "telescope" in finding the way. Simulation and comparison show that the new protocol has led to lower deadline miss ratio, higher energy efficiency and load balance than two existing popular schemes.
4. Finally decision fusion rules under fading channel in WSN are investigated to ensure quality of detection. Local decisions made by local sensors may be corrupted while transmitted to the fusion center via a fading channel. a series of fusion rules are proposed with the assumption of Rayleigh channel model. Likelihood ratio rule has been proved optimal by analysis and simulation. However, it is with high computation complexity and requires instant channel state information, which is not easily available in resource-constrained WSN. We propose three sub-optimal alternatives, which have less computation cost and requires less a priori information. They perform well in their respective SNR ranges. We draw the conclusion that in resource-constrained WSN, a tradeoff should be considered among performance, resource cost and computation complexity while choosing the fusion

rules.

2

Quality of Connectivity Analysis

2.1 Introduction

Connectivity is one of the fundamental properties of WSNs. First, connectivity is a basic requirement for successful communications and it affects the performance of upper-layer protocols, e.g. reliable routings basically function with a well-connected physical topology. Second, for some applications with infrequent transmissions, connectivity is more important than the capacity. For example, a WSN for environmental monitoring requires high quality of connectivity (QoCn) to ensure persistent transmissions of periodic data, but does not need much bandwidth to handle the traffic. In order to ensure certain quality of connectivity before deployment, we need to make it clear how various parameters impact on the connectivity, or the scaling law of QoCn as node density increases, so that the WSNs can be configured with certain level of performance guarantee.

Two nodes are considered to be connected if they are linked either by a single-hop or by a multi-hop, where a single-hop consists of two nodes that can directly communicate with each other and a multi-hop consists of multiple single-hops. Pioneering work dealing with network connectivity are mostly based on simplistic boolean disk model, which means two nodes are connected if and only if their distance is less than a deterministic transmission radius. Unfortunately, the real low-power wireless links are unreliable and asymmetric, suffering from severe propagation impairments such as shadowing, fading and multi-path effect. Boolean disk model is thus too simple and far from reality to describe wireless connections. In this chapter, we consider a "link" from the perspective of the packet-level connection probability. An analytical link model is adopted which matches the real performance of Mica2 Mote [25]. Through an in-depth derivation, local property of individual links and the global network connectivity are bridged. It is shown by both analysis and simulations the impact of various parameters on the QoCn, and how the QoCn scales with the increase of the node density. Furthermore, we introduce largest connected component size and giant component probability as alternative QoCn metrics, in case some applications do not require a strict 100% connectivity.

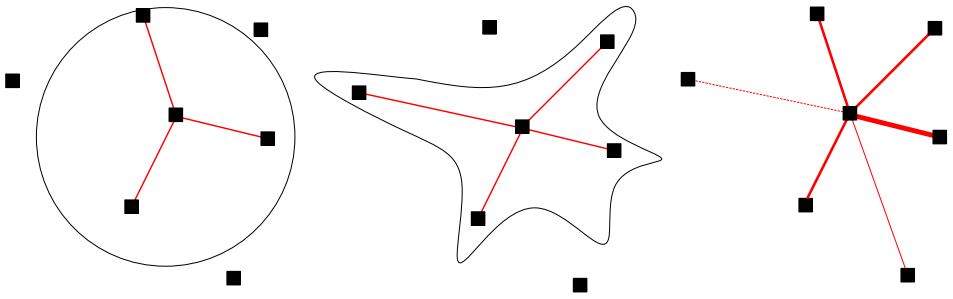
The rest of this chapter is organized as follows. Section 2.2 provides a short overview of both

wireless link model and network connectivity issue. Section 2.3 gives related models and assumption. The detailed analysis of node isolation probability under an analytical link model are presented in Section 2.4. In Section 2.5 we conduct extensive simulations to evaluate the analysis in Section 2.4 and also discuss the impact of various parameters on the global QoCn. Finally we summarize this chapter in section 2.6.

2.2 Related Work

2.2.1 Overview of Wireless Communication Models

A lot of existing research results are based on the boolean disk model as illustrated in Figure 2.1(a). Two nodes are either connected or disconnected depending on the transmitter-receiver distance and the disk radius. Recent work [26, 27] argue that the real characteristics of low-power wireless links differ greatly from the ideal model especially in their unreliability and asymmetry. Some studies have improved the disk model and consider the shadowing effect. The radiation pattern is squeezed and stretched as shown in Figure 2.1(b). It is proved in [26] that irregular radio patterns can achieve connectivity more easily if they can maintain an average number of functioning connections. The two models above are basically deterministic either with assured transmission or assured probability of transmission within the communication range. However, several studies [27–31] have pointed out that the communication range of a wireless node cannot be specified and they propose to use probabilistic link model, as shown in Figure 2.1(c) (the sickness of the line represents the probability of connection), instead of deterministic model. It is found that the successful transmission probability at a given distance s, namely $P(s)$, is a non-monotonically decreasing function to s. P relies not only on the distance s between two nodes but various parameters such as the channel parameters like the path loss exponent, the shadowing variance and the degree of irregularity (DoI) [31], and the radio parameters including modulation, encoding, output power, receiver noise and frame size.



(a) boolean model (b) shadowing model (c) probabilistic model

FIGURE 2.1: Different link models

The third link model discussed above is supported by many experimental studies and significantly affects networks behavior. Ganesan *et al*. [28] provide a wealth of empirical data from studies of large scale, dense wireless network, which demonstrate that even a simple flooding

algorithm entails complex behavior under unreliable links. Woo *et al*. [29] have identified the existence of three distinct reception regions in wireless link : connected, transitional and disconnected. The transitional region is relatively large in size and is characterized by high variance in reception rates and asymmetric connectivity. In a typical sensor networks, a large number of links (even higher than 50%) can be unreliable because of the transitional region. Zhou *et al*. [31] find that radio irregularity has a significant impact on the routing protocols in wireless sensor networks, especially location-based routing, such as geographic forwarding. All these research results lead us to stress the need for realistic link models for wireless sensor networks.

Several studies have proposed communication models based on empirical data and analyzed related phenomena for more accurate evaluation of upper-layer protocols. Woo *et al*. [29] present a simple synthetic link model to generate data under specific radio and environment based on an assumption of Gaussian distribution of the packet reception rate for given transmitter-receiver distance. These synthetic traces are used for simulation of passive link estimators that snoop traffic over the channel and estimate link qualities. Cerpa *et al*. [30] study the relationships between location and communication properties using non-parametric statistical techniques. They provide a probability density function that completely characterizes the relationship and develop a series wireless network models that produce networks of arbitrary size with empirical observed properties. The accuracy of their models is evaluated through a set of communication tasks like connectivity maintenance and routing. Lal *et al*. [32] propose link inefficiency to be a cost metric based on detailed observations of link quality variation and explore how link inefficiency can be measured in an energy efficient way. They find that only a few measurements of the channel are sufficient to obtain a good estimate of the cost metric. Zhou *et al*. [31] establish a radio model called the Radio Irregular Model (RIM) with empirical data obtained from the Mica2 platform. RIM takes into account both the non-isotropic properties of the propagation media and the heterogeneous properties of devices. With this model, they find solutions to improve the communication performance in the presence of radio irregularity. Most important to our work is the analysis of transitional region in [27]. Different from empirical models above which requires specific radio and environment conditions, the analytical model in [27] has a general methodology and can be used for a number of different configurations and hardware designs. The incorporation of different parameters in the link model greatly facilitate our analysis of network connectivity in presence of unreliable and asymmetric links. Detailed description of the link model we use can be found in section 2.4.2.

2.2.2 Overview of Connectivity Issue in Wireless Multi-Hop Networks

One of the first papers on connectivity in wireless multi-hop networks was [33], which investigated how far a node's message percolates for Poisson distributed nodes on an infinitely large area. More recently, Gupta and Kumar [34] performed a fundamental study on the connectivity of uniform distributed nodes in the asymptotic case. They have shown that when the covered area of each node equals to $\frac{\log N + c(N)}{N}$, where N is the number of nodes in the unit disk and $\liminf_{N\to+\infty} c(N) = +\infty$, the resulting wireless network is asymptotically connected with probability one if $c(N) \to +\infty$. Afterwards, various methods and theories have been applied to the study of connectivity. The model of continuum percolation with the Poisson boolean model is commonly

used to study the phase transition phenomenon in wireless connectivity [26,35,36]. Considering nodes with identical range distributed in an infinite two-dimensional plane according to Poisson point process, refereed to Definition 2.1, there is critical node density, above which there will be an infinite cluster almost surely [1] (a.s.), below which there will be bounded clusters a.s.

Although geometric random graph and percolation theory applied to the study of wireless connectivity shows great effectiveness, their basis is a deterministic radio link model, which reduces their charm in real environment. To date, several literatures have investigated the connectivity under realistic radio channel model. Fading and shadowing effects are considered. Miorand et. al [24] analyze connectivity issues in one-dimensional ad hoc networks using queueing theoretical approach. They use $GI/G/\infty$ model to study broadcast percolation problem with general node placement with fading channel. it is found in [37, 38] that when shadowing gets more severe, the link probability at short distances reduces, while increases at large distances. Long-distance connectivity probability will affect the global network connectivity and the routing performance, similar to the small world [39] networks extended with a few long links. It seems that shadowing fading improves global connectivity. However in our study, we find that the effect has been counteracted with the increase in asymmetric links and thus lead to limited improvement of global connectivity. Our work takes into account of more realistic scenarios and builds a bridge over the local link property and global connectivity behavior, which makes it clear to see how local link parameters impact network global performance.

Noisy links are introduced in [40] to study the impact of interference on the connectivity of ad hoc networks using percolation theory. It is found that there is a critical noise coefficient value above which the network is made of disconnected clusters. When the noise coefficient is small enough, the asymptotic connectivity can be achieved. Indubitably, whether two nodes can communicate with each other at any given distance and any moment also depends on the interference condition, which is caused by simultaneous communication between other nodes in the range. Due to interference, communication between two connected nodes may drop to lower bit rate or even become impossible at certain time. However, in our study we regard this case as a medium access issue instead of connectivity problem. In other words, when the link failure is caused by interference or collision, we say that the network MAC is not effective and the link capacity is reduced, instead of saying that the connectivity between two nodes is of lower degree. In fact, a simple collision free MAC protocol, such as TDMA scheme, can help achieve a better connectivity [40] in this sense.

2.3 Models and Assumptions

2.3.1 Node Spatial Distribution Model

Suppose the sensor nodes are scattered from the air vehicle and the process resembles Poisson arrival in a service queue. Thus we use Poisson point process with intensity λ to model the spatial distribution of the nodes. This process is defined as follows :

1. almost surely means the event happens with probability one if the plane goes infinite.

Definition 2.1 *Consider a measure space (X,Σ,μ), where μ is a σ-finite measure. A Poisson Point Process (P. P. P.) with intensity λ is a collection of random variables $N(A,\omega)$, $A \in \Sigma$, $\omega \in \Omega$, defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ such that :*

1. *$N(\cdot,\omega)$ is a counting measure on (X,Σ) for each $\omega \in \Omega$.*
2. *$N(A,\cdot)$ is the number of nodes in subarea A which follows Poisson distribution with mean $\lambda(A)$:*

$$P(N(A)=k)=\frac{\lambda(A)^k}{k\,!}e^{-\lambda(A)} \quad ,all\ A\in\Sigma, \tag{2.1}$$

with an expected value $E(N)=\lambda(A)=\rho(A)|A|$, ρ and $|A|$ are node density and size of subarea A respectively. are independent random variables :

$$P(N(A_1)=k_1 \wedge N(A_2)=k_2 \wedge \ldots \wedge N(A_n)=k_n)=\prod_{i=1}^{n}P(N(A_i)=k_i). \tag{2.2}$$

If ρ is constant over the entire infinitely large area, the process is homogeneous. In other words, the outcome of N only depends on $|A|$ but not on its particular location or shape. Note that a homogenous Poisson point process can be regarded as a case of uniform distribution of k nodes in area A, as k and $|A|$ tends to infinity but the density $\rho = k/|A|$ remains constant.

2.3.2 Network Model and Assumptions

We consider an arbitrary subarea A of an infinite plane, as shown in Figure 2.2. It is modeled as a graph $G(V,E)$ having V nodes and a set of radio links E. Let $G_u(V,E_u)$, or simply G_u, be an undirected graph and $G_d(V,E_d)$, or simply G_d, a directed graph. They model the connectivity graph of a WSN with symmetrical and asymmetrical links, respectively. G_u is said to be connected if for each node pair, there is at least one sequence of edges connecting them. G_d is called weakly connected if replacing all of its directed edges with undirected edges produces a connected (undirected) graph. It is connected if it contains a directed path from z to v or a directed path from v to z for each pair of vertices z,v. It is strongly connected if it contains a directed path from z to v and a directed path from v to z for every pair of vertices u,v. A reduced graph is thereof introduced, referred to as $G_r(V,E_r)$, or simply G_r, which has an undirected link between nodes $z,v \in V$ iff G_d has edges in both directions between z and v.

In WSNs, due to radio irregularity, there are asymmetric links [31]. These links can be unfavorable to upper-layer protocols since they cannot provide feedback channels. Therefore, we prefer using G_r to model the connectivity graph of a WSN so that reliable upper-layer protocols can be supported. Both Asymmetric and symmetric links are discussed in the following sections.

2.4 Network Connectivity Analysis

In this section we provide theoretical analysis for the connectivity in WSNs. Our focus will be on 1-connectivity. Higher orders of connectivity will be discussed in the future work. We are interested in the critical density ρ, at which the network A is 1-connected a.s..

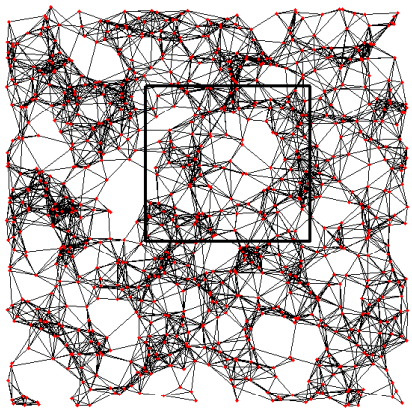


FIGURE 2.2: A subarea of an infinite plane

2.4.1 Node Non-Isolation Probability

In this section, we derive necessary condition for the network to be 1-connected. We denote $P(I)$ as the node isolation probability, which is the probability that a randomly chosen node has no neighbors at all. The probability that none of the nodes in A is isolated, denoted by $P(NI)$ can be the upper bound for the probability that the network is 1-connected , denoted by $P(C)$. Thus we have :

$$P(C) \leq P(NI) \tag{2.3}$$

$$\rho(P(C) = p) \geq \rho(P(NI) = p) \tag{2.4}$$

Equation (2.3) is obvious. Equation (2.4) means under the same network configuration, larger node density is needed to reach 1-connectivity with probability p than to reach node non-isolation with the same probability. We assume the isolation of different nodes in A to be independent events and the number of nodes in A is large, i.e. $\lambda \geq 100$. The conditional probability of non-isolation is given by :

$$P(NI|N = k) = (1 - P(I))^k \tag{2.5}$$

where we simply use N to denote $N(A)$ and λ for $\lambda(A)$. Thus the unconditional node non-isolation probability can be written as :

$$\begin{aligned} P(NI) &= \sum_{k=0}^{\infty} P(NI|N = k) \cdot P(N = k) \\ &= \sum_{k=0}^{\infty} (1 - P(I))^k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda P(I)} \sum_{k=0}^{\infty} \frac{[(1 - P(I))\lambda]^k}{k!} e^{-(1-P(I))\lambda} \\ &= e^{-\rho|A|P(I)} \end{aligned} \tag{2.6}$$

Since $P(I)$ is the probability that a randomly chosen node has no neighbors, we denote D to be the number of neighbors of a node, which is also called node degree. Note that D also follows Poisson distribution according to equation (2.1). Let the expected value of D be D_0, then $P(I)$ can be expressed as :

$$P(I) = P(D = 0) = e^{-D_0} \tag{2.7}$$

In the following, we shall calculate the mean node degree D_0. Let $L(i, j)$ denote the event that there is a link between node i and node j, either symmetric or asymmetric. The distance between the two nodes is known as $s(i, j)$. Given the distance, the probability for $L(i, j)$ is denoted by $P(L|s)$. Since any two nodes in the area could be connected with a probability, D_0 can be calculated by integrating $\rho P(L|s)$ over the whole area :

$$\begin{aligned} D_0 &= \rho \int_0^{2\pi} \int_0^{\infty} P(L|s)\ s\ ds\ d\phi \\ &= 2\pi\rho \int_0^{\infty} P(L|s)\ s\ ds. \end{aligned} \tag{2.8}$$

By plugging equation (2.7) and (2.8) into (2.6), we have :

$$P(NI) = \exp(-\rho|A|\ \exp(-2\pi\rho\ \xi)) \tag{2.9}$$

where $\xi = \int_0^{\infty} P(L|s)\ s\ ds$. Let the left hand side of equation (2.9) be p, then the density ρ satisfies the following transcendental equation :

$$\rho\ e^{-2\pi\rho\ \xi} = \frac{\ln p}{|A|} \tag{2.10}$$

The solution of the equation is

$$\rho(P(NI) = p) = -\frac{1}{2\pi\xi}\text{lambertw}(\frac{-2\pi\xi \ln p}{|A|}) \tag{2.11}$$

where lambertw(x) is the solution w of equation $w\exp(w) = x$. Finally, the problem has transformed to the analysis of ξ, equivalent to the analysis of $P(L|s)$. In other words, the study on the global network connectivity has turned out to be the analysis of the local link probability.

2.4.2 Link Probability Analysis

Depending on different radio link models, the link probability $P(L|s)$ is of different forms. For Boolean disk model as shown in Fig. 2.1(a), $P(L|s)$ is a logical function and equals to either 0 or 1 depending on the communication range. Hence,

$$\xi = \pi r^2 \rho \tag{2.12}$$

where r is communication radius. Under Boolean disk model, the link is perfectly symmetric.

For probabilistic model as shown in Figure 2.1(c), $P(L|s)$ is not deterministic 0 or 1, but a probability within the range of [0,1]. In packet-level perspectives, $P(L|s)$ can be indicated by

by the packet reception rate (PRR) Ψ over a period of time τ and $0 \leq \Psi \leq 1$. In practice, Ψ can be estimated over a moving time window. Here we use it in simulation as a stationary variable to indicate the probability of instant link success or failure. If the radio DoI is low, the links are prone to be symmetric and we have $P(L|s) = \Psi$. On the other hand, if the radio DoI is high which means links suffer severe asymmetry, $P(L|s)$ tends to be Ψ^2. We consider both situations in the following analysis.

Empirical studies have shown that the log-normal shadowing model provides more accurate multi-path channel models than other models like Nakagami and Rayleigh models [27]. According to this model, the received power P_r in dB is given by

$$P_r(s) = P_t - \left(PL(s_0) + 10\eta \log(\frac{s}{s_0}) \right) + X_\sigma \tag{2.13}$$

where P_t is the output transmission power, η is the path loss exponent, X_σ is a Gaussian random variable with mean 0 and variance σ^2 and $PL(s_0)$ is the power decay for the reference distance s_0. Equation(2.13) does not include anisotropic properties of the radio. To incorporate it, we add a coefficient K_i to represent the difference in path loss in different directions. Equation (2.13) can be modified as

$$\begin{aligned} P_r(s) &= P_t - \left(PL(s_0) + 10\eta \log(\frac{s}{s_0}) \right) \cdot K_i + X_\sigma \\ K_i &= \begin{cases} 1, & \text{if } i = 0\,; \\ K_{i-1} + rand \cdot \text{DoI}, & \text{if } 0 < i < 360, i \in N \end{cases} \\ \text{DoI} &= 0.01821 \end{aligned} \tag{2.14}$$

where *rand* is a random variable in [0,1] and DoI represents the degree of irregularity. The reference value of DoI is suggested in [31] to be 0.01821, which is measured in the platform of Mica2 mote. We can generate 360 values for 360 different directions. However, to make the simplicity, we just adopt 4 directions in the simulation.

The channel SNR γ is given by :

$$\gamma_{dB}(s) = P_r(s) - P_n \tag{2.15}$$

where P_n is the noise floor which depends on both the radio and the environment. Considering an ambient temperature of 27°C and no interference signals, the noise floor is -115dB [27]. Note that γ can also be regarded as a Gaussian random variable with mean $\mu(s)$ and variance σ^2 by putting $P_r(s)$ in equation 2.14 into equation 2.15, where $\mu(s)$ is given by :

$$\mu(s) = P_t - \left(PL(s_0) + 10\eta \log(\frac{s}{s_0}) \right) \cdot K_i - P_n. \tag{2.16}$$

On the other hand, given an encoding method E, the PRR is a function of bit-error rate P_e, while P_e is a function of SNR γ, depending on the modulation method, and more generally on transmitting and receiving techniques (diversity, equalization, etc.). With non-coherent FSK, which is the default modulation scheme in Mica2 Mote, we have :

$$P_e = \frac{1}{2} \exp\left(-\frac{\gamma(s)}{2} \frac{1}{0.64} \right), \quad \gamma = 10^{\gamma_{dB}/10}, \tag{2.17}$$

By Manchester encoding, we further have :

$$\begin{aligned}\Psi(s) &= (1-P_e)^{8(2f-h)} \\ &= \left(1-\frac{1}{2}\exp\left(-\frac{\gamma(s)}{2}\frac{1}{0.64}\right)\right)^{8(2f-h)}\end{aligned} \tag{2.18}$$

where f is the frame size including preamble, payload and CRC, h is the preamble, both in byte. For other encoding and modulation methods, the expression of $\Psi(s)$ may be different. With equation (2.18), the PRR of arbitrary pairs of nodes can be obtained. Figure 2.3 shows the analytical model of PRR to distance in the 45° direction. It is clear that there is a large transitional region, in which the PRR is unstable. With the following definitions, we can further outline the transitional region.

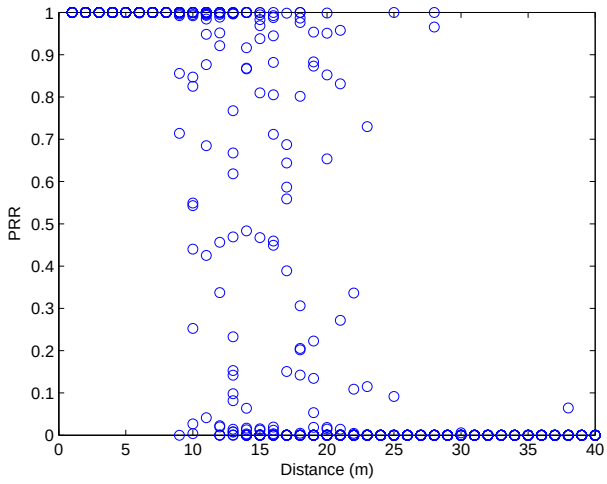


FIGURE 2.3: Analytical PRR to distance, obtained from equation (2.18) and (2.16), P_t=-5dB, η=3, σ=3.3, f=50 bytes, h=2 bytes.

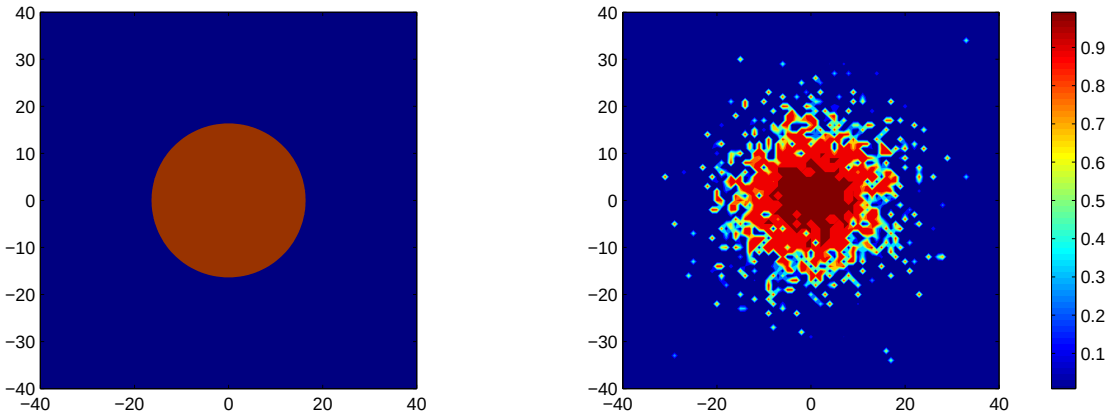
Definition 2.2 *If the link between two nodes has a high probability (e.g. 0.9) of having high PRR (e.g. 0.9), then the two nodes are strongly connected. The upper bound of the connected range is* s_1.

Definition 2.3 *If the link between two nodes has a high probability (e.g. 0.9) of having low PRR (e.g. 0.1), then the two nodes are almost disconnected. The lower bound of the disconnected range is* s_2.

According to definition 2.2 and definition 2.3, the region between s_1 and s_2 is the transitional region, which agrees with empirical observations in literature [27, 29]. The existence of this region has been explained in [27], which is the joint effect of receivers with non-perfect threshold , noisy environment and multi-path effects.

With consideration of different path loss and shadow fading effects along different directions, we draw contour plot of the reception rate for a node located at the center of the region. Figure 2.4 shows the difference of the communication covered area under Boolean disk model and

shadowing fading probabilistic model. Figure 2.4(b) captures the real characteristic of wireless link and shows spatial irregularity as the empirical results in [31]. To calculate $P(NI)$, the value



(a) Boolean disk model

(b) Log-normal shadowing model

FIGURE 2.4: Contour plot for different link models.

of ξ is crucial. As depicted in Figure 2.3, it is difficult to derive a continuous function for Ψ. Instead, we analyze the expected value of Ψ, $E(\Psi)$, using $E(\Psi)$ and $E(\Psi^2)$ for the calculation of respective ξ_Ψ and ξ_{Ψ^2}. We have :

$$\begin{aligned}\xi_\Psi &= \int_0^\infty E(\Psi)\, s\, ds \\ &= \int_0^\infty \int_{-\infty}^\infty \Psi(\gamma) f(\gamma|s) \cdot s\, d\gamma\, ds \qquad (2.19)\\ \xi_{\Psi^2} &= \int_0^\infty E(\Psi^2)\, s\, ds \\ &= \int_0^\infty \int_{-\infty}^\infty \Psi^2(\gamma) f(\gamma|s) \cdot s\, d\gamma\, ds \qquad (2.20)\end{aligned}$$

where $f(\gamma|s)$ denotes the probability density function of SNR. According to (2.18), we draw the PRR as a function of γ. As the power-law relationship entails a sharp threshold, linear approximation turns available. $\Psi(\gamma)$ can be approximated by piecewise linear functions. Similarly, $f(\gamma|s)$ can be evaluated by linear approximation since the interval $[\gamma_0, \gamma_1]$ is relatively narrow compared with $[\mu - 4\sigma, \mu + 4\sigma]$. Hence we have :

$$\begin{aligned}\Psi(\gamma) &= \begin{cases} 0, & \gamma \le \gamma_0\,; \\ k_1\gamma + b_1, & \gamma_0 < \gamma < \gamma_1\,; \\ 1, & \gamma \ge \gamma_1. \end{cases} \\ k_1 &= \frac{0.9-0.1}{\gamma_B - \gamma_A}, b_1 = \frac{0.1\gamma_B - 0.9\gamma_A}{\gamma_B - \gamma_A}, \\ \gamma_A &= \Psi^{-1}(0.1), \gamma_B = \Psi^{-1}(0.9), \\ \gamma_0 &= -\frac{b_1}{k_1}, \gamma_1 = \frac{1-b_1}{k_1}. \qquad (2.21)\end{aligned}$$

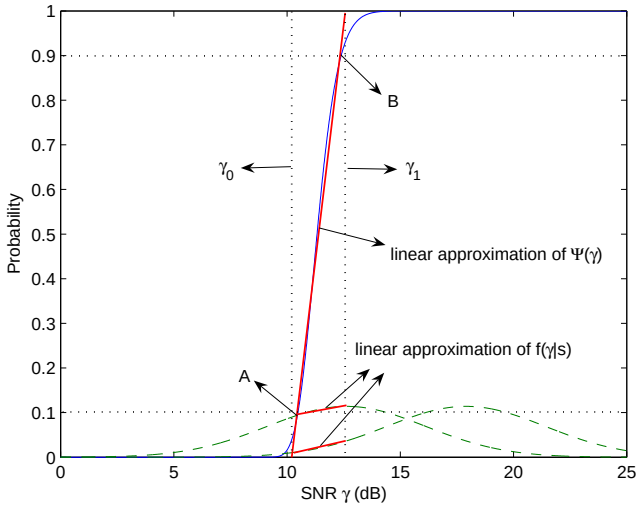


FIGURE 2.5: Linear approximation for $\Psi(\gamma)$ and $f(\gamma|s)$

$$f(\gamma|s) = \begin{cases} k_2(s)\gamma + b_2(s), & \gamma_0 < \gamma < \gamma_1; \\ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu(s))^2}{2\sigma^2}}, & \gamma \le \gamma_0, \gamma \ge \gamma_1. \end{cases}$$
$$k_2(s) = \frac{f(\gamma_1|s) - f(\gamma_0|s)}{\gamma_1 - \gamma_0}, \quad b_2(s) = \frac{f(\gamma_0|s)\gamma_1 - f(\gamma_1|s)\gamma_0}{\gamma_1 - \gamma_0} \tag{2.22}$$

where γ_0 and γ_1 represent the receiver's range of threshold. Figure 2.5 shows the approximation procedure for $\Psi(\gamma)$ and $f(\gamma|s)$. Similar operations can be conducted for calculation of $\Psi^2(\gamma)$. Here we simply list the results :

$$\Psi^2(\gamma) = \begin{cases} 0, & \gamma \le \gamma'_0; \\ k'_1\gamma + b'_1, & \gamma'_0 < \gamma < \gamma'_1; \\ 1, & \gamma \ge \gamma'_1. \end{cases}$$
$$k'_1 = \frac{0.9 - 0.1}{\gamma'_B - \gamma'_A}, b'_1 = \frac{0.1\gamma'_B - 0.9\gamma'_A}{\gamma'_B - \gamma'_A},$$
$$\gamma'_A = \Psi^{-1}(0.1^2), \gamma'_B = \Psi^{-1}(0.9^2),$$
$$\gamma'_0 = -\frac{b'_1}{k'_1}, \gamma'_1 = \frac{1 - b'_1}{k'_1}. \tag{2.23}$$

$$f(\gamma|s) = \begin{cases} k'_2(s)\gamma + b'_2(s), & \gamma'_0 < \gamma < \gamma'_1; \\ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu(s))^2}{2\sigma^2}}, & \gamma \le \gamma'_0, \gamma \ge \gamma'_1. \end{cases}$$
$$k'_2(s) = \frac{f(\gamma'_1|s) - f(\gamma'_0|s)}{\gamma'_1 - \gamma'_0}, \quad b'_2(s) = \frac{f(\gamma'_0|s)\gamma'_1 - f(\gamma'_1|s)\gamma'_0}{\gamma'_1 - \gamma'_0} \tag{2.24}$$

Finally, ξ_Ψ can be approximated by :

$$\begin{aligned}\xi_\Psi &\approx \int_0^{\infty}\left(\int_{\gamma_0}^{\gamma_1}(k_1\gamma+b_\psi)(k_2(s)\gamma+b_2(s))d\gamma+1-\Phi(\frac{\gamma_1-\mu(s)}{\sigma})\right)\,s\,ds \\ &\approx \int_0^{s_2}\int_{\gamma_0}^{\gamma_1}(k_1\gamma+b_\psi)(k_2(s)\gamma+b_2(s))\,s\,d\gamma\,ds+\int_0^{\infty}\left(1-\Phi(\frac{\gamma_1-\mu(s)}{\sigma})\right)\,s\,ds\end{aligned} \tag{2.25}$$

Similarly,

$$\begin{aligned}\xi_{\Psi^2} &\approx \int_0^{\infty}\left(\int_{\gamma_0'}^{\gamma_1'}(k_1'\gamma+b_\psi)(k_2'(s)\gamma+b_2'(s))d\gamma+1-\Phi(\frac{\gamma_1'-\mu(s)}{\sigma})\right)\,s\,ds \\ &\approx \int_0^{s_2}\int_{\gamma_0'}^{\gamma_1'}(k_1'\gamma+b_1')(k_2'(s)\gamma+b_2'(s))\,s\,d\gamma\,ds+\int_0^{\infty}\left(1-\Phi(\frac{\gamma_1'-\mu(s)}{\sigma})\right)\,s\,ds\end{aligned} \tag{2.26}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF), μ is given by equation (2.16) and s_2 is defined in definition 2.3. In fact, the upper bound of the integration can be any value lager than s_2 since the connection probability approaches 0 when the distance between two nodes is larger than s_2. The approximation error of lowering the integration upper bound from ∞ to s_2 is tolerable. Therefore, with equations (2.9), (2.11), (2.25) and (2.26), the values of critical node density and node non-isolation probability can be obtained.

