ANALYSIS AND MODELING OF ROUTING AND SECURITY PROBLEMS IN WIRELESS SENSOR NETWORKS WITH MATHEMATICAL PROGRAMMING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS INSTITUTE
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DAVUT İNCEBACAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
PHILOSOPHY OF DOCTORATE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

DECEMBER 2013

ANALYSIS AND MODELING OF ROUTING AND SECURITY PROBLEMS IN WIRELESS SENSOR NETWORKS WITH MATHEMATICAL PROGRAMMING

Submitted by Davut İNCEBACAK in partial fulfillment of the requirements for the degree of **Philosophy of Doctorate in Information Systems, Middle East Technical University by,**

Prof. Dr. Nazife Baykal _____
Director, Informatics Institute

Prof. Dr. Yasemin Yardımcı Çetin _____
Head of Department, Information Systems

Prof. Dr. Nazife Baykal _____
Supervisor, Information Systems, METU

Assoc. Prof. Dr. Kemal Bıçakcı _____
Co-Supervisor, Computer Engineering, TOBB ETU

Examining Comitte Members:

Assoc. Prof. Dr. Altan Koçyiğit _____
Information Systems, METU

Prof. Dr. Nazife Baykal _____
Information Systems, METU

Assist. Prof. Dr. Erhan Eren _____
Information Systems, METU

Assoc. Prof. Dr. Bülent Tavlı _____
Electrical and Electronics Engineering, TOBB ETU

Assoc. Prof. Dr. Alptekin Temizel _____
Work Based Learning, METU

Date: 02.12.2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: DAVUT İNCEBACAK

Signature :

ABSTRACT

ANALYSIS AND MODELING OF ROUTING AND SECURITY PROBLEMS IN WIRELESS SENSOR NETWORKS WITH MATHEMATICAL PROGRAMMING

İncebacak, Davut
Ph.D., Department of Information Systems
Supervisor : Prof. Dr. Nazife Baykal
Co-Supervisor : Assoc. Prof. Dr. Kemal Bıçakcı

December 2013, 130 pages

Wireless Sensor Networks (WSNs) are composed of battery powered small sensor nodes with limited processing, memory and energy resources. Self organization property together with infrastructureless characteristics of WSNs make them favorable solutions for many applications. Algorithms and protocols developed for WSNs must consider the characteristics and constraints of WSNs but since battery replenishment is not possible or highly challenging for sensor nodes, one of the major concerns in designing network protocols and algorithms is to achieve energy efficiency and to extend the network lifetime. Hence, energy efficient solutions are required for routing and security problems in WSNs. In this dissertation, we model and analyze several routing and security problems in WSNs. We first study the impact of spatial granularity of measurements on the energy requirements of sensor network. We then investigate the energy cost of survivability in the presence of physical attacks. We next study the impact of the number of routing paths on network wide energy balancing under optimal operating conditions. Finally, we investigate the energy cost of route diversity to improve the security of WSNs against adversaries attempting to obtain sensitive sensor data.

We contribute to the literature by developing novel mathematical programming frameworks

and presenting a comprehensive high level analysis of the several routing and security problems in WSNs. The novel mathematical programming frameworks presented in this thesis can be used with minor modifications for future analysis of different routing and security problems in WSNs.

Keywords: Wireless Sensor Networks, Mathematical Programming, Security, Routing, Energy Efficiency

ÖZ

KABLOSUZ ALGILAYICI AĞLARDA YÖNLENDİRME VE GÜVENLİK PROBLEMLERİNİN MATEMATİKSEL PROGRAMLAMA VASITASIYLA MODELLENMESİ VE ANALİZİ

İncebacak, Davut
Doktora, Bilişim Sistemleri Bölümü
Tez Yöneticisi : Prof. Dr. Nazife Baykal
Ortak Tez Yöneticisi : Doç. Dr. Kemal Bıçakcı

Aralık 2013, 130 sayfa

Kablosuz Algılayıcı Ağlar (KAA) pil ile çalışan, sınırlı işleme, hafıza ve enerji kaynaklarına sahip küçük algılayıcılardan oluşur. KAA'da bulunan kendi kendine organize olma ve kablosuz iletişim kullanma özellikleri, KAA'ların birçok uygulamada kullanılabilmesine imkan vermiştir. KAA için geliştirilen algoritmalar ve protokoller KAA'ın özelliklerini ve kısıtlamalarını dikkate almalıdır ancak algılayıcıların pillerinin değişimi genellikle mümkün olmadığı ya da çok zor olduğu için, geliştirilen algoritmalar ve protokoller için dikkate alınan temel tasarım parametrelerinin başında enerji verimliliği ve yaşam süresinin eniyilenmesi gelir. Bundan dolayı KAA'da yönlendirme ve güvenlik problemleri için enerji efektif çözümler gerekir. Bu tezde KAA'da yönlendirme ve güvenlik problemlerinin modellenmesi ve analizi üzerinde çalışmalar yaptık. İlk olarak KAA'ın ölçüm yaptığı parçaların büyüklüğünün KAA'ın enerji harcanımına olan etkisi üzerine çalışma yaptık. Sonrasında fiziksel saldırıların önlenmesinin enerji maliyetini araştırdık. Daha sonra yönlendirme yollarının toplam sayısının enerjinin etkin kullanılması üzerine olan etkisini inceledik. Son olarak KAA'ın veri güvenliğini sağlamak için çoklu yol kullanmanın enerji maliyetini araştırdık.

Literatüre orjinal matematiksel programlama çerçeveleri geliştererek ve KAA'da birçok yönlendirme ve güvenlik problemlerinin kapsamlı analizini yaparak katkıda bulunduk. Bu tezde anlatılan matematiksel programlama çerçeveleri küçük değişikliklerle gelecekte karşılaşılabilecek yönlendirme ve güvenlik problemlerinin analizinde kullanılabilir.

Anahtar Kelimeler: Kablosuz Algilayıcı Ağlar, Matematiksel Programlama, Guvenlik, Yonlendirme, Enerji Etkinligi

dedicated to my wife Tuba and my daughter Zeynep Işık

ACKNOWLEDGMENTS

First of all, I want to present my gratitude to Dr. Kemal Bıçakcı for providing motivation, guidance and all kind of support during my PhD study. I also want to thank Dr. Nazife Baykal, for encouraging me, providing significant support and supervision despite to her intense schedule.

Also, I owe much to the committee members Dr. Altan Koçyiğit, Dr. Erhan Eren, Dr. Alptekin Temizel for helpful comments and discussions. I also thank to Dr. Bülent Tavlı who has continuously been interested in my study for his support, guidance, comments, criticisms and encouragement.

I want to also thank to my colleagues Yasin Uzun, Yusuf Uzunay and İbrahim Arpacı for providing assistance and motivation during my study. I also thank to Ms. Sibel Gulnar, for her efforts regarding to administrative procedures and meeting arrangements.

I want to acknowledge TÜBİTAK (The Scientific and Technological Research Council of Turkey) and Scientific HR Development Program (ÖYP) for providing financial assistance during my PhD study.

I will also never forget the unending support my family have provided me with during all the hard times.

Lastly, I would like to express my very special gratitude to my wife who beared many difficulties and sacrifices in order to help me to finish this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS/ABBREVIATIONS

AWGN: Additive White Gaussian Noise

BS: Base Station

EAO: Eavesdropping Only

GAMS: General Algebraic Modeling System

LP: Linear Programming

MIP: Mixed Integer Programming

NCO: Node Capture Only

NCE: Node Capture and Eavesdropping

ORA: Optimal Role Assignment

QoS: Quality of Service

RDS: Redundant Data Sensing

WSN: Wireless Sensor Network

CHAPTER 1

INTRODUCTION

Wireless Sensor Networks (WSNs) are composed of small form factor devices (sensor nodes) that are instrumented with different types of sensors to carry out certain sensing tasks within a designated operation area. With limited processing, memory and energy resources, sensor nodes are capable of sensing physical phenomena, processing the sensed data, and transmitting the processed information bits via wireless communication. WSNs mostly consist of hundreds to thousands of sensor nodes deployed in a designated area randomly. Sensor nodes can be scattered from an airplane to an area of concern and after that they can form a self-organized network using wireless communication. In this ad hoc network, nodes coordinate to perform distributed sensing of physical phenomena, gathering information from the area of concern and forwarding it to a base station. Characteristics and constraints of WSNs are outlined in Table 1.1 [1].

Table 1.1: Characteristics and constraints of WSNs

Characteristics	**Constraints**
• Sensor nodes are small in terms of size. • The number of sensor nodes in a WSN can be hundreds to thousands. • The main power source of a sensor node is a battery. • Sensor nodes are deployed densely. • Sensor nodes mainly use wireless communication. • Self organization is one of the key concepts in WSNs. • Sensor nodes usually operate unattended.	• Sensor nodes are subject to failures. • The topology of a WSN changes very frequently. • Sensor nodes are limited in energy, processing, and memory. • Security of a WSN is a critical issue. • Redundant data generated by a WSN needs to be optimally eliminated. • Battery replenishment of a sensor node is not possible or highly challenging.

Unique characteristics as outlined in Table 1.1 of WSNs enable them to be used as a platform for several important surveillance and control applications. There are a wide variety of applications that can be realized using WSNs, varying from civilian applications such as precision agriculture to military applications such as military target tracking, battlefield surveillance, and event detection [2]. Precision agriculture monitors field variables such as temperature and humidity in a fine grained form and tries to respond better to these changes than the traditional methods. It enables better control and management of the field by taking the required action in the right time and at the right location. For example, knowing when to irrigate is important for both using scarce water resources effectively and increasing crop production. For example, cotton crops must be irrigated on time; delaying irrigation can causes losses between USD 62/ha and USD 300/ha [3]. A WSN can help to manage irrigation, avoid frost, control quantities of fertilizer, seed in the correct places and with the correct type and arrange harvest schedules [4,5]. In the military applications, WSNs can be used to monitor and gather information about enemy movements, explosions. WSNs can be deployed locations where wireline systems cannot be deployed (e.g. near an enemy, in places that contain toxins).

1.1 Roles of Sensor Nodes

While sensor nodes execute, they can be in three roles:

A Sensing Role: Sensor nodes are designed to sense different parameters, such as light, motion, pressure, electrical fields, sound, etc. If an event occurs, a sensor node in a sensing role captures this event and digitizes it. Afterwards, the sensor node sends the digitized data to the other sensor nodes in the network. In a sensing role, the sensor node spends its energy on sensing and transmission.

A Relay Role: After a sensor node senses an event, usually it is not possible or not feasible to transmit this information directly to the base station, so the sensor node transmits this information to another sensor node. This occurs repeatedly and finally the sensed information reaches its final destination. In a relay role, sensor nodes simply work as forwarders and they spend their energy receiving and transmitting information.

Base Station (BS): There must be at least one sensor node in this role. All the information flow through to this node and it spreads this information to outside world. This node is mostly

assumed to be energy unconstrained. Therefore, its energy dissipation is mostly ignored.

1.2 Routing in Wireless Sensor Networks

In WSNs, a routing protocol discovers and maintains routes between sensor nodes and the base station. Data generated by the sensor nodes is sent towards the base station on a route determined by the routing protocol. The energy consumption of sensor nodes is dominated by communication energy and more energy is required as the distance between two communicating sensor nodes increases; energy efficient communication in a WSN maximizes network lifetime. Therefore, sensor nodes cooperate to send their data through the base station to utilize their energy efficiently. Since it is not always possible or energy-efficient to transmit data directly to the base station, data can be relayed by several sensor nodes until it reaches to the base station.

While energy efficiency must be considered, routing protocols must also be designed by considering the requirements of the WSNs such as single path or multi-path routing, and routing with redundant data elimination.

1.2.1 Single Path and Multi-Path Routing

Single path routing does not allow data splitting in any relay nodes, multi-path routing splits the data into parts without employing data redundancy and sends each part via a different path towards the base station [6]. Multi-path routing can be used to improve the lifetime of WSNs by enabling balanced energy dissipation throughout the network. If each node has a connection to the base station with a single path, energy imbalances occur throughout the network (i.e., relay nodes on heavily utilized paths consume more energy than the other nodes). On the other hand sending data in multiple alternate routes can be used to balance the energy dissipation throughout the network and prevents premature deaths of the nodes on certain paths (e.g., minimum-energy path, minimum hop path) [7]. By distributing the routing burden on multiple paths as opposed to transmitting on only a single path, multi-path routing leads to more balanced energy dissipation in the network but as the number of paths increases so is the complexity of the routing protocol. Hence, it is desirable to limit the number of paths at some point if further increase does not improve energy efficiency significantly.

1.2.2 Routing with Redundant Data Elimination

Communication and computation are the two main energy dissipation categories for a typical WSN and communication energy usually dominates the energy expenditure. Therefore, reducing the amount of data conveyed to the base station is critical for improving energy efficiency. Since sensor nodes are generally deployed densely, highly correlated measurements performed by a multitude of sensor nodes lead to data redundancy. By exploiting the redundancy, it is possible to decrease the amount of data relayed towards the base station. One such redundancy elimination approach is the aggregation of the data coming from a group of sensor nodes [8]. Another related technique is the avoidance of redundant data collection [9]. By assigning sensing tasks to nodes intelligently, no redundant data is generated, thus, energy dissipation for communication is minimized. These energy-efficient strategies not only prolong the network lifetime but can also help to satisfy survivability requirements in case of a security attack.

1.3 Wireless Sensor Network Security

Cooperative and infrastructureless characteristics of WSNs make them suitable for utilization in hostile areas for tactical communication and networking in military applications. Since, in these applications, sensor nodes communicate sensitive information, security is of the utmost importance. However, WSNs can encounter outsider or insider attacks [10].

1.3.1 Outsider Attacks

In an outsider attack, the attacker node is not a part of the WSN. Since communication between sensor nodes occurs over a wireless channel, the attacker can also eavesdrop ongoing communication [11]. By using eavesdropped information, attacker can can perform various types of attacks such as replay attacks (re-sending previously eavesdropped packets) to impair the authenticity of communication or jamming attacks (emitting wireless signals) to interfere with the normal operation of WSNs. Also the main aim of an outsider attack can be to disable sensor nodes by sending bogus packets to drain the battery of receiver nodes or it may capture and physically destroy nodes.

Physical attacks, which aim to render one or more sensor nodes non-operational by physically capturing and destroying them, are among the most serious security threats in WSNs [12, 13]. In case of dense deployment (a desirable property in the design of WSNs) multiple sensor nodes acquire redundant (highly correlated) data. As a result, even if some nodes are destroyed, the remaining nodes can successfully complete the sensing task. However, as the number of nodes in the network decreases, the remaining nodes are burdened with the extra load (energy dissipation).

1.3.2 Insider Attacks

Cooperation is required to route sensed data towards the base station by using other nodes as relays. In security related WSNs applications, sensor nodes communicate sensitive information. Although cryptography is mostly used as a first line of defense to protect sensitive information, the unattended nature of sensor nodes makes them vulnerable to node compromise attacks. After a node compromise, an adversary can perform an insider attack. If sensor nodes are compromised, vital cryptography information such as keys used for encrypting data can be obtained and privacy and integrity of data can also be compromised. If the compromised node is highly used by other nodes as a relay, one node compromise may affect many of the nodes in the WSN. In contrast to disabled nodes, compromised nodes actively perform attacks such as by installing some malicious code on compromised node to disrupt functions of the WSN [10].

1.4 Mathematical Programming

Modeling the optimal allocation of limited resources to achieve an objective is the study area of mathematical programming models. A solution of these mathematical programming models is defined as an optimal solution if it satisfies some conditions aiming to reach a given objective. Linear Programming (LP) and Mixed Integer Programming (MIP) are the subclass of mathematical programming models which use linear functions both for objective and constraints. An objective is given as a linear function with the aim of finding the maximum or minimum by using some variables which must satisfy a finite number of constraints expressed either as linear equalities or linear inequalities. Both LP and MIP are used to find the best

solution considering a given set of constraints which characterize the set of legitimate decisions. Alternative decisions are compared based on their objective function values and the decision with the best value (could be the smallest or the largest depending on the nature of the function) is selected as the optimal decision. There are a wide variety of applications where LP and MIP are used such as transportation, production scheduling, and network flow problems [14]. LP and MIP are powerful tools utilized in many studies for modeling and analyzing WSNs.

1.5 Contribution

Algorithms and protocols developed for WSNs must consider the characteristics and constraints of WSNs but since battery replenishment is not possible or highly challenging for sensor nodes, one of the major concerns in designing routing protocols and algorithms is to achieve energy efficiency and to extend the network lifetime. The lifetime of a sensor node is dependent on the energy in its battery; if the energy in its battery is depleted, the sensor node dies and is excluded from the WSN. The lifetime of a WSN is directly related to the number of active sensor nodes and loss of a sensor node can degrade the lifetime of the WSN significantly. Adding security capabilities to a WSN without considering energy efficiency can lead to inefficient energy dissipation characteristics in the whole network and network lifetime can decrease dramatically. While enhancing the security level of a WSN, energy requirements should not be overlooked. Hence, energy efficient solutions are required for routing and security problems in WSNs.

In this thesis, we firstly investigate the limits for collaborative data gathering from the area of observation to minimize energy consumption. We provide an LP model that is used to find lower bounds for energy consumption by optimally eliminating redundant data in the network and by satisfying the minimal spatial granularity requirements of measurements.

Secondly, we consider a WSN that is tasked with monitoring an area for a predetermined period of time. The area is divided into regions and sensor nodes are deployed uniformly throughout the network. Through this representative application, we investigate the energy cost of combating against physical attacks by using a novel LP framework.

Thirdly, we investigate the impact of multi-path routing on energy efficiency in WSNs. By

developing a novel problem formulation using MIP, we capture the essence of both routing and energy dissipation characteristics of multi-path routing. We analyze the impact of limiting the number of routes from energy efficiency perspective within a general framework.

Finally, we consider a WSN where nodes have sensitive data to be conveyed to the base station. Data is spread out on multiple paths for protection from both active and passive attacks. Data flows on these paths are optimized to minimize the overall energy consumption throughout the network. Furthermore, to avoid premature death of any node within the network, energy dissipation is evenly balanced throughout the network, and hence, the network lifetime is optimized. Within an LP framework, we model the energy dissipation characteristics of path diversity countermeasures against node capture, eavesdropping, and both node capture and eavesdropping attacks. Using the developed LP models, we evaluate the energy cost of these path diversity based countermeasures by benchmarking against the energy cost of unconstrained data flows.

1.6 Thesis Organization

The remainder of the thesis is organized into seven chapters. In Chapter 2, we provide the related work on previous studies that are most relevant to our studies.

In Chapter 3, we provide an overview of the modeling of WSNs and an LP example is solved step by step.

In Chapter 4, the impact of spatial granularity of measurements on the energy requirements of WSNs is investigated. A novel LP framework is developed which allows us to determine almost achievable performance benchmarks in idealized yet practical settings which are achievable when redundancy is totally eliminated.

In Chapter 5, energy cost of mitigating physical attacks is investigated through a novel LP framework. The energy dissipation characteristics of the network for different physical attack scenarios are explored.

In Chapter 6, a novel MIP framework is developed to analyze the impact of the number of routing paths on network wide energy balancing under optimal operating conditions.

In Chapter 7, energy dissipation and data relaying behaviors of three route diversity techniques to mitigate node capture only, eavesdropping only, and node capture and eavesdropping attacks are characterized through a novel LP framework. Effects of node density, network area, level of resilience, and network topology on energy cost are investigated.

In Chapter 8, the contributions of this thesis and possible future works are discussed.

CHAPTER 2

RELATED WORK

Advances in processor, memory and radio technology lead to development of small devices called wireless sensor nodes [1] that are designed to measure and process physical data such as temperature, humidity, light and sound. Sensor nodes have so many restricted resources; therefore the true potential of a sensor is only realized through the cooperation of large number of them in a network environment and easiness of deployment of this network. Unique characteristics of WSNs as outlined in Table 1.1 open broad range of new research areas. There are a wide variety of studies on WSNs. In this chapter, we provide related work on multi-path routing, redundant data elimination, security and mathematical programming in WSNs.

2.1 Multi-path Routing in Wireless Sensor Networks

Energy is one of the primary concern in WSNs. Hence, some of the early works consider how routing operation can be performed energy efficiently [15, 16].

The literature on multi-path routing is extensive and has grown rapidly in recent years [7, 17, 18]. Multi-path routing provides several attractive properties such as security, load/energy balancing, reliability, and quality of service support [18]. Important problems in multi-path routing research include discovering multiple paths, selecting a number of paths among them, and distributing load across these selected paths [7]. We present an overview of the literature on multi-path routing by summarizing the studies most related to our work.

Alternate path routing is different than multi-path routing in the sense that a single path is used in normal operation but alternative paths are kept ready to be used in case the primary

path becomes unavailable [19].

Redundant multi-path routing is another related term that means the data to be conveyed to the base station is transported via multiple paths with added redundancy (*e.g.*, multiple replicas of the data is sent on different paths) [20]. From security point of view, adding redundancy is useful for enhancing resilience against denial of service attacks [21].

Multi-path routing can be used to improve the lifetime of WSNs by enabling balanced energy dissipation throughout the network. Consider the case where each node has a single path to the base station, which leads to energy imbalances throughout the networks (*i.e.*, relay nodes on heavily utilized paths dissipate more energy than the other nodes). Sending data in multiple paths can be used to balance energy dissipation and prevents premature deaths of nodes on certain paths (*e.g.*, minimum-energy path, minimum-hop path). There are several studies which address the problem of imbalanced energy dissipation by designing protocols that use multiple paths (*e.g.*, [22–30]).

In [31], multipath routing is utilized for QoS provisioning. Reliability defined as the packet delivery ratio and delay is used as constraints for QoS. First, they eliminate paths that cause longer delay than QoS requirement then they provide reliability through multipath routing. They model the problem as a probabilistic programming then it is relaxed into a deterministic LP.

In [32], multipath routing scheme with diversity coding is studied to minimize packet drop rate and end-to-end delay, and provide load balancing. By using the multipath routing scheme they aim to alleviate problems resulted from mobility of nodes and unreliable wireless links.

MMSPEED is proposed by [33] in which cross-layer design is used to provide QoS in terms of reliability and timeliness. Delay constraints of different applications is satisfied by constructing separate speed layers over the network. Data packets are forwarded by appropriate speed layer through the destination according to the delay requirement. Reliability is provided sending multiple copies of same packet over several active paths.

In [34], the problem of data distribution across multiple paths is studied with the aim of minimizing the maximum damage when a single link attack occurs in the network. The solution is formulated as a maximum-flow problem that can be solved in a distributed fashion.

In [35], a secure method for choosing multiple paths and distributing data among these paths is presented. The design objective is to minimize the percentage of captured data by an adversary. Each path is assigned with a security parameter that identifies the past performance on reliable data delivery. According to these parameters, multiple paths are constructed and data is distributed among these paths using min-max optimization and game theory.

In [36], an on-demand secure multi-path routing protocol is proposed to protect communication against collaborating malicious nodes. The protocol includes two phases. The first phase achieves neighbor node authentication by Elliptic Curve Cryptography. In the second phase, node-disjoint paths are found between source and destination nodes.

In [37], data is divided into parts and encrypted combinations of these are sent on different paths. Two of the paths are used for signaling and key sharing, thus, at least three paths between source and destination nodes are required. For dividing the message, a channel coding technique is used.

A protocol named H-SPREAD is presented in [38], which is built on SPREAD protocol [39]. H-SPREAD provides security and reliability to network communication by using redundant path routing together with secret sharing schemes. Secret sharing is adopted for splitting data over multiple paths so that even if a certain number of paths are eavesdropped, secrecy of data is not compromised.

Previous studies show that multi-path routing is proposed for energy balancing which prolongs the network lifetime as compared to single-path routing where utilization of a single route between a source node and the base station results in imbalanced energy dissipation. While it is evident that increasing the number of routing paths mitigates the problem of energy over-utilization in a subset of nodes acting as relays, the net effect of the proliferation of multiple routing paths on energy balancing remains unclear. It is imperative to keep the number of routing paths as low as possible without significantly deteriorating the network lifetime; therefore, determination of the optimal number of routing paths in multi-path routing by considering the tradeoff in routing complexity and network lifetime extension is an interesting research problem.

While it is tempting to state that the energy cost of route diversity or multi-path routing for security is high, we are not aware of any clear scientific evidence or convincing analysis to

support such a claim. It can also be argued that (with equal lack of convincing scientific evidence) the energy cost associated with route diversity is small and can easily be neglected. Given the fact that without a proper analysis it is not possible to quantify or even give a rough estimate on the energy cost of route diversity. Hence, investigation of energy cost of route diversity for security can provide valuable foundation for the development of future algorithms.

2.2 Redundancy Elimination

WSNs are generally deployed in high densities in a designated area to perform some sensing tasks. In such scenarios, there can be many sensor nodes generating same sensing information which causes redundancy in the network. Aggregation and sensor scheduling are two ways of eliminating redundancy in the network.

In [40], a theoretical framework is developed to investigate spatial and temporal correlations in WSNs. Design of efficient medium access and reliable event transport is discussed using developed framework.

In [41], policies are developed to find the multiple set of sensor nodes. Policies determine when and which sensor nodes should be powered on. Coverage and connectivity is considered by developing wake-up based topologies. The developed algorithm generates near-optimal (within 2.7% of the optimal) connected-covered wake-up based topologies.

In [42], ”periodic on-off” scheduling scheme is proposed based on social insect colonies. Ant colonies and WSNs are compared and operational states of the sensor nodes are determined by observing tasks of ants. Interactions among ants are used for local decision making by sensor nodes. The proposed ”periodic on-off” scheduling algorithm is compared with ”random on-off” and ”selective on-off” schemes.

In [43], a genetic algorithm with schedule transition as a hybrid is proposed to maximize lifetime of WSNs. Genetic algorithm determines the number of disjoint complete coversets by adopting forward encoding scheme for chromosomes in the population. A coverset can able to monitor whole area. Each coverset is activated by sensor schedule transition operations.

In [44], a power efficient scheduling method based on grid partition of sensing area is pro-

posed. Sensing area is divided into subsensing areas in which number of sensors are located. Grid size changes according to the application requirement. While operating, one sensor node is activated in each sensing areas.

In [45], elimination of redundant sensors is studied while preserving network coverage. For the detection and elimination of redundant sensors, voronoi diagrams are used. Sensor failures and new sensor deployment cases are also considered and algorithms are provided for those cases. Necessary and sufficient conditions are presented for redundant sensor nodes and sensor on the coverage boundary.

Data aggregation is to combine or compress similar or same data originated from different sources. To address the redundancy issue, several data aggregation schemes were proposed in the literature [8, 46–51].

The previous studies have not considered the effects of spatial granularity requirements of the interested area. They deal with the temporal granularity in a limited way using constant traffic generation rate for all sensor nodes. Providing an analysis for redundancy by considering spatial granularity requirements of WSNs is also different from the earlier studies on data aggregation.

2.3 Security in Wireless Sensor Networks

Security is a critical issue for WSNs because nodes usually operate unattended and communication takes place in a broadcast medium. A common and successful technique used against eavesdropping is cryptographic encryption. With encryption, sensor data is scrambled using a key to make eavesdropped data unintelligible to anyone who does not possess the key. Eavesdropping attacks are usually unnoticed (*i.e.*, it is challenging to detect a passive attack), thus, these attacks can succeed without encountering any active defense [52]. Public key encryption [53, 54] is mostly impractical in WSNs because of limited computation capability of sensor nodes. Hence, symmetric key encryption [55, 56] is generally applied in WSNs for the solution of eavesdropping [57–60].

Unattended and dense deployment of WSNs make sensor nodes susceptible to physical capture [61]. In WSNs sensor nodes can be captured (*i.e.*, node capture attacks) and vital cryp-

tography information such as keys can be extracted from them [62]. If keys are captured, this would render encryption useless [11]. Detection of captured nodes and dealing with these nodes is one direction of the solution for the problem of the node capture attacks [62–67]. Another direction is resilience to node capture in terms of even some portion of the WSN is affected from node capture attacks rest of the WSN can continue its operation securely. Key distribution schemes mostly focus on network resilience against node capture attacks [68–72]. Multipath routing algorithms provide solution for secrecy of data routed through the base station in the case of node capture attacks. Data is coded and partitioned at the source node and each part is sent through the base station on different paths. Even some portion of data is compromised by an adversary, it can not be understand [35–38].

The term *physical attack* to refer to the attacks aiming at physically destroying the sensor nodes. We note that in other studies (*e.g.*, [73]) the same term is also used for the concept of tampering with nodes (*i.e.*, capturing a sensor node and gaining direct access to its cryptographic material). By capturing a node, attackers can bypass cryptographic protection and conduct effective denial-of-service attacks [74]. In this thesis, we are not interested in these more advanced attack techniques and their countermeasures.

Physical attacks can be classified into two groups: blind physical attacks and search-based physical attacks [75].

In [12], blind physical attacks are studied. These attacks are performed after detecting the deployment area without considering where each sensor node is located. Sensor nodes may be destroyed blindly using a brute force approach (*e.g.*, by bombing the area). In such a situation, the research problem is to determine the minimum number of sensor nodes together with their location information to achieve the desired lifetime.

In [13], search-based physical attacks, in which sensor nodes are targeted and destroyed individually, are investigated. To defend against these attacks, a protocol is proposed based on the assumption that sensor nodes are able to detect attackers. In order to evaluate the performance of the proposed protocol, a metric called Accumulative Coverage is defined considering that primary success criteria of an attacker is the amount of coverage reduction in the network.

Lifetime optimization of WSNs is one of the most important functional design objectives because WSNs are envisioned to be operating in hostile and harsh environments where human

intervention is risky or costly. In such environments, battery replenishment is not possible or highly challenging. Adding security capabilities to a WSN without considering energy efficiency can lead to inefficient energy dissipation characteristics in the whole network and network lifetime can decrease dramatically. Previous studies focus on security but while enhancing WSN security, energy requirements should not be overlooked. Therefore, energy cost of WSN security enhancement strategies must be investigated.

2.4 Mathematical Programming

Mathematical programming is a powerful tool utilized in many studies to analyze different aspects of WSNs [76]. Here, we overview the previous work most relevant to our work.

In [26], Chang and Tassiulas show that when all packets are relayed from the same route, energy of the nodes on that path drains out more quickly than others. Therefore, they focus on the main objective of maximizing the network lifetime. They formulate the maximum lifetime routing problem as a linear programming problem.

In [29], Cheng et al. propose a general model for evaluating and maximizing lifetime of wireless sensor networks. They state that mostly many-to-one traffic pattern occurs in sensor networks. Therefore, nodes close to base station consume more energy compared to other nodes and this creates a hot-spot near base station. They focus on mitigating the problem of hot spot around the data sink using different deployment strategies.

In [77], a comparison of two multi-hop routing schemes (the first maximizes the minimum lifetime of the nodes and the second minimizes total energy dissipation) through an LP framework is presented.

In [27, 28], both topology insensitive lifetime bounds and topology sensitive lifetime bounds with specified topologies are derived. The formalism of feasible role assignments is a key concept utilized in the derivations. It is shown that a class of role assignment problems can be transformed into LP models.

In [78], the impact of limiting the numbers of incoming and outgoing links on network lifetime is investigated through a Mixed Binary LP framework. The results of this study show that if there are at least three incoming and outgoing links then the decrease in network lifetime is

negligible.

In [79], the problem of coverage optimization through scheduling for WSNs used to control and monitor industrial and manufacturing processes under energy dissipation constraints are considered. A Mixed Integer Programming (MIP) formulation is constructed and techniques to decompose the problem into separate subproblems to reduce the complexity are proposed.

