RELIABLE AND EFFICIENT COMMUNICATIONS IN WIRELESS SENSOR NETWORKS

By

Mai M. Abdelhakim

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Electrical Engineering - Doctor of Philosophy

2014

ABSTRACT

RELIABLE AND EFFICIENT COMMUNICATIONS IN WIRELESS SENSOR NETWORKS

By

Mai M. Abdelhakim

Wireless sensor network (WSN) is a key technology for a wide range of military and civilian applications. Limited by the energy resources and processing capabilities of the sensor nodes, reliable and efficient communications in wireless sensor networks are challenging, especially when the sensors are deployed in hostile environments. This research aims to improve the reliability and efficiency of time-critical communications in WSNs, under both benign and hostile environments.

We start with wireless sensor network with mobile access points (SENMA), where the mobile access points traverse the network to collect information from individual sensors. Due to its routing simplicity and energy efficiency, SENMA has attracted lots of attention from the research community. Here, we study reliable distributed detection in SENMA under Byzantine attacks, where some authenticated sensors are compromised to report fictitious information. The q-out-of-m rule is considered. It is popular in distributed detection and can achieve a good trade-off between the miss detection probability and the false alarm rate. However, a major limitation with this rule is that the optimal scheme parameters can only be obtained through exhaustive search. By exploiting the linear relationship between the scheme parameters and the network size, we propose simple but effective sub-optimal linear approaches. Then, for better flexibility and scalability, we derive a near-optimal closed-form solution based on the central limit theorem. It is proved that the false alarm rate of the q-out-of-m scheme diminishes exponentially as the network size increases, even if the percentage of

malicious nodes remains fixed. This implies that large-scale sensor networks are more reliable under malicious attacks. To further improve the performance under time-varying attacks, we propose an effective malicious node detection scheme for adaptive data fusion; the proposed scheme is analyzed using the entropy-based trust model, and has shown to be optimal from the information theory point of view.

Next, we observe that: while simplifying the routing process, a major limitation with SENMA is that data transmission is limited by the physical speed of the mobile access points (MAs) and the length of their trajectory, resulting in low throughput and large delay. To solve this problem, we propose a novel mobile access coordinated wireless sensor network (MC-WSN) architecture. The proposed MC-WSN can provide reliable and time-sensitive information exchange through hop number control, which is achieved by active network development and topology design. We discuss the optimal topology design for MC-WSN such that the average number of hops between the source and its nearest sink is minimized, and analyze the performance of MC-WSN in terms of throughput, stability, delay, and energy efficiency by exploiting tools in information theory, queuing theory, and radio energy dissipation model. It is shown that MC-WSN achieves much higher throughput and significantly lower delay and energy consumption than that of SENMA.

Finally, motivated by the observation that the number of hops in data transmission has a direct impact on the network performance, we introduce the concept of the N-hop networks. Based on the N-hop concept, we propose a unified framework for wireless networks and discuss general network design criteria. The unified framework reflects the convergence of centralized and ad-hoc networks. It includes all exiting network models as special cases, and makes the analytical characterization of the network performance more tractable. Further study on N-hop networks will be conducted in our future research.

Dedicated to my dear parents and to my beloved husband.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation and gratitude to my advisor, Prof. Tongtong Li, for her guidance and support throughout the years of my PhD studies at Michigan State University. The experience I gained while working with Prof. Li is extremely rewarding. I learned a lot from her broad knowledge and insightful comments. Also, I have always appreciated her thoughtfulness and kindness that have created a family-like atmosphere in her research group. It is certainly an honor to have worked with Prof. Li.

I would like to thank Prof. Hassan Khalil, Prof. Selin Aviyente, and Prof. Guoliang Xing for serving on my committee, and for their helpful comments and insightful discussions.

Thanks to Prof. Jian Ren for his valuable insights and suggestions on security-related issues in research. Thanks to Prof. Hyder Radha for supporting my application at Michigan State University, and for nominating me for the graduate office fellowship in the first year of my PhD program.

I am blessed with a wonderful family without whom this work would not have been possible. I would like to express my profound gratitude to my parents for their endless love and encouragement. They have always motivated me to excel, and have strived to help me achieve my goals since I was a child.

I could not be more grateful to my husband, Mostafa, for being the loving, caring, and supportive person that he is. Without him, my life would not have been complete or enjoyable, and this dissertation would not have been written.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In this chapter, first, a brief overview of wireless sensor networks is provided, illustrating the sensor technology and different network structures. Second, vital performance measures in wireless sensor network design are presented. Third, security and reliability aspects are discussed, pointing-out different security threats and possible countermeasures. Finally, the main contributions of this dissertation are highlighted.

1.1 Overview of Wireless Sensor Networks

Wireless sensor networks (WSNs) have received significant attention form the research community, due to their potential impact on various applications [1–3]. WSNs were initially motivated by military reconnaissance and surveillance applications. Currently, they have been identified as key enabling technology for various civilian applications as well, such as environmental monitoring, emergency response, smart transportation systems, and target tracking. In the following subsections, the sensor technology is presented, then different network structures that can be adopted in WSNs are discussed.

1.1.1 Sensor Technology

In 1980s-1990s, sensor networks were recognized as an essential component in warfare, where sensors were deployed for collaborative detection, reconnaissance, and surveillance purposes [4]. For example, employing a system with multiple radars to collect

information about air targets. At that time, sensors' sizes were large and they had separate units for sensing, processing, and communications.

During the last two decades, along with the advancements in microelectromechnical systems (MEMS), a tremendous improvement in sensors technology has been witnessed. Now, smaller and cheaper sensors are widely available. Each sensor has an integrated sensing, data processing, and wireless communications units.

The sensors platforms can be generally classified into: specialized sensors, generic sensors, high-bandwidth sensors, and gateway-class sensors platforms [5, 6]. The specialized sensors are tiny low-cost sensors with the most constraint resources. Asset tags are examples of specialized sensors. The generic sensors are more powerful than specialized sensors and can serve as their data collectors. Examples on generic sensors can be found in [6–9]. Specialized and generic sensors run an operating system called TinyOS, which is capable of operating on platforms having limited memory and processing capabilities [5].

Among the generic sensor nodes, MICA platforms are very common. They include the MICA2, MICA2DOT, and MICAz platforms. Both MICA2 and MICAz have the IEEE 802.15.4/Zigbee compliant transceiver. The MICA2 [10] platform can be employed in large-scale sensor networks (networks with more than 1000 sensors), security, and surveillance applications. Its RF transmit power ranges from -20dBm to 5dBm, and its receiver sensitivity is -98dBm. The outdoor transmission range is 500ft under a line of sight (LOS) transmission. It operates in the 868/916MHz band and supports data rate of 38.4Kbps. The MICA2 sensor's size is $36mm \times 48mm \times 9$mm. The MICA2DOT sensor has a coin-sized form of diameter 25mm; its platform is mostly similar to that of MICA2 [11].

The MICAz platform can be utilized in indoor building monitoring and security applications, as well as large-scale WSNs [12]. Its transmission power ranges from

-24dBm to 0dBm, and its receiver sensitivity is -94dBm. The outdoor RF transmission range is 75m to 100m with LOS transmission, while the indoor RF range is 20m to 30m. The MICAz RF technology operates in the 2400MHz to 2483.5MHz band and its data rate is 250Kbps. The MICAz sensor's size is $58 \times 32 \times 7$mm excluding the battery pack. In both MICA2 and MICAz sensors, the power is supplied through two AA batteries.

The high-bandwidth sensing platforms provide higher data rate and have more computational capabilities than generic sensor platforms. For example, the "imote" platform, which is developed by Intel [5], supports data rate of 723.2Kbps [6] and its communication range is 30m [13]. The imote incorporates ARM processor and is based on the Bluetooth technology that uses the frequency-hopping spread spectrum technique. The power in the imote is supplied through three AA batteries with recharging capability [7].

Highly powerful sensor platforms are enabled through gateway-class platforms, such as the Intel Stargate platform [5]. Gateway nodes provide links between the sensor network and the conventional infrastructure support, such as Ethernet and WiFi. Stargate is based on Intel XScale (32-bit) microcontroller and has IEEE 802.11b RF module, which supports data rate of $1-11$Mbps [6]. Stargate-class sensors run Linux operating system [5].

1.1.2 Sensor Network Structures

In WSNs, generally the sensors report their readings to a central unit or a sink for processing and final decision making. Due to the limited communication range and power resource of sensors, direct (one-hop) transmissions from a source to its intended destination or a sink might not be possible, especially in fixed large-scale networks. Allowing multihop data transmissions through intermediate relays would solve this

problem and would achieve effective data delivery. The network structure defines how information is exchanged in the network.

Network structures can be generally divided into two categories: (i) distributed or ad-hoc networks, which are virtually structureless (ii) centralized networks with well-defined infrastructure. In ad-hoc networks, there are no prior established routes for data transmission; therefore, the transmission process is random, and theoretically the number of hops can be infinite. In structured network, high-level controllers (such as base stations) are employed to coordinate the network and assist data transfer. In this case, data transfer can go through defined routing paths that are usually established during the network set-up phase. Under normal network conditions, the number of hops along a route from a source to a sink is bounded.

There is also a trend to blend ad-hoc and centralized network models together, resulting in various hybrid network models [14, 15]. For example, in [15], the transmission is made in an ad-hoc mode only if the number of hops is below a certain limit; otherwise, the transmission is made through the centralized base stations. Another representative example is the clustered wireless sensor networks, where the sensors are grouped into clusters with each cluster managed by a cluster head in a centralized manner [16]. The cluster heads are then responsible for routing the information.

Along with the advancements in remote control technologies, Unmanned Aerial Vehicles (UAVs) have been utilized in wireless sensor networks for data collection. For example, in Sensor Networks with Mobile Access points (SENMA) [17], powerful mobile access points (MAs) traverse the network to establish direct communications with each sensor node. In [18], mobile relays are exploited, where each cluster is served by a mobile relay that collects data from its cluster members at predefined locations, then travel over almost a straight line trajectory to send the data to the sink. In [19], multiple locations for data collection (referred to as rendezvous points) are defined,

such that a mobile sink visits the predefined locations and stays at each location for a certain amount of time before it leaves to the next location. In this case, there are three states for the mobile sink, namely, traveling state, waiting state, and data collecting (or harvesting) state. The case when multiple mobile sinks are employed is also investigated in [19].

In many applications, the information transmitted over WSN is critical and time-sensitive. For example, detecting a target in a battlefield, detecting a fire in a building, or monitoring radiation level in the air [20]. Hence, the reliability and efficiency of information generation, transmission, and retrieval is crucial in WSN. In Section 1.2, several vital performance measures and aspects in wireless sensor networks design are presented. Then, in Section 1.3, reliability and security issues that threaten the functionality of a network are discussed.

1.2 Performance Measures in Wireless Sensor Networks

Vital performance measures in wireless sensor networks are summarized in the following.

- *Throughput* The throughput is an important measure of the network performance, as it indicates the amount of information that can be successfully communicated over a network. A throughput of a node is feasible if there exists a communication protocol such that the average transmission rate of a node is equal to its throughput. Throughput affects the buffer dynamics in the nodes [21], and consequently has a direct influence on the stability of the network as well as the delay performance.

The adopted network model largely impacts the performance of the system. For example, in random ad-hoc networks with multiple source-destination pairs, it was shown in [22] that the average throughput of a node vanishes as the number of nodes in the network increases. More specifically, when n is the total number of nodes in the network and W is the maximum rate on any link, then the throughput obtained by each node is $O\left(\frac{W}{\sqrt{n}}\right)$. This indicates that for efficient communications, the network should not be completely structureless.

The throughput of regularly-structured canonical network is considered in [23], and it was shown that proper routing is needed to improve the throughput performance in multihop transmissions. In [24], throughput of multihop many-to-one network with uniformly deployed nodes and time division multiple access (TDMA) protocol is investigated, and the effect of clustering the network on improving the throughput performance is discussed.

Cooperative transmissions and coding at intermediate hop levels can improve the network throughput [25–27]. In [25, 26], the achievable rate of relay channels is provided, where the destination receives the transmitted signal from the source as well as from the intermediate relay(s), then performs joint decoding. In sensor networks, due to the limited communication range and power resource of the sensors, it may not be possible to have the destination receive direct signals from the sources. Direct transmissions would also result in an increased interference region, which would reduce the spectral efficiency of the network.

Allowing nodes to be mobile could improve the throughput performance compared to fixed networks. In [28], the packet stream at the source node is split over multiple mobile relays that in turn deliver the packets to the destination as they get within its communication range. This approach relies on the diversity

and mobility to improve the throughput, at the cost of increased delay and energy consumption.

When the sinks or access points are mobile, like in SENMA, the capacity of a node having a direct communication with the sink could be significantly superior to that of ad-hoc sensor networks [17]. However, since the sources wait for an access point's visit, the throughput is limited by the traversal speed and trajectory length of the access point. In [19], the throughput of a network with mobile sinks is investigated, and the effect of the speed of the sink and its trajectory length on the throughput is highlighted.

- *Delay* The delay is the time consumed until the sensing information is delivered to the destination/sink. It is composed of: transmission delay, processing delay, queuing delay, and data propagation delay along the routing path(s) from the source to its destination. The transmission delay depends on the data rate and the packet size; the processing delay depends on the processing speed in the nodes; the queuing delay depends on the traffic pattern characterizing the buffer dynamics at the nodes; the wireless medium propagation delay depends on the distance between the transmitting and receiving ends as well as the EM wave speed (speed of light).

  For time-sensitive applications, it is crucial to design a network that achieves the required delay performance. In order to guarantee the delay requirement of data transmission over a network, sufficient network control is essential. Proper architecture, topology design, and transmission protocols are needed to assist in balancing the network traffic load and in guaranteeing time-sensitive information exchange.

- *Energy Efficiency and Network Lifetime* Due to the limited energy supply of

sensor nodes, energy efficiency is a fundamental concern in wireless sensor networks. Hence, it is essential to design energy-efficient architectures and protocols to prolong the network lifetime.

The transceiver module is the primary source of energy consumption. Therefore, the routing functions highly contribute to the nodes' power depletion, where each node does not only transmit its packets, but also transmits and receives other nodes' packets to deliver them to the intended destination. Note that SENMA architecture significantly improves the energy-efficiency compared to random ad-hoc networks, since sensors are relieved from the energy-consuming routing functions [17].

Energy consumption determines the network lifetime. There are several definitions for the network lifetime in the literature. In [29], the lifetime is defined as the time until any sensor in the network depletes its energy, or a fraction of the sensors deplete their energy and become nonoperational. In [30], the lifetime is defined as the time until full coverage of the network is no longer provided.

1.3 Reliability and Security

The network should be fast-reacting and self-healing not only to node/link failure conditions that could result from node's power depletion and rough environmental conditions, but also to malicious attacks. This section presents some threats to sensor network operations, and briefly discusses possible countermeasures.

1.3.1 Possible Attacks

Wireless sensor networks are vulnerable to various types of malicious attacks that can severely disrupt the network performance. The malicious attacks in wireless sensor

networks can be classified into two categories:

1. ***Internal attacks***, which are launched by authorized nodes that are compromised by an external intruder. An internal attacker can perform arbitrary behavior to harm the network, such as:

    - Send fictitious sensing reports to the central processing unit or the sink. This is known as *Byzantine attack* [31].
    - Claim different nodes' identities. This is known as *Sybil attack* [32].
    - Launch routing attacks, which include dropping and/or modifying packets as they are being relayed to the destinations through multihop transmissions. For example, a serious threat to WSN is the known as *Sinkhole attack*, where the malicious nodes attract traffic by modifying their routing metrics, then perform selective forwarding, drop, or modify packets [33].
    - Intentionally collide with benign nodes' traffic, resulting in retransmissions and power exhaustion of benign nodes. This attack can be regarded as a special type of *jamming*.

2. ***External attacks***, which are launched by non-authenticated nodes that intentionally intercept and/or disrupt the network operation. External attacks can be further divided into two categories:

    - Passive attacks, such as eavesdropping, traffic interception and analysis.
    - Active attacks, such as jamming, where the attackers transmit high power signals to interfere with the network traffic.

1.3.2 Existing Techniques for Malicious Attacks Mitigation

Internal attacks are more difficult to detect and counteract as compared to external attacks. Cryptographic algorithms, such as authentication and encryption, can counteract most of the external attacks, with exception to jamming [34]. In the following, possible countermeasures that combat some malicious attacks are discussed.

For Byzantine attacks: Recall that in Byzantine attacks some authenticated nodes send false sensing reports to the processing unit or the sink. Byzantine attacks cannot be detected through cryptographic approaches. One way to combat these attacks is through distributed detection, where many nodes collaborate to sense the same phenomenon or target and report their measurements to a central processing unit for data fusion. The data fusion rule should be well designed to ensure reliable decision-making, even if some of the received information is fictitious. We consider distributed detection in SENMA networks under Byzantine attacks in Chapter 2.

For Sybil attacks: A malicious node launching Sybil attack can claim different identities to confuse the processing unit. A node-to-node authentication with location-based keys is proposed in [35] to combat Sybil attacks. In this location-based authentication approach, the keys used in the authentication process are based on the nodes' identifications (IDs) as well as their location. Hence, a node launching Sybil attack would be identified.

For Sinkhole attacks: In a Sinkhole attack, the malicious nodes attract network traffic by modifying their routing metrics, then drop or modify the information as it is being routed to the sink. Location-based authentication can mitigate Sinkhole attacks only if the routing metric is based on the location, and not on the remaining energy or link reliability [35]. Like Byzantine attacks, generally Sinkhole attacks could not be detected through cryptographic means solely. In [36], Sinkhole attacks are detected through monitoring the CPU usage of the nodes; This scheme requires each node to

periodically report its actual CPU usage to a centralized entity, which in turn could identify the malicious nodes. Mitigating Sinkhole attacks through evaluating a trust metric for each node is discussed in [37], where both entropy-based and probability-based trust models are presented for trust evaluation.

For eavesdropping and traffic analysis: Incorporating information encryption in WSNs is essential to thwart eavesdropping and traffic analysis or render them ineffective. The encryption process is classified into two categories, namely, symmetric-key encryption and asymmetric-key (or public key) encryption. In symmetric-key encryption, the same key is used for encryption at the transmitter and decryption at the receiver. In asymmetric-key encryption, the encryption and decryption processes are preformed using different keys.

The operations performed in symmetric-key encryption algorithms are generally less complex than that in asymmetric encryption algorithms. Thus, symmetric-key cryptography is often considered in sensor networks [38]. However, it is noted that symmetric encryption strategies require key distribution and management mechanisms to allow the transmitter and the receiver to securely share the secrete key and periodically update it. A study in [39] demonstrated that symmetric encryption can be easily implemented on MICA2 sensors, and asymmetric encryption could be tractable.

For jamming: Jamming mitigation can be realized through frequency hopping and spread spectrum approaches [40]. Due to the limited number of frequencies that a sensor can operate in, jamming mitigation through spread spectrum techniques could be inapplicable in sensor networks [38]. Other approaches for jamming mitigation include traffic surfing [41], where nodes adapt their communication channel based on the interference. Two approaches are considered in [41]: In the first approach, all the sensors in the network change their channels when high interference (jamming) is detected; in the second approach, only nodes in the interference region choose a

different communication channel. These approaches may work well under fixed-pattern partial-band jamming, but they still face significant challenges under random jamming.

1.4 Major Contributions of the Dissertation

This dissertation aims to improve the reliability and efficiency of time-critical communications in WSNs, under both benign and hostile environments.

In Chapter 2, we explore reliable distributed detection in mobile access wireless sensor networks under Byzantine attacks [42]. We consider the q-out-of-m fusion rule, which is popular in distributed detection and can achieve a good trade-off between the miss detection probability and the false alarm rate. However, a major limitation with it is that the optimal scheme parameters can only be obtained through exhaustive search, making it infeasible for large networks. In this chapter, first, by exploiting the linear relationship between the scheme parameters and the network size, we propose simple but effective sub-optimal linear approaches. Second, for better flexibility and scalability, we derive a near-optimal closed-form solution based on the central limit theorem. Third, subjecting to a miss detection constraint, we prove that the false alarm rate of q-out-of-m diminishes exponentially as the network size increases, even if the percentage of malicious nodes remains fixed. Finally, we propose an effective malicious node detection scheme for adaptive data fusion under time-varying attacks; the proposed scheme is analyzed using the entropy-based trust model, and has shown to be optimal from the information theory point of view. Simulation examples are provided to illustrate the performance of proposed approaches under both static and dynamic attacks.

In Chapter 3, we propose a novel mobile access coordinated wireless sensor network (MC-WSN) architecture for reliable, efficient, and time-sensitive information exchange.

In conventional sensor networks with mobile access points (SENMA), the mobile access (MA) points traverse the network to collect information directly from individual sensors. While simplifying the routing process, a major limitation with SENMA is that a transmission is made only if an MA visits the corresponding source node; thus, data transmission is limited by the physical speed of the MAs and the length of their trajectory, resulting in low throughput and large delay. The proposed MC-WSN effectively resolves this problem through hop number control [43]. More specifically, with active network development and topology design, the number of hops from any sensor to a mobile access can be limited to a pre-specified number. We discuss the optimal topology design for MC-WSN such that the average number of hops between the source and its nearest sink is minimized, and analyze the performance of MC-WSN in terms of throughput, stability, delay, and energy efficiency by exploiting tools in information theory, queuing theory, and radio energy dissipation model. It is shown that under stable system conditions, MC-WSN achieves much higher throughput and significantly lower delay and energy consumption than that of SENMA.

In Chapter 4, we provide a unified framework for quantitative characterization of various wireless networks [44]. We first revisit the evolution of centralized, ad-hoc and hybrid networks, and discuss the trade-off between structure-ensured reliability and efficiency, and ad-hoc enabled flexibility. Motivated by the observation that the number of hops for a basic node in the network to reach the base station or the sink has a direct impact on the network throughput, delay, efficiency and their evaluation techniques, we introduce the concept of the N-hop networks. It can serve as a general framework that includes most existing network models as special cases, and can also make the analytical characterization of the network performance more tractable. Moreover, for the network security, it is observed that hierarchical structure enables easier tracking of user accountability and malicious node detection; on the other hand, the multi-layer

diversity increases the network reliability under unexpected network failure or malicious attacks, and at the same time provides a flexible platform for privacy protection.

In Chapter 5, we summarize the conclusions and present some proposed directions for future research.

Chapter 2

Distributed Detection in Mobile Access Wireless Sensor Networks Under Byzantine Attacks

This chapter explores reliable data fusion in mobile access wireless sensor networks under Byzantine attacks. We consider the q-out-of-m rule, which is popular in distributed detection and can achieve a good trade-off between the miss detection probability and the false alarm rate. However, a major limitation with this rule is that the optimal scheme parameters can only be obtained through exhaustive search, making it infeasible for large networks. In this chapter, first, by exploiting the linear relationship between the scheme parameters and the network size, we propose simple but effective sub-optimal linear approaches. Second, for better flexibility and scalability, we derive a near-optimal closed-form solution based on the central limit theorem. Third, subjecting to a miss detection constraint, we prove that the false alarm rate of the q-out-of-m

rule diminishes exponentially as the network size increases, even if the percentage of malicious nodes remains fixed. Finally, we propose an effective malicious node detection scheme for adaptive data fusion under time-varying attacks; the proposed scheme is analyzed using the entropy-based trust model, and has shown to be optimal from the information theory point of view. Simulation examples are provided to illustrate the performance of proposed approaches under both static and dynamic attacks.

2.1 Introduction

Wireless sensor networks have received significant attention from the research community due to their impact on both military and civilian applications [1,2,47,48]. Limited by the processing capability and power supply of the sensor nodes, incorporating security into wireless sensor networks has been a challenging task [38, 49–53]. A serious threat to wireless sensor networks is the Byzantine attack [31, 54–59], where the adversary has full control over some of the authenticated nodes and can perform arbitrary behavior to disrupt the system.

Distributed detection under malicious attacks has been studied from different perspectives in the literature. In [60], an iterative redundancy approach was proposed to combat the Byzantine failure problems in distributed computation architectures. In this approach, the system chooses a number of nodes to perform the computation in order to reach the required reliability level. If consensus cannot be reached among the selected nodes, then more nodes are recruited for the computation. The algorithm is repeated until the required reliability is achieved. A similar approach use majority voting based on iterative group message exchange can be found in [61]. In [62], credibility-based fault-tolerance is discussed for the volunteer computing systems, where both majority voting and spot-checking scheme are integrated. In spot-checking, training

data is used to detect the malicious machines. The server continues the voting and spot-checking processes until the overall credibility level is reached. While the approaches in [60–62] may work well for ordinary distributed systems, the complexity and delay caused by the iterative processes could be an unaffordable burden for wireless sensor networks (WSNs). In [63], fault-tolerant data acquisition is discussed for clustered sensor networks, in which data fusion is first performed in each cluster to obtain local estimates, and then performed at a higher level to obtain a global estimate from the local estimates provided by each cluster head. Weighted average is utilized for data aggregation, where the weighting factor of each sensor is determined using spatial correlation among neighboring sensors. One possible limitation with this approach is that degraded system performance may occur when a group of neighboring sensors collaborate to send fictitious data. In [64], the median approach is proposed for distributed detection in clustered WSN, under the assumption that the data is always sent through binary symmetric channels, and all nodes have the same, arbitrary-low probability of sending faulty results. The latter assumption could be too ideal, since attackers could send false sensing information with high probabilities.

In this work, we consider reliable data fusion in wireless sensor networks with mobile access points (SENMA) [17] under both static and dynamic Byzantine attacks, in which the malicious nodes report false information with a fixed or time-varying probability, respectively. In SENMA, the mobile access point traverses the network and collects the sensing information from the individual sensor nodes. The major advantage of the SENMA architecture is that it ensures a line of sight path to the access point within the power range of the sensor nodes, allowing the information to be conveyed without routing. This feature makes it a resilient, scalable and energy efficient architecture for wireless sensor networks. In many cases, due to bandwidth and energy limitations, the sensors quantize their sensing result into a single bit [65–68]. The MA receives the

sensing reports and applies the fusion rule to make the final decision. One popular hard fusion rule used in distributed detection is the q-out-of-m scheme [66, 67, 69], in which the mobile access point randomly polls reports from m sensors, then decides that the target is present only if q or more out of the m polled sensors report '1'. It is simple to implement, and can achieve a good trade-off between minimizing the miss detection probability and the false alarm rate. In ideal scenarios, the optimal scheme parameters for the q-out-of-m fusion scheme are obtained through exhaustive search. However, due to its high computational complexity, the optimal q-out-of-m scheme is infeasible as the network size increases and/or the attack behavior changes. To overcome this limitation, effective sub-optimal schemes with low computational complexity are highly desired.

The main contributions in this chapter can be summarized as follows: First, we propose a simplified, linear q-out-of-m scheme that can be easily applied to large size networks. The basic idea is to find the optimal scheme parameters at relatively small network sizes through exhaustive search, and then obtain the fusion parameters for large network size by exploiting the approximately linear relationship between the scheme parameters and the network size. It is observed that the proposed linear approach can achieve satisfying accuracy with low false alarm rate. However, there are chances of violating the problem constraint. To enforce the miss detection constraint and improve the data fusion accuracy, we further propose to use the linear approximation as the initial point for the optimal exhaustive search algorithm. With this enhanced linear approach, near-optimal solutions can be obtained with much lower computational complexity compared with that of the pure exhaustive search approach.

Second, in an effort to search for an easier and more flexible distributed data fusion solutions that can easily adapt to unpredictable environmental changes and cognitive behavior of malicious nodes, we derive a closed-form solution for the q-out-of-m fusion

scheme based on the central limit theorem. It is observed that the closed-form solution is a function of the network size, the percentage of malicious sensors, the malicious nodes' behavior, and the detection accuracy of the sensor nodes. We show that the closed-form solution delivers comparable results with that of the near-optimal solution obtained from the enhanced linear approach.

Third, we perform theoretical analysis for both the linear approach and the closed-form solution. We show that under a fixed percentage of malicious nodes, the false alarm rate for both approaches diminishes exponentially as the network size increases. This analysis reveals an interesting and important result: *even if the percentage of malicious nodes remains unchanged, larger size networks are much more reliable under malicious attacks.* It indicates that the network size plays a critical role in reliable data fusion. Moreover, we also find an upper bound on the percentage of malicious nodes that can be tolerated by the network under the q-out-of-m fusion rule. It turns out that this upper bound is determined by the sensors' detection probability and the attack strategies of the malicious nodes.

Finally, we propose a simple and effective malicious node detection approach, where the malicious sensors are identified by comparing the decisions of the individual sensors with that of the fusion center. It is observed that dynamic attacks generally take longer time and more complex procedures to be detected as compared to static attacks. It is also found that the proposed malicious detection procedure can identify malicious sensors accurately if sufficient observation time is allowed. The proposed approach is analyzed using an entropy-based trust model. We show that under the same system settings, the proposed malicious node detection approach is optimal from the information theory point of view. We further propose to adapt the fusion parameters based on the detected malicious sensors and their estimated probability of attack. It is shown that the proposed adaptive fusion scheme can improve the system performance

significantly under both static and dynamic attack strategies.

2.2 Problem Formulation

2.2.1 Overall System Set-Up

We consider a centralized sensor network architecture, known as SENMA [70], where we assume that the network is composed of n power-limited sensor nodes and a powerful mobile access point. The architecture is illustrated in Figure 2.1. We assume that the nodes are randomly and uniformly distributed over the network, and the mobile access point traverses the network on a predefined trajectory to communicate with all the sensing nodes. The sensor network performs distributed detection. Each sensor node detects the presence of the target object by applying an application-dependent detection algorithm, such as energy detection [71], and sends its one-bit hard decision report to the mobile access point ('1' means that the target is present), which makes the final decision accordingly. This hard decision model is adopted here for two reason: (1) To reduce the transmission and processing burden of the sensor network; (2) To enable more tractable analysis on the effect of the network size on the reliability of the distributed detection under Byzantine attacks.

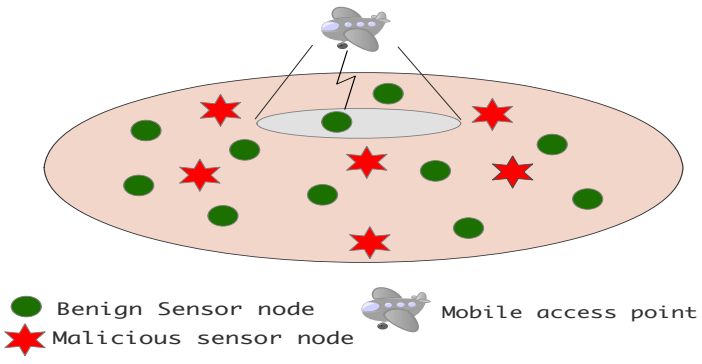


Figure 2.1: SENMA under Byzantine attack.

If the network covers a large area, we divide the area into smaller sections, and apply the fusion rule over nodes that are within the same section. This setting ensures that, statistically, nodes within the same section have the same chance of detecting the target.

We assume that the network contains k malicious sensors. The percentage of malicious sensors, k/n, is denoted by α, which is assumed to be known or can be estimated at the mobile access point (MA). When no prior knowledge of α exists, the MA would start with a majority vote[1]and obtain an estimate for α by comparing the individual sensing reports with the final decision. We assume that each benign sensor node has a false alarm probability[2] P_f and a miss detection probability[3] P_m, while the malicious sensors have their own false alarm probability $\tilde{P}_f$ and miss detection probability $\tilde{P}_m$. These probabilities are determined by the environmental conditions and the sensors' capabilities to detect the target.

The MA uses the binary reports of the sensor nodes to make the final decision on whether the target is present or absent. This distributed detection problem can be modeled using the conventional binary hypothesis test, where the hypothesis H_0 represents the absence of the target, and the hypothesis H_1 represents the presence of the target.

Here, we will first discuss different attack strategies that can be adopted by the malicious sensors, then present the problem formulation based on the q-out-of-m scheme.

[1]In majority vote if more than half of the sensors reported '1', then the final decision is '1'.

[2]The false alarm rate is the conditional probability that the target is said to be present, when it is not.

[3]The miss detection probability is the conditional probability that the target is said to be absent, when it is present.

2.2.2 Modeling of Possible Attack Strategies

There are different attack strategies that could be adopted by the malicious sensors. Let P_o be the probability that each malicious node intentionally reports the opposite information to its actual sensing decision. It is assumed that all malicious nodes have the same probability of attack in a particular sensing period. We classify the possible attack strategies into two categories:

1. *Static Attack:* In this strategy, the malicious nodes send opposite data with an arbitrary probability P_o that is fixed, with $0 < P_o \leq 1$.

2. *Dynamic Attack:* In this strategy, the malicious nodes change P_o after each attacking block, which is composed of one or more sensing periods. More specifically,

$$P_{o_i} = P_{o_{i-1}} + \Delta_1 x - \Delta_2(1-x), \tag{2.1}$$

   where P_{o_i} is the value of P_o in the nth attacking block, x is a Bernoulli random variable that is equal to '1' with probability P_x, Δ_1 and Δ_2 are the increment and decrement step size, respectively.

Taking the malicious nodes' own false alarm and miss detection probabilities into consideration, it turns out that the malicious sensors may have different probabilities of attack when the target is present and when the target is absent. We refer to them as the miss detection attack probability ($P_{a,m}$) and false alarm attack probability ($P_{a,f}$). More specifically, the overall miss detection attack probability is given by: $P_{a,m} = \bar{P_o}(1-\tilde{P}_m) + (1-\bar{P_o})\tilde{P}_m$, where $\bar{P_o} = P_o$ for static attacks, and equals to the average P_{o_i} over all attacking blocks for dynamic attacks. The false alarm attack probability is given by: $P_{a,f} = \bar{P_o}(1-\tilde{P}_f) + (1-\bar{P_o})\tilde{P}_f$. We define P_a as the overall attack probability of malicious sensors. If the state of nature is equal to '1' with probability p,

then $P_a = pP_{a,m} + (1-p)P_{a,f}$. In the special case when the sensor nodes can perfectly detect the state of nature, i.e., $\tilde{P}_f = 0$ and $\tilde{P}_m = 0$, then $P_a = P_{a,m} = P_{a,f} = \bar{P}_o$.

2.2.3 Problem Formulation

For reliable data fusion in SENMA under Byzantine attacks, we propose to use the q-out-of-m fusion rule, in which the MA randomly polls m out of n reports, then decides that the target is present (H_1) only if q or more out of the m polled sensors report '1'. The main reason is that other hard fusion rules, such as OR[4], AND[5], or the majority voting rule, might not achieve the compromise between minimizing the false alarm rate and the miss detection probability [66], especially under malicious attacks. Moreover, the q-out-of-m scheme parameters can be adapted based on the attacking behavior and percentage of malicious sensors. This inherit flexibility makes the q-out-of-m scheme superior to other hard fusion rules.

We aim to obtain m and q to minimize the overall false alarm rate Q_f, while keeping the overall miss detection rate Q_m below a certain predefined value β. That is, our objective is to find the optimal m and q that can minimize Q_f, subject to the constraint $Q_m \leq \beta$. The problem can be formulated as follows:

$$\begin{aligned} &\min_{m,q} Q_f(m,q)m \\ &s.t. \;\; Q_m(m,q) \leq \beta, \;\; 1 \leq q \leq m \leq n, \;\; q, m \in \mathbb{N}. \end{aligned} \tag{2.2}$$

It should be pointed out that there is always a trade-off between minimizing the false alarm rate and the miss detection probability, therefore the parameter q should not be too small nor too large. Large q can improve the false alarm rate, but would increase

[4]OR rule: if at least one sensor reports '1', then the decision is '1'; otherwise, the decision is '0'.

[5]AND rule: if all sensors report '1', then the decision is '1'; otherwise, the decision is '0'.

the miss detection probability. Small q can achieve a higher detection probability, but would increase the false alarm rate.