In the following, we do simulations to verify the theoretical analysis.

2.5 Simulation Results and Discussion

2.5.1 Simulation Settings

In this section, extensive simulations are conducted to validate the analytical results and show how different parameters impact the network connectivity. A total number of n nodes are distributed in a square area A, following a homogeneous P.P.P. with respective node density ρ. The transmission power is the same for each node. To avoid border effect, we use a wrap-around distance model (also called cyclic distance model) instead of usual Euclidean distance model, i.e. a node at the border of A is considered as being close to the nodes on the opposite border of A. Thus it can establish links via the borderline to these nodes. We perform simulations with the increase of node density ρ while changing other related parameters. For each unique setting, we repeat simulation runs independently 500 times and finally evaluate the average performance. In the resulting graph for each run, we evaluate the node non-isolation probability, 1-connectivity probability, largest component size and giant component probability through both numerical simulations and theoretical analysis.

2.5.2 Evaluation Metrics

For different applications, the requirement of QoCn can be different. For example, a surveillance sensor system with one sink node requires the network to be 1-connected so as to collect data

from any node in the network. However to detect or track an intrusion, it is required that only the nodes around the intrusion connect to the sink. Therefore it is argued that for some use of sensor networks, 1-connectivity is too stringent a condition and giant component is suggested as a measure of connectivity instead. In classical random graph $G(n,p)$, every pair of a set of n vertices is chosen to be an edge with probability p. The behavior of the size of the largest component in G when $p = c/n$ for c near 1 receives most interests in the field. For $c < 1$ the size of the largest component is $O(logn)$ a.s.[2]; for $c = 1$ the size of the largest component is $\Theta(n^{2/3})$ a.s.; for $c > 1$ the size of the largest component is a.s. $\Theta(n)$ and the second largest component is a.s. $O(logn)$. When $c > 1$, the largest component, whose order is much larger than any other component, is commonly referred to as the giant component. Such random graphs are fundamental and useful for modeling problems in many applications. Although the link model adopted in this chapter is quite different from classical one, we can still use the concept of giant component. In our simulation, if half of the nodes in the network are connected, a giant component is formed. This definition actually agrees with the third case ($c > 1$) of the above description in random graph theory.

To determine the probability of giant component and one-connectivty of an wireless ad hoc network graph, a simple flooding algorithm can be adopted. A random node is tagged at first, then its neighbors are tagged, which subsequently continue to tag their untagged neighbors, until the corresponding cluster is completely formed. The procedure is repeated for all the untagged nodes until all the nodes in the graph are tagged. By definition, if the size of the largest cluster found is larger than $\frac{1}{2}n$, the giant component exists. For an ad hoc sensor network to be 1-connected, the size of the largest component should equal to the total number of nodes in the network. Theoretically, there is another way to obtain the ratio of 1-connectivity network by using Laplacian matrix. Let $G = (V,E)$ be a graph on vertex with vertices V and edges E. The Laplacian matrix of a graph G with N nodes is an $N \times N$ matrix $L(G) = D(G) - A(G)$, where $D(G)$ is the diagonal matrix of vertex degrees, $D(G) = \text{diag}(d(v_i)), v_i \in V$; $A(G)$ is the adjacent matrix of G. The eigenvalues of $L(G)$ are called the Laplacian eigenvalues, which are all real and nonnegative. The set of all N Laplacian eigenvalues $\lambda_N = 0 \leq \lambda_{N-1} \leq ... \leq \lambda_1$ is called the Laplacian spectrum of a graph G.

Theorem 2.1 *If the graph G has C connected components, then L(G) has exactly C zero eigenvalues (other eigenvalues are positive).*

Based on Theorem 2.1, the graph is 1-connected if $L(G)$ possesses only one zero eigenvalue. The probability for 1-connectivity is estimated from a representative sample of ad hoc sensor networks' connectivity graph as the ratio between 1-connected graph and all sample graphs.

2.5.3 Simulation Results

Figure 2.6 shows the analytical and simulated results under different network and environmental configurations. Each subplot corresponds to a different set of configuration. Figure 2.6(a) serves

2. We use the following notation throughout the thesis : $f = O(g)$ if $\limsup_{n\to\infty} \frac{f(n)}{g(n)} < +\infty$; $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$.

as a benchmark for comparison. The critical node density in different scenarios are reported in Table 2.1. Through comparing and analyzing the results under eight sets of configurations, we draw the following conclusions :

– Throughout all the simulation results, as the node density ρ increases, the network graph becomes denser and $P(C)$, $P(NI)$, $P(G)$ become larger. It is worth noting that the transition from low connectivity to nearly full connected or appearance of a giant component is quite sharp over a small range of ρ. The phenomenon consists with the results in [41] which makes use of the theory of continuum percolation. No matter what kind of radio link model adopted, either Boolean disk model or more realistic model with fading and shadowing, the property which shows phase transition exists. Phase transition gives us a tool for analyzing and determining resource efficient regime of operation for wireless sensor networks. For example, following the settings in Figure 2.6(a), it tells us that for the nodes with identical transmission power, distributed in an area of 20000 m^2 according to homogeneous P.P.P., the node density must be higher than 0.0064/m^2 to form giant component and higher than 0.0125/m^2 to reach 1-connectivity. The density threshold is an energy-efficient point of operation, in that to the left of this threshold the network is disconnected with high probability, and to the right of this threshold, additional energy expenditure results in a negligible increase in the high probability of connectivity.

– Throughout all the simulation results, the node non-isolation probability serves as the upper bound of 1-connectivity probability. Generally, the difference between the two probabilities is non-negligible. However as node density increases, the two probabilities converge to 1. This result agrees with the inequalities (2.3) and (2.4). With respect to critical node density, this means :

$$\rho(P(C)=p)=\rho(P(NI)=p)+\delta,\delta>0.\ \text{As}\ p\to 1,\delta\to 0. \tag{2.27}$$

For $p=90\%$, the critical node densities for each setting are listed in Table 2.1. It is observed that when the shadowing variance σ^2 is of a large value(e.g. $\sigma^2=10^2$), the difference δ becomes very small over the whole node density axis, as shown in Figure 2.6(c). This phenomenon conforms with the description in theorem 2.2 in a special case that $k=1$, namely, $P(G \text{ is 1-connected})=P(D_{min}\geq 1)$. This can be explained by the fact that as σ^2 increases, the increase in the long distance link probability and the decrease in short distance link probability reduces the correlation between links. As a result, the geometric random graph behavior approaches generic random graph behavior and the 1-connectivity probability $P(C)$ approaches the node non-isolation probability $P(NI)$.

Theorem 2.2 *In a random graph $G(V,E)$ with a total number of n nodes, if the minimal degree of each node is k, as $n\to\infty$, the graph G is asymptotically k-connected. It can be denoted as :*

$$P(G \text{ is k-connected})=P(D_{min}\geq k), n\to\infty \tag{2.28}$$

where D_{min} is the minimal degree of a single node.

– It has been believed in some literatures [38, 42] that a large value of σ^2 helps the network to become connected because the number of added long links is more than the number of removed short links. However, it has been neglected that the network also suffers severe link asymmetry as the shadowing variance increases. As σ^2 gets larger, the probability of short-link connection will be affected as well, leading to an increase in asymmetric links. The global QoCn depends on the dominating factor between the increase in long-link probability and the increase in asymmetric links. Therefore, the effect of increase in σ^2 should be two-folded. As shown in Figure 2.6(c) and Figure 2.6(d) compared to the benchmark Figure 2.6(a), when $\sigma = 5$, the 1-connectivity probability $P(C)$ dropped a little ; when $\sigma = 10$, $P(C)$ has been improved. Other metrics show similar tendency. This is because as the value of σ is getting large, the link asymmetry dominates the connectivity performance at first. While σ gets larger, the long-link probability increases, gradually counteracting the effects of link asymmetry. Therefore the global QoCn improves as shown in Figure 2.6(d). For calculation of $P(NI)$, we adopt $E(\Psi^2)$ to describe the link connection probability $P(L|s)$ in case the network suffers severe asymmetric links. This has been proved reasonable as shown in Figure 2.6. When the value of σ^2 is relatively small, the simulated values of $P(NI)$ is closer to the theoretical results obtained by equation (2.19), as shown in Figure 2.6(a), 2.6(b), 2.6(e) and 2.6(g) ; while as the value of σ get large, the simulated values of $P(NI)$ is closer to the theoretical result obtained by equation (2.20), as shown in Figure 2.6(c) and 2.6(d).

– Comparing Figure 2.6(b) with Figure 2.6(a), we can see that the increase in the value of path loss exponent decreases the connectivity of the network. The reason is obvious since the larger the value of η, the faster the decay of the signal strength, resulting in a shortened transmission range. The value of η is usually in the range of [2, 4]. The path loss exponent is the least in an outdoor open space than in an indoor environment or in grass or forest, which shows the deployment environment greatly affect the network QoCn.

– Similar is Figure 2.6(e) compared with Figure 2.6(a). As the transmission power increases, the average transmission range also increases accordingly, and thus a lower node density is sufficient to make the network connected with the same high probability, as shown in Table 2.1. However, it is unreasonable to blindly increase the transmission power for a better connectivity. First, the energy supply of each node is limited. Second, the increase in the number of neighboring nodes will intensify the channel contention and inferences. The selection of optimal transmission power can be referred to as topology control issue and a literature survey can be found in [43].

– Figure 2.6(d) shows the impact of a different encoding scheme. With the same node density, the QoCn is better using SECDED encoding than Manchester encoding. This is due to the error correction capabilities of SECDED, which comes at a cost of energy efficiency with encoding ratio 1:3, while Manchester does not provide error correction and the encoding ratio is 1:2. Complex encoding method can be a way to improve link reliability. However, there must be tradeoffs between memory usage, energy consumption and the performance gain. Besides, Experimental results for different encoding schemes agree with the expected analytical behavior.

– Figure 2.6(g) shows the QoCn performance as the packet size is doubled. It can be observed that the performance merges to 1 at a slightly higher node density. This is because a packet with larger frame size tends to have larger bit error rate. Some literature [44]proposes to find optimal packet sizes for link reliability and energy efficiency.

– The impact of area size $|A|$ on the connectivity is straightforward. As the sub-area size gets larger, to reach the global connectivity requires higher node density, as shown in Figure 2.6(h). The quantity can be estimated via equation (2.10). However, to form a giant component, the probability has not been affected by the size of A.

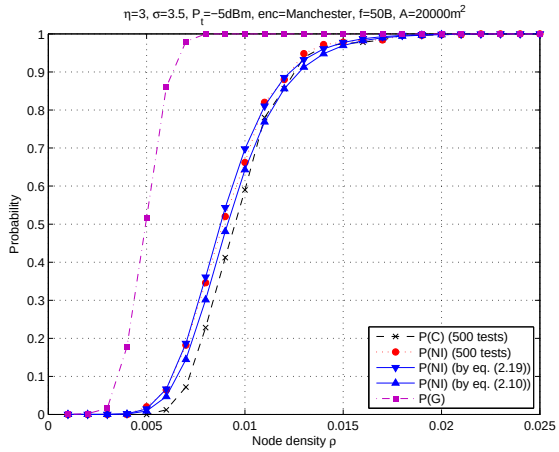
– With the increase in node density, a giant connected component appears first and then "swallowed" by the remaining nodes and clusters. Finally, the network reaches complete 1-connectivity. Therefore, the critical node density of giant component is always smaller than the density required by 1-connectivity. The probability of giant component is related to the largest component size, which will be further discussed in the following section.

TABLE 2.1: Critical density for giant component, 1-connectivity and non-isolation under different settings, comparison of analytical and simulation results.

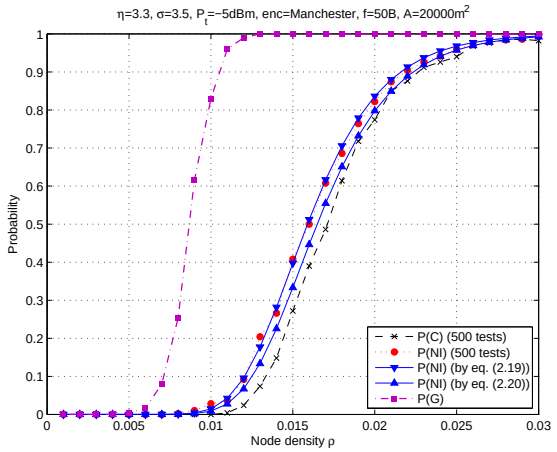
Settings	$\rho(P(G)=90\pm.5\%)$	$\rho(P(C)=90\pm.5\%)$	$\rho(P(NI)=90\pm.5\%)$		
	simulation	simulation	simulation	analysis (2.25)	analysis (2.26)
(a)	$6.37\cdot10^{-3}$	$1.25\cdot10^{-2}$	$1.23\cdot10^{-2}$	$1.22\cdot10^{-2}$	$1.26\cdot10^{-2}$
(b)	$1.06\cdot10^{-2}$	$2.27\cdot10^{-2}$	$2.23\cdot10^{-2}$	$2.21\cdot10^{-2}$	$2.25\cdot10^{-2}$
(c)	$5.86\cdot10^{-3}$	$1.30\cdot10^{-2}$	$1.29\cdot10^{-2}$	$1.12\cdot10^{-2}$	$1.32\cdot10^{-2}$
(d)	$3.90\cdot10^{-3}$	$1.08\cdot10^{-2}$	$1.08\cdot10^{-2}$	$6.00\cdot10^{-3}$	$1.09\cdot10^{-2}$
(e)	$4.95\cdot10^{-3}$	$9.09\cdot10^{-3}$	$9.07\cdot10^{-3}$	$9.07\cdot10^{-3}$	$9.09\cdot10^{-3}$
(f)	$4.93\cdot10^{-3}$	$8.57\cdot10^{-3}$	$8.55\cdot10^{-3}$	$8.00\cdot10^{-3}$	$8.57\cdot10^{-3}$
(g)	$6.72\cdot10^{-3}$	$1.33\cdot10^{-2}$	$1.30\cdot10^{-2}$	$1.32\cdot10^{-2}$	$1.40\cdot10^{-2}$
(h)	$6.37\cdot10^{-3}$	$1.40\cdot10^{-2}$	$1.36\cdot10^{-2}$	$1.34\cdot10^{-2}$	$1.40\cdot10^{-2}$

In summary, a higher transmission power, larger node density and more complicated encoding method can improve the network connectivity. However, it is energy inefficient to sacrifice the transmission power for better connectivity unconditionally ; higher shadowing variance σ^2 can increase the chance of link asymmetry and typically comes with a higher path loss exponent η ; complicated encoding method consumes more system resource (energy, memory space etc.) and takes more time. With reference to the improvement space of each parameter, one should better balance the connectivity performance and the sacrifices of resource and energy. Other techniques such as collaborative radio [45] and multiple antennas [38], although effective in theory to improve global connectivity, need more detailed infrastructure design and proof of their effectiveness in practice. In all simulation cases, we find that the giant component probability converges to 1 much more faster than 1-connectivity. In some applications, a few isolated nodes or small clusters outside the giant component do not change the overall performance. In that case, a lower node density is sufficient for operation. Besides, the largest component size

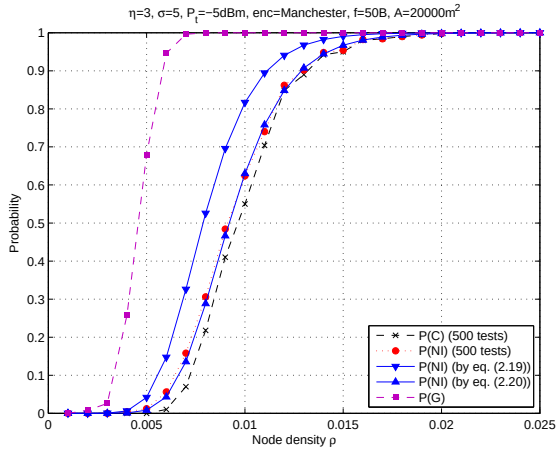
is also an alternative of connectivity measurement. It not only provides information about the network is 1-connected or not, but also the fraction of the connected part. Thus a random size of largest component can be designated satisfying different requirement of QoCn.



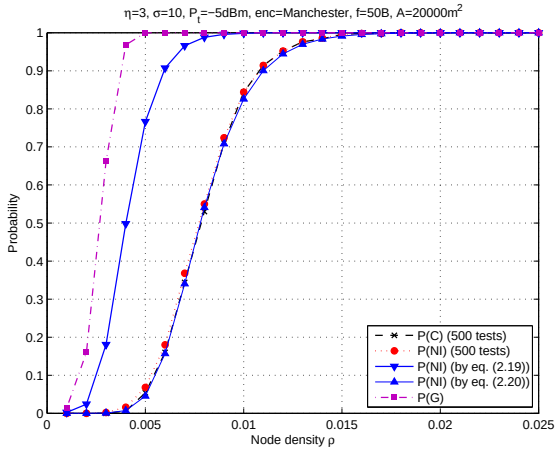
(a)



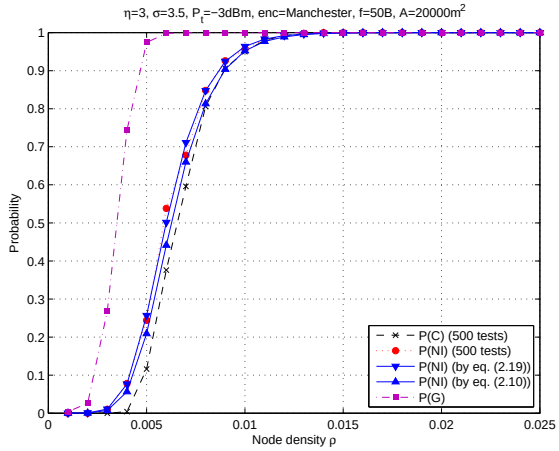
(b)



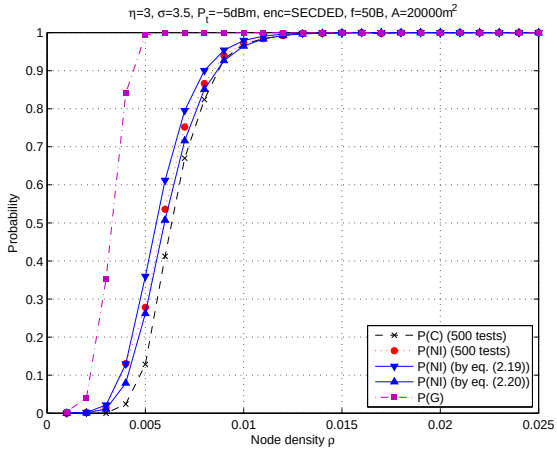
(c)



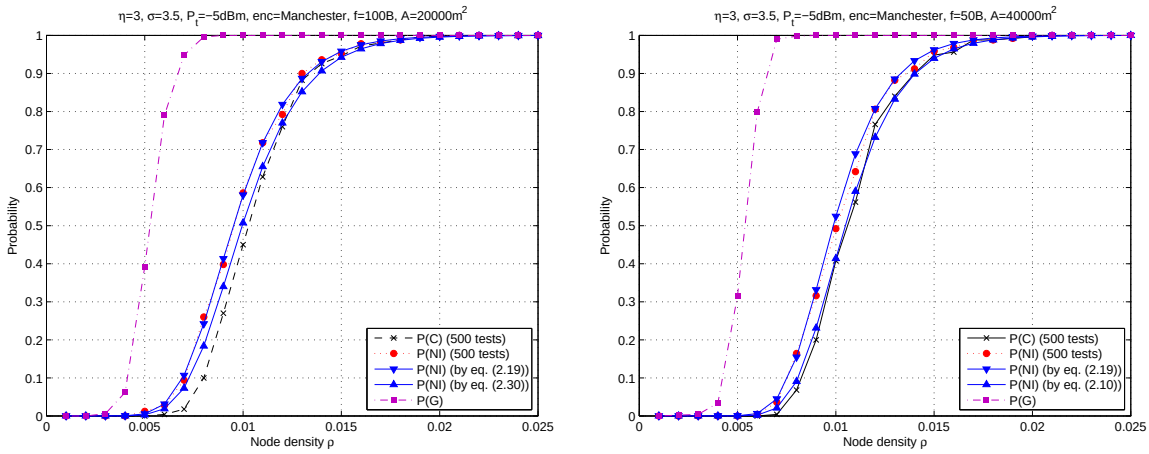
(d)



(e)



(f)



(g) (g)

FIGURE 2.6: Results obtained by both numerical simulations and theoretical analysis under different settings.

2.5.4 Largest Component Size

Let S_L be the largest component size in a random graph. In [46] it is found that when the node number n is large, S_L is the non-zero solution to the following equation :

$$S_L = 1 - \exp\left(D_0 * S_L\right) \tag{2.29}$$

where D_0 is the mean degree of the graph and can be calculated by equation (2.8). Although fast converging series exist to solve equation (2.29), a standard zero finding algorithm like Newton-Raphson method can also be used to find S_L as function of D_0. During simulations, we have found that for larger σ, the value of S_L shifts along the mean degree axis. Therefore, we have tried several function forms to estimate this shift. A good approximation found for this shift is : $2.7\exp\left(-0.3\sigma\right)$. Taking this into account, the largest component size by approximation, S_L, is the non-zero solution to the following equation :

$$S_L = 1 - \exp\left(\widetilde{D}_0 * S_L\right) \tag{2.30}$$

where $\widetilde{D}_0 = D_0 - 2.7\exp\left(-0.3\sigma\right)$.

Figure 2.7 shows the calculated largest component size for different values of σ. As visible in this figure, there is a good match between the simulated and the calculated values of the giant component size.

2.6 Summary

In this chapter, we have presented an analytical procedure of calculating the node non-isolation probability based on a generalized radio link model in wireless sensor networks. The non-isolation probability is the upper bound of 1-connectivity probability. We use a combination

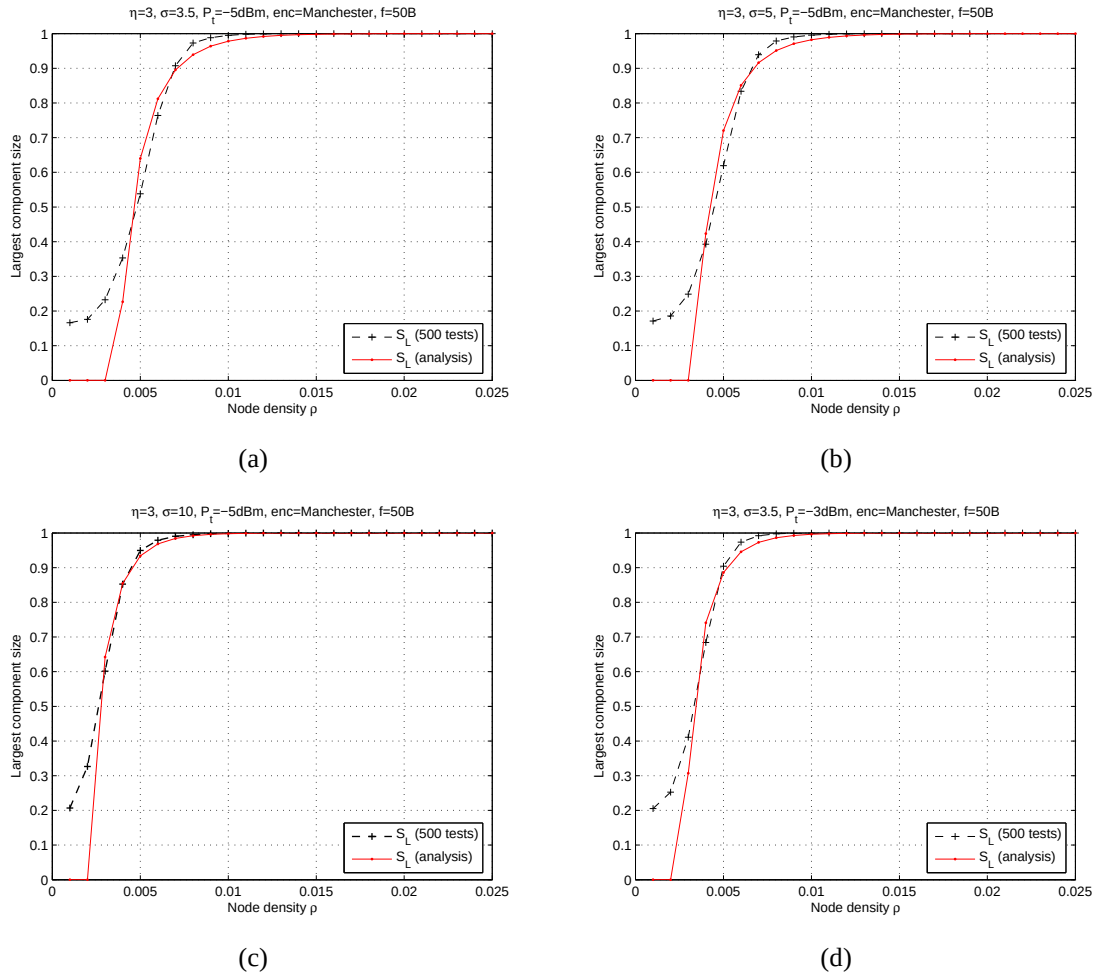


FIGURE 2.7: Giant component size : simulated values v.s. analytical values

of analytical and simulation-based methods to study the impact of various parameters on the connectivity behavior. The results can be applied for practical design and simulation in wireless ad hoc and sensor networks. For example, given the environmental and node configuration, we can determine the minimum density needed to achieve a 1-connected network or less strictly, form a giant component, according to different requirement of QoCn.

Several issues remain for future work in this area. First, this chapter only focus on the connectivity in the spatial domain, it is also interesting to study the temporal dynamics of wireless connectivity. For more challenging dynamic environment, richer time varying link model is required. Second, in this study we assume the node distribution follows homogeneous P.P.P.. However, in many real-world scenarios, the node are in an inhomogeneous distribution. Bettstetter *et al.* in [37] has proposed an algorithm to create random inhomogeneous distribution. We believe the more generalized distribution model will lead to more convinced analytical and numerical results. Third, it will be interesting if we incorporate the context of MAC layer, say IEEE802.15.4, so that connectivity is more information based than pure physical link. Finally, *k*-connectivity is a more generalized measure of global connectivity and provides more redundancy for routing choice. In other words, 1-connectivity is just a special case for k-connectivity. Some analysis

has been conducted in [33], but still need further investigation. The procedure may find some similarities, but the result is of greater generality.

3

Node Deployment with Quality of Coverage and Connectivity Support

3.1 Introduction

The deployment problem asks how to place the least number of sensors in an an area of interest (AoI) to meet some application-specific requirements, e.g. desired coverage and connectivity properties, detection delay, expected lifetime, etc. A WSN designer who takes into account the design issues above obviously deals with more than one nonlinear objective functions or design criteria. However, it is impossible to simultaneously optimize all the objectives because some of them are possibly contradictive. Higher values in one objective results in lower values in another. Thus the focus of the problem is to find the Pareto frontier, the near-optimal non-dominated solutions, which could allow the users of WSNs to make trade-off choices according to their own preferences and some practical limits.

In this chapter, we address how to meet application-specified coverage and connectivity requirements by deploying least sensors. Coverage and connectivity are two fundamental issues in wireless sensor networks. The goal of coverage is to ensure that AoI or the physical state of the target is precisely sensed, while connectivity is to ensure the event in AoI or the state of the target is successfully transmitted to the base station and subsequently captured by remote users. Coverage and connectivity together can be treated as a measure of QoS. In the case where the target field is inaccessible, random deployment is the only method. However, it cannot guarantee the required coverage and connectivity unless the node density is higher than certain threshold [47]. If the target field is accessible, a deterministic sensor deployment strategy should be more effective. Such a strategy would minimize the total number of sensors required and achieve the specific needs of applications in terms of their expected quality of coverage and connectivity.

Most existing works on sensor deployment with guaranteed coverage and connectivity are based on simplistic sensing and communication models, such as the boolean disc model [47–49], with sensing range R_s and communication range R_c. Under disc model, it is well known results that

when $R_c \geq 2R_s$, k-coverage of a convex region implies k-connectivity [48]. However, this result may not hold under probabilistic sensing and communication models. Although disk model facilitates a geometric treatment to the coverage and connectivity problem, it fails to consider collaboration in sensing and stochastic nature of radio communication. Our work is based on a probabilistic sensing model and a shadowing fading link model, which are more realistic compared to boolean disk model.

In this chapter, we take into consideration the geographic feature of the AoI and the target event. In other words, different points of interest (PoIs) in the AoI can be of different importance due to their geographic relation to the events. If there is some *a priori* knowledge about the importance of each PoI, the deployment plan can be more customized. For example, in case of a fire detection application, PoIs around the oil depot are considered highly risky and should be assigned high coverage probability, while the PoIs near the water area can be placed lower importance. Intuitively, locations with higher detection requirement need to be deployed more sensors than those with lower requirement. Meanwhile, the network connectivity should be guaranteed in a probabilistic manner otherwise more power has to be used to ensure communication. This deployment issue can be formulated as a multi-objective optimization problem (MOP). The essential requirement for MOP is to find Pareto-optimal solutions under multiple decision objectives and constraints, from which the user can make a trade-off choice according to their own preference and some practical limits. Among the existing solutions, genetic algorithm (GA) has been recognized as one of the possibly well-suited to MOP [50]. In this chapter, We propose to use NSGA-II [50] to develop the framework for solving the MOP in sensor deployment, and compared the performance results with those obtained by tabu search (TS) heuristic. NSGA-II is one of the most popular GAs and has been shown effective in a number of applications [51, 52]. The algorithm can be run off-line *a priori* and the sensor nodes are distributed following the final computational results. In case the user changes a small part of the detection requirement, it is not efficient to run the algorithm once again over the whole area. We propose to restrict genetic operations in the varied subareas while evaluate the fitness function globally. Simulation shows that this method converges more quickly with little degradation in performance and the variation of sensor nodes are much less than the global way.

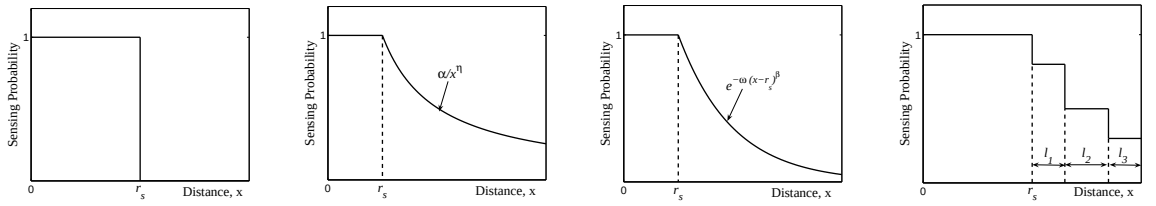
The rest of the chapter is organized as follows. Section 3.2 reviews related work including sensing models and current deployment strategies. Section 3.3 introduces the related models used in this chapter and formulate the problem. In Section 3.4 and Section 3.5, TS and NSGA-II based deployment strategies are specified and analyzed respectively. We present simulation results in Section 3.6 and point out open issues and challenges in Section 3.7. Finally we conclude the chapter in Section 3.8.