In [80], an MIP framework is proposed to analyze capacity and energy consumption of IEEE 802.15.4 cluster-tree hierarchy for organizing transmissions to provide the optimal solution for the network capacity.

In [81], the problem of minimizing the network cost through the minimum number of relay-station installation in continuous data-gathering WSNs is investigated by using an MIP model.

In [82], an MIP based framework for optimizing the placement of RF chargers used for energy harvesting in WSNs is proposed.

In [83], two joint routing and scheduling algorithms which minimize the data delivery latency while enhancing the energy efficiency in WSNs is proposed and investigated through an MIP framework.

In [84], the maximal lifetime scheduling problem in sensor surveillance systems is investigated using an LP framework. The problem is defined as the maximization of the lifetime of a surveillance system given a set of sensors monitoring certain targets in an area. A globally optimal solution is computed using an LP approach.

In [85], the problem of WSN lifetime optimization is investigated through an LP model by considering the energy dissipations of both data communication and computation. A joint routing and compression optimization strategy is proposed which is shown to prolong the network lifetime when compared to no compression or full compression strategies. In [86], the effects of multi-level data compression in conjunction to flow balancing are also investigated.

In [87], the benefits of utilizing multi-domain collaborative WSN paradigm is investigated through an LP framework to quantitatively evaluate the performance of various cooperation strategies. It is shown that multi-domain cooperation in WSNs can increase the lifetime significantly and the lifetime improvement can be as high as an order of magnitude.

In [88], lifetime limits of WSNs improving contextual privacy by transmit range control are analyzed using LP.

In [89], optimal flow-jamming attacks are formulated through an LP framework.

In [90], the network restoration problem in multi-hop multi-channel wireless networks under jamming attacks is investigated. An LP based formulation is developed to model restoration schemes under jamming attack.

Choosing a mathematical programming (i.e. LP and MIP) based analysis method has a number of advantages. One of them is the abstraction from a specific protocol which enables us to investigate energy cost in ideal conditions with optimal routing decisions. Secondly, due to global knowledge in the optimization problem solver, the results can be obtained in an efficient and consistent manner. Like in many previous studies, in this thesis we use the mathematical programming based analysis method.

CHAPTER 3

BACKGROUND

The studies described in this thesis are based on modeling of routing and security problems in WSNs using mathematical programming more specifically Linear Programming (LP) and Mixed Integer Programming (MIP). Before going into more detail about our studies, in this chapter, we present an overview of the modeling of WSNs.

3.1 Models

In order to provide a running example of our LP and MIP frameworks, it is required to use channel propagation and energy dissipation models. This section describes models that were adopted in our frameworks.

3.1.1 Channel Propagation Model

In a wireless channel, the power of electromagnetic signal decreases as the distance between transmitter node and receiver node increases. In this thesis, Friis Free Space Propagation model is used [91, 92]. Friis Free Space Propagation model calculates the average decrease on received power over a distance d as

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{3.1}$$

$P_r(d)$**:** Received power when the distance between two communicating parties is d

d**:** Distance between two communicating parties

P_t**:** Transmit power

G_t**:** Transmitter antenna gains

G_r**:** Receiver antenna gains

λ**:** Wavelength

L**:** System loss factor

Equation 3.1 models the signal attenuation when the transmitter node and the receiver node have a direct line of sight. If the transmitter node and the receiver node don't have a direct line of sight, Two-Ray Ground Propagation model can be used [92, 93]. In this model, the average decrease on the received power over a distance d is calculated as

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_t^2}{d^4} \tag{3.2}$$

$P_r(d)$**:** Received power when the distance between two communicating parties is d

d**:** Distance between two communicating parties

P_t**:** Transmit power

G_t**:** Transmitter antenna gains

G_r**:** Receiver antenna gains

h_t**:** Transmitter antenna height above ground

h_r**:** Receiver antenna height above ground

In the Two-Ray Ground Propagation model, the average decrease on the received power over a distance d is proportional to d^4 which is greater than Friis Free Space Propagation model in which the average decrease on the received power over a distance d is proportional to d^2

3.1.2 Radio Energy Model

Radio energy model is used to determine energy consumption characteristics of nodes while transmitting and receiving data. In our system model, energy consumption of sensor nodes is

dominated by communication energy rather than sensing and processing energy dissipations[1]. In this thesis, we adopt a widely accepted energy model used in many studies[2] [26–29, 95]. In this model, the amount of energy to transmit one bit of data is

$$P_{tx,ij} = \rho + \varepsilon d_{ij}^{\alpha} \tag{3.3}$$

and to receive one bit of data is

$$P_{rx} = \rho, \tag{3.4}$$

where ρ represents the energy dissipated in the electronic circuitry, ε denotes the transmitter efficiency, α represents the path loss exponent, and d_{ij} is the distance between node-i (transmitting node) and node-j (receiving node). In this thesis, for the numerical analysis, we use the standard value of receiver constant (ρ is 50 nJ/bit), the standard value of transmitter constant (ε is 100 pJ/bit/m2) and path loss exponent (α) is chosen as 2 as the ones in [29, 95]. Table 3.1 lists the parameters used in our numerical analysis.

Table 3.1: Description and value of parameters used in the energy model

Parameter	**Description**	**Value**
ρ	Energy dissipated in the electronic circuitry	50 nJ/bit
ε	Transmitters efficiency	100 pJ/bit/m^2
α	Path loss exponent	2

3.2 Assumptions

In our frameworks, we make the following assumptions:

1. The network consists of stationary nodes (both sensor nodes and the base station).
2. The network topology is represented by a directed graph, $G = (V, A)$, where V is the set of all nodes including the base station and W is the set of all nodes except the base station.

[1] For example, communication energy dissipation constitutes 91% of the total energy dissipation in Telos sensor nodes [94].

[2] We emphasize that LP and MIP formulations we present in this thesis can easily be tailored to any other energy model.

3. Data generated at each node, transferred either directly (single-hop) or via relays (multi-hop), terminates at the base station.

4. Network reorganization period is long enough, therefore, the energy costs of topology discovery and route creation operations constitute a small fraction (e.g., less than 1.0 % [96]) of the total network energy dissipation, hence, routing overhead can be neglected in stationary WSNs without leading to significant underestimation of total energy dissipation.

5. A TDMA-based MAC layer is in operation which mitigates interference between active links through a time-slot assignment algorithm which outputs a conflict-free transmission schedule[3].

6. Energy dissipation for idle listening or overhearing in promiscuous mode is negligible[4].

7. The channel is an additive white Gaussian noise (AWGN) channel, with noise power spectral density N_0 over the bandwidth of operation ξ [103].

8. Energy dissipation for transmission and reception is only for data packets[5].

3.3 Background on Mathematical Programming

In this section we provide a brief background on mathematical programming to motivate the use of MIP and LP in our frameworks.

Mathematical programming is a subclass of mathematics that aims to find maximum or minimum of an objective function subjects to linear, non-linear, and integer constraints on the variables [106]. Generally mathematical programming problems include following elements [107]:

[3] A combinatorial interference model can be used to model interference and scheduling constraints can then be modeled by a conflict graph [97, 98]. In [99], it is shown that such an algorithm is possible hence collision free communication is achieved if sufficient bandwidth requirements are satisfied. In fact, in our model, we use a modified version of the sufficient condition presented in [99]. Furthermore, it is also possible to reduce data packet collisions to negligible levels in practical MAC protocols designed with a dynamic TDMA approach [100, 101].

[4] There are many intelligently designed MAC protocols for wireless networks that avoid energy waste in these modes [100, 102]. We assume such a MAC layer is used in our framework.

[5] The framework described in this thesis can be easily extended to model the energy consumption for acknowledgement (ACK) packets as well, however, in general, the size of a typical ACK packet is much smaller than the size of data packets therefore we assume that the receiver is informed of correct or incorrect reception of a data packet through an ideal feedback channel [103–105].

Variables: Variables are used to make decisions and the values of the variables are determined after solving mathematical programming problem. Hence their values aren't known at the beginning of the problem. For example, the values of data flows between sensor nodes are calibrated to find out maximum lifetime of WSN or minimum required energy for the batteries of sensor nodes.

Objective function: Objective function consists of set of variables. Mathematical programming finds out the values of variables with the aim of maximizing or minimizing the objective function. For example, maximizing lifetime of the WSN is an objective function.

Constraints: Constraints determine the solution space for the objective function. In other words, values can only be assigned to variables if the resulting solution satisfies constraints. For example, transmission range of sensor nodes is a constraint and there can't be flow between nodes if the distance between two sensor nodes is larger than transmission range of sensor nodes.

Variable bounds: Variable bounds can also be considered as constraints. Even some values of the variables satisfy constraints, assigning these values to the variables can be infeasible in real life. Hence, the variables are bounded to get realistic results. For example, data flows between nodes can't take negative values.

A mathematical programming problem can be written as in the generic form as

$$\begin{aligned} \text{minimize/maximize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq b_i, \quad i = 1, ..., m \\ & f_j(x) = b_j, \quad j = 1, ..., k. \end{aligned} \tag{3.5}$$

Here variables are defined in the vector $x = x_1, ..., x_n$, objective function is $f_0 : R^n \rightarrow R$, inequality constraints and equality constraints are defined at the functions $f_i : R^n \rightarrow R$, and $f_j : R^n \rightarrow R$, respectively. b_i and b_j are the bounds for the inequality constraints and equality constraints, respectively. A vector x^* is chosen as optimal, if x^* satisfies all inequality constraints and equality constraints by minimizing or maximizing the objective function ($f_0(x^*)$).

As examples of mathematical programming models, both LP and MIP are used to find the best solution considering a given set of constraints, which characterize the set of legitimate

decisions [108]. Alternative decisions are compared based on their objective function values and the one with the best value (could be the smallest or the largest depending on the nature of the function) is selected as the optimal. Although they are used for the same reason, LP and MIP models cannot be used in place of each other in many occasions. Hence, they should not be considered as alternatives. If "Yes/No" type decisions are to be made, we need to use decision variables which take binary and thus integer values. For example in one of our model introduced shortly, a_{ij}^{kl} variable indicates if an arc (i, j) is used in path-l to transmit the data sensed by sensor node-k. It is 1 if the answer is positive and 0 otherwise. As a result, we can say that the type of mathematical model to be used depends on the type of decisions to be made, which leads to an MIP model for our problem.

LP models whose variables take continuous values are relatively easier to solve. This is due to the special geometry of the set of feasible solutions (called the feasible set) of LPs. The vertices of the feasible set are defined by the constraints of the model and it is known that, given a nonempty feasible set, there is always a vertex solution which is optimal. Hence the well-known Simplex Algorithm, which searches the optimal solution among the vertices greedily, is a quite effective solution method for LPs, on the average. Unfortunately, MIP models do not have such a property in general and hence call for more advanced solution algorithms such as branch-and-bound, branch-and-cut, etc. These methods guaranteeing an optimal solution are called exact solution methods. At each step of such algorithms, first the problem without the integrality restrictions on variables (*i.e.*, the LP relaxation) is solved. Then, if an integer variable (*e.g.*, the one which is actually required to be 0 or 1 in the original problem) has a fractional value in the current solution, then the problem is divided into two subproblems by setting that variable's value to 0 and 1, respectively. Then the new problems are solved recursively in the same manner until the optimal solution is found. This basic method can be improved and fastened by incorporating problem specific information in the subproblem creation step.

We use GAMS (General Algebraic Modeling System) [109] for the numerical analysis of the LP and MIP models. GAMS consists of high-performance solvers for solving LP and MIP models efficiently each of which improves upon the basic approach in different ways to attain an increased solution performance. Hence, when we solve our LP and MIP models using GAMS, one of these solvers is used to obtain the best solution. Specific implementation details are beyond the scope of this thesis. Before solving our LP and MIP models using

GAMS, each model is manually solved step by step like in following section (section 3.3.1). By this way, we ensure that each research problem is correctly modeled and each LP and MIP models are ready to extensive analysis using GAMS.

As we have mentioned, general MIP models are computationally difficult problems. They are in NP-hard class according to their computational complexity [108]. Although there are some MIP problems with efficient optimization property (*i.e.*, they can be solved relatively easier due to their special structures), we are not aware of any previous result on the applicability of such a property in our MIP problem. One solution for reducing complexity of the MIP problems is the LP relaxation of the problem which is obtained by relaxing the integrality constraints on binary variables but this solution does not provide the same optimal solution with the original MIP model and the gap is actually significant in many cases. Moreover, the solution times for the MIP problems increase significantly as the instance sizes get larger. Hence, even for medium sized instances, the solution times can be quite high, which can be mitigated using several implementation heuristics. Since our motivation in this thesis is to explore the impact of routing solutions rather than developing specialized efficient solution algorithms for the problem, we accept solutions with relative gap no more than 1.0 %.

Suppose that we have a solution satisfying all integer requirements and has the best objective function value (z^B) found so far. Then the relative gap for this solution measures the distance between z^B and the available best bound for the optimal objective function value (z^L) using the ratio $\frac{|z^B - z^L|}{z^L}$. LP based branch-and-bound algorithms are used for solving MIPs in GAMS [109], thus, z^L is the LP relaxation solution of the MIP problem under consideration. Note that, in general, z^L is not a feasible solution because integer variables are treated as continuous variables (*i.e.*, the occurrence of non-integer values for binary variables is allowed in z^L). The acceptable relative gap can be controlled via the parameter *optcr* in GAMS and when it is set to 0.0, the solution algorithm stops with the exact optimal solution. In this thesis, we let *optcr*=1.0 %, which provides significant time savings in exchange for an immaterial sacrifice for optimality.

3.3.1 Linear Programming Example

In this part, an example LP problem is solved step by step to help better understanding of LP and MIP frameworks in the following chapters.

In this LP problem, we assume that there are N sensor nodes and a single base station in the network. Data generated at each node, transferred either directly (single-hop) or through other sensors acting as relays (multi-hop), terminates at the base station. The network topology is represented as a directed graph $G = (V, A)$. V is the set of all nodes, including the base station as node-0 (n_0). We also define set W, which includes all the nodes except node-0. $A = \{(i, j) : i \in W, j \in V - i\}$ is the set of arcs (links). The amount of data (bits) sent on the directed link (i, j) is denoted as f_{ij}. The LP problem is formulated as maximizing WSN lifetime L subject to the following constraints:

Maximize L
Subject to:

$$f_{ij} = 0 \text{ if } i = j \; \forall (i, j) \in A \tag{3.6}$$

$$\sum_{j \in V} f_{ij} - \sum_{j \in W} f_{ji} - s_i L = 0 \; \forall i \in W \tag{3.7}$$

$$P_{rx} \sum_{j \in W} f_{ji} + \sum_{j \in V} P_{tx,ij} f_{ij} - e_i \leq 0 \; \forall i \in W \tag{3.8}$$

$$f_{ij} = 0 \text{ if } d_{ij} \geq R_{max} \; \forall (i, j) \in A \tag{3.9}$$

$$f_{ij} \geq 0 \; \forall (i, j) \in A \tag{3.10}$$

Figure 3.1: LP model for routing.

We will work on an example topology as shown in Figure 3.2 to illustrate how routing operation is realized to maximize lifetime of the WSN. In this example topology there are three nodes (n_3, n_2, n_1) and a single base station (n_0). As shown in Figure 3.2, all data generated by each sensor node routed through the base station. In the optimization problem, it is required to provide all the options that show how a sensor node can send data to the other sensor nodes and the base station. Figure 3.2 shows all the possible options. We explain constraints in Figure 3.1 in detail using the sample network topology in Figure 3.2.

Constraint 3.6 is used to set flows that do not exist to zero. In other words, there can't be a flow from a node to itself. Therefore,

$$\begin{aligned} f_{11} &= 0, \\ f_{22} &= 0, \\ f_{33} &= 0. \end{aligned} \tag{3.11}$$

Constraint 3.7 is known as flow conservation constraint and this constraint models the sum of

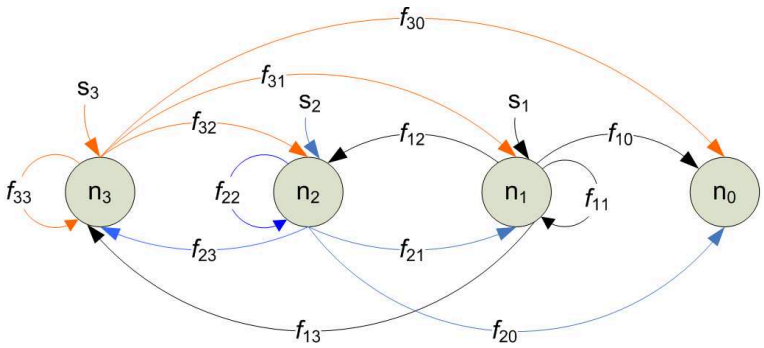

Figure 3.2: Sample WSN topology that consists of three nodes and one base station with all possible transmission options.

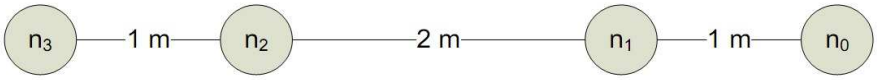

Figure 3.3: Sample WSN topology that consists of three nodes and one base station with distances between nodes.

all data flowing into node-i ($\sum_{j \in W} f_{ji}$) and the total data generated at node-i during network lifetime ($s_i L$) is equal to the sum of all data flowing out of node-i ($\sum_{j \in V} f_{ij}$).

When constraint 3.7 is applied for all sensor nodes of wireless sensor network in Figure 3.2 by assuming that sensor nodes can send data without the restriction on transmission range, the result of constraint 3.7 will be as follows:

$$\begin{aligned} &\textbf{for node-1:} \quad f_{31} + f_{21} + s_1 L = f_{10} + f_{12} + f_{13}, \\ &\textbf{for node-2:} \quad f_{32} + f_{12} + s_2 L = f_{20} + f_{21} + f_{23}, \\ &\textbf{for node-3:} \quad f_{23} + f_{13} + s_3 L = f_{30} + f_{31} + f_{32}. \end{aligned} \tag{3.12}$$

Equations in 3.12 provide all possible transmission flow options to linear programming to be able to find maximum achievable network lifetime L.

Constraint 3.7 closely related with the constraint 3.9. This constraint is about transmission power control capabilities of sensor nodes. If the nodes are capable of sending data without the restriction on transmission range ($R_{max} \rightarrow \infty$) , then transmission flow options will be as equations in 3.12. But if the nodes use a fixed transmission range (R_{max}), then sensor nodes can only send data to other nodes in this transmission range.

Let's assume that distances between sensor nodes in Figure 3.2 are shown in Figure 3.3 and

nodes can adjust their transmission power according to maximum transmission range;

$$R_{max} = 2m.$$

Then the transmission flow options for each node will be as follows:

$$\begin{aligned} &\textbf{for node-1:} \quad f_{21} + s_1 L = f_{10} + f_{12}, \\ &\textbf{for node-2:} \quad f_{32} + f_{12} + s_2 L = f_{21} + f_{23}, \\ &\textbf{for node-3:} \quad f_{23} + s_3 L = f_{32}. \end{aligned} \tag{3.13}$$

As can be concluded from equations in 3.13, using constraint 3.9 with $R_{max} = 2m$; for node-1, since distance between node-1 and node-3 is equal to 3 meters that is bigger than the R_{max}, $f_{31} = 0$ and $f_{13} = 0$. For node-2, since distance between node-2 and the base station is equal to 3 meters, $f_{20} = 0$. For node-3, since distance between node-3 and node-1 is equal to 3 meters and node-3 and the base station is equal to 4 meters, $f_{13} = 0$,$f_{31} = 0$ and $f_{30} = 0$.

Constraint 3.8 is used to determine energy consumption of a node while transmitting and receiving data and it states that energy consumption of the nodes must be less than their battery power. We use energy model described in section 3.1.2. The amount of energy to transmit one bit of data is $P_{tx,ij} = \rho + \varepsilon d_{ij}^{\alpha}$ and to receive one bit of data is $P_{rx} = \rho$, where ρ represents the energy dissipated in the electronic circuitry, ε denotes the transmitters efficiency, α represents the path loss exponent and d_{ij} is the distance between node-i and node-j. We use the standard value of receiver constant (ρ is 50 nJ/bit) and the standard value of transmitter constant (ε is 100 pJ/bit/m^2) given in [29]. Path loss exponent (α) is chosen as 2. Distances between nodes are gathered from Figure 3.3. Sensor nodes do not have a restriction on transmission range ($R_{max} \rightarrow \infty$). After using these parameters, generated equations by using constraint 3.8 will be as follows:

$$\begin{aligned} &\textbf{for node-1:} \quad 50000 f_{31} + 50000 f_{21} + (50000 + 100 * 1^2) f_{10} \\ &\qquad +(50000 + 100 * 2^2) f_{12} + (50000 + 100 * 3^2) f_{13} \leq e_1, \\ &\textbf{for node-2:} \quad 50000 f_{32} + 50000 f_{12} + (50000 + 100 * 3^2) f_{20} \\ &\qquad +(50000 + 100 * 2^2) f_{21} + (50000 + 100 * 1^2) f_{23} \leq e_2, \\ &\textbf{for node-3:} \quad 50000 f_{23} + 50000 f_{13} + (50000 + 100 * 4^2) f_{30} \\ &\qquad +(50000 + 100 * 3^2) f_{31} + (50000 + 100 * 1^2) f_{32} \leq e_3. \end{aligned} \tag{3.14}$$

Table 3.2: The basic components of a GAMS model.

Inputs:	**Outputs:**
• *Sets*	• *Echo Print*
• *Data* (*Parameters*, *Tables*, *Scalars*)	• *Reference Maps*
• *Variables*	• *Equation Listings*
• *Equations*	• *Status Reports*
• *Model and Solve statements*	• *Results*

When equation for node-1 in 3.14 is examined: f_{31} means that node-1 can able to receive some amount of data from node-3. The required energy for this operation is calculated using $P_{rx} = \rho$. Therefore, if node-3 chooses to forward its data to base station using node-1 as a relay, for this operation node-1 dissipates $50000 f_{31}$ pJ energy. Same issue is valid for the flow f_{21}, if node-2 chooses to forward its data to base station using node-1 as a relay, for this operation node-1 dissipates $50000 f_{21}$ pJ energy. f_{12} states that node-1 can able to transmit some amount of data to node-2. Required energy for this operation is calculated using equation $P_{tx,ij} = \rho + \varepsilon d_{ij}^{\alpha}$. If node-1 chooses to forward its data to base station using node-2 as a relay, for this operation node-1 dissipates $(50000 + 100 * 2^2) f_{12}$ pJ energy. Same energy dissipation calculation is used for the f_{13} and f_{10} flows.

Lets consider that nodes can adjust their transmission power according to maximum transmission range ($R_{max} = 2m$). Then the energy consumption characteristics of nodes will be as follows:

$$\begin{aligned}
&\textbf{for node-1:} \quad 50000 f_{21} + (50000 + 100 * 1^2) f_{10} + (50000 + 100 * 2^2) f_{12} \leq e_1, \\
&\textbf{for node-2:} \quad 50000 f_{32} + 50000 f_{12} + (50000 + 100 * 2^2) f_{21} + (50000 + 100 * 1^2) f_{23} \leq e_2, \\
&\textbf{for node-3:} \quad 50000 f_{23} + (50000 + 100 * 1^2) f_{32} \leq e_3.
\end{aligned} \tag{3.15}$$

3.3.1.1 Numerical Analysis Environment

In this section, modeling and solution of example LP problem using GAMS is provided step by step. The basic components of a GAMS model are provided in Table 3.2 [110].

Sets define terms that are going to be used for as an index. Below, we define the set with 4 elements for example topology in Figure 3.2 ($i = 'n0', 'n1', 'n2', 'n3'$).

- ***Set*** *i nodes /n0*n3/;*

Alias is used to assign more than one name for the same set ($i = 'n0','n1','n2','n3'$, $j = 'n0','n1','n2','n3'$).

- ***Alias*** *(i,j);*

The *Scalar* statement is used to declare and (optionally) initialize a GAMS parameter.

- ***Scalar***

  `EAmp` *picoJoule /100/*

  `EElec` *picoJoule /50000/*

  `Prx` *reception energy*

  `Rmax` *maximum transmission range ;*

Parameters are used to store data before the model starts to run. Values of parameters don't change after the model started to run.

- ***Parameters***

  `y(i)` *y coordinate of node-i*

  `x(i)` *x coordinate of node-i*

  `s(i)` *data generated at node-i*

  `d(i,j)` *distance between node-i and node-j*

  `Ptx(i,j)` *consumed energy for transmission of data from node-i to node-j*

  `e(i)` *battery energy of each node ;*

Variables are used to make decisions and the values of the variables are determined after solving mathematical programming problem. Mathematical programming tries to conclude with an optimal solution by assigning appropriate values to the variables. In GAMS, objective variable `L` must not be bounded.

- ***Variables***

  `L` *lifetime*

  Positive Variables

  `f(i,j)` *flows;*

Constraints are defined in *Equations*. First, index of equations are given. Secondly, equation implementations are provided.

- ***Equations***

  `noFlow(i,j)` *no flow*

  `flowBalance(i)` *flow balance*

  `energyConstraint(i)` *energy constraint*

  `transmissonRange(i,j)` *Rmax;*

- `noFlow(i,j)$(ord(i)=ord(j) or ord(i)=1).. f(i,j) =e= 0;`
- `flowBalance(i)$(ord(i)>1).. sum(j,f(j,i)) + s(i)*L =e= sum(j,f(i,j));`
- `energyConstraint(i)$(ord(i)>1)..`
  `e(i) =g= Prx*(sum(j$(ord(j)>1),f(j,i)))+sum(j,(Ptx(i,j)*f(i,j)));`
- `transmissonRange(i,j)$(d(i,j)>Rmax).. f(i,j) =e= 0;`

Model determines which constraints are going to be used in the optimization.

- ***Model*** MaximumLifetime /

  `noFlow`

  `flowBalance`

  `energyConstraint`

  `transmissonRange /;`

File names and locations are defined in *file*.

- `file` *Result /c:\ Result \ LinearTopology-Result.txt/;*

- `file` *Coordinate /c:\Result\LinearTopology-Coordinate.txt/;*

The next step for coding in GAMS is calculation and assignment of parameters' and scalars' values.

- Values of coordinates are assigned as;

  x('n0')=0;

  y('n0')=0;

  x('n1')=1;

  y('n1')=0;

  x('n2')=3;

  y('n2')=0;

  x('n3')=4;

  y('n3')=0;

- Distances between nodes are calculated and assigned to parameter *d(i,j)* as;

  d(i,j) = sqrt(sqr(x(i)-x(j))+sqr(y(i)-y(j)));

- Energy consumption for transmitting one bit of data from node-i to node-j is calculated and assigned to parameter *Ptx(i,j)* as;

  Ptx(i,j)$(ord(i)<> ord(j) and ord(i)<>1) = EElec + EAmp*sqr(d(i,j));

- Battery energy for each node is assigned as;

  e(i)=1e12;

- Energy consumption for receiving one bit of data is assigned to parameter *Prx* as;

  Prx = EElec;

- R_{max} is assigned with a big number ($R_{max} \rightarrow \infty$) as;

  Rmax=100000;

- Data rate of each node is assigned to parameter s_i (1 bit/s) as;

  s(i)$(ord(i)>1)=1;

After all the values of parameters and scalars are assigned, we can run the LP and find the solution as;

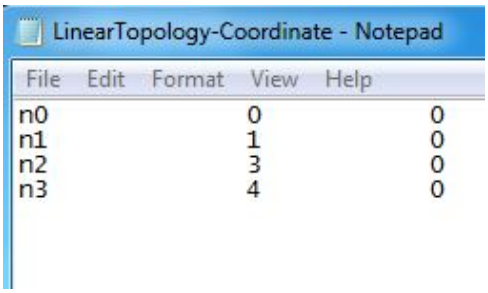

Figure 3.4: Screen shot of generated file for coordinates.

- `Solve MaximumLifetime using lp maximizing L;`

"MaximumLifetime" is the name of the model which is defined at *"Model"* part and *"using lp"* means that use LP for the solution of *"MaximumLifetime"* model and our objective is maximizing lifetime which is defined in the variable *"L"*.

- We can write coordinates, values of flows f_{ij} and lifetime L to files as;

```
put Coordinate;
loop(i,
    put i.tl:4:0 x(i):12:0 y(i):12:0/;
);
put Result;
put 'Lifetime = ' L.l:12:8 /;
loop(i,
    loop(j,
        put$(f.l(i,j)>0) i.tl:4:0 j.tl:4:0 f.l(i,j):12:3 /;
    );
);
```

The generated files are shown in Figure 3.4 and Figure 3.5.

Figure 3.4 shows coordinates of the nodes and and Figure 3.5 shows lifetime of the network and amount of data forwarded by each node through the base station. As results show the network can live $1.9388 * 10^7$ seconds. Since we use s_i as 1 bit/sec, this result also shows that how many total bits each sensor node can gather from the environment. Since LP accomplish the goal of lifetime maximization by load balancing through a combination of intelligent

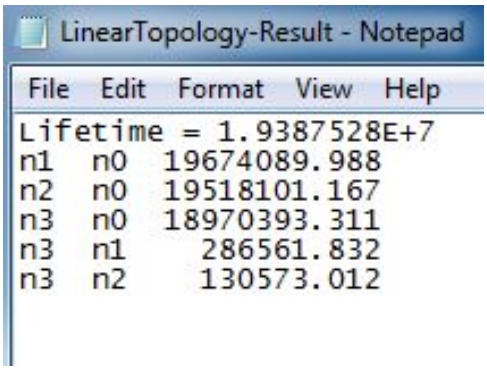

Figure 3.5: Screen shot of generated file for results.

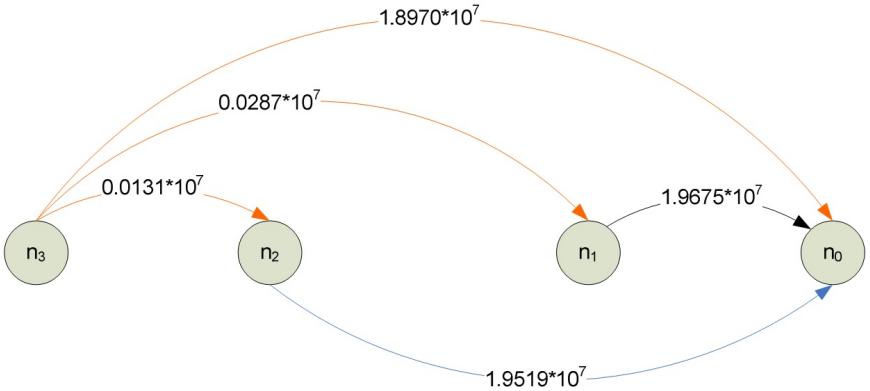

Figure 3.6: Data flows.

routing, Figure 3.5 also shows that how each node sends data through the base station. As illustrated in Figure 3.6, node-1 and node-2 send all data directly to the base station but node-3 doesn't send all its gathered data from the environment directly to base station. It uses node-1 and node-2 as relay for its some traffic.

GAMS implementation and compilation details of this LP Example including GAMS code are provided in the Appendix A.