Define $P_{k,n-k}^{d,m-d}$ as the probability of polling $m-d$ out of $n-k$ benign sensors and d out of k malicious sensors. That is, $P_{k,n-k}^{d,m-d} = \frac{\binom{k}{d}\binom{n-k}{m-d}}{\binom{n}{m}}$. According to our system model, the overall false alarm rate, Q_f, can be expressed as:

$$Q_f = \sum_{\substack{d=\max(0,\\ m+k-n)}}^{k} P_{k,n-k}^{d,m-d} \sum_{c=0}^{d} \binom{d}{c} P_{a,f}^{c} (1-P_{a,f})^{(d-c)} \times \sum_{j=\max(0,q-c)}^{m-d} \binom{m-d}{j} P_f^j (1-P_f)^{(m-d-j)}. \tag{2.3}$$

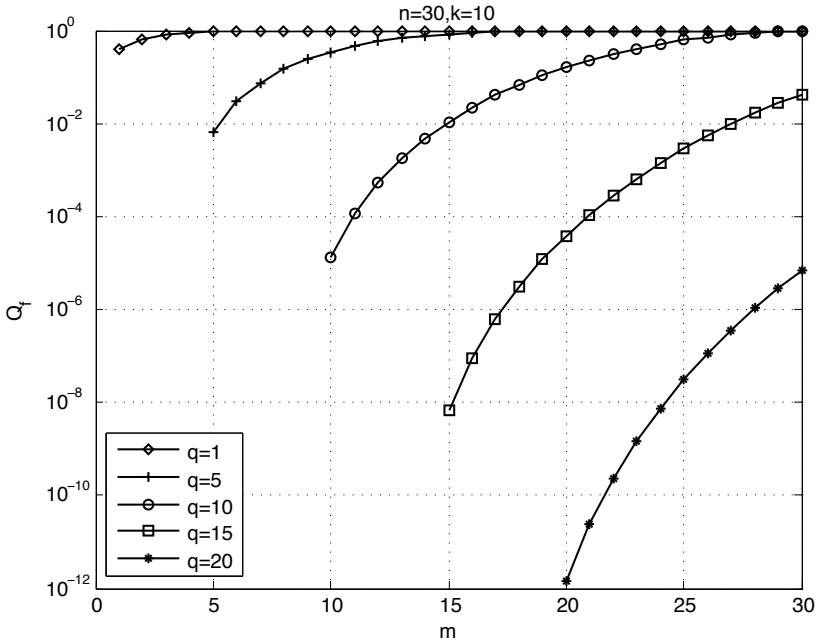
If the m polled sensors contain d out of the k malicious nodes, then the false alarm occurs when c or more out of d malicious sensors attack and $q-c$ or more benign sensors send false alarms, where $0 \le c \le d$. It is noted that the minimum number of malicious reports being polled is $d = \max(0, m+k-n)$. That is, when the number of sensors polled, m, is greater than the number of benign sensors $(n-k)$, then there are at least $m-(n-k)$ malicious reports received by the MA. The overall probability of detection Q_d can be expressed as:

$$Q_d = \sum_{\substack{d=\max(0,\\ m+k-n)}}^{k} P_{k,n-k}^{d,m-d} \sum_{c=0}^{d} \binom{d}{c} (1-P_{a,m})^{c} P_{a,m}^{(d-c)} \times \sum_{j=\max(0,q-c)}^{m-d} \binom{m-d}{j} P_d^j (1-P_d)^{(m-d-j)}, \tag{2.4}$$

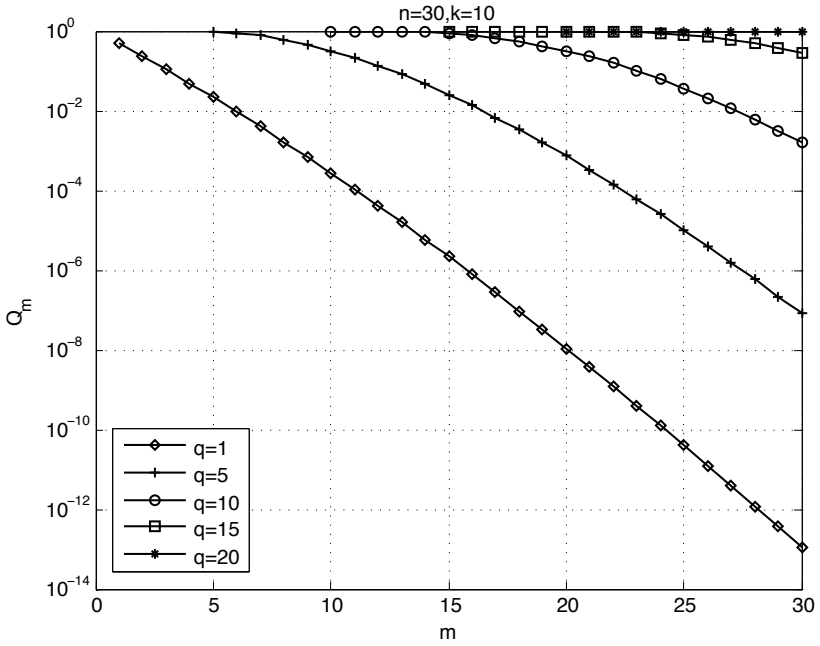
where $P_d = 1 - P_m$ is the detection probability of the benign nodes, when the target is present. The overall miss detection probability Q_m is then obtained by $Q_m = 1 - Q_d$.

Intuitively, if q is greater than the number of benign sensors, i.e., $q > n - k$, and the malicious nodes have high miss detection attack probability ($P_{a,m} \to 1$), then the overall miss detection probability will be very high ($Q_m \to 1$). Thus, q *should not be too large*. On the other hand, if q is less than the number of malicious sensors, i.e., $q < d$, and the malicious nodes have high false alarm attack probability ($P_{a,f} \to 1$), then $Q_f \to 1$. Thus, q *should not be too small*. The trade-off between minimizing the false alarm rate and the miss detection probability is further illustrated in Figures 2.2(a) and 2.2(b), where Q_f and Q_m are obtained for different values of m and q. It is shown that lower q improve the miss detection probability, but degrade the false alarm rate; and, higher q degrades the miss detection probability, but improves the false alarm rate. Note that OR rule corresponds to the case when $m = n$ and $q = 1$. It is clear from Figure 2.2(a) that OR rule results in a very high false alarm rate under Byzantine attack.

It is noted that finding the optimal m and q from (2.2) is a nonlinear integer optimization problem that is hard to be solved theoretically. The optimal approach is to perform exhaustive search over all possible m and q values, and then choose the (m_o, q_o) pair that results in the lowest false alarm rate while satisfying the miss detection constraint. The computational complexity of the optimal exhaustive search is $O(n^2)$ [72], which would be infeasible for real-time data fusion in large networks. Therefore, we aim at finding simpler but accurate methods to obtain the scheme parameters that solve (2.2).



(a) The false alarm rate.



(b) The miss detection probability.

Figure 2.2: The false alarm rate and miss detection probability when $n = 30$, $k = 10$, $P_{a,f} = P_{a,m} = 1$, $P_f = 0.1$, and $P_d = 0.775$.

2.3 Simplified Data Fusion Scheme - The Linear Approach

In this section, we will first highlight some observations based on the optimal q-out-of-m scheme, and then present the simplified algorithms that can be easily applied to large-scale networks.

2.3.1 Observations

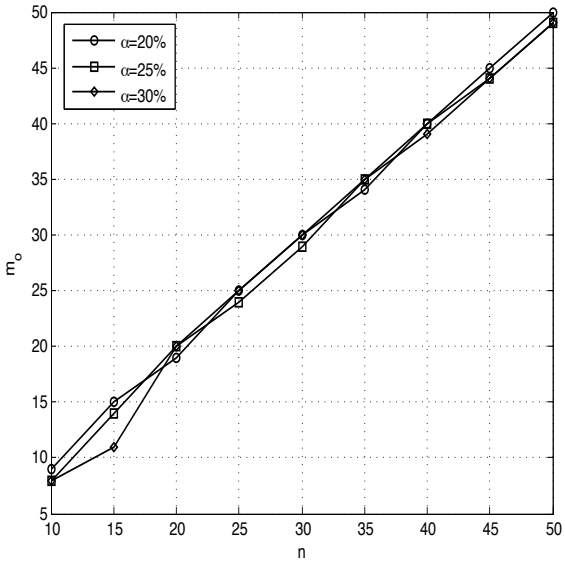
To develop effective sub-optimal schemes with low computational complexity, it is important to know how the parameters m and q change with the system variables, such as α and n. In this section, we consider the case where the malicious sensors attack with probability P_a. We calculate the optimal parameters at different P_a values, under different network sizes and different percentages of malicious sensors. The following observations are made [45]:

Observation 1: The optimal m is almost independent of the percentage of malicious nodes, and has a linear relationship with n. In fact, it is always equal to or very close to n, as shown in Figures 2.3(a) and 2.3(c), which implies that the reports of almost all the sensors should be considered in the optimal q-out-of-m fusion scheme[6]. This observation enables us to reduce the problem to finding the best q when $m = n$, which lowers the computational complexity from $O(n^2)$ to $O(n)$.

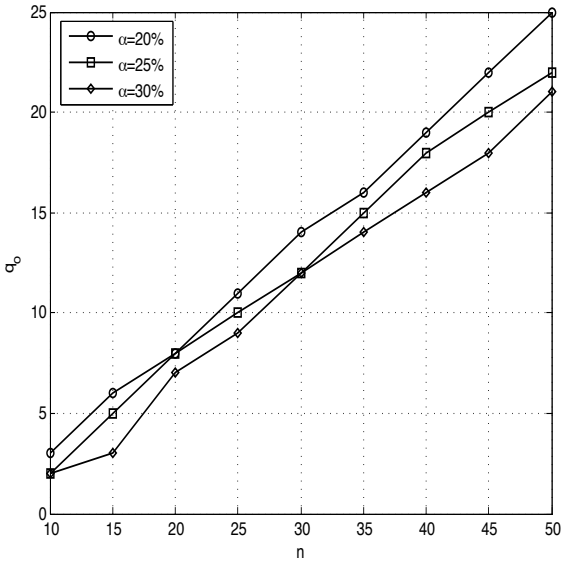
Observation 2: The optimal value of q follows an approximately linear function of n with different slopes depending on the percentage of the malicious nodes and the probability of attack, as shown in Figures 2.3(b) and 2.3(d).

Motivated by these observations, we develop simplified approaches to obtain the

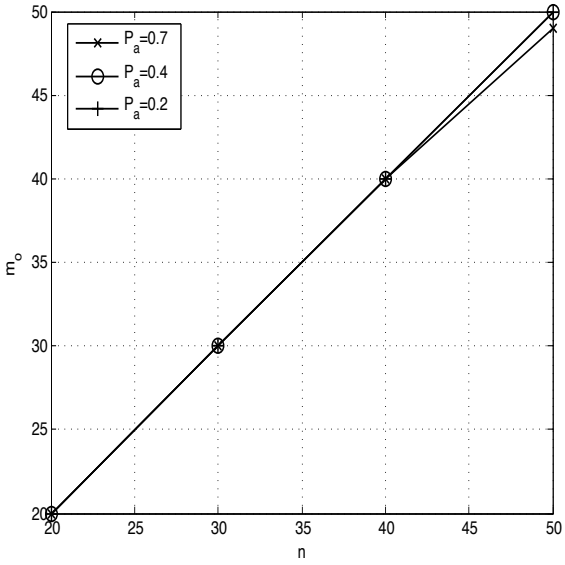
[6]However, this is no longer the case when malicious node detection scheme is employed, as the reports of the malicious sensors would be discarded. Malicious detection will be considered in section 2.6.



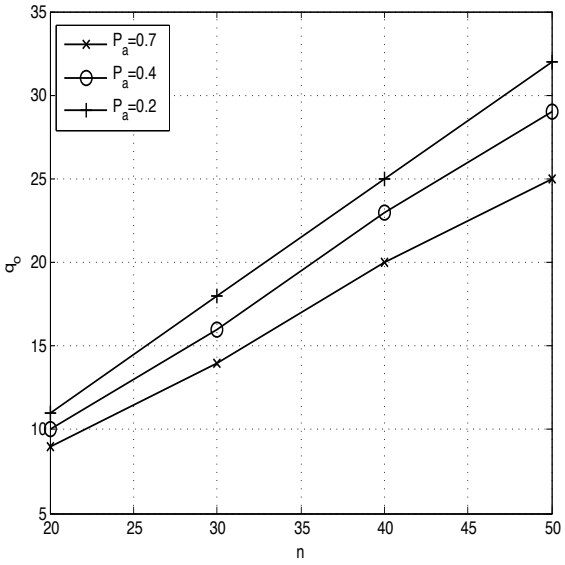
(a) Optimal m vs. n at different α, when $P_a = 1$.



(b) Optimal q vs. n at different α, when $P_a = 1$.



(c) Optimal m vs. n at different P_a, when $\alpha = 25\%$.



(d) Optimal q vs. n at different P_a, when $\alpha = 25\%$.

Figure 2.3: Optimal scheme parameters (m_o, q_o) versus the network size at different percentage of malicious nodes (α) and different probability of attack (P_a), when $\beta = 0.01$, $P_f = 0.1$, $P_d = 0.775$, $P_{a,m} = P_{a,f}$.

q-out-of-m fusion parameters with low complexity that can be easily applied to large network sizes.

2.3.2 The Linear Approach

In this section, we propose a simplified q-out-of-m scheme by exploiting the linear relationship between the scheme parameters and the network size. The main idea is that we can get the optimal scheme parameters at relatively small network sizes, and use them as reference points. These optimal (m, q) pairs for the different network sizes, P_a values, and α ratios, can be obtained and stored in a look-up table, then used to get the suboptimal scheme parameters for large network sizes. We propose to set $m = n$ and use the following linear function of n to obtain q [45]:

$$\hat{q}_{n,\alpha} = \lceil q_{n_0,\alpha} + S_o(\alpha)(n - n_0) \rceil, \tag{2.5}$$

where $S_o(\alpha)$ is the slope of the optimal q_o versus n curve at a particular attack probability given that the percentage of the malicious nodes is α, $\hat{q}_{n,\alpha}$ is the suboptimal q value at a network size n, and $q_{n_o,\alpha}$ is the optimal q value at a relatively small network size n_0 and it serves as a reference point. Both $\hat{q}_{n,\alpha}$ and $q_{n_o,\alpha}$ are at α percent of malicious sensors. $\lceil x \rceil$ is the smallest integer larger than or equal to x. Note that the optimal q depends on the false alarm and miss detection probabilities of the sensor nodes, hence a periodical update of the reference points and related slopes would be required in time-varying environments.

While the linear approach can deliver very good performance, there are chances of violating the problem constraint. Therefore, we propose to enhance the linear approach to guarantee that the best choice of q obtained satisfies the miss detection probability constraint.

2.3.3 Enhanced Linear Approach

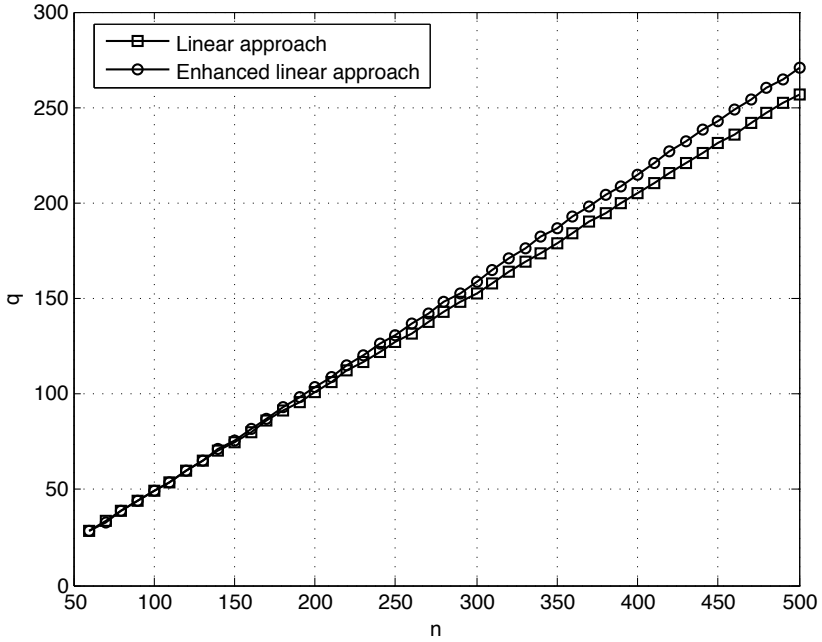
To ensure that the scheme parameter obtained using the linear approximation, $\hat{q}_{n,\alpha}$, results in the lowest false alarm rate and satisfies the miss detection constraint in (2.2), the linear approach is enhanced using an iterative method to find $\hat{q}_{n,\alpha}$[7]. The algorithm works as follows:

1. Set $m = n$ and use the linear approximation in (2.5) as an initial value for $\hat{q}_{n,\alpha}$.
2. Calculate the miss detection probability using (2.4).
3. Increase $\hat{q}_{n,\alpha}$ to $\hat{q}_{n,\alpha} + 1$ if Q_m is below the predefined β. Then, go to step 2.
4. Decrease $\hat{q}_{n,\alpha}$ to $\hat{q}_{n,\alpha} - 1$ if Q_m is above the predefined β. Then, go to step 2.
5. Terminate the iterations when the largest $\hat{q}_{n,\alpha}$ that meets the miss detection constraint is obtained.

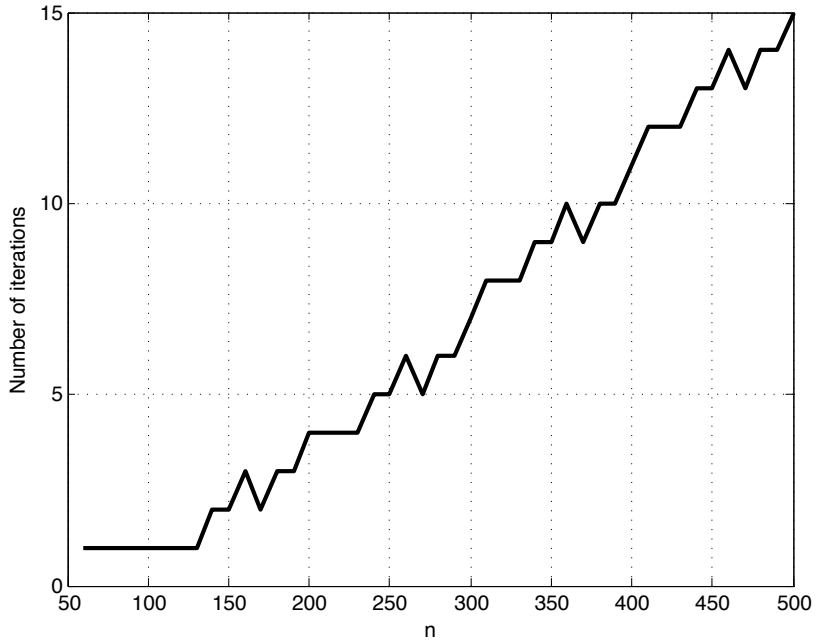
Note that higher values of q lower the false alarm rate. The approach above obtains the largest q that satisfies the miss detection constraint, hence provides a near-optimal solution.

The sub-optimal q values ($\hat{q}_{n,\alpha}$), obtained using the linear and the enhanced linear approaches, are shown in Figure 2.4(a). It is shown that the linear relationship between q and n is also valid for the enhanced linear approach. In Figure 2.4(b), the number of iterations required in the enhanced linear approach is plotted versus the network size. It is shown that the enhanced approach converges after few iterations, and thus it is computationally less intensive than the optimal exhaustive search.

[7]Reprinted from M. Abdelhakim, L. Lightfoot, J. Ren, and T. Li, "Reliable Cooperative Sensing in Cognitive Networks", Wireless Algorithms, Systems, and Applications, WASA'12, Springer Berlin Heidelberg, vol. 7405, pp. 206 - 217 [73], with kind permission from Springer Science and Business Media.



(a) q vs. network size.



(b) Iterations of the enhanced linear approach.

Figure 2.4: The enhanced linear approach at $\alpha = 25\%$.

While the linear approach works quite well, the absence of a well-defined closed-form solution makes it difficult to adapt q based on the environmental conditions and the malicious behavior. To find q for different network settings, the slopes and the reference points should always be updated using exhaustive search. This could be tedious when the environment is fast-varying. To solve this problem, in the following section, we derive a closed-form expression for q.

2.4 A Closed-form Solution

In this section, we derive a closed-form solution of q for the q-out-of-m fusion rule under both static and dynamic attacks. We exploit the observations of the optimal exhaustive search by setting $m = n$, as illustrated in the previous section.

Recall that the malicious sensors have miss detection and false alarm attack probabilities, $P_{a,m}$ and $P_{a,f}$, respectively. For notation simplicity, we assume that these two probabilities are equal, that is, $P_{a,m} = P_{a,f} = P_a$. It is worth mentioning that the analysis can be easily extended to the case when $P_{a,m} \neq P_{a,f}$. We assume that all sensing reports are independent. It is noted that the distribution of each sensing report is determined by the environment and the behavior of the corresponding sensor node. Let the sensing report of node i be $u_i \epsilon \{0, 1\}$, where $i = \{1, ..., n\}$. If node i is benign, then u_i is a Bernoulli random variable characterized by detection probability P_d if the target is present, or the false alarm rate P_f if the target is absent; if node i is malicious, then u_i is a Bernoulli random variable characterized by the parameter $1 - P_a$ if the target is present, or P_a if the target is absent.

The aggregated result at the MA is given by, $U = \sum_{i=1}^{n} u_i$ The random variable U represents the number of 1's that the access point received. To apply the q-out-of-m fusion rule, U is compared to q. If $U \geq q$, the final decision is that the target is present

(i.e., decide H_1); otherwise, the final decision is that the target is absent (i.e., decide H_0).

Our closed-form solution is based on the central limit theorem, where the aggregated result at the access point is approximated as a Gaussian random variable. In fact, we have the following result:

Proposition 2.1 *Suppose a network of size n containing both benign sensors and malicious sensors, where the percentage of malicious sensors is α. The benign sensors have a detection probability P_d, and the malicious sensors attack with a probability P_a. Assuming the q-out-of-m fusion rule is applied subject to a predefined miss detection constraint β, then the lowest false alarm rate can be achieved when $q = \lfloor an + \sqrt{bn}Q^{-1}(1-\beta)\rfloor$. Here, $a = (1-\alpha)\,P_d + \alpha(1-P_a)$, $b = (1-\alpha)\,P_d(1-P_d) + \alpha(1-P_a)P_a$, and $Q^{-1}(.)$ is the inverse Q function*[8].

Proof: Note that U is the summation of independent random variables. If the number of sensors n is large, then U can be approximated as a Gaussian random variable. When the target is present (H_1), $U \sim \mathcal{N}(M_{u,p}, V_{u,p})$, where $M_{u,p}$ and $V_{u,p}$ can be found by summing up the means/variances of the $(n-k)$ benign nodes and the k malicious nodes, respectively. More specifically,

$$\begin{aligned} M_{u,p} &= (n-k)P_d + k(1-P_a) \\ &= [(1-\alpha)\,P_d + \alpha(1-P_a)]\,n, \qquad (2.6) \\ V_{u,p} &= (n-k)P_d(1-P_d) + k(1-P_a)P_a \\ &= [(1-\alpha)\,P_d(1-P_d) + \alpha(1-P_a)P_a]\,n. \qquad (2.7) \end{aligned}$$

Define $a = (1-\alpha)\,P_d + \alpha(1-P_a)$ and $b = (1-\alpha)\,P_d(1-P_d) + \alpha(1-P_a)P_a$, we get

[8]The Q function is defined as: $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{x^2}{2}}\,dx$.

$M_{u,p} = an$ and $V_{u,p} = bn$.

When the target is absent (H_0), $U \sim \mathcal{N}(M_{u,a}, V_{u,a})$, where

$$\begin{aligned} M_{u,a} &= (n-k)P_f + kP_a \\ &= [(1-\alpha)P_f + \alpha P_a]n, \qquad (2.8) \\ V_{u,a} &= (n-k)P_f(1-P_f) + kP_a(1-P_a) \\ &= [(1-\alpha)P_f(1-P_f) + \alpha P_a(1-P_a)]n. \qquad (2.9) \end{aligned}$$

Similarly, define $c = (1-\alpha)P_f + \alpha P_a$ and $d = (1-\alpha)P_f(1-P_f) + \alpha P_a(1-P_a)$, we get $M_{u,a} = cn$ and $V_{u,a} = dn$.

The overall false alarm rate using the Gaussian model is denoted by $\tilde{Q}_f$, and is defined as the conditional probability that U is greater than or equal to q given that the target is absent. That is:

$$\begin{aligned} \tilde{Q}_f &= P(U \geq q | H_0), \\ &= \int_q^n \frac{1}{\sqrt{2\pi V_{u,a}}} e^{-\frac{(X-M_{u,a})^2}{2V_{u,a}}} dX \\ &= Q\left(\frac{q - M_{u,a}}{\sqrt{V_{u,a}}}\right) - Q\left(\frac{n - M_{u,a}}{\sqrt{V_{u,a}}}\right). \qquad (2.10) \end{aligned}$$

Assuming very large network size n, then $Q\left(\frac{n-M_{u,a}}{\sqrt{V_{u,a}}}\right) \to 0$, and it follows that:

$$\tilde{Q}_f \approx Q\left(\frac{q - M_{u,a}}{\sqrt{V_{u,a}}}\right). \qquad (2.11)$$

Similarly, the overall miss detection probability $\tilde{Q}_m$ is defined as the conditional probability that U is less than q given that the target is present. Hence, $\tilde{Q}_m$ can be

expressed as:

$$\begin{aligned}\tilde{Q}_m &= P(U < q|H_1) \\ &= 1 - P(U \geq q|H_1) \\ &\approx 1 - Q\left(\frac{q - M_{u,p}}{\sqrt{V_{u,p}}}\right). \end{aligned} \tag{2.12}$$

There is an obvious trade-off between the false alarm rate and the miss detection probability. It can be noted from equations (2.11) and (2.12) that, increasing q will result in an improved false alarm rate, but degrades the the miss detection probability. Therefore, the miss detection constraint sets an upper bound to the value of q. If the miss detection constraint is $Q_m \leq \beta$, then q should be bounded by[9]:

$$q \leq M_{u,p} + \sqrt{V_{u,p}} Q^{-1}(1 - \beta), \tag{2.13}$$

In order to minimize the false alarm rate, the largest q value is selected. That is:

$$\begin{aligned} q &= \left\lfloor M_{u,p} + \sqrt{V_{u,p}} Q^{-1}(1 - \beta) \right\rfloor \\ &= \lfloor an + \sqrt{bn} Q^{-1}(1 - \beta) \rfloor, \end{aligned} \tag{2.14}$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x. This approach ensures that the q defined in (2.14) minimizes the false alarm rate while satisfying the miss detection constraint. □

In the following, we consider the percentage of malicious nodes that can be tolerated by the network using the q-out-of-m fusion rule. More specifically, we have the following result: *For reliable data fusion using the q-out-of-m rule, the percentage of malicious*

[9]Note that in order to have $Q(x) \geq y$, then $x \leq Q^{-1}(y)$.

nodes has to satisfy $\alpha < \frac{P_d}{P_d + P_a}$*, where* P_d *is the probability of detection of the sensors and* P_a *is the attack probability of the malicious nodes.*

In fact, as discussed in Section 2.2, in order to achieve low false alarm rate, we should have $q > k$. Following *Proposition 2.1*, this implies that:

$$k < M_{u,p} + \sqrt{V_{u,p}} Q^{-1}(1 - \beta) \tag{2.15}$$

Note that the second term on the left hand side of (2.15) is negative, since β is a small number. We have:

$$\begin{aligned} k &< M_{u,p} \\ &< \left[(1 - \alpha) P_d + \alpha(1 - P_a)\right] n. \end{aligned} \tag{2.16}$$

Dividing both sides by n, and note that $\alpha = \frac{k}{n}$, we get:

$$\alpha < \frac{P_d}{P_d + P_a}. \tag{2.17}$$

The values of q obtained by the linear, the enhanced linear, and the closed-form solution, are shown in Figure 2.5. It is observed that the enhanced linear and the derived closed-form solution obtain almost the same value for q, which shows that the closed-form solution obtained using the Gaussian model is accurate and achieves near-optimal solution. The significance of the closed-form solution is that it facilitates instantaneous adaptation of the fusion parameters to the changes in the environmental conditions and the attack strategies. Adaptive fusion will be further discussed in Section 2.6.

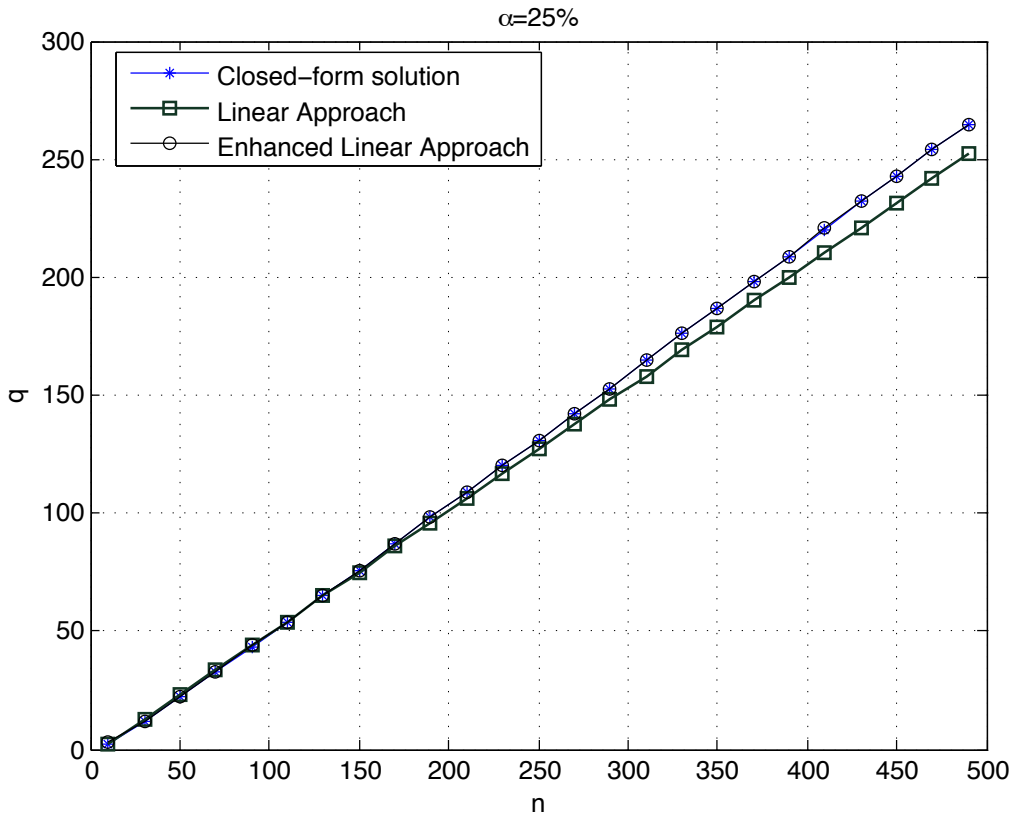


Figure 2.5: The q obtained using linear approach, enhanced linear approach and closed-form solution. Here, the percentage of malicious sensors $\alpha = 25\%$, $P_{a,m} = 1$, $P_{a,f} = 1$, $P_d = 0.775$, and $P_f = 0.1$.

2.5 Analytical Bounds for the Proposed Approaches

In this section, we derive the analytical bound for the q-out-of-m scheme based on the closed-form solution and the linear approach, and *show that the accuracy of the q-out-of-m scheme increases exponentially as the network size increases, even if the percentage of malicious sensors remains the same.* We consider the linear approach first, for which we have:

Proposition 2.2 *Using the linear q-out-of-m approach, for a fixed percentage of malicious sensors, the overall false alarm rate diminishes exponentially as the network*

size n goes to infinity. More specifically, when n is very large and $P_f < \frac{q-k}{n-k}$, then $Q_f \leq \exp\{-(An+B)\}$, where A and B are constants, and $A > 0$.

Proof: In the worst case when $P_{a,f} = 1$, the false alarm probability Q_f can be expressed as:

$$Q_f = \sum_{i=q-k}^{n-k} \binom{n-k}{i} P_f^i (1-P_f)^{n-k-i}. \tag{2.18}$$

It is clear that Q_f is the summation over a binomial probability density function with parameters P_f and $n-k$, where the random variable is the number of benign nodes having false alarm. Recall that the Chernoff bound states that if X is a binomial random variable with mean μ, then:

$$Pr\{X \geq (1+\delta)\mu\} < \left[\frac{\exp^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} \qquad \text{for } \delta > 0. \tag{2.19}$$

Let the random variable X be the number of benign nodes sending false alarm, then $\mu = (n-k)P_f$. Therefore, $Q_f = Pr\{X \geq q-k\}$. By setting $(1+\delta)\mu = q-k$, we get $\delta = \frac{q-k}{(n-k)P_f} - 1$. When $P_f < \frac{q-k}{n-k}$, we have $\delta > 0$. It then follows from (2.19) that:

$$Q_f < \left[\frac{q-k}{\mu}\right]^{-(q-k)} \exp\{q-k-\mu\}. \tag{2.20}$$

By applying the logarithm function followed by the exponential function, we get:

$$Q_f < \exp\left\{-(q-k)\ln\left[\frac{q-k}{\mu}\right] + q - k - \mu\right\}. \tag{2.21}$$

When the linear approach is employed, we have $q = \hat{q}_{n,\alpha}$, where $\hat{q}_{n,\alpha} = q_{n_O,\alpha} + (n-n_o)S_o(\alpha)$. Here, we assume that the percentage of the malicious nodes, α, is fixed. The term $\ln\left[\frac{q-k}{\mu}\right]$ can be written as:

$$\ln\left[\frac{q-k}{\mu}\right] = \ln\left[\frac{q_{n_o,\alpha}+(n-n_o)S_o(\alpha)-\alpha n}{n(1-\alpha)P_f}\right]$$
$$= \ln\left[\frac{q_{n_o,\alpha}-n_oS_o(\alpha)}{n(1-\alpha)P_f}+\frac{S_o(\alpha)-\alpha}{(1-\alpha)P_f}\right]. \quad (2.22)$$

We have:

$$\lim_{n\to\infty}\ln\left[\frac{q-k}{\mu}\right] = \ln\left[\frac{S_o(\alpha)-\alpha}{(1-\alpha)P_f}\right] = Z_o, \quad (2.23)$$

where Z_o is a constant. Following (2.21)-(2.23), when $n \to \infty$, Q_f can be bounded as follows:

$$\begin{aligned} Q_f &< \exp\{-(q-k)Z_o + q - k - \mu\} \\ &< \exp\left\{-(Z_o-1)\left[q_{n_o,\alpha}+(n-n_o)S_o(\alpha)-\alpha n\right]\right. \\ &\quad \left.-n(1-\alpha)P_f\right\} \\ &< \exp\{-(An+B)\}, \end{aligned} \quad (2.24)$$

where $A = (Z_o-1)[S_o(\alpha)-\alpha]+(1-\alpha)P_f$ and $B = (Z_o-1)\left[q_{n_o,\alpha}-n_oS_o(\alpha)\right]$. Note that $A > 0$. In fact, note that $\ln(x) > 1-\frac{1}{x}$ for $x > 1$ [74], following (2.23), we have $Z_o - 1 > -\frac{(1-\alpha)P_f}{S_o(\alpha)-\alpha}$. Then, $A > \left[-\frac{(1-\alpha)P_f}{S_o(\alpha)-\alpha}\right][S_o(\alpha)-\alpha]+(1-\alpha)P_f = 0$.

This proves that, as n goes to infinity, Q_f decreases exponentially with the network size, even if the percentage of malicious nodes is fixed. □

For the closed-form solution, we have:

Proposition 2.3 *For a fixed percentage of malicious nodes and under the same attack strategy, the overall false alarm rate using the closed-form q-out-of-m approach diminishes exponentially as the network size n goes to infinity. More Specifically, for large n*

and $P_f < \frac{q-k}{n-k}$, $\tilde{Q}_f \leq \exp\left[-\frac{1}{2}\frac{(a-c)^2}{d}n\right]$, *where* a, c *and* d *are constants.*

Proof: From equation (2.11), the false alarm rate can be bounded using the Chernoff bound [75] as follows:

$$\tilde{Q}_f \leq \frac{1}{2}\exp\left[-\frac{1}{2}\left(\frac{q-M_{u,a}}{\sqrt{V_{u,a}}}\right)^2\right], \quad \frac{q-M_{u,a}}{\sqrt{V_{u,a}}} > 0. \tag{2.25}$$

The condition $\frac{q-M_{u,a}}{\sqrt{V_{u,a}}} > 0$ is equivalent to $q > M_{u,a}$, and consequently $P_f < \frac{q-kP_a}{n-k}$. In the worst case when $P_a = 1$, this condition becomes $P_f < \frac{q-k}{n-k}$. Following *Proposition 2.1*, set $q = \lfloor an + \sqrt{bn}Q^{-1}(1-\beta)\rfloor$, we get:

$$\begin{aligned}\tilde{Q}_f &\leq \frac{1}{2}\exp\left[-\frac{1}{2}\left(\frac{an+\sqrt{bn}Q^{-1}(1-\beta)-cn}{\sqrt{dn}}\right)^2\right] \\ &\leq \frac{1}{2}\exp\left[-\frac{1}{2}\left(\frac{(a-c)\sqrt{n}+\sqrt{b}Q^{-1}(1-\beta)}{\sqrt{d}}\right)^2\right].\end{aligned} \tag{2.26}$$

If the network size is very large, i.e., $n \to \infty$, then $|(a-c)\sqrt{n}| >> \left|\sqrt{b}Q^{-1}(1-\beta)\right|$ and we obtain:

$$\tilde{Q}_f \leq \frac{1}{2}\exp\left[-\frac{1}{2}\frac{(a-c)^2}{d}n\right]. \tag{2.27}$$

If α is fixed and the attack strategy is the same, then a and c are constants and the false alarm rate decreases exponentially as the network size increases. This proves the proposition. □

Discussions: (i) Our analytical results provided in this section highlight the impact of the q-out-of-m fusion scheme on large-scale networks that are more reliable under malicious attacks. They indicate that when the q-out-of-m rule is used for data fusion,

then the false alarm rate diminishes exponentially as the network size increases even if the percentage of malicious sensors remains the same. This implies that for a fixed α, we can improve the network performance significantly by increasing the network size.

(ii) **Explanation on the condition in *Propositions 2.2* and *2.3*:** The condition "$P_f < \frac{q-k}{n-k}$" is equivalent to "$q - k > P_f(n - k)$". The physical meaning of this condition can be explained in two ways: First, in order to have arbitrarily low overall false alarm probability by the q-out-of-m fusion rule, the individual false alarm rates of the benign nodes should be less than a certain limit. This limit is equal to the ratio between the least number of benign nodes in the set of q nodes relied on in the q-out-of-m scheme $(q - k)$, to the total number of benign nodes $(n - k)$. Second, since each benign node has a none zero false alarm rate P_f, to reduce the overall false alarm rate, sufficient number of benign nodes need to be taken into account so the the "averaged" result will lead to a low overall false alarm rate. Mathematically, this condition can also be understood as: if $P_f > \frac{q-k}{n-k}$ (i.e., $\delta < 0$), then the overall false alarm rate would be very high and cannot decrease beyond a certain level. In fact, we can write $Pr\{X \geq (1 + \delta)\mu\} = Pr\{(1 - |\delta|)\mu \leq X < \mu\} + Pr\{X \geq \mu\}$. The term $Pr\{X \geq \mu\}$ is generally high. If we approximate X as a normal distribution, then $Pr\{X \geq \mu\} = \frac{1}{2}$. Therefore, if $\delta < 0$, the overall false alarm $Q_f \geq \frac{1}{2}$.