3.2 Related Work

3.2.1 Sensing Models

A sensor transforms environmental stimuli into electrical signals. The quality of the resulting signal mainly depends on three factors. First is the distance between the sensor and the actual

event. For example, the amplitude of sound waves decreases quadratically with increasing distance to the location of the sound-producing event. When an acoustic sensor node has a very large distance to the sound event, the sensor readings become indistinguishable from the case of no sound event occurring at all. A second aspect of sensing quality is directionality. In an idealized scenario, a sensor has the same sensitivity in all directions; however, in practice, often certain directions are preferred. This can be either by construction (for example, visual sensor) or as a result of sensor deployment, e.g. a node's acoustic sensor is obstructed by other node components. A third aspect is constituted by the possibility that the same sensor can generate different outputs for the same environmental stimulus at different times, for example, due to temperature variations the sensing circuitry is exposed to. Generalizing this observation, the signal delivered by a sensor for an external event at a certain distance is not a fixed value, but a distance-dependent random variable. Most sensing models used in the literature focus on the first aspect and assume omnidirectional sensing and no random variations. These assumptions are generally reasonable and make some problem less complicated. Some commonly used sensing models are illustrated in Figure 3.1.



(a) Boolean disk model [53] (b) Power model [54] (c) Exponential model [55] (d) Staircase model [56]

FIGURE 3.1: Some of the sensing models used in the literature.

Boolean disk model :

$$S(s_i,P) = \begin{cases} 1, & d(s_i,P) \leq r_s \\ 0, & d(s_i,P) > r_s. \end{cases} \tag{3.1}$$

Power model :

$$S(s_i,P) = \begin{cases} 1, & d(s_i,P) \leq r_s \\ \frac{\alpha}{d(s_i,P)^{\eta}}, & r_s < d(s_i,P) < r_u \\ 0, & d(s_i,P) \geq r_u. \end{cases} \tag{3.2}$$

Exponential model :

$$S(s_i,P) = \begin{cases} 1, & d(s_i,P) \leq r_s \\ e^{-\omega(d(s_i,P)-r_s)^{\beta}}, & r_s < d(s_i,P) < r_u \\ 0, & d(s_i,P) \geq r_u. \end{cases} \tag{3.3}$$

Staircase model :

$$S(s_i,P) = \begin{cases} 1, & d(s_i,P) \leq r_s \\ S_{l_1}, & r_s < d(s_i,P) \leq l_1 \\ S_{l_2}, & l_1 < d(s_i,P) \leq l_2 \\ S_{l_3}, & l_2 < d(s_i,P) \leq l_3. \end{cases} \tag{3.4}$$

3.2.2 Overview of Deployment Methods

Deployment strategies can be classified into three categorizations : random deployment, regular deployment and on-demand deployment. The choice of the deployment scheme depends highly on the type of sensors, application and the environment that the sensors will operate in. Controlled node deployment (regular deployment and on-demand deployment) is viable and often necessary when sensors are expensive or when their operation is significantly affected by their position. Such scenarios include populating an area with highly precise seismic nodes, underwater WSN applications, and placing imaging and video sensors. Instances include "a line in the sand" [4] sensor network for target tracking, CitySense [57] network for urban monitoring, Soil Monitoring [58] sensor network etc., wherein sensors are hand-placed at selected spots prior to network operation. On the other hand, in some applications random distribution of nodes is the only feasible option. This is particularly true for harsh environments such as a battle field or a disaster region. Depending on the node distribution and the level of redundancy, random node deployment can possibly achieve the required performance goals. In the following we describe each methodology and highlight their strengths and limitations. Besides, popular strategies and techniques for node deployment in WSNs can be found in a survey paper [59].

Random deployment : In applications of WSNs in reconnaissance missions during combat and disaster recovery, deterministic deployment of sensors is very risky and/or infeasible. It is widely expected that sensors will be dropped by helicopter, grenade launchers or clustered bombs. Such means of deployment lead to random spreading of sensors. Although the node density can be controlled to some extent, the specific requirements of applications are uncertain to be guaranteed. Random deployment can be modeled by point process, for example, Poisson point process (see definition 2.1) or binomial process. Figure 3.2 and Figure 3.3 illustrates the node spatial distribution generated by homogeneous and inhomogeneous Poisson point process (Poisson cluster process) respectively. Poisson point processes are popular, for example, for modeling the number of stars in a certain space area or the number of bacteria cultures on a Petri dish. The striking feature of such a Poisson point process is that it matches the intuitive notion most people have of "random deployments".

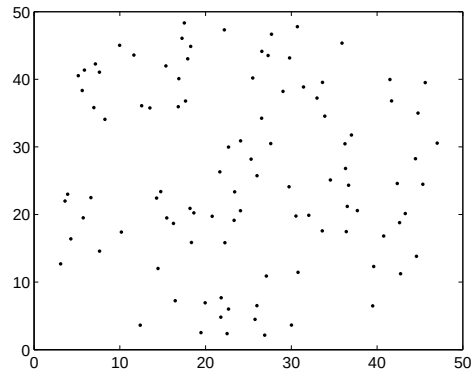


FIGURE 3.2: Random deployment generated by homogeneous Poisson point process.

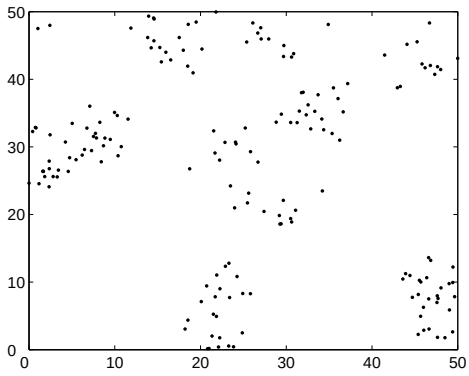


FIGURE 3.3: Random deployment generated by Poisson cluster process.

Regular deployment : In contrary to random deployment, regular deployment follows some regular geometric patterns. In WSNs, the regular deployment method is commonly adopted in an indoor environment or somewhere easy for people to access. There are various regular patterns. Researchers are most interested in the pattern that can guarantee coverage and connectivity with minimal number of sensor nodes. For years, works in this research domain are based on a result presented in 1939, which states that the regular triangular lattice pattern (triangle pattern in short) is asymptotically optimal in terms of the number of discs needed to achieve full coverage [60]. This result naturally provides six connectivity only when $R_c \geq \sqrt{3}R_s$. However, general values of R_c/R_s in practice can be anything positive. For example, while the reliable communication range of the Extreme Scale Mote platform is 30m, the sensing range of the acoustics sensor for detecting an All Terrain Vehicle is 55m [61]. In this case, R_c/R_s is far less than $\sqrt{3}$. Progresses in exploring optimal patterns under different values of R_c/R_s have been made in recent years. In [53] and [49], the asymptotically optimal pattern to achieve k-connectivity ($k \leq 6$) and full coverage is explored. Figure 3.4 and Figure 3.5 illustrate two popular regular deployment patterns : square lattice and equilateral triangle lattice.

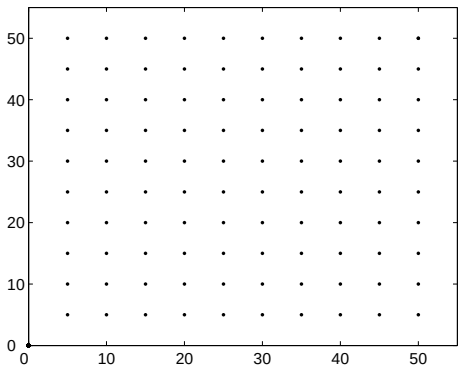


FIGURE 3.4: Regular deployment pattern following square lattice.

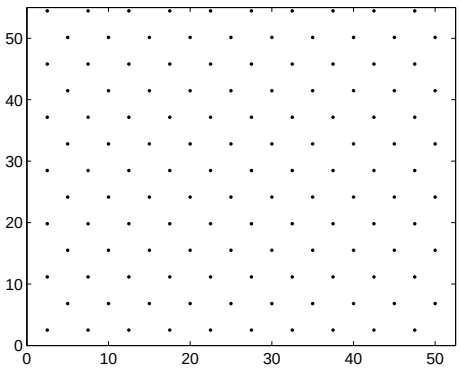


FIGURE 3.5: Regular deployment pattern following equilateral triangle lattice.

On-demand deployment : Although random deployment and regular deployment are effective in some environment, they can not provide service guarantee to meet differentiated requirements of applications. On-demand deployment is specifically designed to meet the requirement of the application. The final deployment pattern is usually obtained by optimization of several objectives. The most common objectives include increasing the coverage, achieving strong network connectivity [62], extending the network lifetime [63] and/or boosting the data fidelity. A number of secondary objectives such as tolerance of node failure [64] and load balancing have also been considered. Most of the work strives to maximize the design objectives using the least amount of resources, e.g., number of nodes, since typical sensor nodes still cost around $100 each [25].

In this chapter, we particularly cares about the performance of coverage and connectivity. In the literature, the art gallery problem [65,66] aims to use the minimum number of guards to watch a polygon area, which is identical to the deployment problem. However, it is assumed that a guard can watch any point as long as line-of-sight exists and it does not address the communication

issue between guards. The problem of finding optimal deployment with guaranteed coverage and connectivity has been addressed in several literatures. Bai *et al*. [49] propose optimal deployment patterns to achieve full coverage and k-connectivity ($k \leq 6$) under different ratios of sensor communication range over sensing range. However, their work is under the assumption of disk sensing and communication model. Under non-disc models, their "optimal" deployment patterns may be too conservative. Several studies [67–69] model a sensing field as grid points and discuss how to place sensors on some grid points to satisfy certain coverage requirements. Reference [67] discusses how to place two types of sensors with different costs and sensing ranges such that every grid point is covered by sensors and the total cost is minimized. The work in [68] considers a probabilistic sensing model and discusses how to place sensors in a field possibly with obstacles such that every grid point is covered with a minimum confidence level. In [69], the objective is to place sensors to ensure that every grid point is covered by different sensors ; the result is to distinguish from different grid points for localization applications. However, these works do not address the relationship between sensors's communication range R_c and sensing range R_s and most of them use Boolean disk sensing model.

A few literatures have replaced disk sensing model by probabilistic [70–72] or stochastic models. Zou *et al*. [70] present two heuristic algorithms to optimize the number of sensors while minimizing the overall miss probability and they specifically deal with the problem of uncertainty in sensor locations. In [71], Zhang *et al*. formulate the differentiated deployment problem as an integer linear programming problem and also propose a heuristic algorithm to solve it. Propositions in [70, 71] only satisfy coverage requirement but do not consider network connectivity. Aitsaadi *et al*. [72] use TS to find optimal deployment with differentiated coverage and connectivity. Their approach is testified to outperform solutions in [70, 71]. However, they adopt disk model as the communication model and their final solution has no diversity. There are also literatures addressing mobile sensor deployment [73], while in this work, we only focus on static sensor node deployment, i.e., sensors do not move after deployed.

3.3 Preliminaries and Problem Formulation

3.3.1 Sensing and Communication Models

We adopt exponential sensing model as described in 3.3. Specifically, the probability that a target/event at point P be sensed by sensor s_i is :

$$c_{s_i}(P) = \begin{cases} 1, & h(s_i,P) \leq r_s \\ e^{-\omega(h(s_i,P)-r_s)^\beta}, & r_s < h(s_i,P) < r_u \\ 0, & h(s_i,P) \geq r_u. \end{cases} \tag{3.5}$$

where $h(s_i,P)$ is the distance between s_i and P, ω and β are parameters that measure the decay of sensing probability with distance, r_s and r_u are two thresholds denoting boolean sensing radius and maximum sensing range respectively. This model reflects the behavior of range sensing devices such as infrared and ultra sound sensors.

Let **S** be the set of nodes whose sensing range cover the point P, expressible as $\mathbf{S} = \{s_i | h(s_i, P) < r_u\}$. The total coverage probability of the point P, also referred to as 1-coverage probability, is defined as :

$$C_{\mathbf{S}}(P) = 1 - \prod_{i=1}^{\|\mathbf{S}\|}(1 - c_{s_i}(P)) \tag{3.6}$$

which is the probability that point P is covered by at least one sensor.

For radio link model, we adopt PRR-based link model as described in Chapter 2, Section 3.8. It is expressible as :

$$p(h) = \left(1 - \frac{1}{2}\exp\left(-\frac{\gamma(h)}{2}\frac{1}{0.64}\right)\right)^{8f} \tag{3.7}$$

The explanation and value of the notations can be found in Chapter 2, Section 3.8.

3.3.2 Problem Formulation

Figure 3.7 illustrates the deployment problem. The to-be-deployed area is associated with differentiated requirements for quality of converge. The lighter color implies a higher value of coverage requirement. After deployment, the requirement should be satisfied and the resulting network should meet the connectivity requirement as well. The to-be-deployed area $\mathscr{A}$ is first discretized into $X \times Y$ grids, as shown in Figure 3.6. The grid dimension depends on the user's precision requirement and computational capability . If the area is highly fine-grained, each grid can be regarded as a PoI. To facilitate mathematical description, we refer to each grid by its barycenter. The distance between two grids is defined as the distance between their barycenters. As mentioned before, the coverage requirement of different PoI in the area $\mathscr{A}$ is different characterized by their geographic relation to events or targets. Thus, we assume that each PoI P_i in $\mathscr{A}$ is associated with a minimum coverage threshold, denoted R_i, $0 \leq R_i \leq 1$. The deployment strategy can thus be denoted by a $X \times Y$ boolean matrix D, with the element $d_i = 1$ indicating a deployed sensor in respective PoI while $d_i = 0$ no sensor deployed. After deployment, each PoI is actually covered with probability C_i, defined by (3.6). We use $G(V,E)$ to represent the communication graph after deployment, in which V is the vertex set and E is the edge set denoting communication links. Finally, under given connectivity requirement R_{conn}, we formulate the MOP as :

$$\begin{aligned} &min \;\; N = \sum_{i=1}^{X\times Y} d_i, \\ &min \;\; \sum_{i=1}^{X\times Y} \Delta_i, \\ &s.t. \;\; Conn(G(V,E)) \geq R_{conn}, \\ &d_i \in \{0,1\}. \end{aligned} \tag{3.8}$$

where

$$\Delta_i = \begin{cases} (R_i - C_i)/R_i, & R_i > C_i\,; \\ 0, & R_i \leq C_i. \end{cases} \tag{3.9}$$

and *Conn* is defined as the one-connectivity probability of the resulting network. In problem (3.8), the first objective is to minimize the number of nodes while the second objective is to minimize the difference between the required coverage threshold and the after-deployment coverage probability. The constraint is that the 1-connectivity probability of the resulting network should be greater than user requirement R_{conn}.

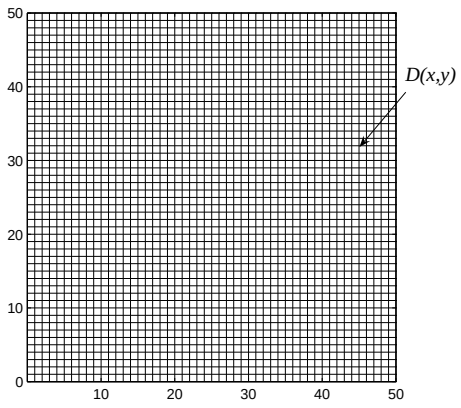
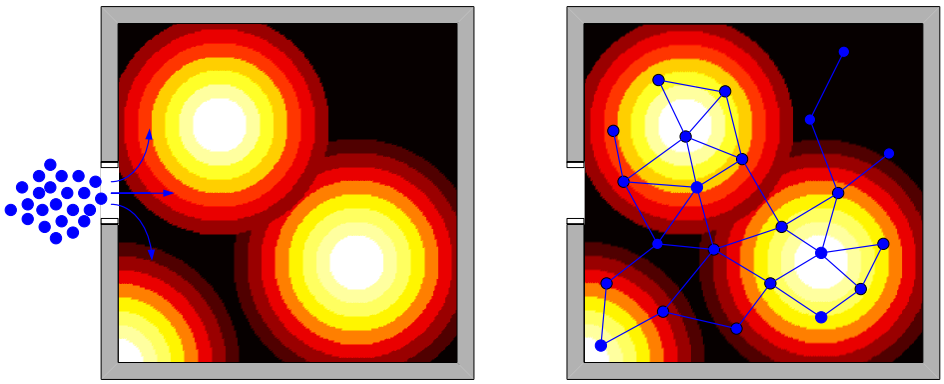


FIGURE 3.6: Illustration of the area discretization.



(a) Before deployment (b) After deployment

FIGURE 3.7: Illustration of the deployment problem.

The dimension of the solution space of the above problem is $2^{X \times Y}$. To obtain the optimal or sub-optimal solutions in polynomial time, we resort to heuristic approaches. In the following, we use two different multi-objective heuristic methods to solve the problem : TS heuristic and NSGA-II. The pros and cons of each algorithm are analyzed and the simulation results are compared in Section 3.6.

3.4 Method 1 : TS Based Deployment Strategy

TS is a local search optimization technique which tries to minimize a cost function $F(s)$, by iteratively moving from a solution s_i to a solution s_{i+1} in the neighborhood of s_i (according to a neighborhood function $H(s_i)$) until a stopping criterion is satisfied or a predetermined number N of iterations is reached.

3.4.1 Initial Solution

The convergence of TS method depends on the judicious choice of the initial solution. Ideally, the first solution has to be close to the optimal one, otherwise, since the maximum number of iterations is fixed, the algorithm may stop before reaching the best solution. Consider that the decision, $D(x,y)$, of deploying a sensor in a point $P(x,y)$ is a random variable, which follows a Bernoulli distribution with parameter θ. The binary form of the decision rule motivates the choice of the Bernoulli law. Precisely, $D(x,y)$ can assume a value of 1 with a probability of θ and the value of 0 with a probability of $1-\theta$. The parameter θ, associated to a point $P(x,y)$, is chosen as the percentage of the points located in the vicinity of $P(x,y)$ and not receiving the required probabilities of detection. The vicinity, denoted $\mathscr{S}$, is defined as the set of neighbor points located inside the maximum monitoring circle of a sensor which would be placed in P. Formally,

$$\theta = \frac{1}{|\mathscr{S}(P_i)|} \sum_{i=1}^{|\mathscr{S}(P_i)|} 1_{\{C_i<R_i\}}, \tag{3.10}$$

where $|\mathscr{S}(P_i)|$ is the number of nodes in $\mathscr{S}$, and $1_{\{.\}}$ is the indicating function, which is equal to 1 if the condition is true and 0 otherwise. The initialization of the TS approach follows these steps :

1. The initial TS solution is started assuming zero deployed sensors. Thus, for any P, Bernoulli parameters are computed using equation (3.10) with $C(x,y) = 0$.
2. Generate a list, L, including all points of A. L is a decreasingly sorted list of points according to their Bernoulli parameters.
3. In L, select the point P with the highest Bernoulli parameter, and remove it from the list. If the actual detection probability $C(x,y)$ associated to P is lower than R, then a decision to deploy a sensor in P is randomly generated through a Bernoulli decision rule with parameter θ.
4. If the Bernoulli decision is to deploy a sensor ($D(x,y) = 1$), then (1) the probabilities of detection for all points in the vicinity of P (the set $\mathscr{S}$) are recomputed, and (2) L is updated.
5. If L is not empty, go back to Step 3).

When the stop criterion of step 5) is satisfied, the resulting positions of deployed sensors are considered as the initial solution s_0. The latter one is saved in the algorithm's memory, called the tabu list. The goal of the tabu list is not to block the method in a local minimum of the cost function. The size of the tabu list is denoted by $tabu\_length$.

3.4.2 Neighborhood Exploration

After obtaining the initial solution, we begin the procedure of neighborhood exploration. The neighborhood exploration will be conducted for *stop_length* times, which is to limit the iteration rounds. The neighborhood of a solution s_i is defined as the solution set that can be generated by solution s_i through some basic transformations. We propose two neighboring generation methods, namely suppression oriented stage and additional oriented stage. These two methods alternate in the successive iterations of our TS approach. Both stages are detailed hereafter.

Suppression oriented stage :

The aim of this stage is to suppress some sensors among those deployed in over-covered areas. The method proceeds with the following steps :

1. Compute the Bernoulli parameters for all P using the equation (3.11) and assuming the deployment obtained in the last TS iteration.

$$\theta_2 = \frac{1}{|\mathscr{S}(P(x,y))|} \sum_{(i,j)\in\mathscr{S}(P(x,y))} \left[\left(1 - \frac{R(i,j)}{C(i,j)}\right) \times 1_{\{C(i,j)<R(x,y)\}} \right] \tag{3.11}$$

2. Generate a list, L_{supp}, including all the points where a sensor is deployed ($D(x,y) = 1$). The list is then decreasingly sorted according to θ_2.
3. In L_{supp}, select and remove the point with the highest θ_2, and randomly generated the decision to suppress the sensor through a Bernoulli decision rule with parameter θ_2 if the connectivity requirement is satisfied, $conn(s_i) > R_{conn}$.
4. If the Bernoulli decision is to suppress the sensor ($D(x,y) = 0$), then the probabilities of detection for all points in the intersection of sensing and communication vicinity of P are recomputed. In this case, the list L_{supp} is updated.
5. If L_{supp} is not empty, go back to Step 3). Once the stop criteria on step 5) is satisfied, the next TS iteration alternates toward an additional stage.

Additional oriented stage :

The aim of this stage is to add more sensors to the actual deployed ones in under-covered areas. The execution is very similar to the initialization stage, except step 1) and step 2) are replaced by the following steps :

1. Compute the Bernoulli parameters for all P using the equation (3.10) and assuming the deployment obtained in the last TS iteration.
2. Generate a list L_{add} including all the points where a sensor is not deployed. The list one is then decreasingly sorted according to the resulting Bernoulli parameters.

Steps 3), 4) and 5), as detailed in initialization stage, are repeatedly executed until that the list L_{add} is empty.

3.4.3 Cost Function

After the neighborhood exploration during the nth iteration, an elected solution must be chosen among the explored candidates with number of Ng. This solution, which cannot be in the tabu list, is the one which provided by minimizing a given cost function F. The cost function reflects two objectives of the optimization problem described in 3.8, minimize the number of deployed sensors and reduce the gap between the generated and requested event detection probabilities. The first objective could be quantified by counting the number of deployed sensors. Formally, as defined in section $D(x,y) = 1$, if a sensor is deployed in the point $P(x,y)$. Otherwise, $D(x,y) = 0$. In order to minimize the cost function, F includes the following term :

$$Cost\_supp = \sum_{\{x,y\} \in \mathscr{A}} \Delta(x,y) \tag{3.12}$$

$$Cost\_add = \sum_{\{x,y\} \in \mathscr{A}} D(x,y) + \sum_{\{x,y\} \in \mathscr{A}} \Delta(x,y) \tag{3.13}$$

According to the expression of penalty function, a successful deployment solution should lead to obtain detection probabilities higher than (or ideally equal to) the required detection thresholds. If it is not satisfied, the penalty function value translates how far is the solution to the required thresholds. This is exactly the second objective of the optimization problem. From the above objective expressions, the authors define two cost functions *Cost_supp* and *Cost_add*. The function *Cost_supp* is used to choose the best next iteration solution in the case of a suppression oriented stage. The function *Cost_supp* is formulated using only the equation (3.12). On the other hand, the cost function *Cost_add* is used in the case of additional oriented stage. In this case, the two both terms associated to each objective of optimization problem are integrated into the cost function.

The pseudo code of TS based deployment strategy is illustrated in Algorithm 33.

3.5 Method 2 : NSGA-II Based Deployment Strategy

There are a number of heuristic methods possible to solve problem (3.8), among which we choose a popular genetic algorithm, NSGA-II. As well-known, GAs have been successfully applied in many areas, especially in numerical and combinatorial optimizations [51,52]. NSGA-II follows some basic steps of GA. The population is first initialized and then sorted based on nondomination into each front by using a fast sorting algorithm. Each individual in each front is assigned a fitness (or rank) value equal to its nondomination level. Once the nondominated sorting is completed the crowding distance is also assigned. Parents are selected from the population by using binary tournament selection based on the rank and crowding distance. The offspring population is combined with the current generation population and selection is performed to set the individuals of the next generation. The selected parents generate offspring by

Algorithm 3.4.1: TS

Compute the initial solution s_0 ;
$s_{current} = s_0$;
$tabu\_list = \{s_0\}$;
$i = 0$;
bool $suppress\_sensor = true$;
while $i < stop\_length$ **do**
    $neighbor\_solution = \varnothing$;
    if $suppress\_sensor == true$ & $conn(s_i) > R_{conn}$ **then**
        for j=1 **to** N **do**
            $s_i^j = Fun\_supp(s_i)$;
            $neighbor\_solution = neighbor\_solution + \{s_i^j\}$;
            $Cost(s_i^j) = Cost\_supp(s_i^j)$;
        end
    else
        for j=1 **to** N **do**
            $s_i^j = Fun\_add(s_i)$;
            $neighbor\_solution = neighbor\_solution + \{s_i^j\}$;
            $Cost(s_i^j) = Cost\_add(s_i^j)$;
        end
    end
    for j=1 **to** Ng **do**
        if $s_i^j \in tabu\_list$ **then**
            $neighbor\_solution = neighbor\_solution - \{s_i^j\}$;
        end
    end
    find s_i^{best} in the *neighbor_solution*, which satisfy $Cost(s_i^{best}) = min(Cost(s_i^{1 \to N}))$;
    $s_{i+1} = s_i^{best}$;
    if $Cost(s_i^{best}) < Cost(s_{current})$ **then**
        $s_{current} = s_i^{best}$;
    end
    $tabu\_list = tabu\_list + \{s_i^{best}\}$;
    $suppress\_sensor = false$;
end

using crossover and mutation operators. The new generation is controlled by each front subsequently until the population size exceeds the current population size. Since all the previous and current best individuals are added in the combined population, elitism is ensured in the NSGA-II.

3.5.1 Initial Population

Each GA requires an initial population P_0 to serve as the starting point. To have more diversity in the initial population, we use both random and greedy approaches to generate P_0. Random deployment is calculated as follows :

$$d_i = \begin{cases} 1, & rand < K \cdot R_i\,; \\ 0, & \text{otherwise} \end{cases} \tag{3.14}$$

in which K is a tunable parameter controlling the node density and $rand \in [0,1]$. The greedy approach is similar to the initialization procedure in TS. The decision to deploy a sensor on P_i follows Bernoulli distribution $\mathscr{B}(\theta)$, whose parameter θ associated to P_i is calculated as follows :

$$\theta = \frac{1}{|\mathbf{S}(P_i)|} \sum_{i=1}^{|\mathbf{S}(P_i)|} 1_{\{C_i < R_i\}}, \tag{3.15}$$

where $|\mathbf{S}(P_i)|$ is the number of nodes in $\mathbf{S}$, and $1_{\{\cdot\}}$ is the indicating function, which is equal to 1 if the condition is true and 0 otherwise. First, there is no node deployed. For any PoI P_i, $C_i = 0$. Select a random PoI and compute the value of θ following (3.15). Second, generate a list L to include all points of $\mathscr{A}$ in decreasing order of the value of θ. Third, select the point P_i with highest θ and remove it from L, and compare the actual sensing probability C_i with the requirement R_i. If $C_i < R_i$, deploy a sensor in P_i following $\mathscr{B}(\theta)$. If the decision is to deploy a sensor, recompute the coverage probability in the sensing range of P_i and update L. Repeat the third step until the list L is empty.

3.5.2 Objective Evaluation, Fast Sorting and Crowding Distance Comparison

After the initial population is generated, we compute all the objectives for each individual solution and evaluate the constraints, following equation 3.16.

$$\begin{aligned} F_1 &= \sum_{i=1}^{X \times Y} d_i, \;\; F_2 = \sum_{i=1}^{X \times Y} \Delta_i, \\ Ct &= Conn(G). \end{aligned} \tag{3.16}$$

After computing the objectives and constraints, each individual solution in the population is ranked with nondominated criteria and sorted into different fronts by using a fast sorting algorithm. Crowding distance is computed for differentiating the solutions of the same rank. Thus the diversity among nondominated solutions can be obtained.

3.5.3 Genetic Operations

An intermediate population of size n_p is created by employing the following genetic operators :

Selection : First, parents are selected from the population by using a binary tournament selection based on the rank and crowding distance. This is based on the principle that parents with better chromosomes can reproduce better offsprings.

Crossover : We choose multiple crossover points, whose locations are calculated using a random number generator (RNG), to create a new population with probability p_c.

Mutation : Newly reproduced chromosomes are transferred to the mutation pool. With mutation probability p_m, randomly chosen chromosomes are mutated. A new population is thus generated. To achieve good convergence, p_c is usually set a larger value than p_m.

After the genetic operation, we combine the parents and offspring populations into a whole population of size $2n_p$ and continue the procedures in section 3.5.2 to obtain a new population of size n_p, which is then used for the genetic operation in section 3.5.3. The iteration stops when the termination criterion, e.g. a maximum generation *gen*, is reached.

Algorithm 3.5.1: Genetic algorithm

```
1  Generate the initial population P_0, size pop_size ;
2  Evaluate objective values F_1(P_0), F_2(P_0) ;
3  sort(P_0) ;
4  for i = 0 to max_gen do
5  |   for j = 1 to pop_size do
6  |   |   select two parents p_1 and p_2 from P_i ;
7  |   |   offspring_j = selection(p_1, p_2) ;
8  |   |   with a probability α_c, perform
9  |   |       offspring_j = crossover(p_1, p_2) ;
10 |   |   with a probability α_m, perform
11 |   |       offspring_j = mutate(p_1, p_2) ;
12 |   end
13 |   Q_i = offspring ;
14 |   R_i = P_i ⋃ Q_i, size 2 * pop_size ;
15 |   ℱ = sort(R_i) ;
16 |   P_{i+1} = replace(ℱ) ;   \* select pop_size elements from ℱ *\
17 end
```

3.5.4 Deployment Result Update

Suppose an optimal solution has been obtained using the above NSGA-II algorithm. There is a question that when the coverage thresholds at some PoIs vary, is it required to run the above algorithm once again ? As illustrated in Figure 3.8, the coverage thresholds of PoIs inside the ellipse vary while those outside the ellipse remain the same. One way is to take it as a brand new issue and follow the original steps to reach an optimal solution. Due to the nature of genetic

algorithm, this treatment may result in a major change of the node deployment over the whole AoI. In order to preserve the original placement of most sensors, another way is to clip out the varied subarea and run the algorithm independently in the subarea. However, this will bring a problem that nodes outside the subarea cannot contribute to the coverage or connectivity inside, which may result in larger number of nodes than required. Therefore, we propose to do genetic operations in the varied subarea while evaluate the objectives still in a global way. This will ensure that nodes outside the varied subarea cooperate with nodes inside in both sensing coverage and communications.

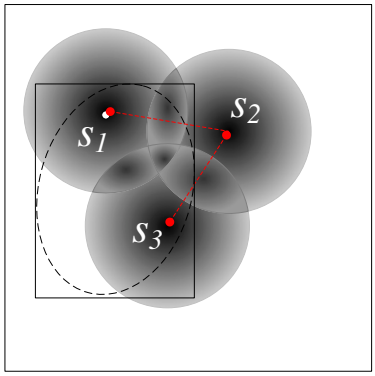


FIGURE 3.8: The coverage requirement in the circled area is varied.