CHAPTER 4

SPATIAL GRANULARITY AND ENERGY REQUIREMENTS IN WIRELESS SENSOR NETWORKS

The amount of measurement data collected in a WSN is affected from three characteristics of the network design and the application domain:

1) Redundancy

2) Spatial granularity

3) Temporal granularity

Informally speaking, redundancy means that the same (or correlated) physical data is measured by several sensors. This duplication may be desired to mitigate the reliability problems caused by low-cost error-prone sensor nodes. However to decrease the communication cost, such redundancy needs to be eliminated by in-network processing (data aggregation) techniques before transmission to the base station takes place. Spatial granularity is based on the in-field geographic deployment of the sensor nodes. Physical properties of the field are mostly different in distinct locations where each location represents a spatial granule. Therefore, measurements made by sensor nodes at each spatial granule also show differences. In WSNs, different physical phenomena generally require sensing at different spatial granularities. A spatial granule can be as big as the whole area if the whole area exhibits exactly the same physical property or as small as sensing range of each sensor node. Temporal granularity is based on the measurements of physical phenomena in different times with the same set of sensor nodes. The time interval between two measurement events represents a temporal granule and making measurements at the same spatial granule using different temporal granules might yield different results. Length of temporal granule is determined according to the characteristics of application.

In this part, we investigate the relationship between battery requirement of sensor nodes and the measurement characteristics introduced above. We focus on spatial granularity in our investigation and secondarily address the issue of redundancy in this context. As a motivating example for our research we choose precision agriculture.

Since not all parts of the field shows the same properties, precision agriculture techniques should also enable every part of the field to be behaved differently. To achieve this, it needs intensive information from the field. By equipping nodes with different sensors which are capable of sensing temperature, humidity, sun light intensity, etc., WSN can help to manage irrigation, avoid frost, control quantity of fertilizer, seed correct places and correct type and arrange harvest schedule [4, 111]. In precision agriculture, monitored field includes several spatial granules with different sizes and properties. After the deployment, WSN is used to understand field properties e.g., which spatial granule of the field is productive, wet, dry or not valuable to seed. After detecting these, a more appropriate seeding can be realized compared to classical agriculture. Then each spatial granule can be continuously monitored against humidity, temperature, fertilization and even disease. According to measurements, necessary actions can be applied to each spatial granule in different ways. For instance, some of the spatial granule in the field may need more irrigation than others. Only distributed WSNs can provide the spatial granularity of measurements needed. Simply putting a few powerful sensors do not provide this functionality.

In this study, we are motivated by the example application (i.e., precision agriculture) introduced in previous section and we find optimum energy levels that can be achieved by assuming that sensor nodes on the same spatial granule coordinate in an energy efficient manner and a single measurement result is returned to the base station for that particular spatial granule.

Since sensor nodes have limited energy resources, one of the major concerns in designing network protocols is to achieve energy efficiency and to extend the network lifetime. In our study, we investigate the limits for collaborative data gathering from the area of observation to minimize the energy consumption. We provide a linear programming (LP) model that is used to find out lower bounds for energy by optimally eliminating redundant data in the network and by satisfying the minimal spatial granularity requirements of measurements.

4.1 Problem Definition

Self organization nature of sensor networks enable them to operate in harsh environments but also force them to work with limited battery resource since replenishment of battery is not always feasible. After the deployment, sensor network gathers information from an area of observation by making measurements and relay it to an energy unconstrained base station possibly using multi-hop communication. We assume that minimum amount of measurement data that should be collected from each spatial granule is pre-determined according to application needs. We also assume that total number of spatial granules and their locations are also known. The sensor network must possess some capabilities to complete its task of collecting this information. Energy is the most critical resource that may prevent the completion of this task. Informally speaking, the problem now is the specification of which node collects how much of measurement data and how this data is relayed to the base station to accomplish the task with minimum energy required for each sensor node.

The optimization problem is formulated as an LP problem, below. Table 4.1 lists the parameters we use in the formulation. The objective function is the minimization of "e_i - battery energy of a sensor node " subject to the following constraints:

Minimize *battery*
Subject to:

$$f_{ij} \geq 0 \ \forall (i,j) \in A \tag{4.1}$$

$$f_{ij} = 0 \text{ if } \ i = j \ \forall (i,j) \in A \tag{4.2}$$

$$\sum_{j \in V} f_{ij} - \sum_{j \in W} f_{ji} - s_i = 0 \ \forall i \in W \tag{4.3}$$

$$P_{rx} \sum_{j \in W} f_{ji} + \sum_{j \in V} P_{tx,ij} f_{ij} - e_i \leq 0 \forall i \in W \tag{4.4}$$

$$\sum_{i | i \in W \cup (x_i, y_i) \in R_k} s_i = D_k \ \forall k \in A \tag{4.5}$$

$$s_i = D_k \ \forall i \in W, k \in A, (x_i, y_i) \in R_k \tag{4.6}$$

$$e_i \geq 0 \ \forall i \in W \tag{4.7}$$

$$s_i \geq 0 \ \forall i \in W \tag{4.8}$$

$$e_i = battery \ \forall i \in W \tag{4.9}$$

Figure 4.1: LP model for redundant and non-redundant data routing.

Table 4.1: Terminology for LP Formulation

Variable	Description
N	Number of nodes
Z	Set of the regions in the network
f_{ij}	Flow from node-i to node-j
s_i	Data generated at node-i in one day
P_{rx}	Energy consumption for receiving one bit of data
$P_{tx,ij}$	Energy consumption for transmitting one bit of data from node-i to node-j
e_i	Energy requirement for sensor node-i
G	Directed graph that represents network topology
V	Set of nodes, including the base station as node-1
W	Set of nodes, except the base station (node-1)
A	Set of edges (links)
D_k	Amount of data collected in k^{th} spatial granule
R_k	Set of all coordinates of k^{th} spatial granule
N_Z	Total number of regions
(x_i, y_i)	Coordinates of node-i

Constraint 4.1 states that all flows are non-negative. Constraint 4.2 is used to set flows that do not exist: there can't be a flow from base station to other nodes or from a node to itself. Constraint 4.3 states that the difference between the data flowing out of node-i and the data flowing into node-i is the data generated at node-i.

Constraint 4.4 states that for all nodes except the base station the energy consumed for transmission and receiving is equal to or less than the energy stored in batteries. In our model, we ignore the energy spent for sensing which is negligible as compared to energy spent for communication [29]. We use energy model described in section 3.1.2 and parameters in energy model listed in 3.1. $P_{tx,ij}$ represents the energy cost of transmitting one bit data between node-i and node-j. P_{rx} is the energy cost of receiving one bit data.

Equation 4.5 formulates the case of optimal cooperation between sensor nodes in each spatial granule for collecting measurement data i.e., no redundancy exists in the transmitted data. Total network area is divided into k different spatial granules and amount of data generated on each granule is D_k.

Equation 4.6 formulates the operation of sensor network in which sensor nodes gather data

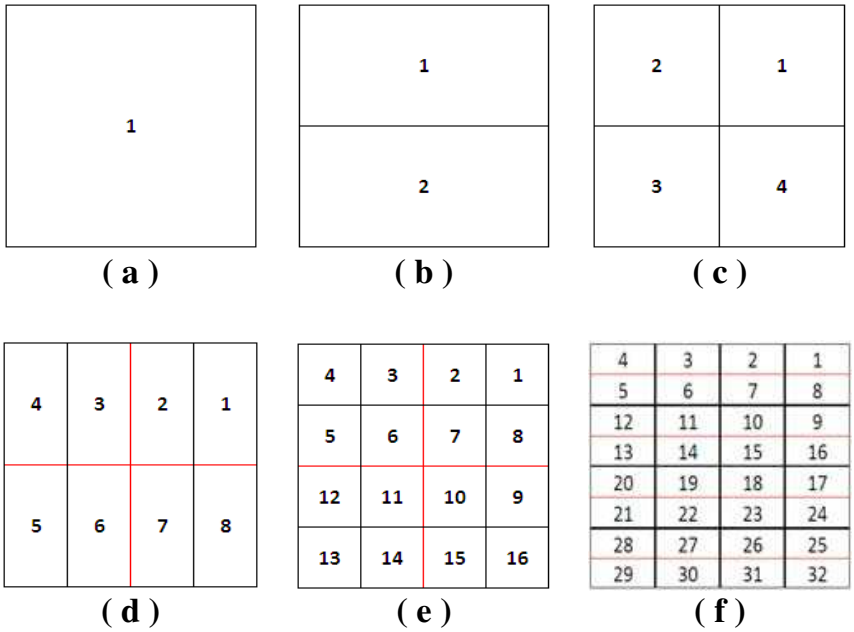

Figure 4.2: Map of spatial granules.

from their spatial granule without any cooperation with other nodes for gathering data. Therefore, each sensor node in the same spatial granule collects the same amount of data and redundant data is transmitted if more than one node is in the same spatial granule. Constraint 4.7 states that battery power for each node must be higher than zero. Constraint 4.8 implies that a sensor node can be used only as a relay without generating sensor data. Constraint 4.9 is used to assign equal energy to each sensor node.

4.2 Analysis

For the numerical analysis we assume that sensor nodes are deployed in a 200 meter x 200 meter square area that includes 1, 2, 4, 8, 16 or 32 spatial granules as shown in Figure 4.2. For each granule, we assume that 1600 bits of data needs to be collected. It is possible to assign different amount of data to each spatial granule according to the application characteristics but for simplicity we consider equal data distribution in the analysis. For example, if the area is divided in two spatial granules, we have to gather 1600 bits data from the first spatial granule and 1600 bits data from the second spatial granule to monitor these spatial granules.

For the deployment of sensor nodes, we use uniform random deployment strategy. In this strategy, density of sensor nodes is same for each spatial granule. We assume that locations of

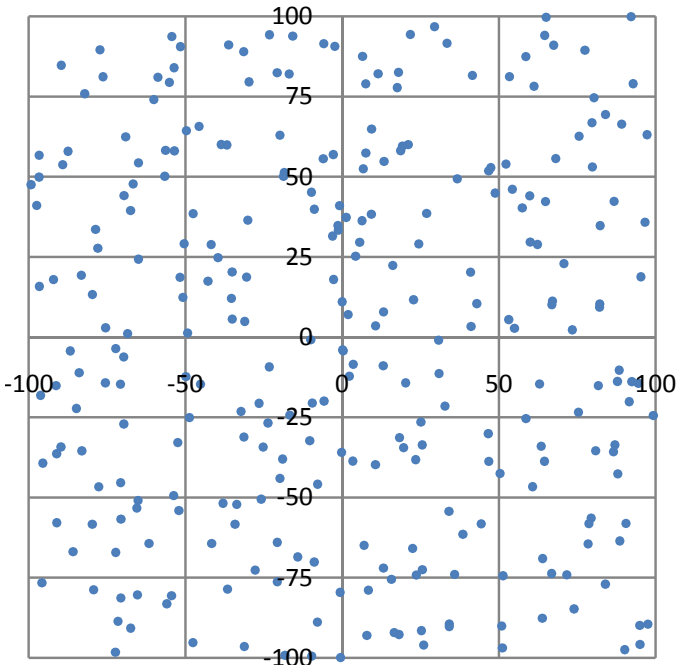

Figure 4.3: Deployment example (N=288).

Table 4.2: Parameters used in the analysis

Parameter	**Value**
Network area	200 m X 200 m
N	32, 64, 96, 128, 160, 192, 224, 256, 288
N_Z	1, 2, 4, 8, 16, 32
D_k	1600 bits

the nodes are known and the base station is located at the center of the area. Sensor nodes do not have a restriction on transmission range. We use energy model described in section 3.1.2 and the communication parameters are listed in Table 3.1. Figure 4.3 demonstrates a random deployment scenario with a total number of 288 nodes. The same number of nodes (i.e., 9) is deployed for each of 32 spatial granules illustrated in this figure. Table 4.2 lists the parameters used in our analysis.

4.2.1 Analysis for Redundant Data Elimination

First of all, we provide analysis results for the investigation of the impact of eliminating redundancy on the battery requirements. To model the optimal case of gathering data in which sensor nodes in the same spatial granule operate in a coordinated fashion and redundancy is totally eliminated. (4.1), (4.2), (4.3), (4.4), (4.5), (4.7), (4.8), and (4.9) are used as the

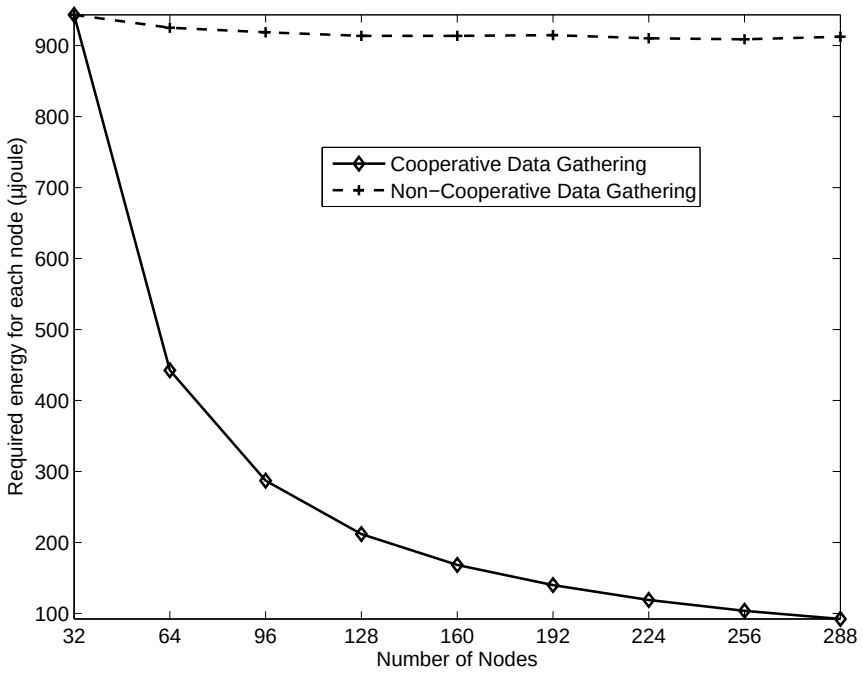

Figure 4.4: Effect of redundancy on energy requirements for sensor nodes.

constraints of our first linear program. For comparison reasons, we construct a second model with (4.1), (4.2), (4.3), (4.4), (4.6), (4.7), (4.8), and (4.9) used as the constraints in which no cooperation exists and redundancy elimination is not considered. In this analysis, the area is divided into 32 spatial granules. Number of sensors deployed on the area is changed between 32 and 288 in 32-node increments. Sensor nodes are always evenly distributed to spatial granules. Each models is solved for 100 random topologies and the results are averaged.

Figure 4.4 shows the amount of battery energy required for each node as a function of number of sensor nodes deployed in a 200m x 200m area. In cooperative data gathering case, 1600 bits of data is gathered for each granule whereas in non-cooperative case, the same amount of data is gathered from each individual sensor node, separately.

When 32 nodes are deployed to the area, since each granule includes only one sensor node, redundant data does not exist. Therefore, both models give the same result. However doubling the number of nodes reduces the energy requirement by more than 50% in cooperative case which is in contrast to non-cooperative operation. Without eliminating redundancy, the decrease for minimum battery requirements we can obtain by increasing node density is only marginal and quickly saturates. The slight improvement is due to increasing number of relaying options as the number of nodes in the network increases.

As we further increase number of nodes in the area, the difference in energy requirement per sensor node for two models also grows accordingly. Hence, the benefits of eliminating redundancy become more significant as the network becomes more-densely populated.

An interesting result here is that when each spatial granule hosts n sensors, the decrease in minimum energy requirements due to elimination of redundant data is even larger than n times. For instance when 9 nodes are deployed in each granule (sum up to 288 nodes in total), minimum energy required is 912,5 and 92,3 micro joules for non-cooperative and cooperative cases, respectively (the ratio is 9.88). To explain the reason for such a behavior, we note that both models provide data relaying in the optimal way. However, the net gain of jointly optimizing the collaboration of data gathering together with data relaying is greater than optimizing these two tasks independently.

4.2.2 Analysis for Effects of Spatial Granularity

After our analysis for the impact of redundant data elimination, in this section, we investigate the effect of spatial granularity of measurements on the minimum energy requirements. In the second analysis, we use the same set of parameters given in Table4.2. But this time, number of spatial granules in the network is changed between 1 and 32 as shown in Figure 1. For all analysis we assume redundancy is totally eliminated thus (4.1), (4.2), (4.3), (4.4), (4.5), (4.7), (4.8), and (4.9) are used as constraints for the linear program. Again, results of analysis are depicted after 100 iterations.

Figure 4.5 shows the amount of battery energy required per each node as a function of number of spatial granules for different number of nodes deployed in the area.

As can be seen in figure 4.5, battery energy required for each node increases as a function of number of spatial granules. A more careful treatment reveals that the increase in minimum energy requirement is slightly more than the increase in number of spatial granules. Thus, the relationship is not linear in strict sense. For instance, battery requirement per node for a 32-node 16-granule network (437,4 μJ) is more than twice the battery requirement for a 32-node 8-granule network (192,7 μJ). We see that this increase can almost be balanced by a proportional increase in number of nodes deployed in the network. For instance, battery requirement for 64-node 16-granule network (206,5 μJ) is slightly more than the one for a

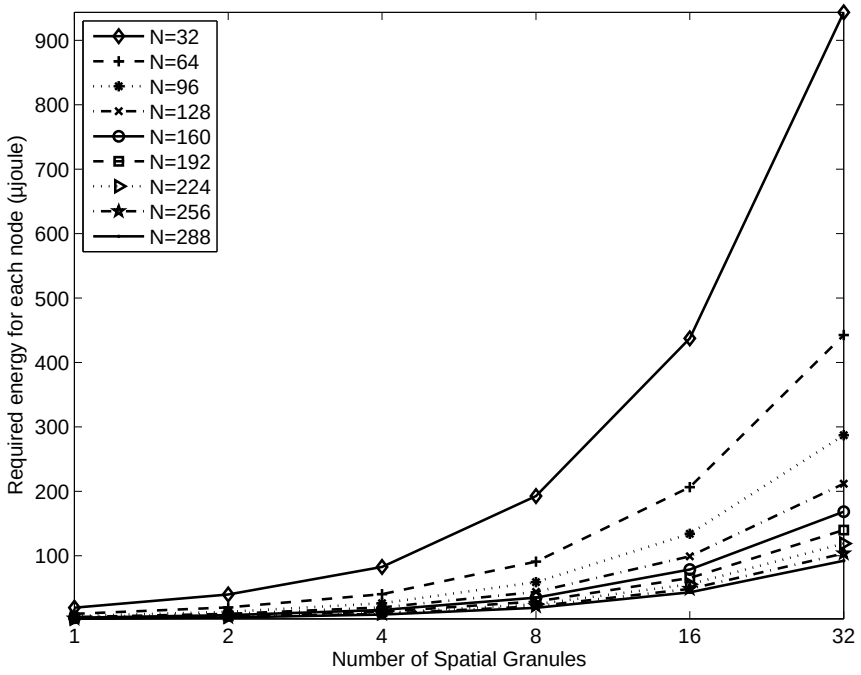

Figure 4.5: Effect of redundancy on energy requirements for sensor nodes.

32-node 8-granule network (192,7 μJ). For 32-node deployment, required energy for each sensor node increases about 48.5 times when the number of spatial granules is increased from 1 to 32. For 288-node deployment, the same ratio is less; 42.5 times. As we increase the node density, the network has more options both for data gathering and for relaying to base station thus more stringent spatial granularity requirements can be satisfied in a more energy-efficient way.

4.2.3 Analysis for Effects of Energy Distribution

In the analysis presented in section 4.2.1 and 4.2.2, we assume that each sensor node is loaded with same amount of energy. However if we want to minimize the total amount of energy spent in the entire network this may not be the optimal strategy because energy requirement for a sensor node may change according to its location and closeness to the base station. To find out the energy saving that is possible if different amounts of energy can be freely assigned to sensor nodes, we need to change the objective of the optimization problem as the minimization of total energy ($\sum_i e_i \; i \in W$) consumed in the network.

Figure 4.6 shows the total energy spent in the WSN as a function of number of spatial granules for two different cases. For both cases there are 288 nodes in the network. First case is for

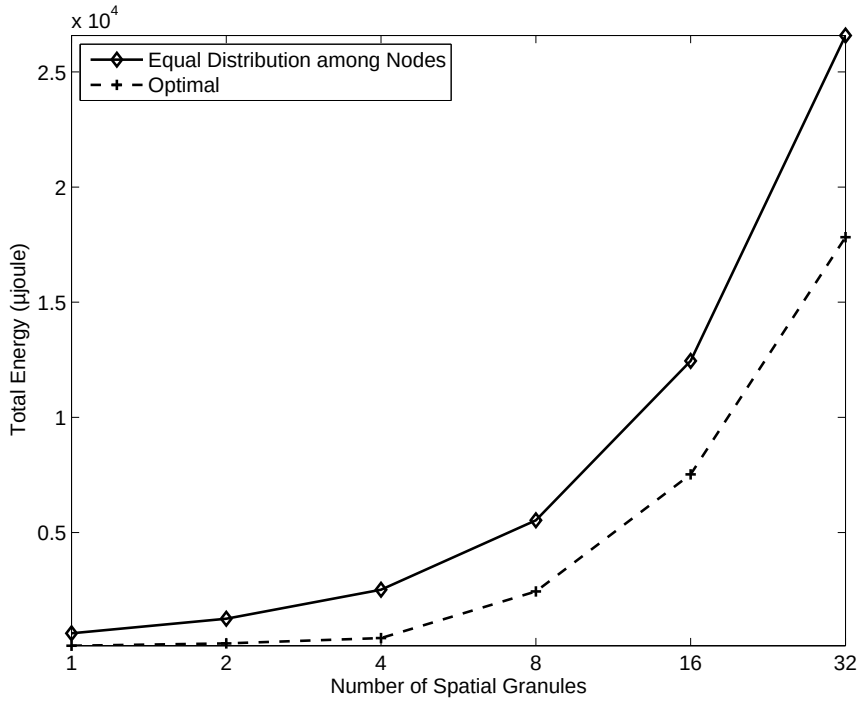

Figure 4.6: Effect of redundancy on energy requirements for sensor nodes.

equal energy distribution among all sensor nodes as was done in previous sections. Second case is our new scenario; the objective function is changed as described above and constraint 4.9 assigning equal energy to each sensor node is not used hence (4.1), (4.2), (4.3), (4.4), (4.5), (4.7), and (4.8) are used as the constraints of our new linear program.

Figure 4.6 reveals that a significant amount of energy saving can be obtained if total energy can be freely distributed among the nodes in the network. This saving in absolute numbers grows proportionally with number of spatial granules however the ratio of energy overhead due to equal energy distribution to the total energy spent in the network decreases as number of spatial granules in the network increases. For instance, total energy requirement when nodes are charged with the same battery is 65% and 49% more than the total energy spent in the optimal energy distribution case for 16-spatial granule and 32-spatial granule networks, respectively.

Figure 4.7 shows the energy consumed in each spatial granule when the total energy is optimally distributed among sensor nodes. It is interesting to see that there is not a strict inverse relationship between amount of energy required in each spatial granule and the distance of that spatial granule to the base station. More specifically, energy requirement for spatial granules with indices 10, 11, 22 and 23 is the maximum. A plausible explanation of this result is

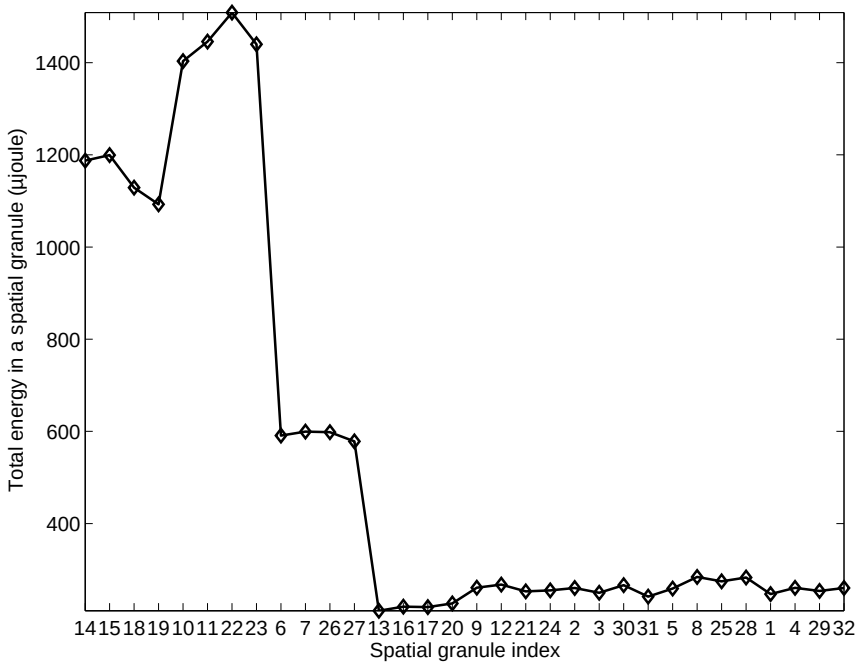

Figure 4.7: Effect of redundancy on energy requirements for sensor nodes.

that minimization of energy spent in the network is realized when the sensor nodes located on these more-distant granules take a larger role in data relaying than the nodes closer to the base station.

CHAPTER 5

PHYSICAL ATTACKS IN WIRELESS SENSOR NETWORKS

Physical attacks are a class of security attacks on WSNs in which sensor nodes are physically destroyed and rendered permanently non-operational [12, 13]. Physical attacks are easy to conduct, yet, their effects can be catastrophic (*i.e.*, these attacks can destroy vital nodes in the network and lead to inefficient energy dissipation trends in the whole network). Self organization is an effective way of countering against these attacks assuming that the number of nodes adversaries can destroy is not limitless. If the network is designed for survivability then, in case of a security attack remaining operational nodes can reorganize by taking over the functions (*e.g.*, relaying function) of the attacked nodes. Redundancy is another key factor for survivability; even if one or more nodes are dead, remaining nodes that share the same observation window (*i.e.*, if two nodes sense highly correlated data then their observation windows are overlapping) can perform the sensing functionality. However, such additional functions increase the energy dissipation of the remaining nodes.

In this chapter, we consider a WSN that is tasked with monitoring an area for a predetermined period of time. The area is divided into regions and sensor nodes are deployed uniformly throughout the network (*i.e.*, initially there are equal number of nodes in each region). Through this representative application, we investigate the energy cost of combating against physical attacks by using a novel LP framework.

5.1 System Model

In this section, we first define the research problem informally. Then, the formalization is carried out with LP modeling.

5.1.1 Motivation and Problem Definition

During their operational lifetime, sensor nodes can take at least two different roles: sensing role and relaying role. In the sensing role, a sensor node gathers data from the environment and either transmits the data directly to the base station or relays it to another sensor node acting as a relay. It is possible that some of the data can be transmitted directly to the base station and some of the data is conveyed via relay nodes. In the relay role, a sensor node forwards the data it received from either the source node or from another relay node to the base station or yet to another relay node. Data can be relayed successively until final destination (*i.e.*, the base station) is reached. Using wireless communication, sensor nodes form a network in self-organized fashion. This means nodes could collaborate and exchange the roles when needed.

To accomplish the task of monitoring a region for a pre-determined amount of time, the network must possess certain resources and capabilities. Energy is usually the most critical resource. A sensor node can satisfy its assigned tasks as long as it has sufficient energy resources. Sensor nodes deployed in hostile areas are vulnerable to physical attacks. If an attacker destroys a sensor node, other nodes have to bear the extra burden to fulfill the responsibility of the destroyed node. If the possibility of such a physical attack has not been considered, in case of an attack even though the remaining nodes are capable of taking over the roles of the destroyed nodes, their energy limitations can prevent them to accomplish the task. In other words, the self organizing nature of the network enable the network to operate with the remaining sensor nodes but these sensor nodes cannot continue to monitor the operation area as planned since their batteries do not last long enough.

We consider a WSN in which sensor nodes are deployed to monitor an area (a battlefield for instance) which is composed of a certain number of non-overlapping regions. The network must be able to collect data from the area for a pre-determined period of time. Before the deployment, each sensor node should be charged with the energy that is sufficient to accomplish the task. Initial energy of the nodes would not be enough to function long enough if the possibility of physical attacks was not considered because node failures may lead to inefficient energy dissipation trends. Informally speaking, the research problem is finding the minimum amount of additional energy to collect information collaboratively for a period of time from a designated area if certain number nodes fail to operate due to a physical attack.

In our threat model, we assume the base station is physically well-protected. We further assume that the attacker is not capable of destroying all of the nodes in a given region (*i.e.*, sensor nodes are camouflaged, hence, it is not possible to locate all of them in a reasonable amount of time). We consider two attack models: (i) uniform attack and (ii) non-uniform attack. In uniform attack, we consider a more powerful adversary which picks a certain number of nodes in each region and renders the targeted nodes inoperable. In non-uniform attack, nodes only in a single region are destroyed by the attacker.

5.1.2 Linear Programming Framework

In our framework, we assume that there are N sensor nodes and a single base station in the network. Data generated at each node, transferred either directly (single-hop) or through other sensors acting as relays (multi-hop), terminates at the base station. The network topology is represented as a directed graph $G = (V, A)$. V is the set of all nodes, including the base station as node-1. We also define set W, which includes all the nodes except node-1. $A = \{(i, j) : i \in W, j \in V - i\}$ is the set of arcs (links). The amount of data sent on the directed link (i, j) is denoted as f_{ij}.

We consider a uniform random deployment scenario in which sensor nodes are deployed over a rectangular area that includes N_Z number of regions to be monitored (Z denotes the set of regions and the members of set Z are denoted by Z_k). The set of all sensor nodes located within region-Z_k are denoted with W_k. Number of days that sensor node-i monitors region-Z_k is denoted as d_i and the total time region-Z_k monitored by all nodes located in it is denoted by D_k. In each region, s_i unit of raw data per day is to be conveyed to the base station. We also define a set F_k, which consists of failed (destroyed) nodes in region-Z_k and the number of elements in set F_k is M_k (the union of all F_k's constitute the set F).

We use energy model described in section 3.1.2 and the communication parameters are listed in 3.1. $P_{tx,ij}$ represents the energy cost of transmitting one bit data between node-i and node-j. P_{rx} is the energy cost of receiving one bit data.

The optimization problem is formulated as an LP problem. Figure 5.1 presents the basis of our formulation, which is similar to the models in earlier work [29] (we modify this basic model to suit for our needs and expand it with additional constraints). Since the objective is

to minimize *battery*, the problem is the minimization of the maximum battery requirement of the nodes in the network by finding the f_{ij}'s (flows) that satisfy the constraints.

Minimize *battery*
Subject to:

$$f_{ij} \geq 0 \; \forall (i,j) \in A \tag{5.1}$$

$$f_{ij} = 0 \text{ if } \; i = j \; \forall (i,j) \in A \tag{5.2}$$

$$\sum_{j \in V} f_{ij} - \sum_{j \in W} f_{ji} - d_i s_i = 0 \; \forall i \in W \tag{5.3}$$

$$P_{rx} \sum_{j \in W} f_{ji} + \sum_{j \in V} P_{tx,ij} f_{ij} - e_i \leq 0 \forall i \in W \tag{5.4}$$

$$e_i = battery \; \forall i \in W \tag{5.5}$$

Figure 5.1: LP model as the basis for investigating the energy cost of mitigating physical attacks in WSNs.

Equation 5.1 states that all flows are non-negative. Equation 5.2 is used to eliminate infinite loops - there cannot be a flow from the base station to other nodes or from a node to itself. Equation 5.3 states that the difference between the data flowing out of node-i and the data flowing into node-i is the data generated at node-i. Equation 5.4 states that for all nodes except the base station the energy consumed for transmission and receipt of data is equal to or less than the energy stored in batteries. Equation 5.5 is used to assign equal energy to each sensor node. As noted earlier, the model presented in Figure 5.1 and explained so far is the basic model for flow balancing (*i.e.*, all data generated at the sensor nodes eventually terminate at the base station) and energy minimization (*i.e.*, to minimize the maximum energy dissipation of nodes, all sensor nodes are forced to dissipate their energies in a balanced fashion).