2.6 Malicious Node Detection and Adaptive Fusion

In this section, we propose to enhance the system performance through malicious node detection, where the hostile behavior is identified and the malicious sensors are discarded from the final decision making. Furthermore, we propose an adaptive fusion procedure, where the fusion parameters are tuned based on the attack behavior and the percentage of the malicious sensors.

2.6.1 The Malicious Node Detection Scheme

Let I_{mal} be the set of the malicious nodes, and O_{N_s} denotes the reports of all nodes till the sensing period N_s. When the attack strategy is known, and the percentage of malicious nodes is fixed, a traditional approach to find the malicious set, I_{mal}, is to maximize the *a posterior* probability of I_{mal} given the observations O_{N_s} [76]. That is, the detected malicious set $\hat{I}_{mal} = \arg\max_{I_{mal}} P(I_{mal}|O_{N_s})$, where $P(I_{mal}|O_{N_s})$ is the conditional probability that the malicious set is I_{mal} given all the reports O_{N_s}. However, this detection approach is difficult to be implemented since it requires searching over all possible sets of I_{mal}.

In this section, we propose a simple malicious node detection scheme, where the sensors' decision reports are used to identify the malicious nodes and estimate their attack behavior. Let $P_{a,f}(i)$ and $P_{a,m}(i)$ denote the probabilities that the ith node attacks when the target is absent and present, respectively. Let $\hat{P}_{a,f}(i)$ and $\hat{P}_{a,m}(i)$ be their estimated versions. We estimate $P_{a,f}(i)$ and $P_{a,m}(i)$ by using two counters for each node at the mobile access point. More specifically, for node i,

- $T_{i,o}$: represents the number of times node i sends '0' when the final decision is '1'.
- $T_{i,1}$: represents the number of times node i sends '1' when the final decision is '0'.

These counters are updated after each sensing period by comparing the final decision (obtained using the q-out-of-m rule) with the individual sensing reports.

Assuming the observation interval is N sensing periods, and the number of observations where the access point decides that the target is present and absent are N_1 and N_0, respectively. Then, if the node is benign, $\frac{T_{i,o}}{N_1}$ and $\frac{T_{i,1}}{N_0}$ would be indications for the ith node's miss detection probability and false alarm rate, respectively. On the other

hand, if node i is malicious, $\frac{T_{i,o}}{N_1}$ and $\frac{T_{i,1}}{N_0}$ will be estimates for $P_{a,m}(i)$ and $P_{a,f}(i)$, respectively.

More specifically, $\frac{T_{i,o}}{N} = Q_f(1-P_f)(1-p) + (1-Q_m)P_m p$, where $p = Prob\{H_1\}$ is the probability that the target is present. When both Q_f and Q_m are very low, then $p = \frac{N_1}{N}$ and $\frac{T_{i,o}}{N} \simeq P_m\frac{N_1}{N}$, which implies that $\frac{T_{i,o}}{N_1} \simeq P_m$. We can write, $\frac{T_{i,o}}{N_1} < P_m + \delta_{m,0}$, where $\delta_{m,0}$ mainly depends on the false alarm rate of the access point and the individual sensor. Similarly, $\frac{T_{i,1}}{N} = Q_m(1-P_m)p + (1-Q_f)P_f(1-p) \simeq P_f\frac{N_0}{N}$. That is, $\frac{T_{i,1}}{N_0} < P_f + \delta_{f,0}$, where $\delta_{f,0}$ mainly depends on the miss detection probability at the access point as well as the individual sensor.

We define the thresholds $\lambda_{p,f}$ and $\lambda_{p,m}$ as:

$$\lambda_{p,f} = P_f + \delta_{f,0}, \qquad \lambda_{p,m} = P_m + \delta_{m,0}, \tag{2.28}$$

where P_f and P_m are the benign nodes' false alarm and miss detection probabilities, $\delta_{f,0}$ and $\delta_{m,0}$ represent the tolerance in the estimated false alarm rate and miss detection probability of the nodes. Since, the fusion rule at the access point keeps the miss detection constraint, i.e. Q_m has low value, $\delta_{0,f}$ is a considerable small value. The miss detection constraint could be met at the expense of high false alarm rate, especially at small network sizes. Thus, we choose $\delta_{f,0} < \delta_{m,0}$. The parameters $\delta_{f,0}$, $\delta_{m,0}$ should also account for the tolerance in the benign nodes' miss detection and false alarm rates that would result from the changes in the environmental conditions and/or the sensors' capabilities. We assume that the MA can estimate P_f and P_m.

The malicious node detection procedure has two levels:

- *Level 1: Discard the suspicious reports* If $\frac{T_{i,o}}{N_1} \geq \lambda_{p,m}$ or $\frac{T_{i,1}}{N_0} \geq \lambda_{p,f}$, the node's report is discarded from the current decision process, but its counters will continue to be updated in the next sensing periods.

- *Level 2: Discard the unreliable nodes* If $\frac{T_{i,o}}{N_1} \geq P_m + \delta_1$ or $\frac{T_{i,1}}{N_0} \geq P_f + \delta_2$, where δ_1 and δ_2 are relatively large, then the corresponding node will be discarded from the sensing process. The nodes' counters would be calculated to estimate the attack probability, but they will not be involved in the final decision making process.

It should be noted that N needs to be greater than or equal to a certain threshold N_{th} before taking the decision to discard any node. N_{th} should be chosen to ensure the accuracy of the time averages, $\hat{P}_{a,f}(i)$ and $\hat{P}_{a,m}(i)$. When $P_{a,f}(i)$ and $P_{a,m}(i)$ are in the orders of 10^{-1}, it is safe to choose $N_{th} \geq 100$. As will be illustrated in Section 2.7, the detection of the malicious nodes launching dynamic attacks is generally more difficult and takes longer time than the detection of the malicious nodes performing static attacks.

2.6.2 The Adaptive Fusion Algorithm

Adaptive fusion can be achieved by updating the value of the q-out-of-m fusion parameters based on the average probability of attack. Recall that $\hat{I}_{mal}$ is the set of detected malicious nodes, then $\left|\hat{I}_{mal}\right|$ is the total number of sensors detected to be malicious. The estimated average attack probability is given by:

$$\hat{P}_a = \frac{1}{\left|\hat{I}_{mal}\right|} \sum_{i=1}^{\left|\hat{I}_{mal}\right|} \hat{P}_a(\hat{I}_{mal}(i)) \tag{2.29}$$

where $\hat{I}_{mal}(i)$ is the ith detected malicious sensor and $\hat{P}_a(\hat{I}_{mal}(i)) = \frac{T_{\hat{I}_{mal}(i),o} + T_{\hat{I}_{mal}(i),1}}{N}$. Then, q is tuned using equation (2.14) with the new problem settings, where $n - \left|\hat{I}_{mal}\right| \Rightarrow n$, $k - \left|\hat{I}_{mal}\right| \Rightarrow k$, $\alpha = k/n$ and $P_a = \hat{P}_a$.

To deal with the malicious nodes who disguise themselves as benign nodes for a long periods between attacks, we can reset the counters periodically, and take the history of the nodes into account through initial conditions. More specifically, we define $N_{th,2}$ as the observation interval after which the counters are reset to the initial conditions. $N_{t,0}$ and $N_{t,1}$ are the total sensing periods performed in the network when the final decision is '0' and '1', respectively. The history of each node i is always reflected in the percentage of its false alarm reports ($I^o_{f,i}$) and the percentage of its miss detection reports ($I^o_{m,i}$). The adaptive fusion algorithm based on the malicious node detection is further illustrated in Table 2.1. As will be shown in the simulation section, adaptive fusion with malicious node detection can improve the system performance significantly.

We define η_d and η_f as the detection accuracy and false alarm rate of the malicious node detection scheme, respectively. That is,

$$\eta_d \triangleq \frac{N_{MM}}{k}, \quad \eta_f \triangleq \frac{N_{BM}}{n-k}, \tag{2.30}$$

where N_{MM} is the number of malicious nodes detected to be malicious, N_{BM} is the number of benign nodes mistakenly regarded as malicious, k is the total number of malicious sensors and $(n-k)$ is the number of benign sensors. Note that $\left|\hat{I}_{mal}\right| = N_{MM} + N_{BM}$. It will be shown in Section 2.7 that with sufficient observation time, the proposed detection scheme can achieve high η_d and low η_f under static and dynamic attacks.

2.6.3 Analysis From the Entropy Point of View

In the proposed malicious node detection approach, each node's behavior is determined based on the uncertainty in the accuracy of its sensing report. Since uncertainty is generally measured by entropy, in this section, we analyze the proposed approach

Table 2.1: Adaptive fusion with malicious node detection

Initialize all counters (N, N_1, N_0, $N_{t,1}$, $N_{t,0}$, $N_{r,1}$, $N_{r,0}$, $T_{i,o}$, $T_{i,1}$, $I^o_{m,i}$, $I^o_{f,i}$) to zeros
After each sensing period, do:
Set N to $N+1$
for i from 1 to n
  if decision of node i is not equal to the final q-out-of-m decision
   check if node i reports '0' and the final decision is '1'
      Set $T_{i,o}$ to $T_{i,o}+1$, $N_{t,1}$ to $N_{t,1}+1$, and N_1 to N_1+1
   otherwise check if node i report '1' and the final decision is '0'
      Set $T_{i,1}$ to $T_{i,1}+1$, $N_{t,0}$ to $N_{t,0}+1$, and N_0 to N_0+1
  end if
end for
If the observation intervals N is greater than or equal N_{th}, then check for each node i:
{
  if $\frac{T_{i,o}}{N_1} + I^o_{m,i} \geq P_m + \delta_{0,m}$ or $\frac{T_{i,1}}{N_0} + I^o_{f,i} \geq P_f + \delta_{0,f}$
      discard the reports of node i from the current decision process
  end if
  if $\frac{T_{i,o}}{N_1} + I^o_{f,i} \geq P_m + \delta_1$ or $\frac{T_{i,1}}{N_0} + I^o_{f,i} \geq P_f + \delta_2$
      discard the reports of node i from the sensing process
  end if
  Estimate the attack probability of each node: $\hat{P}_a(i) = \frac{T_{i,o}+T_{i,1}}{N} + I^o_{f,i} + I^o_{m,i}$
  Estimate the average probability of attack using (2.29)
  Update q based on the new settings using (2.14)
}
If $N > N_{th,2}$, do the following for all nodes in the sensing process
{
  Update the history: Set $I^o_{m,i}$ to $\frac{T_{i,o}+I^o_{m,i}N_{r,1}}{N_{t,1}}$ and $I^o_{f,i}$ to $\frac{T_{i,1}+I^o_{f,i}N_{r,0}}{N_{t,0}}$
  Set $N_{r,1} = N_{t,1}$
  Set $N_{r,0} = N_{t,0}$
  Reset the counters $T_{i,o}$, $T_{i,1}$
  Reset N, N_1, N_0 to zero
}

using the entropy-based trust model [37].

First, for each node $i \in \{1, 2, ..., n\}$, we define two trust metrics $Trust_f(i)$ and $Trust_m(i)$ to represent the uncertainty in the node's accuracy when the target is absent and present, respectively.

$$Trust_f(i) \triangleq \begin{cases} 1 - H(\hat{P}_{a,f}(i)), & \text{if } \hat{P}_{a,f}(i) < 0.5; \\ \\ H(\hat{P}_{a,f}(i)) - 1, & \text{if } \hat{P}_{a,f}(i) \geq 0.5. \end{cases} \tag{2.31}$$

where $H(\hat{P}_{a,f}(i))$ is the entropy which represents the uncertainty that node i intentionally reports a false '1' when the actual state of nature is '0'. That is,

$$\begin{aligned} H\left(\hat{P}_{a,f}(i)\right) &= -\hat{P}_{a,f}(i) \log_2 \left[\hat{P}_{a,f}(i)\right] \\ &\quad - \left[1 - \hat{P}_{a,f}(i)\right] \log_2 \left[1 - \hat{P}_{a,f}(i)\right]. \end{aligned} \tag{2.32}$$

$Trust_m(i)$ is defined in a similar way by replacing $\hat{P}_{a,f}(i)$ in equation (2.31) with $\hat{P}_{a,m}(i)$.

The entropy trust metrics are in the range $[-1, 1]$, where negative values mean that the attack probability of the corresponding node is greater than 0.5. The trust metrics are equal to '1' when the corresponding node is benign with a perfect detection accuracy. Figure 2.6 plots the trust metrics ($Trust_f(i)/Trust_m(i)$) versus the $\hat{P}_{a,f}(i)/\hat{P}_{a,m}(i)$.

Note that P_f and P_m are generally small quantities, and we can assume $P_f + \delta_{0,f} < 1/2$ and $P_m + \delta_{0,m} < 1/2$. Define $\lambda_{e,f}$ and $\lambda_{e,m}$ as:

$$\lambda_{e,f} \triangleq 1 - H(P_f + \delta_{0,f}), \qquad \lambda_{e,m} \triangleq 1 - H(P_m + \delta_{0,m}). \tag{2.33}$$

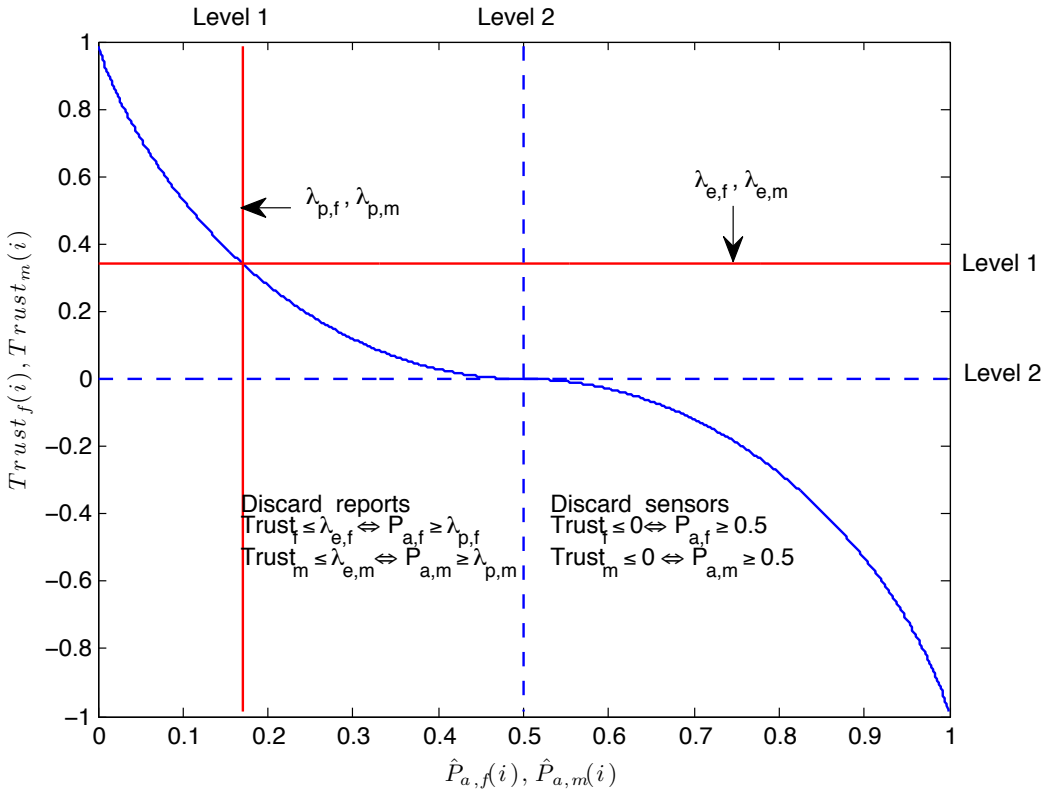


Figure 2.6: The trust metrics $(Trust_f(i)/Trust_m(i))$ vs. the $\hat{P}_{a,f}(i)/\hat{P}_{a,m}(i)$.

Compare (2.33) and (2.28), we can see that the proposed malicious node detection approach can be mapped to the entropy-based trust model. The equivalence is further illustrated in Table 2.2.

Our discussions above show that under the same settings for $\delta_{0,f}$, $\delta_{0,m}$, δ_1 and δ_2, the proposed malicious node detection scheme is equivalent to the detection approach based on the entropy-based trust model. This implies that the proposed malicious node detection scheme is optimal from the information theory point of view.

Table 2.2: Equivalence between the entropy based trust model and the proposed malicious node detection

Cases	Entropy-based trust model vs. the proposed malicious node detection approach
1. Discard suspicious reports	$Trust_f(i) \leq \lambda_{e,f} \Leftrightarrow \hat{P}_{a,f}(i) \geq \lambda_{p,f}$
	$Trust_m(i) \leq \lambda_{e,m} \Leftrightarrow \hat{P}_{a,m}(i) \geq \lambda_{p,m}$
2. Discard unreliable nodes	$Trust_f(i) \leq 0 \Leftrightarrow \hat{P}_{a,f}(i) \geq 0.5$
	$Trust_m(i) \leq 0 \Leftrightarrow \hat{P}_{a,m}(i) \geq 0.5$

2.7 Simulation Results

In this section, we illustrate the performance of the proposed approaches through simulation examples. In the simulations, we assume that the miss detection limit is $\beta = 0.01$, the hypothesis H_1 happens with probability $p = 0.5$. The false alarm rate and the miss detection probability of each benign sensor are assumed to be $P_f = 0.1$ and $P_d = 0.775$. These false alarm and miss detection values are obtained assuming that the sensors employ energy detection, when the SNR level is 5 dB and the time-bandwidth product is 5 [77]. For the static attack strategy, we set $P_o = 1$. For the dynamic attack strategy, we set $\Delta_1 = \Delta_2 = 0.2$, $P_{o_1} = 0.7$, $P_x = 0.5$, and the number of sensing periods per attacking block is $T = 10$. The cumulative distribution function (CDF) of P_a is shown in Figure 2.7. It can be seen that P_a is spread over wide range of values.

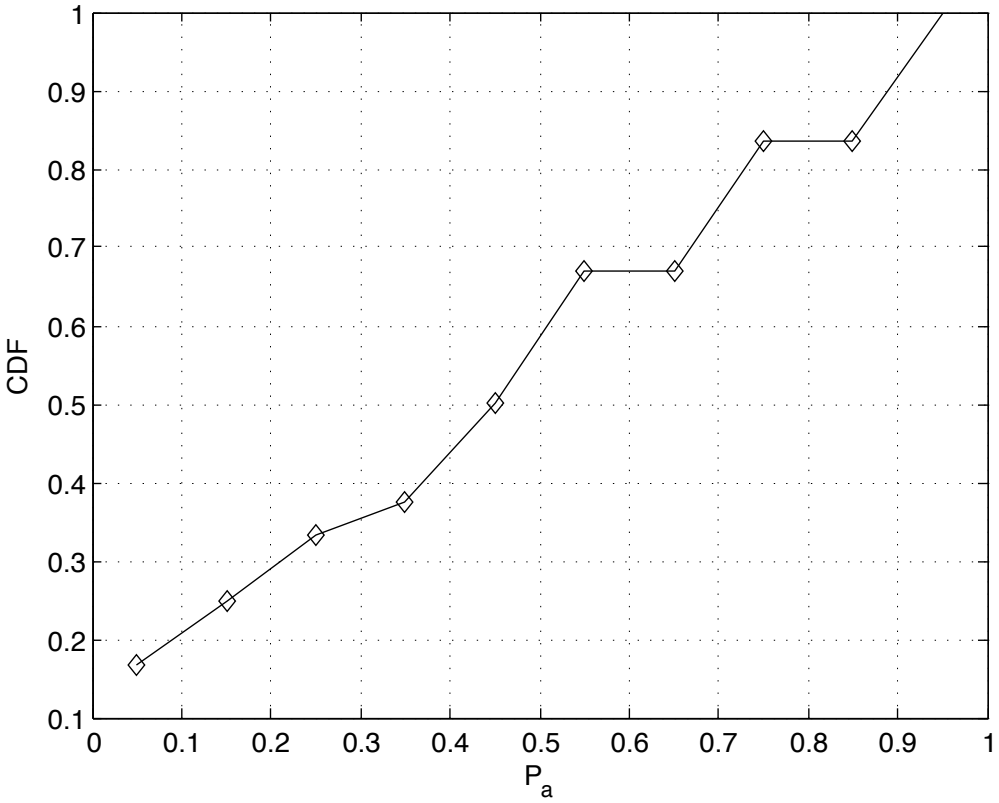


Figure 2.7: CDF of P_a for dynamic attack strategy with $\Delta_1 = \Delta_2 = 0.2$, initial P_{a_1}=0.7, $P_x = 0.5$.

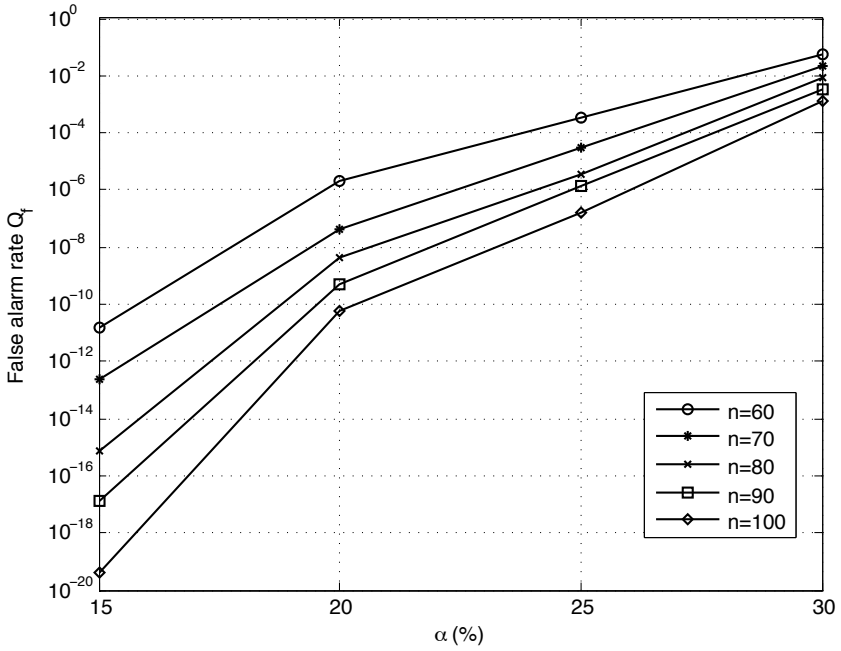
Example 2.1 - *Linear approaches and comparison with existing methods* In this example, the performance of the linear and the enhanced linear approaches are evaluated, and we also compare them with existing AND rule, OR rule, and majority voting fusion approaches. We assume that the malicious nodes can detect the target perfectly and always report false information (i.e. $P_{a,m} = 1$, $P_{a,f} = 1$). At different values of α, $S_o(\alpha)$ is calculated from Figure 2.3. The reference points are at $n_o = 35$. The false alarm rate and the corresponding miss detection probability of the linear approach for different values of n and α are shown in Figures 2.8(a) and 2.8(b), respectively. It is clear that in most cases, the miss detection constraint is met. Slight increase in the miss detection over β happens at $\alpha = 15\%$ and 25%. One solution to

this problem is to use the enhanced linear approach discussed in Section 2.3.3.

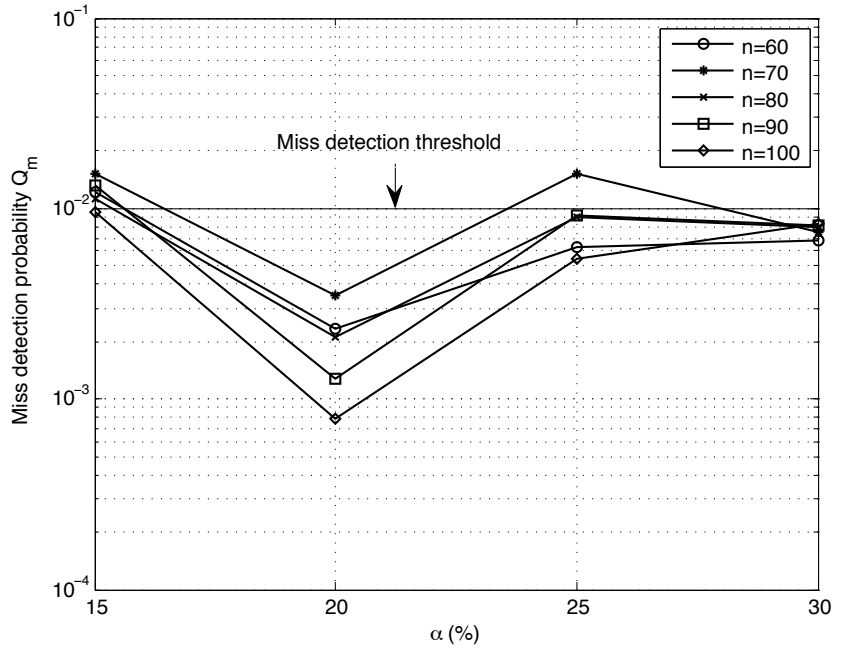
Figures 2.9(a) and 2.9(b) show the false alarm rate and the miss detection probability, respectively, at different percentages of malicious nodes, when the iterative enhanced linear approach is used to find the fusion parameter. Comparing Figure 2.9(b) with Figure 2.8(b), it can be seen that the miss detection constraint is enforced when the enhanced procedure is applied. It can also be shown from both Figures 2.8(a) and 2.9(a) that the false alarm rate improves as the network size increases even under the same percentage of malicious nodes.

In comparison with existing approaches, it is shown in Figure 2.9(b) that the majority voting rule cannot guarantee the miss detection requirement. Moreover, AND rule results in a very high miss detection probability, although it can achieve low false alarm rate. On the other hand, OR rule results in a very high false alarm rate, although it can achieve low miss detection probability. Hence, they are not reliable under malicious attacks.

Example 2.2 - *The closed-form solution without malicious node detection* In this example, we assume that $\alpha = 25\%$, and the malicious nodes have $\tilde{P}_f = P_f$ and $\tilde{P}_m = 1 - P_d$. The access point assumes that the attack probability is '1', and obtain q accordingly. We assume that the percentage of malicious sensors is known or can be estimated at the access point. It is shown in Figure 2.10(a) that when the malicious detection approach is not employed, the performance is worst under the static attack strategy with $P_o = 1$. It is observed that the false alarm rate is lower for the considered dynamic attacks as compared to the static attack. This is because that when the probability of attack is time varying, it could be very low in some sensing periods. It can be observed from Figure 2.10(a) that at a fixed percentage of malicious nodes, the false alarm rate decreases rapidly as the network size increases. This echoes our

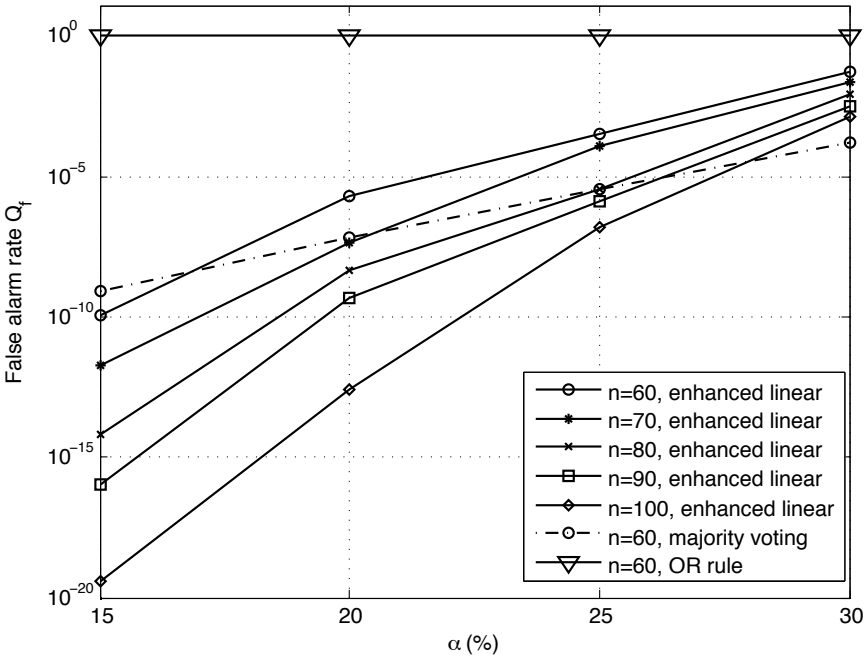


(a) The false alarm rate.

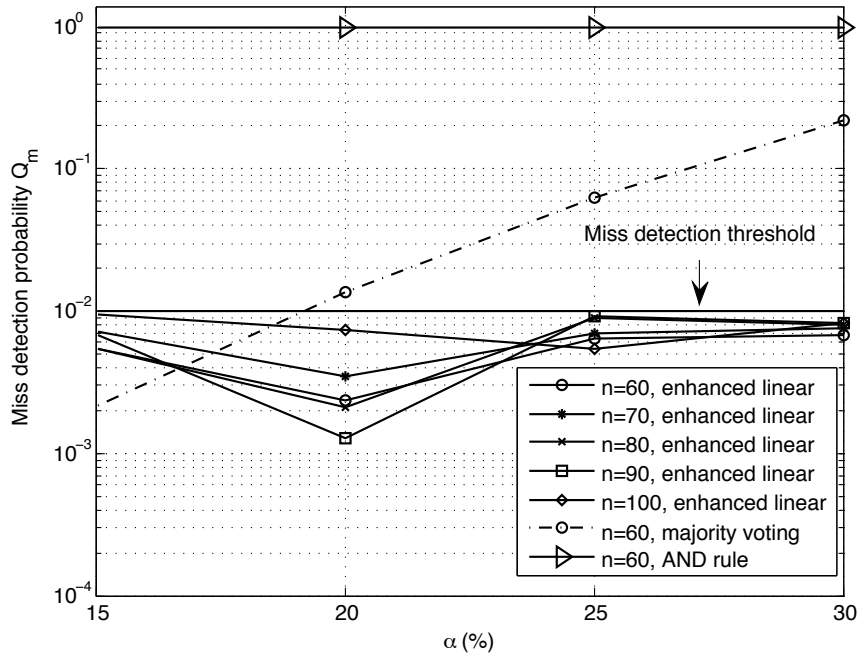


(b) The miss detection probability.

Figure 2.8: The false alarm rate and miss detection probability using the linear approach.



(a) The false alarm rate.



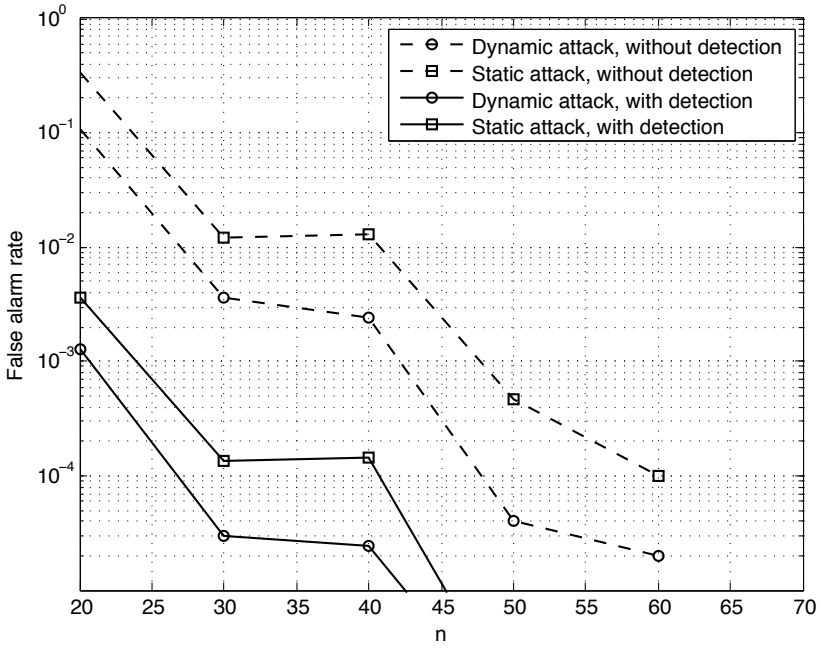
(b) The miss detection probability.

Figure 2.9: The false alarm rate and the miss detection probability using the enhanced linear approach, and comparisons with AND rule, OR rule and majority voting rule. In general, AND rule results in very high miss detection probability, although it can achieve low false alarm rate. On the other hand, OR rule results in a very high false alarm rate, although it can achieve low miss detection probability.

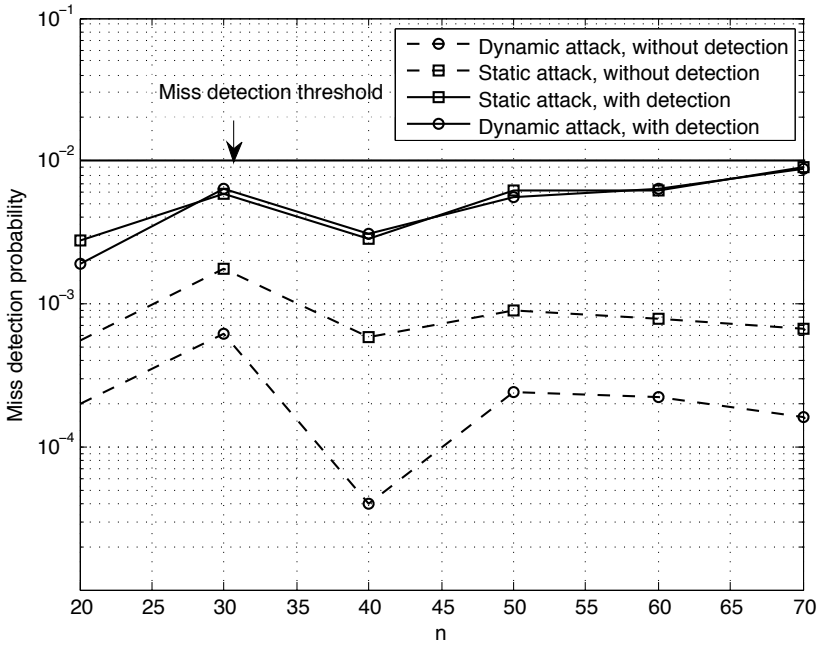
analytical results presented in Section 2.5. The miss detection probability versus the network size is plotted in Figure 2.10(b). It is clear that the miss detection constraint is met, and there is a good margin for improvement that can be achieved using the adaptive fusion scheme. This margin is mainly due to the choice of q, where the access point assumed that the attack probability is '1'.

Example 2.3 - *Adaptive fusion: closed-form solution with malicious node detection* In this example, we use the same settings as in Example 2.2. Malicious node detection is applied and the value of q is adapted based on the detected malicious behavior. Here, we use $\delta_{0,f} = 0.07$, $\delta_{0,m} = 0.2$, $\delta_1 = 0.4$, and $\delta_2 = 0.3$. Figure 2.10(a) shows the overall false alarm rate averaged over 10^4 observation periods when $N_{th} = 100$. We assume that $N_{th,2}$ is larger than the considered observation interval; hence, the counters will not be reset. The results are further averaged over 10^2 iterations to get more accurate results. It can be seen that significant performance improvement is achieved for both static and dynamic attacks when the adaptive fusion with malicious node detection is employed. Figure 2.10(b) shows that the miss detection constraint is satisfied for all cases. The non-smoothness of the curves is mainly due to tuning the integer-valued scheme parameters to satisfy the miss detection constraint. It should be emphasized that the thresholds $\delta_{0,f}$, $\delta_{0,m}$ have a direct impact of the performance of the malicious node detection scheme and they could be further optimized to improve the performance. In the simulations, we added the condition that when $\hat{P}_{a,f}(i) > \lambda_{p,f}$ or $\hat{P}_{a,m}(i) > \lambda_{p,m}$ $\forall i$, the access point will discard node i only if $\hat{P}_{a,f}(i) > 0.5$ or $\hat{P}_{a,m}(i) > 0.5$

In Figure 2.11, we show the effect of the observation threshold N_{th} on the false alarm rate of the malicious node detection scheme (η_f) for both static and dynamic attacks when the adaptive fusion is employed. Here, we set $n = 30$. The results are



(a) The false alarm rate.



(b) The miss detection probability.

Figure 2.10: The false alarm rate and miss detection probability under static and dynamic attacks with and without the malicious node detection scheme.