3.6 Performance Evaluation

In this section, we conduct numerical experiments to evaluate the performance of the deployment strategy based on TS and NSGA-II. We compare the two algorithm with random and grid deployment. In case the coverage threshold in a small area varies after deployment, both global and local operations are conducted to redeploy the sensors. In addition, we compare the efficiency of the strategies according to multiple metrics.

3.6.1 Simulation Settings

Unless otherwise stated, we use the following settings : (i) $\omega = 0.4$, $\beta = 1.2$, $r_s = 2$m, $r_u = 10$m. (ii) The transmission power increases from -16.9dB to -3.2dB, corresponding to equivalent communication ranges from 5m to 11m. The equivalent sensing and communication range under disk models are calculated as below :

$$\begin{aligned} ER_s &= \int_0^{r_u} c \, dh \\ ER_c &= \int_0^{+\infty} p \, dh \end{aligned} \tag{3.17}$$

where c is the sensing probability of a sensor within distance h defined in (3.5), and p is the communication probability defined in (3.7). Following (3.17), the equivalent sensing and communication range ($P_t = -6.9$dB) under disk models equal to 4m and 9m respectively, as shown in Figure 3.10 and Figure 3.10. (iii) The deployment region is discretized into 50×50 grids and the center of each grid is a PoI. The coverage thresholds of the AoI are non-uniformly distributed and vary from 0.1 to 0.99, as illustrated in the first sub-figure of Figure 3.18, with light color indicating a large threshold value while dark color on the contrary. (iv) Random deployment approach follows (3.14) and the best is chosen out of 10 tries. For TS based strategy, $stop\_length = 100$, $tabu\_length = 10$, $Ng = 15$; For NSGA-II based strategy, $n_p = 50$, $p_c = 0.75$, $p_m = 0.1$, $gen = 50$; (v) The connectivity constraint is set to 95%.

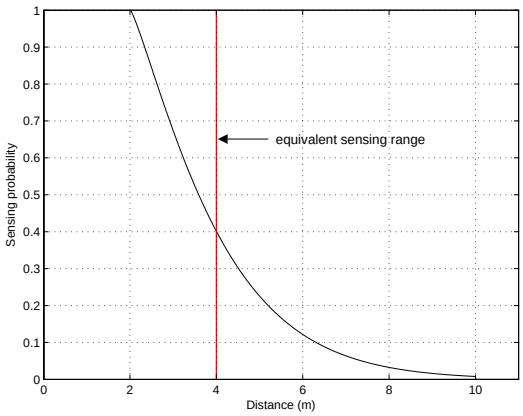


FIGURE 3.9: Sensing model, indicating equivalent sensing range under disk model.

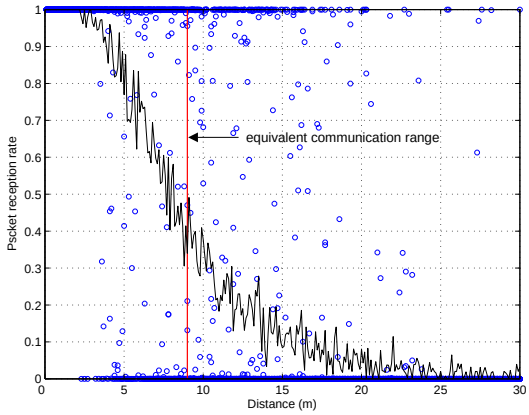


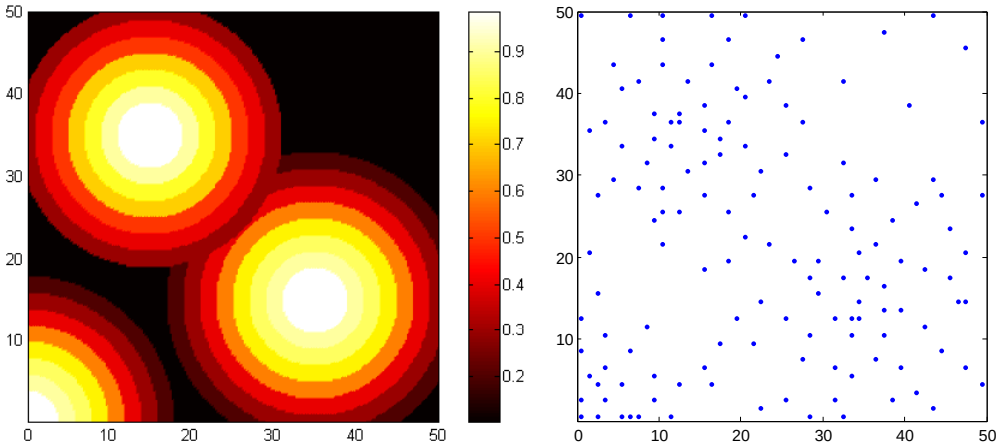
FIGURE 3.10: Communication model, indicating equivalent communication range under disk model, with $P_t = -6.9$dB.

3.6.2 Simulation Results

Figure 3.11 is one of the results obtained by TS. It is clear that nodes are densely deployed in the area with high sensing requirement, while sparse in the area with lower requirement. However, for a specified requirement, TS can provide one optimal solution. It is not flexible for multi-objective optimizations.

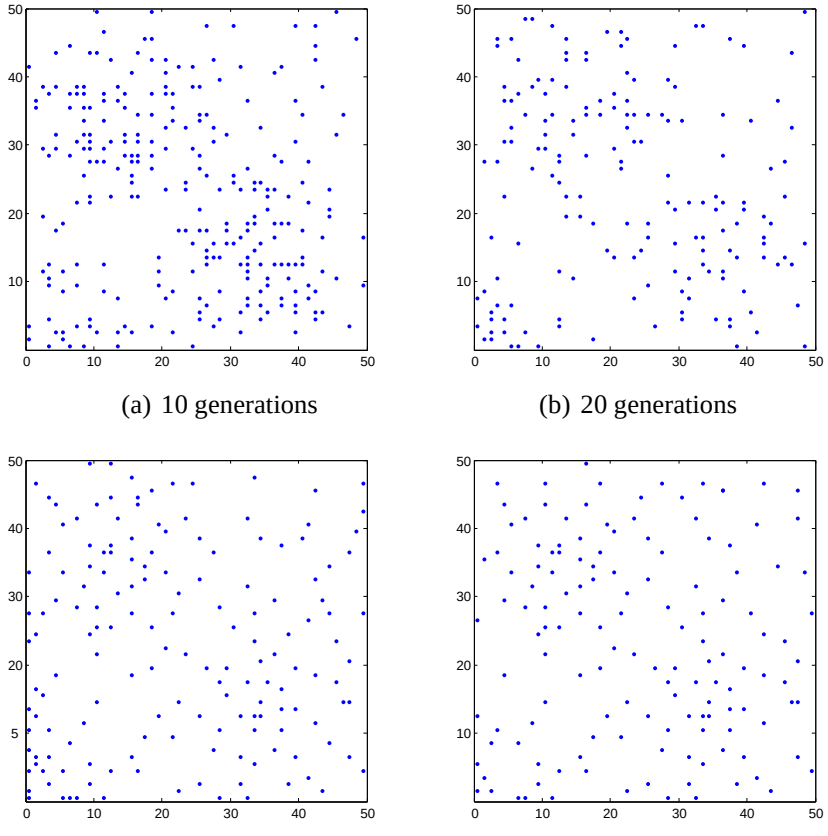
Fig. 3.12 captures a snapshot of the evolution procedure of GA. Clearly, as the evolution generation increases, the resulting deployment gradually approaches the optimal. It is worth noting that Fig. 3.12 only shows one of the diverse solutions while through GA, we can obtain a group of non-dominated Pareto solutions. The application user is able to make tradeoff choices between different solutions according to their practical considerations.

Figure 3.13 plots the number of nodes required by four deployment methods under different ER_c/ER_s ratios. For each method, as the ER_c/ER_s ratio increase from 5/4 to 11/4, the number of nodes required gradually decreases and finally converges to a minimum value. This is because



(a) (b)

FIGURE 3.11: The deployment strategy obtained by TS.



(a) 10 generations (b) 20 generations

(c) 30 generations (d) 50 generations

FIGURE 3.12: Evolution procedure of GA.

the sensing parameters remain unchanged. There should be a minimum number of nodes guaranteeing the coverage requirement. Further increase in ER_s cannot do any help to the coverage but result in a waste of power instead.

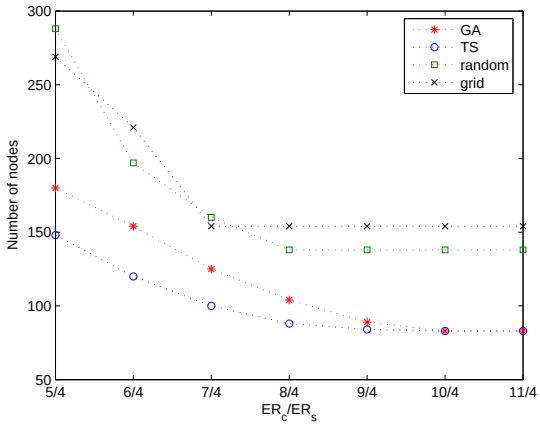


FIGURE 3.13: Required node number of vs. ER_c/ER_s ratio.

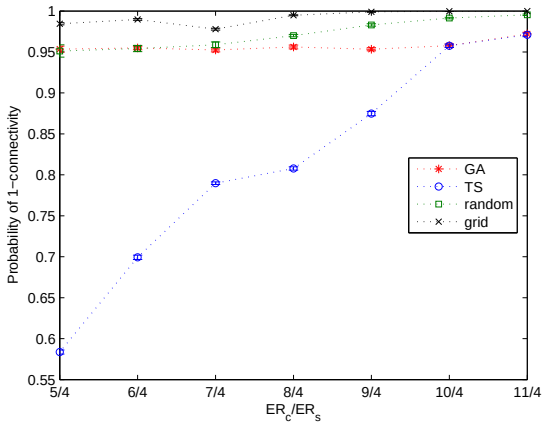


FIGURE 3.14: 1-connectivity probability vs. ER_c/ER_s ratio.

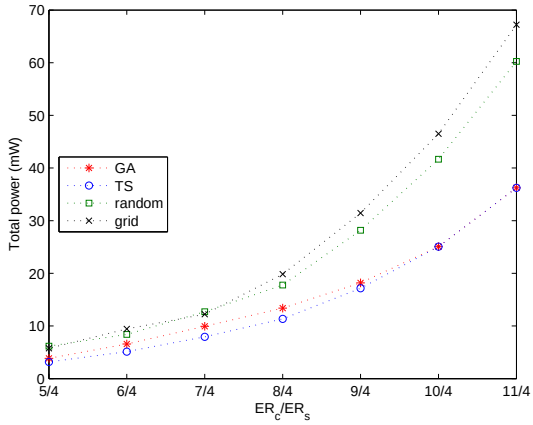


FIGURE 3.15: Total power of different deployment vs. ER_c/ER_s ratio.

Although TS based strategy requires least sensors, it cannot satisfy the connectivity requirement as shown in Figure 3.14. This is because TS approach does not support multi-objective optimization problem very well. The strategy proposed in [72] ensures a connected network under boolean model in the initialization phase. However, it does not maintain the connectivity in the afterward searching procedure. The 1-connectivity probability is plotted in Figure 3.14, which indicates that the resulting network by GA approach can ensure a connected network with the probability around 95%.

As ER_c/ER_s increases, more transmission power is required. As shown in Figure 3.15, the total power for the network increases in spite of the number of nodes decreasing. It is worth noting that in case $ER_c/ER_s < 2$, the increase in the total power is gentle while in case $ER_c/ER_s > 2$,

the increase becomes steep. This implies that there should be a tradeoff between transmission power setting and the total number of nodes required, preventing the total power from going to the steeply increase range.

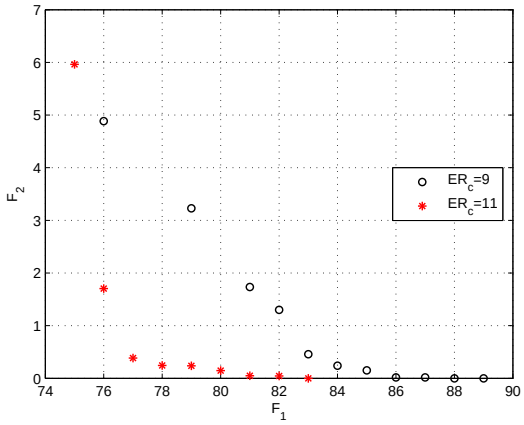


FIGURE 3.16: Nondominated solutions with $ER_c = 9$ and $ER_c = 11$, $Ct \geq 95\%$.

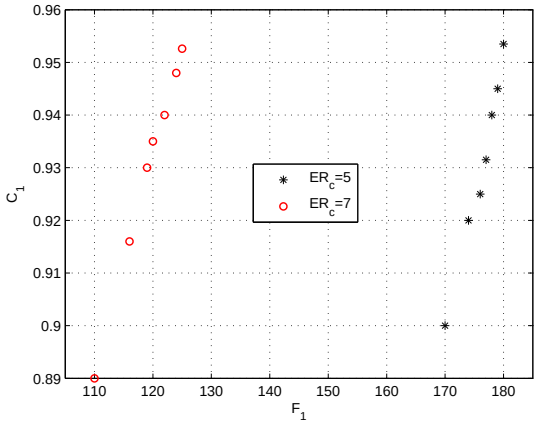


FIGURE 3.17: Nondominated solutions with $ER_c = 5$ and $ER_c = 7$, $F_2 = 0$.

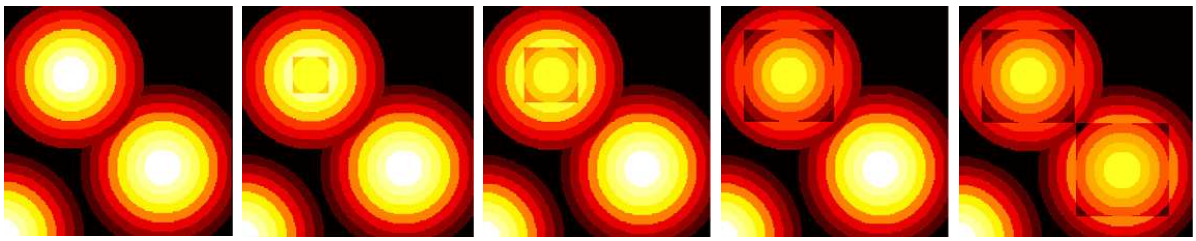


FIGURE 3.18: The coverage requirement in subareas is varied : four situations.

Figure 3.16 and Figure 3.17 show the obtained nondominated solutions using NSGA-II. In Figure 3.16, all the solutions satisfy the constraint Ct, nondominated solutions with tradeoffs in F_1 and F_2 are presented. Similarly in Figure 3.17, all the solutions satisfy $F_2 = 0$, nondominated solutions with tradeoffs in F_1 and Ct are given for users to choose an appropriate one according to their practical conditions and limits.

In case that the coverage requirement in some subarea vary, we do both global and local optimizations for the updated problem. As shown in Figure 3.18, the first sub-figure denotes the original requirement, while the 4 sub-figures afterwards are new requirement with subareas variation. Figure 3.19 shows respective performance. We can see that the total number of nodes required is almost the same for the two approaches. Local approach converges more quickly and it is easier to find a novel solution while the global approach which searches in a wider range, lacks of efficiency and sometimes may not reach the same level as local approach with the same time complexity. An obvious advantage of local approach lies in the varied number of nodes in the network. For global genetic operation, it is highly possible that the deployment in the area without threshold variation be destroyed to satisfy a global search of optimal solution. On the contrary, local approach restricts the search in the localized area. Thus the placement outside is

not affected. Since the fitness function is still evaluated in a global way, the local approach will not degrade too much, if compared with the optimal solutions. Both local and global approaches satisfy the connectivity constraints as shown in Figure 3.20.

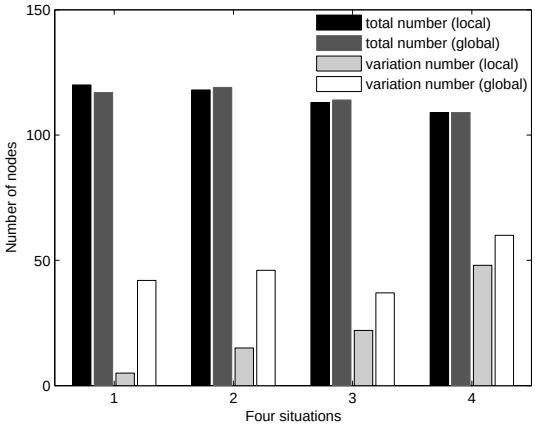


FIGURE 3.19: Total and varied number of nodes by local and global computations

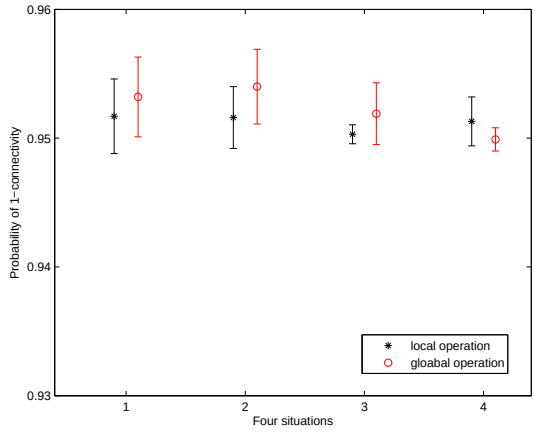


FIGURE 3.20: 1-connectivity probability by local and global computations

3.6.3 Complexity Analysis

The worst-case complexity of NSGA-II is $O(mn_p^2)$, where m is the number of objectives while for TS it is $O(kXY)$, where k is a constant related to iteration time and sensing range. It is worth noting that if the deployment area is large or the grid granularity is small, the complexity of NSGA-II based strategy will not be affected but the complexity of TS based method will increase. In this sense, NSGA-II based deployment strategy is more adaptive and robust to the dimension and granularity of the area.

3.7 Open Issues and Research Challenges

In this section, we highlight open research problems, identify the issues involved and report the ongoing work. First, the genetic operator used in this work is relatively simple and direct. However, to achieve higher searching efficiency, more complicated operators should be explored, e.g. hybrid algorithm for the genetic operations. Second, this work studies the deployment problem in a 2D area, which can be extended to 3D situations. However, there may be new challenging issue in 3D deployment problem such as description of 3D converge and connectivity, and the enlarging of the solution space. Third, in this work we did not consider the sensor scheduling. However, for many long-term unattended applications, energy constraint is a major stumbling block that limits the long-term sustainability of WSNs. Therefore, an intuitive way to save energy is to make the sensor nodes work at a low duty cycle, in which they schedule themselves to be active for a brief period of time and then stays dormant for a long time as illustrated in Figure 3.21, where T_w is the sleeping duration for each sensor in each period and t_1, t_2, $t3$ are the waking time of each senor. Our ongoing work is to simultaneously optimize the deployment and scheduling of sensor nodes so that application-specified requirements (converge, connectivity, network lifetime etc.) can be satisfied.

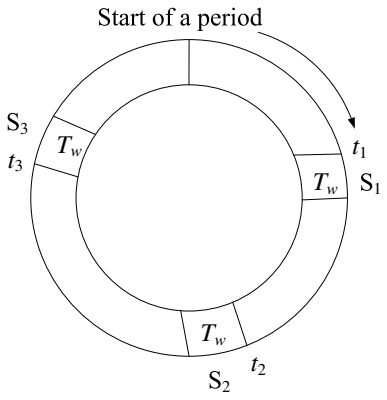


FIGURE 3.21: Illustration of a cyclic sleep schedule for 3 sensors.

3.8 Summary

The deployment strategy for achieving differentiated coverage and probabilistic connectivity in wireless sensor networks is studied in this chapter. The deployment problem is formulated into a multi-objective optimization problem and multi-objective metaheuristics are proposed to solve the problem. Solutions based on TS heuristic and NSGA-II are specified and analyzed. Simulation results show that NSGA-II based strategy can meet the desired coverage requirements and maintain connectivity in a probabilistic manner with a relatively small number of sensors. It also provide diverse solutions for tradeoff choices between multiple objectives. In addition, for the applications in which the coverage requirement varies in some subareas, a local genetic operation is more time-efficient and needs less variation in the original disposal than a renewed

global optimization. Our solutions to the deployment issue can also be applied to solve on-duty sensor set selection and scheduling problem if the sensors have already been densely deployed.

4

Routing with Real-Time QoS Support

4.1 Introduction

In WSNs, as single node has limited power, it is common for the data being transmitted to the base station through multi-hops. While we have discussed how to maintain coverage and connectivity in the former chapters, and suppose now nodes have sensed the data we are interested and the network connectivity is fine, it is the responsibility of the routing protocol to relay the data through multi-hops to the BS within certain delay bound. Supporting real-time QoS in WSN can be addressed from different layers and mechanisms, among which routing protocol has always played a crucial role in supporting end-to-end QoS.

Providing real-time QoS in network layer is to enable transmissions of periodic or sporadic messages within predefined deadlines in a reliable fashion ; timeliness is especially important for crucial alarm messages. Out-of-date data are usually irrelevant and may even lead to negative effects to the system monitoring and control. Since the wireless channel is random and time-varying, conventional deterministic QoS measures should be replaced by probabilistic ones. An important performance measure is the deadline miss ratio (DMR) which is defined as the ratio of messages that cannot meet deadlines [74]. Moreover, sensor nodes usually use battery for energy supply. Hence, energy efficiency is also an important design goal. It is usually defined by the energy consumed per successfully transmitted packet. Furthermore, in order to avoid network topology holes and achieve a longer network lifetime, node load and energy balance need to be considered. The design in this chapter is oriented to more demanding applications which emphasize packet delivery timeliness and end-to-end QoS, e.g., alarm messages should be transmitted from sensor nodes to control center in time so as to take prompt actions [5]. Energy efficiency and load balance are also among the design goals.

It is known from the literature that for system simplicity most existing routing protocols are based on 1-hop neighborhood information. However, it is expected that multi-hop information can lead to better performance in many issues including routing, message broadcasting and channel access scheduling [75–77]. For computing 2-hop neighborhood information in wireless ad hoc and sensor networks, some distributed algorithms and efficient information exchange

schemes are reported in [78, 79]. In a network of n nodes, computing 1-hop neighbors with $O(n)$ messages is trivial while computing 2-hop neighbors seems to increase the complexity and overheads. However, a complexity analysis reported in [78] has shown that every node can obtain the knowledge of 2-hop neighborhood by a total of $O(n)$ messages, each of $O(\log n)$ bits, which could be enough to address the ID and geographic position of nodes.

It is very likely that a system can perform better if more information is available and effectively utilized. By a preliminary study of the asymptotic performance of a generic routing with multi-hop routing information, it is observed that the number of hops required from the source to sink decreases significantly from 1-hop to 2-hop information based routing. However, the further gain from 2-hop based decision to 3-hop based decision is less attractive, especially if complexity increase is also taken into account. In this chapter, we propose a 2-hop information based real-time routing protocol and show its improvement over 1-hop based protocol SPEED [80]. The choice of two hops is a tradeoff between performance improvement and the complexity cost. The idea of 2-hop routing is straightforward but how to use or integrate the information effectively so as to improve energy and real-time performance is generally nontrivial. The resulting design has the following novel features :

1. Compared with existing protocols that utilize only 1-hop neighborhood information, it achieves lower deadline miss ratio and also higher energy efficiency.
2. A mechanism is embedded which can release nodes that are frequently chosen as packet forwarder. An improvement of energy balance throughout the network is achieved.
3. The simulation is built on Mica2-based [25] lossy link model, energy model and CSMA/CA MAC setting (similar to B-MAC [20]) which are very close to real systems.

The rest of this chapter is organized as follows. Section 4.2 discusses related routing protocols for real-time QoS in WSN and explains the motivations. Section 4.4 presents our design. The performance of proposed protocol is reported in Section 4.5. Simulations and comparisons have shown its effectiveness. In Section 4.6, we discuss possible enhancement. Finally, Section 4.7 concludes the chapter.

4.2 Related Work

4.2.1 Real-time Related Routing Protocols

Generally speaking, there are three classes of routing policies that favor end-to-end delay performance guarantee in WSNs : (i) tree based routing, (ii) optimal routing based on shortest-path-first (SPF) principle by the knowledge of whole network topology, and (iii) geographic routing by the knowledge of node position.

Tree based routing is popular in industrial WSN setting. ZigBee has provided a hierarchical tree routing scheme in which packets travel along the edges of the tree network. This approach suits the many-to-one traffic model and does not need routing table. End-to-end QoS (delay, energy consumption, etc.) can be estimated by the depth of the tree. However, the hierarchical tree routing can be very inefficient when two nodes in different branches but mutual radio range want

to communicate with each other since packets must travel through the ZigBee coordinators. AODV routing is thus suggested as a supplement in this case. As proposed in [81], another solution is to look up the neighbor table in routing decisions so as to avoid long path and thus shorten the worst-case delay. Another drawback of tree based routing is the problem of node energy consumption balancing. Nodes near the root of the tree will consume much more energy than the other and consequently lead to network topology holes.

In [82], a tree-based QoS-aware protocol is proposed for WSNs. It finds a least-cost and energy-efficient path that can meet end-to-end delay requirement during the connection. In addition, a class-based queueing model is employed to serve both best-effort and real-time traffics. Their approach however does not consider the impact of channel access delay. Besides, the use of class-based priority queuing mechanism is too complicated and costly for resource limited sensor nodes.

A tree routing protocol is often not optimal as it does not choose the shortest path. AODV is one of the optimal routing based protocols by SPF principle. However, additional overhead (e.g., extra packet and energy consumption) will be introduced in order to maintain the routing table. AODV is a reactive routing which is more favorable when communication is required infrequently. The route discovery on demand adds additional latency to packet transmission. This has been investigated in [83] and an AODV variant is proposed after introducing a new routing metric in evaluating path efficiency which includes end-to-end delay and energy consumption. As a result, the network lifetime is prolonged and end-to-end delivery ratio is improved for real-time embedded systems.

Geographic routing is popular in WSNs since it does not need to maintain routing table and consequently can reduce network energy consumption. Resulting algorithms are highly scalable [84]. Different greedy forwarding strategies can be used in geographic routing protocols as shown in Figure4.1. However, geographic routing protocols are in general not optimal since most of them are based on 1-hop decision. In addition, determining node position will introduce some overheads and energy consumption. Several solutions exist for finding coordinates, e.g., using GPS in outdoor environments. Note that for resource-limited WSNs, using GPS can be a problem as the required positioning chips will increase the price and energy consumption. This problem can be alleviated by using positioning chips only in some nodes, while other nodes calculate positions with the assistance of their neighbors. On the other hand, existing localization techniques such as triangulation, multilateration and diffusion [84] can provide GPS-free solutions. Some ranging techniques have also been specified in the IEEE 802.15.4a standard [85], e.g., estimating distance by measuring the difference of propagation delays.

1. Compass routing : Let t be the destination node. Current node u finds the next relay node v such that the angle $\angle vst$ is the smallest among all neighbors of s in a given topology.
2. Random compass routing : Let v_1 be the node on the above of line st such that $\angle v_1st$ is the smallest among all such neighbors of s and v_2 to be nodes below line st that minimizes the angle $\angle v_2st$. Then node s randomly choose v_1 or v_2 to forward the packet.
3. Greedy routing : s finds the next relay node v such that the distance vt is the smallest among all neighbors of s in a given topology.

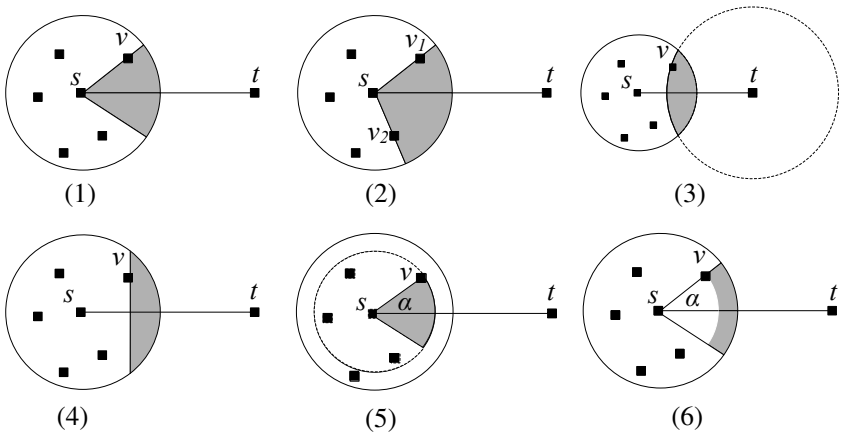


FIGURE 4.1: Various forwarding methods for geographic routing.

4. Most forwarding routing : Current node s finds the next relay node v such that $\|v't\|$ is the smallest among all neighbors of s in a given topology, where v' is the projection of v on segment st.
5. Nearest neighbor routing : Given a parameter angle α, node s finds the nearest node v as forwarding node among all neighbors of s in a given topology such that $\angle vst \leq \alpha$.
6. Farthest neighbor routing : Given a parameter angle α, node s finds the farthest node v as forwarding node among all neighbors of s in a given topology such that $\angle vst < \alpha$.

In geographic routing, the heuristic greedy forwarding protocol SPEED [80] is the first one addressing real-time guarantees for WSNs. Relay velocity toward a next-hop node is identified by dividing the distance progress by its estimated forwarding delay. Packet deadline is mapped to a velocity requirement. The node with the largest relay velocity higher than the velocity requirement is selected in the highest probability. If there is no neighbor node that can meet the requirement, the packet is dropped probabilistically to regulate network workload. Meanwhile, back-pressure packet re-routing in large-delay link is conducted to divert and reduce packets injected to a congested area. MM-SPEED [15] extends SPEED by defining multiple delivery velocities for packets with different deadlines in supporting different QoS. RPAR [16] is another variant of SPEED. A node will adaptively change its transmission power by the progress towards destination and packet's due time in order to meet the required velocity in the most energy-efficient way. Note that all the above protocols are based on 1-hop neighborhood information.

In our proposed scheme, we also adopt the approach of mapping packet deadline to a velocity, which is known as a good metric to delay constrained packet delivery. However, our routing decision will be made based on 2-hop neighborhood information and corresponding metrics. It is therefore named as Two-Hop Velocity based Routing (THVR). Note that generally speaking it is also possible to employ other metrics, e.g., by packet lifetime or hop count, to design routing protocols. The idea of 2-hop information based routing is generic and applicable. Here, we will focus on THVR. The routing design and details are given in Section 4.4. The performance comparisons of current popular real-time routing protocols are listed in Table 4.1.

TABLE 4.1: A comparison of the discussed routing protocols

Name	Type	Link reliability	Energy efficiency	Scalability
Zigbee routing [81]	cluster-tree	N/A	N/A	good
improved AODV [83]	SPF	considered	high	low
EA-QoS [82]	SPF	N/A	high	low
SPEED [80]	geographic	N/A	N/A	good
MMSPEED [15]	geographic	high	N/A	good
RAPR [16]	geographic	considered	high	good

4.2.2 Computing 2-hop Neighborhoods

In this section, we discuss how to compute 2-hop neighborhoods through a distributed algorithm. Before discussing the algorithm, two important definitions are introduced as follows :

Definition 4.1 *A MIS (Maximal Independent Set) of a graph $G = (V,E)$ is a subset $S \subseteq V$ such that $\forall e = \{u,v\} \in E$, either $u \in S$ or $v \in S$.*

Definition 4.2 *A CDS (Connected Dominating Set) of a graph $G = (V,E)$ is a subset $C \subseteq V$ such that each node in $V - C$ is adjacent to some node in C, and C induces a connected subgraph.*

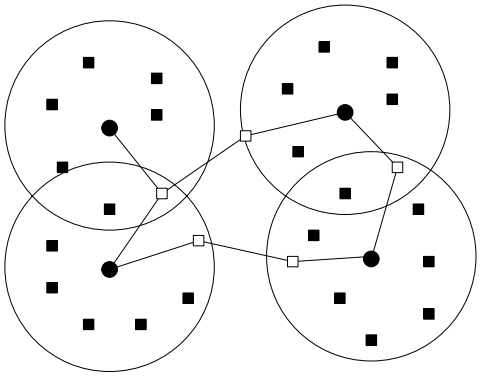


FIGURE 4.2: Illustration of MIS and CDS, where black circles are MIS nodes which are connected through empty squares, forming a CDS [78].