Failed nodes cannot participate in data gathering or relaying, thus, any flows originating or flowing through such a node should be set to zero. Equation 5.6 incorporates this restriction into our model:

$$f_{ij} = 0 \text{ if } \; (i \in F \; or \; j \in F) \; \forall (i,j) \in A \tag{5.6}$$

To model the optimal case of monitoring the area in which sensor nodes in the same region operate in a coordinated fashion and redundancy is totally eliminated, we introduce the following constraint:

$$\sum_{i \in W_k} d_i = D_k \forall Z_k \in Z \tag{5.7}$$

Equation 5.7 formulates the case of optimal cooperation between sensor nodes in each region for monitoring the region (*i.e.*, no redundancy exists in the transmitted data). Total network area is divided into N_Z non-overlapping regions and region-Z_k must be monitored D_k days by the sensor nodes that are located in region-Z_k. Note that in each region-Z_k the set of nodes located in that region (W_k) are essentially observing the same phenomena (*i.e.*, they would acquire redundant data if they transmitted data simultaneously). The model consisting of constraints presented in Equations 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, and 5.7 is called Optimal Role Assignment (ORA) Model. In ORA model only a single node performs the sensing operation at a time in each region (*i.e.*, nodes take turns for sensing the data).

Equation 5.8 formulates the operation of WSN without any cooperation in data acquisition (*i.e.*, the redundant data acquisition case). Each sensor node in the same region-Z_k collects the same amount of data and all redundant data is conveyed to the base station.

$$d_i = D_k \ \forall i \in W_k, Z_k \in Z \tag{5.8}$$

The set of constraints presented in Equations 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, and 5.8 are used to model the network operation mode where nodes do not perform sensing collaboratively (*i.e.*, each sensor node performs sensing and sends the data to the base station without exploiting the data redundancy). We call this model Redundant Data Sensing (RDS) model. Note that in RDS model energy minimization is achieved by optimizing only f_{ij}'s (*i.e.*, d_i's are fixed), however, in ORA model both flows (f_{ij}'s) and monitoring times (d_i) are jointly optimized. Note that Equation 5.7 and Equation 5.8 are not used simultaneously. All system variables with their acronyms and descriptions are presented in Table 5.1.

5.2 Analysis

In our analysis, we solve the LP problems introduced in previous section for a simple exemplifying topology with a square area of size 200 m x 200 m. As shown in Figure 5.2, the area consists of 32 regions (*i.e.*, $N_Z = 32$). The task of the sensor nodes in each region is to monitor it for 50 days (*i.e.*, $D_k = 50$ days). Note that we opt to adopt a uniform monitoring time for all regions. In each day, same amount of data ($s_i = 1\ Mb$) is collected for each region. Each region has the same number of randomly deployed sensor nodes. We use 192 nodes ($N = 192$) in total and the base station is at the center (*i.e.*, there are six sensor nodes in

Table 5.1: Terminology for LP Formulations

Variable	Description
N	Number of nodes
Z	Set of the regions in the network
f_{ij}	Flow from node-i to node-j
s_i	Data generated at node-i in one day
d_i	Number of days that node-i monitors a region
P_{rx}	Energy consumption for receiving one bit of data
$P_{tx,ij}$	Energy consumption for transmitting one bit of data from node-i to node-j
ρ	Energy dissipated in the electronic circuitry
ε	Transmitters efficiency
α	Path loss exponent
e_i	Energy requirement for sensor node-i
G	Directed graph that represents network topology
V	Set of nodes, including the base station as node-1
W	Set of nodes, except the base station (node-1)
A	Set of edges (links)
D_k	Total number of days that region-Z_k must be monitored
Z_k	A member of set Z
N_Z	Total number of regions
W_k	Set of sensor nodes in region-Z_k
N_k	Total number of sensor nodes in region-Z_k
F	Set of failed nodes in the network
M	Total number of failed nodes in the network
F_k	Set of failed nodes in region-Z_k
M_k	Total number of failed nodes in region-Z_k

each region; $N_k = 6$). All nodes perform the relaying operation collaboratively and there are no restrictions on their transmission ranges (*e.g.*, any node-$i1$ in any region-Z_{k1} can send data to any other node-$i2$ in any other region-Z_{k2}). We use GAMS [109] to solve the LP models. All data points are the averages of the results of 100 random topologies. The parameters used in the analysis are presented in Table 5.2.

To evaluate the benefits of eliminating data redundancy we first perform an analysis without any node failures ($M_k = 0,\ \forall Z_k \in Z$) by using ORA and RDS models. The required battery energy for monitoring the network is found to be 4.37 J and 28.58 J for ORA and RDS models, respectively. In RDS model, amount of data collected is six times more than ORA

4	3	2	1
5	6	7	8
12	11	10	9
13	14	15	16
20	19	18	17
21	22	23	24
28	27	26	25
29	30	31	32

Figure 5.2: Region map in our analysis

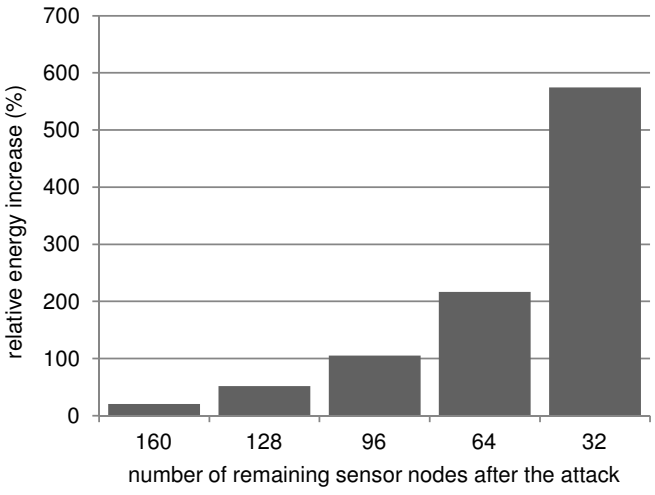

Figure 5.3: Relative energy increase per node as a function of number of remaining nodes in the network.

Table 5.2: Parameters used in the analysis

Parameter	**Value**
Network area	200 m X 200 m
N	192
N_Z	32
D_k	50 day
s_i	1 Mb / day

model but corresponding energy requirement increases slightly more than six times. In the ORA model, sensor nodes monitor the area by taking turns, however, they always participate in data relaying. The solution of the LP model with the chosen parameter set provides the optimum amount of time assigned to each sensor node to monitor their regions. For example, sensor nodes in region-1 has the following monitoring times; node-1: 2.52 days, node-2: 2.48 days, node-3: 15.09 days, node-4: 2.83 days, node-5: 23.78 days, and node-6: 3.30 days. Depending on their locations, some nodes use most of their energies for relaying data, thus, they take less role in sensing and generating data and dissipate less energy for it. In the remaining analysis, we concentrate on the ORA model and do not consider RDS model.

In Figure 5.3, relative energy overhead (*i.e.*, percentage energy increase when compared to the case where none of the nodes fail – $M_k = 0 \ \forall Z_k \in Z$) is presented as a function of number of remaining nodes in the network for the case of uniform attack. In uniform attacks, equal number of nodes are dead in each region (*e.g.*, for 128 remaining nodes in the network a total of 64 nodes, two from each region, are dead). Percentage energy overhead grows from 20 % (when only one sensor node dies in each region – 160 sensor nodes remain in the network) to 574 % (when five sensor nodes die in each region – 32 sensor nodes remain in the network).

In Figure 5.4, relative energy overhead for each region is plotted for different number of nodes failed due to a nonuniform attack. In this case, indicated number of nodes are dead only in one region – none of the nodes in other regions are dead. After nodes are failed due to a physical attack, the network reorganizes itself and remaining sensor nodes update their sensing and relaying patterns. We assume that at most five sensor nodes fail to operate in each region after the attack and one node is enough to cover and monitor the region. In Figure 5.4, regions are ordered in groups of 4 according to their distance from the base station.

Figure 5.4 reveals an interesting energy dissipation trend, described as follows. Sensor nodes

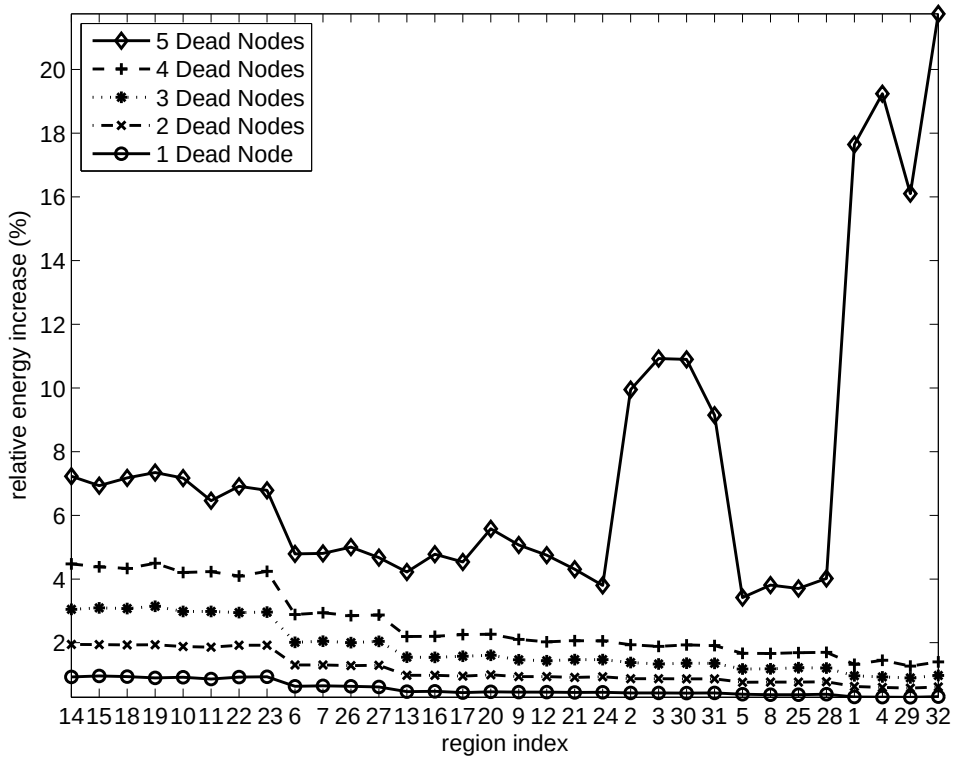

Figure 5.4: Relative energy increase per node as a function of region index for different number of dead nodes in each region.

close to the base station directly send most of their data to the base station but sensor nodes far away from the base station need to transmit via a multi-hop route (*i.e.*, directly sending collected data to the base station is not energy efficient). As a result, sensor nodes close to the base station become more heavily used than the others. For this reason, attacks against regions close to the base station cause higher energy overhead per sensor node when compared to the attacks performed to other regions of the network. There is an exception to this mechanism when there are five failed nodes in each region (*i.e.*, there is only one surviving node in each region; $M_k = 5$). In this situation, the minimum energy requirement of the network is the highest when attacks are directed against the farthest regions from the base station. The reason for such behavior is that network cannot share the burden of monitoring a region when there is only one node left alive in the region (the nodes in other regions can cooperate only for relaying the data out of the region and the remaining operational node in the region has to get the data out of the region on its own). The burden of getting the data out of the farthest regions is especially heavy due to the extended distances to the base station.

CHAPTER 6

OPTIMAL NUMBER OF ROUTING PATHS IN MULTI-PATH ROUTING TO MINIMIZE ENERGY CONSUMPTION IN WIRELESS SENSOR NETWORKS

Earlier studies revealed that multi-path routing improves the energy efficiency of WSNs, however the impact of the number of paths on the level of energy efficiency is not well understood (*i.e.*, what amount of energy efficiency improvement should be expected with each additional path?). As the number of paths increases so is the complexity of the routing protocol. Hence, it is desirable to limit the number of paths at some point if further increase does not improve energy efficiency significantly, yet, an analysis of the impact of limiting the number of paths on energy efficiency has not been performed in the literature.

In this chapter, our goal is to investigate the impact of multi-path routing on energy efficiency in WSNs. By developing a novel problem formulation using Mathematical Programming, we capture the essence of both routing and energy dissipation characteristics of multi-path routing. We analyze the impact of limiting the number of routes from energy efficiency perspective within a general framework and without considering any specific protocol or algorithm. This approach abstracts us away from the protocol-specific overhead or implementation details. Characterization of the impact of limiting the number of routes on energy balancing in WSNs is a novel research contribution and may provide valuable insights for the design of future protocols.

There are several related concepts in this area which can also be referred as multi-path routing and may be confused with the meaning of multi-path routing as it is used in this chapter. In this chapter, we use the term multi-path routing for partitioning the data into groups of data

packets without employing data redundancy and sending each group of packets via a different path towards the base station. Alternate path routing is different than multi-path routing in the sense that a single path is used in normal operation but alternative paths are kept ready to be used in case the primary path becomes unavailable. Redundant multi-path routing is another related term that means the data to be conveyed to the base station is transported via multiple paths with added redundancy (*e.g.*, multiple replicas of the data are sent on different paths) [21].

6.1 Model

In our network model, there is a single base station and N sensor nodes in the network. Each sensor node-i creates the same number of data packets (s_i) with packet length L_P bits at each round to be conveyed to the base station (*i.e.*, sensor nodes create CBR flows). Data packets are treated as indivisible data units (*i.e.*, data packets are neither fragmented nor combined with other data packets until they reach the base station). Time is organized into rounds with duration T_{rnd} and the total number of rounds is M_{rnd}. The network topology is represented by a directed graph, $G = (V, A)$, where V is the set of all nodes including the base station as node-0. We also define set W which includes all nodes except node-0 (*i.e.*, $W = V \setminus \{0\}$). Each node can forward its generated data towards the base station using at most N_P paths. Different paths can be either disjoint or braided (*i.e.*, two paths can share common links in their chains of links forming paths). $A = \{(i, j) : i \in W, j \in V - i,\ d_{ij} \leq R_{max}\}$ is the ordered set of arcs. Note that the definition of A implies that no node sends data to itself or to a node that is separated from it beyond the maximum transmission range R_{max}. Data generated at node-k forwarded on the l^{th} path flowing from node-i to node-j is represented as f_{ij}^{kl}. Moreover, b_k^l is the total amount of data packets generated by sensor node $k \in W$ and transmitted on the l^{th} configuration to the base station and a_{ij}^{kl} indicates if arc $(i, j) \in A$ is used in the l^{th} routing configuration originated at sensor $k \in W$.

We use the energy model described in section 3.1.2 and parameters in this energy model listed in 3.1. We assume that the transmission energy model is such that the bit error rate (BER) is constant and same for all links [103, 112, 113]. The amount of energy to transmit L_P bits of data between node-i and node-j is $P_{tx,ij}(L_P)$) and to receive L_P bits of data is $P_{rx}(L_P)$. Packet error rate is $\chi = (1 - (1 - \varphi)^{L_P})$, where φ is the BER. Each packet has to be transmitted

$\lambda = 1/(1-\chi)$ times (average packet retransmission rate), on the average, for successful delivery of the packets. The interference range of a transmission from node-i to node-j is γd_{ij}, where γ is the interference range multiplier [99] and d_{ij} is the distance between node-i and node-j. To model interference between links we define a binary interference matrix, I^i_{jm}, presented in Equation 6.1. If node-i is in the interference region of the transmission from node-j to node-m (*i.e.*, $\gamma d_{jm} \geq d_{ji}$), then node-$i$ is blocked from receiving any data because any such flow to node-i results in a conflict (packet collusion). Therefore, if $I^i_{jm} = 1$ then node-i has a conflict with the flow on arc (j, m) (node-i is sharing the bandwidth with the flow on arc (j, m)). On the other hand, if $I^i_{jm} = 0$ then flow on arc (j, m) is not conflicting with node-i. Generally speaking, interference range is equal to or greater than transmission range (*i.e.*, $\gamma \geq 1$) [99, 114]. This means depending on the value of γ, node-j's transmission to node-m can interfere with node-i even if the distance between node-j and node-m is less than the distance between node-j and node-i.

$$I^i_{jm} = \begin{cases} 1 & \text{if } \gamma d_{jm} \geq d_{ji} \; \forall j \in W, \forall m \in V \setminus \{j\} \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

The objective of the optimization problem is to minimize the maximum energy requirement (E) of sensor nodes. The network flow is modeled in the form of a series of constraints presented in Figure 6.1. All system variables with their acronyms and descriptions are presented in Table 6.1.

In Figure 6.1, constraint (6.2) limits the energy used by each sensor node for data transmission and reception by the total battery energy allocated to it. In fact, the objective is to minimize the energy dissipation of the maximum energy dissipating sensor node. The expression in the parenthesis gives the energy dissipation of node-i on packet transmission and reception for conveying source node-k's data on its l'th routing path. Summation over k and l gives the total energy dissipation of node-i. Energy dissipation for retransmissions are incorporated into the model through the multiplication of the whole expression by λ. If there is no retransmission, then $\lambda = 1$.

Constraint (6.3) is known as the flow conservation constraint, which is satisfied for all i (all nodes including the base station), k (sensor nodes), and l (routing paths). If node-i is the source node ($i = k$) then the difference between the sum of outgoing flows and the sum of incoming flows is the total amount of packets injected into the network by source node-k on

its l'th routing path (b_k^l). If $i = 0$ (the base station) then the all packets generated at each node-k and transmitted on path-l (b_k^l) reach the base station. If $i \neq k$ and $i \neq 0$ then the sum of incoming flows is equal to the sum of outgoing flows (node-i is a relay node for source node-k's flow on its l'th path). In summary, constraint (6.3) ensures that all flow generated at each node-k and transmitted on path-l reach the base station.

Constraint (6.4) ensures that data generated at sensor node-k and routed out to the rest of the network does not loop back to node-k. In other words, the sum of flows generated at node-k and received by node-k itself is zero. Note that constraint (6.10) ensures that all flows are non-negative, hence, constraint (6.4) together with constraint (6.10) dictates that the value of any flow creating any possible loop is exactly zero.

Constraint (6.5) guarantees that each sensor $k \in W$ generates and sends exactly a total of $s_k M_{rnd}$ packets to the base station. The total amount of data packets generated at node-k is routed to the base station by using at most N_P paths and the amount of data injected by node-k into each one of the paths is denoted as b_k^l, hence, the summation over l for each k gives the total amount of data generated at node-k. Since both the flows and the amount of data injected on each path are integer variables the packets cannot be split (all packets are created as L_P bits long and reach the base station with the same length as they are formed). However, different paths can be used in a periodic time interleaved fashion. It is also possible that different paths are used to convey data in an aperiodic sequential arrangement. For example, if $s_k = 1$ packet, $M_{rnd} = 3600$ rounds, $N_P = 3$ paths, $b_5^1 = 1800$ packets, $b_5^2 = 1200$ packets, and $b_5^3 = 600$ packets then node-5 can create a cyclic structure of length 6 rounds. At each cycle of six rounds, three data packets, two data packets, and one data packet are conveyed to the base station using the first path, the second path, and the third path, respectively. Alternatively, node-5 can convey all its data from round 1 to round 1800 on its first path, from round 1801 to round 3000 on its second path, and from round 3001 to round 3600 on its third path. In our model, we do not impose any timing restriction on scheduling. We determine the optimal paths and the amount of data transported on each path throughout the entire network operation as specified by N_P, s_k, M_{rnd} and other parameters. In fact, all feasible schedules that do not violate flow constraints in our model are equivalent.

Constraint (6.6) ensures that an arc $(i, j) \in A$ is marked as used for conveying data generated at node-k on its l'th path only if there is positive flow on $(i, j) \in A$ ($a_{ij}^{kl} = 1$ if $f_{ij}^{kl} > 0$). Note

Minimize E
Subject to:

$$\lambda \sum_{l=1}^{N_P} \sum_{k \in W} \left(\sum_{(i,j) \in A} P_{tx,ij}(L_P) f_{ij}^{kl} + \sum_{(j,i) \in A} P_{rx}(L_P) f_{ji}^{kl} \right) \leq E \quad \forall i \in W \tag{6.2}$$

$$\sum_{(i,j) \in A} f_{ij}^{kl} - \sum_{(j,i) \in A} f_{ji}^{kl} = \begin{cases} b_k^l & i = k \\ -b_k^l & i = 0 \\ 0 & \text{otherwise} \end{cases}$$
$$\forall i \in V, k \in W, l = 1, .., N_P \tag{6.3}$$

$$\sum_{(j,k) \in A} f_{jk}^{kl} = 0 \quad \forall k \in W, l = 1, ..., N_P \tag{6.4}$$

$$\sum_{l=1}^{N_P} b_k^l = s_k M_{rnd} \quad \forall k \in W \tag{6.5}$$

$$f_{ij}^{kl} \leq s_k M_{rnd} a_{ij}^{kl} \quad \forall (i,j) \in A, k \in W, l = 1, .., N_P \tag{6.6}$$

$$\sum_{(i,j) \in A} a_{ij}^{kl} \leq 1 \quad \forall i, k \in W, l = 1, .., N_P \tag{6.7}$$

$$\sum_{(j,0) \in A} f_{j0}^{kl+1} \leq \sum_{(j,0) \in A} f_{j0}^{kl} \quad \forall k \in W, l = 1, .., N_P - 1 \tag{6.8}$$

$$\lambda \frac{L_P}{\xi} \sum_{l=1}^{N_P} \sum_{k \in W} \left(\sum_{(i,j) \in A} f_{ij}^{kl} + \sum_{(j,i) \in A} f_{ij}^{kl} + \sum_{(j,m) \in A \setminus \{i\}} f_{jm}^{kl} I_{jm}^{i} \right)$$
$$\leq M_{rnd} T_{rnd} \forall i \in V \tag{6.9}$$

$$f_{ij}^{kl} \geq 0 \quad \forall (i,j) \in A, k \in W, l = 1, .., N_P \tag{6.10}$$

$$a_{ij}^{kl} \in \{0, 1\} \quad \forall (i,j) \in A, k \in W, l = 1, .., N_P \tag{6.11}$$

$$b_k^l \geq 0 \quad \forall k \in W, l = 1, .., N_P \tag{6.12}$$

Figure 6.1: MIP model for multi-path routing

Table 6.1: Terminology for MIP formulation

Variable	Description
N	Number of nodes
f_{ij}^{kl}	Number of data packets generated at node-k forwarded on the l^{th} path flowing from node-i to node-j
s_i	Number of data packets generated at node-i
$P_{rx}(L_P)$	Energy consumption for receiving L_P bits of data
$P_{tx,ij}(L_P)$	Energy consumption for transmitting L_P bits of data from node-i to node-j
d_{ij}	Distance between node-i and node-j
E	Battery energy of each sensor node
$G = (V, A)$	Directed graph that represents network topology
V	Set of nodes, including the base station as node-0
W	Set of nodes, except the base station (node-0)
A	Set of edges (links)
l	Set of paths
a_{ij}^{kl}	Binary variable to determine if arc $(i, j) \in A$ is used in the l^{th} routing configuration originated at sensor $k \in W$.
b_k^l	Total amount of data sensed by sensor $k \in W$ and transmitted on the l^{th} configuration to the base station
R	Radius of deployment area
R_{max}	Maximum transmission range
N_p	Number of paths
L_P	Packet length in bits
M_{rnd}	Number of rounds
T_{rnd}	Round duration
ξ	Channel bandwidth
γ	Interference range multiplier
φ	BER (bit error rate)
χ	Packet error rate
λ	Average packet retransmission rate

that the value of f_{ij}^{kl} can at most be $s_k M_{rnd}$. Such a case happens if node-k uses only one routing path (*i.e.*, $b_k^1 = f_{ij}^{k1}$ and $b_k^m = 0$ for $m > 1$). If $f_{ij}^{kl} = 0$ then binary variable a_{ij}^{kl} can be either one or zero (*i.e.*, by itself constraint (6.6) does not force a_{ij}^{kl} to neither one nor zero for $f_{ij}^{kl} = 0$). However, constraint (6.7) as described below forces a_{ij}^{kl} to be zero if both options are feasible. Hence, constraint (6.6) in conjunction with constraint (6.7)results in $a_{ij}^{kl} = 0$ if $f_{ij}^{kl} = 0$.

The flow on each configuration is guaranteed to be non-bifurcated by constraint (6.7). Note that constraint (6.7) must be satisfied for all values of i, k, and l. Consider one of possible combinations; $(i, k, l) = (3, 3, 1)$. For this example, constraint (6.7) states that data generated

at node-3 designated to flow on its first routing path transmitted by node-3 (the first hop of the path) can have only one receiver (*i.e.*, there should be only one second hop node, which can be the base station or another node acting as a relay). The summation over arc set $(3, j)$ guarantees that only one of the arcs $(3, j)$ have non-zero flow because the sum is equal to or less than one. In the same example, assume that $j = 7$ (i.e., $f_{37}^{31} = b_3^1$ and $f_{3m}^{31} = 0$ for all $m \neq 7$). As the second hop relay, node-7 can transmit the data it received from node-3 (f_{37}^{31}) to only one of its neighbors (dictated again by constraint (6.7)). If the third hop relay is node-8 then $f_{78}^{31} = f_{37}^{31} = b_3^1$ and $f_{7m}^{31} = 0$ for all $m \neq 8$. Continuing is this manner, data injected by source node-3 to its first path reaches the base station without being split into multiple branches. In other words, constraint (6.7) enables the construction of an unbroken and non-branching logical pipe (path) from the source to the base station for transportation of data. Indeed, N_P is the upper limit on the number of such pipes for each source node. The maximum number of such pipes in a network of M sensor nodes can be MN_P.

Constraint (6.8) is used to have a logical ordering of the configurations for originator nodes. Constraint (6.8) implies that $b_k^1 \geq b_k^2 \geq b_k^3 \geq ... \geq b_k^{N_P}$ (*i.e.*, number of packets conveyed on the l'th path of source node-k is greater than or equal to number of packets conveyed on its $(l + 1)$'th path).

To address bandwidth limitations in a broadcast medium, we need to make sure that the bandwidth used to transmit and receive at each node is limited by the available channel bandwidth. Such a constraint should take the shared capacity into consideration. For node-i, we refer to the flows around node-i which are not flowing into or flowing out of node-i but affecting the available bandwidth available to node-i as interfering flows. Constraint (6.9) guarantees that for each node (including the base station) the aggregate amount of incoming flows, outgoing flows, and interfering flows can be scheduled within the given time frame (T_{rnd} sec/round $\times$ M_{rnd} rounds = $T_{rnd}M_{rnd}$ sec). The summation over l, k, (i, j), and (j, m) gives the total number of packets sharing the capacity of node-i. Multiplication by L_P (bits/packet) converts the number of packets to number of bits. Division by ξ (bits/sec) transforms number of bits to seconds. Scaling with λ is for the extra time needed due to retransmissions. This constraint is a modified version of the sufficient condition given in [99]. We note that in the numerical analysis, we choose the parameters affecting constraint (6.9) in such a way that the maximum value of the left hand side of the inequality is more than an order of magnitude lower than the right hand side value, therefore, construction of a conflict-free

transmission schedule through a non-complicated time-slot assignment algorithm is possible[1].

Finally, constraint (6.10), constraint (6.11), and constraint (6.12) are nonnegativity constraints for the variables of the model.

The objective is to minimize E, which is the energy of the battery in each node. Once the parameter N_P is set, the solution of the model gives the set of paths each node uses to forward its data and the amount of data transported on each of these paths in a way that the energy required by the most energy consuming node is minimized. As a result, all nodes transmit their data in order to keep the required battery energy per sensor node as low as possible. In other words, all nodes dissipate their energies in the most balanced fashion. Sensor nodes are not required to use exactly N_P paths (*e.g.*, it is possible for a node use only two paths for transporting all its generated data even if $N_P > 2$). Furthermore, the amount of data flow on each path (b_k^l) is also determined by the MIP framework to optimize energy dissipation.

6.2 Analysis

In our analysis, we investigate two deployment scenarios: (i) linear deployment in which nodes are deployed equidistantly on a line[2] and (ii) uniform random distribution in which N nodes are deployed in a disc of radius R. We assume that there is a single base station located at one end in linear deployments and at the center in disc deployments. The communication parameters are listed in 3.1. For random deployment scenarios, each problem is solved for 100 random topologies and the results are averaged. The parameters used in the analysis are presented in Table 6.2.

A small-scale WSN topology is presented in Figure 6.2 to illustrate the network dynamics clearly[3]. When there is no limit on the number of paths used by each sensor node ($N_P \rightarrow \infty$), the required battery energy for each node is 5.86 J (*i.e.*, energy dissipations of all nodes are exactly balanced). For the optimal case, node-4 and node-5 use three paths and other nodes use a single path. When the number of paths used by each sensor node is upper limited by 2

[1] It is also shown that well designed Carrier Sense Multiple Access (CSMA) based MAC protocols are highly successful in reducing the collision rate to negligible levels provided that the network traffic is much lower (*e.g.*, an order of magnitude) than the available capacity [115, 116].

[2] There are many applications for linear sensor network deployments including border surveillance, highway traffic monitoring, safeguarding railway tracks, oil and natural gas pipeline protection, structural monitoring and surveillance of bridges and long hallways [117].

[3] We prefer line topology to avoid more complex flow patterns in Figure 6.2.

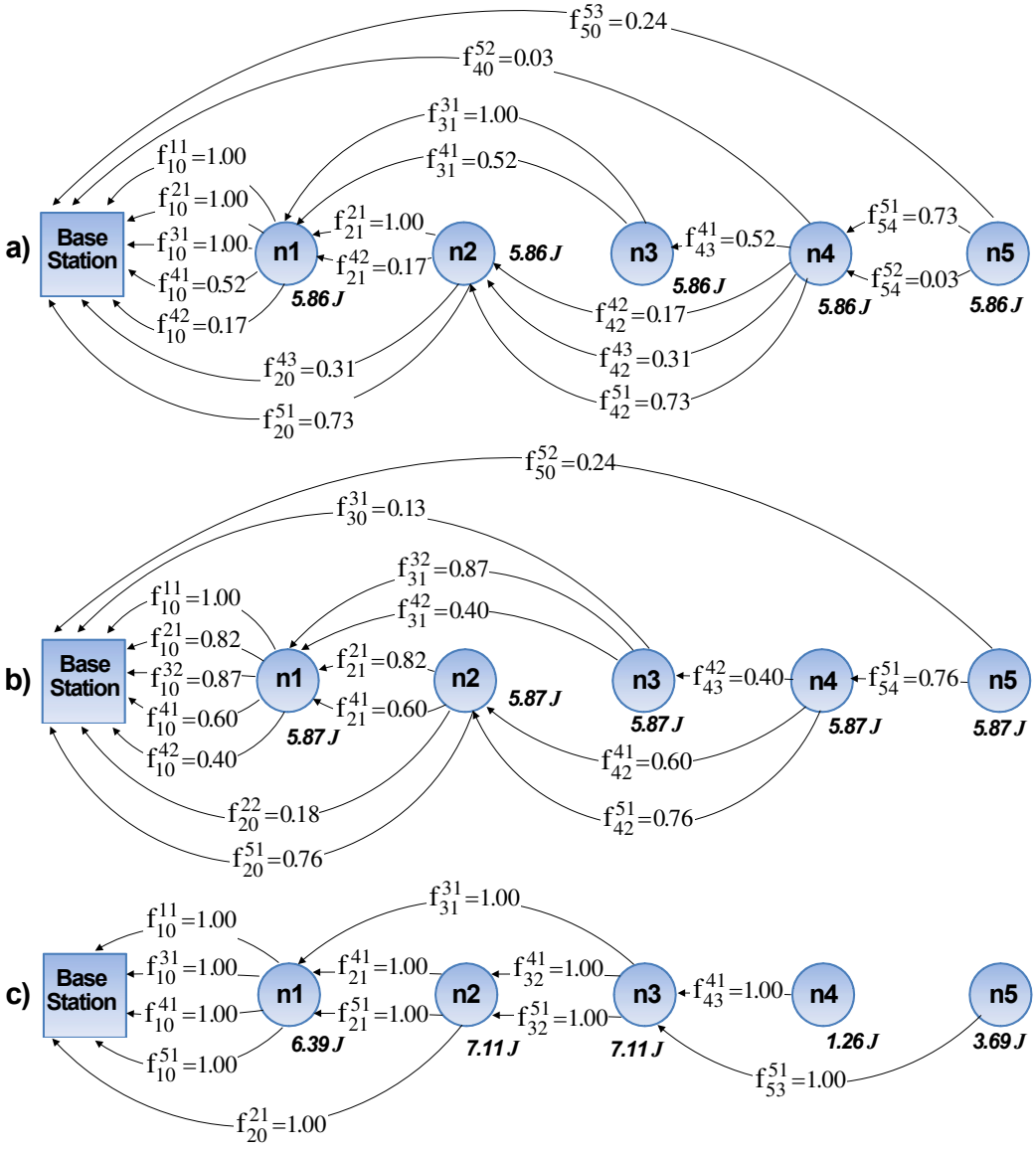

Figure 6.2: Optimal flows that minimize energy dissipation in (one dimensional) linear topology is illustrated by using the example topology.