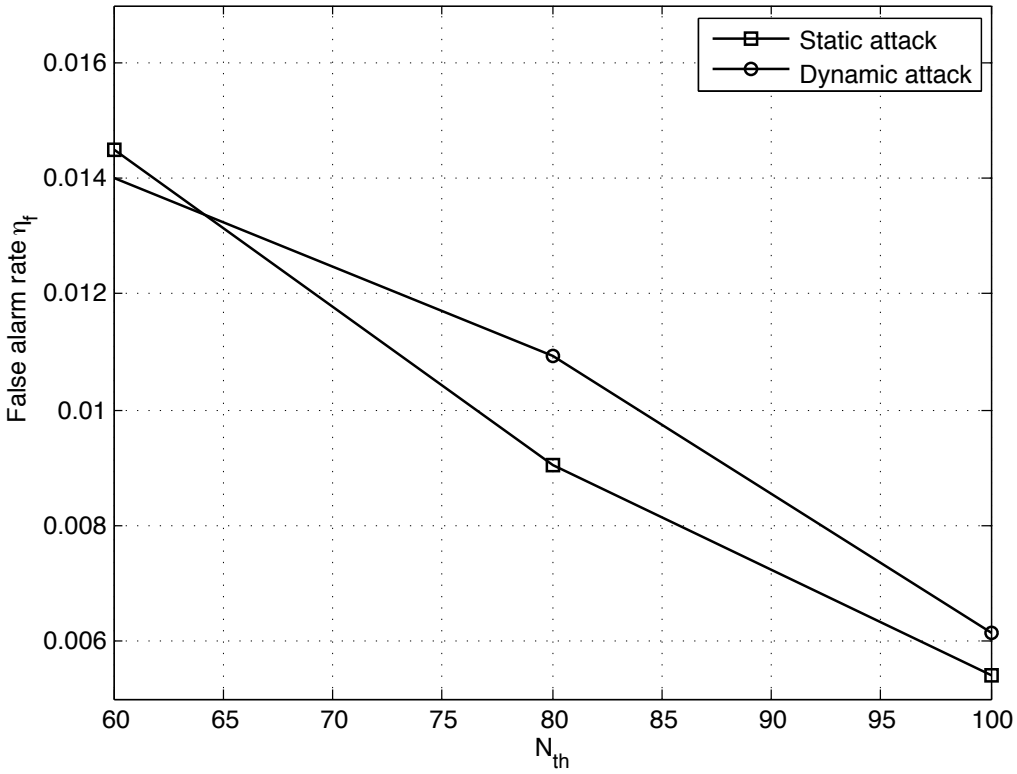
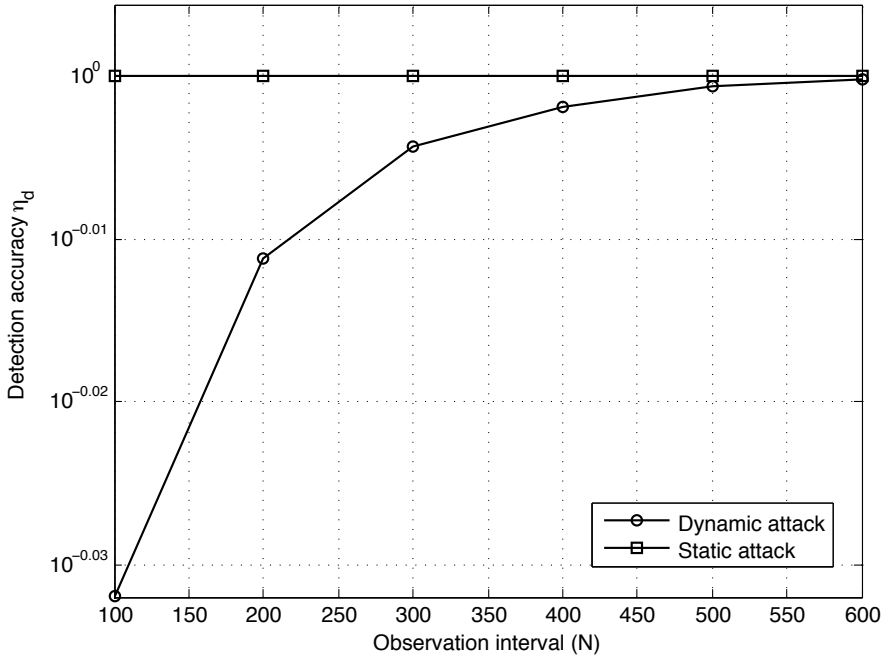


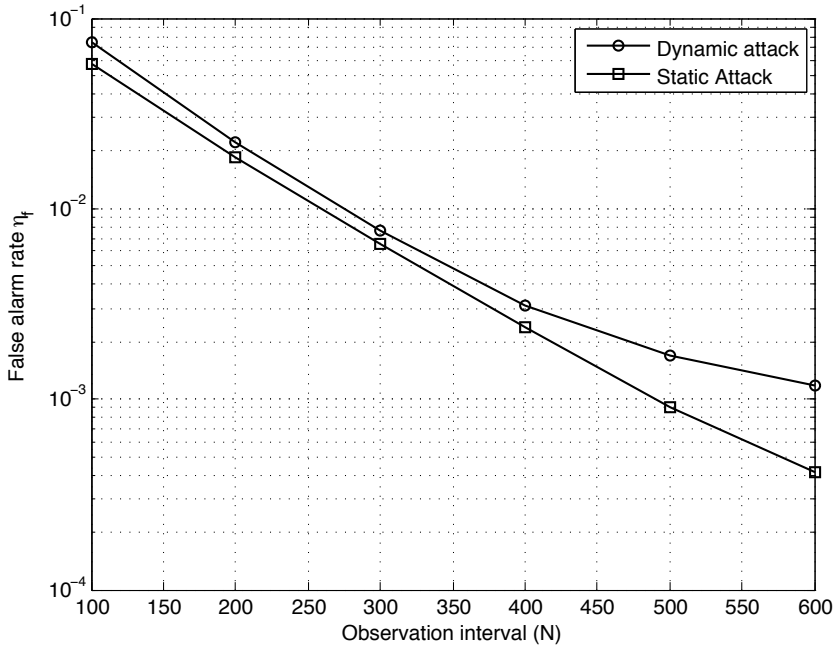
Figure 2.11: The malicious node detection false alarm rate η_f vs. the observation threshold N_{th} for static and dynamic attacks, when $n = 30$ and $\alpha = 25\%$. The results are the average of $N = 10^3$ observations, each is further averaged over 10^3 iterations.

the average of 10^3 observations each is further averaged over 10^3 iterations. Larger N_{th} would generally result in a lower η_f, since collecting more observations will result in more accurate statistics.

In Figure 2.12(a), we show the effect of the observation interval N on the detection accuracy of the malicious node detection scheme (η_d). It is clear that malicious nodes launching dynamic attack require longer observation interval to be detected than nodes adopting static attack. In general, it is clear from Figure 2.12(a) that the proposed malicious node detection scheme is efficient and provides very high detection accuracy. In Figure 2.12(b), the false alarm rate of the malicious node detection scheme is plotted versus the observation interval. As expected, it is shown that η_F decreases as more



(a) The detection accuracy



(b) The false alarm rate.

Figure 2.12: The effect of the observation interval (N) on the detection accuracy η_d and the false alarm rate η_f for static and dynamic attacks using $N_{th} = 100$ and $n = 30$. The results are averaged over 4×10^3 iterations.

observations are available at the access point.

2.8 Summary

In this chapter, we considered the q-out-of-m fusion rule for SENMA networks under Byzantine attacks. Both static and dynamic attack strategies were discussed. We proposed simplified q-out-of-m fusion schemes by exploiting the linear relationship between the scheme parameters and the network size. We also derived a near-optimal closed-form solution for the fusion threshold based on the central limit theorem. An important observation is that, even if the percentage of malicious sensors remains fixed, the false alarm rate diminishes exponentially with the network size. This implies that for a fixed percentage of malicious nodes, we can improve the network performance significantly by increasing the density of the nodes. Furthermore, we obtained an upper bound on the percentage of malicious nodes that can be tolerated using the q-out-of-m rule. It is found that the upper bound is determined by the sensors' detection probability and the attack strategies of the malicious nodes. Finally, we proposed an effective malicious node detection scheme for adaptive data fusion under time-varying attacks. The detection procedure is analyzed using the entropy-based trust model, and has shown to be optimal from the information theory point of view. It was observed that nodes launching dynamic attacks take longer time and more complex procedures to be detected as compared to those conducting static attacks. The adaptive fusion procedure has shown to provide significant improvement in the system performance under both static and dynamic attacks. Further research can be conducted on adaptive detection under Byzantine attacks with soft decision reports.

Chapter 3

Mobile Access Coordinated Wireless Sensor Networks – Design and Analysis

In this chapter, we propose a novel mobile access coordinated wireless sensor network (MC-WSN) architecture for reliable, efficient, and time-sensitive information exchange. In conventional sensor networks with mobile access points (SENMA), the mobile access (MA) points traverse the network to collect information directly from individual sensors. While simplifying the routing process, a major limitation with SENMA is that a transmission is made only if an MA visits the corresponding source node; thus, data transmission is limited by the physical speed of the MAs and the length of their trajectory, resulting in low throughput and large delay. The proposed MC-WSN effectively resolves this problem through hop number control. More specifically, with active network development and topology design, the number of hops from any sensor to a mobile access can be limited to a pre-specified number. In this chapter,

we discuss the optimal topology design for MC-WSN such that the average number of hops between the source and its nearest sink is minimized, and analyze the performance of MC-WSN in terms of throughput, stability, delay, and energy efficiency by exploiting tools in information theory, queuing theory, and radio energy dissipation model. It is shown that under stable system conditions, MC-WSN achieves much higher throughput and significantly lower delay and energy consumption than that of SENMA. The effectiveness of the proposed approaches are demonstrated through simulation results.

3.1 Introduction

Wireless sensor networks (WSNs) have been identified as a key enabling technology for various military and civilian applications, such as reconnaissance, surveillance, environmental monitoring, emergency response, smart transportation, and target tracking. Along with recent advances in remote control technologies, Unmanned Aerial Vehicles (UAVs) have been utilized in wireless sensor networks for data collection [17, 19], as well as for sensor management and network coordination. Network deployment through UAV has also been explored in literature [79, 80].

For efficient and reliable communication over large-scale networks, sensor network with mobile access points (SENMA) was proposed in [17]. In SENMA, the mobile access points (MAs) traverse the network to collect the sensing information directly from the sensor nodes. SENMA has been considered for military applications, where small low-altitude unmanned aerial vehicles (UAVs) serve as the mobile access points that collect sensing information for surveillance, reconnaissance and collaborative spectrum sensing [81]. When the energy consumption at the MAs is not of a concern, SENMA improves the energy efficiency of the individual sensor nodes over ad-hoc networks by relieving sensors from complex and energy-consuming routing functions.

While simplifying the routing process, a major limitation with SENMA is that a transmission is made only if an MA visits the corresponding source node; thus, data transmission is largely limited by the physical speed of the MAs and the length of their trajectory, resulting in low throughput and large delay. This makes SENMA undesirable for time-sensitive applications.

As an effort to solve this problem, in this work, we propose a mobile access coordinated wireless sensor networks (MC-WSN) for time-sensitive, reliable, and energy-efficient information exchange. In MC-WSN, the whole network is divided into cells, each is covered by one MA, and served with powerful center cluster head (CCH) located at the middle of the cell, and multiple ring cluster heads (RCHs) uniformly distributed along a ring in the cell. The MAs coordinate the network through deploying, replacing and recharging the nodes. They are also responsible for enhancing the network security, by detecting compromised nodes then replacing them. Data transmission from sensor nodes to the MA goes through simple routing with cluster heads (CHs), CCH or RCHs serving as relay nodes. As in SENMA, the senors are not involved in the routing process. A major feature of MC-WSN is that: Through active network deployment and topology design, the number of hops from any sensor to the MA can be limited to a pre-specified number. As will be shown, the hop number control, in turn, results in better system performance in throughput, delay, energy efficiency, and security management.

For performance evaluation of the proposed architecture, we first discuss optimal topology design for MC-WSN such that the average number of hops between the source and its nearest sink is minimized, then analyze the performance of MC-WSN in terms of throughput, stability, delay, and energy efficiency.

As an important measure of network performance, throughput is generally defined as the amount of information that can be successfully transmitted over a network,

and is largely determined by the network model and transmission protocols. Existing work on throughput analysis is versatile [22–24, 82–85], including one-hop centralized cases [82, 83] and ad-hoc cases [22, 84, 85]. There are also research on systems with mobile nodes [28, 86] and systems with mobile access points, like SENMA [17]. In SENMA, as there is a direct link between each sensor and the mobile sink, the system throughput is significantly superior to that of ad-hoc sensor networks [17].

In [22], the throughput of random ad-hoc networks is studied. It was shown that the throughput obtained by each node vanishes as the number of nodes in the network increases. More specifically, for an ad-hoc network containing n nodes, the throughput obtainable by each node is $O\left(\frac{W}{\sqrt{n}}\right)$ bit-meters/sec, where W is the maximum capacity of each link in the network. Note that the size or density of an ad-hoc network or a wireless sensor network plays a critical role in network performance. This result indicates that for reliable and efficient communications, the network cannot be completely structureless, but should have a well-defined structure while maintaining sufficient flexibility. This thought has actually been reflected in the merging of centralized and ad-hoc networks, leading to ad-hoc networks with structures, known as hybrid networks [44, 87]. As will be shown in Section 3.2, the proposed MC-WSN is also an example of hybrid network: it has a hierarchical structure supported by the CCH, RCHs, and CHs; at the same time, it also allows partially ad-hoc routing for network flexibility and diversity.

In this chapter, we analyze the throughput of MC-WSN under both single path and multiplath routing. We evaluate the average per node throughput and compare it with that of SENMA. It is observed that the throughput of MC-WSN is independent of the physical speed of the MA, and hence is orders of magnitude higher than that of the conventional SENMA.

Throughput is closely related to network stability and delay performance. For a

system to be stable, the arrival rate at each node cannot be larger than the service rate, which is bounded by the corresponding throughput. At the same time, to ensure bounded delay in the transmission, the system has to be stable.

The major challenge for stability and delay analysis in wireless networks lies in the dependency among different queues along the routing path. For example, we have dependency between the inter-arrival times (which measures the time difference between two successive arrivals) and service times at each of the intermediate queues, and dependency between service times at different queues. Due to these dependencies, network analysis becomes highly complicated and intractable. Fortunately, it was observed that when the network is densely connected with moderate to heavy traffic loads, the dependencies between the inter-arrival times and service times can be eliminated by merging or multiplexing multiple packet streams at each link. This is known as the *Klienrock independence assumption*, and has shown to be a valid model for network analysis [88].

In this work, based on the *Klienrock independence assumption* and queuing modeling/analyzing theorems (mainly *Burke's theorem* and *Little's Theorem*), we establish the queuing model for the CHs in MC-WSN, analyze the stability and delay of the network, while highlighting their relationship with the throughput.

The major contributions in this chapter can be summarized as follows:

- We propose a reliable and efficient mobile access coordinated WSN (MC-WSN) architecture for time-sensitive information exchange. The MAs coordinate the network through node deployment, replacement, recharging, malicious node detection, and data collection. The energy efficiency for individual sensors is maximized as they are not involved in the routing process, and do not need to receive beacon signals from the MA. Through active network deployment, the number of hops from any sensor to its corresponding MA can be limited to a pre-specified

number. The hop number control ensures efficient system performance, and also makes the quantitative characterization of MC-WSN (in terms of throughput, stability, delay, and energy efficiency) more tractable.

- We present an optimal topology design for MC-WSN such that the average number of hops between a sensor and its nearest sink is minimized.

- We calculate the throughput of MC-WSN considering both single path and multiplath routing between each source and its corresponding sink. More specifically: (i) we analyze the throughput from an information theoretic perspective, and show that as the packet length gets large, the throughput approximately equals to the average normalized information that passes through the channel between a source and its sink; (ii) we illustrate the effect of the number of hops on the throughput, and show that the throughput diminishes exponentially as the number of hops increases; (iii) we show that the throughput of MC-WSN is independent of the physical speed of the MA and the length of its trajectory, and is orders or magnitude higher than that of SENMA.

- We establish the queuing model for the cluster heads based on the *Klienrock independence assumption* and *Burke's theorem*. We prove that the traffic at each CH can be modeled as an independent M/M/1 queue, by showing that: (i) the service process of each queue can be modeled as a Poisson process and is independent from node to node; (ii) the arrival process at each queue can be modeled as a Poisson process. We calculate the arrival rate and service rate of individual CHs, and derive the necessary conditions for the stability of MC-WSN. It is shown that the system stability largely relies on the arrival rate and the throughput, which maps to conditions on scheduling, number of channels, signal to noise ratio (SNR), as well as the bound on the number of CHs that can

be served by each sink (either CCH or RCH).

- We conduct delay analysis for MC-WSN and calculate the average delay for a packet to reach its nearest sink in both single path case and multipath case. It is shown that the hop number control and the network uniformity achieved by MC-WSN can largely simplify the delay analysis.

- We calculate the energy dissipated in the individual sensor nodes, and show that MC-WSN achieves higher energy efficiency over SENMA.

Our analysis are demonstrated through numerical results. It is shown that under stable system conditions, MC-WSN achieves much higher throughput and considerably lower delay and energy consumption over SENMA. Overall, the hierarchical and heterogeneous structure makes MC-WSN a highly resilient, reliable, and scalable architecture. Moreover, the methods used here for network design and analysis provide insight for more general network modeling and evaluation.

3.1.1 Related Work on Network Performance Analysis – Throughput, Stability, and Delay

Throughput performance The network model has a direct impact on the attained performance. For ad-hoc networks with multiple source-destination pairs, it was shown in [22] that the throughput of each node diminishes as the number of nodes in the network increases. The throughput analysis of one-hop many-to-one scenario was considered in [82,83], where multiple nodes directly communicate with a sink under a random multiple access protocol. More specifically, in [82], the throughput, capacity, and stability regions of random multiple access were analyzed under multipacket reception channel model (which allows simultaneous successful receptions of packets). It

was shown that as the packet size gets large, the asymptotic capacity region becomes equal to the throughput region. In [83], the asymptotic stable throughput (i.e., the limiting stable throughput as the number of users goes to infinity) of opportunistic slotted Aloha system was obtained. The results were applied to CDMA-based networks, where an optimal transmission protocol was analyzed in the case of a large spreading gain. Defining the channel utilization as the normalized throughput, i.e., the throughput divided by the maximum throughput, it was shown that slotted Aloha can achieve $1 - O\left(\frac{\log(N_t)}{N_t}\right)$ channel utilization, where N_t is the maximum number of simultaneous transmissions that would satisfy the signal to interference and noise ratio (SINR) requirement.

Throughput of multihop many-to-one single-path regularly-structured canonical network, where there is no merging of different paths/chains connected to the sink, was considered in [23]. It was shown that with the random IEEE 802.11 MAC protocol, the throughput generally does not reach the maximum link capacity. However, proper routing can significantly improve the throughput performance. The throughput of multihop many-to-one network with uniformly deployed nodes and time division multiple access (TDMA) protocol was studied in [24]; it was shown that a throughput of $\frac{W}{n}$ can be achieved in a single-hop many-to-one scenario, however it is generally not achievable in multihop scenarios. The effect of clustering on improving the throughput was also investigated in [24], along with discussions on the trade-offs between throughput and energy efficiency.

Allowing nodes to be mobile could improve the network throughput performance compared to fixed ad-hoc networks. In [28], the effect of the mobility of the nodes on the throughput performance was discussed, considering that the mobile nodes transmit only when they are close to the destination. The main idea in [28] is that for a random source-destination pair, the source splits the packet stream to as many mobile relays

as possible, then a relay delivers the data when it gets close to the destination. Data transmission is made over at most two hops, and the successful reception is determined based on having the SINR value greater than a certain threshold. This approach mainly relies on the diversity as well as the nodes' mobility to improve the throughput. Under this setting, it was shown that the achieved throughput is independent of the number of nodes in the network, and hence is significantly superior to that of fixed ad-hoc networks. Possible limitations of this approach are the large energy consumption and complexity in dynamic scheduling associated with the mobility, the high cost of diversity, the large delay that would depend on the velocity of the nodes [85], and series security issues.

In sensor networks with mobile access points, SENMA, a sensor can only transmit when an access point is within the sensor's communication range [17]. Hence, the throughput is limited by the access point's traversal speed and its trajectory length. In [19], mobile sinks are employed for data collection; each mobile sinks visits limited number of pre-defined collection points in the network. Each sensors routes its information close to the nearest collection point through multihop routing, then data is delivered to the sink that visits the corresponding location. The throughput of this network model was investigated in [19], and the effect of the speed of the sinks and their trajectory length on the throughput was discussed.

Stability and delay performance Stability ensures that the quantity of interest, such as the queue length or delay, is kept within a bounded region or has a limiting distribution [89]. In other words, to ensure a bounded delay performance, the system should be stable. The stability analysis of systems employing Aloha MAC protocol or its variant can be found in [89–92]. Stability region was obtained by means of stochastic dominance [90, 91, 93], where an auxiliary dominant system is modeled in which the

lengths of the queues are larger than, or equal to, that of the original system. Hence, the stability of the dominant system presents sufficient conditions on the stability of the original system. For stable system operation, the arrival rate to each queue in the system should be lower than its service rate [89, 90, 93]. Note that the maximum service rate at a node is equal to the probability of successful transmission, i.e., the throughput.

Delay analysis for wireless networks has been considered in previous work from different perspectives and considering different communication models. In [94], the delay-limited capacity was investigated for a single-hop single transmitter-receiver pair. That is, the analysis was mainly focused on the coding delay, and did not consider the queuing and the multiple access delays. Similar work can be found in [95]. In [91], stability and delay were characterized for simple two-user single-hop network using Aloha multiple access with multipacket reception capability, and the transmission probability that minimizes the delay was obtained.

Analyzing stability and delay in large-scale multihop networks is challenging, due to the complex interactions and dependencies between the nodes/queues along the routing path(s) between a source to its destination. These challenges were highlighted in [88]. There are several existing approaches attempting to analyze the network performance taking into account dependencies between different queues [96–99]. However, most of the work in this area mainly focus on small-scale networks and/or consider special assumptions, such as deterministic service process.

An attempt to find lower bound on the delay for multihop transmissions with multiple source-destination pairs and assuming fixed and equal link capacities can be found in [100]. The approach relies on obtaining a reduced system model, in which each group of adjacent nodes representing a bottleneck (only few nodes can transmit at the same time) are modeled by a single queue. Then, a lower bound on the delay is

obtained based on the reduced system.

Tractable analysis for densely connected networks became possible with the introduction of the *Klienrock independence assumption*, which allows us to model each node in the network separately when there is sufficient merging/multiplexing of data streams at each link [88, 101]. Motivated by the *Klienrock independence assumption*, an end-to-end delay distribution was analyzed in [102] when traffic aggregation is performed at the relay nodes. In [103], the delay of multihop ad-hoc networks was analyzed using the diffusion approximation, where each node is characterized as a separate queue with general distributions for the arrival and service processes (G/G/1 queue). In [104], the end-to-end delay distribution in multihop WSNs was obtained by modeling each hop along a route as a Geo/PH/1/M queue, in which the inter-arrival times are geometrically distributed and the service times are phase-type (PH) distributed. Note that the geometric distribution is the discrete analogue of the exponential distribution. Through empirical results, it was verified that the arrival process can be accurately represented by a Poisson distribution, which has exponentially distributed inter-arrival times.

The delay performance in networks utilizing mobile nodes, such as in [17, 19, 28], will generally depend on the velocity of the corresponding nodes [85]. Hence, these network models would be inefficient for time-sensitive applications. In this chapter, after describing the proposed MC-WSN architecture and topology design, we present the network analysis. As will be shown, the MC-WSN performance is independent of the physical speed of the MA.

3.2 The Proposed Mobile Access Coordinated Wireless Sensor Network (MC-WSN)

In this section, we describe the proposed MC-WSN architecture and highlight its major features.

3.2.1 General Description

We assume the network is divided into cells of radius d_c. Each cell contains a single powerful mobile access point (MA) and n uniformly deployed sensor nodes (SNs) that are arranged into N_{CH} clusters. Each cluster is managed by a cluster head (CH), to which all the cluster members report their data. CHs then route the data to the MA [43, 78]. A powerful center cluster head (CCH) is employed in the middle of each cell, and K powerful ring cluster heads (RCH) are placed on a ring of radius R_t. The CCH and RCHs can establish direct communication with the MA or with other RCHs that are closer to the MA. All nodes within a distance R_o from the CCH route their data to the MA through the CCH. All other nodes route their data to the MA through the nearest RCH. If a sensor is within the MA's coverage range, then direct communications can take place when permitted or needed. After receiving the data of the sensors, the MA delivers it to a Base Station (BS). The overall network architecture is illustrated in Figure 3.1. As will be illustrated in Section 3.3, the number of hops from any sensor to the MA can be limited to a pre-specified number through the deployment of CCH and RCHs.

In the proposed MC-WSN architecture, the MA coordinates the sensors and resolves the node deployment issue as well as the energy consumption problem of wireless sensor networks. More specifically, the MAs are responsible for: (i) deploying nodes, (ii) replacing and recharging nodes, (iii) detecting malicious sensors, then removing and

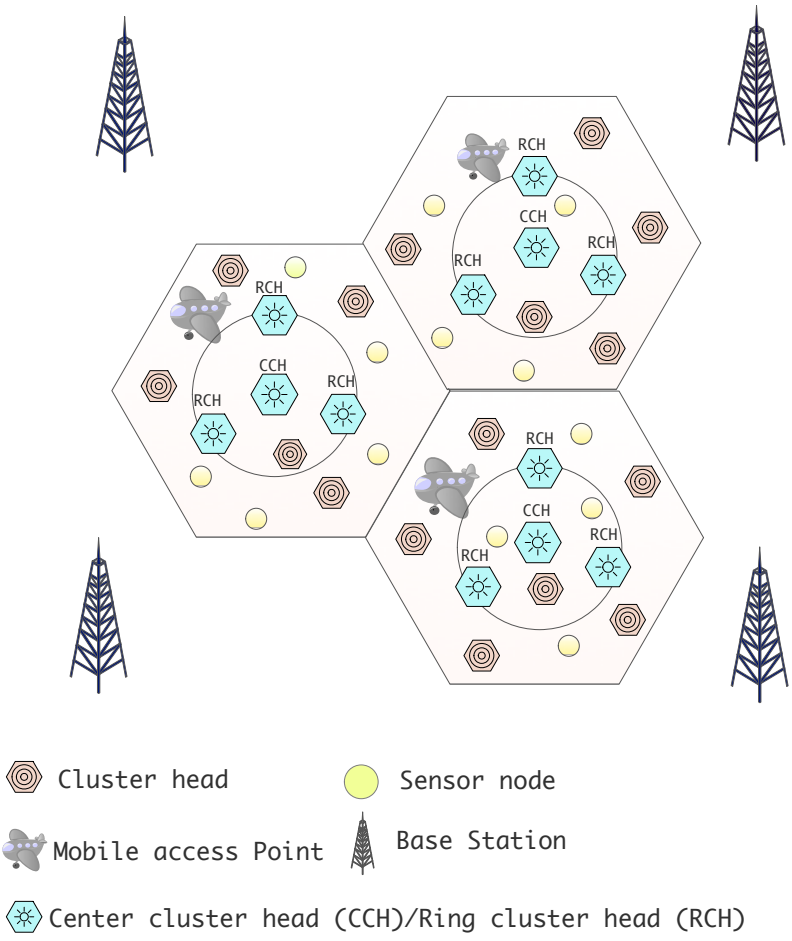


Figure 3.1: Proposed MC-WSN architecture.

replacing them, (iv) collecting the information from sensors and delivering it to a BS.

When an MA needs to be recharged or reloaded, it sends a request to the MA base. The base will send a new MA to the cell, and the substituted MA will be called back to the base for maintenance services. The MAs can move on the ground, and can also fly at low altitude. Each MA traverses its cell mainly for replacing or recharging low-energy sensor nodes and cluster heads, as well as removing the malicious nodes. The recharging can be performed in a wireless manner [105]. The MA moves physically for data collection only in the case when the routing paths do not work.

Data collection from the sensors can be event based or periodic. Data transmissions

from SNs to CHs, between CHs, and from CCH/RCHs to the MA are made over different channels to avoid interference between different communication links. Let the communication range of each sensor node and CH be r_c and R_c, respectively. CHs have larger storage capacity and longer communication range than SNs, i.e., $R_c > r_c$. We assume shortest path routing between the CHs and the CCH/RCHs. Note that the sensors are not involved in the inter-cluster routing to minimize their energy consumption.

Due to the MA-assisted active network deployment, we can assume that the nodes are uniformly distributed in the network. It is therefore reasonable to place the powerful RCHs at evenly spaced locations on the ring R_t. To maximize the throughput and minimize the delay of data transmission from the sensors to the MA, the number of hops needed in routing should be minimized. In Section 3.3, we discuss network topology design and obtain the optimal R_t and R_o that minimize the number of hops.

3.2.2 Major Features

The main advantages of MC-WSN lie in: (i) multi-functionality of the mobile access; (ii) hop number control through topology design; and (iii) hierarchical and heterogeneous node deployment. More specifically, MC-WSN has the following features:

- *Controlled network development and prolonged network lifetime* The proposed MC-WSN allows the MAs to manage the deployment of SNs and CHs. That is, the MA can add more nodes, relocate or replace exiting nodes. In addition, it can recharge or replace low-energy nodes. When a node has low remaining energy, it sends a control message to the MA notifying it with its energy level. The MA can then check and make the decision to replace the node or recharge it. Being coordinated by the MA, the MC-WSN architecture resolves the network deployment issue and can actively prolong the network lifetime.

- *Time-sensitive data transmission* In conventional SENMA, a transmission is made only if an MA visits the corresponding source node; thus, data transmission is limited by the physical speed of the MAs and the length of their trajectory, resulting in low throughput and large delay. In MC-WSN, the delay is effectively managed through hop number control, and is independent of the physical speed of the MA.

- *Enhanced network security* First, the MAs can detect malicious SNs and CHs and replace them [45]. When the MA receives data from a node, it first authenticates the source and checks its identity. If the source passes the authentication procedure, the MA monitors the reports of each individual node and compares it with the final decision obtained through data fusion. Based on the observations over multiple sensing periods, the malicious nodes can be detected and removed [42]. Second, with hop number control, the delay from a sensor to the MA is limited within a pre-specified time duration under regular network conditions. If the actual delay is significantly larger, then an unexpected network event or network failure is detected. Third, it is difficult to get the MA itself compromised or destroyed, since it is much more powerful than other network nodes, and it moves randomly in the network where its location can be kept private [106].

- *Efficient energy consumption* The SNs have the most limited resources in wireless sensor networks. In the proposed MC-WSN, SNs only communicate with their nearest CHs, and are not involved in any inter-cluster routing. Also, unlike SENMA, SNs in MC-WSN do not need to receive the periodic beacon signal from the MA, and hence the energy efficiency is further improved. Note that the beacon signals in SENMA are used to notify SNs of the presence of the MA and

to indicate which sensor to transmit.

- *Enhanced network resilience, reliability and scalability:* MC-WSN is a self-healing architecture, where the CCH and RCHs represent different options for data transmission to the MA. The diversity in multipath routing increases the resilience of the network. In the worst case when the routing paths do not work, the MA can traverse its cell for data collection. Overall, the hierarchical and heterogeneous structure makes the MC-WSN a highly resilient, reliable, and scalable architecture.

3.3 Network Topology Design

In this section, we investigate network topology design of MC-WSN, and calculate the optimal radius R_o and the ring radius R_t that minimize the average number of hops from any CH to the MA. Note that under shortest path routing, the number of hops is proportional to the distance between the source and the sink. To minimize the number of hops, *we design the topology such that the average distance between a cluster head and its nearest sink is minimized.*

In the proposed MC-WSN architecture, the average squared distance between any source and the corresponding sink (CCH/RCH) can be expressed as:

$$\begin{aligned}\bar{d}^2 &= 2K\left[\int_{\theta=0}^{\pi/K}\int_{x=0}^{R_o} x^2 f_X(x) f_\theta(\theta) dx d\theta + \right. \\ &\quad \int_{\theta=0}^{\pi/K}\int_{x=R_0}^{R_t}\left[x^2 - 2xR_t\cos(\theta) + R_t^2\right] f_X(x) f_\theta(\theta) dx d\theta + \\ &\quad \left.\int_{\theta=0}^{\pi/K}\int_{x=R_t}^{d_c}\left[x^2 - 2xR_t\cos(\theta) + R_t^2\right] f_X(x) f_\theta(\theta) dx d\theta\right], \end{aligned} \tag{3.1}$$

where x is the distance from any CH to the center of the cell, and θ is the angle from

the CCH, as illustrated in Figure 3.2. Here, $f_X(x)$ is the PDF of x. Assuming that the CHs are uniformly distributed in a circle of radius d_c, then $f_X(x)$ can be approximated by $f_X(x) = \frac{2x}{d_c^2}$, and the PDF of θ is modeled as $f_\theta(\theta) = \frac{1}{2\pi}$, $\forall \theta \in [0, 2\pi]$.

Recall that K is the number of RCHs. Assume $K > 1$, and set

$$\frac{\partial \bar{d}^2}{\partial R_o} = 0, \quad \frac{\partial \bar{d}^2}{\partial R_t} = 0. \tag{3.2}$$

We get the optimal $R_o = \frac{\pi R_t}{2K \sin(\frac{\pi}{K})}$, and $R_t = \frac{\sqrt{3}-1}{\pi} K \sin(\frac{\pi}{K}) d_c = 0.233K \sin(\frac{\pi}{K}) d_c$. It follows that $R_o = 0.366 d_c$. In summary, we have the following result.

Proposition 3.1 *Assuming a circular cell of radius d_c, to minimize the number of hops in the MC-WSN architecture with one CCH and K RCHs, where $K > 1$, data transmission should be arranged as follows: (1) The CHs within a distance $R_o = 0.366\ d_c$ from the center of the cell deliver their data to the MA through the CCH. (2) The CHs at a distance x from CCH, where $R_o \leq x < d_c$, deliver their data to the MA through the nearest RCH on the ring of radius $R_t = 0.233K \sin(\frac{\pi}{K}) d_c$.*

With the optimal topology, the average squared distance from a CH to its nearest sink is $\bar{d}^2 = 0.5 d_c^2 - 0.047 d_c^2 K^2 [\sin(\frac{\pi}{K})]^2$. Assuming shortest path routing is available, the average number of hops can be expressed as $N_{hop} = \frac{\bar{d}}{R_c}$, where R_c is the communication range of the cluster heads. Note that as K increases, $\bar{d}$ and consequently N_{hop} decrease. As can be seen, the maximum number of hops can be limited to a pre-specified number through the deployment of RCHs.

3.4 Throughput Analysis

In this section, we analyze the throughput of the multihop MC-WSN architecture. After introducing the definition of the throughput in the single hop case, we analyze

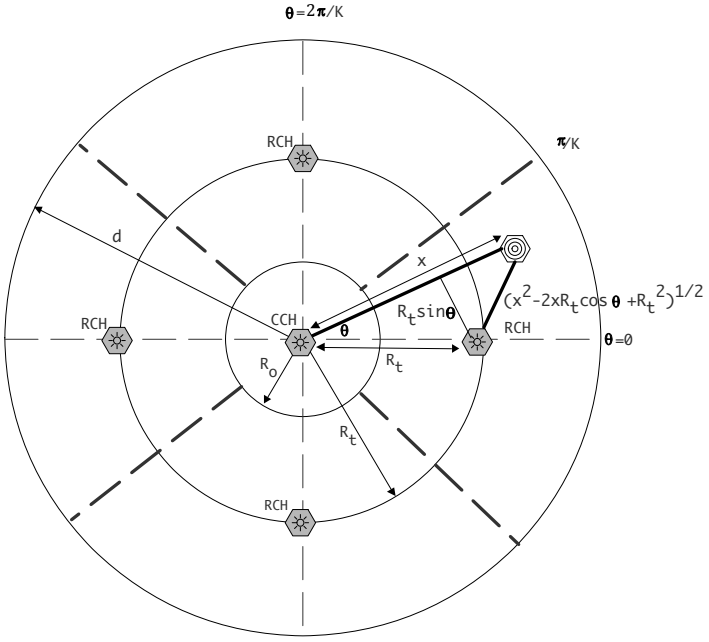


Figure 3.2: MC-WSN with four powerful RCHs.

the multihop throughput under both single path and multipath routing.

3.4.1 Definition of the Throughput

We start with the single hop case. Assuming node i is transmitting to sink k, where $k \in \{0, 1, ..., K\}$. *The throughput of node i to sink k*, $T_{i,k}$, is defined as the average number of packets per slot that are initiated by node i and successfully delivered to the intended receiver k [92]. Define $R_S^k(\nu)$ as the set of nodes that have their packets successfully delivered to sink k in slot ν, where S is the set of nodes scheduled to transmit. Then, $T_{i,k}$ can be expressed as:

$$\begin{aligned} T_{i,k} &= E\left[\lim_{T\to\infty}\frac{1}{T}\sum_{\nu=1}^{T} I[i \in R_S^k(\nu)]\right] \\ &= \lim_{T\to\infty}\frac{1}{T}\sum_{\nu=1}^{T} Pr\{i \in R_S^k(\nu)\}, \end{aligned} \tag{3.3}$$

where $I(.)$ is the indication function.

Let t_i^k be a binary flag indicating that node i transmits data to sink k: $t_i^k = 1$ means that sensor i is scheduled to transmit its data to the sink k, otherwise $t_i^k = 0$. Similarly, let r_i^k be a binary flag indicating that the data of node i is successfully received at the intended destination k (CCH or RCH). Note that the transmission from the powerful CCH/RCH to the MA can be made at high-power and high-rate. Also, with the active network deployment performed by the MA, the data from each sensor to its CH can be transmitted over a single hop using a collision-free MAC protocol. Thus, we focus on data transmission from the CH of the originating node to its corresponding CCH/RCH. Assume that the packet reception from slot to slot is an i.i.d process, then it follows that:

$$T_{i,k} = Pr\{r_i^k = 1 | t_i^k = 1\} Pr\{t_i^k = 1\}. \tag{3.4}$$

In the following, we analyze $T_{i,k}$ from the information theory perspective, by discussing the relationship between $T_{i,k}$ and the mutual information between the packet transmitted from CH i and the packet received at sink k.