It has been proved in [78] that CDS can be constructed distributely with $O(n)$ messages, where the message length is $O(\log n)$. Nodes first get their 1-hop neighbor's ID and positions through broadcasting, then compute their 2-hop neighbors by Algorithm 17.

Theorem 4.1 *In a network with total number of n nodes, each node can obtain its 1-hop neighborhoods with the necessary condition that the network exchange O(n) messages in total.*

Proof. Each node broadcasts their ID and position information as a message. Since the number of neighbors of each node is always less than the total node number n, this theorem holds. ■

Theorem 4.2 *In a network with total number of n nodes, each node can obtain its 2-hop neighborhoods with the necessary condition that the network exchange O(n) messages in total.*

Proof. See literature [78]. ■

Algorithm 4.2.1: Compute two-hop neighbors

```
1  node v sends a packet with its < ID, position > to adjacent MIS node ;
2  MIS node broadcasts the packet with < ID, position, counter >, set counter = 2 ;
3  while counter > 0 do
4  |   if common node hears the packet then
5  |   |   if v is in its 1-hop list then
6  |   |   |   deletes the packet ;
7  |   |   else
8  |   |   |   add v to its 2-hop list ;
9  |   |   end
10 |   else if CDS node hears the packet then
11 |   |   if v is in its 1-hop list then
12 |   |   |   retransmit the packet, counter − 1 ;
13 |   |   else
14 |   |   |   add v to its 2-hop list ;
15 |   |   end
16 |   end
17 end
```

With theorem and theorem, we can draw the conclusion that there is no increase in the order of magnitude of the communication overheads from computing 1-hop neighborhoods to 2-hop neighborhoods. Actually, the knowledge of 2-hop neighborhoods has been assumed in many protocols and algorithms for routing [86], broadcasting [76] and media access [79].

4.3 Performance Gain of Geographic Routings with *k*-Hop Neighborhood Information

In this section, we extend the conventional geographic greedy routing to the utilization of *k*-hop neighborhood information to investigate the potential improvement when more information is allowed. In principle, the neighbor node which can forward the packet by the most progress is selected as the next hop forwarder. Given *k*-hop neighborhood information, the one in next forwarding which will lead to a node, after *k*-hop routing, closest to the destination is chosen. In generalization, it is chosen by its potential advantage in terms of *k*-hop transmissions identified by the neighborhood information in the corresponding range. The determination follows the same concept of preference in the proximity to destination.

As we know, in geographic routing, a selection of forwarding candidate only by the preference of shortest distance to destination can simply fall into a routing deadlock [86]. In other words, the packet may be passed between two nodes at the dead end. This can be resolved by many

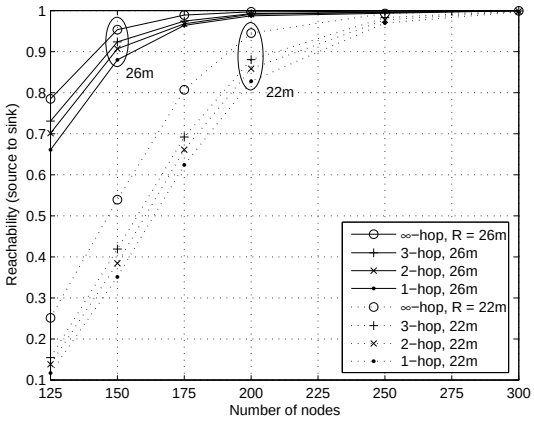


FIGURE 4.3: End-to-end reachability in *k*-hop searching.

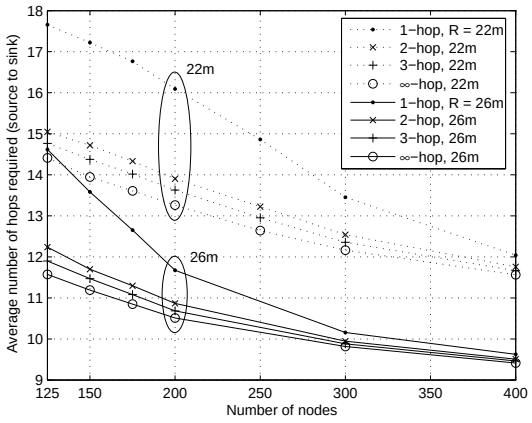


FIGURE 4.4: Average number of hops required in *k*-hop searching.

different techniques. A simple and effective solution is to blacklist the visited end node and eliminate it from the forwarding candidate list since a loop-back occurs. Here, we follow this approach. Besides, in this Section, the link model adopted is simply a boolean disk, which means if the packet transmission is in the effective radio range, R, the delivery is considered successful.

In the performance comparison, the number of hops or transmissions required from source to destination (sink) in the packet forwarding is in our major concern as it reflects both the routing delay and energy consumed during transmissions. Here, we will not take into account MAC and queueing delay in order to focus on the issue of routing. The reachability to be reported soon is defined by the ratio of packets from source which can reach the sink finally. When we find the packet is always buffered at a specific node or possibly looped in an isolated region and thus cannot go into the right direction after a large number of hops, it will be considered as routing failure. This indicator can reflect the routing capability of a path searching scheme. However, significantly, it also depends on the node density and can help to indicate the level of network connectivity.

Nodes in the WSNs are uniformly distributed in a geographical area of (200m, 200m), while the source and sink are fixed at the location of (30m, 30m) and (170m, 170m) respectively. Clearly shown in Figure 4.3, as the number of nodes increases, the reachability from source to sink increases as well generally due to the resulting better network connectivity. We conduct two sets of simulations with different effective radio range, R, of 26m and 22m respectively, and plot in parallel for reference. As expected, when the number of nodes is insufficient, the reachability will be low even by ∞-hop searching due to disjoint clusters or low connectivity between source and sink. It is therefore unsuitable for network formation.

As shown in Figure 4.3, when the number of nodes is above 150 in the case $R = 26$m, the reachability is at a satisfactory level approximately equal to or above 0.9. Reasonably, in the case $R = 22$m, a higher node density is required to reach this reachability. Detailed studies of critical density for wireless network connectivity can be found in [34, 87]. Here, we will focus

on the comparison among different *k*-hop searching results. As observable in Figure 4.3, there is an improvement in reachability from 1-hop to 2-hop path searching and also an improvement from 2-hop to 3-hop searching by the fact that some routing difficulties can be earlier identified and better resolved, or even avoided, when more neighborhood and connectivity information is provided. However, generally speaking, the improvement is decreasing as the number of nodes is increasing. For example, the reachability tends to 1 in all the curves when the number of nodes goes to 300.

It should be noted that the number of hops required from source to sink is an important indicator from the point of view of end-to-end packet delay and energy consumption. This is at the core of our study. As shown in Figure 4.4, when the number of nodes increases, the number of hops required decreases as there can be more and better forwarding choices to reach the sink with potentially less hops. An important observation here is that there is a significant improvement in the number of hops required from 1-hop to 2-hop searching, while the improvement from 2-hop to 3-hop searching and even more is relatively small. For example, we can clearly see this in the case $R = 26$m and the number of nodes equals to 200, in which the reachability is almost 1 as known from Figure 4.3. Generally speaking, in this system, 2-hop searching is a good option in different k values by its attractive gain in the reduction of required number of hops. Meanwhile, the system complexity is concerned. Note that, as expected, the gain will get decreased when the number of nodes increases. In addition, by comparing the curves of $R = 26$m and $R = 22$m, as the effective radio range is reduced, under same node density, the number of hops required increases generally as in Figure 4.4 since the network is now in a lower connectivity. In Figure 4.3, the reachability in $R = 22$m drops significantly, especially in the region of relatively low node density, e.g. 125 to 175.

To have a better understanding on what kind of scenarios the 2-hop searching will take a smaller number of hops than the 1-hop searching does, we study the node topology and resulting routing paths by details. Figure 4.5 and 4.6 depict two typical examples observed. As shown in Figure 4.5, both searching can find a routing path from source to sink eventually. However, the two searchings will go into different paths since node X. Node V is chosen as the next forwarder in the 2-hop searching scheme instead of node Y, although Y is in fact closer to the sink. The reason is the 2-hop searching has identified that in X's 2-hop range, via V, node W is even closer to sink than node Z. So, node V is better than node Y from the point of view of 2-hop resulting progress. However, the 1-hop searching will simply choose node Y. This effect often helps the 2-hop searching to drive into a shorter path than that of the 1-hop searching.

Figure 4.6 shows an example of the 1-hop searching in which a possible routing deadlock occurs and is finally resolved. After forwarded to node Y from node X, the packet will be routed back to X. As now Y will be blacklisted, the packet is thus forwarded to node Z although Y is at a closer position to sink than Z. However, by the 2-hop searching, when the packet arrives X, in the next forwarder selection, the potential progress in 2-hop range will be checked. Clearly, Y is found as a dead end as it will be back to X. The 2-hop searching can thus bypass Y and forward the packet from X to Z directly despite Y is in fact closer to sink than Z. However, comparatively, the 1-hop searching will encounter more hops (transmissions) and consequently a delay in the deadlock scenario due to the difference in capability. It is worth pointing out that although the 1-hop and 2-hop searching may enjoy close reachability, the number of hops

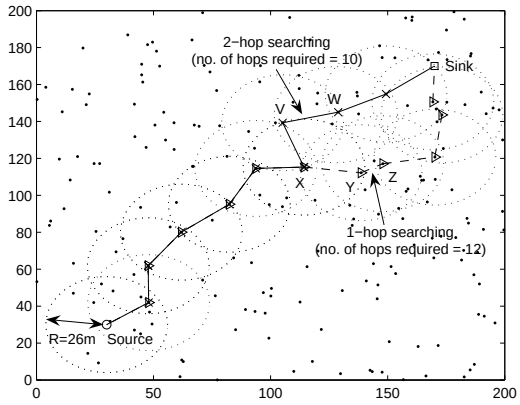


FIGURE 4.5: An example to show the typical scenario in which the 1-hop searching encounters more hops than the 2-hop searching does.

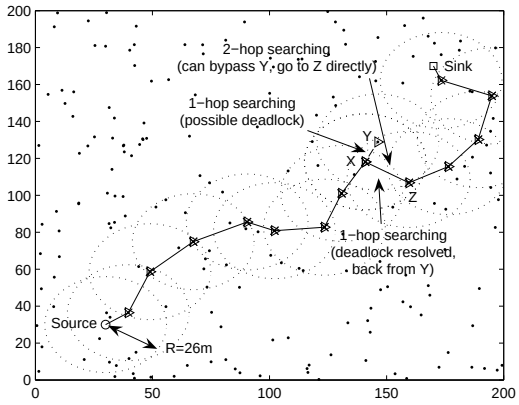


FIGURE 4.6: During the possible deadlock, the 1-hop searching scheme will encounter more transmissions although the deadlock can be resolved.

required from source to sink can be quite different generally.

Inspired by the study in this section, we design 2-hop information based real-time routing protocol named THVR.

4.4 Design of THVR for RT-WSNs

Although 2-hop information based routing is intuitively helpful to improve the routing decision, an explicit mechanism is necessary. It is worth noting that THVR primarily aims at lowering packet DMR for demanding real-time WSNs but will also consider energy utilization efficiency that has not been explicitly addressed in SPEED and MM-SPEED.

As assumed in most geographic routing algorithms, each node in the network is aware of the geographic location of itself and the destination, via GPS or other localization techniques [84, 85] as mentioned in Section 4.2. The information can be further exchanged among 2-hop neighbors [79, 88]. Thus, each node is aware of its immediate and 2-hop neighbors, and their locations. This is achieved by two rounds of HELLO messages. First, each node informs its neighbors about its existence (ID, position, remaining energy, etc.). Next, each node sends message to all its neighbors informing about its 1-hop neighbors. If the network is static or with low mobility, this could be done at one stroke until there is node failure. Otherwise in a mobile network, each node periodically emits additional HELLO messages to maintain 2-hop information. Too old entries are removed from the neighbor table, as corresponding nodes have moved out of 1-hop or 2-hop range.

To be detailed below, our design is mainly composed of three components : (i) forwarding metric, (ii) delay estimation and update, and (iii) initiative drop control.

4.4.1 Forwarding Metric

To begin with, some definitions are introduced. For each node i, $\mathbf{N}(i)$ is used to denote the set of its 1-hop neighbors. The source and destination nodes are labeled by S and D respectively. The distance between a pair of nodes i and j is denoted by $d(i,j)$. Consequently, the required end-to-end packet delivery velocity for deadline, t_{set}, is defined as :

$$S_{set} = \frac{d(S,D)}{t_{set}}. \tag{4.1}$$

$\mathbf{F}(i)$ is defined as the set of node i's potential forwarders which will make a progress towards the destination, i.e.,

$$\mathbf{F}(i) \triangleq \{j | d(i,D) - d(j,D) > 0, j \in \mathbf{N}(i)\}. \tag{4.2}$$

$\mathbf{F_2}(i,j)$ is defined to represent the set of corresponding 2-hop potential forwarders, i.e.,

$$\mathbf{F_2}(i,j) \triangleq \{k | d(j,D) - d(k,D) > 0, j \in \mathbf{F}(i), k \in \mathbf{N}(j)\}. \tag{4.3}$$

An illustration of node's neighbor set, 1-hop and 2-hop forwarder set is shown in Figure 4.7.

In SPEED, the core component SNGF (stateless non-deterministic geographic forwarding) works as follows. Upon receiving a packet, node i calculates the velocity provided by each of the forwarding nodes in $\mathbf{F}(i)$, which is expressible as :

$$S_i^j = \frac{d(i,D) - d(j,D)}{Delay_i^j} \tag{4.4}$$

where $j \in \mathbf{F}(i)$ and $Delay_i^j$ denotes the estimated hop delay between i and j. If there exists j such that $S_i^j \geq S_{set}$, it will be chosen as the forwarder with probability $P(j)$ following the discrete exponential distribution below [80] :

$$P(j) = \frac{(S_i^j)^K}{\sum_{j=1}^{N} (S_i^j)^K} \tag{4.5}$$

where N is the number of candidates in $\mathbf{F}(i)$ and K is a weighting exponent to tradeoff between load balance and optimal delivery delay. A larger K will lead to a shorter end-to-end delay while a smaller one can achieve a better load balance.

In our proposed THVR, similarly to SPEED, by 2-hop information, node i will calculate the velocity provided by each of the 2-hop forwarding pairs $\{\mathbf{F}(i), \mathbf{F}_2(i,j)\}$, i.e.,

$$S_i^{j \to k} = \frac{d(i,D) - d(k,D)}{Delay_i^j + Delay_j^k} \tag{4.6}$$

where $j \in \mathbf{F}(i)$ and $k \in \mathbf{F}_2(i,j)$. For node pair (j,k) satisfying $S_i^{j \to k} \geq S_{set}$, we denote it by set $\mathbf{S}$. Beyond comparing the potential forwarding velocities, we also take into account node's remaining energy level, and thus define the following new joint metric :

$$ve_i^{j \to k} = C \times \frac{S_i^{j \to k}}{\sum\limits_{(j,k) \in \mathbf{S}} S_i^{j \to k}} + (1 - C) \times \frac{E_j / E_j^0}{\sum_j (E_j / E_j^0)} \tag{4.7}$$

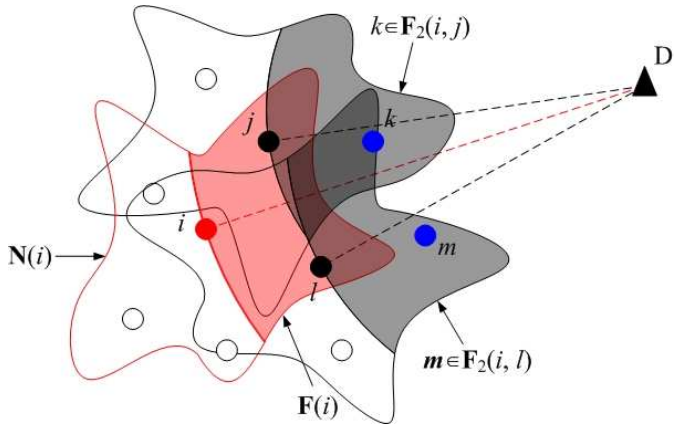


FIGURE 4.7: Illustration of node's neighbor set, 1-hop and 2-hop forwarder set.

where E_j is the remaining energy of forwarder candidate j while E_j^0 is its initial energy, and $C \in [0,1]$ is the weighting factor incorporating energy level into the joint metric. Note that larger C tends to favor end-to-end delay performance, while smaller one can distribute traffics to nodes in higher energy level and result in a better energy balance. Clearly, a setting of C relies on deadline requirements. The larger the deadline, the smaller C could be.

By (4.7), the node in $\mathbf{F}(i)$ (e.g., node j) with the largest *ve* will be chosen as the forwarder. The routing then proceeds and the mechanism is repeated at the selected node iteratively. In THVR, the sender will search the largest velocity in 2-hop neighborhood before making the forwarding decision. However, in SPEED [80], it is only 1-hop optimized. For example, if there is a topology hole after the first forwarding node, SPEED will get a critical problem and have to activate back-pressure re-routing. By THVR, this kind of problems can be alleviated. Inherently, THVR has 1-hop more prediction capability as using a "telescope" in finding the path. Generally speaking, even if the starting choice is not the globally optimized one, it may still have a better chance to gradually be corrected due to the farther sight and view.

4.4.2 Delay Estimation

From (4.6), it is observable that packet delay estimation from sender to its potential forwarder has played an important role in the velocity. In general, the delay of a packet from a node i to its immediate forwarder j is expressible as :

$$Delay_i^j = (Delay_{MAC} + Delay_{TX}) \times C_i^j \tag{4.8}$$

where $Delay_{MAC}$, $Delay_{TX}$, and C_i^j are used to represent the MAC delay, transmission time and transmission count respectively. The transmission time includes the queueing delay (depending on the load of the node) and the packet transmission time (determined by the packet/ACK size and the bandwidth). The transmission count refers to the number of retransmissions involved since ARQ (Automatic Repeat-reQuest) is adopted when packet fails to be transmitted due to collision or lossy link.

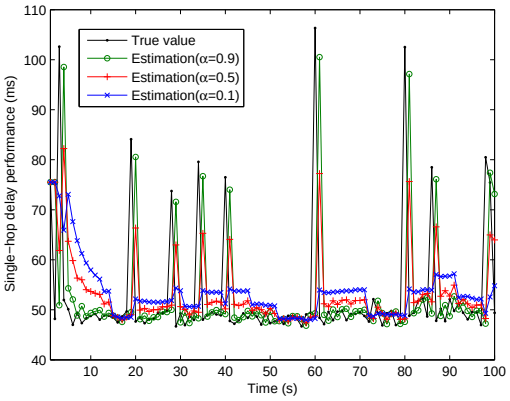


FIGURE 4.8: Delay estimation performance under different values of α.

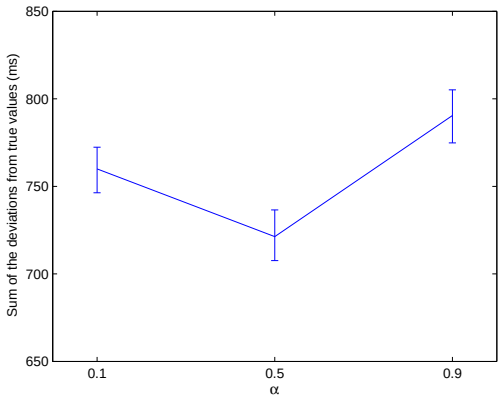


FIGURE 4.9: Deviation from true values under different values of α. The 90% confidence interval is also plotted.

To have packet delay estimation in identifying (4.6), we adopt the method of window mean with exponentially weighted moving average (WMEWMA), which has been shown in [29] with its best estimation performance among existing techniques similar to the round-trip time estimation of TCP. The estimate of $Delay_i^j$ for time instant $(t+1)$ is given by :

$$Delay_i^j(t+1) = \alpha M_i^j(t) + \frac{1-\alpha}{T} \sum_{k=\max(1,t-T)}^{t-1} Delay_i^j(k) \tag{4.9}$$

where T is the time window, $M_i^j(t)$ is the newly measured delay (known from the most recent packet), and $0 < \alpha < 1$ is the tunable weighting coefficient. It is clear that a large α will emphasize $M_i^j(t)$ and fits the case where delay variance is small, while a small α is more suitable if the variance is significant. A demonstration of the delay estimates under different α is plotted in Figure 4.8, while the sum of deviations is indicated in Figure 4.9. With a small α ($\alpha = 0.1$), the delay estimate is insensitive and too slow to capture the system's immediate fluctuation and thus may result in a big deviation sum. However, when α is too large ($\alpha = 0.9$), the update to the delay estimate appears too rigorous while nearly ignoring the historic average and results in an even larger deviation sum. As indicated in Figure 4.9, the deviation is the lowest when α is set to 0.5 which is quite robust generally. Note that it is also possible to design an adaptive tuning mechanism with reference to encountered delay variance. However, we will not go into the detail in this chapter.

To identify the link delay of a packet, a sender will stamp the time the packet is first sent and then compare that with the time when an ACK is received. On the other hand, to update the link delay information to other nodes in the routing path, after receiving the ACK with delay information from its forwarder, the node will initiate a feedback packet, which contains the updated delay of the forwarding link, to its parent node, i.e., the one who chose it as a forwarder. Meanwhile, other neighboring nodes which can overhear the feedback will also update their delay records. Figure 4.10 shows an example of link delay update when node G is chosen as the forwarder of node E. $Delay_E^G$ is first updated at E after receiving ACK from G and then feedback to node A.

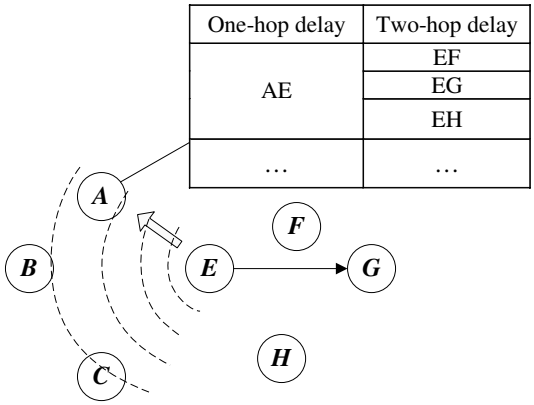


FIGURE 4.10: An example of two-hop delay update.

Nodes B and C overhear the feedback. As a result, the delay field EG in their records, e.g., a 2-hop delay table, will be updated by the new information and (4.9). It should be noted that the 2-hop information will enlarge the table of delay profile stored in each node. This needs to be considered if sensor nodes employed have very limited memory.

4.4.3 Initiative Drop Control

If no node in the 2-hop forwarding set can provide the required velocity, the following initiative drop control will be conducted. To begin with, some technical details are defined. Let e_j be the packet loss ratio of node j ($j \in \mathbf{F}(i)$) and N be the number of nodes in $\mathbf{F}(i)$.[3] We define the following forwarding probability of node i, denoted by u_i, as :

$$u_i = \begin{cases} 1 - K_1 \frac{\sum_{j=1}^{N} e_j}{N}, & \frac{d(i,D)}{d(S,D)} > \frac{1}{2} \\ 1 - K_2 \frac{\sum_{j=1}^{N} e_j}{N}, & \frac{d(i,D)}{d(S,D)} \leq \frac{1}{2} \end{cases} \tag{4.10}$$

where K_1 and K_2 are proportional gains with $K_1 > K_2 > 0$, $0 \leq u_i \leq 1$. The forwarding probability is jointly decided by the loss ratio in the forwarding set and the node position. Firstly, the node that is close to the destination has higher forwarding probability. This is designed by the fact that a packet near the destination has already traveled a long way along the routing and many nodes have consumed energy to relay it, thus it is worthwhile to try more efforts and see whether we can finally deliver it successfully. Although the current hop may not be able to meet the required velocity, it is possible to meet the end-to-end requirement finally if the coming hops may have relatively short delays. However, if the packet is still at a node near the source that cannot meet the velocity, from the point of view of energy utilization efficiency, it will be more efficient to drop it earlier.

Secondly, by (4.10), the node whose forwarding candidates have lower average loss ratio has higher forwarding probability. As shown in Figure 4.11, link layer collects the node packet loss

3. Note that e_j can be obtained by reading LQI (link quality indicator) field in some sensor nodes (e.g., MicaZ [25]).

ratio and feeds it back to the dropping controller, which calculates the forwarding probability according to (4.10). For WSNs, the broadcast nature of the wireless medium allows snooping on the channel. Losses can be known by tracking the link sequence number in the packets from each source. Various low-power listening mechanisms exist [29] that would enable snooping at a much lower cost. An alternative approach is to use received signal strength as an indication of link quality. Note that the controller is a proportional controller and the function of the control loop is to force the loss ratio of neighbors to converge to the set point, e.g., 0. The output of the controller is deterministic and binary. If the output of the controller is 1, the node will forward the packet to the candidate that provides largest *ve* regardless of the velocity. Otherwise, dropping is made to maintain the delay requirement.

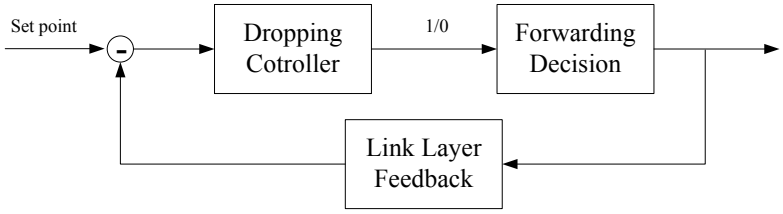


FIGURE 4.11: Initiative drop control.

4.5 Performance Evaluation

The proposed THVR is simulated in Prowler-Rmase [89,90]. Prowler is a probabilistic wireless network simulator, capable of simulating wireless distributed systems, from the application to the physical communication layers. It provides simple yet realistic radio/MAC models based on the Berkeley mote platform, and supports an event-driven structure similar to TinyOS/NesC. Implemented in Matlab, there are graphical interfaces for on-line debugging and toolboxes for off-line parameter optimization and tuning. Rmase has extended Prowler to more options of topologies, application models and routing designs.

To be close to practical WSNs and realistic implementation, we set the MAC layer, link quality model and energy consumption parameters according to Mica2 Motes [25] with MPR400 (915MHz) radio. Nodes are distributed in a 200m$\times$200m area following Poisson point process with node density $\rho = 0.005$ node/m^2. This node density is chosen by the method described in [91] to ensure a high level of network connectivity. To simulate multi-hop transmissions with a large enough number of hop counts, we limit the source nodes to the left-lower corner of the above region, while the sink is fixed at location (200m, 200m). The size of the neighbor table for each node is set to 400 bytes for all the tested protocols, which is found sufficient to store neighbor information within two hops. Note that for THVR, relatively each node needs more memory to maintain two-hop information. Practically, the average number of neighbors within two hops is around 20 with average node degree 6 under our simulation settings. Theoretically, without losing the generality, if we consider an ideal node deployment over a grid, the number of h-hop neighbors of a node is given by $(2h+1)^2-1$. So the memory space complexity is polynomial with $O(h^2)$, which is acceptable in practice for $h = 2$.

4.5.1 MAC Settings

Following the default CSMA scheme (similar to B-MAC [20]) in Mica2 Motes, to initiate a packet transmission, a sensor node will generate a random initial waiting time uniformly distributed in the range of [200, 328] bit-time (by Mica2 Motes, one bit-time equals to 1/40000 second) and start a timer. Upon timer expiration, the channel is sensed. If it is found idle, a packet is transmitted. Otherwise, it backoffs and then continues the sensing until the channel is found idle. The backoff time is uniformly distributed in [100, 130] bit-time. To improve delivery reliability, ARQ is employed here. If the total number of transmission count is greater than 7, the packet will be dropped. This is for avoiding excessive tries to a bad link or a too busy channel.

4.5.2 Link Model

We adopt the packet reception rate (PRR) model [27] for lossy WSN links. It is built on experimental measures of practical systems with respect to statistics of wireless channel. With the standard non-coherent FSK modulation and Manchester encoding, the PRR, $0 \leq p(s) \leq 1$, of a wireless link is expressible as :

$$p(s) = \left(1 - \frac{1}{2}\exp\left(-\frac{\gamma(s)}{2}\frac{1}{0.64}\right)\right)^{8(2f-l)} \tag{4.11}$$

where s is the transmitter-receiver distance, $\gamma(s)$ is the signal-to-noise ratio (SNR), and f is the frame size which equals to 50 bytes including preamble l (2 bytes), payload and CRC (2 bytes). Here, we adopt the setting of [27]. Note that the maximum packet size allowed is 241 bytes for Mica2. This model takes into account both distance-dependent path loss and log-normal shadowing in characterizing wireless links. For transmitting power P_t, the SNR, $\gamma(s)$, is expressible as :

$$\gamma(s)_{dB} = P_{t\,dB} - PL(s)_{dB} - P_{n\,dB} \tag{4.12}$$

where, according to MICA2 radios, $P_{t\,dB}$ is set at 0 dBm, the noise floor $P_{n\,dB}$ is at -115 dBm, and the path loss $PL(s)_{dB}$ is modeled as :

$$PL(s)_{dB} = PL(s_0)_{dB} + 10n\log_{10}(s/s_0) + X_{\sigma\,dB} \tag{4.13}$$

where n is the path loss exponent, s_0 is the reference distance (1 meter), and X_σ denotes the log-normal shadowing with zero mean and variance σ^2. In the coming simulations, we set $n = 3$ and $\sigma = 3$. For each transmission, a random number x is generated and then compared with $p(s)$ as in a lossy WSN link. When $x \leq p(s)$, the packet is assumed successfully transmitted. Otherwise, it is considered lost and a retransmission will be initiated.

4.5.3 Pre-study of Node Density and Network Connectivity

To support multi-hop delivery and offer an effective WSN, it is necessary to have a pre-study of the relationship between node density and network connectivity [37]. Consider that the transmission power of each node is fixed at 0 dBm, the link reliability follows (4.11), and nodes are

distributed in area A following a homogeneous Poisson point process with intensity λ, we want to know which node density value ρ ($\rho = \lambda/A$) is the most efficient operating point to offer a good quality of connectivity that is commonly indicated by the node non-isolation probability [4], giant component size [5] and average node degree [6] [38].

We perform simulations with the increase of node density ρ while keeping all other parameters the same when conducing the routing test. For each density, we repeat simulation runs independently 500 times and finally evaluate the average performance. As shown in Figure 4.12, it is found that at node density $\rho = 0.005$, the non-isolation probability is above 0.8 and the giant component size is greater than 0.99. Meanwhile, the average node degree is less than 7. Note that too many neighbors will increase the chance of collision and enlarge forwarding table. We choose this density as the operating point for its high guarantee of connectivity and appropriate number of neighbors. This point can be estimated either via simulation or theoretical derivations. The analytical result is plotted by blue dotted line, which closely matches the simulated one. Related derivation procedure can be referred to [38].