Table 6.2: Parameters used in analysis

Parameter	Values
N	5-50
Deployment Scenarios	Linear (1-D) Equidistant, Disc (2-D) Random
Inter-node Distance (1-D)	30 m, 10 m
Network Radius R (2-D)	100 m – 1600 m
s_i	1 packet
N_p	1 - ∞
R_{max}	$R - R/2$
L_P	2048 bits (256 Bytes)
M_{rnd}	3600 rounds
T_{rnd}	60 s
ξ	256 Kbps
γ	1.7
φ	10^{-4}
χ	0.18
λ	1.22

($N_P = 2$), the required energy for each sensor node becomes 5.87 J (*i.e.*, percentage energy overhead is 0.14 % with respect to the $N_P \rightarrow \infty$ case). Note that all sensor nodes spend the same amount of energy, however, energy dissipation is slightly higher than the $N_P \rightarrow \infty$ case due to the suboptimal path selection.

For the case of $N_P = 1$ (*i.e.*, single path routing), energy overhead becomes 21.30 % when compared to $N_P \rightarrow \infty$ case – energy requirement for the maximum energy dissipating node (node-2) is 7.11 J[4]. Unlike $N_P \rightarrow \infty$ and $N_P = 2$ cases, in $N_P = 1$ case, sensor nodes do not spend equal amount of energy (*e.g.*, 6.39 J for node-1 and 7.11 J for node-2). Hence, we observe that single path routing cannot lead to a balanced energy dissipation regime in the network, which leads to over-utilization of some nodes' batteries.

In Figure 6.3, energy overhead (with respect to $N_P \rightarrow \infty$ case) as a function of inter-node separation is presented for linear topology and for $N_P = 1$ and $N_P = 2$ cases with number of nodes ranging from 20 nodes to 50 nodes. All nodes can transmit to and receive from any other node in the network because nodes' transmission ranges are not limited in this scenario (*i.e.*, $R_{max} \rightarrow \infty$). Energy overhead values of all $N_P = 2$ curves are always less than 1.00 %. On the other hand, energy overhead of single path routing stays in 5.58 % - 11.52 % band.

[4] We investigate line topologies by varying the number of sensor nodes and inter-node distance to confirm the effects of limiting number of routing paths observed in the small scale line topology in Figure 6.2 also holds for larger line topologies with different inter-node separation values.

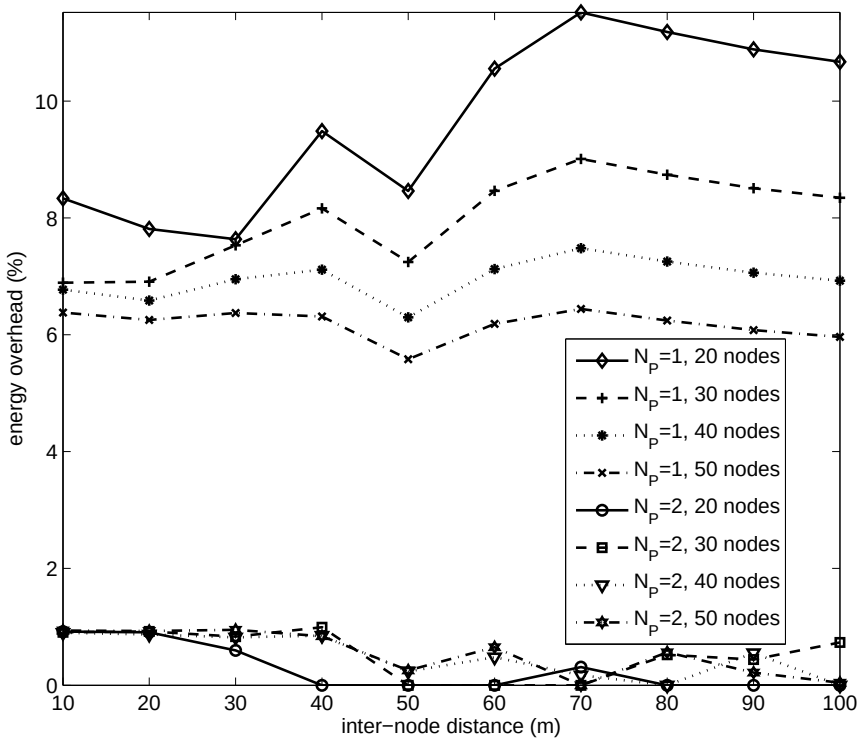

Figure 6.3: Percentage energy overhead with respect to the $N_P \rightarrow \infty$ case as a function of inter-node distance in (one dimensional) linear topology ($R_{max} \rightarrow \infty$)

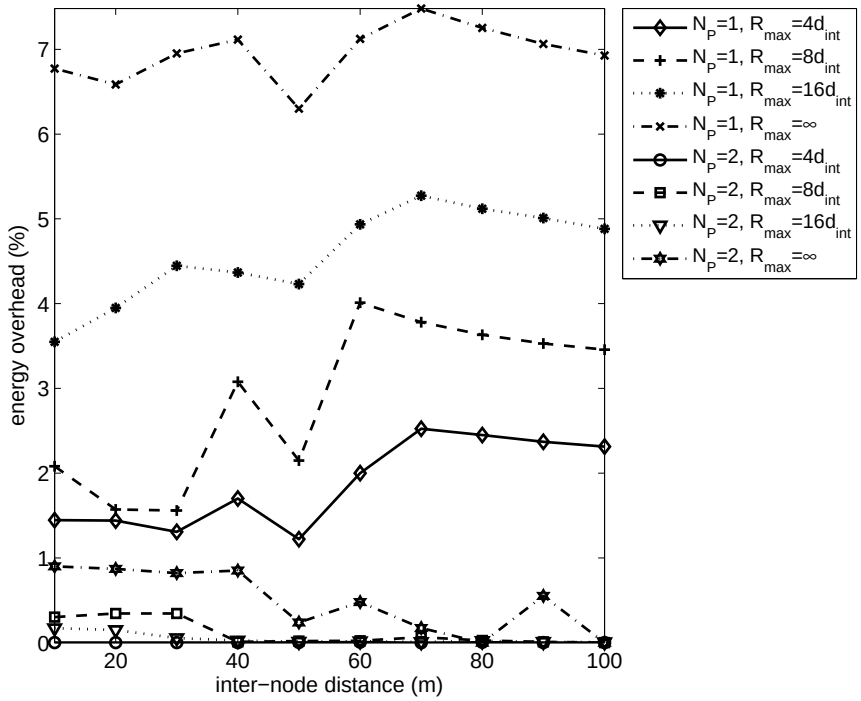

Figure 6.4: Percentage energy overhead with respect to the $N_P \rightarrow \infty$ case as a function of inter-node distance in (one dimensional) linear topology ($N = 40$).

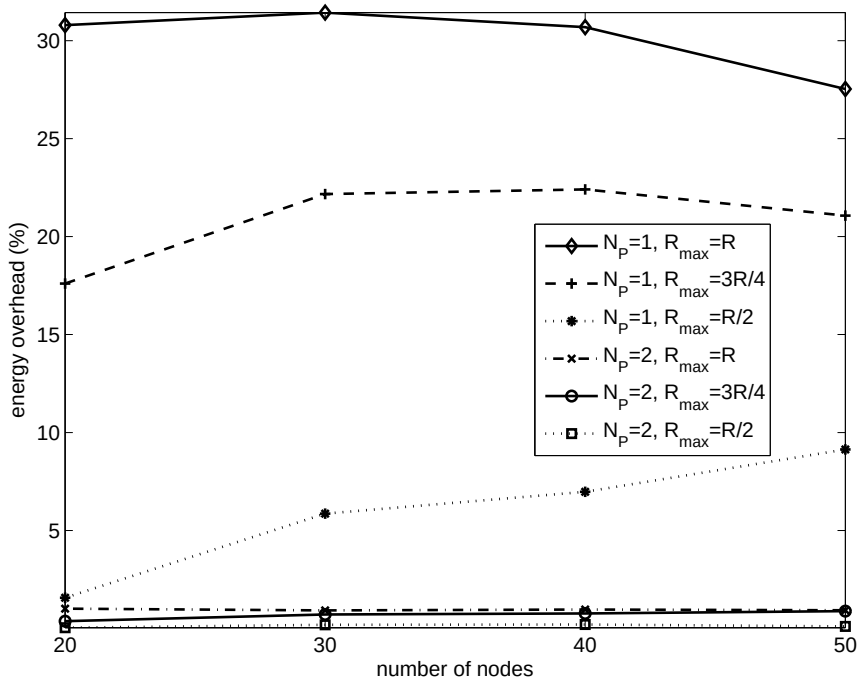

Figure 6.5: Percentage energy overhead with respect to the $N_P \rightarrow \infty$ case as a function of number of sensor nodes in (two dimensional) disc topology with $R = 200$ m

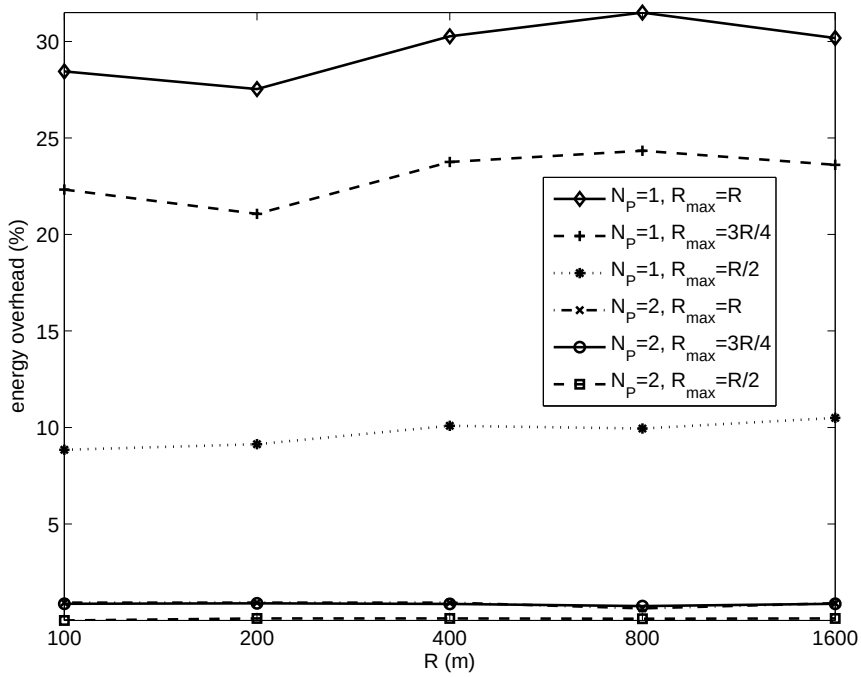

Figure 6.6: Percentage energy overhead with respect to the $N_P \rightarrow \infty$ case as a function of network radius in (two dimensional) disc topology with 50 sensor nodes

In Figure 6.4, the impact of R_{max} on energy overhead in linear topology with 40 nodes for $N_P = 1$ and $N_P = 2$ is presented. Energy overhead values of all $N_P = 2$ curves are less than 1.00 %. On the other hand, energy overhead of single path routing stays in 1.22 % - 7.48 % band. In single path routing ($N_P = 1$), energy overhead is lower for lower R_{max} because for lower R_{max}, $N_P \rightarrow \infty$ case is not as effective as it is with higher R_{max} (*i.e.*, the number of paths to choose from decreases as R_{max} decreases which narrows the options available for energy balancing).

In Figure 6.5 and Figure 6.6, energy overheads as functions of number of sensor nodes and disc radius are presented, respectively, for different maximum node transmission ranges (R_{max}) and N_P's[5]. In Figure 6.5, for $N_P = 1$ as R_{max} decreases the energy overhead also decreases. This is because for smaller R_{max} values even with $N_P \rightarrow \infty$, energy balancing is not as effective as in the case of $R_{max} \rightarrow \infty$ due to more limited routing options. In single path routing, energy overheads are in 31.43 % - 27.53 % and 1.57 % - 9.13 % bands for $R_{max} = R$ (200 m) and $R_{max} = R/2$ (100 m), respectively. The characteristics of energy overhead exhibit similar trends in Figure 6.6 (*i.e.*, energy overhead is dominated by R_{max}). Both in Figure 6.5 and Figure 6.6 energy overheads of all two-path routing scenarios are less than 1.00 %.

In all topologies explored in this chapter, our experiments revealed that energy overhead values for $N_P > 2$ (not presented in the figures) are always less than 1.00 %.

[5] In Figure 6.5 and Figure 6.6, we present results on two dimensional networks to generalize our results in one dimensional networks to two dimensional networks.

CHAPTER 7

ROUTE DIVERSITY FOR SECURITY IN WIRELESS SENSOR NETWORKS

Security is a critical issue for WSNs because nodes usually operate unattended and communication takes place in a broadcast medium. A common and successful technique used against eavesdropping is cryptographic encryption. With encryption, sensor data is scrambled using a key to make eavesdropped data unintelligible to anyone who does not possess the key. However there is always a possibility that a single technique may be flawed or cracked (*e.g.*, IEEE 802.11 WEP protocol). Moreover, in WSNs sensor nodes can be captured (*i.e.*, node capture attacks) and vital cryptography information such as keys can be extracted from them. If keys are captured, this would render encryption useless. Eavesdropping attacks are usually unnoticed (*i.e.*, it is challenging to detect a passive attack), thus, these attacks can succeed without encountering any active defense [52]. Layered approach to security and "defense-in-depth" strategy mandate that alternative or complementary techniques be used.

Security can be enhanced by using route diversity (*i.e.*, multi-path routing) which exploits multi-hop characteristics of WSNs by providing multiple paths between the source node and the base station to split data on these paths. Route diversity can be used as a standalone security countermeasure or in conjunction with encryption. While encryption is a good defense against attackers who are already eavesdropping wireless transmission, route diversity makes eavesdropping more difficult in the first place[1] [52]. By splitting data along different paths, an adversary has to capture portions routed through different paths to construct one node's data which requires more effort than extracting information in single path case. In other words, an adversary needs to spend more resources to collect data from the network if route diversity is

[1] While route diversity can also be useful against denial of service attacks [21], our main motivation in this chapter is to study it in the context of data confidentiality.

implemented.

Our goal in this chapter is to investigate the energy impact of route diversity countermeasures in WSNs. There are at least two types of energy overhead route diversity brings. First of these is due to the need to discover, establish, and maintain multiple routes instead of a single route between sensor nodes and the base station. In WSNs consisting of stationary nodes, this is a one-time operation for a substantial amount of time [95], hence, can be ignored. There is a second factor which makes the energy cost of route diversity more than the energy cost of single route paradigm. When data is split into multiple parts and forwarded via multiple routes, it is no longer possible to carry all data in energy-optimal paths. Some portion of the data should be transmitted towards the base station via less efficient paths. In fact, all routes can be sub-optimal from an energy efficiency perspective when compared to a single energy-efficient route. Broadly speaking, we can say that the second factor is more significant than the first one because it is applicable not only for a limited period of network operation but during the entire network lifetime. In other words, since WSNs generally exhibit stationary topology and connectivity behavior, route updates are infrequent. On the other hand, data transport by using multiple paths is a continuous operation spanning the entire network lifetime.

In this chapter we make the first attempt to carry out an analysis on energy cost of route diversity for security by building a framework using Linear Programming (LP). LP is a technique to solve the problem of maximizing or minimizing a linear function whose variables are required to satisfy a finite set of constraints that are expressed either as linear equalities or linear inequalities. In the context of WSNs, LP approach has previously been applied in many studies to model the unique characteristics of WSNs and to determine the optimal solutions to problems that are specific to WSNs (*e.g.*, [26–29, 77–79, 88–90]).

In this chapter, we consider a WSN where nodes have sensitive data to be conveyed to the base station. Data is spread out on multiple paths to mitigate both active and passive attacks. Data flows on these paths are optimized to minimize the overall energy consumption throughout the network. Furthermore, to avoid premature death of any node within the network energy dissipation is evenly balanced throughout the network, hence, the network lifetime is optimized. Within an LP framework, we model the energy dissipation characteristics of route diversity countermeasures against node capture (NCO), eavesdropping (EAO), and both node capture and eavesdropping (NCE) attacks. Using the developed LP models, we evaluate the

energy cost of these countermeasures by benchmarking against the energy requirements of unconstrained optimal case.

7.1 Model

Our main goal in this chapter is to investigate the minimum energy requirements for transferring data via multiple paths in WSNs purported as security enhancement strategies against node capture and eavesdropping attacks. In this section we formulate the system model with the objective function (minimization of energy dissipation) and three sets of problem constraints leading to route diversity for resistance against node capture and/or eavesdropping attacks. The threat model we use in this study is presented first in the following subsection.

7.1.1 Security Assumptions and Threat Model

In our model, we consider a WSN consisting of a base station and a number of sensor nodes distributed over an operation area. Sensor nodes generate data that must be sent to the base station possibly by using other nodes as relays. The goal of the attacker is to capture the sensitive data (*e.g.*, surveillance images) collected by some specific target node(s). The attacker aims at obtaining the complete data or a high portion of it[2]. He is not in a physically close location to the target sensor node(s). However he is in proximity to the WSN so that he can attempt to obtain the data either by passively eavesdropping the links and/or by actively capturing the relaying nodes. We assume the base station is physically well-protected. Attacker's physical location and which sensor nodes are specifically targeted by him are not known. We further assume that the attacker is not capable of compromising all of the nodes and can not listen to the whole network.

Protecting the confidentiality of the sensor data against such a threat is conventionally achieved by encryption. For WSNs, encryption can be implemented either in link layer (*i.e.*, hop-by-hop encryption) or in upper layers (*i.e.*, end-to-end encryption).

Link layer encryption when applied to WSNs has the following problem. Sensor nodes are deployed in an unattended environment and do not include strong tamper resistance hardware,

[2] If, under other assumptions, an attacker only needs to get a small portion of data, then route diversity could be a weakness as it spreads data throughout the network.

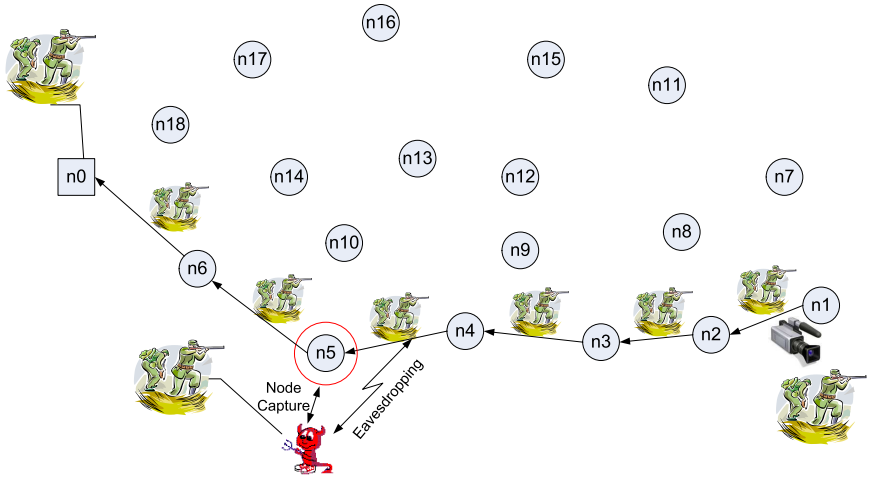

Figure 7.1: Attacker in single path routing scenario.

hence, they are vulnerable to node compromise [118]. Once nodes are compromised, vital cryptography information such as keys can be extracted from them. If nodes convey all their data by using a single path then the attacker can capture the complete data coming from a particular sensor node by compromising any single node used as a relay.

Data encryption in general has other implementation difficulties when applied to WSNs. Firstly, key establishment, an essential service for all cryptographic schemes, has scalability problems [118]. Secondly, sensor nodes have limited computation and energy resources, thus, especially public key encryption which could ease the key management problem may not always be a viable option.

Regardless of whether or which encryption approach is adopted, multi-path routing proves itself as an effective countermeasure [52]. As shown in Figure 7.1, an attacker can capture the complete data by eavesdropping on a single link or by compromising a single node when all data is routed through the same path. On the other hand, multi-path routing enables the flow of data from a sensor node to the base station through multiple paths between them (see Figure 7.2). Each sensor node spreads its data out to multiple paths so that the attacker needs to spend more resources to collect the data (*i.e.*, he has to listen to more links or he has to capture more nodes). Within this threat model, the security goal is to make it more difficult to obtain the sensitive data. We assume that attack costs are additive as in [52] (*i.e.*, capturing R nodes and eavesdropping K links are R times and K times more difficult than capturing a node and eavesdropping a link, respectively). We defer discussion of alternative attack cost models to Section 7.3.

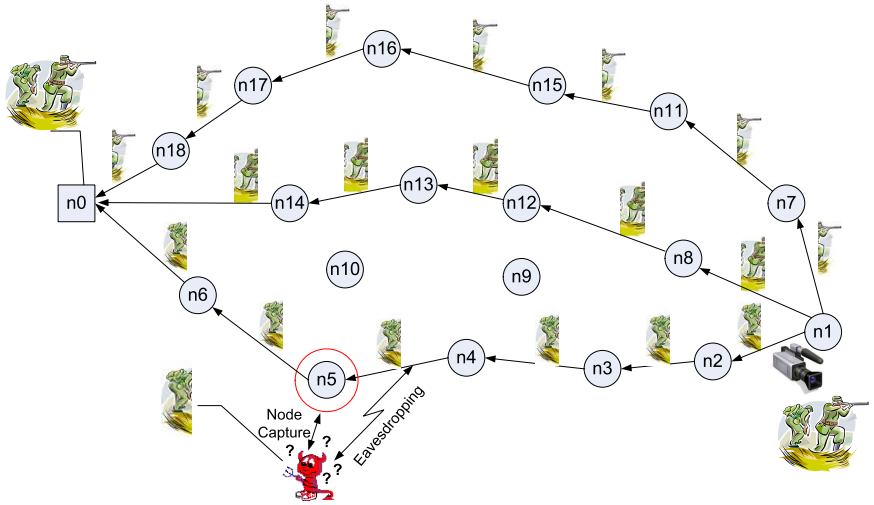

Figure 7.2: Attacker in multi-path routing scenario.

Based on this threat model, our goal is to assess the energy cost of different multi-path routing countermeasures. Previously, a number of studies investigated the energy cost of data encryption in WSNs [96, 119, 120].

7.1.2 System Model

In our framework, there is a single base station. We assume that time is organized into rounds. Each sensor node-i creates the same amount of data (s_i) at each round to be conveyed to the base station.

Data generated at each node terminates at the base station either by being transferred directly (single hop) or through other nodes acting as relays (multi-hop). The network topology is represented by a multiple edge directed graph $G = (V, A)$. V is the set of all nodes, including the base station as node-0. We also define set W which includes all nodes except node-0. $A = \{(k, i, j) : k \in W, i \in W, j \in V - i\}$ is the set of edges. Note that there are multiple directed edges between nodes.

We use energy model described in section 3.1.2 and the communication parameters are listed in 3.1. In this chapter, the amount of energy to transmit one bit of data is represented as $E^B_{tr,ij}$ and to receive one bit of data is represented as E^B_{rc}.

The definition of a set of edges in a graph representing a network, usually, is a subset of V^2 (flows are represented as f_{ij} instead of f^k_{ij}). We explain the reason of our uncommon edge

definition as follows. In our LP model, the flows originated at different source nodes should be identifiable at each relay node in order to be able to assert constraints working separately on each source's data. Hence, the set of edges used in our framework is a subset of V^3 (data generated at node-k flowing from node-i to node-j is represented as f_{ij}^k).

We define a node-limit variable (L_{node}) to limit the amount of data flowing through a node (*i.e.*, $\sum_{j \in W} f_{ji}^k \leq L_{node}$). We also define a link-limit variable (L_{link}) to limit the amount of data flow on a link (*i.e.*, $f_{ij}^k \leq L_{link}$). Both L_{node} and L_{link} are used to limit the amount of flows originated from different nodes separately (*e.g.*, for the flows between node-i and node-j the inequality $L_{link} \geq f_{ij}^k$ must be satisfied for each k independently). We assume link and node-limits are effective throughout the whole network. All system variables with their acronyms and descriptions are presented in Table 7.1.

7.1.3 The Base LP Model

In this subsection we present a novel LP model which forms the base for the rest of our formulations. In the subsequent subsections, extensions of the base model by adding other constraints are presented. The optimization problem for minimizing the maximum energy requirement of the sensor nodes (*battery*) is presented in Figure 7.3.

Equation 7.1 states that all flows are non-negative. Equation 7.2 is used to eliminate infinite loops – there cannot be a flow from a node to itself. Equation 7.3 states that all data originated at node-i is routed out to the rest of the network including the base station. Equation 7.4 states that when a node (node-i) is relaying data of another node (node-k) sum of all node-k's data flowing into node-i (either directly from node-k or via other relay nodes) equals to sum of all node-k's data flowing out of node-i (either directly to the base station or to other relay nodes). Note that Equation 7.4 must be satisfied at each node for all other nodes excluding the base station (in total there are $N - 2$ constraints to be satisfied at each node excluding node-0). Equation 7.5 is used to determine the number of packets required to convey the flow on edge (k, i, j). Equation 7.6 and 7.7 give the total energy dissipation of node-i for reception and transmission, respectively. Note that energy dissipation for packet overhead is also accounted for transmission and reception. Packet retransmissions due to packet errors are modeled with λ parameter which is related to the packet error rate (χ) by $\lambda = 1/(1 - \chi)$. Equation 7.8 and Equation 7.9 are used to compute the total amount of time spent for reception

Table 7.1: Terminology for LP Formulations

Variable	Description
N	Number of nodes
f_{ij}^{k}	Flow from node-i to node-j that carries data generated at node-k (bits)
$G = (V, A)$	Directed graph that represents network topology
V	Set of nodes, including the base station as node-0
W	Set of nodes, except the base station (node-0)
A	Set of edges
s_i	Amount of data generated at node-i (bits)
E_{rc}^{B}	Energy consumption for receiving one bit of data
$E_{rc,i}^{T}$	Total reception energy consumption of node-i
$E_{tr,ij}^{B}$	Energy consumption for transmitting one bit of data from node-i to node-j
$E_{tr,i}^{T}$	Total transmission energy consumption of node-i
P_{sl}	Power dissipation in sleep mode
$E_{sl,i}^{T}$	Total sleep energy consumption of node-i
e_i	Battery energy of each sensor node
B_P	Packet size (bits)
B_H	Overhead per packet (bits)
B_D	Payload per packet (bits)
λ	Average packet retransmission rate
χ	Packet error rate
ξ	Bandwidth (bits per second)
g_{ij}^{k}	Number of packets required to contain f_{ij}^{k}
t_{rn}	Duration of a round (s)
$t_{rc,i}$	Total reception time for node-i
$t_{tr,i}$	Total transmission time for node-i
$t_{sl,i}$	Total sleep time for node-i
$t_{if,i}$	Total interference time for node-i
I_{jl}^{i}	Interference function
γ	Ratio of interference range to transmission range
L_{link}	Link-limit value limiting the maximum amount of data that can flow over any link for data generated at each node
L_{node}	Node-limit value limiting the maximum amount of data that can pass through any node for data generated at each node

Minimize *battery*
Subject to:

$$f_{ij}^k \geq 0 \ \forall (k,i,j) \in A \tag{7.1}$$

$$f_{ij}^k = 0 \text{ if } \ i = j \ \forall (k,i,j) \in A \tag{7.2}$$

$$\sum_{j \in V} f_{ij}^k = s_i \text{ if } i = k, \ \forall i \in W, \ \forall k \in W \tag{7.3}$$

$$\sum_{j \in V} f_{ij}^k = \sum_{j \in W} f_{ji}^k \text{ if } i \neq k, \ \forall i \in W, \forall k \in W \tag{7.4}$$

$$g_{ij}^k \geq f_{ij}^k / B_D \ \forall (k,i,j) \in A \tag{7.5}$$

$$E_{rc,i}^T = \lambda E_{rc}^B \sum_{j \in W} \sum_{k \in W} (f_{ji}^k + g_{ji}^k B_H) \ \forall i \in W \tag{7.6}$$

$$E_{tr,i}^T = \lambda \sum_{j \in V} \sum_{k \in W} E_{tr,ij}^B (f_{ij}^k + g_{ij}^k B_H) \ \forall i \in W \tag{7.7}$$

$$\lambda (\sum_{j \in W} \sum_{k \in W} (f_{ji}^k + g_{ji}^k B_H))/\xi = t_{rc,i} \ \forall i \in W \tag{7.8}$$

$$\lambda (\sum_{j \in V} \sum_{k \in W} (f_{ij}^k + g_{ij}^k B_H))/\xi = t_{tr,i} \ \forall i \in W \tag{7.9}$$

$$t_{sl,i} = t_{rn} - (t_{rc,i} + t_{tr,i}) \ \forall i \in W \tag{7.10}$$

$$E_{sl,i}^T = P_{sl} t_{sl,i} \ \forall i \in W \tag{7.11}$$

$$E_{rc,i}^T + E_{tr,i}^T + E_{sl,i}^T \leq e_i \ \forall i \in W \tag{7.12}$$

$$e_i = battery \ \forall i \in W \tag{7.13}$$

$$\lambda (\sum_{j \in W} \sum_{l \in V} I_{jl}^i)/\xi = t_{if,i} \ \forall i \in V \tag{7.14}$$

$$(t_{rc,i} + t_{tr,i} + t_{if,i}) \leq t_{rn} \ \forall i \in V \tag{7.15}$$

Figure 7.3: The Base LP model.

and transmission at node-i, respectively, by the total amount of bits (payload and overhead) transmitted and received by node-i to the channel bandwidth (ξ). By subtracting the sum of reception time and transmission time from the duration of a round (t_{rn}), total sleep time ($t_{sl,i}$) is obtained (Equation 7.10) and the total sleep energy is obtained by multiplying sleep power (P_{sl}) and total sleep time (Equation 7.11). Equation 7.12 states that for all nodes except the base station energy consumed for transmission, reception and sleep is equal to or less than the energy stored in batteries. Energy dissipation for idle listening or overhearing in promiscuous mode is considered negligible. There are many intelligently designed MAC protocols for WSNs that avoid the energy waste in these modes [121]. We assume such a MAC layer is used.