For each slot, define X_i^k as the transmitted packet from CH i to sink k, where $X_i^k = 0$ means that node i is not transmitting. Let $\tilde{X}_i^k$ be the non-zero packets of X_i^k, then $X_i^k = t_i^k \tilde{X}_i^k$ [82]. Assuming that sink k receives packets from multiple nodes in a collision-free manner. Define $\mathbf{Y}^k$ as the received vector at sink k, where the ith element in $\mathbf{Y}^k$ is the received packet from CH i. Let $\mathbf{r}^k$ be the vector whose ith element is r_i^k. It has been shown in [82] that the mutual information between X_i^k and $\mathbf{Y}^k$ can be written as a function of the throughput of CH i to sink k ($T_{i,k}$) as follows:

$$\mathbf{I}(X_i^k, \mathbf{Y}^k) = \mathbf{I}(t_i^k, \mathbf{r}^k) + H(\tilde{X}_i^k) T_{i,k}, \tag{3.5}$$

where $\mathbf{I}(x, y)$ is the mutual information between x and y, and $H(x)$ is the entropy

of x. Let $\mathbf{I}_p^k = \mathbf{I}(X_i^k, \mathbf{Y}^k)/H(\tilde{X}_i^k)$, which is measured in number of packets per slot. In general, $T_{i,k} \leq \mathbf{I}_p^k$. Note that t_i^k is binary, i.e., $H(t_i^k) \leq 1$, which implies that $\mathbf{I}(t_i^k, \mathbf{r}^k) \leq H(t_i^k) \leq 1$. As a result, if the packet length gets large, i.e., $H(\tilde{X}_i^k) \to \infty$, then we have $T_{i,k} \simeq \mathbf{I}_p^k$.

From the information theory perceptive, this shows that: $T_{i,k}$ is the average normalized information (measured in packets per slot) passed through the channel between CH i and sink k.

3.4.2 Multihop Single Path Routing Case

In this subsection, we analyze the throughput of a node in the case when there is a pre-defined, multihop single path from each CH to its corresponding sink.

Consider that CH i requires N_i^k hops to reach sink k. N_i^k is based on the network architecture, topology, and routing scheme. Let the ideal or shortest path from CH i to sink k be $i_{N_i^k} \to i_{N_i^k-1} \to ...i_1 \to i_0$, where $i_{N_i^k}$ is the source CH i and i_0 is the sink k. This is illustrated in Figure 3.3. Let $t_{i,h}^k$ be a binary flag at hop h, indicating that CH i_h is scheduled to relay a packet of CH i to CH i_{h-1} along the route to sink k. Also, let $r_{i,h}^k$ be a binary flag indicating that the data of CH i is successfully received at CH i_{h-1} along the same route to sink k. It follows that, at each particular time slot, we have:

$$Pr\{r_{i,h}^k = 1\} = Pr\{r_{i,h}^k = 1 | t_{i,h}^k = 1\} Pr\{t_{i,h}^k = 1\}. \tag{3.6}$$

Consider that a packet of CH i is received at sink k in slot ν. This implies that there exists a scheduling slot vector $\boldsymbol{\nu} = [\nu - \Delta\nu_{N_i^k-1}, ..., \nu - \Delta\nu_1, \nu]$, such that all nodes along the routing path from i to the sink successfully transmit the packet of node i. More specifically, node i_h is scheduled to transmit in slot $\nu - \Delta\nu_{h-1}$, where

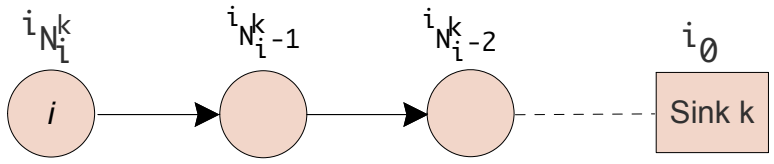


Figure 3.3: Multihop single path between node i and sink k.

$\Delta\nu_x > \Delta\nu_y$, $\forall x > y$ and $\Delta\nu_0 = 0$. Along slot vector $\boldsymbol{\nu}$, define the transmission flag of CH i as $t_i^k(\boldsymbol{\nu})$, such that $t_i^k(\boldsymbol{\nu}) = [1, .., 1]$ when CH i transmits a packet to sink k and the transmission at the last hop (at CH i_1) occurs in slot ν. Note that if the relay at the last hop along the transmission path from i to the sink transmits the packet of node i, then it implies that all intermediate hops were scheduled to transmit in prior slots. That is, we have

$$Pr\{t_i^k(\boldsymbol{\nu}) = 1\} = Pr\{t_{i,1}^k(\nu) = 1, ..., t_{i,N_i^k}^k(\nu - \Delta\nu_{N_i^k - 1}) = 1\}. \tag{3.7}$$

Omit the slot index, (3.7) can be simplified as: $Pr\{t_i^k = 1\} = Pr\{t_{i,1}^k = 1, ..., t_{i,N_i^k}^k = 1\}$.

For the throughput calculation here, we do not consider retransmissions of packets. Assuming that there exists a schedule such that the source CH and all its intermediate relays are assigned time slots to transmit/forward the source's data, and assuming that the transmissions in all slots are i.i.d, then we can drop the slot index from the throughput expression. In the case when the amplify-and-forward protocol is adopted in the relaying process, which implies that $r_{i,h}^k$'s are independent at different hops, it

follows from (3.4) and (3.6) that:

$$\begin{aligned} T_{i,k} &= Pr\{t_{i,1}^k = 1, ..., t_{i,N_i^k}^k = 1\} \prod_{h=1}^{N_i^k} Pr\{r_{i,h}^k = 1 | t_{i,h}^k = 1\}, \\ &= Pr\{t_i^k = 1\} \prod_{h=1}^{N_i^k} Pr\{r_{i,h}^k = 1 | t_{i,h}^k = 1\}. \end{aligned} \tag{3.8}$$

Note that if decode-and-forward is employed at the intermediate CHs instead of the amplify-and-forward, then the errors in one hop can be corrected at another hop experiencing better channel conditions. This is at the expense of increased complexity and delay at all hops.

It is noted from equation (3.8) that the throughput depends on the employed PHY, MAC, routing protocols as well as the network environment. t_i^k is related to the MAC protocol, while r_i^k is related to the PHY protocol. The routing protocol determines the path and the number of hops from a source to its destination.

Denote N_{intf} as the minimum separation between links for bandwidth reuse. That is, when a transmission is made by a CH, other nodes within a distance of $N_{intf}R_c$ from the transmitting CH should remain silent or use another orthogonal channel. Let n_k be the number of nodes connected to sink k. Then we have the following result.

Lemma 3.1 *When TDMA is used, each node connected to sink k can transmit with a probability* $P(t_i^k = 1) \geq \frac{1}{N_{intf}\ n_k}$. *If hybrid TDMA/FDMA is used, and* N_{Freq} *is the number of frequencies available for simultaneous CHs transmissions within the same interference region, then* $P(t_i^k = 1) \geq \frac{N_{Freq}}{N_{intf}\ n_k}$.

Proof: The proof is provided in Appendix A. □

We now evaluate *the probability of successful reception*, which can be viewed as

a condition on the signal to interference and noise ratio SINR. Let P_i be the power of node i that is exponentially distributed with mean $\bar{P}_i$. That is, $Pr\{P_i = x\} = \bar{P}^{-1} \exp\left\{-\bar{P}^{-1}x\right\}$. Assume $\bar{P}_i = \bar{P}\ \forall i$. Suppose a transmission is made from l_i to l_j, where l_i and l_j are the locations of the transmitting and receiving nodes, respectively, and $L_{i,j} = |l_i - l_j|$ is the distance between them. The SINR in the transmission from i to j, $SINR_{i,j}$, can be expressed as $SINR_{i,j} = \frac{L_{i,j}^{-\beta} P_i}{N_o + \sum_{\substack{x \in X^i \\ x \neq i}} L_{x,j}^{-\beta} P_x}$, where N_o is the noise power, X^i is the set of all radios transmitting on the same channel and in the same time slot as node i, and $\beta \geq 2$ is the path loss exponent ($\beta = 2$ in free space environment). In structured networks, the assignment of channels and time slots can be managed to minimize the interference. In this case, the interference term becomes negligible, and we get $SINR_{i,j} = \frac{L_{i,j}^{-\beta} P_i}{N_o}$. Hence, we use $SINR$ and SNR interchangeably.

We can write

$$Pr\{r_{i,h}^k = 1 | t_{i,h}^k = 1\} = Pr\{SINR_{i_h, i_{h-1}} > \gamma\}, \tag{3.9}$$

where γ is the SINR threshold for successful transmission. Note that if the transmitter power is fixed and is affected by a Rayleigh fading channel, the received power will be exponentially distributed [107]. In other words, this model is equivalent to having a fixed-power transmitted signal passing through a Rayleigh fading channel. In both cases, the *received SINR will be exponentially distributed* [108]. Define $\lambda_{i,h} = \gamma N_o \left[L_{i_h, i_{h-1}}\right]^{\beta}$ as the minimum transmit power of node i_h to guarantee the SINR

threshold at hop $h-1$. We have,

$$\begin{aligned} Pr\{SINR_{i_h,i_{h-1}} > \gamma\} &= Pr\{P_{i_h} > \lambda_{i,h}\} \\ &= \int_{s=\lambda_{i,h}}^{\infty} \frac{1}{\bar{P}} \exp\left\{-\frac{1}{\bar{P}}s\right\} ds \\ &= \exp\left\{-\gamma \frac{N_o}{\bar{P}} \left[L_{i_h,i_{h-1}}\right]^{\beta}\right\}. \end{aligned} \tag{3.10}$$

Note that the average SNR at hop h can be expressed as: $\overline{SNR}_h = \frac{\bar{P}\left[L_{i_h,i_{h-1}}\right]^{-\beta}}{N_o}$. If $L_{i_h,i_{h-1}} = L \; \forall h$, then $\overline{SNR}_h = SNR$ and $Pr\{SINR_{i_h,i_{h-1}} > \gamma\} = \exp\left\{-\frac{\gamma}{SNR}\right\} \; \forall h$. From (3.8) - (3.10), we get

$$\begin{aligned} T_{i,k} &= Pr\{t_i^k = 1\} \prod_{h=1}^{N_i^k} \exp\left\{-\gamma \frac{N_o}{\bar{P}} \left[L_{i_h,i_{h-1}}\right]^{\beta}\right\} \\ &= Pr\{t_i^k = 1\} \exp\left\{-\gamma \frac{N_o}{\bar{P}} \sum_{h=1}^{N_i^k} \left[L_{i_h,i_{h-1}}\right]^{\beta}\right\}. \end{aligned} \tag{3.11}$$

Theorem 3.1 *In a multihop MC-WSN network, assuming exponentially distributed transmit powers, the throughput of CH i along a predefined single routing path to sink k is:*

$$T_{i,k} = Pr\{t_i^k = 1\} \exp\left\{-\kappa \sum_{h=1}^{N_i^k} \left[L_{i_h,i_{h-1}}\right]^{\beta}\right\}, \tag{3.12}$$

where N_i^k is the number of hops in CH i's transmission, $Pr\{t_i^k = 1\}$ is the probability that CH i and all its intermediate relaying nodes are scheduled to transmit the data of CH i to sink k, β is the path loss exponent of the channel, $L_{x,y}$ is the distance between nodes x and y, and $\kappa = \gamma \frac{N_o}{\bar{P}}$.

Remark 3.1 *It can be seen from* Theorem 3.1 *that if the hops are equidistant, the*

throughput will decrease as the number of hops increases. More specifically, when $L_{i_{h-1},i_h} = L,\ \forall h \in \{1, 2, .., N_i^k\}$*, we get* $T_{i,k} \propto \exp\{-N_i^k\}$*. It follows that:*

$$\lim_{N_i^k \to \infty} T_{i,k} = 0. \tag{3.13}$$

This result justifies our motivation of limiting the number of hops from each sensor to the MA to a pre-specified number through the topology design and deployment of CCH and RCHs. With hop number control, we can have better control and management over the system throughput, delay, security, and energy efficiency.

Remark 3.2 *It is worth mentioning that if the distance between the source and the sink is fixed, then larger number of hops would correspond to lower per-hop distance, and consequently resulting in an improved performance at low* SNR *values. However, this would request higher node density, and hence an increase in the number of nodes in each cell. In this chapter, under the assumption that the number of nodes in each cell is fixed, we will mainly consider the case of fixed per-hop distance.*

Now we obtain the overall average per node throughput. Define P_{A_k} as the probability that a cluster head lies in the coverage area of sink k. That is, its nearest sink is sink k. Following *Lemma 3.1*, we set $P(t_i^k = 1) = \frac{N_{Freq}}{N_{intf}\ n_k}$, which is a conservative measure for the per node transmission probability. Recall that N_{CH} is the total number of CHs, then the number of CHs that transmit to sink k is $n_k = P_{A_k} N_{CH}$. Hence, the overall average per node transmission probability in the cell, $\bar{P}_t$, can be expressed as:

$$\begin{aligned} \bar{P}_t &= \sum_{k=0}^{K} P_{A_k} \frac{N_{Freq}}{N_{intf}\ n_k} = \sum_{k=0}^{K} P_{A_k} \frac{N_{Freq}}{N_{intf} P_{A_k} N_{CH}} \\ &= (K+1) \frac{N_{Freq}}{N_{intf}\ N_{CH}}, \end{aligned} \tag{3.14}$$

where N_{intf} is the bandwidth reuse measure, and N_{Freq} is the number of frequencies available for simultaneous cluster head transmissions. For equidistant hops with length R_c, the overall average per node throughput is expressed as

$$\bar{T} = \bar{P}_t \exp\left\{-\kappa N_{hop} R_c^{\beta}\right\}, \tag{3.15}$$

where N_{hop} is the average number of hops from a CH to its corresponding sink in each cell, and is obtained in Section 3.3.

3.4.3 Multihop Multipath Routing Case

In the previous subsection, we considered the case when there is a single pre-defined path between a CH and a sink. Note that, in general, the transmission can go through different paths due to the existence of network diversity. In this section, we formulate the throughput for the multipath case. We have the following result:

Theorem 3.2 *Let N be the maximum number of hops from a CH to its sink along any routing path. Consider that for each hop number $l \in \{1, 2, ..., N\}$, there are $P_{i,l}$ possible l-hop paths from CH i to sink k. Let $T(i|N_i^k = l, \mathcal{P}_i^k = p)$ be the throughput that can be achieved along one of the l-hop paths from source i to sink k assuming the path $\mathcal{P}_i^k = p$, then the throughput of node i can be calculated as:*

$$T_{i,k} = \sum_{l=1}^{N} \sum_{p=1}^{P_{i,l}} T(i|N_i^k = l, \mathcal{P}_i^k = p) \Pr\{\mathcal{P}_i^k = p|N_i^k = l\} \Pr\{N_i^k = l\}. \tag{3.16}$$

Here, l-hop path means a path that consists of l hops. It is noted that $T(i|N_i^k = l, \mathcal{P}_i^k = p)$ can be obtained from *Theorem 3.1* by substituting $N_i^k = l$, which is the number of hops along the particular path $\mathcal{P}_i^k = p$. The term $\Pr\{\mathcal{P}_i^k = p|N_i^k = l\}$ depends on the routing protocol. It should be emphasized that when multiple routes

are enabled, the utilized scheduling protocol, and hence $P(t_i^k = 1)$, could be different than that in the single routing path case.

3.4.4 Total Network Throughput

The *network throughput*, Υ, is defined as the average number of packets received successfully from all clusters per unit time.

Let $\mathcal{N}^k$ be the set of CHs that transmit to sink k. Following *Theorems 3.1* and *3.2*, the total throughput of the proposed MC-WSN architecture with K RCHs and a CCH can be obtained as:

$$\begin{aligned}
\Upsilon &= \sum_{k=0}^{K} \sum_{i \in \mathcal{N}^k} T_{i,k} \\
&= \sum_{k=0}^{K} \sum_{i \in \mathcal{N}^k} \sum_{l=1}^{N} \sum_{p=1}^{\mathcal{P}_{i,l}} T(i|N_i^k = l, \mathcal{P}_i^k = p) \Pr\{\mathcal{P}_i^k = p | N_i^k = l\} \Pr\{N_i^k = l\} \\
&= \sum_{k=0}^{K} \sum_{i \in \mathcal{N}^k} \sum_{l=1}^{N} \sum_{p=1}^{\mathcal{P}_{i,l}} p_i^k(p) \exp\left\{ -\kappa \sum_{h=1}^{l} \left[L_{i_h^k, i_{h-1}^k}(p) \right]^{\beta} \right\} \Pr\{\mathcal{P}_i^k = p | N_i^k = l\} \\
&\qquad \times \Pr\{N_i^k = l\}, \qquad (3.17)
\end{aligned}$$

where n_k is the number of nodes connected to sink k, $L_{i_h^k, i_{h-1}^k}(p)$ is the length between CHs i_h^k and i_{h-1}^k along path p, and $p_i^k(p)$ is the transmission probability of CH i along path p to sink k.

3.5 System Stability and Delay Analysis

In this section, we analyze the stability and delay of MC-WSN by exploiting tools in queuing theory. After introducing the independence assumption and modeling theorems, we establish the queuing model of the CHs, and then perform the stability and

delay analysis.

3.5.1 Queue Independence Assumption and Modeling Theorems

The difficulty in the system stability and delay analysis in communication networks is mainly attributed to the dependency between different queues along the routing path of a packet. Let the arrival time of packet j at a queue be T_j. Then the *inter-arrival time* between packets j and $j+1$ is $A_j = T_{j+1} - T_j$. The *service time* at a node is generally defined as the duration between the time the packet is at the head of the node's queue until it is successfully transmitted. In other words, the service time equals to the packet length divided by the service rate.

3.5.1.1 Klienrock Independence Assumption

In networks of tandem queues, there is generally a correlation between the inter-arrival times and the packet lengths/service times at the intermediate queues [88]. For example, if the packets retain their lengths when they are forwarded at different hops, considering that the link rates are fixed, then we have: (i) dependency between the inter-arrival times and service times at each of the intermediate queues; (ii) dependency between service times at different queues. Due to these dependencies, network analysis becomes highly complicated and intractable.

However, it was observed that when the network is densely connected with moderate to heavy traffic loads, these dependency can be removed [101]. In other words, the dependencies between the inter-arrival times and service times can largely be eliminated by merging multiple packet streams on each link [88]. This is known as the *Klienrock independence assumption. More specifically, in a densely connected network*

with moderate to heavy traffic loads, if we have Poisson arrival processes at the entry points of the network, and exponentially distributed service times at each link, then the multiplexing of the independent Poisson packet streams at every node has the effect similar to restoring the independence between the inter-arrival times and service times [88].

The underlying argument is that: if packets received by a node from different sources are ordered in the queue by the order they arrive in a first-come first-served manner, the resulted queues through the packet multiplexing/merging process become independent. The idea here is similar to the interleaving process in communication systems, which randomizes consecutive symbols and validates the independence assumption among all the symbols.

The independence assumption was verified through experiments in [88] using different network topologies (star, diamond, and k-connect networks) under uniform and non-uniform traffic. It was shown that *the independence assumption provides a valid model for network analysis*. It should be noted that, having an exponentially distributed packet lengths and deterministic service process is equivalent to having fixed packet lengths and Poisson service process. Both cases will result in an exponentially distributed service time. The independence assumption allows us to treat each node in the network independently as an M/M/1 queue [88], and hence enables tractable network analysis.

3.5.1.2 Burke's Theorem and Little's Theorem

Next, we introduce two important queue modeling and analyzing theorems that will be used in our analysis in the following subsections. The first one is the *Burke's theorem* [101], which describes the relationship between the arrival flow and the service flow.

***Burke's theorem*:** *Consider an M/M/1, M/M/m, or M/M/∞ system with Poisson arrival process of rate λ_x, then the departure process is Poisson with rate λ_x.*

The second one is the well-known *Little's theorem* [101], which formulates the average delay per packet as a function of the average arrival rate and the number of packets in the system.

***Little's Theorem*:** *Let the steady state average number of packets in a system be N_x and the average packet arrival rate be λ_x, then the average delay per packet in the system $D_x = \frac{N_x}{\lambda_x}$.*

In the next subsection, we characterize the queuing model of cluster heads in MC-WSN.

3.5.2 Queuing Model Characterization for MC-WSN

3.5.2.1 Modeling the Arrival and Service Processes

In this subsection, we provide the queuing model for each individual CHs in MC-WSN, and show that the *Klienrock independence assumption* provides an accurate model for stability and delay analysis of the MC-WSN network. More specifically, we have the following result:

Theorem 3.3 *(i) The service process of each queue can be modeled as a Poisson process and is independent from node to node. (ii) The arrival process at each queue can be modeled as a Poisson process.*

Proof: (i) Service Process: *Due to the exponentially distributed SNR*, different links in the MC-WSN multihop transmissions have different service times that are independent from link to link. Recall that the probability of successful transmission (i.e., the throughput) between node i and j is $T_{i,j}$. Then, the number of packets successfully transmitted from i to j in c slots can be modeled as a Binomial random

variable with parameters c and $T_{i,j}$ [109]. According to the *law of small numbers*, when large time interval is considered, i.e., $c \to \infty$, the Binomial distribution with parameters c and $T_{i,j}$ converges to a Poisson distribution with parameter $S = cT_{i,j}$ [110].

(ii) Arrival process: All the CHs can be divided into two groups: (a) CHs that only transmit packets generated from their own clusters. (b) CHs that serve as relays for other CHs, and hence transmit their generated traffic and also the relay traffic. Without loss of generality, consider two CHs i and j, where CH i receives data from its cluster members (sensors) only, while CH j receives data from its cluster members as well as from CH i. Note that in general the aggregation of several independent and identically distributed traffic can be accurately approximated as a Poisson process [101] (p. 165). Hence, the arrival process of packets from sensors to their corresponding CHs can be modeled as a Poisson process. That is, CH i has a Poisson arrival process.

Next, we will show that CH j has an overall Poisson arrival process as well. Since the service process from each CH is Poisson, therefore CH i is an M/M/1 queue. It follows from the *Burke's theorem* that the departures process of CH i is Poisson distributed. The Poisson departures from CH i arrive at CH j and are merged with data from sensors in cluster j, which is also Poisson. Since the summation of independent Poisson process of rates $\{\lambda_1, ..., \lambda_n\}$ is a Poisson process of rate $\lambda_t = \sum_i^n \lambda_i$ [101], then the overall arrival process at CH j has a Poisson distribution. This proof can be directly extended to CHs that serve as a relay for more than one CH. □

Based on the discussions above, each CH in the network can be modeled as an independent M/M/1 queue. Our stability and delay analysis are based on this model.

3.5.2.2 Calculation of Arrival and Service Rates

Here, we calculate the arrival and service rates of CHs by considering different traffic loads at the CHs in the network. To do this, we first group the CHs based on their

locations and the number of hops to their corresponding sink (either the CCH or an RCH).

- For the CCH: Due to the uniformity of the MC-WSN structure achieved by the MA, it is reasonable to assume that all CHs at the same hop level from the CCH carry approximately the same amount of traffic. Hence, for delay analysis, we do not distinguish between nodes within the same hop level from the CCH.

- For the RCHs: The traffic around the RCH could be different due to the unequal areas. More specifically, within a particular RCH coverage area (illustrated in Figure 3.2), the outer region, where $x > R_t$, and the inner region, where $R_o < x \leq R_t$, have different traffic loads. This is because the area of the outer region is larger than that of the inner region, which corresponds to larger number of hops and more CHs in the outer region. Therefore, when analyzing the performance of the CHs around the RCH, we identify the nodes by their hop level as well as their region from the RCH (inner or outer region). Nodes within the same hop level of a particular region from a RCH are not distinguished.

From the discussions above, without loss of generality, we define the following:

- $g^O_{h,k}$ is the group of nodes in the hth hop level from sink k and in the outer region. Similarly, $g^I_{h,k}$ is the group of nodes in the hth hop level from sink k and in the inner region. The superscript O and I are omitted when referring to the CCH.

- $\lambda^O_{i,h,k}$ and $\lambda^I_{i,h,k}$ are the total arrival rates at CH $i \in g^O_{h,k}$ and $i \in g^I_{h,k}$, respectively.

- $s^O_{i,h,k}$ and $s^I_{i,h,k}$ are the service rates at CH $i \in g^O_{h,k}$ and $i \in g^I_{h,k}$, respectively.

Take a CH at the hth hop from the outer region of sink k as an example. Based on the independence assumption, it can be modeled as an M/M/1 queue with total arrival

rate $\lambda^O_{i,h,k}$ and service rate $s^O_{i,h,k}$.

The total arrival rate at a CH is the sum of the arrivals of packets from its own cluster members and the arrivals of packets forwarded from other cluster heads to be delivered to the nearest sink. We refer to the former as the "*generated arrival rate*", denoted by $\tilde{\lambda}_{g,i}$, while the latter is referred to as "*forwarded arrivals rates*", denoted by $\tilde{\lambda}^O_{f,i,k}$ or $\tilde{\lambda}^I_{f,i,k}$, depending on where the CH resides. Following our discussions in the previous subsection, we assume that the traffic generated from each cluster in the network follows a Poisson process with equal rates, and is independent of the hop level or the location in the network. That is, $\tilde{\lambda}_{g,i} = \lambda \ \forall i$. In the following, we consider the analysis of CHs in the outer regions of the sinks (RCHs). The analysis of CHs in the inner regions as well as those within the coverage area of the CCH can be performed in a similar manner.

We characterize the forwarded traffic to CH $i \in g^O_{h,k}$ based on the *Burke's theorem*. Let $N^O_{f,h,k}$ be the number of cluster heads that forward their data through CH $i \in g^O_{h,k}$ on their route to sink k, as illustrated in Figure 3.4. It follows that the forwarded traffic to CH $i \in g^O_{h,k}$ is:

$$\tilde{\lambda}^O_{f,i,k} = N^O_{f,h,k}\lambda, \tag{3.18}$$

Hence, the total arrival rate to CH i is:

$$\begin{aligned} \lambda^O_{i,h,k} &= \lambda_{g,i} + \tilde{\lambda}^O_{f,i,k} \\ &= \left(1 + N^O_{f,h,k}\right)\lambda, \quad i \in g^O_{h,k},\ h \in \{1, ..., N\}, \end{aligned} \tag{3.19}$$

where N is the largest number of hops from a CH to its sink along any routing path.

In the following, we model the service rate of the CHs. Let T_{slot} be the slot duration in seconds, then the average service rate in packets per second from node i to j is $\frac{T_{i,j}}{T_{slot}}$. It follows from *Theorem 3.1* that the throughput of a direct one-hop transmission from

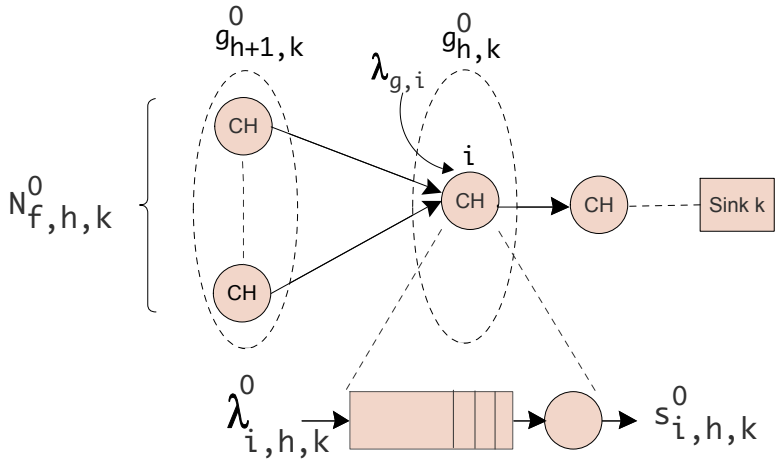


Figure 3.4: Model of CH $i \in g^O_{h,k}$.

a CH at the hth hop to a CH at the $(h-1)$th hop is:

$$s^O_{i,h,k} = \frac{1}{T_{slot}} Pr\{t^{k,O}_{i_h,j_{h-1}} = 1\} Pr\{SINR^{k,O}_{i_h,i_{h-1}} > \gamma\}, \tag{3.20}$$

where $Pr\{t^{k,O}_{i_h,j_{h-1}} = 1\}$ is the probability that CH $i \in g^O_{h,k}$ is scheduled to transmit a packet to CH $j \in g^O_{h-1,k}$, and $Pr\{SINR^{k,O}_{i_h,j_{h-1}} > \gamma\}$ is the probability of successful reception of a packet at hop level $h-1$. Following *Theorem 3.1*, under exponentially distributed SNR, we have:

$$s^O_{i,h,k} = \frac{1}{T_{slot}} Pr\{t^{k,O}_{i_h,j_{h-1}} = 1\} \exp\left\{-\frac{\gamma}{\overline{SNR}_h}\right\}, \tag{3.21}$$

where $\overline{SNR}_h = \dfrac{\bar{P}L^{-\beta}_{i_{h-1},j_h}}{N_o}$. For equidistant hops, we have $\overline{SNR}_h = SNR,\ \forall h$.

3.5.3 Stability Analysis

Assuming that the arrival and departure processes are stationary, then for a system to be stable, the service rate must be larger than the arrival rate at each queue [83, 89, 90, 93, 111]. That is, for $i \in g^O_{h,k}$ and $\forall h, k$ we must have:

$$s^O_{i,h,k} > \lambda^O_{i,h,k} \quad \Rightarrow \quad s^O_{i,h,k} > \left(1 + N^O_{f,h,k}\right) \lambda. \tag{3.22}$$

The stability condition would impose a requirement on how often a node is scheduled to transmit. Intuitively, nodes closer to the sink should be scheduled more often than other nodes, due to the larger amount of traffic they relay to the sink. Alternatively, for a particular scheduling, the stability will impose an upper bound on the rate at which traffic is generated λ. Following (3.21) - (3.22), we have the following result for any CH $i \in g^O_{h,k}$. Similar results can be obtained for nodes in other regions.

Proposition 3.2 *(Node stability analysis) For the node buffer to be stable, a CH $i \in g^O_{h,k}$ should be scheduled to transmit to the nearest CH $j \in g^O_{h-1,k}$ with a probability*

$$Pr\{t^{k,O}_{i_h,j_{h-1}} = 1\} > \frac{T_{slot}\left(1 + N^O_{f,h,k}\right)\lambda}{\exp\left\{-\frac{\gamma}{SNR}\right\}}, \quad \forall i, k. \tag{3.23}$$

Corollary 3.1 *For a particular scheduling protocol, to ensure node stability for any CH $i \in g^O_{h,k}$, the arrival rate of the self-generating traffic of each cluster must satisfy:*

$$\lambda < \arg_{k,h} min \; Pr\{t^{k,O}_{i_h,j_{h-1}} = 1\} \frac{\exp\left\{-\frac{\gamma}{SNR}\right\}}{T_{slot}\left(1 + N^O_{f,h,k}\right)}, \quad \forall i. \tag{3.24}$$

Remark 3.3 *As can be seen, the system stability is guaranteed as long as the transmission probability is above a certain threshold. This condition, in turn, can be fulfilled by providing sufficient channels and/or utilizing signal processing techniques for simul-*

taneous transmissions. Note that the stability condition can also be mapped to a lower bound on the transmission power at each CH. However, this is not recommended due to two reasons: (i) The transmission power is generally limited. (ii) Increasing the power would result in increased interference, which could reduce the frequency reuse efficiency.

Recall that the probability at which a CH is scheduled to transmits its *own* traffic to sink k is lower bounded by $P\{t_i^k = 1\} \geq \frac{N_{Freq}}{N_{intf}\ n_k}\ \forall k, i$, where n_k is the total number of clusters served by sink k. In other words, there is a scheduled time of length equal to or less than $\frac{N_{intf}\ n_k}{N_{Freq}}$ slots, where each CH can send *one* of its own generated packets to sink k. At each hop, a single CH transmits its own traffic as well as the relayed traffic from other CHs. That is, it transmits in a total of $\left(1 + N_{f,h,k}^O\right)$ slots in a single scheduling period. Thus, we have:

$$Pr\{t_{i_h,j_{h-1}}^{k,O} = 1\} \geq \frac{\left(1 + N_{f,h,k}^O\right) N_{Freq}}{N_{intf}\ n_k}, \quad i \in g_{h,k}^O. \tag{3.25}$$

When the lower bound on the transmission probability in (3.25) is higher than that in (3.23), then the stability is guaranteed through proper scheduling. We have the following result:

Corollary 3.2 *In the worst case when the scheduling protocol satisfies (3.25) with equality, then a necessary condition to ensure stability is:*

$$\frac{\left(1 + N_{f,h,k}^O\right) N_{Freq}}{N_{intf}\ n_k} > \frac{T_{slot}\left(1 + N_{f,h,k}^O\right)\lambda}{\exp\left\{-\frac{\gamma}{SNR}\right\}}. \tag{3.26}$$

It follows that for system stability, the number of clusters within the service area of

sink k must be bounded as follows:

$$n_k < \frac{N_{Freq} \exp\left\{-\frac{\gamma}{SNR}\right\}}{N_{intf} \lambda T_{slot}}, \qquad \forall k. \tag{3.27}$$

Remark 3.4 *Based on N_{Freq}, the arrival rate λ, and the average link throughput, the number of RCHs K can be chosen such that (3.27) is satisfied.*

3.5.4 Delay Analysis

Based on the *Klienrock independence assumption*, the traffic at each CH can be modeled as an M/M/1 queue whose rates are obtained as illustrated in the previous subsections. We define the utilization factor of CH $i \in g^O_{h,k}$ as:

$$\rho^O_{i,h,k} = \frac{\lambda^O_{i,h,k}}{s^O_{i,h,k}}, \tag{3.28}$$

where $\lambda^O_{i,h,k}$ and $s^O_{i,h,k}$ are the arrival rate and the service rate of CH $i \in g^O_{h,k}$. Hence, the expected number of packets in the queue at CH i is [101]:

$$\mathcal{N}^O_{i,h,k} = \frac{\rho^O_{i,h,k}}{1 - \rho^O_{i,h,k}} = \frac{\lambda^O_{i,h,k}}{s^O_{i,h,k} - \lambda^O_{i,h,k}}. \tag{3.29}$$

The average delay per packet (in seconds) along the queue at CH $i \in g^O_{h,k}$ is obtained using *Little's theorem* [101] as:

$$D^O_{i,h,k} = \frac{\mathcal{N}^O_{i,h,k}}{\lambda^O_{i,h,k}} = \frac{1}{s^O_{i,h,k} - \lambda^O_{i,h,k}}. \tag{3.30}$$

The delay in a transmission from a CH to a sink is the sum of the delays encountered at all intermediate hops along the route to the sink. Let $\bar{D}(i \in g^O_{h,k})$ be the average

delay per packet of node $i \in g^O_{h,k}$, thus we have

$$\bar{D}(i \in g^O_{h,k}) = \sum_{\substack{j=1 \\ x \in g^O_{j,k}}}^{h} D^O_{x,j,k}. \tag{3.31}$$

Delay analysis for CHs in other regions can be performed similarly.

Let $\mathbb{N}^O_{h,k}$ be the number of nodes at the hth hop from RCH k in the outer region, and $\mathbb{N}^I_{h,k}$ are those in the inner region. Also, let $\mathbb{N}_{h,0}$ be the number of nodes at the hth hop level from the CCH ($k = 0$). Define, $\mathbb{N}^O_k$ and $\mathbb{N}^I_k$ as the maximum number of hops to RCH k from the outer and inner regions, respectively, while $\mathbb{N}_0$ is the maximum number of hops to the CCH from a CH in the region $x < R_o$. Assuming that all CHs have data to transmit, we get the overall average delay in the cell by summing the delay encountered by a transmission from each CH to the nearest sink, then dividing by the number of CHs in the cell. In summary, we have the following proposition.

Proposition 3.3 *(Single path case) The average delay of a packet in the network to reach its corresponding stationary sink (CCH/RCH) along a predefined single routing path can be expressed as:*

$$\begin{aligned} \bar{\mathbb{D}} \;=\; \frac{1}{N_{CH}} \Bigg[K \Bigg(\sum_{h=1}^{\mathbb{N}^O_k} \mathbb{N}^O_{h,k} \bar{D}(i \in g^O_{h,k}) + \sum_{h=1}^{\mathbb{N}^I_k} \mathbb{N}^I_{h,k} \bar{D}(i \in g^I_{h,k}) \Bigg) \\ + \sum_{h=1}^{\mathbb{N}_0} \mathbb{N}_{h,0} \bar{D}(i \in g_{h,k}) \Bigg]. \end{aligned} \tag{3.32}$$

Note that due to the symmetry of the architecture, we get the delay of traffic around a single RCH, multiply by K, then add it to the delay of packets in the CCH region; the result is then divided by the number of CHs in the network to obtain the overall average

delay per packet. The calculations of $N^O_{f,h,k}$, $N^I_{f,h,k}$, $\mathbb{N}^O_{h,k}$, and $\mathbb{N}^I_{h,k}$ are provided in Appendix B.