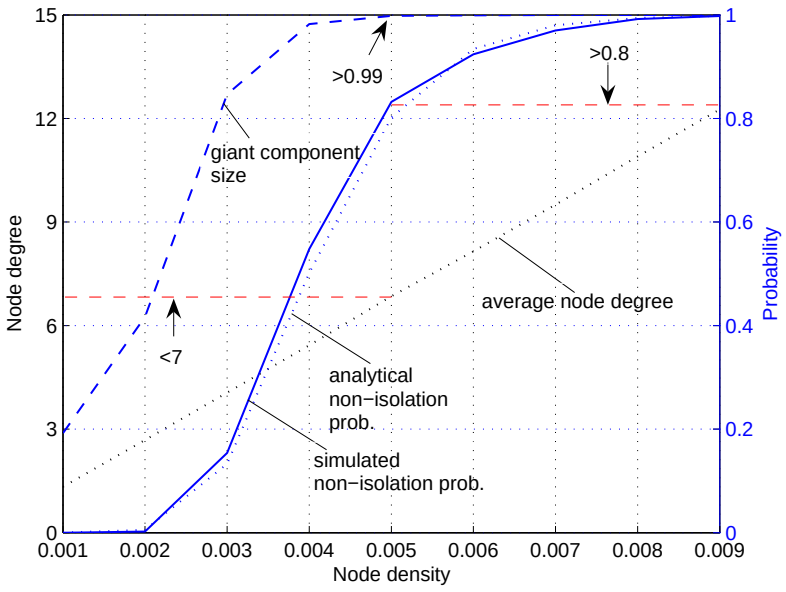


FIGURE 4.12: How to choose an efficient operating point under the transmission power of 0 dBm.

4.5.4 Energy Model

In WSNs, the energy consumed in a node is mainly due to packet transmission (E_{tx}), reception (E_{rx}) and channel sensing (E_{cs}) to check whether it is clear. The total energy consumed is thus

4. It is the probability that no node in the network is isolated.
5. It is defined by the size of the largest connected component in the network.
6. It is defined by the average number of neighbor nodes.

expressible as :

$$\begin{aligned} E &= E_{tx} + E_{rx} + E_{cs} \\ &= V \times (f \cdot I_{tx} T_{tx} + f \cdot I_{rx} T_{rx} + I_{cs} T_{cs}) \end{aligned} \tag{4.14}$$

where f is the packet size (i.e., 50 bytes), I_{tx}, I_{rx} and I_{cs} denote the current required during transmission, reception and channel sensing respectively, V is the voltage supply (by default, 3V), and T_{tx}, T_{rx} and T_{cs} refer to the corresponding activity durations, as listed in Table 4.2 [20].

TABLE 4.2: Mica2 Motes based Energy Model [20]

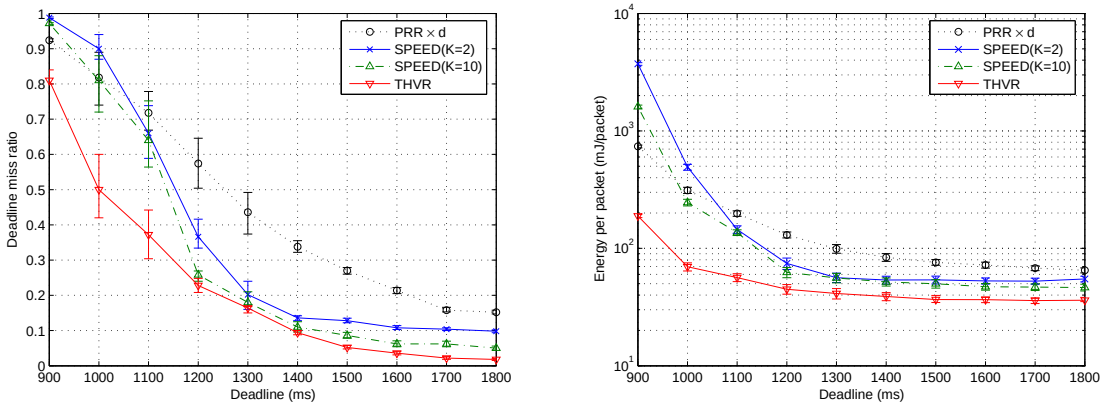
Operations	Duration (ms)	Current (mA)
Transmit 1 byte (0 dBm) :	0.416 (T_{tx})	20 (I_{tx})
Receive 1 byte :	0.416 (T_{rx})	15 (I_{rx})
Channel Sensing :	0.35 (T_{cs})	15 (I_{cs})

4.5.5 Simulation Results

In this section, a detailed performance investigation of THVR is conducted and compared with SPEED [80] and the well-known PRR-distance-product routing metric, $PRR \times d$, proposed in [92], in which d is the distance traversed towards destination, which was verified superior to simple greedy geographic routing.

In the first set of simulation, we consider there is one source node located at (20m, 20m) while the sink is at (200m, 200m). The source generates a CBR flow at 1 packet/s with packet frame size equal to 50 bytes (including preamble, payload and CRC). The value of C in (4.7) is set at 0.9 to emphasize end-to-end delay performance. In each run, 500 packets are transmitted. Figure 4.13 shows the result under different deadline requirements ranging from 900 to 1800 ms. As shown in Figure 4.13(a), it is clear that with an increase of deadline, the DMR of a protocol has decreased generally by the fact that more packets can finally be forwarded to the destination due to a longer allowable duration. As the deadline increases, their DMRs will converge to some corresponding levels. It can be observed that THVR has lower DMR than all the other in general. When deadline is stringent (e.g., less than 1200 ms), the advantage is especially significant. Generally speaking, SPEED with $K = 10$ has smaller DMR than that with $K = 2$ as expected. On the other hand, the metric $PRR \times d$ is known good at choosing a link for better reliability and routes in best effort. However, it lacks an explicit consideration of packet timeliness and delay performance. Clearly, compared with the other two protocols, the proposed THVR is able to enhance real-time delivery by an effective integration of 2-hop information. It has inherently a higher capability in path finding.

Figure 4.13(b) shows the energy consumed per packet successfully transmitted. The consumption has similar tendency and characteristic as that in DMR, Figure 4.13(a). By a high tolerance of packet delay (e.g., deadline larger than 1400 ms), DMR tends to be stable and the number of packets successfully transmitted from end to end is also quite stable. This supports the convergence of overall energy consumption. In comparing to the other protocols, Figure 4.13(b)



(a) DMR with deadline varied

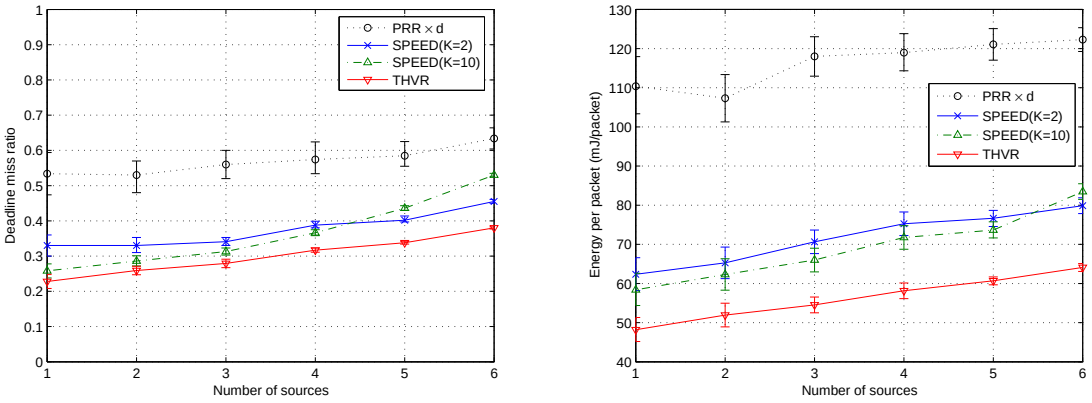
(b) Energy consumed per successfully transmitted packet

FIGURE 4.13: Performance of (i) PRR-distance-product routing, (ii) SPEED, with $K = 2$ and 10 in (4.5) respectively, and (iii) THVR, each with 90% confidence interval.

clearly shows that THVR is more energy efficient. One of the major reasons is that THVR has a better capability in forwarding packets to a small delay path. This has resulted in a smaller DMR, smaller retransmission rate, and also higher energy utilization efficiency. Besides, the initiative drop control has a positive effect in energy saving. As indicated by Figure 4.13, THVR outperforms SPEED and also $PRR \times d$ geographic routing in both DMR and energy efficiency performance under the workload of single CBR flow.

Furthermore, we investigate the performance of THVR under different workload. Figure 4.14(a) shows the DMR in which the number of sources increases from 1 to 6. Each source generates a CBR flow at 1 packet/s while the deadline requirement is fixed at 1200 ms. The source nodes are located in the left bottom area, as highlighted in Figure 4.15(a) and labeled with ID {190, 116, 1, 112, 93, 158} respectively. On the other hand, Figure 4.14(b) shows the energy consumption performance of the three protocols. It is clear that as the number of sources increases, both the DMR and energy consumption increase generally. The increase in DMR is resulted by the increased channel busy probability, packet collisions at MAC, and network congestion, due to the increased number of sources and consequent traffics. The comparison indicates that THVR has lower DMR and also lower energy consumption per successfully transmitted packet, as shown in Figure 4.14(a) and Figure 4.14(b) respectively. This reflects the general improvement by THVR. It is worth pointing out that as the number of sources increases from 4 to 6, SPEED with $K = 2$ will outperform SPEED with $K = 10$. This is due to the benefit of load balance with $K = 2$ in the case the workload is heavy and traffic congestion is more likely to happen.

In addition, we conduct the following investigations to study the energy balance performance of the proposed cost function (4.7). First, deadline is relaxed to a large value of 2000 ms so that a larger end-to-end delay is allowable. The value of C in THVR is set at 0.7 so as to have a larger weighting on the factor of residual energy in forwarder selection. Six source nodes are located in the left bottom area as usual. Simulation stops when a total of 400 packets are successfully delivered within deadline by each source. Figure 4.15 depicts the energy consumption distribu-



(a) DMR with number of sources varied

(b) Energy consumed per successfully transmitted packet

FIGURE 4.14: Performance of (i) $(PRR \times d)$-routing, (ii) SPEED, and (iii) THVR, each with 90% confidence interval indicated, while the number of source nodes increases from 1 to 6.

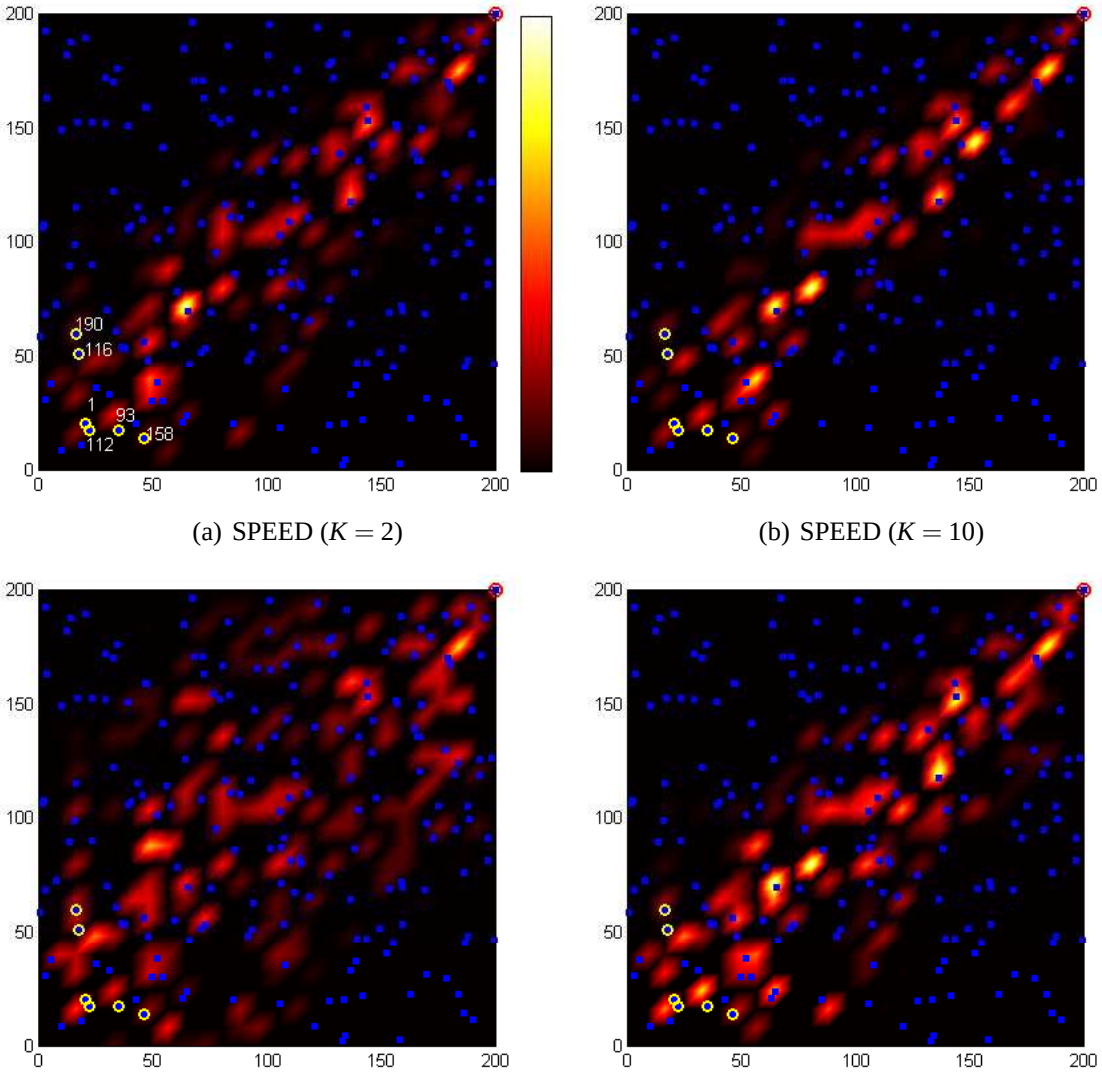
tion and magnitude of nodes in the WSNs. In a comparison of Figure 4.15(a) and 4.15(b), it is observable that SPEED with $K = 10$ has energy consumption footprints more centralized along the diagonal path while the setting of $K = 2$ is able to spread the footprints to a wider area. Comparing the four distributions in Figure 4.15, it is clear that THVR has the most even energy consumption that is shared among a large number of nodes. It can be expected that THVR will have a longer system lifetime due to the better balancing. On the other hand, their delay performance in the WSNs is shown in Table 4.3 under the simulation. THVR has the lowest DMR and energy consumption per successfully transmitted packet. However, it is worth noting that the value of C should be carefully chosen in the delay and load balance tradeoff. Otherwise, the end-to-end delay performance could be over-sacrificed and consequently much degraded.

TABLE 4.3: Deadline Miss Ratio and Energy Consumption under Figure 4.15.

Routing protocols	Deadline miss ratio	Energy per packet (mJ/packet)
$PRR \times d$	0.3144	67.74
SPEED ($K = 2$)	0.1724	55.50
SPEED ($K = 10$)	0.2077	57.59
THVR	0.0967	54.49

4.6 Discussions

Generally, an instant 2-hop delay updating will induce more overheads than that required for 1-hop information updating. This issue will impact our 2-hop based design as well. It can be observed from Figure 4.10 that a further feedback will be sent from a child node to its parent node as aforementioned. We measure the total amount of overheads (ACK packets) encountered



(a) SPEED ($K = 2$)

(b) SPEED ($K = 10$)

(c) THVR ($C = 0.7$)

(d) $PRR \times d$

FIGURE 4.15: Energy consumption distribution and magnitude in different protocols.

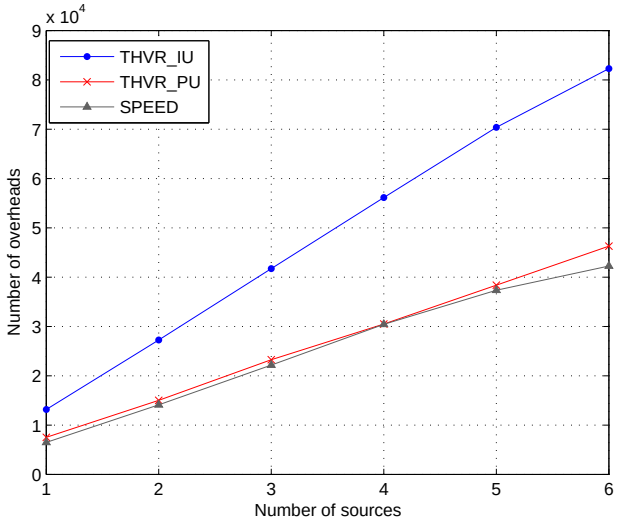


FIGURE 4.16: Comparison of the number of overheads encountered in SPEED and THVR with instant update (IU) and with piggybacked update (PU).

and plot it (labeled by "THVR_IU") in Figure 4.16, and compare to that required in SPEED. In our case, it is nearly two times of that by SPEED [7]. However, one can consider to reduce the overheads by piggybacking the updated information in conventional ACK packets but without further (extra) feedbacks. Consequently, these data will be piggybacked and sent together only when ACK is to be transmitted. This helps to maintain in a small number of feedback packets despite the fact that the resulting ACK size will be larger. By this approach, simulation shows that the number of overheads encountered in the WSNs is almost the same as that in SPEED. Theoretically, they are in the same amount. The slight difference is due to simulation randomness. The result is plotted in Figure 4.16 and labeled by "THVR_PU". Note that a drawback of this piggyback solution is that the 2-hop delay information may not be updated frequently enough. However, since the link delay estimation is based on the combination of the historical average and most recent one, there could be only minor difference to the estimation performance even when the update is not immediate and especially in WSNs with low mobility.

4.7 Summary

In this chapter, a 2-hop neighborhood information based geographic routing protocol is proposed to enhance the service quality of real-time packet delivery for WSNs. We adopt the approach of mapping packet deadline to the velocity as SPEED ; however, the routing decision is made based on the 2-hop velocity integrated with energy balancing mechanism. An energy-efficient packet drop control is incorporated to enhance energy utilization efficiency while keeping low packet deadline miss ratio. The actual characteristics of physical and MAC layers are

7. Note that in SPEED, two methods can be used for updating neighborhood information : (i) periodic beacon announcement, or (ii) ACK piggyback. Here, we adopt the latter, by which update is embedded in ACK message.

captured in the simulation studies. Simulation results show that, compared with SPEED and the $(PRR \times d)$-routing which both only utilize 1-hop information, THVR has achieved lower end-to-end deadline miss ratio and higher energy utilization efficiency.

5

Detection Fusion with Quality of Information Support

5.1 Introduction

In contrast to wireless ad hoc networks that are deployed for generic data service, WSNs are being deployed to support specific applications. Sensor-enabled applications typically seek the occurrence of events or patterns of interest, and take an action. However, the effectiveness and the subsequent impact of these actions will depend on the goodness of the information collected like the accuracy of a set of measurements, the false alarm in the reporting of an intruder, the timeliness of the reporting of a threshold crossing etc. In other words, how well the application accomplish their tasks will depend on the quality of information (QoI).

In [93], QoI has been given a formal definition as follows, similar to QoS definition in ITU :

Definition 5.1 *QoI is the collective effect of information characteristics (or attributes) that determines the degree by which the information is fit-to-use for a purpose.*

The concept of "fit-to-use" here is to imply any metadata about the raw information that can be used by the application to decide whether it is within the desired tolerance bound to the application or compatible with the application's requirement. QoI captures the sensor-derived information to describe the important feature of event of interest (e.g., the detection of a fire or intruder) sufficient well to allow related mission to perform tasks at a desired level of effectiveness.

In this chapter, we study QoI for a particular class of sensor-enabled applications that relate to event detection. Each sensor node deployed in AoI preprocesses and extracts information from the raw observations. They also have the ability to communication with each other or a central base station (fusion center) via wireless channels. Fusion center jointly processes data from local sensors and form a global and more precise situational assessment. This avoids large volume of raw observations directly sending to the fusion center, resulting in waste of power and channel resource. In such a system, QoI can be described as quantifiable attributes like

the system detection probability, false alarm rate and the performance robustness in different noisy environment. If the raw data observed at local sensors are accessible entirely in the fusion center, the problem is reduced to the classical hypothesis testing problem at the fusion center with multiple data samples supplied by multiple sensors.

Distributed detection has been intensively studied in the past few decades. Optimal local quantizer and fusion rule design can be found in vast literature under either the Bayesian criterion or the Neyman-Pearson criterion. However, there is an implicit assumption in the majority of existing results : the transmissions between local sensors and the fusion center are error-free. Although by proper encoding and decoding, any noisy channel can be made reliable, provided the information rate is less than that prescribed by the channel capacity. This is difficult to guarantee in WSNs with limited resources (energy, computing ability etc.). As with the distributed detection problem, two different problems need to be addressed : the design of a fusion rule at the fusion center and the design of a distributed signal processing algorithm at the local sensors. In this chapter, we will mainly focus on the former, considering local decision has been made. Channel fading effects are taken into account in the fusion rule designs. It is worth noting that while designing fusion rules for WSNs, there should be tradeoff considerations between the algorithm effectiveness and computational complexity.

The rest of the chapter is organized as follows. Section 5.2 reviews related work. Section 5.3 introduces the system models and problem definition. In Section 5.4, four fusion rules are designed. We study the statistics of the fusion rules in Section 5.5. Performance evaluation, including numerical examples, is contained in Section 5.6. We summarize the chapter in Section 5.7.

5.2 Related Work

The pioneering work in distributed detection is accomplished by Tenney and Sandell [94]. They have studied the optimal sensor decision rules in a detection system composed of a fusion center and two sensors. It has been proved that optimal is likelihood ratio decision, and the optimal decision threshold of local sensors are coupled. Chair and Varsheny [95] brought the design of fusion rules into the framework of classical hypothesis testing. In [95], local decisions are viewed as the observations of the fusion center while the fusion rules are considered as general hypothesis testing decision rules. The optimal fusion rules were derived for Bayes fusion systems. It has been argued in [96] that when the sensor observations are correlated, the optimal fusion rule cannot be simplified to likelihood ratio decision. Therefore for simplicity, most literatures assume the observation from sensors are independent. This assumption holds when sensors are apart from each other.

More recently, decision fusion under a communication constraint has been considered. The constraint, however, is often in the form of the total number of bits allowed [97, 98]. The actual transmission is still idealized, i.e., the information sent from local sensors is assumed intact at the fusion center. While this assumption may be reasonable for some applications, it may not be realistic for many WSNs where the transmitted information has to endure both channel fading

and noise/interference. In [99], a non-asymptotic approach is adopted to study the tradeoffs between several important parameters like number of sensors, degree of quantization at each local sensor, and SNR. Although it considers the communication constraints in WSNs, it fails to capture the fading characteristic of wireless channel. Decision fusion with non-ideal channels has been investigated in [100]. The local decisions are transmitted over noisy channels so that they may not be correctly received at the fusion center. Yet the channel model is simplified as a binary channel thus does not allow a full integration of signal transmission into information processing.

In this chapter, a series of fusion rules are proposed under the assumption of Rayleigh channel model. The performance, resource cost and computational complexity are extensively compared with different SNR ranges. It is worth noting that in resource-constraint WSNs, tradeoffs should be made among the above metrics in choosing the optimal rules.

5.3 System Model and Problem Formation

The three-layer model for distributed detection system in the presence of fading channels is illustrated in Figure. 5.1. There are two hypotheses, H_1 and H_0, under test. Each sensor obtains its own observations, processes them and makes a local decision. These local decisions are transmitted via fading and noisy channels to a fusion center. In a traditional distributed detection system model, all local decisions are assumed to be recovered perfectly at the fusion center. For WSNs operating in a fading environment, channel fading and noise impairment may render the received decisions at the fusion center unreliable, especially in resource constrained applications. Toward this end, a channel layer must be incorporated into our model to allow for the development of channel aware decision fusion rules that have proved to be more energy efficient. The model shown in Figure 5.1 is described below.

Local sensor layer : All local sensors collect observations generated under a specific hypothesis. In this paper, we assume that the observations are independent of each other across sensors conditioned on any hypothesis. After receiving its observation, each sensor makes a hard (binary) decision : $u_k = 1$ is sent if H_1 is decided, and $u_k = -1$ is sent otherwise, where $k = 1, ..., K$. The detection performance of each local sensor node can be characterized by its corresponding probability of false alarm and detection, denoted by P_{fk} and P_{dk}, respectively, for the kth sensor : $P_{dk} = P(u_k = 1|H_1)$, $P_{fk} = P(u_k = -1|H_0)$.

$$H_0 : x_k = N_k,$$
$$H_1 : x_k = S + N_k$$

In general, these pairs need not be identical and they are functions of SNR's as well as thresholds at local sensors.

Fading channel layer : Decisions at local sensors, denoted by u_k for $k = 1, 2, ..., K$ are transmitted over parallel channels that are assumed to undergo independent fading. In this paper, we assume flat fading channels between local sensors and the fusion center. This assumption is reasonable because most WSNs operate at short range due to power and energy limitations.

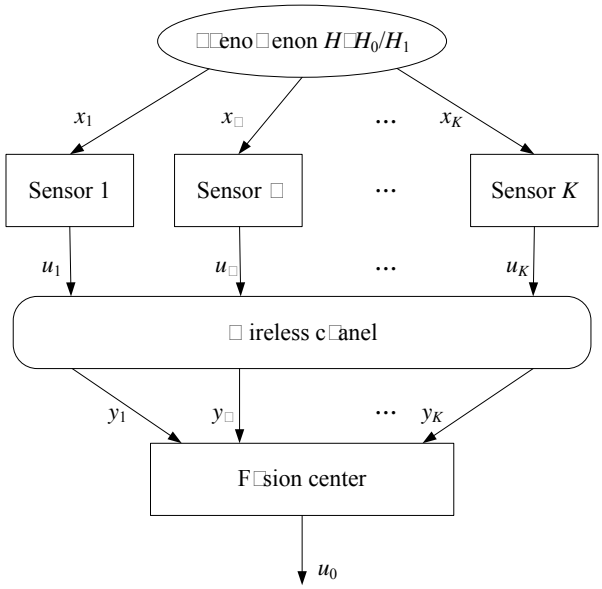


FIGURE 5.1: Paralleled distributed detection system under fading channel

We further assume phase coherent reception, thus the effect of a fading channel is further simplified as a real scalar multiplication given that the transmitted signal is assumed to be binary . This phase coherent reception can be either accomplished through limited training for stationary channels or at a small cost of SNR degradation, by employing differential encoding for fast fading channels which results in the same signal model. The statistics of the real scalar, denoted by r_k, is determined by the fading type. For example, for homogeneous scattering background, Rayleigh distribution best describes the envelope of a fading signal. In the development of fusion rules, the gain of the fading channel is considered as a constant during the transmission of a single local decision. We assume that the channel noise is additive white Gaussian and uncorrelated from channel to channel. For simplicity, we assume that the noise variances are identical for different channels. To summarize, each local decision is transmitted through a fading channel and the output of the channel (or input to the fusion center) for the kth sensor is

$$y_k = r_k u_k + n_k \tag{5.1}$$

where r_k is the fading channel gain which follows Rayleigh distribution and n_k is a zero-mean Gaussian random variable with variance σ^2. Suppose the signal power is normalized to 1, namely, $E(r_k^2) = 1$, then the PDF of r_k is :

$$p(r_k) = \frac{r_k}{b^2} e^{-\frac{r_k^2}{2b^2}} = 2r_k e^{-r_k^2} \tag{5.2}$$

It is obvious that the $SNR(dB) = 10\log(1/\sigma^2)$. If other fading models are adopted, i.e., Ricean fading model, the major results of this study will not be affected as well.

Fusion center : Based on the received data y_k for all k , the fusion center decides which hypothesis is more likely to be true. This is done by constructing and evaluating a fusion statistic

using the observations as well as some system parameters, if available. A global decision u_0 is finally obtained with the system detection probability P_F and system false alarm rate P_F, and $P_D = P(u_0 = 1|H_1)$, $P_F = P(u_0 = -1|H_1)$.

The local decision information may be corrupted through the wireless channel, which may degrade the detection system performance. Suppose the local decision rule is already known, the major issue is how to design the decision rules in the fusion center to maintain the QoI and counteract the effect of channel fading.

5.4 Fusion Rules Design

In this section, we first analyze the likelihood ratio based fusion rule with fading channels in the classical Bayes fusion framework. Then this optimal fusion rule is approximated under different SNR ranges and three sub-optimal rules are obtained with less *a prior* knowledge, resource cost and less complexity.

5.4.1 Likelihood Ratio based Fusion Rule

Suppose local sensor decisions have been made, the optimal fusion rule which minimize the Bayesian risk is given by the likelihood ratio test. Therefore, we have Theorem 5.1.

Theorem 5.1 *By assuming complete knowledge regarding the Rayleigh fading channel and the local sensor performance indices, the P_{fk} and P_{dk} values. Assuming conditional independence of observations at the sensors and that each local sensor makes a binary decision, the final LR rule is :*

$$\Lambda_{LR}(\boldsymbol{y}) = \log \frac{f(\boldsymbol{y}|H_1)}{f(\boldsymbol{y}|H_0)} = \sum_{k=1}^{K} \log \frac{P_{dk} e^{-\frac{(y_k - r_k)^2}{2\sigma^2}} + (1 - P_{dk}) e^{-\frac{(y_k + r_k)^2}{2\sigma^2}}}{P_{fk} e^{-\frac{(y_k - r_k)^2}{2\sigma^2}} + (1 - P_{fk}) e^{-\frac{(y_k + r_k)^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{5.3}$$

where $\boldsymbol{y} = [y_1, ..., y_K]^T$ is a vector containing data received from K sensors, and τ is the test threshold.

Proof. Since the each sensor's local decision is independent, we have :

$$\log \frac{f(\mathbf{y}|H_1)}{f(\mathbf{y}|H_0)} = \log \prod_{k=1}^{K} \frac{f(y_k|H_1)}{f(y_k|H_0)} = \sum_{k=1}^{K} \log \frac{f(y_k|H_1)}{f(y_k|H_0)} \tag{5.4}$$

where $f(y_k|H_1)$ can be obtained by using total probability formula :

$$\begin{aligned} f(y_k|H_1) &= dp(y_k|H_1)/dy_k \\ &= d(p(y_k|u_k = 1)p(u_k = 1|H_1) + p(y_k|u_k = -1)p(u_k = -1|H_1))/dy_k \\ &= P_{dk} f(y_k|u_k = 1) + (1 - P_{dk}) f(y_k|u_k = -1) \\ &= P_{dk} e^{-\frac{(y_k - r_k)^2}{2\sigma^2}} + (1 - P_{dk}) e^{-\frac{(y_k + r_k)^2}{2\sigma^2}} \end{aligned} \tag{5.5}$$

Similarly :

$$f(y_k|H_0) = P_{fk}e^{-\frac{(y_k-r_k)^2}{2\sigma^2}} + (1-P_{fk})e^{-\frac{(y_k+r_k)^2}{2\sigma^2}} \tag{5.6}$$

Plugging equations (5.5) and (5.6) into equation (5.4), we obtain (5.3). ■

An implicit assumption is that all the channel outputs are co-phased. This assumption allows us to deal exclusively with real observations. While the form of the LR based fusion rule Λ_{LR} is straightforward to implement, it does need both the local sensor performance indices P_{dk} and P_{fk} and complete channel knowledge r_k and n_k. In practical, it is difficult to estimate the channel state information in WSNs. Sub-optimal fusion rules that relieve the above requirements are more desirable.