Equation 7.13 is used to assign equal energy to each sensor node. Since the objective is to minimize *battery*, our problem is the minimization of the maximum battery requirement of the nodes in the network by finding the values of f_{ij}^k's that satisfy the constraints and minimize the energy consumption. To minimize the maximum energy dissipation, all nodes dissipate their energies in a balanced fashion in this model.

To take channel bandwidth limitations into consideration in a broadcast medium, we need to determine interfering flows for each node. Total interference time for node-i ($t_{if,i}$) due to the transmissions of other nodes are computed by using Equation 7.14. Interference function (I_{jl}^i) is presented in Equation 7.16. If node-i is in the interference region of the transmission from node-j to node-l, then the value of interference function for node-i (I_{jl}^i) takes the value of total amount of bits carried from node-j to node-l, otherwise it is zero. Generally speaking, interference range is equal to or greater than transmission range (*i.e.*, $\gamma \geq 1$). This means depending on the value of γ, node-j's transmission to node-l can interfere with node-i even if the distance between node-j and node-l is less than the distance between node-j and node-i.

$$I_{jl}^i = \begin{cases} \sum_{k \in W}(f_{jl}^k + g_{jl}^k B_H) & \text{if } \gamma d(j,l) \geq d(j,i) \\ & \forall j \in W, \forall l \in V \\ 0 & \text{otherwise} \end{cases} \tag{7.16}$$

Equation 7.15 states that the total time for transmission, reception, and interference at node-i cannot be larger than the duration of a round. In other words, for all nodes including the base station the aggregate rate of incoming flows, outgoing flows, and interfering flows is upper

bounded by the channel bandwidth. This constraint is a modified version of the sufficient condition given in [99].

7.1.4 Mitigating Node Capture Only (NCO) Attacks

In this subsection, we extend the base LP model described above by introducing an additional constraint that limits the amount of flow passing through each node so that Node Capture Only (NCO) attacks are mitigated. When deployed in hostile environments, sensor nodes are vulnerable to node capture attacks. If captured node is used as a relay then data coming from other nodes can be obtained. Even if the data is encrypted, an adversary can use cryptographic keys in the captured node to decrypt it back if encryption is implemented in the link layer.

If all sensor nodes only relay at most a fraction of the data of other nodes, then without capturing multiple nodes it is not possible to construct data pertaining to other sensor nodes. To model such a limitation for each node we introduce a node limit variable (L_{node}), which is a fraction of s_i. The amount of data originated at any node-i and flowing into any relay node cannot be more than L_{node}). We call $L_{node} = s_i/R$ limit as R-degree-node-resilience (*i.e.*, at least R nodes need to be captured to construct any sensor node's data). Note that the total amount of data relayed by any sensor node can be larger than L_{node} because the limit works for separate sources' data independently. In this model, only the node capture attacks are considered. The optimization problem for minimizing the maximum energy requirement of the sensor nodes (*battery*) is constructed by augmenting Equation 7.17 to the base LP framework presented in Figure 7.3.

$$\sum_{j \in W} f_{ji}^{k} \leq L_{node} \text{ if } i \neq k, \forall i \in W, \forall k \in W \tag{7.17}$$

Equation 7.17 states that the sum of all flows that carry node-k's data flowing into node-i is upper limited by L_{node}. We point out the impossibility to put a node-limit constraint on a node for its own data. In other words, an attacker who captures a node always obtains all data generated by that node. However, in most cases such an attack would not be meaningful because if the attacker is capable of capturing the node he can also collect the node's sensitive data directly from the environment (*e.g.*, with his own camera).

Figure 7.4 is used to illustrate the dynamics of node limit (L_{node}) on a toy example, a small scale network consisting of one base station (node-0) and three sensor nodes. In this example

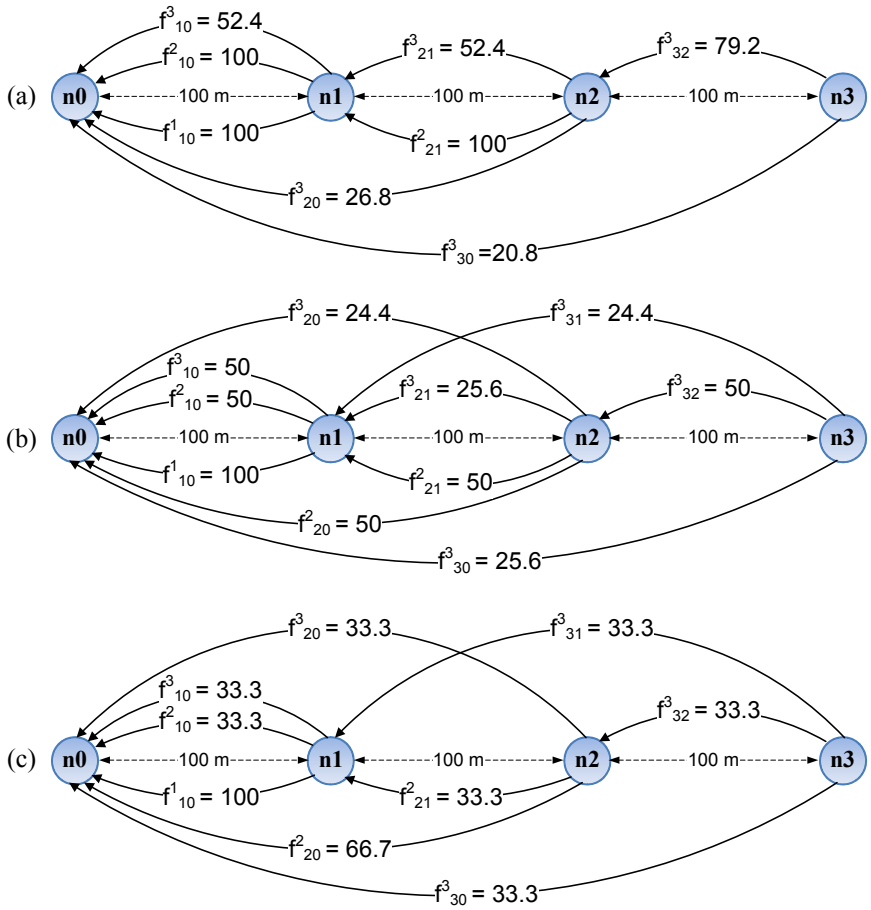

Figure 7.4: Illustration of network flows for NCO

a small scale network and a large inter-node distance is chosen to better illustrate the concept. Table 7.2 presents the parameters used in the example (N is 4 and inter-node distance is 100 m). Figure 7.4(a) presents optimal flows for $L_{node} \rightarrow \infty$ (no L_{node} limit). Note that flows are presented in percent form (*e.g.*, $f_{21}^3 = 52.4$ means that 52.4 % of node-3's data flows from node-2 to node-1). If no restrictions apply on L_{node}, an attacker can capture all of node-2's data and more than half of node-3's data by capturing node-1 only.

Figure 7.4(b) and Figure 7.4(c) present respectively the flows minimizing maximum energy dissipation for $L_{node} = s_i/2$ and $L_{node} = s_i/3$. Observe that none of the nodes carry more than 50.0 % of the data generated at any other node for $L_{node} = s_i/2$. Similarly, nodes carry at most 33.3% of any other node's data for $L_{node} = s_i/3$. As node-limit decreases the deviation from the optimal flows increases, which leads to the increased energy requirements (*i.e.*, normalized energy requirements for $L_{node} \rightarrow \infty$, $L_{node} = s_i/2$, and $L_{node} = s_i/3$ are 1.00, 1.41, and 1.73, respectively).

Table 7.2: List of parameters used in the toy example and analysis

Parameter	Values
N	4-125
Deployment Scenarios	Linear (1-D) Equidistant, Square (2-D) Random
Inter-node Distance (1-D)	100 m, 10 m
Network Size (2-D)	100 m × 100 m − 3200 m × 3200 m
L_{link}	$s_i/2 - s_i/6, s_i/8, s_i/16, s_i/32$
L_{node}	$s_i/2 - s_i/6, s_i/8, s_i/16, s_i/32$
Data generated at each round s_i	1.0 Mbits
Duration of a round t_{rn}	10 minutes
Sleep power P_{sl}	0.00 Watts
Channel bandwidth ξ	1.0 Mbps
Packet error rate χ	0.001
Packet size B_P	800 bits
Packet overhead B_H	80 bits
Packet payload B_D	720 bits
Interference range ratio γ	1.7

7.1.5 Mitigating Eavesdropping Only (EAO) Attacks

In this subsection, we extend the base LP model described earlier by introducing an additional constraint that limits the amount of flow passing over each link so that EAvesdropping Only (EAO) attacks are mitigated. If encryption is infeasible or broken, or there is a leak of the key, an eavesdropper may try to capture as much sensitive data as possible by passively listening to the links without any active attempt to capture nodes. Since sensor nodes relay data of other nodes, an eavesdropper can capture data coming from targeted sensor node(s) by listening only to a single link if necessary countermeasures are not in effect. If each node splits its data into multiple parts routed through different paths so that on any link only a fraction of the data is sent, then an effective countermeasure against an eavesdropper is implemented.

In our second LP formulation we restrict the amount of data on each link by setting a limit (L_{link}). This formulation is most appropriate for the case when the cost of eavesdropping multiple links is equal to the sum of listening to each of these links. If attack costs are not additive, then alternative link limit formulations should be preferred (see Section 7.3).

Our objective is again minimization of required energy for each node (*battery*), now subject

to the constraints presented in Figure 7.3 plus Equation 7.18.

$$f_{ij}^{k} \leq L_{link} \; \forall (k, i, j) \in A \tag{7.18}$$

Equation 7.18 states that the maximum amount of node-k's data that can be sent over a link between any two nodes is limited by L_{link}. Thus, even if an eavesdropper can listen to a link between two nodes, he can capture at most a limited (L_{link}) amount of node-k's data. Unlike NCO case, since eavesdropping attacks can be conducted without being physically close to the attacked node, we put L_{link} constraint for all links without any exception (links of source nodes are also constrained).

The link limit value (L_{link}) is a fraction of s_i. If $L_{link} = s_i/K$, then by eavesdropping to a single link an adversary can capture at most $1/K$'th of any source node's data. Therefore, we call $L_{link} = s_i/K$ limit as K-degree-link-resilience due to the fact that any adversary needs to capture data from at least K links to construct any node's complete data.

The toy example used earlier for NCO is also useful to grasp the basic idea behind EAO framework (see Figure 7.5). The optimal flow distribution without any L_{link} or L_{node} restriction is presented again in Figure 7.5(a). If L_{link} constraint is not in effect ($L_{link} \rightarrow \infty$), an attacker can capture all of node-2's data and more than half of node-3's data by listening only the link between node-1 and node-2.

None of the links carry more than 50.0 % and 33.3 % of the data generated at any node for $L_{link} = s_i/2$ and $L_{link} = s_i/3$, respectively. As link-limit becomes more stringent, maximum energy requirement increases (*e.g.*, normalized energy requirements for $L_{link} \rightarrow \infty$, $L_{link} = s_i/2$, and $L_{link} = s_i/3$ are 1.00, 1.88, and 3.96, respectively). In contrast to NCO, the link constraint in EAO applies to the data generated and directly transmitted to the base station, leading to reverse flows (data flowing away from the base station). This is the main reason for the significantly higher energy overhead of EAO in comparison to the overhead of NCO.

7.1.6 Mitigating Node Capture and Eavesdropping (NCE) Attacks

As our third and last model, we extend the base LP model by considering a stronger notion of security in the presence of both Node Capture and Eavesdropping (NCE) attacks. We call $L_{node} = s_i/R, L_{link} = s_i/K$ limit as $R \times K$-degree-joint-resilience (an attacker can get at most L_{link} amount of s_i by listening to a link and can get at most L_{node} amount of s_i of other

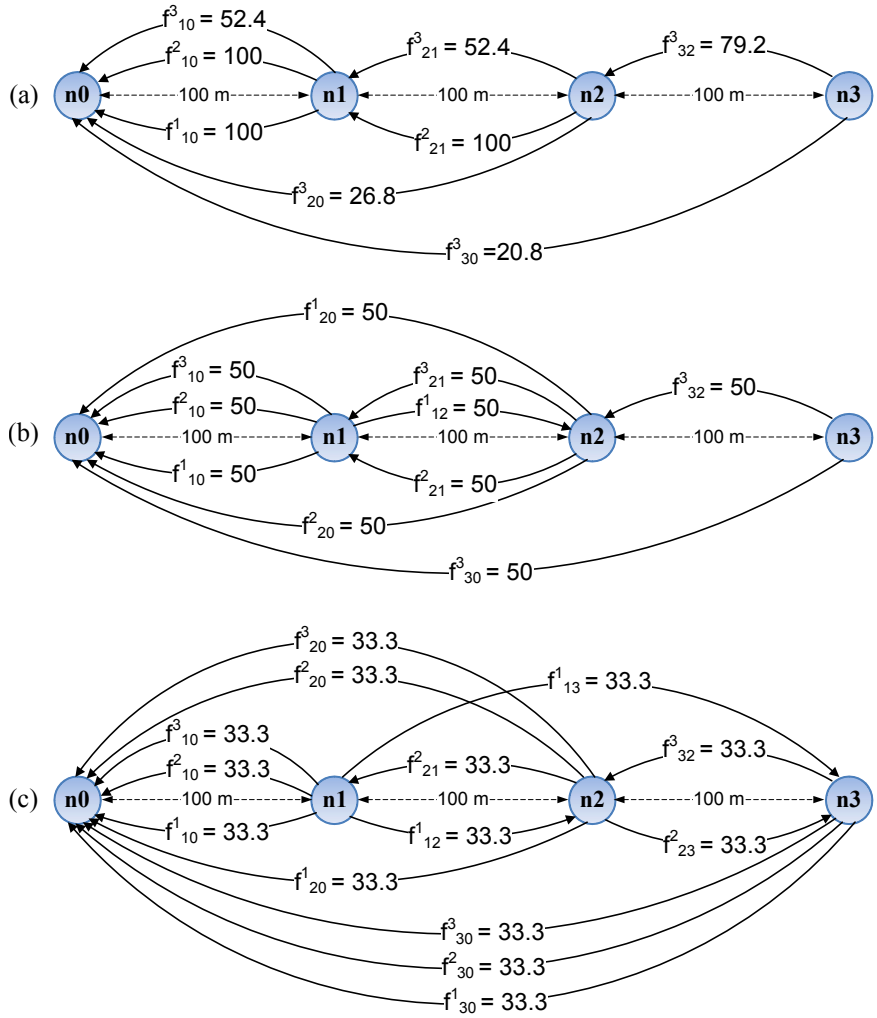

Figure 7.5: Illustration of network flows for EAO

sensor nodes by capturing any one of the sensor nodes). Our objective in NCE case is also to minimize the maximum energy requirement of sensor nodes (*battery*), now subject to the constraints presented in Figure 7.3 combined with both Equation 7.17 and Equation 7.18. To restrict the amount of data on any link and on any node Equation 7.17 and Equation 7.18 are used together.

In our toy example if both L_{node} and L_{link} constraints are in effect, all flows and maximum energy requirements are identical with EAO case shown in Figure 7.5 due to the fact that link-limit completely dominates energy-efficient optimal routing behavior in this topology. Observe that the optimal flows for EAO satisfying $L_{link} = s_i/2$ and $L_{link} = s_i/3$ also satisfy NCO constraint for $L_{node} = s_i/2$ and $L_{node} = s_i/3$, respectively, however, the optimal flows for NCO do not satisfy EAO constraints. As we will discuss in the next section, in larger networks we observe that link-limit still dominates but complete domination does not happen and the impact of node-limit is also present.

7.2 Analysis

We use GAMS [109] for numerical analysis of LP models. Two topologies are used in our analysis: 1-Dimensional (1-D) line topology and 2-Dimensional (2-D) square topology. In line topology, nodes are placed equidistantly over a line with an inter-node distance of 10 m and the base station is placed at one end of the line. In square topology, nodes are randomly placed within a square shaped area of predetermined size and the base station is located at the center. The number of deployed nodes is varied between 10 and 125 nodes. We investigate the square topology further by varying the network size. Different values of L_{link} and L_{node} are used to characterize different route diversity behaviors. The parameters used in the analysis are presented in Table 7.2. The communication parameters are same as the ones used in [29]. For random deployment scenarios, each problem is solved for 100 times (*i.e.*, 100 independent random topologies are generated for the same parameter set) and the results are averaged. Rather than presenting the absolute energy dissipation values, we opt to present the percentage increase in the energy dissipation when compared to a network with the same parameters (network size, number of nodes, topology, etc.) without any restriction defined by link-limit or node-limit ($L_{link} \rightarrow \infty, L_{node} \rightarrow \infty$).

7.2.1 Analysis of NCO

In this subsection we analyze the energy overhead of route diversity countermeasures against NCO attacks as a function of the number of nodes, network size, and node-limit in line and square network topologies. In Figure 7.6 the variation of energy overhead for different L_{node} values is presented as a function of number of nodes in a line network. Limiting the amount of data that can be relayed through a node (limit imposed by L_{node}) does not put any restriction on nodes sending all their data directly to the base station. Therefore, the increase in energy overhead is mainly due to the nodes which are farther away from the base station and convey their data by multi-hop relaying. Hence for all L_{node} values the energy overhead increases as network size gets larger.

Figure 7.7 shows the variation of energy overhead as a function of number of nodes for different L_{node} values in a 400 m $\times$ 400 m square network. As number of nodes increases (leading to higher node density), the energy overhead decreases (except the increase from 10-node to 20-node when $L_{node} \leq s_i/4$).

The optimal operation of low density networks when L_{node} constraint applies results in sending most of the data directly to the base station. As the node density increases, the adverse effects of L_{node} constraint on energy dissipation reduces because of the availability of increased number of links in the network. To put it into more precise terms, L_{node} constraint prevents the network from taking full advantage of the optimal energy efficient routes, thus, it leads to extra energy dissipation. When more energy balancing routes are available in higher density networks, satisfying the L_{node} constraint with less energy overhead becomes possible.

In the 20-node network, the trade-off between sending data directly to the base station and sending data cooperatively towards the base station to reduce the effects of L_{node} constraint is more visible. It is a better choice to send data cooperatively towards the base station for $L_{node} = s_i/2$ and $L_{node} = s_i/3$. In these cases, the energy overhead of the 20-node networks is less than the energy overhead of the 10-node network. On the other hand, for $L_{node} = s_i/4$, $L_{node} = s_i/5$, and $L_{node} = s_i/6$, optimum energy balancing could be achieved by sending most of the data directly to the base station since the number of available routes becomes insufficient to address more strict L_{node} constraints. As a result, for these cases the energy overheads in the 20-node network are higher than the energy overheads in the 10-node network.

Figure 7.8 presents the energy overhead as a function of L_{node} in a 400 m $\times$ 400 m square network with 75, 100, and 125 nodes. Energy overhead decreases as the level of resilience decreases. As the number of nodes increases while the network area is kept constant the energy overhead decreases because the network with higher number of nodes have more routing options to balance the energy dissipation and mitigate the effects of L_{node} constraint. Hence, higher node density networks organize flows more efficiently (*i.e.*, with less energy overhead) against the pressure induced by node-limit constraints. For example, the energy overheads of 75-, 100-, and 125-node deployments for 2-degree-node-resilience (*i.e.*, $L_{node} = s_i/2$) are 0.78 %, 0.45 %, and 0.34 %, respectively, whereas for 32-degree-node-resilience, the energy overheads of 75-, 100-, and 125-node deployments are 120.00 %, 87.00 %, and 54.61 %, respectively.

Figure 7.9 shows the energy overhead as a function of the network area for different L_{node} values. The number of nodes in the network is kept constant as 100. For $L_{node} \rightarrow \infty$, the percentage of data sent directly to the base station decreases as the area increases. Since in NCO case there is no limit on the amount of data directly sent to the base station (*i.e.*, L_{link} constraint is not in effect), for smaller network areas deviation from the optimal flows is less. For example, in the 100 m $\times$ 100 m network L_{node} restriction does not apply for approximately 70 % of the generated data which is sent directly to the base station. For larger networks, the percentage of direct transmissions to the base station is lower and more deviation from the optimal flows (because of L_{node} limit) results in higher energy overheads.

7.2.2 Analysis of EAO

In this subsection we explore energy overhead characteristics of route diversity countermeasures against EAO attacks. In Figure 7.10 the variation of energy overhead for different L_{link} values is presented as a function of number of nodes in a line network. The maximum energy overhead is observed for $L_{link} = s_i/6$ in the 10-node network. As the network size increases the energy overhead decreases. The reason for this behavior is that for small networks a significant portion of the nodes (*e.g.*, 4 out of 10) send their data directly to the base station if $L_{link} \rightarrow \infty$, hence, limiting the amount of data on these links results in large deviations from the optimal flows. On the other hand, in larger networks only a few nodes do not use multi-hop forwarding when $L_{link} \rightarrow \infty$ (*e.g.*, 2 out of 100), therefore, flows created due to the

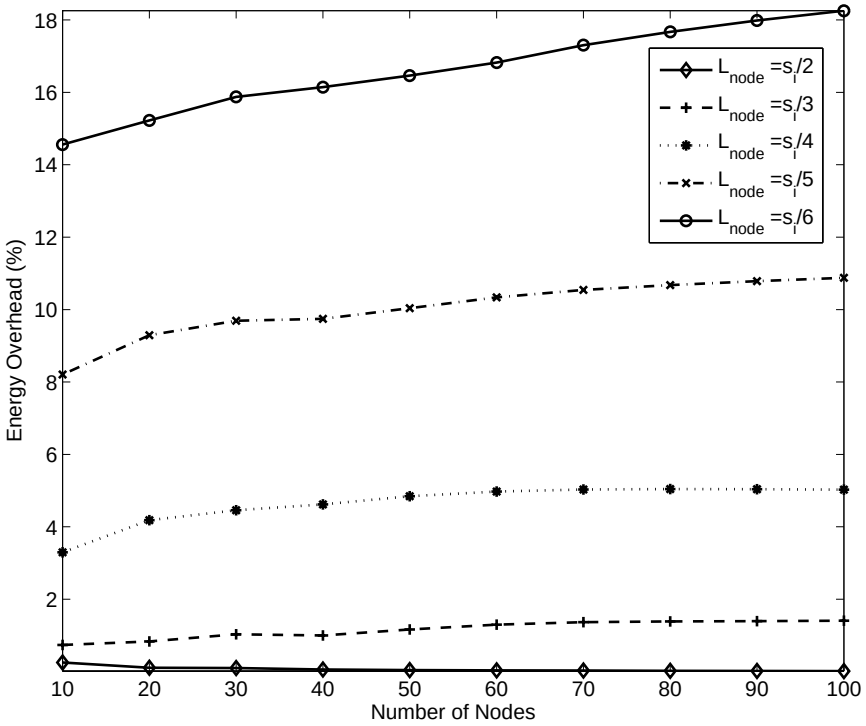

Figure 7.6: Relative energy overhead with respect to the unrestricted case ($L_{node} \rightarrow \infty$) as a function of number of nodes and L_{node} in a line network with inter-node distance of 10 m.

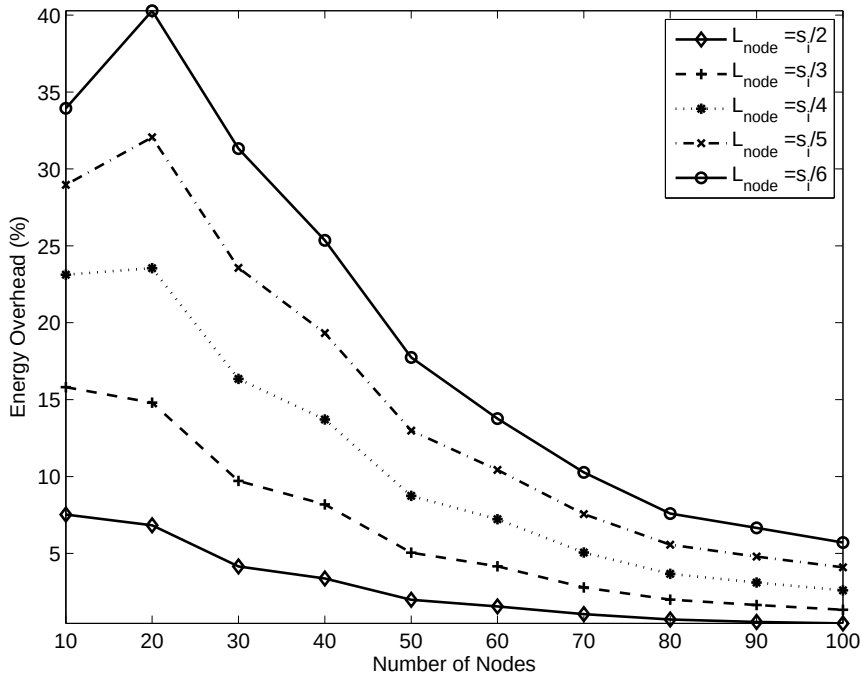

Figure 7.7: Relative energy overhead with respect to the unrestricted case ($L_{node} \rightarrow \infty$) as a function of number of nodes and L_{node} in a 400 m $\times$ 400 m square topology.

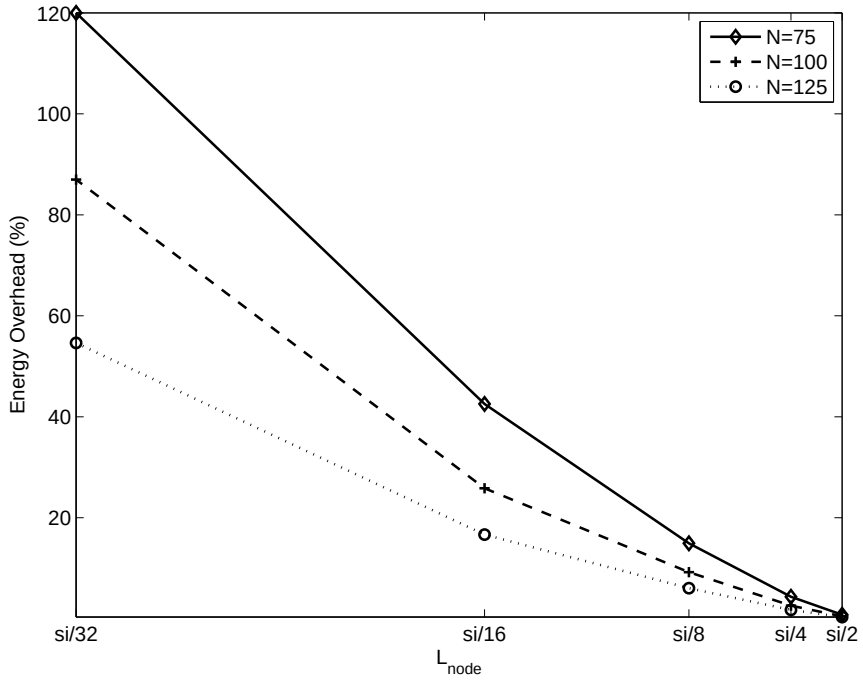

Figure 7.8: Relative energy overhead with respect to the unrestricted case ($L_{node} \rightarrow \infty$) as a function L_{node} in a 400 m $\times$ 400 m square topology.

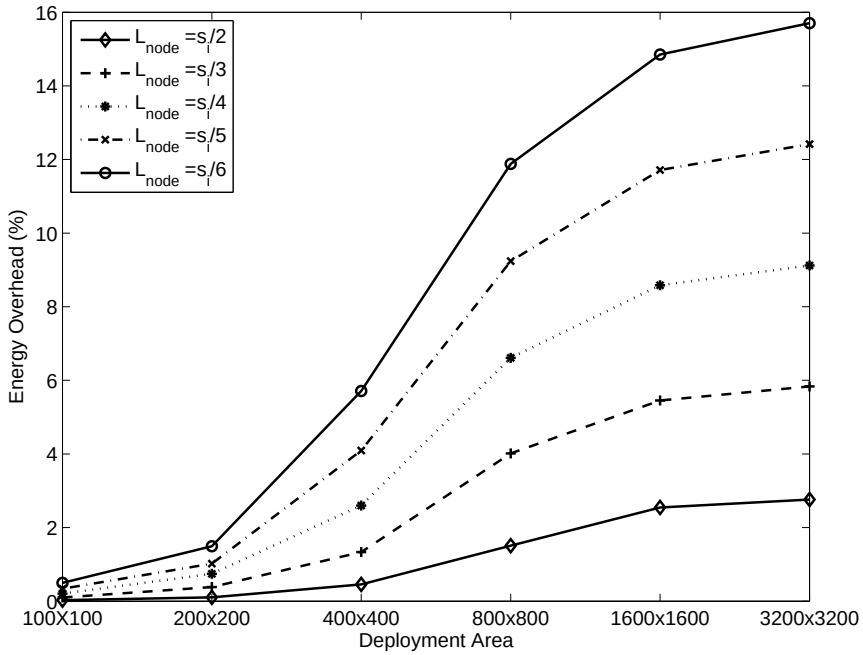

Figure 7.9: Relative energy overhead with respect to the unrestricted case ($L_{node} \rightarrow \infty$) as a function of network area and L_{node} in a square topology with 100 nodes deployment.

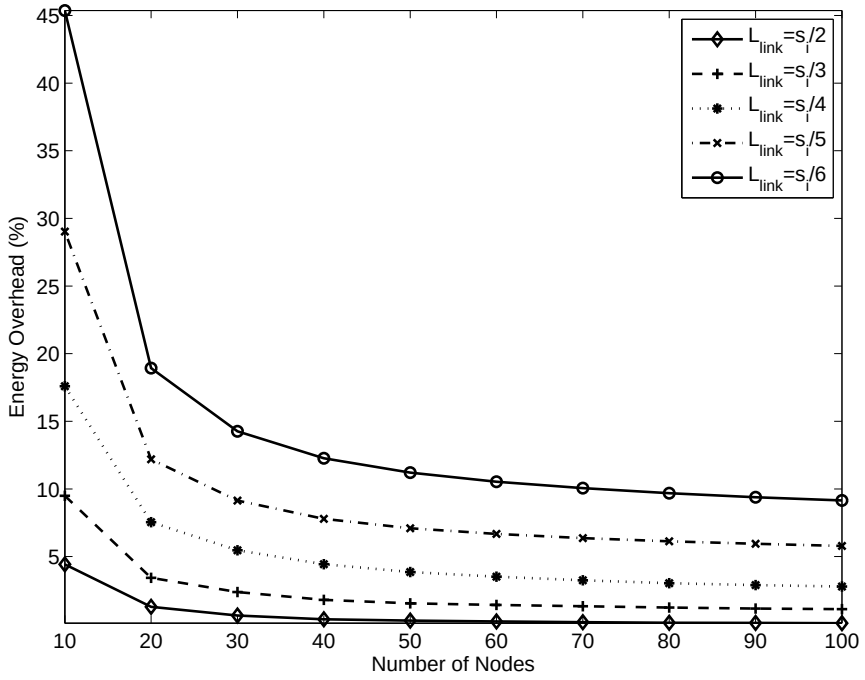

Figure 7.10: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$) as a function of number of nodes and L_{link} in a line network with inter-node distance of 10 m.

link-limit constraints exhibit relatively less deviation from unrestricted cases.