Under routing diversity, the result for the single path case can be extended to the multipath case as follows:

Proposition 3.4 (Multipath case) Let N be the maximum number of hops from a CH to its sink along any routing path. Consider that for each hop number $l \in \{1, 2, ..., N\}$, there are $P_{i,l}$ possible l-hop paths from CH i to sink k. Let $\bar{D}_{i,k}(N^k_i = l, \mathcal{P}^k_i = p)$ be the average delay along one of the l-hop paths from source i to sink k assuming the path $\mathcal{P}^k_i = p$, then the overall average delay of node i's packet can be calculated as:

$$\bar{D}_{i,k} = \sum_{l=1}^{N} \sum_{p=1}^{P_{i,l}} \bar{D}_{i,k}(N^k_i = l, \mathcal{P}^k_i = p) \Pr\{\mathcal{P}^k_i = p | N^k_i = l\} \Pr\{N^k_i = l\}. \quad (3.33)$$

Let $\mathcal{N}^k$ be the set of CHs that transmit to sink k, then the overall average per packet delay in the network can be expressed as:

$$\bar{\mathbb{D}} = \frac{1}{N_{CH}} \sum_{k=0}^{K} \sum_{i \in \mathcal{N}^k} \bar{D}_{i,k}. \quad (3.34)$$

3.6 Simulation Results

In this section, we demonstrate the performance of MC-WSN through simulation examples. First, we show the effect of the number of RCHs on the average number of hops in data transmission. Then, we illustrate the per node throughput and the delay performance of the MC-WSN, and compare them to that of SENMA.

In the simulations, we use the following parameters: the communication range of the cluster heads is $R_c = 30$m and that of sensors is $r_c = 15$m, the optimal values for R_o and R_t are set according to *Proposition 3.1*, the path loss exponent is $\beta = 2$, the

SINR threshold $\gamma = 5$dB, and the bandwidth reuse measure $N_{intf} = 2$. Assuming the packet size is 16 bytes and the data rate is 5kbps, then the packet duration will be 25.6ms. The slot duration equals to the packet duration, i.e., we set $T_{slot} = 25.6$ms. Note that the same slot duration will be needed if the packet size is 128 bytes, and the data rate 40kbps.

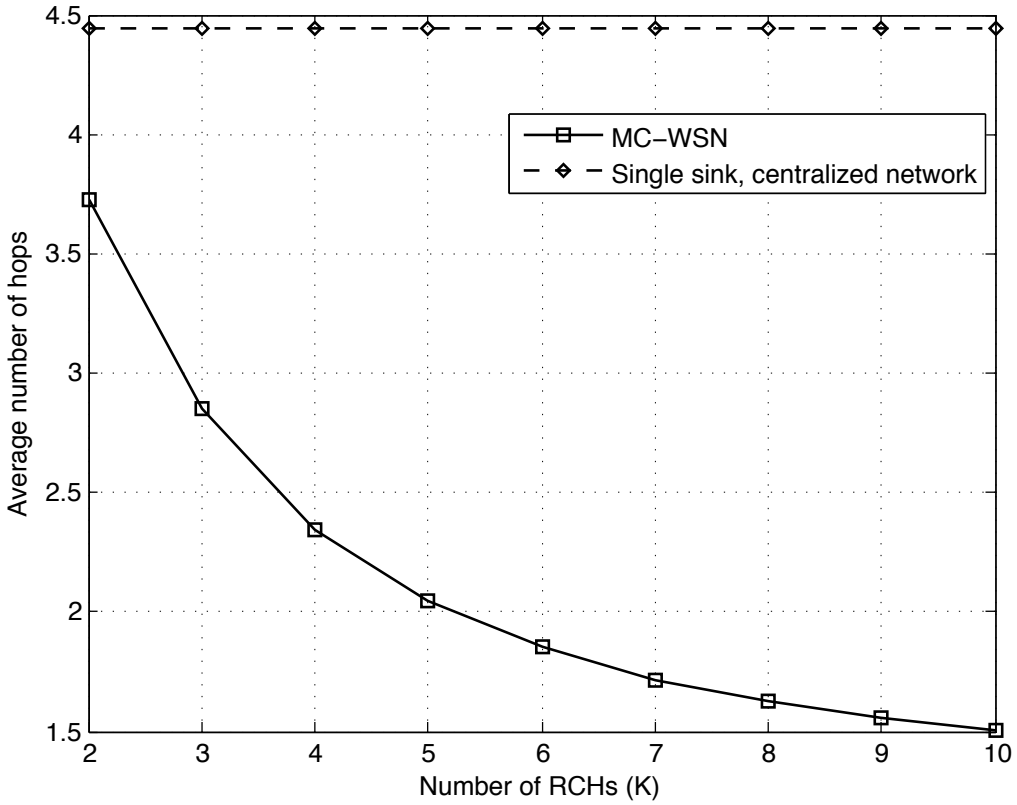


Figure 3.5: Average number of hops from a CH to its nearest sink versus the number of RCHs (K), when $d_c = 200$m and $R_c = 30$m.

Example 3.1 - *Hop number control* Figure 3.5 shows the average number of hops versus the number of RCHs (K) in a stable system. As expected, when K increases, the average number of hops decreases. It is noted that in the case when only the CCH is employed, which corresponds to the traditional centralized networks,

the average number of hops is $\frac{2d}{3R_c}$. Under the same settings used in Figure 3.5, it is clear that data transmission in MC-WSN can be performed effectively through smaller number of hops as compared to the traditional centralized network model with a single sink.

Example 3.2 - *Throughput comparison* In this example, we evaluate the overall average per node throughput of MC-WSN and compare it to that of SENMA for different network cell sizes, d_c. Define the density of the sensor nodes (SNs) and the cluster heads (CHs) as $\rho_{SN} = \frac{n}{\pi d_c^2}$ and $\rho_{CH} = \frac{N_{CH}}{\pi d_c^2}$, respectively. Here, we set $\rho_{SN} = 0.0283$, $\rho_{CH} = 0.0014$, and assume $SNR = 8$dB. In SENMA, the transmission probability of any sensor can be evaluated as: $P(t_{SENMA} = 1) = \frac{T_{slot}}{\frac{L_{MA}}{V_{MA}} + nT_{slot}}$, where V_{MA} is the speed of the MA, L_{MA} is the length of the MA trajectory, and T_{slot} is the slot duration assigned to each node for transmission. We set $V_{MA} = 30$m/s, which is relatively high. The length of the MA trajectory in SENMA can be expressed as: $L_{MA} = 2\pi \sum_{l=0}^{\lceil \frac{d_c}{2r_c} \rceil - 1} (d_c - (2l+1)r_c) + 2r_c(\lceil \frac{d_c}{2r_c} \rceil - 1)$ [112]. The MA flies at an altitude H_S. Therefore, the per node throughput in SENMA is $P(t_{SENMA} = 1) \exp\left\{-\gamma H_S^{\beta} \frac{N_o}{P}\right\}$.

In Figure 3.6, the overall average per node throughput of MC-WSN with $K = 6$ and SENMA architecture are plotted versus the network cell radius. For MC-WSN, we consider the cases when $N_{Freq} = 1$ and 4. It is shown that the throughput of MC-WSN is superior to that of SENMA. This is because the transmission of the nodes in the SENMA architecture depends on the speed of the MA and its trajectory length. It can be seen from Figure 3.6 that as the number of orthogonal frequencies increases, the throughput of MC-WSN can be further improved.

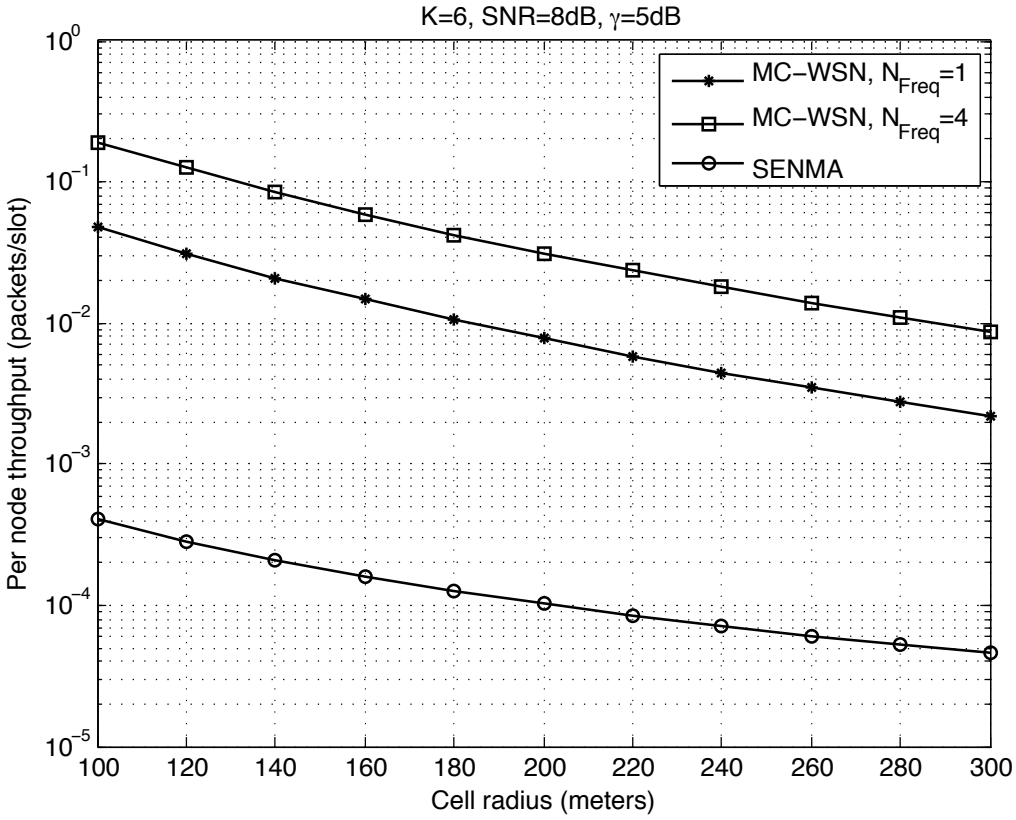


Figure 3.6: Average per node throughput in packets per slot vs. the cell radius for MC-WSN and SENMA. Here, $K = 6$, $V_{MA} = 30$m/s, $\rho_{SN} = 0.0283$, $\rho_{CH} = 0.0014$, $SNR = \frac{\bar{P}}{N_o} R_c{}^{-\beta} = 8$dB, $N_{intf} = 2$, $R_c = 30$m, $r_c = 15$m, and $T_{slot} = 25.6$ms.

Example 3.3 - *Delay comparison* In this example, we compare the average delay per packet of MC-WSN and SENMA. We set the cell radius to $d_c = 200$m, the number of cluster heads $N_{CH} = 200$, the number of RCHs $K = 6$, and the number of frequencies available for simultaneous transmissions $N_{Freq} = 1, 4$.

First, the upper bound on the rate λ the guarantees stability is shown in Figure 3.7. Here, we assume (3.25) holds with equality. As can be seen, higher data generation rates can be supported with higher SNR values. Also, as N_{Freq} increases, even higher rates can be tolerated at the same SNR level.

Next, we obtain the average delay per packet. Denote the upper bound on λ as

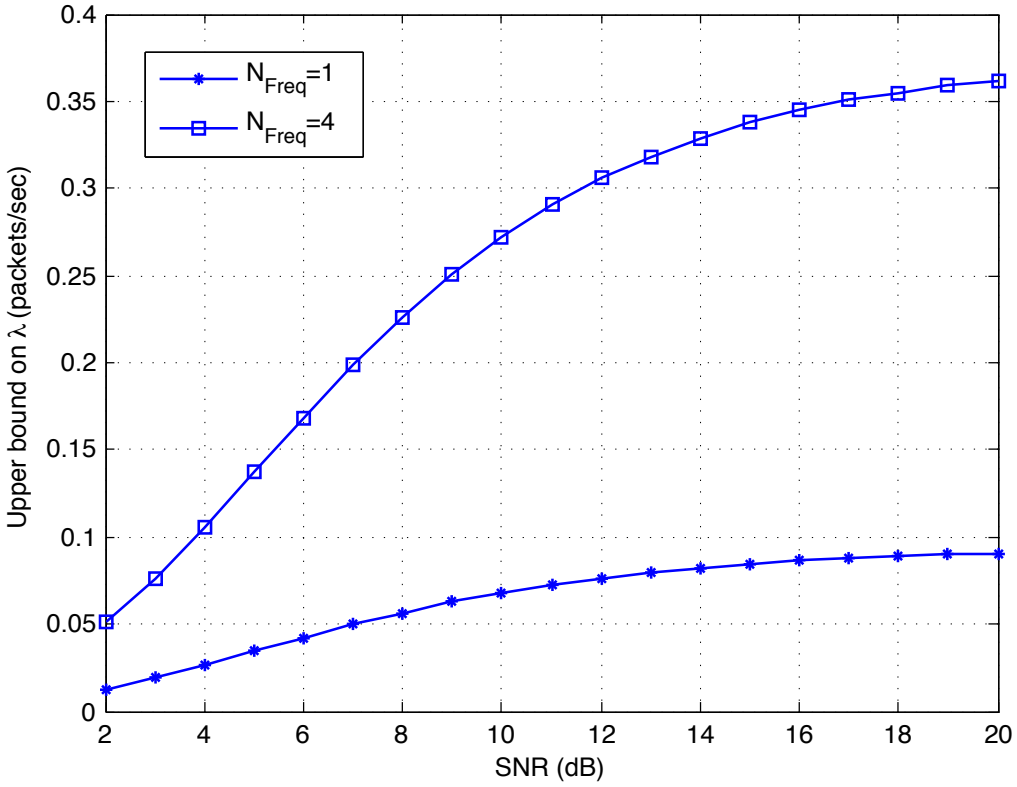


Figure 3.7: Upper bound on packet generation rate in each cluster (λ) for MC-WSN when $d_c = 200$m, $N_{CH} = 200$, and $K = 6$.

λ_{UB}, and set $\lambda = 0.9\lambda_{UB}$. The transmission probability is obtained from (3.25). For MC-WSN, we mainly consider the delay in the transmissions from the source cluster head until its corresponding sink (CCH/RCH). The delay from a sensor to its CH and from the CCH/RCH to the MA are negligible when compared to the queuing and transmission delays of the intermediate multihop transmissions. For SENMA, the delay in packet transmission is mainly dominated by the waiting time until the MA visits the source sensor; a node can be anywhere along the trajectory, hence the average delay for a node to transmit to the MA is $D_S = \frac{L_{MA}}{2V_{MA}}$. In the delay calculations for SENMA, we ignore the transmission time of signals from the sensors to the MA, and the transmission time of the MA beacon signal that notifies the sensors to transmit,

as well as the waiting time of the MA at each location for data collection.

The delay versus the SNR is shown in Figure 3.8. It is clear that MC-WSN provides orders of magnitude lower delay than that of SENMA, and even lower delays are possible when larger number of orthogonal frequencies, N_{Freq}, is available.

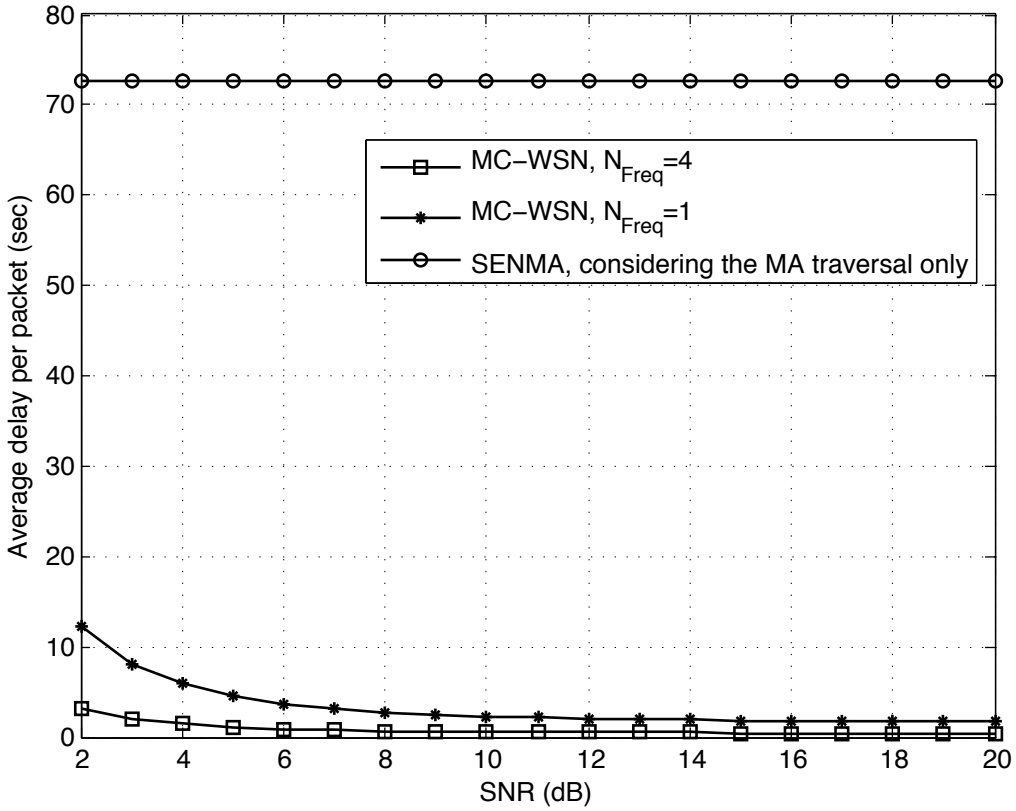


Figure 3.8: Average delay of MC-WSN and SENMA vs. received SNR. Here, d_c = 200m, $N_{CH} = 200$, $K = 6$, and $V_{MA} = 30$m/s.

Example 3.4 - *Energy efficiency* We focus on the energy dissipated in the individual sensor nodes (SNs), since they have the most limited resources. We use the circuitry radio energy dissipation model to evaluate the energy efficiency [113]. Assume that the radius of the cluster be r_c, and let E_{tx} and E_{rx} be the energy dissipated in the transmitter and receiver electronics of the SNs, respectively. Then, in MC-WSN,

the maximum energy dissipated in a sensor to transmit a bit to its corresponding CH is $E_{SN,M} = E_{tx} + \epsilon_{pa} r_c^{\beta}$ (J/bit), where ϵ_{pa} is the energy consumed by the power amplifier, β is the path loss exponent.

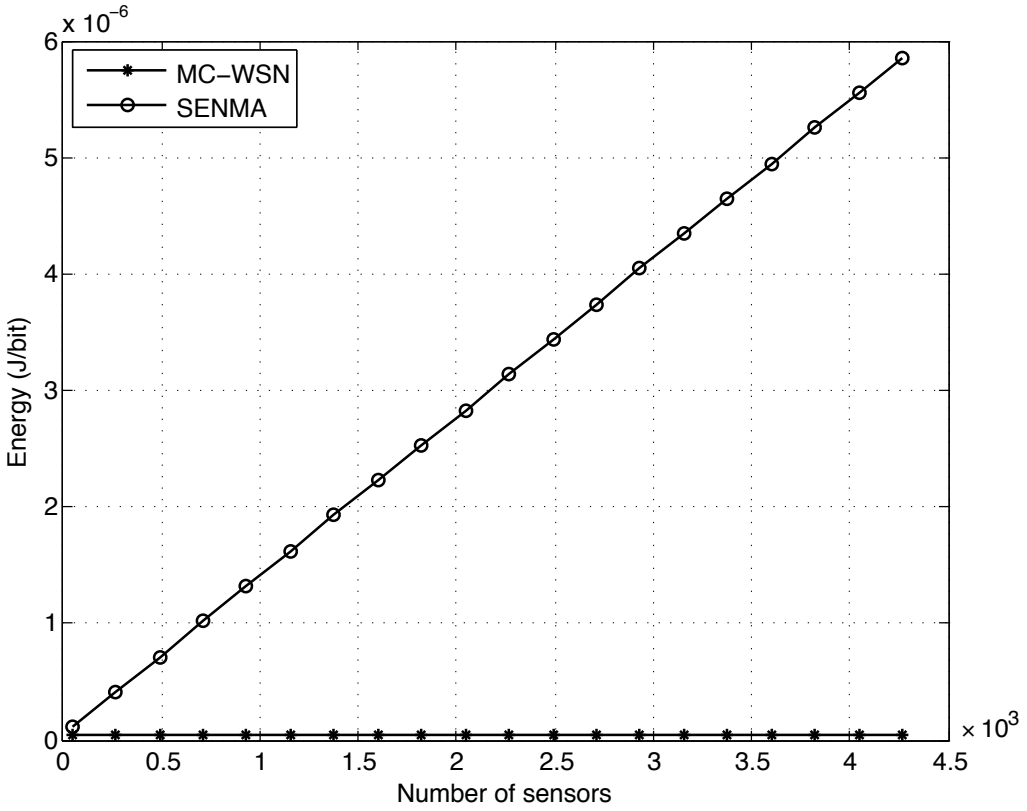


Figure 3.9: The energy dissipation (J/bit) vs. the number of SNs in the MC-WSN and SENMA networks, when $d_c = 100$m, $r_c = r = 15$m, $H_S = 10$m, $\beta = 2$, $E_{tx} = E_{rx} = 50$ nJ/bit, and $\epsilon = 10$ pJ/bit/m^2.

In SENMA, each SN must first receive a beacon signal from the MA in order to report its data. Assume the access point traverses the network at a height H_S broadcasting beacon signals at random locations. The coverage area of the access point is modeled as a circle of radius r. Therefore, the energy dissipated by a sensor to report a single bit to the MA is $E_{SN,S} = E_{tx} + \epsilon_{pa} H_S^{\beta} + E_{rx} \pi r^2 \frac{n}{A_T}$ [17], where A_T is the area of the cell. Figure 3.9 shows $E_{SN,M}$ and $E_{SN,S}$ as the number of sensor

nodes in the network increases. In this example, we set $d_c = 100$m, $r_c = r = 15$m, $H_S = 10$m, $\beta = 2$, $E_{tx} = E_{rx} = 50$ nJ/bit, and $\epsilon = 10$ pJ/bit/m^2. It can be seen from the figure that MC-WSN is significantly more energy-efficient than SENMA, and the energy efficiency gains increases as the density of the sensors increases.

It should be noted that energy dissipation in CHs and MAs are ignored here. However, if their energy dissipation is taken into account, the MC-WSN would still be more efficient than SENMA architecture. This is because, in SENMA the MAs are assumed to traverse the network continuously leading to a very high energy consumption.

3.7 Summary

In this chapter, a mobile access coordinated wireless sensor networks (MC-WSN) architecture was proposed for reliable, efficient, and time-sensitive information exchange. MC-WSN exploits the MAs to coordinate the network through deploying, replacing, and recharging nodes, as well as detecting malicious nodes and replacing them. The hierarchical and heterogeneous structure makes the MC-WSN a highly resilient, reliable, and scalable architecture. We provided the optimal topology design for MC-WSN such that the average number of hops from any sensor to the MA is minimized. We analyzed the performance of MC-WSN in terms of throughput, stability, delay, and energy efficiency. It was shown that with active network deployment and hop number control, MC-WSN achieves much higher throughput and considerably lower delay and energy consumption over the conventional SENMA. Moreover, our analysis also indicated that with hop number control, network analysis does become more tractable.

Chapter 4

N-Hop Networks – A General Framework for Wireless Systems

This chapter aims to provide a unified framework for quantitative characterization of various wireless networks. We first revisit the evolution of centralized, ad-hoc and hybrid networks, and discuss the trade-off between structure-ensured reliability and efficiency, and ad-hoc enabled flexibility. Motivated by the observation that the number of hops for a basic node in the network to reach the base station or the sink has a direct impact on the network capacity, delay, efficiency and their evaluation techniques, we introduce the concept of the N-hop networks. It can serve as a general framework that includes most existing network models as special cases, and can also make the analytical characterization of the network performance more tractable. Moreover, for the network security, it is observed that hierarchical structure enables easier tracking of user accountability and malicious node detection; on the other hand, the multi-layer diversity increases the network reliability under unexpected network failure or malicious attacks, and at the same time provides a flexible platform for privacy protection. Finally, we discuss some possible topics for further research.

4.1 Preface

Communications rely on networks. Today's wireless networks are generally divided into two categories: centralized networks with well-defined infrastructure, and distributed or ad-hoc networks which are virtually structureless. There is also a trend to blend these two structures together, resulting in various hybrid networks [14, 15]. In this chapter, we try to summarize the general design criteria of wireless networks, and come up with a unified framework that can include most of the existing systems as special cases, and makes quantitative characterization of wireless networks more tractable. To do this, we first examine some representative systems in the literature.

4.2 The Evolution of Wireless Communication Systems

The development of mobile telephony traces back to the late 1910s, when a group of German engineers started the experiments on telephony via radio links, and tested on the military trains between Berlin-Zossen in 1918 [114]. The first handheld radio transceivers, also called walkie-talkies, were the backpacked Motorola SCR-300 [115], developed in 1940; later refined and widely used during the World War II (1939). Right after the war, engineers in Bell Labs invented a system to allow mobile users to place and receive telephone calls from automobiles, leading to the inauguration of mobile services in 1946 in St. Louis, Missouri.

After that, a wide range of incompatible mobile services, supported by analog techniques, were provided in urban areas, each offering very limited coverage through a base station that has only a few channels.

4.2.1 Cellular Systems

The concept of cellular technology, which exploited low-power transmitters and allowed wide range frequency reuse, was introduced and developed by Bell Labs engineers from the late 1940s to early 1970s [116]. While the first hexagonal cell concept was proposed in 1947 [117], the full development and implementation of the cellular technology, including both frequency reuse and call handover, took more than two decades. The cellular technology made the mobile services affordable to ordinary people, and led to the revolutional widespread of wireless communications.

The first generation (1G) cellular systems (1970s), represented by AMPS (Advanced Mobile Phone System) in the US (later on evolved to IS-41) and ETACS (European Total Access Communication System) in Europe, relied on analog technologies and mainly provided voice services. Started in late 1980s and deployed in 1990s, the second generation (2G) cellular systems (United States Digital Cellular (USDC) IS-136, CDMA IS-95, and GSM) all used digital coding and modulation techniques. The 2G systems increased the network capacity by about three times. As they were designed before the wide spread of the internet, they mainly supported voice-centric services and limited data-services, like short messages and Fax.

Began in late 1990s, the 3G systems (UMTS WCDMA, CDMA 2000, and TD-SCDMA) supported high-speed multimedia services and seamless global roaming. Wireless access became available throughout the earth, with the proud claim of "anywhere, anytime, anything". The communication quality was further enhanced by the OFDM technique, leading to the high speed, high quality 4G systems, represented by WiMAX (Worldwide Interoperability for Microwave Access) [118] and LTE (Long Term Evolution) [119]. Today, with the coexistence of 3G and 4G, we can have real-time multimedia communications through world-wide networks.

4.2.2 Ad-hoc Networks

The walkie-talkies, which are still widely used today in military, public safety, businesses, outdoor recreation and the like, actually form a complete mesh network, where any two users, within the device power range, can communicate directly in a structureless manner.

Going beyond this one-hop communication mode and allowing multihop routing process, the self-configuring infrastructureless wireless ad-hoc network has attracted lots of attention from the research community. The research on ad-hoc networks was mainly driven by the growth of laptops and 802.11/Wi-Fi wireless networking, and the advent of all kinds of wireless sensors, leading to the areas of MANET (Mobile Ad-hoc Network) [120] and WSN (Wireless Sensor Network) [121], respectively. MANET has been widely deployed as local area networks in businesses, universities, airports and places alike, for convenient wireless internet access and internet-assisted communications. At the same time, wireless sensor networks have seen wide use in both military and civilian applications, such as health monitoring [122], intelligent transportation systems [3], target detection and tracking especially in unattended and possibly hostile areas.

4.2.3 The Merging Ground for Cellular and Ad-hoc – Hybrid Networks

While the structureless ad-hoc networks provide excellent flexibility with reliable performance for small scale networks, scalability proved to be a serious challenge for large-scale ad-hoc networks due to the uncertainty, complexity, as well as the delay and energy concerns in the routing process. The problems become even worse when the devices are mobile.

This observation leads to ad-hoc networks with local structures, known as *hybrid networks*. One representative example is the clustered wireless sensor networks, where the sensors are grouped into clusters, with each cluster managed by a cluster head in a centralized manner [16]. The routing responsibility is fulfilled only by the cluster heads, and not ordinary sensor nodes. Similar ideas are developed for the mobile ad-hoc network (MANET) [120], include multi-hop cellular network (MCN), integrated cellular and ad-hoc relaying systems (iCAR) [87], self organizing product radio networks (SOPRANO) [123], etc.

At the same time, *hybrid networks* also raised from the cellular networks. This is mainly motivated by the following two observations: (i) For todays centralized network, the mobile will generally lose network connection once the BS is not functioning, since each mobile is typically connected to only one BS. (ii) Even if two mobiles are spatially close, they cannot establish direct communication, but have to communicate through the BS, leading to unnecessary resource waste. That is, traditional centralized network does not have sufficient diversity and endpoint communication flexibility. As a result, recent wireless MAN and LAN standards, such as WiMAX 802.16 [118] and WiFi 802.11s [124], have incorporated the mesh capability to the wireless network nodes, which allows each node to forward the traffic of other nodes in the network in a planned yet an ad-hoc manner.

Other representative examples include Ad-hoc GSM (A-GSM) [125], cellular networks with device-to-device (D2D) communications [126–129], and iCAR [87], where the main idea is to improve transmission flexibility, viability, capacity, and traffic balance by allowing device-to-device and/or device-to-relay station communications.

From our discussions above, it can be seen that hybrid networks serve as a merging ground for centralized and ad-hoc networks, as shown in Figure 4.1, and stimulate the research on different kinds of heterogeneous networks (Hetnets). On the mobile

side [126, 130, 131], one visible trend is the real-time video communications that drives for the trade of memory for capacity. The Femtocells in LTE-advanced [130], for example, can be used to store popular videos so that they do not need to be requested through the BS. On the sensor side, an interesting move is sensor networks with robot-like mobile access points [17, 78].

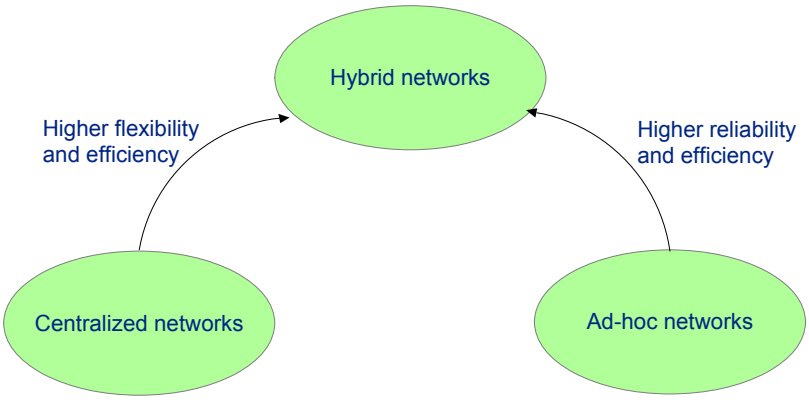


Figure 4.1: Merging of centralized and ad-hoc networks.

4.3 General Design Criteria

The evolution of the centralized and ad-hoc networks to hybrid networks reveals that: for wireless communications, we would need both network centric management as well as ad-hoc flexibility. Based on this observation, we can summarize the general network design criteria as follows:

The network needs to have a well-organized infrastructure to ensure the reliability (including both transmission accuracy and security), capacity, energy efficiency as well as time efficiency. At the same time, the network should provide sufficient flexibility by allowing authorized ad-hoc communications among the nodes or devices. More specifically,

- The infrastructure needs to be *hierarchical* for efficient management. The density of the devices gets higher as their level gets lower.

- The infrastructure needs to provide *sufficient diversity* at each level to combat hostile attacks or unexpected system failures. More specifically, devices or basic nodes (BNs) at each level can communicate with two or more upper level devices.

- The infrastructure needs to provide *sufficient flexibility*.

  - Once authenticated, neighboring devices (either relay stations or basic nodes) at the same level can communicate directly with each other within their transmission range without going through higher layer nodes.

  - When permitted, each device can communicate directly with higher layer nodes within its communication range to minimize the number of hops needed to reach a base station (BS) or sink.

  - Under special cases when a BN cannot access the network directly, as long as an agreement is reached between two BNs (both BN should be authenticated if possible), one BN can serve as the relay for another BN.

From a biomimetic perspective, these criteria can be largely verified in the design of the human body. Consider the circulatory system, in which the extracellular fluid is transported through parts of the body in two stages [132]. The first stage is the movement of the blood in the blood vessels; the second stage is the movement of fluid between the blood capillaries and the intercellular space between the tissue cells. The latter stage is also called micro-circulation, it is for the transport of nutrition to the tissues and removal of cell excreta. The first stage is centralized and well structured with good diversity. Even if some vessels are not functioning well (e.g. blocked), as long as they are within a certain threshold, the human body will continue to function.

In the second stage, the micro-circulation, the exchange of water, nutrients and other substance between the plasma in blood and interstitial fluid in the tissue is mainly done through diffusion, which results from thermal motion of the water molecule and dissolved substances in the fluid. To make it short, it is random!

As can be seen, the circulatory system in human body is an excellent combination of a well-structured "backbone" network and numerous small random networks at the endpoints. It ensures transmission efficiency, diversity and endpoint service flexibility, and thus provides a very good example for network design.

4.4 The Concept of N-hop Networks

With the general design criteria in mind, we now try to come up with a unified framework for wireless networks that could cover most of the existing systems as special cases, and makes quantitative network characterizations (such as throughput, delay, and efficiency) more tractable.

We first look at some examples. In strictly centralized networks, which is widely adopted in cellular communication systems, the mobile reaches the base station (BS) in one hop. In the one layer relay-assisted networks, the basic node either reaches the BS directly in one hop, or reaches the base station through the relay in two hops [133]. In the pure ad-hoc networks or sensor networks, there is generally no specific limit on the number of hops for a basic node to reach a sink.

For any wireless network, let the *minimum* number of hops for a basic node (i.e., the terminal, such as a mobile or a sensor) i to reach the base station (BS) or the sink be N_i. Define $N = \max\{N_1, N_2, ...N_n\}$, where n is the number of basic nodes. N is an important characterization of how closely the basic nodes are connected to the BS or the sink, and it has a direct impact on network capacity, reliability, delay,

efficiency, as well as their evaluation techniques. Here we introduce the concept of *N-hop networks*: a wireless network is said to be an N-hop network if every basic node (BN) can reach the BS or the sink within N hops under normal network conditions. By normal conditions, we mean that there are no hostile attacks, or severe, unexpected system failures.

Based on this definition, if $N = 1$, we obtain the strictly centralized network. For some sensor networks with mobile access points, we also have $N = 1$, see the SENMA in [17] for example. In SENMA, with well designed mobile access trajectory, there is no routing and all the sensors can reach the mobile access in one hop. If $N = 2$, we get the relay-assisted network; if $N = \infty$, it reduces to the pure ad-hoc network. Actually, almost all the existing systems fall into this unified framework. As will be seen later, with this framework, analytical evaluation of the throughput, delay, and efficiency becomes more tractable.

Denote the total number of hops for a node i to reach its destination, node j, as $N_{i,j}$. For an N-hop network, we have $1 \leq N_{i,j} \leq 2N+N_{i,j}^c$, where $N_{i,j}^c$ is the number of hops required for the intercell communications between the two base stations connected to nodes i and j, respectively. For the complete (local) mesh network where any two nodes can communicate directly, we always have $N_{i,j} = 1$ for any source-destination pair (i,j).

Due to possible link failure conditions and/or malicious attacks, the number of hops for a node to reach the sink could be more than N. For this reason, we extend the definition of N-hop networks to *α-level N-hop networks*, which is characterized by: $Pr\{\text{BN can reach the sink within N-hops}\} = \alpha$. The level α can be used as an indicator of how smooth the network is operating.

Next, we provide two examples to further illustrate the $N - hop$ network: one on mobile network, and one on sensor network.

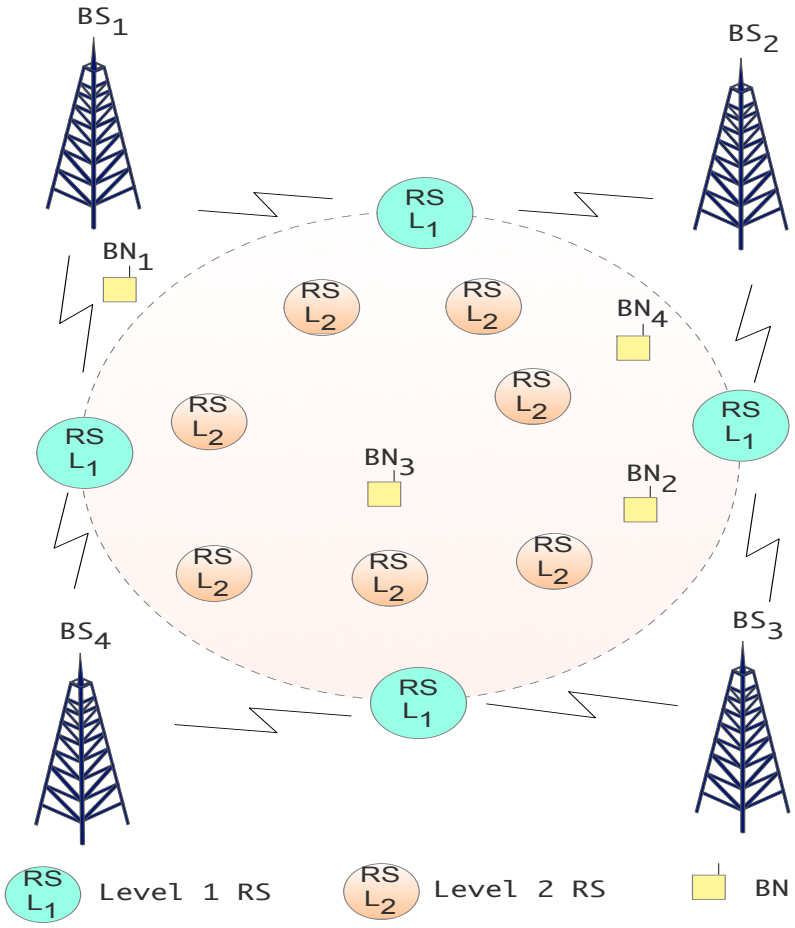


Figure 4.2: A 3-hop mobile network.