5.4.2 Two-Stage Fusion Rule

The fusion rule specified in (5.3) jointly considers the effects of the fading channel and the local sensor output to achieve optimal performance. A direct alternative is to separate this into a two stage process : first y_k is used to infer about u_k, and then, the optimal fusion rule based on u_k (assuming that the estimates are reliable). We obtain

Proposition 5.1 *As channel SNR* $\to \infty$*, the optimal* Λ_{LR} *reduces to* Λ_{TS}*, i.e.,*

$$\Lambda_{TS}(\boldsymbol{y}) = \sum_{k\in\mathscr{S}^+} \log\frac{P_{dk}}{P_{fk}} + \sum_{k\in\mathscr{S}^-} \log\frac{1-P_{dk}}{1-P_{fk}} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{5.7}$$

where $\mathscr{S}^+ = \{k : y_k > 0\}$, $\mathscr{S}^- = \{k : y_k < 0\}$.

Proof. As $SNR \to \infty$, $\sigma^2 \to 0$, $n_k \to 0$, we have :

$$\Lambda_{LR}(\mathbf{y}) \approx \sum_{k\in\mathscr{S}^+} \log\frac{P_{dk} + (1-P_{dk})e^{\frac{2y_kr_k}{\sigma^2}}}{P_{fk} + (1-P_{fk})e^{\frac{2y_kr_k}{\sigma^2}}} + \sum_{k\in\mathscr{S}^-} \log\frac{P_{dk}e^{-\frac{2y_kr_k}{\sigma^2}} + (1-P_{dk})}{P_{fk}e^{-\frac{2y_kr_k}{\sigma^2}} + (1-P_{fk})} \tag{5.8}$$

for $k \in \mathscr{S}^+$, $e^{\frac{2y_kr_k}{\sigma^2}} \gg 1$; for $k \in \mathscr{S}-$, $e^{-\frac{2y_kr_k}{\sigma^2}} \gg 1$. Since the numerator and the denominator of both fractions are derivable, we apply L'Hospital rule here and have :

$$\lim_{\sigma^2\to\infty} \Lambda_{LR}(\mathbf{y}) = \sum_{k\in\mathscr{S}^+} \log\frac{P_{dk}}{P_{fk}} + \sum_{k\in\mathscr{S}^-} \log\frac{1-P_{dk}}{1-P_{fk}} = \Lambda_{TS} \tag{5.9}$$

■

Notice that Λ_{TS} does not require any knowledge regarding the channel gain but does require P_{dk} and P_{fk} for all k. Further, this two-stage approach falls into the conventional thinking that separates the communication and signal processing aspects. We show later through numerical examples that this two-stage approach suffers significant performance loss at low to moderate channel SNR.

5.4.3 Weighed Average Fusion Rule

Proposition 5.2 *As channel SNR* $\to 0$*, the optimal* Λ_{LR} *reduces to* $\hat{\Lambda}_{WED}$ *:*

$$\hat{\Lambda}_{WED}(\boldsymbol{y}) = \sum_{k=1}^{K} 2(P_{dk} - P_{fk}) \frac{r_k y_k}{\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{5.10}$$

Proof. As $SNR \to \infty$, $\sigma^2 \to \infty$, we have $-\frac{(y_k - r_k)^2}{2\sigma^2} \to 0$, $-\frac{(y_k + r_k)^2}{2\sigma^2} \to 0$. Since e^x approximates its first-order Taylor series when $x \to 0$, namely, $e^x \approx 1 - x$, we have :

$$e^{-\frac{(y_k - r_k)^2}{2\sigma^2}} \approx 1 - \frac{(y_k - r_k)^2}{2\sigma^2} \tag{5.11a}$$

$$e^{-\frac{(y_k + r_k)^2}{2\sigma^2}} \approx 1 - \frac{(y_k + r_k)^2}{2\sigma^2} \tag{5.11b}$$

plugging equations (5.11a) and (5.11b) into equation (5.3), we have :

$$\begin{aligned} \lim_{\sigma^2 \to 0} \Lambda_{LR}(\mathbf{y}) &\approx \sum_{k=1}^{K} \log \frac{P_{dk}(1 - \frac{(y_k - r_k)^2}{2\sigma^2}) + (1 - P_{dk})(1 - \frac{(y_k + r_k)^2}{2\sigma^2})}{P_{fk}(1 - \frac{(y_k - r_k)^2}{2\sigma^2}) + (1 - P_{fk})(1 - \frac{(y_k + r_k)^2}{2\sigma^2})} \\ &= \sum_{k=1}^{K} \left(\log(1 + P_{dk} \frac{2 r_k y_k}{\sigma^2}) - \log(1 + P_{fk} \frac{2 r_k y_k}{\sigma^2}) \right) \end{aligned} \tag{5.12}$$

As $\sigma^2 \to \infty$, $\frac{(y_k - r_k)^2}{2\sigma^2} \to 0$. Since $\log(1+x)$ approximates its first-order Taylor series when $x \to 0$, namely, $\log(1+x) \approx x$, we have :

$$\begin{aligned} \lim_{\sigma^2 \to 0} \Lambda_{LR}(\mathbf{y}) &\approx \sum_{k=1}^{K} \left(2 P_{dk} \frac{r_k y_k}{\sigma^2} - 2 P_{fk} \frac{r_k y_k}{\sigma^2} \right) \\ &= \sum_{k=1}^{K} 2(P_{dk} - P_{fk}) \frac{r_k y_k}{\sigma^2} = \hat{\Lambda}_{WED} \end{aligned} \tag{5.13}$$

■

Further, if the local sensors are identical, i.e., P_{dk} and P_{fk} are the same for all k's, then $\hat{\Lambda}_{WED}$ further reduces to a form analogous to a maximum ratio combiner :

$$\Lambda_{WED}(\mathbf{y}) = \frac{1}{K} \sum_{k=1}^{K} r_k y_k \tag{5.14}$$

The factor $\frac{1}{K}$ in Λ_{WED} does not affect the detection performance but is introduced for the convenience of performance analysis. Notice that the form of Λ_{WED} in equation (5.10) does not require the knowledge of P_{dk} and P_{fk} provided $P_{dk} - P_{fk} > 0$, i.e., the local detectors are unbiased. Knowledge of the channel gain is, however, required.

5.4.4 Average Fusion Rule

While Λ_{TS} and Λ_{WED} relieve some of the requirements compared with the optimal likelihood ratio based fusion rule, they still need some information either about the local sensors or the channel statistics. Further, we note that these fusion statistics, Λ_{TS} and Λ_{WED}, as approximations to the optimal LR based fusion rule at high and low SNR cases, may suffer performance loss for SNR outside those ranges. It would be very important to investigate other robust alternatives that operate well for the non-extreme SNR range while requiring the same or even less amount of information regarding the channel and/or the sensors.

Proposition 5.3 *When the channel statistics are the same for each sensor, Λ_{WED} can be further reduced to Λ_{ED} :*

$$\Lambda_{ED}(\boldsymbol{y}) = \frac{1}{K}\sum_{k=1}^{K} y_k \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{5.15}$$

Proof. It can be easily obtained by equation 5.14. ■

Since y_k's are assumed to be phase coherent outputs of each channel, we still require the phase of the fading channel but no other information regarding the channel/ sensor is needed. While this heuristic and the simple fusion rule does relieve most of the requirements compared with the optimal LR based fusion rule, its usefulness largely depends on its performance compared with the optimal fusion rule as well as the first two alternatives.

5.4.5 Physical Meaning Analysis

Figure 5.2 shows the physical realization of the above mentioned fusion rules. We can see from Figure 5.2 that each fusion rule can be regarded as a weight sum of local information. The difference lies in the weight factors. The respective weight factors are listed in Table 5.1.

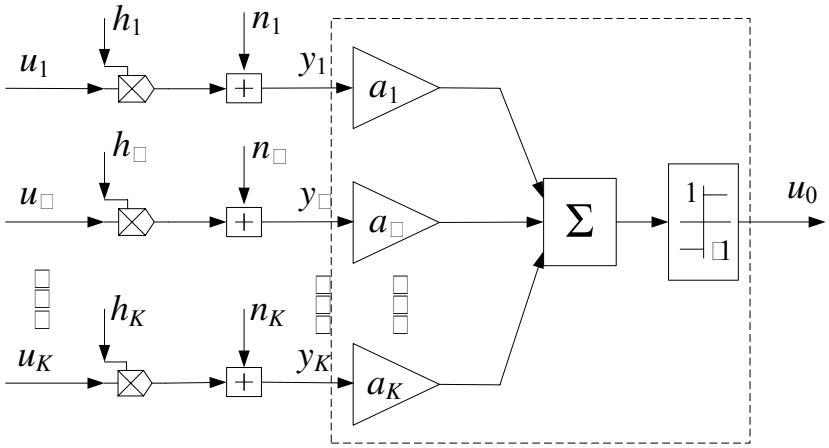


FIGURE 5.2: The physical realization of different fusion rules.

TABLE 5.1: The amplifier input of Λ_{LR}, Λ_{TS},Λ_{WED},Λ_{ED} in Figure 5.2.

Λ_{LR}	$a_k = \log \frac{P_{dk} e^{-\frac{(y_k - r_k)^2}{2\sigma^2}} + (1-P_{dk}) e^{-\frac{(y_k + r_k)^2}{2\sigma^2}}}{P_{fk} e^{-\frac{(y_k - r_k)^2}{2\sigma^2}} + (1-P_{fk}) e^{-\frac{(y_k + r_k)^2}{2\sigma^2}}}$
Λ_{TS}	$a_k = \begin{cases} \log \frac{P_{dk}}{P_{fk}}, & k \in \mathscr{S}_0, \mathscr{S}_0 = \{k : y_k < 0\}; \\ \log \frac{1-P_{dk}}{1-P_{fk}}, & k \in \mathscr{S}_1, \mathscr{S}_1 = \{k : y_k > 0\} \end{cases}$
Λ_{WED}	$a_k = r_k y_k$
Λ_{ED}	$a_k = y_k$

5.5 The Statistics Analysis of Λ_{TS}, Λ_{WED} and Λ_{ED}

The statistic analysis of the fusion rules provides a theoretical way of studying their performance. In this section, the statistics of Λ_{TS}, Λ_{WED} and Λ_{ED} are investigated.

5.5.1 The Statistics of Λ_{TS}

Recall equation (5.9), if all the sensors are identical, i.e., for all k, $P_{dk} = P_d$, $P_{fk} = P_f$. Define $K_1 = |\mathscr{S}^+|$, $K_0 = |\mathscr{S}^-|$, and $K = K_1 + K_0$. Then (5.9) becomes :

$$\begin{aligned} \Lambda_{TS} &= K_1 \log \frac{P_d}{P_f} + (K - K_1) \log \frac{1-P_d}{1-P_f} \\ &= K_1 \log \frac{P_d(1-P_f)}{P_f(1-P_d)} + K \log \frac{1-P_d}{1-P_f} \end{aligned} \tag{5.16}$$

which is an affine function of K_1 when $P_d > P_f$. They have equivalent statistics. If all the sensors are identical, all y_k are independently and identically distributed (i.i.d). Therefore, K_1 is Binomial distribution, namely, $K_1 \sim \mathscr{B}(K, p)$, where p is defined as $p = P(y_k \geq 0)$. We denote p_0 and p_1 as the success probabilities under H_0 and H_1, respectively. The closed-form solutions are provided in Lemma5.1.

Lemma 5.1 *The probabilities of a nonnegative observation y_k under hypotheses H_0 and H_1 are :*

$$\begin{aligned} p_0 &= p(y_k \geq 0 | H_0) = \frac{1}{2} + \frac{P_f - 1/2}{\sqrt{1 + 2\sigma^2}} \\ p_1 &= p(y_k \geq 0 | H_1) = \frac{1}{2} + \frac{P_d - 1/2}{\sqrt{1 + 2\sigma^2}} \end{aligned}$$

Proof. Applying the total probability formula, we have :

$$\begin{aligned} p_1 &= p(y_k \geq 0|H_1) \\ &= p(u_k = 1|H_1)p(y_k \geq 0|u_k = 1) + p(u_k = -1|H_1)p(y_k \geq 0|u_k = -1) \\ &= P_d p(y_k \geq 0|u_k = 1) + (1 - P_d)p(y_k \geq 0|u_k = -1) \end{aligned} \tag{5.17}$$

$$\begin{aligned} \text{where} p(y_k \geq 0|u_k = 1) &= p(y_k = r_k + n_k \geq 0) \\ &= \int_0^\infty \int_0^\infty f(x) f(y-x) dx dy \\ &= \int_0^\infty \int_0^\infty 2xe^{-x^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}} dx dy \\ &= \frac{1}{2}\left(1 + \frac{1}{\sqrt{1+2\sigma^2}}\right) \end{aligned} \tag{5.18}$$

$$\text{Similarly}, p(y_k \geq 0|u_k = -1) = \frac{1}{2}\left(1 - \frac{1}{\sqrt{1+2\sigma^2}}\right) \tag{5.19}$$

Substituting (5.18) and (5.19) into (5.17), we have :

$$\begin{aligned} p_1 &= p(y_k \geq 0|H_1) \\ &= p(u_k = 1|H_1)p(y_k \geq 0|u_k = 1) + p(u_k = 0|H_1)p(y_k \geq 0|u_k = -1) \\ &= P_d p(y_k \geq 0|u_k = 1) + (1 - P_d)p(y_k \geq 0|u_k = -1) \\ &= \frac{1}{2} + \frac{2P_d - 1}{2\sqrt{1+2\sigma^2}} \end{aligned} \tag{5.20}$$

$$\text{Similarly}, p_0 = p(y_k \geq 0|H_0) = \frac{1}{2} + \frac{2P_f - 1}{2\sqrt{1+2\sigma^2}} \tag{5.21}$$

■

Since Λ_{TS} and K_1 have identical statistics, Λ_{TS} is binomial (K, p_0) and binomial (K, p_1) distribution under H_0 and H_1 respectively.

5.5.2 The distribution of Λ_{WED} and Λ_{ED}

Theorem 5.2 *(Central Limit Theorem) Let $X_1, X_2, X_3, ..., X_n$ be a sequence of n i.i.d. random variables each having finite values of expectation μ and variance σ^2. As the sample size n increases, the distribution of the sample average of these random variables $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ approaches the normal distribution with a mean μ and variance σ^2/n, namely, $\bar{X} \sim N(\mu, \sigma^2/n)$.*

Proof. According to CLT, Λ_{WED} and Λ_{ED} approaches normal distribution when K is large, their respective mean and variance under H_0 and H_1 can be obtained by the expressions of Λ_{WED} and Λ_{ED}. ■

5.5.3 Summary of the first-order and second-order statistics of Λ_{TS}, Λ_{WED} and Λ_{ED}

The first-order and second-order statistics of Λ_{TS}, Λ_{WED} and Λ_{ED} are summarized in Table 5.2. With these statistics, system detection probability and false alarm rate can be easily calculated.

TABLE 5.2: The mean and variance of Λ_{TS}, Λ_{WED}, Λ_{ED} under H_0 and H_1 with K sensors.

	Λ_{TS}	Λ_{WED}	Λ_{ED}
$E_0\ (H_0)$	$K\left(\frac{1}{2}+\frac{P_f-1/2}{\sqrt{1+2\sigma^2}}\right)$	$2P_f-1$	$\frac{\sqrt{\pi}}{2}(2P_f-1)$
$E_1\ (H_1)$	$K\left(\frac{1}{2}+\frac{P_d-1/2}{\sqrt{1+2\sigma^2}}\right)$	$2P_d-1$	$\frac{\sqrt{\pi}}{2}(2P_d-1)$
$V_0\ (H_0)$	$K\left[\frac{1}{4}-\frac{(P_f-1/2)^2}{1+2\sigma^2}\right]$	$\frac{1}{K}[1+\sigma^2+4P_f(1-P_f)]$	$\frac{1}{K}[1-\frac{\pi}{4}+\sigma^2+\pi P_f(1-P_f)]$
$V_1\ (H_1)$	$K\left[\frac{1}{4}-\frac{(P_d-1/2)^2}{1+2\sigma^2}\right]$	$\frac{1}{K}[1+\sigma^2+4P_d(1-P_d)]$	$\frac{1}{K}[1-\frac{\pi}{4}+\sigma^2+\pi P_d(1-P_d)]$

5.6 Simulations and Performance Analysis

While it is clear that the LR-based fusion rule provides the best detection performance, it is interesting to see how much performance degradation the other three simple alternatives suffer, and among these three, which one provides the best and most robust detection performance. While analytical results are most desirable, the problem is, in general, intractable. The Λ_{WED} and Λ_{ED} fusion rules, however, are amenable to asymptotic analysis because of their simple expression in the form of a sum of some random variables that are independent of each other. In the case of identical sensors and fading statistics, these independent random variables are also identical to each other, which leads to the direct application of the central limit theorem for asymptotic analysis. We emphasize that the Λ_{WED} and Λ_{ED} are perhaps more desirable because of their performance advantage compared with Λ_{TS} for low to medium SNR values. Most WSNs operating using on-board battery supply are energy limited. Given that RF communication is the most energy consuming function of a sensor node, it is, therefore, imperative to use as little power as possible for data transmission, which usually results in modest SNR values at the fusion center receiver. Throughout this section, we will assume a Rayleigh fading channel for both analysis and numerical simulation. Other fading types, such as Ricean fading, can be used instead though the analysis is more involved.

5.6.1 ROC curve analysis

In the following, we denote the performance at the network level as system level detection probability and false alarm rate, denoted by P_D and P_F, to distinguish them from sensor level P_{dk}

and P_{fk}. We assume in this section that the sensors are identical to each other, thus $P_{fk} = P_f$ and $P_{dk} = P_d$ for all k. Therefore, both Λ_{WED} and Λ_{ED} fusion statistics are sums of i.i.d. random variables which allows direct application of the CLT. This converts the decision fusion problem into hypothesizing between two Gaussian distributions which can lead to a lot of insight. In order to use the CLT, we need the first and second order statistics which are derived and summarized in Table 5.2.

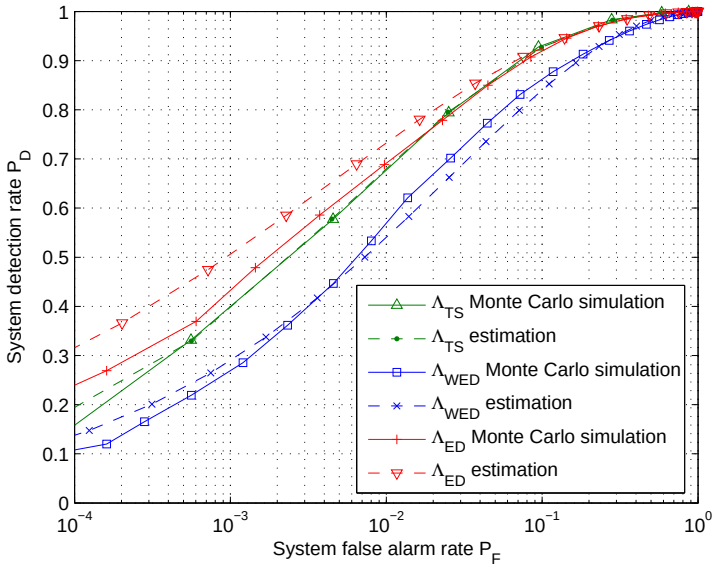


FIGURE 5.3: System ROC curve : Monte Carlo simulation vs. numerical approximation.

Given the above statistics, the probabilities of detection and false alarm can be easily obtained using the $Q(\cdot)$ function, defined as the complimentary distribution function of standard Gaussian. Figure 5.3 presents the receiver operating characteristic (ROC) curves obtained both by Monte Carlo simulation and numerical approximation using CLT. In this example, the total number of sensors is 10 with sensor level $P_f = 0.1$ and $P_d = 0.6$ and channel SNR equal to 5dB. While some discrepancy exists due to the relax in the CLT condition (the value of K is not large enough), the approximations using CLT match relatively well to the corresponding simulation results.

Figure 5.4 shows the ROC curves of all the fusion rules by Monte Carlo simulation. The settings are the same with those in Figure 5.3. Obviously, the optimal Λ_{LR} provides the uniformly most powerful detection performance. However it requires instantaneous CSI and sensor indices. On the other hand, under this settings, Λ_{ED} performs just a little inferior to Λ_{LR} but requires least information *a priori* and computational complexity. We further discuss their performance under different SNR values in Section 5.6.3.

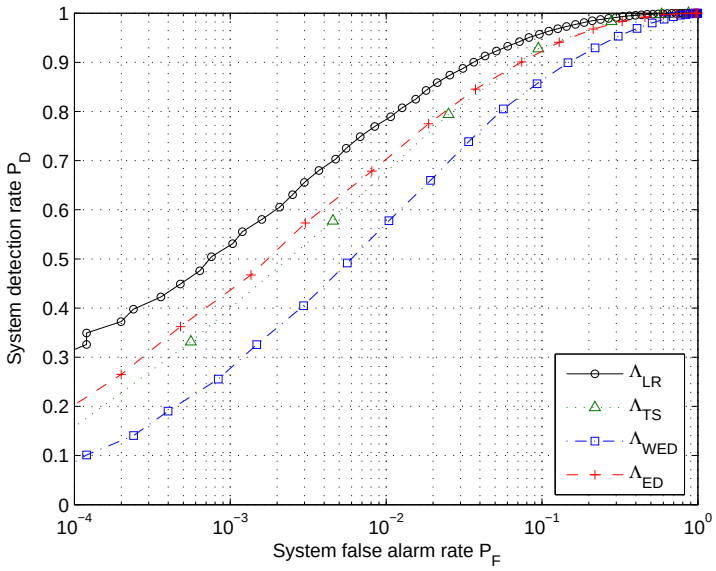


FIGURE 5.4: System ROC curves of the four fusion rules by Monte Carlo simulation.

5.6.2 Deflection Coefficient

Deflection coefficient (DC) is proposed in [101] to be a measure of the detection performance. DC can be regarded as the system output SNR. The larger the DC is, the better is the detection performance. Assume x is the observed values and $S(x)$ is the output statistics. The definition of DC $D(S)$ is as below :

$$D(S) = \frac{[E_1(S) - E_0(S)]^2}{V_0(S)} \tag{5.22}$$

where E_1 and E_0 are the expectations under H_1 and H_0, while V_0 is the variance under H_0. There are various of advantages of DC as a measure of detection performance :

1. By using the deflection criterion the system $S(x)$ is characterized by only one number instead of by a curve. It is in general much easier to calculate.
2. There is a relation between the LR receiver and the optimal receiver in terms of deflection. First the system $S(x)$ maximizing 5.22 is none other than the LR $L(x)$, i.e. ratio of the probability densities functions of x under H_1 and H_0, respectively. Furthermore, if there are some constraints on $S(x)$, the receiver maximizing the deflection under these constraints is the receiver the output of which is the mean square estimation of $L(x)$.
3. DC concerns the asymptotic case. If the components of x are i.i.d. random variables, it is possible to use CLT, to assume that the output is Gaussian. Then it becomes again possible to calculate the ROC curve and it appears that maximizing the deflection becomes equivalent to optimizing the performances calculated in terms of ROC.

Inevitably, the deflection criterion has some limitations with respect to its use in detection problems. First, although DC can be treated as system output SNR, its physical meaning is not quite clear. If there are two systems S_1 and S_2, their deflections satisfy $D(S_1) < D(S_2)$. It is then impossible to prove generally that for a given false alarm probability, the detection probability of

S_2 is greater than that of S_1. This can be true for some values of a and untrue for other ones. In other words a greater deflection does not necessarily give a better performance in terms of ROC curves. This does not mean that the deflection criterion must be completely rejected. It must be used with some care.

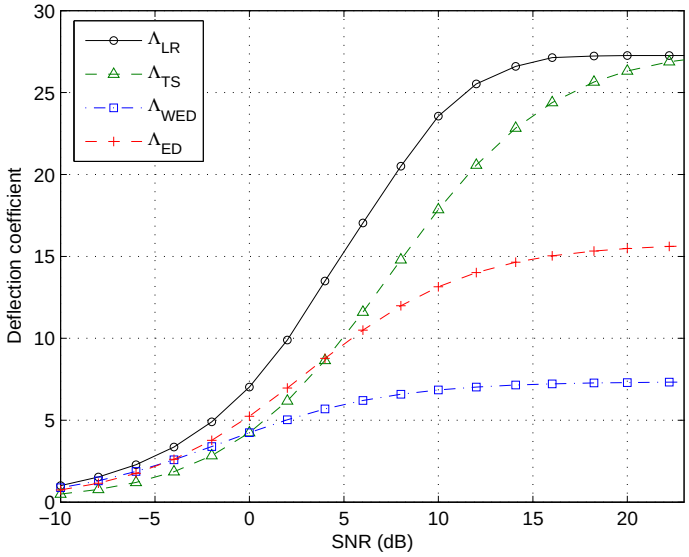


FIGURE 5.5: The performance of DC as channel SNR increases.

Figure 5.5 shows the performance of different fusion rules in terms of the DC. The settings are the same as in Figure 5.3 and Figure 5.4. As we can see, at very low channel SNR, the performance of Λ_{LR} reduces to Λ_{WED} and Λ_{ED} ; for high channel SNR, Λ_{TS} has close performance as the Λ_{LR}. Along the SNR axis, the performance of Λ_{ED} is the most robust. Even though it is an interesting metric and it completely characterizes the detection performance under the Gaussian assumption, in general, it can not be proved that a greater deflection always leads to a better performance in terms of ROC curves. Therefore, the DC performance can be regarded as a good complement to ROC curves as shown in Figure 5.4.

5.6.3 Detection performance in term of different SNRs

To better understand the performance difference as a function of channel SNR, Figure 5.6 gives the probability of detection as a function of the average channel SNR for a constant system false alarm rate of $P_F = 0.01$. Other settings is identical to the above example. From this figure, it is clear that performances are consistent with the results in Figure 5.5 using the DC for the same set of parameters. The jumpy behavior of the Λ_{LR} can be explained by its statistics. As we have

discussed in Section 5.5.1, the Λ_{LR} is binomial distributed. We have :

$$P_D = \sum_{i=K_t}^{K} p_1^i (1-p_1)^{K-i} \tag{5.23}$$

$$P_F = \sum_{i=K_t}^{K} p_0^i (1-p_0)^{K-i} \tag{5.24}$$

where K_t is the threshold and ranges in $[0,K]$. According to equation (5.23), the system level P_D is only a function of the threshold K_t. As we know, there are only finite integer numbers from which K_t can take values. The value of K_t is determined by the average channel SNR and the system level false alarm rate P_F judging from equation (5.21) and equation (5.24). Even for different SNR values, and hence different K_t values, could still be the same due to its finite alphabet property. Thus, with the same K_t, P_D remains constant for a certain range of channel SNRs.

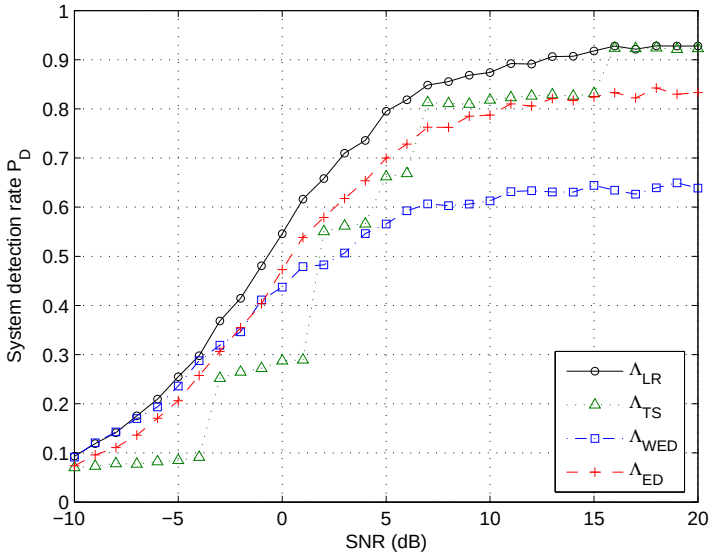


FIGURE 5.6: The performance of P_D as channel SNR increases.

To have a more intuitive understanding of how much performance degradation the three sub-optimal alternatives suffer compared to the optimal Λ_{LR}, the respective difference in P_D is shown in Figure 5.7. The smaller the difference values, the closer the performance approaches the optimal Λ_{LR}. It is obvious that at low SNR, Λ_{WED} has the smallest difference with Λ_{LR} ; however when the SNR is larger than 7dB, Λ_{TS} is closer to optimal and Λ_{WED} performs the worst. It can be noted that Λ_{WED} is the most robust as channel SNR varies and its average performance seems the best.

Generally speaking, the performance of Λ_{LR} is optimal with all SNR values ; Λ_{TS} is sub-optimal when SNR is large ; Λ_{WED} is sub-optimal when SNR is small ; Λ_{ED} performs robust with different range of SNR. With regard to computational complexity and resource cost, Λ_{LR} is the

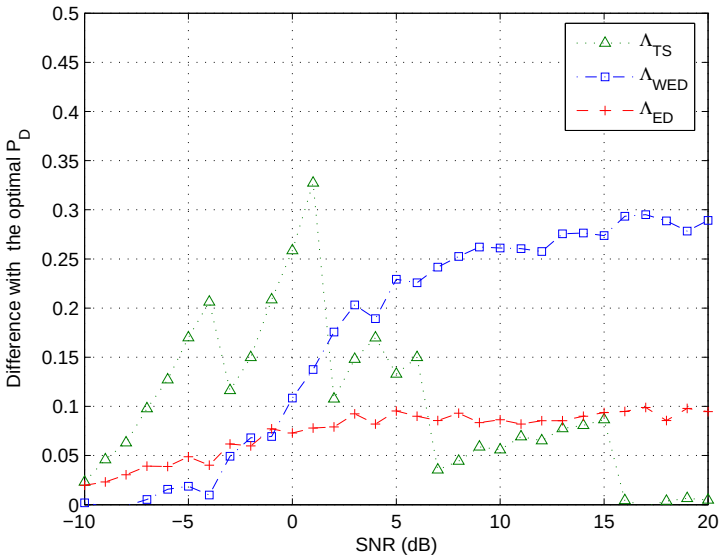


FIGURE 5.7: Difference of P_D (Λ_{TS}, Λ_{WED} and Λ_{ED} compared with Λ_{LR}).

most complicated and requires instant channel state information while Λ_{ED} is the simplest and requires least resources. In WSNs, although the algorithm performance is important, other metrics must also taken into consideration. To meet application's different requirement of QoI, there must be tradeoffs while choosing among the different rules.

5.6.4 Detection performance in term of different number of nodes

The system detection probability as a function of total number of sensors is shown in Figure 5.8. As we can see, when k is very large, the detection probability at the fusion center approaches 1, even when the local sensors have a modest detection performance with $P_d = 0.6$ and $P_f = 0.1$, and the average channel SNR is low (5dB). This is due to the accumulation of information from a large number of sensors. Note that no matter how many sensors are employed, the Λ_{LR} fusion rule outperforms the other three sub-optimal fusion rules. With the specific system parameters in this example, the performance of Λ_{ED} is quite robust and outperforms both Λ_{WED} and Λ_{TS} fusion rules, regardless of the scaling factor .

Figure 5.9 shows another example when SNR is large (30 dB). Compared to Figure 5.8, the rising of P_D in Figure 5.9 is much faster as the number of nodes increases. It is worth noting that different from Figure 5.8, the performance of Λ_{TS} is almost the same as Λ_{LR}. These results consist with Figure 5.6 and also verify the correctness of Proposition 5.4.2.