As L_{link} decreases (the degree of resilience increases), energy overhead of the network increases. By decreasing the maximum allowed amount of data on the links, the network is forced to use less energy efficient (suboptimal) routes. For example, if a node is very close to the base station it sends all of its data directly to the base station if $L_{link} \rightarrow \infty$, however, with $L_{link} = s_i/K$, the node needs to split its data at least into K parts and forward them towards the base station by using multiple routes. As K increases the deviation from the optimal flows also increases.

For the maximum network lifetime all nodes should cooperate and dissipate their energies in a balanced fashion (*i.e.*, the energy burden of data forwarding is evenly shared among all nodes and none of the nodes dies prematurely). However, when we put limitations on the links, we observe that some sensor nodes are left with residual energy. For example, in the 10-node network with $L_{link} = s_i/4$, node-1 (the node closest to the base station) does not spend 17.50 % of its initial energy while all other sensor nodes use all their energies. When $L_{link} = s_i/5$ and $L_{link} = s_i/6$ in the 10-node network, not only first node but also the second node is also left with some residual energy (*i.e.*, for $L_{link} = s_i/5$ residual energy is 30.42 %

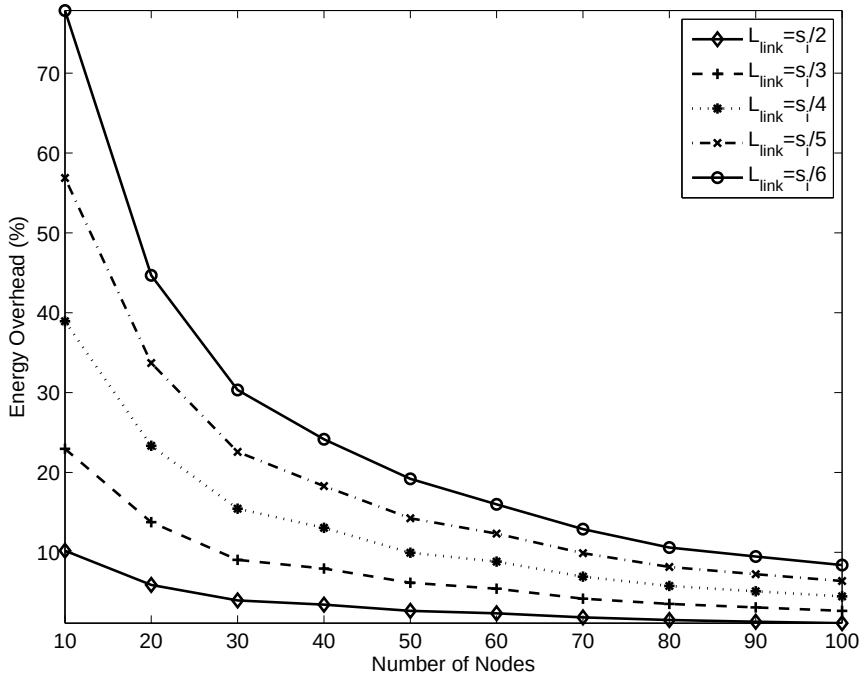

Figure 7.11: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$) as a function of number of nodes and L_{link} in a 400 m $\times$ 400 m square topology.

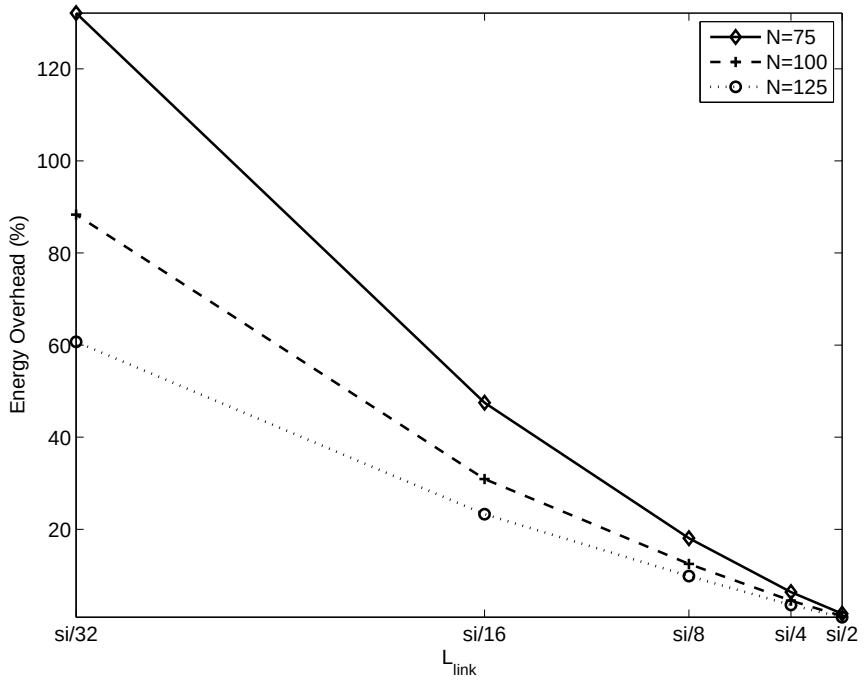

Figure 7.12: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$) as a function of L_{link} in a 400 m $\times$ 400 m square topology.

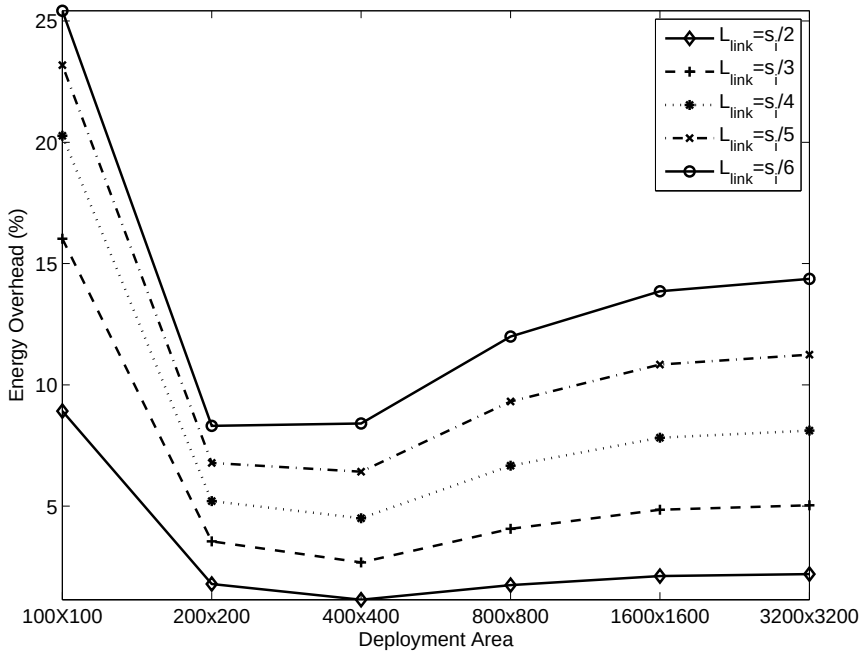

Figure 7.13: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$) as a function of network area and L_{link} in a square topology with 100 nodes deployment.

in the first node and 3.21 % in the second node). In the 100-node network with $L_{link} = s_i/5$ residual energy is 58.38 % in the first node and 10.05 % in the second node and all other nodes completely deplete their energies. Nevertheless, nodes cooperate to share the energy burden of multi-hop routing evenly except possibly one or two nodes. These nodes are left with residual energies because using up the residual part does not help to decrease the energy dissipation of other nodes.

We present the energy overhead as a function of the number of nodes in a square network of size 400 m $\times$ 400 m in Figure 7.11 for various L_{link} values. Despite the fact that network size is kept constant while the number of nodes increases (unlike linear deployment), the network characteristics in 2-D case is similar to 1-D case which is in contrast to NCO case. As the number of nodes increases the energy overhead decreases. The reason for such a behavior is that denser networks are more tolerant to link-limit constraint as there are more routes available. Lower L_{link} values result in higher energy overhead.

In Figure 7.12, the energy overhead as a function of link-limit in a 400 m $\times$ 400 m square network with 75, 100, and 125 nodes is presented. The energy cost of route diversity increases substantially with the increasing degree of link resilience.

In Figure 7.13, the energy overhead is presented as a function of the network area for different L_{link} values. The number of nodes in the network is kept constant as 100. The curves in Figure 7.13 are bimodal (*i.e.*, not monotonically increasing or decreasing) due to two mechanisms working in opposite directions. When link-limit is not in effect, most of the nodes in the 100 m × 100 m network send most of their data directly to the base station. The amount of data that can be sent through the direct link between a sensor node and the base station is limited by the value of L_{link}. On the other hand, in the 3200 m × 3200 m network nodes (especially ones farther away from the base station) tend to send most of their data to a limited number of relay nodes to be conveyed to the base station. Hence, for both small and large sized networks enforcing link-limit results in larger deviations from the optimal energy balancing flows. In the 200 m × 200 m and 400 m × 400 m networks the energy overheads are lower than the other networks due to the reduced impact of two mechanisms dominating the opposite ends of the network size.

7.2.3 Analysis of NCE

In this subsection we investigate the energy overhead trends for countermeasures against NCE attacks. In Figure 7.14, the variation of energy overheads for different L_{link} and L_{node} values as a function of number of nodes in a line network are presented. The similarity between Figure 7.14 and Figure 7.10 and the dissimilarity between Figure 7.14 and Figure 7.6 show that link-limit is dominant over node-limit. Yet, node-limit also manifests its impact by increasing the energy overhead. For example, the overheads for EAO and NCE in the 10-node line network are 45.36 % ($L_{link} = s_i/6$) and 50.92 % ($L_{node} = L_{link} = s_i/6$), respectively, and the difference is due to the impact of node-limit. NCE has a bi-modal energy overhead behavior. The energy overhead decreases as the network gets larger up to a certain level thereafter the energy overhead starts increasing slowly. Such behavior is due to the combination of monotonic behaviors of EAO and NCO (*i.e.*, EAO overhead is monotonically decreasing while NCO overhead is monotonically increasing as the number of nodes increases).

Figure 7.15 shows the variation of energy overhead as a function of number of nodes for different L_{link} and L_{node} values in a 400 m × 400 m square network. NCE characteristic in a 400 m × 400 m square network is also dominated by link-limit as manifested by the similarity between Figure 7.11 and Figure 7.15. Yet, the impact of node-limit can also be observed.

In Figure 7.16, the variation of energy overhead as a function of link-limit and node-limit in a 400 m × 400 m network with 75-, 100-, and 125-node deployments is presented. The energy overheads of the networks decreases as less stringent joint resilience constraints are enforced.

Figure 7.17 shows the energy overhead as a function of the network area for different L_{link} and L_{node} values. The number of nodes in the network is kept constant as 100. The energy overhead has the maximum for the 100 m × 100 m network with $L_{link} = L_{node} = s_i/6$ and has the minimum for the 400 m × 400 m network with $L_{link} = L_{node} = s_i/2$.

In the LP models, the link-limit constraint (Equation 7.18) puts a limitation on all links including direct links to the base station. On the other hand, the node-limit constraint (Equation 7.17) does not apply to the base station hence the flows generated and sent directly to the base station are not restricted. Limiting the amount of data sent to the base station without using a relay has a more significant impact on energy overhead than limiting the amount of data passing over relay nodes. Because in the former case alternate multi-hop paths lead to higher energy inefficiency than using alternate relay nodes in the later case (as an example for the inefficiency, consider the reverse flows from node-1 in Figure 7.5(c)). This the main reason of the difference between the energy overheads of EAO and NCO (*i.e.*, EAO energy overheads are, generally, larger than NCO overheads for the same network settings). By the same token, NCE energy overhead characteristics is dominated by EAO with minor impact from NCO.

7.3 Discussion

Generally speaking, WSNs are assumed to be consisting of stationary sensor nodes and unlike mobile ad hoc networks topology changes are not frequent. Thus, topology discovery and route creation are one-time operations and for substantial amount of time (rounds/epochs) these functions are not repeated [96]. If the epoch durations (network reorganization cycle time) are long enough then the energy costs of these operations constitute a small fraction (less than 1 %) of the total network energy dissipation [96]. On the other hand, in highly dynamic topologies network organization energy costs can be as high as 60 % of the total energy dissipation [101]. Considering that routing overhead can be neglected in stationary WSNs without leading to significant underestimation of total energy dissipation, our characterization

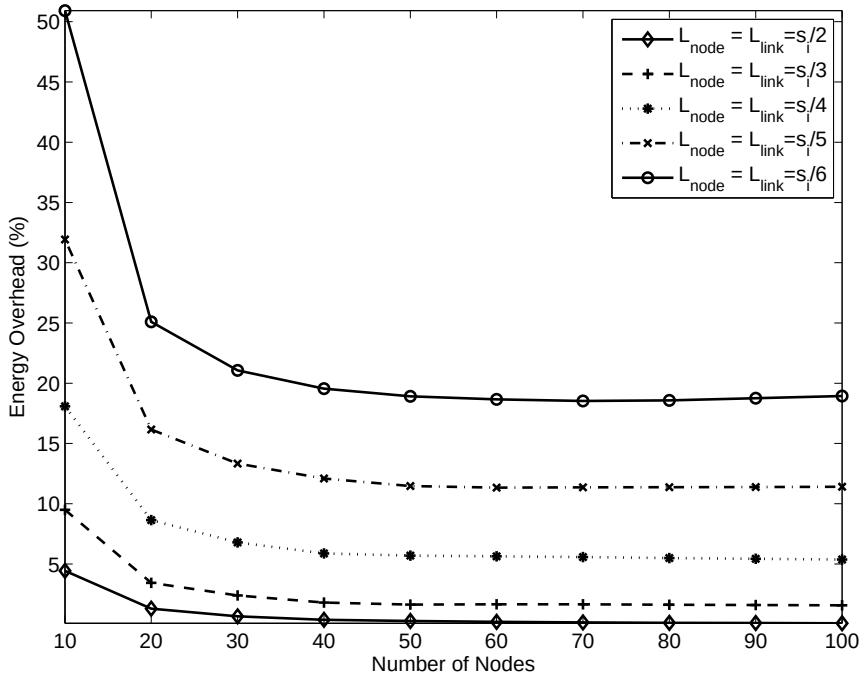

Figure 7.14: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$ and $L_{node} \rightarrow \infty$) as a function of number of nodes, L_{link}, and L_{node} in a line network with inter-node distance of 10 m.

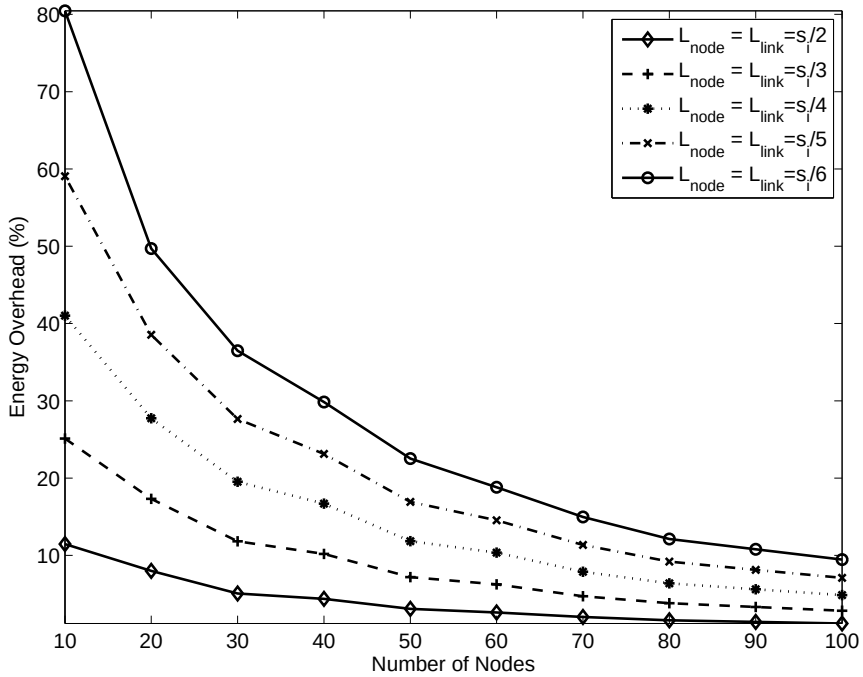

Figure 7.15: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$ and $L_{node} \rightarrow \infty$) as a function of number of nodes, L_{link}, and L_{node} in a 400 m $\times$ 400 m square topology.

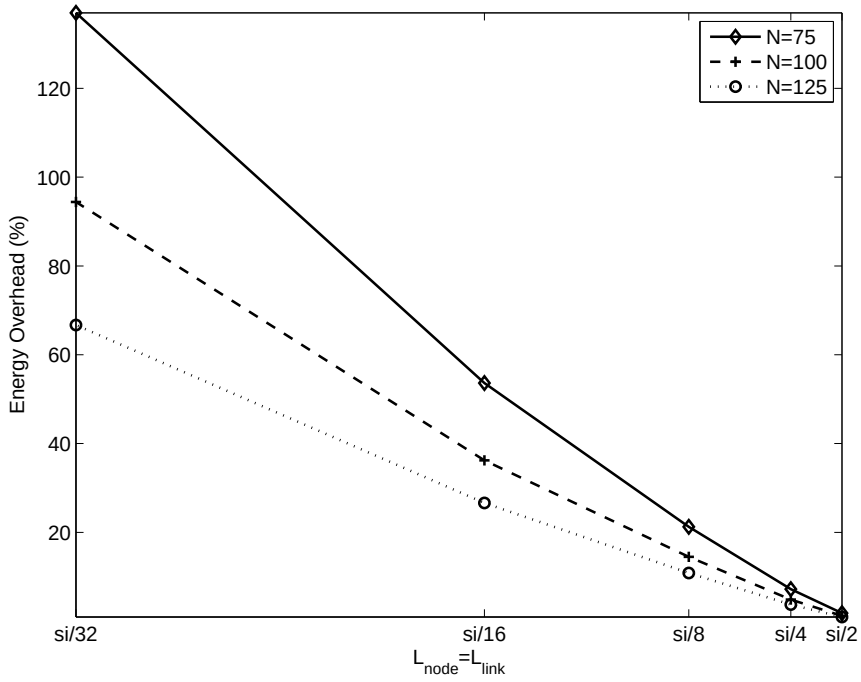

Figure 7.16: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$ and $L_{node} \rightarrow \infty$) as a function L_{link} and L_{node} in a 400 m $\times$ 400 m square topology.

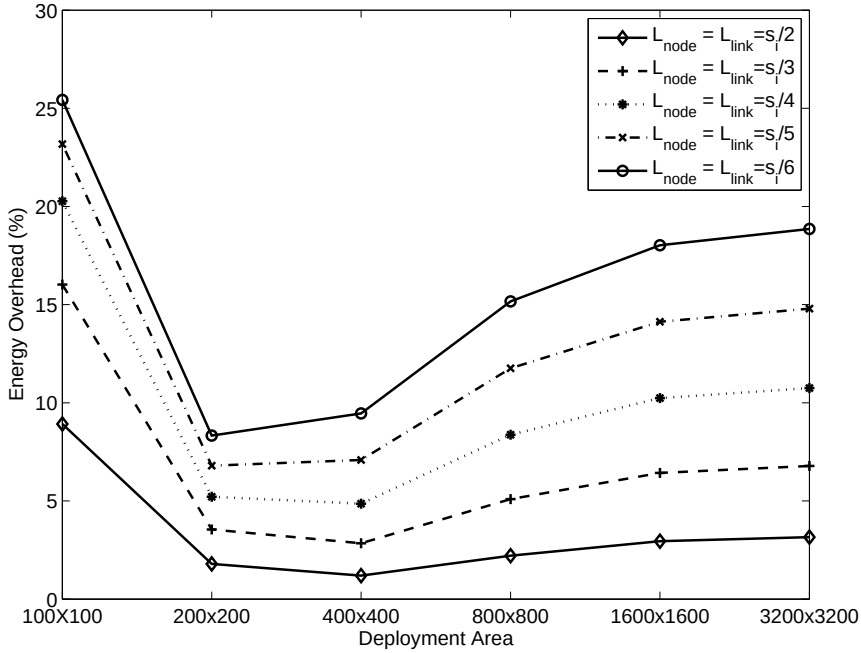

Figure 7.17: Relative energy overhead with respect to the unrestricted case ($L_{link} \rightarrow \infty$ and $L_{node} \rightarrow \infty$) as a function of network area, L_{link}, and L_{node} in a square topology with 100 nodes deployment.

of energy overhead due to route diversity is based on realistic assumptions.

In our model route diversity is enforced by restricting the amount of flow on each link and the amount of flow passing through each node. Thus, we indirectly cause the network to split the data generated at each node and route the data through multiple paths. In other words, by selecting $L_{link} = s_i/K$ the network is forced to route the data of each node via at least K different links. However, routing data through more links can be a better alternative for energy balancing and certain nodes can split their data into more parts and route it through more links. A similar argument holds for node-limit ($L_{node} = s_i/R$). More concisely, we can say that L_{link} and L_{node} determines the "minimum" level of route diversity.

The rationale behind limiting the amount of data flow on each link through the use of L_{link} is to prevent an eavesdropper from capturing the data of any sensor node by listening to a single link. An eavesdropper has to capture data from at least K links to assemble any source node's complete data in a network with $L_{link} = s_i/K$. However the cost of eavesdropping is not always additive. We distinguish the following three cases:

1. Additive cost: Eavesdropping K links is K times more difficult than eavesdropping a single link. This is a realistic assumption for instance when communication on each link is encrypted with a separate key which can be broken with a time-consuming brute force attack. As another example, consider a single commodity hardware which can listen to only one link at a time.

2. Sub-additive cost: Eavesdropping K links is less than K times difficult than eavesdropping a single link. A typical example of it can be seen when there is no encryption and eavesdropping all of the incoming links of a particular node is achieved by using a single receiver. An interesting special case is the one in which eavesdropper can listen to all incoming links of a node at the same time except the incoming links of the base station. From key management point of view, nodes can secure their direct links with the base station more easily because this can be achieved using encryption with a single secret key shared with the base station. The constraint for NCO (Equation 7.17) can be used to model this kind of limit since for our purposes eavesdropping all incoming links of a node (except the base station) is equivalent to a capture attack conducted on that node. To cover other cases, as the link constraint Equation 7.18 should be replaced by Equation 7.19. The set E_i includes all the nodes whose communication to node-i

can be eavesdropped simultaneously.

$$\sum_{j \in E_i} f_{ji}^{k} \leq L_{link} \ \forall i \in V, \forall k \in W \tag{7.19}$$

3. Super-additive cost: Eavesdropping K links is more than K times difficult than eavesdropping a single link (*e.g.*, when the coordination for reconstructing the data crossing several links poses additional costs [52]).

On the other hand, costs of node capture attacks can be assumed as additive if such attacks are performed by physically compromising the nodes and remote attacks exploiting software vulnerabilities of sensor nodes are not viable. As long as $L_{node} = s_i/R$ is in effect, an adversary has to capture at least R nodes to construct data of other sensor nodes which can be assumed as R times more difficult than capturing a single node.

CHAPTER 8

CONCLUSION AND FUTURE WORK

The most precious resource in wireless sensor networks is energy. Data reduction by the use of local processing is the key for energy efficiency. One factor that dictates the maximum amount of reduction is the spatial granularity of the environment phenomena. Understanding environmental conditions is the primary purpose of applying wireless sensor network technology but a lack of data of adequate granularity hinders the accurate understanding.

In Chapter 4, we presented a linear programming framework that models the operation of a wireless sensor network satisfying the spatial granularity requirements of measurements in the most energy-efficient way. Using our model we quantitatively evaluate the effect of spatial granularity and redundancy of measurements on minimum energy requirements. The framework we have presented and the results of our analysis makes easier to understand the tradeoffs between spatial granularity and energy requirements in wireless sensor networks.

In Chapter 5, through a novel LP framework we investigate the energy overhead arising due to physical attacks in WSNs. We model optimal network behavior to balance the energy dissipation throughout the network so that incapacitation of any node (due to a physical attack) does not result in an overwhelming energy cost for the whole network. We consider two attack scenarios: (i) uniform attack, where all network regions are equally affected by the attack (*i.e.*, the same number of nodes are dead in each region) and (ii) non-uniform attack, where only the nodes in a single region are affected. Our results show that the energy cost of uniform attacks can efficiently be shared by the network (*e.g.*, if one sixth of the nodes are incapacitated by an attack the energy overhead is one fifth of the energy required to accomplish the task without any node failure). However, for non-uniform attacks the energy overhead depends on the targeted region. Physical attacks targeting the regions closer to the base station lead to more

energy overhead when compared to the attacks targeting other regions provided that there are multiple nodes left operational in the targeted region after the attack. On the other hand, if there is only one node left in the targeted region then the attacks on the furthest regions from the base station result in the highest energy overhead due to the lack of available redundancy, which limits the network wide cooperation to only the relaying operation for the attacked region.

Note that in Chapter 5 we characterize the energy overhead due to node failures induced by physical attacks. However, similar energy overhead characteristics are observed if the nodes fail with the same spatial pattern due to natural factors. Nevertheless, nature cannot target the nodes which lead to the most damage in terms of the energy overhead for the remaining nodes (unlike the case of intelligent attackers).

In our models, nodes are either active or failed for the entire network operation period (*i.e.*, we do not investigate the case where nodes operate some time and then fail). Since it is not possible to know the exact timing of an attack, a reasonable strategy is to allocate enough resources to mitigate the effects of the worst-case scenario in which the attack is conducted just after network starts operating and damages nodes permanently.

The LP framework we present in Chapter 5 can easily be tailored to accommodate other aspects of physical attacks in WSNs. For instance, a natural extension of our analysis would be to examine energy dissipation characteristics when the base station is not located at the center of the monitored region.

In Chapter 6, we presented an MIP framework to investigate the energy dissipation of WSNs as a function of number of routing paths. We explored various WSN scenarios in both one dimensional and two dimensional network topologies by sampling the parameter space through the developed model.

Our analysis revealed that single path routing may lead to more than 30.00 % energy overhead due to lack of sufficient number of energy balancing routes. On the other hand, multi-path routing with only two paths results in near-optimal values with at most 1.00 % energy overhead. Thus, our main conclusion is that use of more than two paths for energy balancing in multi-path routing for WSNs does not bring any significant benefit from an energy efficiency perspective. The MIP framework we presented in Chapter 6 can easily be tailored to

accommodate other aspects of multi-path routing in WSNs. Nevertheless, the concept of an end-to-end path should exist for our results to be relevant.

In Chapter 7, we investigated the energy overhead characteristics of route diversity countermeasures for resilience against node capture only (NCO), eavesdropping only (EAO), and node capture and eavesdropping (NCE) attacks. We developed an LP framework that is capable of jointly modeling energy dissipation and route diversity in WSNs. Through the developed framework we conducted a comprehensive analysis by exploring the design space in a systematic fashion. Characterization of the energy overhead of route diversity countermeasures in WSNs is important for understanding the feasibility of utilization of such security techniques. A brief summary of our results are itemized as follows:

1. Energy overhead of route diversity increases with the level of resilience. If the level of resilience is low, then the energy overhead is also low. On the other hand, for high degrees of resilience energy overhead can be prohibitive (*i.e.*, higher than the energy dissipation of unconstrained networks).

2. For maximal network lifetime all nodes should cooperate and use their energies in a way that prevents premature death of any node. Route diversity imposed on the network results in energy imbalances among the nodes (*i.e.*, some nodes do not deplete their initial energies). However the extent of the imbalance is limited. We observed that only a few nodes exhibit such a behavior.

3. In line networks: (a) for EAO case the energy overhead decreases as the number of nodes in the network increases due to the fact that in larger networks nodes can adapt to link-limit with less deviation from the optimal energy-efficient routes; (b) for NCO case direct transmission to the base station eliminates data splitting to satisfy L_{node} constraint, however, in larger networks direct transmission to the base station results in higher energy inefficiency, therefore, the energy overhead increases with the number of nodes in the network; (c) for NCE case similar to EAO case as the network size increases the energy overhead decreases.

4. In square networks: (a) EAO and NCE cases reach their minimum energy overhead values for specific network area values where the composite impact of direct transmission and multi-hop relaying behaviors (arising due to the effects of L_{link} and L_{node}) reaches

a relative minimum; (b) for NCO case energy overhead is a monotonically increasing function of the network area because there is less deviation from the optimal flows in smaller size networks where direct transmission to the base station is the dominant mode of operation not affected by node-limit; (c) for all three countermeasures as the number of nodes increases while the area is kept constant energy overhead decreases because when there are more nodes in the network, there are more links and paths in the network which enable the network to satisfy the constraints with less energy overhead.

5. NCE energy overhead behavior is mainly shaped by link-limit, yet, the impact of node-limit is also observable as a minor factor. It is interesting to observe that mitigation of a passive attack (*i.e.*, eavesdropping) yields more energy overhead than mitigation of an active attack (*i.e.*, node-capture) under the assumption of additive attack costs.

Unlike the previous studies we do not propose a new network protocol for routing and security problems that we studied in this thesis. Instead, we analyze the routing and security problems from energy efficiency perspective within a general framework and without going into the details of specific protocols or algorithms. The results we report in this thesis represent optimistic performance bounds because we do not include any protocol specific control message exchange in our analysis. Yet, our results reveal the benchmarks that can be used to evaluate and compare similar routing and security problems in various settings. In this sense, we contribute to the literature by presenting a comprehensive high level analysis of the several routing and security problems which captures the essence of both security and networking perspectives.

In summary, in this thesis, we investigate the energy dissipation characteristics of several routing and security strategies in WSNs in terms of redundancy elimination, multi-path routing, physical attacks and enhancing data confidentiality against passive and active adversaries. We develop novel mathematical programming frameworks and present a comprehensive high level analysis of these routing and security strategies. The mathematical programming frameworks presented in this thesis are efficient from computational point of view and can be used with minor modifications for future analysis of different routing and security problems in WSNs.

8.1 Future Work

Routing and security problems continue to evolve in WSNs. Therefore, the research areas for modeling and analysis of these problems must also evolve. We consider the following problems in WSNs are as future works.

While cooperating, sensor nodes relay data of other sensor nodes through the base station. In many cases, one sensor node can be used by many sensor nodes in the network as a relay. This leads to special class of security threat after a node capture attack since one node capture can result in privacy and integrity of many sensor nodes which use captured node as a relay to be compromised. As a future study, we will investigate energy efficient minimization of effects of a node capture attack by modeling the node capture problem from the point of affected sensor nodes.

It is straightforward to extend LP framework presented in chapter 7 as a future study to investigate the energy dissipation characteristics of redundant multi-path routing. As a simple example, consider the case when $(X-1)$ redundant copies of the data[1] are also transmitted towards the base station over multiple paths. If our analysis presented in chapter 7 indicates that the overhead percentage due to multi-path routing (X-path routing) is $k\%$, then the overhead of multi-path routing combined with redundant data transfer will be $(X\times k)\%+((X-1)\times 100)\%$ (the overhead of multi-path routing is multiplied by X since it is defined as a ratio of the cost for unconstrained data flow which is X times more due to carrying redundant copies).