Example 4.1 - *A 3-hop mobile network* In this network, multiple base stations (BSs) and two levels of relay stations (RSs) (level 1 and level 2) are employed to serve the basic nodes (BNs) - the mobiles, as illustrated in Figure 4.2. Level 1 RSs have larger coverage area and storage capacity than level 2 RSs, but level 2 RSs have much higher distribution density in the network. Devices or nodes at each level can communicate with two or more upper level devices. Within their transmission range, neighboring devices (either RSs or BNs) at the same level can communicate directly with each other without going through higher layer nodes. At the same time, each device can communicate directly with the highest level higher layer nodes within its

communication range to minimize the number of hops needed to reach a BS. This is a 3-hop network. The tolerance of the network to system failures or hostile attacks is determined by its inherent diversity.

Example 4.2 - *Mobile Access Coordinated - Wireless Sensor Network (MC-WSN)* In the proposed MC-WSN architecture described in Chapter 3, data transmission from sensor nodes to the mobile access point (MA) goes through simple routing with the center cluster head (CCH) or the ring cluster heads (RCHs). MC-WSN is an example of hybrid network: it has a hierarchical structure supported by the CCH, RCHs, and CHs; at the same time, it also allows partially ad-hoc routing for network flexibility and diversity. Through active network deployment and topology design, the number of hops from any sensor to the MA can be limited to a *pre-specified number* N [43]. The MC-WSN architecture is illustrated in Figure 3.1.

The examples above illustrate that the N-hop network does provide a general framework for the characterization of centralized, ad-hoc, as well as hybrid networks.

4.5 Analytical Evaluation of the Network Performance

In this section, we provide a quantitative characterization of wireless networks under the N-hop framework, in terms of throughput, delay, and energy efficiency.

4.5.1 Throughput

Consider an N-hop network that contains n basic nodes. For each individual BN i, the throughput, T_i, is defined as the average number of packets per slot that are

initiated by node i and successfully delivered to the intended receiver [91]. For an N-hop network, when the receiver is the BS or the sink, the number of hops from BN i to the BS satisfies $1 \leq N_i \leq N$. Note that the transmission can always go through different paths due to the existence of network diversity. We assume that for each hop number $l \in \{1, 2, ..., N\}$, there are $P_{i,l}$ possible l-hop paths from BN i to the BS. Let $T(i|N_i = l, \mathcal{P}_i = p)$ be the throughput that can be achieved along one of the l-hop paths $\mathcal{P}_i = p$, then the throughput of node i can be calculated as:

$$\begin{aligned} T_i &= \sum_{l=1}^{N} \sum_{p=1}^{\mathcal{P}_{i,l}} T(i|N_i = l, \mathcal{P}_i = p) \Pr\{\mathcal{P}_i = p|N_i = l\} \\ &\quad \times \Pr\{N_i = l\}. \end{aligned} \tag{4.1}$$

The overall network throughput can be obtained as $\sum_{i=1}^{n} T_i$.

It should be noted that the throughput of node i along $\mathcal{P}_i = p$, $T(i|N_i = l, \mathcal{P}_i = p)$, mainly depends on the probability of successful transmission at each hop, which is generally characterized by the probability that the signal to noise and interference ratio ($SINR$) is above a certain threshold γ [134]. More specifically, referring to *Theorems 3.1 and 3.2* in Chapter 3 and setting $K = 1$, the throughput from node i to the sink given a certain routing path can be expressed as:

$$T(i|N_i = l, \mathcal{P}_i = p) = Pr\{t_i = 1\} \exp\left\{ -\gamma \frac{N_o}{\bar{P}} \sum_{h=1}^{l} \left[L_{i_h, i_{h-1}} \right]^{\beta} \right\}, \tag{4.2}$$

where $Pr\{t_i = 1\}$ is the probability that node i transmits a packet to the sink, $L_{i_h, i_{h-1}}$ is the distance between the transmitting and receiving nodes at the hth hop along the routing path from source i to the sink, $\bar{P}$ is the average transmission power, and N_o is the noise power.

Assume a relatively flat noise power along the transmission path, when the trans-

mission power of the nodes is low, the throughput improves as the number of hops increases. This is due to the reduced path loss at each hop as compared to longer distance transmission. On the other hand, when the transmission power is large, the throughput improves as the number of hops decreases, because of reduced propagation errors. This is illustrated in Figures 4.3, where the per-node throughput versus the transmit signal to noise power ratio is shown. In Figures 4.3(a) and 4.3(b), the bandwidth reuse measure along the path (N_{intf}) equals to N_i and 3, respectively, and $Pr\{t_i = 1\} = \frac{1}{\min\{N_{intf}, N_i\}}$.

From the discussions above, we can see that under a particular power constraint, there exists an optimal number of hops for throughput maximization. Consider a single source-destination pair. Let $N_{intf} = N_i$, i.e., there is no bandwidth reuse along the path from the source to the destination, then $Pr\{t_i = 1\} = \frac{1}{N_i}$. Assume equidistant hops of length $\frac{L_t}{N_i}$, where L_t is the distance between the source and the intended destination. The optimal number of hops, N_{opt}, that maximizes the throughput is obtained by setting $\frac{\partial T(i|N_i, \mathcal{P}_i)}{\partial N_i} = 0$. It follows from (4.2) that

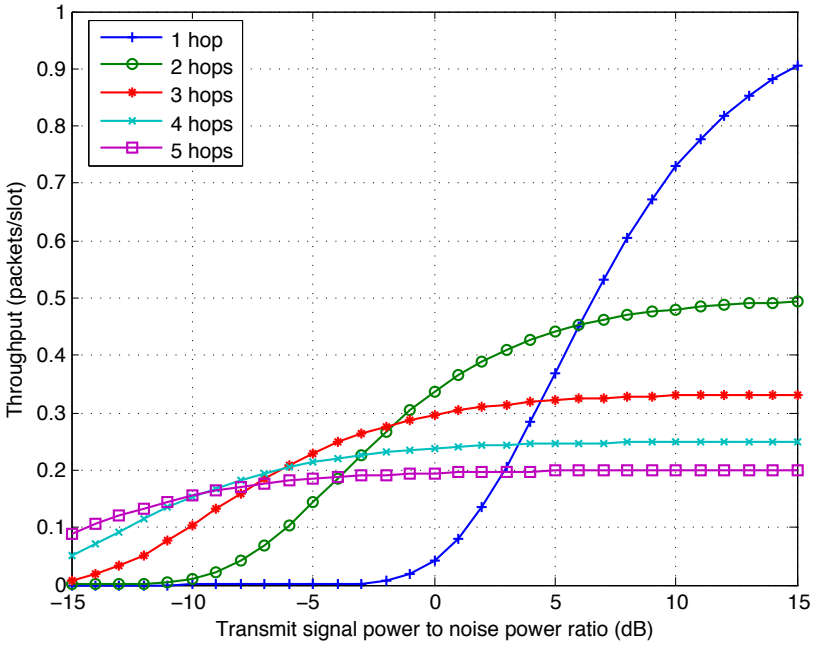
$$\left[-\frac{1}{(N_i)^2} + (N_i)^{-1-\beta}\frac{(\beta-1)\gamma}{\bar{P}/N_o}L_t^{\beta}\right]\exp\left\{-\gamma\frac{N_o}{\bar{P}}N_i\left[\frac{L_t}{N_i}\right]^{\beta}\right\} = 0. \tag{4.3}$$

Therefore, we get:

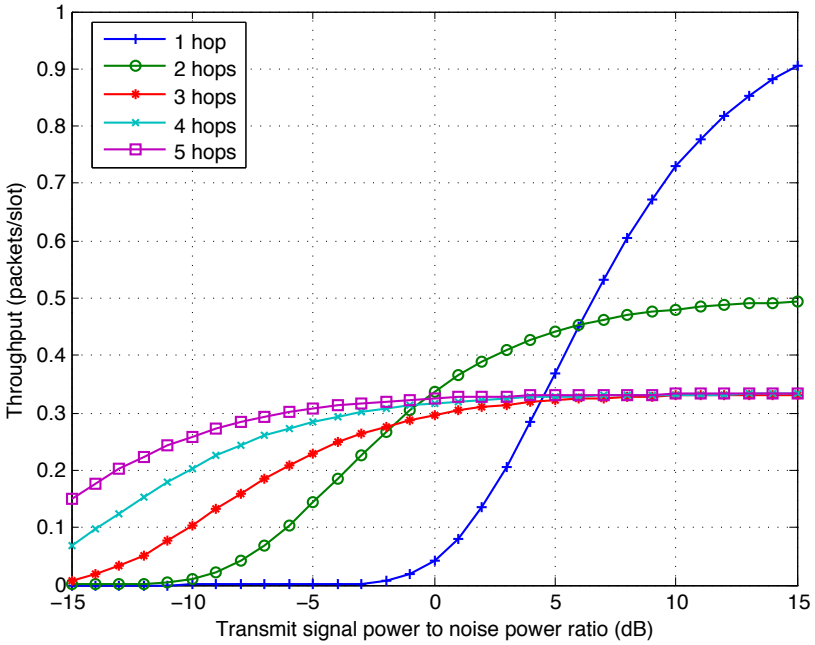
$$N_{opt} = \left[\frac{(\beta-1)\gamma}{\bar{P}/N_o}L_t^{\beta}\right]^{\frac{1}{\beta-1}}. \tag{4.4}$$

N_{opt} versus the transmit signal to noise power radio ($\bar{P}/N_o$) is shown in Figure 4.4. This result indicates that the optimal hop number versus the transmission power provides a critical reference for network structure design.

Next, we will discuss the impact of the network structure and routing flexibility on the throughput performance.



(a) Bandwidth reuse measure $N_{intf} = N_i$.



(b) Bandwidth reuse measure $N_{intf} = 3$.

Figure 4.3: Per-node throughput $T(i|N_i, \mathcal{P}_i)$ vs. the average transmit power per noise power ratio for different number of hops, assuming AWGN channel, path loss exponent is 4, SINR threshold is $\gamma = 5dB$, the hops are equidistant, distance between transmitter and receiver is normalized to 1m. The transmit power is exponentially distributed.

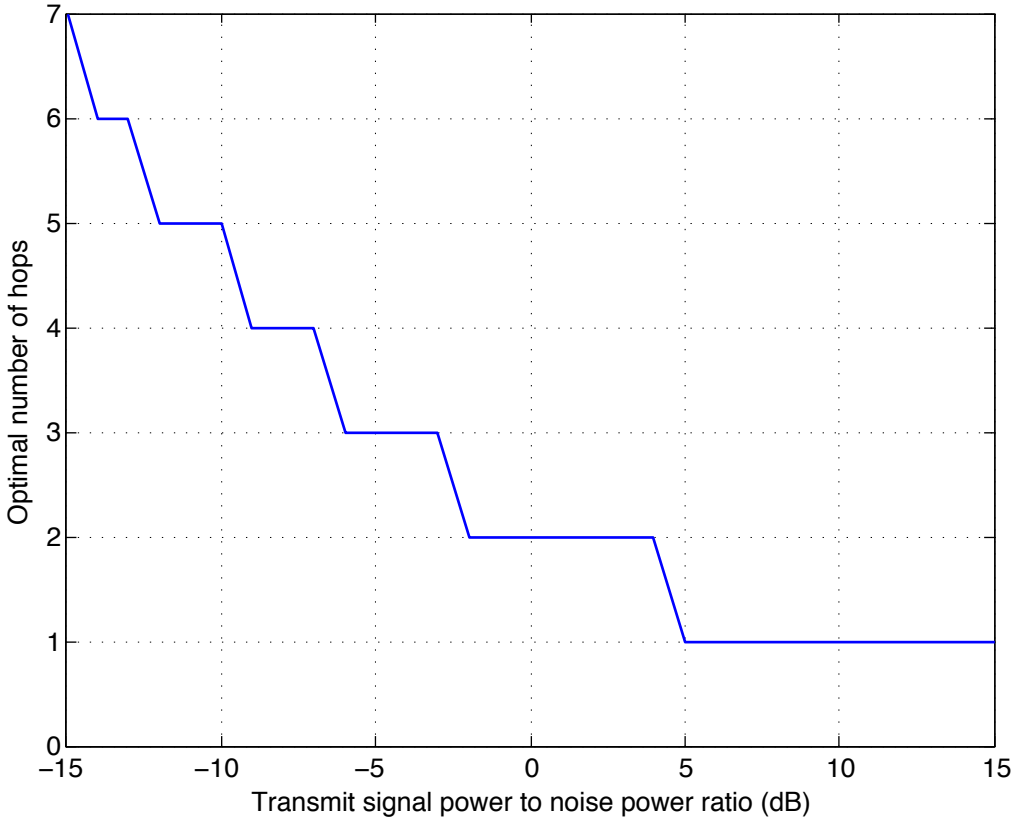


Figure 4.4: Optimal number of hops obtained by rounding (4.4) to the nearest integer. Here, $\gamma = 5dB$.

Example 4.3 ***Structured versus structureless network models*** In an N-hop structured network, under normal network conditions we always have $N_i \leq N$. On the other hand, for structureless networks, due to the absence of the infrastructure support, generally it is hard to put a limit on the maximum number of hops needed in the data delivery process. Assume that the per-hop distance is fixed, it follows from (4.2) and also from *Remark 3.1* in Chapter 3, that $\lim_{N_i \to \infty} T_i = 0$. It can hence be seen that comparing with structureless network, the N-hop network secures the throughput for each node by limiting the number of hops in the data delivery process.

Example 4.4 ***Routing flexibility*** Now, let us look at the impact of the routing flexibility on the throughput. Although structured network is highly desired as mentioned earlier, it is important to have routing flexibility around the endpoints (BNs) to enhance the overall network efficiency. That is, neighboring BNs can communicate directly or use simple routing through the lowest level relay stations without going to the higher layer stations. This is possible with the advances in radio technology that enable today's wireless devices to have cognitive abilities, which help them learn about the environment, detect their neighbors, and determine the available spectrum bands to utilize [46]. Thus, BS intervention in coordinating the communications between endpoints could be reduced.

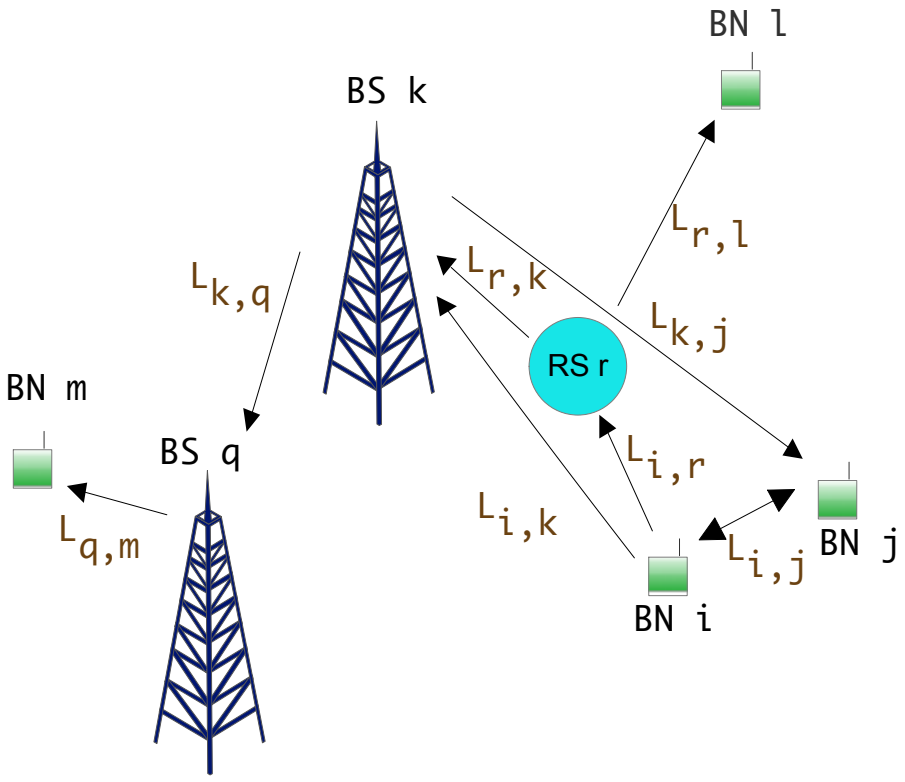


Figure 4.5: Routing flexibility: Scenario 1: BN i and BN j communicate directly. Scenario 2: BN i and BN l communicate through RS r. Scenario 3: BN i and BN m communicate through RS r and the BSs k and q.

Consider an example when BN i wants to communicate with BNs j, l and m. BN m is out of the footprint (or the cell) covered by BS k that serves BNs i, j, and l, as illustrated in Figure 4.5. We consider the following three scenarios:

(i) Communication between BN i and BN j. With the proposed network model, direct communication can be established between BN i and BN j. In this case, the throughput is $T_{i_1} \propto \exp\{-L_{i,j}^{\beta}\}$. However, in strictly centralized networks including the current cellular systems, where routing flexibility is not employed, BN i has to reach the BS first to communicate with BN j. In this case, the throughput will become $T_{i_2} \propto \exp\{-(L_{i,k}^{\beta} + L_{k,j}^{\beta})\}$. Clearly, since $L_{i,j} < L_{i,k} + L_{k,j}$, then $T_{i_1} > T_{i_2}$.

(ii) Communication between BN i and BN l. In our proposed model, the communication can be made through RS r (BN $i \Leftrightarrow$ RS $r \Leftrightarrow$ BN l). In this case, the throughput will be $T_{i_1} \propto \exp\{-(L_{i,r}^{\beta} + L_{r,l}^{\beta})\}$. If the transmission is made through the BS, then we need BN $i \Leftrightarrow$ BS $k \Leftrightarrow$ BN l or BN $i \Leftrightarrow$ RS $r \Leftrightarrow$ BS $k \Leftrightarrow$ BN l. Considering the case when the transmission is made through BN $i \Leftrightarrow$ BS $k \Leftrightarrow$ BN l, then the throughput will be $T_{i_2} \propto \exp\{-(L_{i,k}^{\beta} + L_{k,l}^{\beta})\}$. When the same transmit power is used in both cases, we have $T_{i_1} > T_{i_2}$.

(iii) Communication between BN i and BN m. Since BN m is not in the same cell as BN i, the communication is made through the BSs, i.e., BN $i \Leftrightarrow$ BS $k \Leftrightarrow$ BS $q \Leftrightarrow$ BN l. In this case, $T_i \propto \exp\{-(L_{i,k}^{\beta} + L_{k,q}^{\beta} + L_{q,m}^{\beta})\}$.

In this example, BSs are only involved in communications to nodes out of its footprint. Considering the possibility of using low-power transmissions over unlicensed band in scenarios (i) and (ii), the overall network capacity can potentially be increased by allowing endpoint routing flexibility.

4.5.2 Delay

The quantification of the delay in N-hop networks involves both information theory and queuing theory. The former studies the maximum rate at which each node can transmit over the channel, by considering noise and interference effects, but ignoring the queuing delay that could be experienced at intermediate relays/queues. The latter considers the queuing delay with possible random arrival and departure times at the intermediate relays.

The delay in one-hop communication is mainly composed of three parts:

(i) The *queuing delay* is the time between the packet arrives at a node, until it reaches the head of the queue where it can be transmitted. Little's theorem [101] formulates the average queuing delay as $D_q = \frac{Q_L}{\lambda_q}$, where Q_L is the average number of packets in the queue, and λ_q is the rate at which the packets arrive to the queue.

Nodes with finite storage can be modeled as M/M/1/B queues, where the arrivals are memoryless Poisson process with rate λ_q, the service times are exponentially distributed with rate ϑ, and the storage of each queue is B packets. Let $\mathbb{P}_B$ be the probability that the queue is full. Note that each node receives a packet only when it is not full, hence the effective arrival rate is $\lambda_e = \lambda_q(1 - \mathbb{P}_B)$. With the effective arrival rate and the mean queue length, the queuing delay can then be obtained using the *Little's theorem* [101].

(ii) The *back-off delay* occurs when a packet is not successfully received due to either full receiver buffer or collisions in the transmission; in this case, the transmitter will retransmit the packet after a back-off time. Collision happens when two or more interfering nodes access the channel at the same time. In structured networks, back-off time can be minimized due to the presence of a centralized control on data transmission, which allows interfering nodes to transmit on a different time slots or different frequency bands. The back-off time can be assumed to be exponentially distributed [135]. The

node stays in the back-off state until the channel is idle and the receiver buffer is no longer full.

(iii) The transmission delay is the difference between the time data is encoded and transmitted until it is successfully recovered/decoded at the receiver [136]. It depends on the size of the packet and the transmission rate, which is bounded by the information-theoretic capacity.

In an N-hop wireless network, the average delay of a transmission, $\bar{D}$, can be calculated as:

$$\bar{D} = \sum_{l=1}^{N} \sum_{p=1}^{\mathcal{P}_{i,l}} \sum_{m=1}^{l} \bar{D}_{p,m} \Pr\{\mathcal{P}_i = p | N_i = l\} \Pr\{N_i = l\}, \tag{4.5}$$

where N_i denotes the actual number of hops for a transmission, and $\bar{D}_{p,m}$ is the average delay in the mth hop along the path p.

4.5.3 Energy Efficiency

As in Chapter 3, we use the circuitry radio energy dissipation model [113] to evaluate the energy efficiency. In this model, each receiving node consumes E_{rx} (J/bit), and each transmitting node consumes $E_{tx}+\epsilon_{pa}L^{\beta}$ (J/bit), where ϵ_{pa} is the energy consumed by the power amplifier, β is the path loss exponent, L is the per-hop distance, and E_{tx} is the energy dissipated in the transmitter electronics. Then the total energy dissipated at a one-hop communication is $E_{tx} + \epsilon_{pa}L^{\beta} + E_{rx}$ (J/bit).

In an N-hop network, the average energy consumption for a packet transmission, $\bar{E}$, can be calculated as:

$$\bar{E} = \sum_{l=1}^{N} \sum_{p=1}^{\mathcal{P}_{i,l}} \sum_{m=1}^{l} \bar{E}_{p,m} \Pr\{\mathcal{P}_i = p | N_i = l\} \Pr\{N_i = l\}, \tag{4.6}$$

where $\bar{E}_{p,m}$ is the average energy consumed at the mth hop of path p.

Example 4.5 - *Energy efficiency versus the number of hops* In this example, we compare the average energy dissipation for two network models: (i) SENMA architecture, which is a one-hop centralized network with mobile access, (ii) MC-WSN architecture, which is described in Chapter 3, with $K = 3$ RCHs and shortest path routing. For energy comparison, in Chapter 3 we mainly focused on the energy dissipated at a sensor node to transmit to its cluster head. Here, we focus on the energy dissipated in the multihop transmissions from a cluster to a sink in MC-WSN. The result is shown in Figure 4.6. It is clear that the N-hop MC-WSN is much more efficient than SENMA networks. The reason is that in SENMA, each SN must first receive a beacon signal from the MA in order to report its data. All sensors within the coverage area of the MA receive the beacon signal, and only one sensor responds each time [17]. The energy dissipation during the beacon reception process contributes significantly to the overall energy consumption for each transmission in SENMA.

Remark 4.1 *For the α-level N-hop network, the number of hops can be greater than N with probability $(1-\alpha)$. In this case, we can extend (4.1), (4.5), (4.6) accordingly, by changing N to N_{max}, which is the maximum number of hops in a cell. Equations (4.1), (4.5), (4.6) can also be extended directly to the node-to-node communication case.*

4.6 Security Perspectives

4.6.1 Delay-assisted Network Failure/Attack Detection

Consider a particular node i, under normal network conditions, N_i is the number of hops from BN i to the sink, then the delay of node i's transmission to the sink is $D_i = \sum_{k=1}^{N_i} d_k$, where d_k is the delay in hop k. Note that the average delay is given in (4.5). If the actual delay D_i is much larger than the average delay $\bar{D}_i$, then this

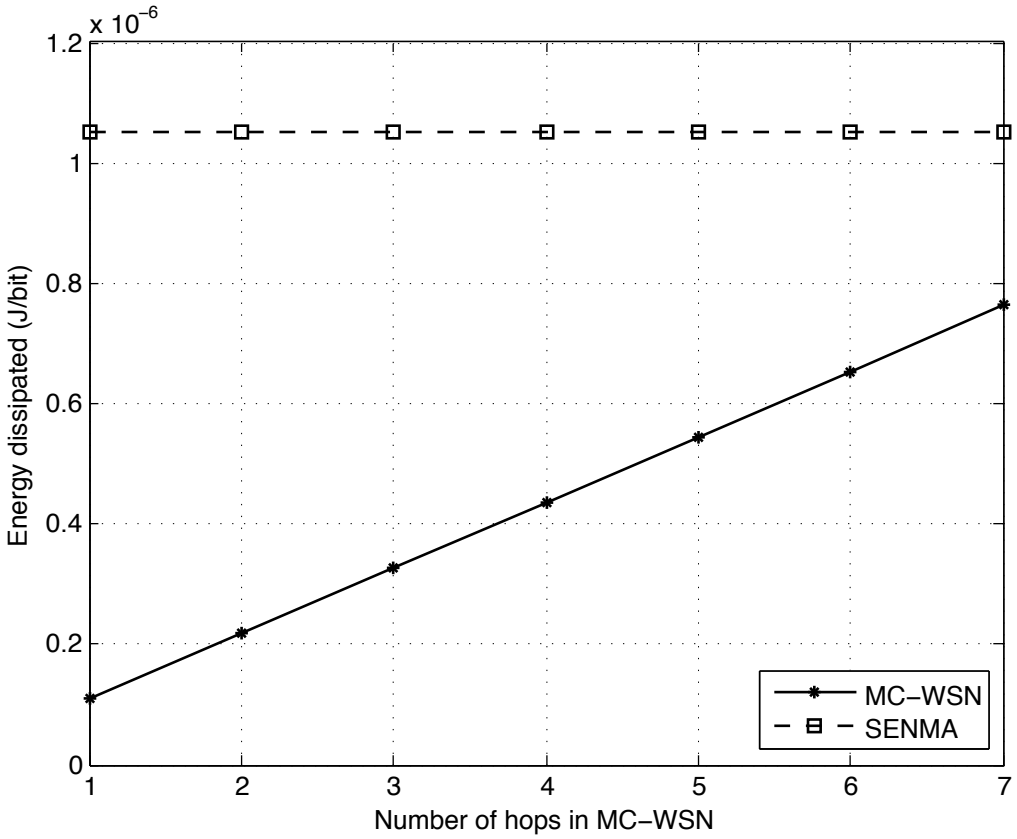


Figure 4.6: The energy dissipation (J/bit) vs. the number hops in N-hop MC-WSN, and comparison with the single hop SENMA network. Here, we set the cell radius $d_c = 100$m; for the MC-WSN, the per-hop distance for CHs is 30m; for SENMA, the per-hop distance and the MA coverage radius are equal to 10m; the path loss exponent $\beta = 2$, $n = 2000$, $E_{tx} = E_{rx} = 50$ nJ/bit, and $\epsilon_{pa} = 10$ pJ/bit/m^2.

indicates that either additional hops are utilized at the data delivery, or there is an unexpected large back-off time. In other words, the ratio between the actual delay and the average delay of a transmission can be used as an indicator for the detection of unexpected network failures or hostile attacks.

When the network synchronization is achieved, the detection of network failure problems can be implemented by including a time stamp in each packet representing the transmission time of the data. To compute the delay, the sink then compares the time stamp with the actual time of reception. If the delay is greater than a certain

threshold, then an exceptional behavior is detected.

Let the actual delay of BN i's transmission at time t be $\tilde{D}_i(t)$. The detection problem can be modeled using the binary hypothesis test. Let H_0 be the null hypothesis that represents normal network conditions ($\tilde{D}_i(t) \leq \bar{D}_i + \delta$), and H_1 be the alternative hypothesis that represents exceptional network behavior ($\tilde{D}_i(t) > \bar{D}_i + \delta$). That is,

$$H_0 : \quad \tilde{D}_i(t) \leq \bar{D}_i + \delta, \tag{4.7}$$

$$H_1 : \quad \tilde{D}_i(t) > \bar{D}_i + \delta, \tag{4.8}$$

where δ is a pre-defined parameter that reflects the time fluctuation in the system caused by queuing delay, and possible retransmissions due to the channel environment. Let $Z_i(t)$ be a binary indicator, such that it is equal to '1' if H_o is true at time t, and equal to '0' when H_1 is true. That is,

$$Z_i(t) = \begin{cases} 1 & H_o \text{ is true for node i at time t,} \\ 0 & H_1 \text{ is true for node i at time t.} \end{cases} \tag{4.9}$$

Define $\alpha(t) = \sum_{i=1}^{n} Z_i(t)/n$. $\alpha(t)$ can serve as an indicator on how well the network is functioning at time t. Ideally, the network should provide sufficient diversity to ensure that $\alpha(t)$ is close to unity at any time instant t.

When an exceptional network condition is detected, more measurements can be scheduled or requested for the network to locate the specific communication failure.

4.6.2 Access Authentication: Accountability and Privacy Protection

For N-hop networks, authentication can be implemented through a centralized authentication center (AuC) in the device-level. The hybrid network structure and routing diversity also enables the N-hop network to support network-level authentication.

For the device-level authentication, the authentication service is initiated by the fixed network and can be implemented through a simple challenge-response based authentication protocol. The authentication requires a shared secret key between each device and the centralized authentication center (AuC).

When a wireless device A needs to initiate a communication with another wireless device B, A makes an initial access request to B. The access request should contain the device identity ID, the BS that A accesses and can be authenticated through the AuC. After receiving the access request, B works as a proxy and forwards A's access request to the fixed BS and the AuC for A to be authenticated. The AuC then issues a random access $Challenge$ and send it to A through the BS and B. Upon receiving the $Challenge$, A computes the response $Response = E_{k_A}(Response)$ based on the $Challenge$ and the secret key k_A shared between A and the AuC. The computed response will be send back to the AuC through B and the BS for authentication. If the authentication is successful, the communication between A and B can be established. Otherwise, the access request from A will be rejected by B. This process only provides authentication of A to B. If two way authentication is required, B can be authenticated to A following the same procedures.

For the network-level authentication, the authentication can be split into two phases: the end-user device authentication to network access point (NAP) (such as BS, CH etc.) and the authentication between the NAPs through a mutually trusted

network server in the hierarchical structure such as the AuC. The end-user device authentication to the NAP can either be performed by the NAP locally or through the AuC. In both scenarios, the NAP can be viewed as a proxy for the end-user device and can provide end-user privacy protection to hide the communication events between the source and the destination.

In addition to authentication and accountability services, compared with traditional centralized network, the routing diversity in hybrid networks make the transmissions more robust under unexpected network failure or hostile attacks. At the same time, the routing diversity can also be exploited to achieve better privacy protection.

4.7 Summary

In this chapter, we first revisited the evolution of wireless systems and discussed the general design criteria of wireless networks. It was observed that in order to achieve a good balance among capacity, reliability, delay and flexibility, a network should be sufficiently structured and at the same time should provide adequate ad-hoc flexibility. On the evolution of wireless networks, this is reflected as the merging of centralized and ad-hoc networks, leading to the development of hybrid networks. In an effort to provide a unified framework for existing wireless systems, especially hybrid networks, we introduced the concept of N-hop networks. Under the N-hop framework, we discussed the analytical characterization of network performance in terms of throughput, delay, and energy efficiency, and also looked into the security perspectives on the balance between user accountability and privacy protection. It was shown that the N-hop framework includes most of the existing systems as special cases, and provides a flexible and tractable platform for network design, management, and performance evaluation. We also provided some related topics that may lead to further research.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This dissertation considered improving the reliability and efficiency of time-critical communications in wireless sensor networks, under both benign and hostile environments. The main conclusions are summarized in the following.

First, in Chapter 2, we considered the q-out-of-m fusion rule for reliable distributed detection in sensor networks with mobile access points (SENMA) under Byzantine attacks. Both static and dynamic attack strategies were discussed, where malicious sensors attack with fixed and time-varying probability, respectively. By exploiting the linear relationship between the network size and the scheme parameters, simple and effective q-out-of-m linear approaches were developed. We also derived a near-optimal closed-form solution for the fusion threshold based on the central limit theorem. Furthermore, we obtained an upper bound on the percentage of malicious nodes that can be tolerated using the q-out-of-m rule. It was found that the upper bound is determined by the sensors' detection capability and the attack probability of the malicious nodes.

We proved that the false alarm rate diminishes exponentially with the network size, even if the percentage of malicious sensors remains fixed. This implies that for a fixed percentage of malicious nodes, we can improve the network performance significantly by increasing the density of the nodes. To further improve the reliability of the data

fusion process, we proposed an effective malicious node detection scheme for adaptive data fusion under time-varying attacks, where the malicious sensors are identified and discarded then the fusion parameters are updated accordingly. We analyzed the detection procedure using the entropy-based trust model, and showed that it is optimal from the information theory point of view. It was observed that nodes launching dynamic attacks take longer time and more complex procedures to be detected as compared to those conducting static attacks. The adaptive fusion procedure has shown to provide a significant improvement in the system performance under both static and dynamic attacks.

Next, in Chapter 3, we proposed a mobile access coordinated wireless sensor networks (MC-WSN) architecture for reliable, efficient, and time-sensitive information exchange. The proposed MC-WSN exploits the mobile access points (MAs) to coordinate the network through deploying, replacing, and recharging nodes, as well as detecting malicious nodes and replacing them. Not only does MC-WSN resolve the network deployment problem, but it also prolongs the network lifetime actively and provides an efficient framework for time-sensitive information exchange. The hierarchical and heterogeneous structure makes MC-WSN a highly resilient, reliable, and scalable architecture. We provided the optimal topology design for MC-WSN such that the average number of hops from any sensor to the MA is minimized, and analyzed the performance of MC-WSN in terms of throughput, stability, delay, and energy efficiency. It was shown that with active network deployment and hop number control, MC-WSN achieves much higher throughput and considerably lower delay and energy consumption over the conventional SENMA.

Finally, in Chapter 4, we introduced the concept of the N-hop networks. Based on the N-hop concept, we proposed a unified framework for wireless networks and discussed general network design criteria for reliable and efficient communications. It was

shown that the N-hop framework includes existing network models as special cases, and provides a flexible and tractable platform for network design, management, and performance evaluation. Quantitative characterization of N-hop wireless networks was provided, along with discussions on different security perspectives. It was observed that in order to achieve a good balance between efficiency and reliability, a network should be sufficiently structured and at the same time should provide adequate ad-hoc flexibility. More specifically, the wireless network should have: (i) hierarchical structure, for efficient management, tracking of user accountability, as well as malicious node detection; (ii) multi-layer diversity, for higher reliability under unexpected network failure or malicious attacks; (iii) endpoint routing flexibility, for efficient utilization of the available resources.

5.2 Discussions for Future Work

We propose the following directions for future research.

Adaptive data fusion with soft decisions under malicious attacks with varying sensors' detection capabilities

- In Chapter 2; interesting results had been drawn on the effect of the network size on the reliability of the distributed detection under Byzantine attacks. For this purpose, hard decision model was mainly considered, where sensors quantize each observation into a single bit. Further research can be conducted to model the worst case Byzantine attacks when soft decisions are used, and on developing reliable distributed detection approaches in this case.

- It is also interesting to study the effect of possible varying detection capabilities of sensors on the reliability of the data fusion, and the effect of dividing the

cell into smaller regions over which the fusion rule can be applied with highest accuracy.

Security enhancements for MC-WSN under malicious attacks

- In Chapter 3, an MC-WSN architecture was proposed for reliable and time-sensitive information exchange. Detecting malicious sensors under general attack models in the multihop MC-WSN needs further investigations. More specifically, further research is needed to develop secure schemes in MC-WSN to combat different security threats, including various routing attacks and jamming.

Trade-off between security and efficiency

- Security is generally achieved at the cost of reduced efficiency. For example, to combat Byzantine attacks we need to employ large number of sensors for distributed detection; this would result in increasing the delay in the final decision making process. Also, to improve the privacy protection in multihop networks, routing diversity should be employed, which requires more resources to route the information to the intended destination. It is important to quantify the trade-off between the network efficiency versus its security strength, and explore secure schemes that can achieve a good compromise between them.

Further analysis for the N-hop networks

- Time-sensitive applications would impose delay constraints on data transmissions. Therefore, further analysis for the N-hop networks under delay constraints is important. Particularly, characterizing the multihop network-level capacity under both delay and power constraints remains an overwhelming research problem that needs further investigations.

APPENDICES

Appendix A Transmission Probability – Proof of Lemma 3.1

In this appendix, we obtain the uniform transmission probability of CHs within the coverage area of sink $k \in [0, 1, ..., K]$ in MC-WSN. We show that when hybrid TDMA/FDMA is used, the transmission probability $P(t_i^k = 1) \geq \frac{N_{Freq}}{N_{intf}\ n_k}$, $\forall i \in \mathcal{N}^k$, where n_k is the number CHs transmitting to sink k, $\mathcal{N}^k$ is the set of CHs within the coverage area of sink k, N_{intf} is the bandwidth reuse measure, and N_{Freq} is the number of frequencies available for simultaneous CHs transmissions within the same interference region.

The length of the TDMA schedule is the number of slots needed for the sink to receive one packet from all CHs within its coverage area. Since CHs within one hop from the sink relay the traffic of all other CHs within the sink's coverage area, then the largest length of a scheduling period can be obtained by finding the number of slots these CHs need to forward all the traffic they have (one packet from each source) to the sink. Here, we assume that each node has a packet to transmit and all packets are of the same importance, i.e. periodic data collection is considered. Note that in event driven scenarios, the length of the TDMA schedule could be less than that in the periodic data collection case.