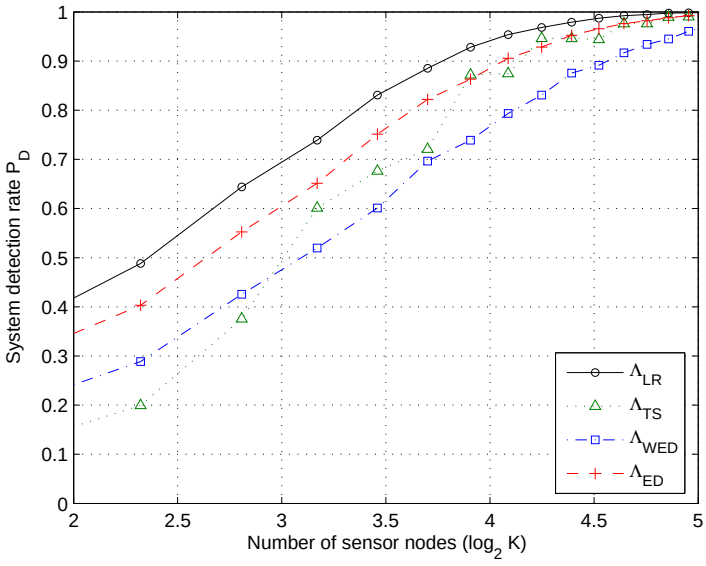


FIGURE 5.8: The performance of P_D as the number of nodes increases (SNR=5 dB).

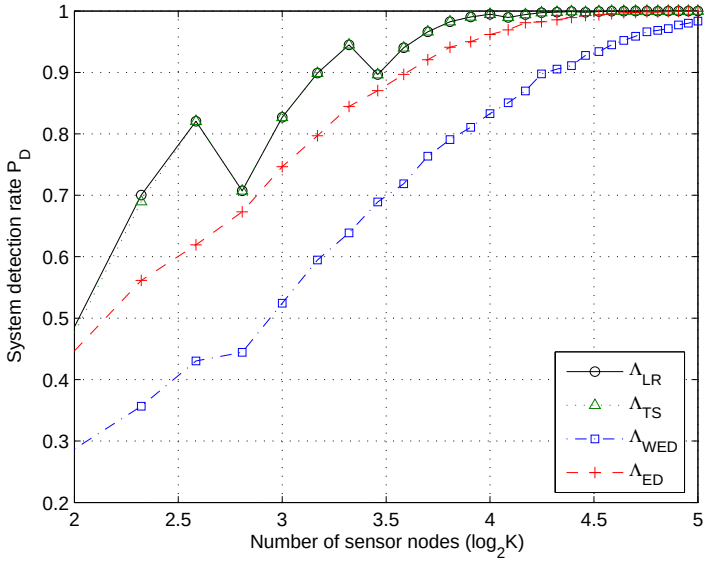


FIGURE 5.9: The performance of P_D as the number of nodes increases (SNR=30 dB)

5.6.5 System detection performance with different local sensors

In Sections 5.6.1, 5.6.2, 5.6.3 and 5.6.4, we gave examples where the detection performance are assumed identical for all sensors. In practice, however, it is often the case that different sensors have different detection performances. In this section, we investigate and compare the performances of different fusion rules in these more complicated and practical scenarios. First, we assume that all the wireless channels between local sensors and the fusion center have the same average SNR. There are totally 10 sensors. All the sensors have the same false alarm rate $P_f = 0.1$. However, they have different detection probabilities. In this particular example, we assume that $\vec{P}_d = [0.2, 0.28, 0.36, ..., 0.92]$, where $\vec{P}_d = [P_{d1}, P_{d2}, P_{d3}, ..., P_{dK}]$. The simulation results are shown in Figure 5.10. The performance tends to be identical to those in Figure 5.6, which implies that node difference does not have much impact on the performance of fusion rules.

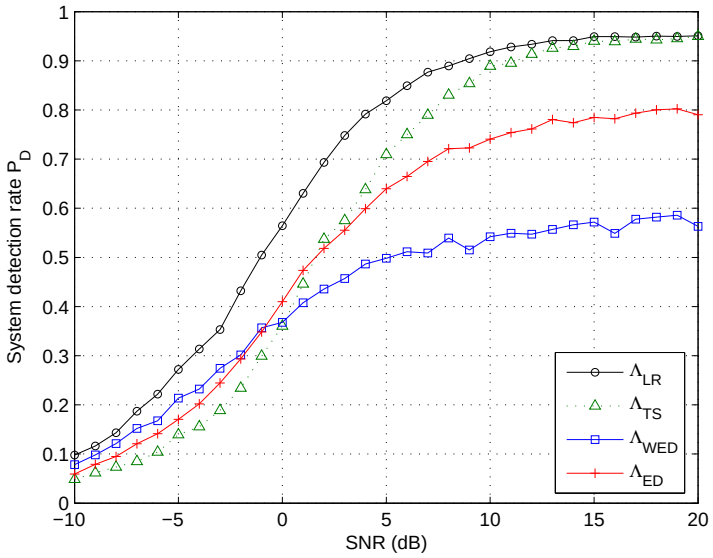


FIGURE 5.10: System detection probability as a function of average channel SNR with 10 sensors whose detection performances are different.

5.6.6 Summary of the Simulation Results

The proposed fusion rules are extensively compared, as shown in Table 5.3, in term of the performance, a priori information required, complexity and respective application scopes.

5.7 Summary

The problem of fusing binary decisions transmitted over fading and noisy channels in WSNs is studied in this chapter. Local decisions made by local sensor nodes may be lost or corrupted

TABLE 5.3: Comparison of four fusion rules.

Fusion rules	Performance	Information *a priori*	Complexity	Application
Λ_{LR}	optimal	CSI and sensor indices	highest	traditional wired network
Λ_{TS}	sub-optimal at large SNR	sensor indices	high	noiseless WSNs
Λ_{WED}	sub-optimal at low SNR	CSI	low	noisy WSNs
Λ_{ED}	robust	none	lowest	dynamic WSNs

while transmitted to the fusion center via a fading channel. We propose a series of fusion rules under the assumption of Rayleigh channel model. Likelihood ratio rule has been shown optimal through theoretical analysis and simulation. However, it consumes system resource and requires good knowledge of local and channel information, which is not easily available in resource-constrained WSNs. We propose three sub-optimal alternatives, which have less computation and information cost. They perform well in their respective SNR range. Finally we find that in resource-constrained WSNs, a tradeoff should be considered among performance, resource cost and computation complexity while choosing the fusion rules.

6

Conclusions and Future work

6.1 Conclusions

Supporting QoS in WSN is a requirement instead of a complement for WSN-based applications. Diverse QoS requirements can be guaranteed by taking effective measures in different protocol layers or functional components. For different applications, the focusing QoS metrics can be different from one another. This dissertation is a preliminary and immature exploration of the problem. We centers on QoS support for the task of event detection. When the network is deployed, the quality of coverage and connectivity should be satisfied ; as the critical data is transmitted through multi-hops, the end-to-end delay and reliability should be guaranteed while energy efficiency and load balance should be considered ; during the phase of data processing, information accuracy should be provided to ensure a high detection rate. Respective studies are summarized as follows :

– To better understand how various parameters impact on the quality of connectivity, we analyze the connectivity of WSNs in a non-isotropic log-normal shadowing environment. An explicit expression of node non-isolation probability is derived as the upper bound of one-connectivity, based on an analytical link model which incorporates important parameters such as path loss exponent, shadowing variance of channel, modulation, encoding method etc. A tight lower bound for the minimum node density that is necessary to obtain an almost surely connected network is also given. Besides, we find giant component size a good relaxed measure of connectivity in some applications that do not require full connectivity.

– To meet the quality of coverage and connectivity together, a fine deployment strategy should be developed. Assuming each point in the area is associated with a coverage threshold, which must be satisfied after nodes are deployed, the resulting deployment problem is formulated as a multi-objective optimization problem, which seeks to minimize both the gap between the generated coverage probability and the required threshold and the number of deployed nodes with the constraint of maintaining the network connectivity. Heuristic methods based on tabu search (TS) and generic algorithm (GA) are proposed. Simulations show that GA and TS deployment outperforms random and regular lattice deployment. Furthermore, GA provides

diverse solutions with better tradeoffs between two objectives.

– To satisfy real-time and reliable delivery requirement, a two-hop neighborhood information based real-time routing protocol is proposed. The packet deadline is mapped to a network-wide velocity while routing decision is made based on the novel two-hop velocity integrated with energy balancing mechanism. Initiative drop control is embedded to enhance energy utilization efficiency while reducing packet deadline miss ratio. The proposed routing has one-hop more prediction capability as using a "telescope" in finding the way. Simulation and comparison show that the new protocol has led to lower deadline miss ratio, higher energy efficiency and load balance than two existing popular schemes.

– Finally decision fusion rules under fading channel in WSN are investigated to ensure high quality of detection. Local decisions made by local sensors may be corrupted while transmitted to the fusion center via a fading channel. a series of fusion rules are proposed with the assumption of Rayleigh channel model. Likelihood ratio rule has been proved optimal by analysis and simulation. However, it is with high computation complexity and requires instant channel state information, which is not easily available in resource-constrained WSN. We propose three sub-optimal alternatives, which have less computation cost and requires less a priori information. They perform well in their respective SNR ranges. We draw the conclusion that in resource-constrained WSN, a tradeoff should be considered among performance, resource cost and computation complexity while choosing the fusion rules.

6.2 Future Work

QoS has been intensively investigated in wired networks. However for WSN, QoS has been given quite a lot of new contents due to its unique characteristics and challenges. Although many researches have been made in this domain, there are still problems left open. Based on the work of this dissertation, the following problems can be further investigated :

– Although maintaining full sensing range of the AoI guarantees immediate response to emergent event, it is not favorable due to its high energy consumption. In some real-time applications, an immediate response is not required and the detection within certain delay bound should be tolerable. Therefore, designers are willing to sacrifice some delay in detection in exchange for prolonged system lifetime. The challenge is how to guarantee the required quality of surveillance in an energy-saving mode.

– Most of the existing WSN protocol designs aim at multiple source to single destination model. It should be noted that the result may not be applicable to a more sophisticated multi-source multi-sink system. Conflicts can occur among crossed tasks in the network. It is challenging to satisfy the QoS requirements of multiple tasks simultaneously. In wireless sensor-actuator networks (WSANs), sensors have multiple potential destinations (e.g. actuators) in event reporting. Besides, actuators could be mobile in conducting actions. The resulting heterogenous system is quite different from a traditional data collection WSN model

with single static sink. Note that, in WSANs, actuators are often assumed resource-rich and with high energy and communication capability. This implies a possibility of using them as fast relays to transmit data collected from sensors to a destination directly or in a few hops for time-critical service. Interesting explorations are expected.

– Cross-layer design should be a promising tendency in the QoS-based research. In this dissertation, information from MAC layer are taken as feedbacks to routing decision making. Further co-design of routing and detection fusion are expected to get the idea that where and when to do detection fusion can satisfy QoS requirement by the greatest extent.

Table des figures

Bibliographie

[1] M. Weiser, "The computer for the 21st century," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 3, no. 3, pp. 3–11, 1999.

[2] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *ACM WSNA*, 2002, pp. 88–97.

[3] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, "Codeblue : an ad hoc sensor network infrastructure for emergency medical care," in *Int. Workshop Wearable Implantable Body Sens. Networks*, Apr. 2004.

[4] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, and M. Gouda, "A line in the sand : A wireless sensor network for target detection, classification, and tracking," *Computer Networks*, vol. 46, no. 5, pp. 605–634, 2004.

[5] N. Ota and P. Wright, "Trends in wireless sensor networks for manufacturing," *Int. J. Manufacturing Research*, vol. 1, no. 1, pp. 3–17, 2006.

[6] K. Kim, J. Jun, S. Kim, and B. Sung, "Medical asset tracking application with wireless sensor networks," in *SENSORCOMM*, 2008, pp. 531–536.

[7] S. Nourizadeh, C. Deroussent, Y. Song, and J. Thomesse, "Medical and home automation sensor networks for senior citizens telehomecare," in *ICC Communications Workshops*, 2009, pp. 1–5.

[8] R. Aylward and J. Paradiso, "A compact, high-speed, wearable sensor network for biomotion capture and interactive media," in *Proceedings of IPSN*, 2007, pp. 380–389.

[9] H. Karl and A. Willig, *Protocols and architectures for wireless sensor networks*. John Wiley & Sons Ltd, 2005.

[10] L. Rossi, B. Krishnamachari, and C. Kuo, "Monitoring of diffusion processes with pde models in wireless sensor networks," in *Defense and Security Symposium*, 2004, pp. 12–16.

[11] C. Wan, A. Campbell, and L. Krishnamurthy, "PSFQ : a reliable transport protocol for wireless sensor networks," in *ACM WSNA*, 2002, pp. 1–11.

[12] F. Stann and J. Heidemann, "Rmst : Reliable data transport in sensor networks," in *IEEE SNPA*, 2003, pp. 102–112.

[13] Y. Sankarasubramaniam, Ö. Akan, and I. Akyildiz, "ESRT : event-to-sink reliable transport in wireless sensor networks," in *ACM MobiHoc*, 2003, pp. 177–188.

[14] T. He, J. Stankovic, C. Lu, and T. Abdelzaher, "SPEED : a stateless protocol for real-time communication in sensor networks," in *ICDCS*, May 2003, pp. 46–55.

[15] E. Felemban, C. G. Lee, and E. Ekici, "MMSPEED : Multipath multi-speed protocol for QoS guarantee of reliability and timeliness in wireless sensor network," *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 738–754, 2006.

[16] O. Chipara, Z. He, G. Xing, Q. Chen, X. Wang, C. Lu, J. Stankovic, and T. Abdelzaher, "Real-time power-aware routing in sensor network," in *IEEE IWQoS*, Jun. 2006, pp. 83–92.

[17] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks : Attacks and countermeasures," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 293–315, 2003.

[18] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks," *IEEE/ACM Trans. Networking*, vol. 12, no. 3, pp. 493–506, June 2004.

[19] T. van Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *ACM Sensys*, Nov. 2003, pp. 65–72.

[20] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *ACM SenSys*, 2004, pp. 95–107.

[21] Y. Liu, I. Elhanany, and H. Qi, "An energy-efficient qos-aware media access control protocol for wireless sensor networks," in *IEEE MASS*, 2005, pp. 189–191.

[22] M. Caccamo, L. Zhang, L. Sha, and G. Buttazzo, "An implicit prioritized access protocol for wireless sensor networks," in *IEEE RTSS*, Dec. 2002, pp. 39–48.

[23] I. Rhee, A. Warrier, M. Aia, and J. Min, "Z-MAC : a hybrid MAC for wireless sensor networks," in *ACM Sensys*, Nov. 2005, pp. 90–101.

[24] IEEE Std 802.15.4, "Part 15.4 : Wireless medium access (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (WPANs)," IEEE-SA Standards Board, 2006.

[25] Crossbow Motes. http ://www.xbow.com.

[26] I. G., W. K., R. S., and M. G., "Continuum percolation of wireless ad hoc communication networks," *Statistical Mechanics and its Applications*, vol. 325, pp. 577–600, 2003.

[27] M. Zuniga and B. Krishnamachari, "Analyzing the transitional region in low power wireless links," in *IEEE SECON*, 2004, pp. 517–526.

[28] D. Ganesan, B. Krishnamachari, D. Woo, A. ; Culler, and S. Estrin, D.and Wicker, "Complex behavior at scale : An experimental study of low-power wireless sensor networks," *Technical Report CSD-TR 02-0013, UCLA*, 2002.

[29] A. Woo, T. Tong, and D. Culler, "Taming the underlying challenges of reliable multihop routing in sensor networks," in *ACM Sensys*, Nov. 2003, pp. 14–27.

[30] A. Cerpa, J. Wong, L. Kuang, M. Potkonjak, and D. Estrin, "Statistical model of lossy links in wireless sensor networks," in *ACM/IEEE IPSN*, 2005, pp. 81–88.

[31] G. Zhou, T. He, S. Krishnamurthy, and J. Stankovic, "Models and solutions for radio irregularity in wireless sensor networks," *ACM Trans. Senser Networks*, vol. 2, no. 2, pp. 221–262, 2006.

[32] D. Lal, A. Manjeshwar, F. Herrmann, E. Uysal-Biyikoglu, and A. Keshavarzian, "Measurement and characterization of link quality metrics in energy constrained wireless sensor networks," in *IEEE GLOBECOM*, 2003, pp. 446–452.

[33] Y. Cheng and R. T.G., "Critical connectivity phenomena in multihop radio models," *IEEE Trans. on Communication*, vol. 7, no. 37, pp. 770–777, 1989.

[34] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," *Stochastic Analysis, Control, Optimization and Applications*, pp. 547–566, 1998.

[35] F. M., B. L., B. J., C. M., and M. R., "Percolation in multi-hop wireless networks," *IEEE Trans. on Information Theory*, 2003.

[36] K. B., W. S.B., and B. R., "Phase transition phenomena in wireless ad-hoc networks," in *IEEE GLOBECOM*, 2001, pp. 2921–2925.

[37] C. Bettstetter, "On the connectivity of ad hoc networks," *The Computer Journal*, vol. 47, no. 4, pp. 432–447, 2004.

[38] R. Hekmat and V. P. Mieghem, "Connectivity in wireless ad-hoc networks with a log-normal radio model," *Mobile Networks and Applications*, vol. 11, no. 3, pp. 351–360, 2006.

[39] A. Helmy, "Small worlds in wireless networks," *IEEE Communications Letters*, vol. 7, no. 10, pp. 490–492, 2003.

[40] O. Dousse, F. Baccelli, and P. Thiran, "Impact of interferences on connectivity in ad hoc networks," *IEEE/ACM Trans. Networking*, vol. 13, no. 2, pp. 425–436, 2005.

[41] I. Glauche, W. Krause, R. Sollacher, and M. Greiner, "Continuum percolation of wireless ad hoc communication networks," *Physica A : Statistical Mechanics and its Applications*, vol. 325, no. 3-4, pp. 577–600, 2003.

[42] C. Bettstetter and C. Hartmann, "Connectivity of wireless multihop networks in a shadow fading environment," *Wireless Networks*, vol. 11, no. 5, pp. 571–579, 2005.

[43] P. Santi, "Topology control in wireless ad hoc and sensor networks," *ACM Computing Surveys (CSUR)*, vol. 37, no. 2, pp. 164–194, 2005.

[44] Y. Sankarasubramaniam, I. Akyildiz, and S. McLaughlin, "Energy efficiency based packet size optimization in wireless sensor networks," in *IEEE WSNA*, 2003, pp. 1–8.

[45] S. Song, D. Goeckel, and D. Towsley, "Collaboration improves the connectivity of wireless networks," in *IEEE INFOCOM*, 2006, pp. 1–11.

[46] B. Bollobas, *Random graphs*. Cambridge University Press, 2001.

[47] H. Ammari and S. Das, "Integrated coverage and connectivity in wireless sensor networks : A two-dimensional percolation problem," *IEEE Trans. Comput.*, vol. 57, no. 10, pp. 1423–1434, 2008.

[48] X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration in wireless sensor networks," in *ACM SenSys*, 2003, pp. 28–39.

[49] X. Bai, D. Xuan, Z. Yun, T. Lai, and W. Jia, "Complete optimal deployment patterns for full-coverage and k-connectivity ($k \leq 6$) wireless sensor networks," in *ACM MobiHoc*, 2008, pp. 401–410.

[50] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm : NSGA-II," *IEEE Trans. on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[51] R. K. amd H. Liu and H.-H. Chen, "Reduced complexity intrusion detection in sensor networks using genetic algorithm," in *IEEE ICC*, 2009.

[52] M. Alabau, L. Idoumghar, and R. Schott, "New hybrid genetic algorithms for the frequency assignment problem," *IEEE Trans. Broadcast.*, vol. 48, no. 1, pp. 27–34, 2002.

[53] X. Bai, Z. Yun, D. Xuan, T. Lai, and W. Jia, "Deploying Four-Connectivity And Full-Coverage Wireless Sensor Networks," in *IEEE Infocom*, 2008, pp. 296–300.

[54] Y. Zou and K. Chakrabarty, "A distributed coverage-and connectivity-centric technique for selecting active nodes in wireless sensor networks," *IEEE Trans. Computers*, vol. 54, no. 8, pp. 978–991, 2005.

[55] N. Ahmed, S. Kanhere, and S. Jha, "Probabilistic coverage in wireless sensor networks," in *IEEE LCN*, 2005, pp. 672–681.

[56] B. Liu and D. Towsley, "A study of the coverage of large-scale sensor networks," in *IEEE MASS*, 2004, pp. 475–483.

[57] R. Murty, A. Gosain, M. Tierney, A. Brody, A. Fahad, J. Bers, and M. Welsh, "Citysense : A vision for an urban-scale wireless networking testbed," in *IEEE International Conference on Technologies for Homeland Security*, 2008.

[58] R. Cardell-Oliver, M. Kranz, K. Smettem, and K. Mayer, "A reactive soil moisture sensor network : Design and field evaluation," *International journal of distributed sensor networks*, vol. 1, no. 2, pp. 149–162, 2005.

[59] M. Younis and K. Akkaya, "Strategies and techniques for node placement in wireless sensor networks : A survey," *Ad Hoc Networks*, vol. 6, no. 4, pp. 621–655, 2008.

[60] R. Kershner, "The number of circles covering a set," *American Journal of Mathematics*, 1939.

[61] A. Arora, R. Ramnath, E. Ertin, P. Sinha, S. Bapat, V. Naik, V. Kulathumani, H. Zhang, H. Cao, M. Sridharan *et al.*, "Exscal : Elements of an extreme scale wireless sensor network," in *IEEE RTSCA*, 2005, pp. 102–108.

[62] X. Liu, Q. Wang, W. He, M. Caccamo, and L. Sha, "Optimal real-time sampling rate assignment for wireless sensor networks," *ACM Trans. Sensor Networks (TOSN)*, vol. 2, no. 2, pp. 263–295, 2006.

[63] Y. Chen, C. Chuah, and Q. Zhao, "Sensor placement for maximizing lifetime per unit cost in wireless sensor networks," in *IEEE MILCOM*, 2005, pp. 1–6.

[64] J. Bredin, E. Demaine, M. Hajiaghayi, and D. Rus, "Deploying sensor networks with guaranteed capacity and fault tolerance," in *ACM MobiHoc*, 2005, pp. 309–319.

[65] T. C. Shermer, "Recent results in art galleries," *Proceedings of the IEEE*, vol. 80, no. 9, pp. 1384–1399, 1992.

[66] H. González-Baños, "A randomized art-gallery algorithm for sensor placement," in *ACM SoCG*, 2001, pp. 232–240.

[67] K. Chakrabarty, S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks," *IEEE Transactions on Computers*, vol. 51, p. 12, 2002.

[68] S. Dhillon and K. Chakrabarty, "Sensor placement for effective coverage and surveillance in distributed sensor networks," in *IEEE WCNC*, vol. 3, 2003, pp. 1609–1614.

[69] F. Lin and P. Chiu, "A near-optimal sensor placement algorithm to achieve complete coverage/discrimination in sensor networks," *IEEE Communications Letters*, vol. 9, no. 1, p. 43, 2005.

[70] Y. Zou and K. Chakrabarty, "Uncertainty-aware and coverage-oriented deployment for sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 788–798, 2004.

[71] J. Zhang, T. Yan, and S. Son, "Deployment strategies for differentiated detection in wireless sensor networks," in *IEEE SECON*, 2006, pp. 316–325.

[72] N. Aitsaadi, N. Achir, K. Boussetta, and G. Pujolle, "Heuristic Deployment to Achieve Both Differentiated Detection and Connectivity in WSN," in *IEEE VTC Spring*, 2008, pp. 123–127.

[73] G. Tan, S. Jarvis, and A. Kermarrec, "Connectivity-guaranteed and obstacle-adaptive deployment schemes for mobile sensor networks," in *IEEE ICDCS*, 2008, pp. 429–437.

[74] J. Stankovic, T. Abdelzaher, C. Lu, L. Sha, and J. Hou, "Real-time communication and coordination in embedded sensor networks," *Proceedings of IEEE*, vol. 91, no. 7, pp. 1002–1022, 2003.

[75] M. A. Spohn and J. J. Garcia-Luna-Aceves, "Enhancing broadcase operations in ad hoc networks with two-hop connected dominating sets," in *IEEE MASS*, 2004, pp. 543–545.

[76] W. Lou and J. Wu, "On reducing broadcast redundancy in ad hoc wireless networks," *IEEE Trans. Mobile Comput.*, vol. 1, no. 2, pp. 111–122, 2002.

[77] F. Ingelrest, D. Simplot-Ryl, and I. Stojmenovic, "Smaller connected dominating sets in ad hoc and sensor networks based on coverage by two-hop neighbors," in *Comm. Syst. Software and Middleware (COMSWARE)*, 2007, pp. 1–8.

[78] G. Calinescu, "Computing 2-hop neighborhoods in ad hoc wireless networks," in *AdHoc-Now*, 2003, pp. 175–186.

[79] V. Rajendran, K. Obraczka1, and J. J. Garcia-Luna-Aceves, "Energy-efficient, collision-free medium access control for wireless sensor networks," *Wireless Networks*, vol. 12, no. 1, pp. 63–78, Feb. 2006.

[80] T. He, J. Stankovic, C. Lu, and T. Abdelzaher, "A spatiotemporal communication protocol for wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 16, no. 10, pp. 995–1006, 2005.

[81] B. Nefzi and Y.-Q. Song, "Performance analysis and improvement of zigbee routing protocol," in *IFAC FET*, Nov. 2007, pp. 199–206.

[82] K. Akkaya and M. Younis, "An energy-aware QoS routing protocol for wireless sensor networks," in *IEEE ICDCS*, May 2003, pp. 710–715.

[83] N. Boughanmi and Y.-Q. Song, "A new routing metric for satisfying both energy and delay constraints in wireless sensor networks," *J. Signal Process. Syst.*, vol. 51, no. 2, pp. 137–143, 2008.

[84] I. Stojmenovic, *Handbook of Sensor Networks : Algorithms and Architectures*. Hoboken, New Jersey : John Wiley & Sons, Inc., 2005.

[85] IEEE P802.15.4aD/4 (Amendment of IEEE Std 802.15.4), "Part 15.4 : Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs)," IEEE-SA, Jul. 2006.

[86] I. Stojmenovic and X. Lin, "Loop-free hybrid single-path/flooding routing algorithm with guaranteed delivery for wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 12, no. 10, pp. 1023–1032, 2001.

[87] B. Krishnamachari, S. Wicker, R. Bejar, and M. Pearlman, *Critical Density Thresholds in Distributed Wireless Networks*. Book chapter in Communications, Information and Network Security, 2002.

[88] L. Bao and J. Garcia-Luna-Aceves, "Transmission scheduling in ad hoc networks with directional antennas," in *ACM Mobicom*, 2002, pp. 48–58.

[89] Network Embedded Systems Technology, Vanderbilt University. Prowler : probabilistic wireless network simulator. http ://www.isis.vanderbilt.edu/projects/nest/prowler.

[90] Y. Zhang. Rmase : routing modeling application simulation environment. http ://www2.parc.com/isl/groups/era/nest/Rmase.

[91] Y. Li, Y.-Q. Song, R. Schott, Z. Wang, and Y. Sun, "Impact of link unreliability and asymmetry on the quality of connectivity in large-scale sensor networks," *J. Sensors*, vol. 8, no. 10, pp. 6674–6691, 2008.

[92] K. Seada, M. Zuniga, A. Helmy, and B. Krishnamachari, "Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks," in *ACM SenSys*, 2004, pp. 108–121.

[93] S. Zahedi, M. Srivastava, and C. Bisdikian, "A framework for quality of information analysis for detection-oriented sensor network deployments."

[94] R. Tenney and N. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerospace and Electronic Systems*, vol. 17, no. 1, pp. 501–510, 1981.

[95] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerospace and Electronic Systems*, vol. 22, no. 1, pp. 98–101, 1986.

[96] A. Reibman and L. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerospace and Electronic Systems*, pp. 24–30, 1987.

[97] J. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 407–416, 2003.

[98] Y. Yuan and M. Kam, "Distributed decision fusion with a random-access channel for sensor network applications," *IEEE Trans. Instrumentation and Measurement*, vol. 53, no. 4, pp. 1339–1344, 2004.

[99] S. Aldosari and J. Moura, "Fusion in sensor networks with communication constraints," in *IEEE/ACM IPSN*, 2004, pp. 108–115.

[100] S. Thomopoulos and L. Zhang, “Distributed decision fusion with networking delays and channel errors,” *Information Science*, vol. 66, pp. 91–118, 1992.

[101] B. Picinbono, “On deflection as a performance criterion in detection,” *IEEE Trans. Aerospace and Electronic Systems*, vol. 31, no. 3, pp. 1072–1081, 1995.

Nancy-Université
Université Henri Poincaré

A NANCY, le 02 novembre 2010

No étudiant : 26006725

LI YANJUN
38 ZHEDA ROAD
ZHEJIANG UNIVERSITY
310027 ZHEJIANG
CHINE
CHINE

Madame,

Par décision en date du 30 octobre 2010, vous avez été autorisée à présenter en soutenance vos travaux en vue de l'obtention du diplôme :

DOCTORAT UHP INFORMATIQUE

La soutenance aura lieu le 10 novembre 2010 à 14h00 à l'adresse suivante :

Faculté des Sciences & Technologies - Bâtiment LORIA (amphithéatre C) - Campus Victor Grignard
54500 Vandoeuvre-lès-Nancy

La soutenance sera publique.

Je vous prie d'agréer, Madame, l'expression de mes salutations distinguées.

Le Président de l'Université

J.-P. FINANCE

24 - 30 rue Lionnois | BP 60120
54003 Nancy Cedex | FRANCE
Tél. : +33(0)3 83 68 20 00
Fax : +33(0)3 83 68 21 00

info@uhp-nancy.fr | www.uhp-nancy.fr

Résumé

Cette thèse présente nos travaux à la fois théoriques et techniques sur la fourniture de la qualité de service dans les réseaux de capteurs sans fil, travaux développés principalement pour la détection en temps réel d'événements.
Le premier problème fondamental pour assurer une qualité de service est la connectivité d'un réseau. La probabilité de non-isolation de nœud est donnée garantissant une borne supérieure de 1-connectivité du réseau. Un deuxième problème traité concerne la considération conjointe du problème de connectivité de communication et de couverture de détection. Ce problème étant formalisé comme un problème d'optimisation multi-objectif, un algorithme heuristique du type génétique a été développé, permettant ainsi d'aider au déploiement. Pour assurer la communication des données en temps réel et de façon fiable, un protocole de routage basé sur SPEED a été développé. Les simulations ont montré une amélioration notable de performances par rapport aux solutions existantes. Afin d'assurer la fiabilité de l'information finale, un ensemble de règles de fusion de décision a été proposé. Quant à son implémentation réelle avec moins de complexité sur des nœud de ressources limitées, trois alternatives sousoptimales ont été proposées et qui donnent des performances satisfaisantes dans des plages de rapport signal sur bruit.

Mots clés : Réseaux de capteurs sans fil, Qualité de service, Couverture, Connectivité, déploiement, routage, fusion de données.

Abstract

The fundamental theories and key technologies of QoS support for event detection in wireless sensor networks are studied in this dissertation.
Firstly, to ensure the quality of connectivity, an explicit expression of node non-isolation probability is derived as the upper bound of one-connectivity. A tight lower bound for the minimum node density is also given for obtaining an almost surely connected network. Secondly, to meet the quality of coverage and connectivity together, a fine deployment strategy is developed. The deployment problem is formulated as a multi-objective optimization problem. Heuristic methods based on tabu search and generic algorithms are proposed. Thirdly, to satisfy real-time and reliable delivery requirement, a two-hop neighborhood information based real-time routing protocol is proposed. The simulations show that the proposed routing protocol has a significant improvement in performance compared to existing solutions. Finally, decision fusion rules under fading channel are investigated to ensure high quality of information. We propose three sub-optimal alternatives to the optimal likelihood ratio rule. They have less computation cost and require less a priori information and perform well in their respective SNR ranges.

Keywords: wireless sensor networks, quality of service, coverage, connectivity, deployment, routing, decision fusion.