First, consider $N_{Freq} = 1$. Then, a CH close to the sink can transmit only if other CHs within the interference region are silent. Hence, the length of the transmission schedule to sink k is:

$$S_k \leq \sum_{h=1}^{N_{intf}} \mathbb{N}_{h,k}(N_{f,h,k} + 1), \tag{A-1}$$

where $\mathbb{N}_{h,k}$ is the number of CHs at hop level h from sink k, and $(N_{f,h,k} + 1)$ is the total number of packets a node at hop level h sends to sink k. Note that the inequality

is mainly due to considering the largest number of interfering neighbors, which is when a CH arbitrary close to the sink location is considered [24].

Recall that N_{CH} is the total number of CHs in the cell. Let A_T be the total area of the cell, and A_k be the coverage area of sink k; then, $n_k = \frac{A_k}{A_T} N_{CH}$. Let $A_{h,k}$ be the area served by sink k until hop level h only. Hence, we have[1]:

$$\mathbb{N}_{h,k} \simeq \frac{\left(A_{h,k} - A_{h-1,k}\right)}{A_k} n_k, \quad N_{f,h,k} \simeq \frac{A_k - A_{h,k}}{A_k} \frac{n_k}{\mathbb{N}_{h,k}}. \tag{A-2}$$

When $N_{intf} = 2$, it follows from (A-1) that [24]:

$$\begin{aligned} S_k &\leq \frac{\left(A_{1,k}\right)}{A_k} n_k \left(\frac{A_k - A_{1,k}}{A_k} \frac{n_k}{\frac{\left(A_{1,k}\right)}{A_k} n_k} + 1 \right) \\ &\quad \frac{\left(A_{2,k} - A_{1,k}\right)}{A_k} n_k \left(\frac{A_k - A_{2,k}}{A_k} \frac{n_k}{\frac{\left(A_{2,k} - A_{1,k}\right)}{A_k} n_k} + 1 \right) \\ &\leq n_k + n_k \left(1 - \frac{A_{1,k}}{A_k}\right) \\ &\leq n_k \left(2 - \frac{A_{1,k}}{A_k}\right). \end{aligned} \tag{A-3}$$

Since $A_{1,k} < A_k$, then $0 < \frac{A_{1,k}}{A_k} < 1$, and we have $S_k \leq 2n_k$. Similarly, for general N_{intf} one can prove that:

$$S_k \leq N_{intf}\ n_k. \tag{A-4}$$

Note that if one of the nodes within one hop from sink k transmits the data of CH i to the sink, then this indicates that all intermediate CHs within the routing path from

[1]Note that $\mathbb{N}_{h,k}$ and $N_{f,h,k}$ are integer values in general, that is why we have the semi-equal sign in A-2.

i to the sink have transmitted this data. That is

$$Pr\{t_i^k = 1\} = Pr\{t_{i,1}^k = 1, ..., t_{i,N_i^k}^k = 1\} = Pr\{t_{i,1}^k = 1\}. \tag{A-5}$$

In other words, within a scheduling period of length S_k, all CHs would have transmitted their packets to the sink. Hence, the transmission probability $P(t_i^k = 1) = \frac{1}{S_k}$. Thus,

$$P(t_i^k = 1) \geq \frac{1}{N_{intf}\ n_k}. \tag{A-6}$$

For general N_{Freq}, we have

$$P(t_i^k = 1) \geq \frac{N_{Freq}}{N_{intf}\ n_k}. \tag{A-7}$$

Appendix B Traffic Load Calculations used in Proposition 3.3

In this appendix, we calculate the traffic around each sink (CCH/RCH) in MC-WSN, by obtaining the number of clusters at each hop level and the amount of traffic required to be forwarded on average by each cluster head at each hop level. More specifically, we get $N^O_{f,h,k}$, $N^I_{f,h,k}$, $\mathbb{N}^O_{h,k}$, and $\mathbb{N}^I_{h,k}$, $\forall h, k$.

Traffic around the CCH: Recall that all nodes within the radius R_o route their traffic to the CCH, and the maximum hop distance is R_c. Let A_T be the total area of the cell, $A_0 = \pi R_0^2$ be the area served by the CCH, and $A_{h,0} = \pi \min\{hR_c, R_o\}^2$ be the area until the hth hop level from the CCH. Therefore, the total number of cluster heads at the hth hop level from the CCH is:

$$\mathbb{N}_{h,0} \simeq \frac{\left(A_{h,0} - A_{h-1,0}\right)}{A_T} N_{CH}. \tag{B-1}$$

Each cluster head at hop level h from the CCH forwards the traffic of $N_{f,h,0}$ cluster heads at higher hop levels, where

$$N_{f,h,0} \simeq \frac{A_0 - A_{h,0}}{A_T} \frac{N_{CH}}{\mathbb{N}_{h,0}}. \tag{B-2}$$

We can obtain the total arrival rate at any CH within the service region of the CCH by substituting with $N_{f,h,0}$ in (3.19), which is then used in equations (3.30)-(3.32) to calculate the average delay per packet.

Traffic around the RCHs: Here, we analyze the traffic around the RCHs. Without loss of generality, we consider the RCH at $\theta = 0$. Note that, due to the uniformity of the network, other RCHs will have the same amount of traffic at each of the served nodes.

A CH with polar coordinates (x, θ), measured from the center of the cell, is at a distance from the RCH equals to $\sqrt{x^2 - 2xR_t\cos(\theta) + R_t^2}$, as illustrated in Figure 3.2. The farthest node at the hth hop level from a RCH is at a distance:

$$\sqrt{x^2 - 2xR_t\cos(\theta) + R_t^2} = hR_c. \tag{B-3}$$

That is,

$$\begin{aligned} x^2 - 2xR_t\cos(\theta) + R_t^2 &= (hR_c)^2, \\ x^2 - 2xR_t\cos(\theta) + R_t^2 - (hR_c)^2 &= 0. \end{aligned} \tag{B-4}$$

The solutions of x, denoted as $x_O(\theta, h)$ and $x_I(\theta, h)$, are functions of the angel θ and the hop level h, and are expressed as:

$$\begin{aligned} x_O(\theta, h) &= R_t\cos(\theta) + \sqrt{(hR_c)^2 - R_t^2\sin^2(\theta)}, \\ x_I(\theta, h) &= R_t\cos(\theta) - \sqrt{(hR_c)^2 - R_t^2\sin^2(\theta)}, \end{aligned} \tag{B-5}$$

where $x_O(\theta, h)$ corresponds to a CH at the outer region of the RCH (i.e., $x_O(\theta, h) \geq R_t$), and $x_I(\theta, h)$ corresponds to a CH at the inner region of the RCH (i.e., $x_I(\theta, h) < R_t$). In the following, we obtain $N^O_{f,h,k}$, $N^I_{f,h,k}$, $\mathbb{N}^O_{h,k}$, and $\mathbb{N}^I_{h,k}$, $\forall k \in \{1, ..., K\}$. Due to the symmetry of the architecture, we get the traffic in the inner and outer regions of the RCH at $\theta = 0$ focusing on the region where $\theta = [0, \frac{\pi}{K}]$. It can be seen from

equation (B-5) that the condition on θ for a valid solution is

$$\theta \leq \sin^{-1}\left(\frac{hR_c}{R_t}\right). \tag{B-6}$$

In the following, we consider the outer region and the inner region separately.

1. For the outer region, we have $x_O(\theta, h) \geq R_t$. It follows from (B-5) that:

$$\theta \;\leq\; \cos^{-1}\left(1 - \frac{(hR_c)^2}{2R_t^2}\right). \tag{B-7}$$

   There are two cases in the outer region:

   (a) *When the number of hops to the RCH is* $h \leq \lfloor\frac{d_c - R_t}{R_c}\rfloor$: In this case, the number of cluster heads up to hop level h is:

$$\mathbf{N}_{h,k}^{T,O} \simeq 2N_{CH}\int_{\theta=0}^{\theta_1(h)}\int_{x=R_t}^{x_O(\theta,h)} f_X(x)f_\theta(\theta)dxd\theta, \tag{B-8}$$

   where

$$\theta_1(h) = \min\left\{\frac{\pi}{K}, \sin^{-1}\left(\frac{hR_c}{R_t}\right), \cos^{-1}\left(1 - \frac{(hR_c)^2}{2R_t^2}\right)\right\}. \tag{B-9}$$

   This can be seen from equations (B-6) and (B-7), and from the fact that the maximum θ within the RCH region under consideration is $\frac{\pi}{K}$.

   (b) *When the number of hops to the RCH is* $h > \lfloor\frac{d_c - R_t}{R_c}\rfloor$: In this case, there is a lower bound on the angle θ such that $x_O(\theta, h) \leq d_c$. That is,

$$\begin{aligned} R_t\cos(\theta) + \sqrt{(hR_c)^2 - R_t^2\sin^2(\theta)} \;&\leq\; d_c, \\ d_c^2 + R_t^2 - (hR_c)^2 \;&\geq\; 2d_cR_t\cos(\theta). \end{aligned} \tag{B-10}$$

It follows that

$$\cos(\theta) \leq \frac{d_c^2 + R_t^2 - (hR_c)^2}{2R_t d_c},$$
$$\theta \geq \cos^{-1}\left(\frac{d_c^2 + R_t^2 - (hR_c)^2}{2R_t d_c}\right). \quad \text{(B-11)}$$

Note that since $0 \leq \theta \leq \frac{\pi}{K}$ and $K \geq 2$, then θ is inversely proportional to $\cos(\theta)$. From the above discussion, the number of cluster heads up to the hth hop level from the outer region of the RCH, where $h > \lfloor \frac{d_c - R_t}{R_c} \rfloor$, is:

$$\mathbf{N}_{h,k}^{T,O} \simeq 2N_{CH}\left[\int_{\theta=0}^{\theta_1(h)}\int_{x=R_t}^{x_O(\theta,h)} f_X(x) f_\theta(\theta) dx d\theta - \int_{\theta=0}^{\theta_2(h)}\int_{x=d_c}^{x_O(\theta,h)} f_X(x) f_\theta(\theta) dx d\theta\right]. \quad \text{(B-12)}$$

where $\theta_1(h)$ is defined in (B-9), and $\theta_2(h) = \cos^{-1}\left(\frac{d_c^2 + R_t^2 - (hR_c)^2}{2R_t d_c}\right)$.

2. For the inner region, we have $x_I(\theta, h) < R_t$. It follows from (B-5) that:

$$\theta \leq \cos^{-1}\left(1 - \frac{(hR_c)^2}{2R_t^2}\right). \quad \text{(B-13)}$$

Note that this condition is the same as the one in (B-7). There are two cases in the inner region:

(a) *When the number of hops to the RCH is* $h \leq \lfloor \frac{R_t - R_o}{R_c} \rfloor$: In this case, the number of cluster heads up to the hth hop level from the inner region is:

$$\mathbf{N}_{h,k}^{T,I} \simeq 2N_{CH}\int_{\theta=0}^{\theta_1(h)}\int_{x=x_I(\theta,h)}^{R_t} f_X(x) f_\theta(\theta) dx d\theta, \quad \text{(B-14)}$$

where $\theta_1(h)$ is given in (B-9).

(b) *When the number of hops to the RCH is* $h > \lfloor \frac{R_t - R_o}{R_c} \rfloor$: In this case, the integration is only over the area where $x \geq R_o$, which imposes a lower bound on θ. That is, by following a similar procedure, we have:

$$\begin{aligned} R_t \cos(\theta) - \sqrt{(hR_c)^2 - R_t^2 \sin^2(\theta)} &\geq R_o, \\ (R_t \cos(\theta) - R_o)^2 &\geq (hR_c)^2 - R_t^2 \sin^2(\theta), \\ R_t^2 + R_o^2 - 2R_t R_o \cos(\theta) &\geq (hR_c)^2. \end{aligned} \tag{B-15}$$

It follows that,

$$\begin{aligned} \cos(\theta) &\leq \frac{R_t^2 + R_o^2 - (hR_c)^2}{2R_o R_t}, \\ \theta &\geq \cos^{-1}\left(\frac{R_t^2 + R_o^2 - (hR_c)^2}{2R_o R_t}\right). \end{aligned} \tag{B-16}$$

Therefore, the total number of cluster heads in the inner region up to the hth hop level from the RCH, where $h > \lfloor \frac{R_t - R_o}{R_c} \rfloor$, is:

$$\begin{aligned} \mathbf{N}_{h,k}^{T,I} \simeq 2N_{CH} \Bigg[&\int_{\theta=0}^{\theta_1(h)} \int_{x=x_I(\theta,h)}^{R_t} f_X(x) f_\theta(\theta) dx d\theta \\ &- \int_{\theta=0}^{\theta_3(h)} \int_{x=x_I(\theta,i)}^{R_o} f_X(x) f_\theta(\theta) dx d\theta \Bigg], \end{aligned} \tag{B-17}$$

where $\theta_3(h) = \max\left\{0, \cos^{-1}\left(\frac{R_t^2 + R_o^2 - (hR_c)^2}{2R_t R_o}\right)\right\}$.

Note that the above integrals over θ do not have closed-form solution, but can be evaluated numerically.

From the above discussions, we obtain the number of cluster heads at each hop

level from the outer and inner regions as follows:

$$\begin{aligned} \mathbb{N}^{O}_{h,k} &= \mathbf{N}^{T,O}_{h,k} - \mathbf{N}^{T,O}_{h-1,k}, \\ \mathbb{N}^{I}_{h,k} &= \mathbf{N}^{T,I}_{h,k} - \mathbf{N}^{T,I}_{h-1,k}. \end{aligned} \tag{B-18}$$

We assume that the traffic load is divided evenly among the CHs at each hop level. Therefore, the average amount of traffic that a cluster head at the hth hop level from the outer/inner region ($N^{O}_{f,h,k}/N^{I}_{f,h,k}$) forwards is:

$$\begin{aligned} N^{O}_{f,h,k} &= \frac{\mathbf{N}^{T,O}_{\mathbb{N}^{O}_{k},k} - \mathbf{N}^{T,O}_{h,k}}{\mathbb{N}^{O}_{h,k}}, \\ N^{I}_{f,h,k} &= \frac{\mathbf{N}^{T,I}_{\mathbb{N}^{I}_{k},k} - \mathbf{N}^{T,I}_{h,k}}{\mathbb{N}^{I}_{h,k}}, \end{aligned} \tag{B-19}$$

where $\mathbb{N}^{O}_{k}$ and $\mathbb{N}^{I}_{k}$ are the maximum number of hops to RCH k from the outer and inner regions, respectively.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] A. Bharathidasas and V. Anand, "Sensor networks: An overview," *Technical report, Dept. of Computer Science, University of California at Davis*, 2002.

[2] C. Chong and S. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247 – 2056, Aug. 2003.

[3] M. Tubaishat, P. Zhuang, Q. Qi, and Y. Shang, "Wireless sensor networks in intelligent transportation systems," *Wireless Communications and Mobile Computing*, vol. 9, no. 3, pp. 287 – 302, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1002/wcm.v9:3

[4] C.-Y. Chong and S. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247 – 1256, 2003.

[5] J. Hill, M. Horton, R. Kling, and L. Krishnamurthy, "The platforms enabling wireless sensor networks," *Communications of the ACM*, vol. 47, no. 6, pp. 41 – 46, Jun. 2004. [Online]. Available: http://doi.acm.org/10.1145/990680.990705

[6] B. Tavli, K. Bicakci, R. Zilan, and J. Barcelo-Ordinas, "A survey of visual sensor network platforms," *Multimedia Tools and Applications*, vol. 60, pp. 689 – 726, 2012. [Online]. Available: http://dx.doi.org/10.1007/s11042-011-0840-z

[7] D. T. Fokum, D. Victor, S. Frost, D. Gary, and J. Minden, "An evaluation of sensing platforms used for sensor network research," Tech. Rep., 2007.

[8] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," *Fourth International Symposium on Information Processing in Sensor Networks, IPSN'05*, pp. 364 – 369, Apr. 2005.

[9] G. Zhou, C. Huang, T. Yan, T. He, J. A. Stankovic, and T. F. Abdelzaher, "Mmsn: Multi-frequency media access control for wireless sensor networks," *Proceedings 25th IEEE International Conference on Computer Communications INFOCOM'06*, pp. 1 – 13, Apr. 2006.

[10] Crossbow, "Mica2: Wireless measurement system," 2004.

[11] ——, "Mica2dot: Wireless microsensor mote," 2003.

[12] ——, "Micaz: Wireless measurement system," 2004.

[13] L. Nachman, R. Kling, R. Adler, J. Huang, and V. Hummel, "The Intel® mote platform: a bluetooth-based sensor network for industrial monitoring," *Fourth International Symposium on Information Processing in Sensor Networks, IPSN'05*, pp. 437 – 442, Apr. 2005.

[14] A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 3, pp. 657 – 667, Mar. 2005.

[15] G. Zhang, Y. Xu, X. Wang, and M. Guizani, "Capacity of hybrid wireless networks with directional antenna and delay constraint," *IEEE Transactions on Communications*, vol. 58, no. 7, pp. 2097 – 2106, Jul. 2010.

[16] C.-C. Shen, C. Srisathapornphat, and C. Jaikaeo, "Sensor information networking architecture and applications," *IEEE Personal Communications*, vol. 8, no. 4, pp. 52 – 59, 2001.

[17] G. Mergen, Z. Qing, and L. Tong, "Sensor networks with mobile access: Energy and capacity considerations," *IEEE Transactions on Communications*, vol. 54, no. 11, pp. 2033 – 2044, Nov. 2006.

[18] W. Liu, K. Lu, J. Wang, L. Huang, and D. Wu, "On the throughput capacity of wireless sensor networks with mobile relays," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1801 – 1809, May 2012.

[19] W. Liu, K. Lu, J. Wang, G. Xing, and L. Huang, "Performance analysis of wireless sensor networks with mobile sinks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 6, pp. 2777 – 2788, Jul. 2012.

[20] C. Avram, D. Radu, A. Astilean, and V. Cosma, "Ant routing protocol in a zigbee ad hoc sensors network for radiation level monitoring," *IEEE International Conference on Automation Quality and Testing Robotics, AQTR'10*, vol. 3, pp. 1 – 6, May.

[21] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135 – 1149, May 2002.

[22] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388 – 404, Mar. 2000.

[23] C. P. Chan, S. C. Liew, and A. Chan, "Many-to-one throughput capacity of IEEE 802.11 multihop wireless networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 4, pp. 514 – 527, April 2009.

[24] E. J. Duarte-Melo and M. Liu, "Data-gathering wireless sensor networks: organization and capacity," *Computer Networks*, vol. 43, no. 4, pp. 519 – 537, 2003, wireless Sensor Networks. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128603003578

[25] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572 – 584, Sep. 1979.

[26] P. Gupta and P. Kumar, "Towards an information theory of large networks: an achievable rate region," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1877 – 1894, Aug. 2003.

[27] H. El Gamal, "On the scaling laws of dense wireless sensor networks: the data gathering channel," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 1229 – 1234, Mar. 2005.

[28] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477 – 486, Aug. 2002.

[29] S. Gandham, M. Dawande, R. Prakash, and S. Venkatesan, "Energy efficient schemes for wireless sensor networks with multiple mobile base stations," *IEEE Global Telecommunications Conference, GLOBECOM'03*, vol. 1, pp. 377 – 381, Dec. 2003.

[30] J. Luo and J.-P. Hubaux, "Joint mobility and routing for lifetime elongation in wireless sensor networks," *IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM'05*, vol. 3, pp. 1735 – 1746, Mar. 2005.

[31] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of byzantine attacks," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 16 – 29, Jan. 2009.

[32] J. Newsome, E. Shi, D. Song, and A. Perrig, "The sybil attack in sensor networks: analysis & defenses," *Proceedings of the 3rd international symposium on Information processing in sensor networks*, pp. 259 – 268, 2004. [Online]. Available: http://doi.acm.org/10.1145/984622.984660

[33] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," *IEEE International Workshop on Sensor Network Protocols and Applications*, pp. 113 – 127, May 2003.

[34] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," *IEEE Communications Surveys Tutorials*, vol. 11, no. 2, pp. 52 – 73, 2009.

[35] Y. Zhang, W. Liu, W. Lou, and Y. Fang, "Location-based compromise-tolerant security mechanisms for wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 247 – 260, Feb.

[36] C. Chen, M. Song, and G. Hsieh, "Intrusion detection of sinkhole attacks in large-scale wireless sensor networks," *IEEE International Conference on Wireless Communications, Networking and Information Security, WCNIS'10*, pp. 711 – 716, Jun.

[37] Y. L. Sun, W. Yu, Z. Han, and K. Liu, "Information theoretic framework of trust modeling and evaluation for ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 305 – 317, Feb. 2006.

[38] D. Martins and H. Guyennet, "Wireless sensor network attacks and security mechanisms: A short survey," *13th International Conference on Network-Based Information Systems, NBiS'10*, pp. 313 – 320, Sept. 2010.

[39] D. Malan, "Crypto for tiny objects," Tech. Rep., 2004.

[40] L. Zhang and T. Li, "Anti-jamming message-driven frequency hopping; part ii: Capacity analysis under disguised jamming," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 80 – 88, 2013.

[41] W. Xu, W. Trappe, and Y. Zhang, "Defending wireless sensor networks from radio interference through channel adaptation," *ACM Transactions on Sensor Networks*, vol. 4, no. 4, pp. 1 – 34, Aug. 2008. [Online]. Available: http://doi.acm.org/10.1145/1387663.1387664

[42] M. Abdelhakim, L. Lightfoot, J. Ren, and T. Li, "Distributed detection in mobile access wireless sensor networks under byzantine attacks," *IEEE Transactions on Parallel and Distributed Systems, accepted*.

[43] M. Abdelhakim, J. Ren, and T. Li, "Mobile access coordinated wireless sensor networks – topology design and throughput analysis," *IEEE Global Communications Conference, GLOBECOM'13*, 2013.

[44] T. Li, M. Abdelhakim, and J. Ren, "N-hop networks — a general framework for wireless systems," *IEEE Wireless Communications Magazine, accepted*.

[45] M. Abdelhakim, L. Lightfoot, and T. Li, "Reliable data fusion in wireless sensor networks under byzantine attacks," *IEEE Military Communications Conference, MILCOM'11*, Nov. 2011.

[46] M. Abdelhakim, L. Zhang, J. Ren, and T. Li, "Cooperative sensing in cognitive networks under malicious attack," *IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP'11*, pp. 3004 – 3007, May 2011.

[47] Y.-C. Wang and Y.-C. Tseng, "Distributed deployment schemes for mobile wireless sensor networks to ensure multilevel coverage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 9, pp. 1280 – 1294, Sept. 2008.

[48] P. Barooah, H. Chenji, R. Stoleru, and T. Kalmar-Nagy, "Cut detection in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 3, pp. 483 – 490, Mar. 2012.

[49] A. Perrig, R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler, "Spins: security protocols for sensor networks," *Wireless Networks*, vol. 8, pp. 521 – 534, Sept. 2002. [Online]. Available: http://dx.doi.org/10.1023/A:1016598314198

[50] C. Karlof, N. Sastry, and D. Wagner, "Tinysec: a link layer security architecture for wireless sensor networks," *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pp. 162 – 175, 2004. [Online]. Available: http://doi.acm.org/10.1145/1031495.1031515

[51] L. Lightfoot, J. Ren, and T. Li, "An energy efficient link-layer security protocol for wireless sensor networks," *IEEE International Conference on Electro/Information Technology, EIT 2007*, pp. 233 – 238, May 2007.

[52] I. Rodhe, C. Rohner, and A. Achtzehn, "n-lqa: n-layers query authentication in sensor networks," *IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems, MASS'07*, pp. 1 – 6, Oct. 2007.

[53] W. Zhang, N. Subramanian, and G. Wang, "Lightweight and compromise-resilient message authentication in sensor networks," *IEEE 27th Conference on Computer Communications, INFOCOM'08*, pp. 1418 – 1426, Apr. 2008.

[54] H. Chan, A. Perrig, and D. Song, "Secure hierarchical in-network aggregation in sensor networks," *Proceedings of the 13th ACM conference on Computer and communications security, ACM CCS'06*, pp. 278 – 287, 2006.

[55] H. Kumar, D. Sarma, and A. Kar, "Security threats in wireless sensor networks," *IEEE Aerospace and Electronic Systems Magazine*, vol. 23, no. 6, pp. 39 – 45, Jun. 2008.

[56] B. Awerbuch, R. Curtmola, H. D., N.-R. C., and R. H., "Mitigating byzantine attacks in ad hoc wireless networks," *Technical report version 1*, Mar. 2004.

[57] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of byzantine attack in large wireless sensor networks," *IEEE Military Communications Conference, MILCOM'06*, pp. 1 – 4, Oct. 2006.

[58] O. Kosut and L. Tong, "Distributed source coding in the presence of byzantine sensors," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2550 – 2565, Jun. 2008.

[59] G. Latif-Shabgahi, "A novel algorithm for weighted average voting used in fault tolerant computing systems," *Microprocessors and Microsystems*, vol. 28, no. 7, pp. 357 – 361, 2004.

[60] Y. Brun, G. Edwards, J. Y. Bang, and N. Medvidovic, "Smart redundancy for distributed computation," *2011 31st International Conference on Distributed Computing Systems, ICDCS'11*, pp. 665 – 676, Jun. 2011.

[61] S. Wang, K. Yan, and H. Hsieh, "The byzantine agreement under mobile network," *IEEE International Conference on, Networking, Sensing and Control*, vol. 1, pp. 46 – 51, Mar. 2004.

[62] L. F. G. Sarmenta, "Sabotage-tolerance mechanisms for volunteer computing systems," *Future Generation Computer Systems*, vol. 18, pp. 561 – 572, 2002.

[63] P. Sridhar, A. Madni, and M. Jamshidi, "Hierarchical aggregation and intelligent monitoring and control in fault-tolerant wireless sensor networks," *IEEE Systems Journal*, vol. 1, no. 1, pp. 38 – 54, Sept. 2007.

[64] Q. Tian and E. Coyle, "Optimal distributed detection in clustered wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3892 – 3904, Jul. 2007.

[65] A. Rawat, P. Anand, H. Chen, and P. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 774 – 786, Feb. 2011.

[66] R. Viswanathan and V. Aalo, "On counting rules in distributed detection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 5, pp. 772 – 775, May 1989.

[67] R. Niu and P. Varshney, "Performance analysis of distributed detection in a random sensor field," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 339 – 349, Jan. 2008.

[68] V. Aalo and G. Efthymoglou, "Decision fusion schemes for wireless sensor networks operating in a nakagami-m fading environment," *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'09*, pp. 2720 – 2724, Sept. 2009.

[69] H. Wang, L. Lightfoot, and T. Li, "On phy-layer security of cognitive radio: Collaborative sensing under malicious attacks," *44th Annual Conference on Information Sciences and Systems, CISS'12*, pp. 1 – 6, Mar. 2010.

[70] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," *IEEE Military Communications Conference, MILCOM'03*, vol. 1, pp. 688 – 693, Oct. 2003.

[71] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523 – 531, Apr. 1967.

[72] M. R. Fellows, F. V. Fomin, D. Lokshtanov, F. Rosamond, S. Saurabh, and Y. Villanger, "Local search: Is brute-force avoidable?" *Journal of Computer and System Sciences*, vol. 78, no. 3, pp. 707 – 719, 2012.

[73] M. Abdelhakim, J. Ren, and T. Li, "Reliable cooperative sensing in cognitive networks," *Wireless Algorithms, Systems, and Applications, WASA'12, Springer Berlin Heidelberg*, vol. 7405, pp. 206 – 217, 2012.

[74] E. R. Love, "64.4 some logarithm inequalities," *The Mathematical Gazette*, vol. 64, no. 427, pp. 55 – 57, 1980. [Online]. Available: http://www.jstor.org/stable/3615890

[75] U. Madhow, *Fundamentals of digital communication.* Cambridge University Press, 2008, p. 483.

[76] W. Wang, H. Li, Y. Sun, and Z. Han, "Catchit: Detect malicious nodes in collaborative spectrum sensing," *IEEE Global Telecommunications Conference, GLOBECOM'09*, pp. 1 – 6, 2009.

[77] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN'05.*, pp. 131 – 136, 2005.

[78] M. Abdelhakim, L. Lightfoot, J. Ren, and T. Li, "Architecture design of mobile access coordinated wireless sensor networks," *IEEE International Conference on Communications, ICC'13*, pp. 1720 – 1724, Jun. 2013.

[79] I. Maza, F. Caballero, J. Capitan, J. Martinez-de Dios, and A. Ollero, "A distributed architecture for a robotic platform with aerial sensor transportation and self-deployment capabilities," *Journal of Field Robotics*, vol. 28, no. 3, pp. 303 – 328, 2011. [Online]. Available: http://dx.doi.org/10.1002/rob.20383

[80] P. Corke, S. Hrabar, R. Peterson, D. Rus, S. Saripalli, and G. Sukhatme, "Autonomous deployment and repair of a sensor network using an unmanned aerial vehicle," *IEEE International Conference on Robotics and Automation, ICRA'04*, vol. 4, pp. 3602 – 3608, 26-May 1, 2004.

[81] H.-C. Chen, D. Kung, D. Hague, M. Muccio, and B. Poland, "Collaborative compressive spectrum sensing in a *UAV* environment," *IEEE Military Communications Conference, MILCOM'11*, Nov. 2011.

[82] J. Luo and A. Ephremides, "On the throughput, capacity, and stability regions of random multiple access," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2593 – 2607, Jun. 2006.

[83] G. Mergen and L. Tong, "Maximum asymptotic stable throughput of opportunistic slotted ALOHA and applications to CDMA networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1159 – 1163, Apr. 2007.

[84] A. Behzad and I. Rubin, "High transmission power increases the capacity of ad hoc wireless networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 156 – 165, 2006.

[85] A. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," *Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, INFOCOM'04*, vol. 1, 2004.

[86] M. Nekoui and H. Pishro-Nik, "Throughput scaling laws for vehicular ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2895 – 2905, 2012.

[87] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2105 – 2115, 2001.

[88] L. Kleinrock, *Communication nets; stochastic message flow and delay.* McGraw-Hill, 1964.

[89] W. Szpankowski, *Stability Conditions for Some Multiqueue Distributed Systems: Buffered Random Access Systems*, ser. Technical report. Purdue University, Department of Computer Sciences, 1992, no. 29. [Online]. Available: http://books.google.com/books?id=DcrHuAAACAAJ

[90] R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple-access system," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 918 – 930, Sep. 1988.

[91] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite-user slotted ALOHA with multipacket reception," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2636 – 2656, Jul. 2005.

[92] H. Wang and T. Li, "Hybrid ALOHA: A novel mac protocol," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5821 – 5832, Dec. 2007.

[93] R. M. Loynes, "The stability of a queue with non-independent inter-arrival and service times," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 58, pp. 497 – 520, 1962.

[94] S. Hanly and D. Tse, "Multiaccess fading channels. ii. delay-limited capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2816 – 2831, 1998.

[95] R. Negi and J. Cioffi, "Delay-constrained capacity with causal feedback," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2478 – 2494, 2002.

[96] B. W. Conolly, "The waiting time process for a certain correlated queue," *Operations Research*, vol. 16, no. 5, pp. 1006 – 1015, 1968. [Online]. Available: http://www.jstor.org/stable/168493

[97] K. Fendick, V. Saksena, and W. Whitt, "Dependence in packet queues," *IEEE Transactions on Communications*, vol. 37, no. 11, pp. 1173 – 1183, 1989.

[98] E. Modiano, J. Wieselthier, and A. Ephremides, "A simple analysis of average queueing delay in tree networks," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 660 – 664, 1996.

[99] M. Neely, C. Rohrs, and E. Modiano, "Equivalent models for queueing analysis of deterministic service time tree networks," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3576 – 3584, 2005.

[100] G. Gupta and N. Shroff, "Delay analysis for multi-hop wireless networks," *IEEE INFOCOM*, pp. 2356 – 2364, 2009.

[101] D. Bertsekas and R. Gallager, *Data Networks*. Prentice-Hall, 1992.

[102] L. Galluccio and S. Palazzo, "End-to-end delay and network lifetime analysis in a wireless sensor network performing data aggregation," *IEEE Global Telecommunications Conference, GLOBECOM'09*, pp. 1 – 6, 2009.

[103] N. Bisnik and A. A. Abouzeid, "Queuing network models for delay analysis of multihop wireless ad hoc networks," *Ad Hoc Networks*, vol. 7, no. 1, pp. 79 – 97, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870507001813

[104] Y. Wang, M. C. Vuran, and S. Goddard, "Cross-layer analysis of the end-to-end delay distribution in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 305 – 318, 2012.

[105] A. Sample, D. Yeager, P. Powledge, A. Mamishev, and J. Smith, "Design of an rfid-based battery-free programmable sensing platform," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 11, pp. 2608 – 2615, 2008.

[106] J. Ren, Y. Li, and T. Li, "Spm: Source privacy for mobile ad hoc networks," *EURASIP Journal on Wireless Communications and Networking*, 2010.

[107] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "An Aloha protocol for multihop mobile wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 421 – 436, 2006.

[108] S. Choudhury and J. D. Gibson, "Ergodic capacity, outage capacity, and information transmission over Rayleigh fading channels," *Information Theory and Applications Workshop*, 2007.

[109] A. Abdrabou and W. Zhuang, "Service time approximation in IEEE 802.11 single-hop ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 1, pp. 305 – 313, 2008.

[110] C. Goldschmidt, "The chen-stein method for convergence of distributions," *Masters-level essay, University of Cambridge*, 2000.

[111] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913 – 931, 1994.

[112] A. Somasundara, A. Kansal, D. Jea, D. Estrin, and M. Srivastava, "Controllably mobile infrastructure for low energy embedded networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 8, pp. 958 – 973, Aug. 2006.

[113] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660 – 670, Oct. 2002.

[114] "1900 to 1999," *http://www.deutsches-telefon-museum.eu/1900.htm*, Dec. 2007.

[115] L. H. Anderson, "The first walkie-talkie radio - an affectionate look back in time and some thoughts about the first true fabled walkie-talkie," Jun. 2005.

[116] R. Frenkiel, "Creating cellular: A history of the AMPS project (1971-1983) [History of Communications]," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 14 – 24, 2010.

[117] D. H. Ring, "Mobile telephony - wide area coverage - case 20564," 1947.

[118] "IEEE Standard for Local and metropolitan area networks-Part 16: Air Interface for Fixed Broadband Wireless Access Systems," *IEEE Std 802.16-2004*, 2004.

[119] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 10)," *3GPP TS 36.300 V10.4.0 (2011-06)*, 2011.

[120] I. Chlamtac, M. Conti, and J. J.-N. Liu, "Mobile ad hoc networking: imperatives and challenges," *Ad Hoc Networks*, vol. 1, no. 1, pp. 13 – 64, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870503000131

[121] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102 – 114, 2002.

[122] F. Hu, M. Jiang, L. Celentano, and Y. Xiao, "Robust medical ad hoc sensor networks (masn) with wavelet-based ECG data mining," *Ad Hoc Networks*, vol. 6, no. 7, pp. 986 – 1012, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S157087050700145X

[123] A. Zadeh, B. Jabbari, R. Pickholtz, and B. Vojcic, "Self-organizing packet radio ad hoc networks with overlay (soprano)," *IEEE Communications Magazine*, vol. 40, no. 6, pp. 149 – 157, 2002.

[124] "IEEE Draft Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 10: Mesh Networking," *IEEE P802.11s/D10.0, March 2011*, pp. 1 – 379, 2011.

[125] G. Neonakis Aggelou and R. Tafazolli, "On the relaying capability of next-generation GSM cellular networks," *IEEE Personal Communications*, vol. 8, no. 1, pp. 40 – 47, 2001.

[126] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to lte-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42 – 49, Dec. 2009.

[127] P. Janis, C. Yu, K. Doppler, C. Ribeiro, C. Wijting, K. Hugl, O. Tirkkonen, and V. Koivunen, "Device-to-device communication underlaying cellular communications systems," *International Journal of Communications, Network and System Sciences*, vol. 2, no. 3, pp. 169 – 178, Jun. 2009.

[128] M.-H. Han, B.-G. Kim, and J.-W. Lee, "Subchannel and Transmission Mode Scheduling for D2D Communication in OFDMA Networks," *IEEE Vehicular Technology Conference, VTC 2012*, pp. 1 – 5, 2012.

[129] H.-Y. Hsieh and R. Sivakumar, "On using peer-to-peer communication in cellular wireless data networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 1, pp. 57 – 72, 2004.

[130] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10 – 21, 2011.

[131] J. Sydir and R. Taori, "An evolved cellular system architecture incorporating relay stations," *IEEE Communications Magazine*, vol. 47, no. 6, pp. 115 – 121, 2009.

[132] A. C. Guyton and J. E. Hall, *Textbook of medical physiology.* Philadelphia, Saunders, 2000.

[133] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572 – 584, sep 1979.

[134] Y. Chen and J. Andrews, "An upper bound on multihop transmission capacity with dynamic routing selection," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3751 – 3765, june 2012.

[135] N. Bisnik and A. Abouzeid, "Queuing network models for delay analysis of multihop wireless ad hoc networks," *Proceedings of the international conference on Wireless communications and mobile computing, IWCMC'06*, pp. 773 – 778, 2006. [Online]. Available: http://doi.acm.org/10.1145/1143549.1143704

[136] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135 – 1149, 2002.