Symmetric key encryption necessitates usage of same symmetric key between communicating pairs. Therefore, key establishment is required before communication. A trivial solution for key establishment is using same key in the entire network and an extreme solution is assigning a unique symmetric key between each pair of nodes. First solution can reveal the shared key after a single node compromise and encryption becomes useless in the entire network. Since WSNs include high number of sensor nodes, second solution doesn't scale well. If n sensor nodes are deployed, each sensor node is required to store $n-1$ symmetric keys and $n(n-1)/2$ keys are needed in the entire network. In [122], a random-key predistribution scheme was proposed for the solution of key establishment. In the random-key predistribution scheme, before deployment of WSN, each sensor node is loaded with a subset of symmetric

[1] Alternatively, an (X,Y) secret sharing scheme can be implemented so that an attacker cannot learn anything about the message even if he obtains up to $Y-1$ shares out of X [38, 39].

keys from a large pool of symmetric keys. After deployment, if two sensor nodes want to communicate, they search their stored symmetric keys. If they have a common key, sensor nodes can communicate securely. The random-key predistribution scheme doesn't guarantee that every pair of sensor nodes communicates but if the key sharing probability is sufficiently high, a securely connected network can be obtained and compromise of one node affects small portion of the WSN.

As a future work, we will consider a WSN where sensor nodes communicate sensitive data encrypted against eavesdropping attacks using symmetric key encryption. Previous studies focused on secure communication between sensor nodes using symmetric key encryption and connectivity between sensor nodes. While enhancing security against eavesdropping, lifetime requirements of WSNs should also be considered. We will develop a mathematical programming framework to find out maximum lifetime using symmetric key encryption with random-key predistribution scheme.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.

[2] J. Sarangapani. *Wireless Ad-hoc Sensor Networks - Protocols, Performance and Control*. CRC Press, 2007.

[3] G. Vellidis, M. Tucker, C. Perry, C. Kvien, and C. Bednarz. A real-time wireless smart sensor array for scheduling irrigation. *Computers and Electronics in Agriculture*, 61(1):44–50, 2008.

[4] K. P. Ferentinos, T. A. Tsiligiridis, and K. G. Arvanitis. Energy optimization of wireless sensor networks for environmental measurements. In *Proceedings of the International Conference on Computational Intelligence for Measurment Systems and Applicatons (CIMSA)*, pages 250–255, 2005.

[5] S. Singh, M. Woo, and C. S. Raghavendra. Topology optimization in wireless sensor networks for precision agriculture applications. In *Proceedings of the International Conference on Sensor Technologies and Applications (SensorComm).*, pages 526–530. IEEE, 2007.

[6] E. Hyytiä and J. Virtamo. On optimality of single-path routes in massively dense wireless multi-hop networks. In *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*, pages 28–35. ACM, 2007.

[7] M. Tariquea, K.E. Tepeb, S. Adibic, and S. Erfanib. Survey of multipath routing protocols for mobile ad hoc networks. *Journal of Network and Computer Applications*, 32:1125–1143, 2009.

[8] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. TAG: A tiny aggregation service for ad-hoc sensor networks. *ACM SIGOPS Operating Systems Review*, 36:131–146, 2002.

[9] L. Wang and Y. Xiao. A survey of energy-efficient scheduling mechanisms in sensor networks. *Mobile Networks and Applications*, 11(5):723–740, 2006.

[10] H. Chan and A. Perrig. Designing secure sensor networks. *IEEE Wireless Communications*, 11:38–43, 2004.

[11] A. Perrig, J. Stankovic, and D. Wagner. Security in wireless sensor networks. *Communications of the ACM*, 47(6):53–57, 2004.

[12] X. Wang, W. Gu, S. Challeppan, K. Schoseck, and D. Xuan. Lifetime optimization of sensor networks under physical attacks. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 5, pages 3295–3301, 2005.

[13] W. Gu, X. Wang, S. Chellappan, D. Xuan, and T. Lai. Defending against search-based physical attacks in sensor networks. In *Proceedings of the IEEE International Conference on Mobile AdHoc and Sensor Systems (MASS)*, pages 520–527, 2005.

[14] Saul I. Gass. *Linear programming methods and applications fifth edition*. McGraw-Hill, 1985.

[15] S. Singh, M.Woo, and C. S. Raghavendra. Power-aware routing in mobile ad hoc networks. In *Proceedings of the 4th Annual IEEE/ACM International Conference Mobile Computing and Networking*, volume 5, pages 181–190, 1998.

[16] I. Stojmenovic and X. Lin. Power-aware localized routing in wireless networks. *IEEE Transactions Parallel Distributed Systems*, 12(11):1122–1133, 2001.

[17] M. Radi, B. Dezfouli, K. A. Bakar, and M. Lee. Multipath routing in wireless sensor networks: Survey and research challenges. *Sensors*, 12(1):650–685, 2012.

[18] E. Stavrou and A. Pitsillides. A survey on secure multipath routing protocols in WSNs. *Computer Networks*, 54:2215–2238, 2010.

[19] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin. Highly-resilient, energy-efficient multipath routing in wireless sensor networks. *ACM Mobile Computing and Communications Review*, 5:10–24, 2001.

[20] B. Littlewood and L. Strigini. Redundancy and diversity in security. In *Proceedings of the European Symposium on Research Computer Security (ESORICS), LNCS 3193*, pages 423–438, 2004.

[21] K. Bicakci and B. Tavli. Denial-of-service attacks and countermeasures in IEEE 802.11 wireless networks. *Computer Standards & Interfaces*, 31:931–941, 2009.

[22] S. Li, X. Ma, X. Wang, and M. Tan. Energy-efficient multipath routing in wireless sensor network considering wireless interference. *Journal of Control Theory and Applications*, 9:127–132, 2011.

[23] B. Yahya and J. Ben-Othman. REER: Robust and energy efficient multipath routing protocol for wireless sensor networks. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1–7, 2009.

[24] Z. Wang, E. Bulut, and B. K. Szymanski. Energy efficient collision aware multipath routing for wireless sensor networks. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pages 1–5, 2009.

[25] Y. M. Lu and V. W. S. Wong. An energy-efficient multipath routing protocol for wireless sensor networks. *International Journal of Communication Systems*, 20(7):747–766, 2007.

[26] J. H. Chang and L. Tassiulas. Maximum lifetime routing in wireless sensor networks. *IEEE/ACM Transaction on Networking*, 12:609–619, 2004.

[27] M. Bhardwaj, T. Garnett, and A. Chandrakasan. Upper bounds on the lifetime of sensor networks. In *Proceedings of the IEEE International Conference on Communications (ICC)*, 2001.

[28] M. Bhardwaj and A. Chandrakasan. Bounding the lifetime of sensor networks via optimal role assignments. In *IEEE International Conference on Computer Communications (INFOCOM)*, 2002.

[29] Z. Cheng, M. Perillo, and W.B. Heinzelman. General network lifetime and cost models for evaluating sensor network deployment strategies. *IEEE Transactions on Mobile Computing*, 7:484–497, 2008.

[30] L. Zhang, Z. Zhao, Y. Shu, L. Wang, and O. WW Yang. Load balancing of multipath source routing in ad hoc networks. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pages 3197–3201, 2002.

[31] X. Huang and Y. Fang. Multiconstrained QoS multipath routing in wireless sensor networks. *Wireless Networks*, 14(4):465–478, 2008.

[32] A. Tsirigos and Z. J. Haas. Multipath routing in mobile ad hoc networks or how to route in the presence of frequent topology changes. In *Proceedings of the IEEE Military Communications Conference(MILCOM)*, volume 2, pages 878–883, 2001.

[33] E. Felemban, C. Lee, and E. Ekici. MMSPEED: multipath multi-speed protocol for qos guarantee of reliability and. timeliness in wireless sensor networks. *IEEE Transactions on Mobile Computing*, 5(6):738–754, 2006.

[34] P.P.C. Lee, V. Misra, and D. Rubenstein. Distributed algorithms for secure multipath routing. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, volume 3, pages 1952–1963, 2005.

[35] L. Chen and J. Leneutre. On multipath routing in multihop wireless networks: security, performance, and their tradeoff. *EURASIP Journal on Wireless Communications and Networking*, Article ID 946493, 2009.

[36] R. Mavropodi, P. Kotzanikolaou, and C. Douligeris. SecMR: secure multipath routing protocol for ad hoc networks. *Ad Hoc Networks*, 5:87–99, 2007.

[37] J.B. Othman and L. Mokdad. Enhancing data security in ad hoc networks based on multipath routing. *Journal of Parallel and Distributed Computing*, 70:309–316, 2010.

[38] W. Lou and Y. Kwon. H-SPREAD: a hybrid multipath scheme for secure and reliable data collection in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 55:1320–1330, 2006.

[39] W. Lou, W. Liu, and Y. Fang. SPREAD: Enhancing data confidentiality in mobile ad hoc networks. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, volume 4, pages 2404–2413, 2004.

[40] M. C. Vuran, Özgür B. Akan, and I. F. Akyildiz. Spatio-temporal correlation: theory and applications for wireless sensor networks. *Computer Networks*, 45(3):245–259, 2004.

[41] Rajagopal Iyengar, Koushik Kar, and Suman Banerjee. Low-coordination topologies for redundancy in sensor networks. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, pages 332–342, 2005.

[42] C.T. Cheng, C.K. Tse, and F.C.M. Lau. A scheduling scheme for wireless sensor networks based on social insect colonies. *IET Communications*, 3(5):714–722, 2009.

[43] X. Hu, J. Zhang, Y. Yu, H. Chung, Y. L. Li, Y. H. Shi, and X. N. Luo. Hybrid genetic algorithm using a forward encoding scheme for lifetime maximization of wireless sensor networks. *IEEE Transactions on Evolutionary Computation*, 14(5):766–781, 2010.

[44] H. F. Lu, Y. C. Chang, H. H. Hu, and J. L. Chen. Power-efficient scheduling method in sensor networks. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 5, pages 4705–4710, 2004.

[45] B. Cărbunar, A. Grama, J. Vitek, and O. Cărbunar. Redundancy and coverage detection in sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 2(1):94–128, 2006.

[46] K. Kalpakis, K. Dasgupta, and P. Namjoshi. Maximum lifetime data gathering and aggregation in wireless sensor networks. In *Proceedings of the IEEE International Conference on Networking*, 2002.

[47] B. Krishnamachari, D. Estrin, and S. Wicker. Modelling data-centric routing in wireless sensor networks. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, 2002.

[48] B. Krishnamachari, D. Estrin, and S. Wicker. The impact of data aggregation in wireless sensor networks. In *Proceedings of the International Conference on Distributed Computing Systems Workshops*, 2002.

[49] D. Perovic, R. Shah, K. Ramchandran, and J. Rabaey. Data funneling: Routing with aggregation and compression for wireless sensor networks. In *Proceedings of the IEEE International Workshop on Sensor Network Protocols and Applications*, 2003.

[50] K. Dasgupta, K. Kalpakis, and P. Namjoshi. An efficient clustering-based heuristic for data gathering and aggregation in sensor networks. In *Proceedings of the IEEE Wireless Communications and Networking Conference*, 2003.

[51] E. Cayirci. Data aggregation and dilution by modulus addressing in wireless sensor networks. *IEEE Communications Letters*, 7:355–357, 2011.

[52] Haim Zlatokrilov and Hanoch Levy. Session privacy enhancement by traffic dispersion. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pages 1–12, 2006.

[53] W. Diffie and M. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, 1976.

[54] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[55] Federal Information Processing Standard Publication# 180. Secure hash standard. *National Institute of Standards and Technology, US Department of Commerce*, 56:57–71, 1993.

[56] R. L. Rivest and R. W. Baldwin. The rc5, rc5-cbc, rc5-cbc-pad, and rc5-cts algorithms. *RFC*, 1996.

[57] S. Basagni, K. Herrin, D. Bruschi, and E. Rosti. Secure pebblenets. In *Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*, pages 156–163, 2001.

[58] A. Perrig, R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler. Spins: Security protocols for sensor networks. *Wireless networks*, 8(5):521–534, 2002.

[59] R. Blom. An optimal class of symmetric key generation systems. In *Advances in cryptology*, pages 335–338. Springer, 1985.

[60] F. P. Miller, A. F. Vandome, and J. McBrewster. *Advanced Encryption Standard*. Alpha Press, 2009.

[61] E. Shi and A. Perrig. Designing secure sensor networks. *IEEE Wireless Communications*, 11(6):38–43, 2004.

[62] C. Hartung, J. Balasalle, and R. Han. Node compromise in sensor networks: The need for secure systems. *Department of Computer Science University of Colorado at Boulder*, 2005.

[63] M. Conti, R. D. Pietro, L. V. Mancini, and A. Mei. Emergent properties: detection of the node-capture attack in mobile wireless sensor networks. In *Proceedings of the ACM conference on Wireless network security*, pages 214–219, 2008.

[64] M. Conti, R. D. Pietro, L. V. Mancini, and A. Mei. Mobility and cooperation to thwart node capture attacks in manets. *EURASIP Journal on Wireless Communications and Networking*, 2009.

[65] G. D. Meulenaer and F. X. Standaert. Stealthy compromise of wireless sensor nodes with power analysis attacks. In *Mobile Lightweight Wireless Systems*, pages 229–242. Springer, 2010.

[66] C. Krauß, F. Stumpf, and C. Eckert. Detecting node compromise in hybrid wireless sensor networks using attestation techniques. In *Security and Privacy in Ad-hoc and Sensor Networks*, pages 203–217. Springer, 2007.

[67] H. Song, L. Xie, S. Zhu, and G. Cao. Sensor node compromise detection: the location perspective. In *Proceedings of the International conference on Wireless communications and mobile computing*, pages 242–247, 2007.

[68] S. P. Chan, R. Poovendran, and M. T. Sun. A key management scheme in distributed sensor networks using attack probabilities. In *Proceedings of the IEEE Global Telecommunications Conference, (GLOBECOM'05)*, 2005.

[69] D. Huang, M. Mehta, D. Medhi, and L. Harn. Location-aware key management scheme for wireless sensor networks. In *Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pages 29–42, 2004.

[70] Z. Yu and Y. Guan. A robust group-based key management scheme for wireless sensor networks. In *Proceedings of the IEEE Wireless Communications and Networking Conference*, volume 4, pages 1915–1920, 2005.

[71] W. Du, J. Deng, Y. S. Han, P. K. Varshney, J. Katz, and A. Khalili. A pairwise key predistribution scheme for wireless sensor networks. *ACM Transactions on Information and System Security (TISSEC)*, 8(2):228–258, 2005.

[72] D. Liu, P. Ning, and R. Li. Establishing pairwise keys in distributed sensor networks. *ACM Transactions on Information and System Security (TISSEC)*, 8(1):41–77, 2005.

[73] A. Becher, Z. Benenson, and M. Dornseif. Tampering with motes: Real-world physical attacks on wireless sensor networks. In *Proceedings of the 3rd International Conference on Security in Pervasive Computing (SPC)*, pages 104–118, 2006.

[74] K. Bicakci, C. Gamage, B. Crispo, , and A. S. Tanenbaum. One-time sensors: a novel concept to mitigate node-capture attacks. In *Proceedings of the European Workshop on Security and Privacy in Ad hoc and Sensor Networks (ESAS), LNCS 3813*, pages 80–90, 2005.

[75] X. Wang, S. Chellappan, W. Gu, W. Yu, and D. Xuan. Policy-driven physical attacks in sensor networks: modeling and measurement. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, volume 2, pages 671–678, 2006.

[76] F. Ishmanov, A. S. Malik, and S. M. Kim. Energy consumption balancing (ECB) issues and mechanisms in wireless sensor networks (WSNs): a comprehensive overview. *European Transactions on Telecommunications*, 22:151–167, 2011.

[77] S. Ergen and P. Varaiya. On multi-hop routing for energy efficiency. *IEEE Communications Letters*, 9:880–881, 2005.

[78] B. Tavli, M. B. Akgun, and K. Bicakci. Impact of limiting number of links on the lifetime of wireless sensor networks. *IEEE Communications Letters*, 15:43 – 45, 2011.

[79] L. Palopoli, R. Passerone, and T. Rizano. Scalable offline optimization of industrial wireless sensor networks. *IEEE Transactions on Industrial Informatics*, 7:328–339, 2011.

[80] F. Theoleyre and B. Darties. Capacity and energy-consumption optimization for the cluster-tree topology in IEEE 802.15.4. *IEEE Communications Letters*, 15:816–818, 2011.

[81] C. Prommak and S. Modhirun. Optimal wireless sensor network design for efficient energy utilization. In *Proceedings of the IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA)*, pages 814–819, 2011.

[82] M. Erol-Kantarci and H. T. Mouftah. Mission-aware placement of RF-based power transmitters in wireless sensor networks. In *Proceedings of the IEEE Symposium on Computers and Communications (ISCC)*, pages 12–17, 2012.

[83] D. H. Tran and D. S. Kim. Minimum latency and energy efficiency routing with lossy link awareness in wireless sensor networks. In *Proceedings of the IEEE International Workshop on Factory Communication Systems (WFCS)*, pages 75 –78, 2012.

[84] H. Liu, X. Jia, P. J. Wan, C. W. Yi, K. Makki, and N. Pissinou. Maximizing lifetime of sensor surveillance systems. *IEEE/ACM Transactions on Networking*, 15:334–345, 2007.

[85] B. Tavli, M. Kayaalp, O. Ceylan, and I.E. Bagci. Data processing and communication strategies for lifetime optimization in wireless sensor networks. *AEU: International Journal of Electronics and Communications*, 64:992–998, 2010.

[86] B. Tavli, I. E. Bagci, and O. Ceylan. Optimal data compression and forwarding in wireless sensor networks. *IEEE Communications Letters*, 14:408–410, 2010.

[87] K. Bicakci and B. Tavli. Prolonging network lifetime with multi-domain cooperation strategies in wireless sensor networks. *Ad Hoc Networks*, 8:582–596, 2010.

[88] B. Tavli, M. M. Ozciloglu, and K. Bicakci. Mitigation of compromising privacy by transmission range control in wireless sensor networks. *IEEE Communications Letters*, 14:1104–1106, 2010.

[89] P. Tague, D. Slater, R. Poovendran, and G. Noubir. Linear programming models for jamming attacks on network trafic flows. In *Proceedings of the IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pages 207–216, 2008.

[90] S. Jiang and Y. Xue. Optimal wireless network restoration under jamming attack. In *Proceedings of the IEEE International Conference on Computer Communications and Networks (ICCCN)*, pages 1–6, 2009.

[91] H. T. Friis. A note on a simple transmission formula. *IRE*, 34(5):254–256, 1946.

[92] W. Heinzelman. Application-specific protocol architectures for wireless networks [Ph. D. Thesis]. *Boston: Massachusetts Institute of Technology*, 2000.

[93] A. Neskovic, N. Neskovic, and G. Paunovic. Modern approaches in modeling of mobile radio systems propagation environment. *IEEE Communications Surveys*, 3(3):2–12, 2000.

[94] M. Rahimi, R. Baer, O.I. Iroezi, J.C. Garcia, J. Warrior, D. Estrin, and M.Srivastava. Cyclops: in situ image sensing and interpretation in wireless sensor networks. In *Proceedings of the ACM Conference on Embedded Networked Sensor Systems (SenSys)*, pages 192–204, 2005.

[95] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. An application specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1:660–670, 2002.

[96] K. Bicakci, H. Gultekin, and B. Tavli. The impact of one-time energy costs on network lifetime in wireless sensor networks. *IEEE Communications Letters*, 13:905–907, 2009.

[97] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu. Impact of interference on multihop wireless network performance. In *Proceedings of the ACM Annual international conference on Mobile computing and networking (MOBICOM)*, pages 66–80, 2003.

[98] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46:388 –404, 2000.

[99] M. Cheng, X. Gong, and L. Cai. Joint routing and link rate allocation under bandwidth and energy constraints in sensor networks. *IEEE Transactions on Wireless Communications*, 8:3770–3779, 2009.

[100] I. Demirkol, C. Ersoy, and F. Alagoz. MAC protocols for wireless sensor networks: a survey. *IEEE Communications Magazine*, 44:115 – 121, 2006.

[101] B. Tavli and W. Heinzelman. Energy and spatial reuse efficient network-wide real-time data broadcasting in mobile ad hoc networks. *IEEE Transactions on Mobile Computing*, 5:1297–1312, 2006.

[102] B. Tavli and W. Heinzelman. Energy-efficient real-time multicast routing in mobile ad hoc networks. *IEEE Transactions on Computers*, 60:707–722, 2011.

[103] R. Madan, S. Cui, S. Lall, and A. J. Goldsmith. Modeling and optimization of transmission schemes in energy-constrained wireless sensor networks. *IEEE/ACM Transactions on Networking*, 15:1359–1372, 2007.

[104] Y. Sankarasubramaniam, I. F. Akyildiz, and S. W. McLaughlin. Energy efficiency based packet size optimization in wireless sensor networks. In *Proceedings of the IEEE International Workshop on Sensor Network Protocols and Applications (SNPA)*, pages 1–8, 2003.

[105] K. Seada, M. Zuniga, A. Helmy, and B. Krishnamachari. Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks. In *Proceedings of the International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 108–121, 2004.

[106] G. B. Dantzig. *Linear programming and extensions*. Princeton university press, 1998.

[107] J. W. Chinneck. Practical optimization: a gentle introduction. *Electronic document: http://www. sce. carleton. ca/faculty/chinneck/po. html*, 2004.

[108] L. Wolsey. *Integer Programming*. Wiley Interscience Publication, 1998.

[109] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman. *GAMS: A User's Guide*. The Scientific Press, 1998.

[110] R. E. Rosenthal. Gams–a users guide. gams development corporation, washington, dc. *Available online at the following website: http://www. gams. com*, 2012.

[111] X. Wang, W. Gu, S. Challeppan, K. Schoseck, and D. Xuan. Topology optimization in wireless sensor networks for precision agriculture applications. In *Proceedings of the International Conference on Sensor Technologies and Applications*, volume 5, pages 526–530, 2007.

[112] R. Madan, S. Cui, S. Lal, and A. Goldsmith. Cross-layer design for lifetime maximization in interference-limited wireless sensor networks. *IEEE Transactions on Wireless Communications*, 5:3142–3152, 2006.

[113] A. Karnik, A. Iyer, and C. Rosenberg. Throughput-optimal configuration of fixed wireless networks. *IEEE/ACM Transactions on Networking*, 16:1161–1174, 2008.

[114] I. Rhee, A. Warrier, M. Aia, J. Min, and M. L. Sichitiu. Z-MAC: a hybrid MAC for wireless sensor networks. *IEEE/ACM Transactions on Networking*, 16:511–524, 2008.

[115] K. Duffy, D. Malone, and D. Leith. Modeling the 802.11 distributed coordination function in non-saturated conditions. *IEEE Communications Letters*, 9:715–717, 2005.

[116] D. Malone, K. Duffy, and D. Leith. Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions. *IEEE/ACM Transactions on Networking*, 15:159–172, 2007.

[117] X. Liu and P. Mohapatra. On the deployment of wireless sensor nodes. In *Proceedings of the International Workshop on Measurement, Modelling, and Performance Analysis of Wireless Sensor Networks (SenMetrics)*, 2005.

[118] H. Chan and A. Perrig. Security and privacy in sensor networks. *IEEE Computer*, 36:103–105, 2003.

[119] A. Wander, N. Gura, H. Eberle, V. Gupta, and S. C. Shantz. Energy analysis of public-key cryptography for wireless sensor networks. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PERCOM)*, pages 324–328, 2005.

[120] P. Langendoerfer K. Piotrowski and S. Peter. How public key cryptography influences wireless sensor node lifetime. In *Proceedings of the ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN)*, pages 169–176, 2006.

[121] M. R. Ahmad, E. Dutkiewicz, and X. Huang. A survey of low duty cycle MAC protocols in wireless sensor networks. In *Emerging Communications for Wireless Sensor Networks*. InTech, 2011.

[122] L. Eschenauer and V. D. Gligor. A key-management scheme for distributed sensor networks. In *Proceedings of the 9th ACM conference on Computer and communications security*, 2002.

APPENDIX A

GAMS IMPLEMENTATION and COMPILATION DETAILS of an LP EXAMPLE

GAMS code of LP Example

```
*------Options--------
option limrow=100;
*---------------------

* Sets define terms that are going to be used for as an  index
* Below, we define a set with 4 elements.
* i={'n0','n1','n2','n3' }

Set i nodes /n0*n3/;

* Alias is used to assign more than one name for the same set.
* i={'n0','n1''n2''n3''n4' }
* j={'n0','n1''n2''n3''n4' }
Alias (i,j);


* The scalar statement is used to declare and (optionally)
* initialize a GAMS parameter

Scalar
         EAmp picoJoule /100/
         EElec picoJoule /50000/
         Prx reception energy
         Rmax maximum transmission range
;


* Parameters is used to store data before model is running.
* Values of parameters dont changes after model is running

Parameters

         y(i) y coordinate of node-i
         x(i) x coordinate of node-i
         s(i) data generated at node-i
         d(i,j) distance between node-i and node-j
         Ptx(i,j) consumed energy for transmission of data from node-i to node-j
         e(i)  battery power of each node
;


*Variables are used to make decisions and the values of the variables are
*determined after solving mathematical programming problem.
*Mathematical programming tries to conclude with an optimal solution
*by assigning appropriate values to the variables.

Variables
         L lifetime

Positive Variables

         f(i,j) flow;

*Constraints are defined in equations. First, index of equations are given.
* Secondly, equation implementations are provided.

Equations
         noFlow(i,j)  no flow
         flowBalance(i) flow balance
         energyConstraint(i) energy constraint
         transmissonRange(i,j) R_max;
```

```
noFlow(i,j)$(ord(i)=ord(j) or ord(i)=1)..     f(i,j) =e= 0;
flowBalance(i)$(ord(i)>1)..            sum(j,f(j,i)) + s(i)*L =e= sum(j,f(i,j));
energyConstraint(i)$(ord(i)>1)..
            e(i) =g= Prx*(sum(j$(ord(j) > 1),f(j,i)))+sum(j,(Ptx(i,j)*f(i,j)));
transmissonRange(i,j)$(d(i,j)>Rmax)..        f(i,j) =e= 0;


* Model determines which constraints are going to be used in the optimization.

Model MaximumLifetime /
         noFlow
         flowBalance
         energyConstraint
         transmissonRange
/;


*-------------------------Files-------------------------------------------

file Result /c:\Result\LinearTopology-Result.txt/;

file Coordinate /c:\Result\LinearTopology-Coordinate.txt/;

*-------------------------------------------------------------------------


* ------------- Coordinates of nodes ----------------------------------

         x('n0')=0;
         y('n0')=0;

         x('n1')=1;
         y('n1')=0;

         x('n2')=3;
         y('n2')=0;

         x('n3')=4;
         y('n3')=0;
*-------------------------------------------------------------------------


* Calculate distances between nodes and store them in parameter d(i,j)

         d(i,j) = sqrt(sqr(x(i)-x(j))+sqr(y(i)-y(j)));
*-------------------------------------------------------------------------



*-------Energy consumption for transmitting one bit---------
*-------------of data from node-i to node-j-----------------

         Ptx(i,j)$(ord(i)<> ord(j) and ord(i)<>1) = EElec + EAmp*sqr(d(i,j));
         e(i)=1e12;

*-------------------------------------------------------------------------

*---- Energy consumption for receiving one bit of data -------------------
         Prx = EElec;
*-------------------------------------------------------------------------

*----- Rmax is assigned with a big number.
```

```
 Rmax=100000;
*-------------------------------------------------------------------------

*------------------------- Data rate of each node ----------------------
        s(i)$(ord(i)>1)=1;
*-------------------------------------------------------------------------



*---------Run the model and find solution---------------

Solve MaximumLifetime using lp maximizing L;
*-------------------------------------------------------------------------


* the set element labels are identified using their set
* identifier and the suffix (.tl)
*----Coordinates are written to the "LinearTopology_Coordinate.txt" file---
put Coordinate;
loop(i,
                put i.tl:4:0 x(i):12:0  y(i):12:0/;
);

* (.l) is used to write of value of a variable
*----Solution is written to the "LinearTopology_Result.txt" file---
put Result;
put 'Lifetime = '  L.l:12:8 /;

loop(i,
        loop(j,
                put$(f.l(i,j)>0) i.tl:4:0 j.tl:4:0 f.l(i,j):12:3 /;
        );
);
```

GAMS Log of LP Example

```
--- Starting compilation
--- Tutorial.gms(161) 3 Mb
--- Starting execution: elapsed 0:00:00.044
--- Tutorial.gms(133) 4 Mb
--- Generating LP model MaximumLifetime
--- Tutorial.gms(140) 4 Mb
---   13 rows  17 columns  46 non-zeroes
--- Executing CPLEX: elapsed 0:00:00.118

IBM ILOG CPLEX   Mar 17, 2012 23.8.2 WEX 31442.32372 WEI x86_64/MS Windows
Cplex 12.4.0.0

Reading data...
Starting Cplex...
Tried aggregator 1 time.
LP Presolve eliminated 7 rows and 7 columns.
Aggregator did 1 substitutions.
Reduced LP has 5 rows, 9 columns, and 31 nonzeros.
Presolve time =    0.00 sec.
Initializing dual steep norms . . .

Iteration      Dual Objective            In Variable           Out Variable
     1 sI           0.000000              f(n1.n0)energyConstraint( slack
     2        19960079.840319              f(n3.n2)energyConstraint( slack
     3        19387528.155052              f(n3.n1)energyConstraint( slack
LP status(1): optimal

Optimal solution found.
Objective :    19387528.155052

--- Restarting execution
--- Tutorial.gms(140) 2 Mb
--- Reading solution for model MaximumLifetime
--- Executing after solve: elapsed 0:00:00.323
--- Tutorial.gms(157) 3 Mb
--- Putfile Result c:\Result\LinearTopology-Result.txt
--- Putfile Coordinate c:\Result\LinearTopology-Coordinate.txt
*** Status: Normal completion
--- Job Tutorial.gms Stop 10/10/13 17:32:27 elapsed 0:00:00.440
```

GAMS Compilation Details of LP Example

```
 1  *------Options--------
 2  option limrow=100;
 3  *---------------------
 4
 5  * Sets define terms that are going to be used for as an  index
 6  * Below, we define a set with 4 elements.
 7  * i={'n0','n1','n2','n3' }
 8
 9  Set i nodes /n0*n3/;
10
11  * Alias is used to assign more than one name for the same set.
12  * i={'n0','n1''n2''n3''n4' }
13  * j={'n0','n1''n2''n3''n4' }
14  Alias (i,j);
15
16
17  * The scalar statement is used to declare and (optionally)
18  * initialize a GAMS parameter
19
20  Scalar
21           EAmp picoJoule /100/
22           EElec picoJoule /50000/
23           Prx reception energy
24           Rmax maximum transmission range
25  ;
26
27
28  * Parameters is used to store data before model is running.
29  * Values of parameters dont changes after model is running
30
31  Parameters
32
33           y(i) y coordinate of node-i
34           x(i) x coordinate of node-i
35           s(i) data generated at node-i
36           d(i,j) distance between node-i and node-j
37           Ptx(i,j) consumed energy for transmission of data from node-i to
     node-j
38           e(i)  battery power of each node
39  ;
40
41
42  *Variables are used to make decisions and the values of the variables are
43  *determined after solving mathematical programming problem.
44  *Mathematical programming tries to conclude with an optimal solution
45  *by assigning appropriate values to the variables.
46
47  Variables
48           L lifetime
49
50  Positive Variables
51
52           f(i,j) flow;
53
54  *Constraints are defined in equations. First, index of equations are given
     .
55  * Secondly, equation implementations are provided.
56
57  Equations
```

```
58          noFlow(i,j)  no flow
59          flowBalance(i) flow balance
60          energyConstraint(i) energy constraint
61          transmissonRange(i,j) R_max;
62
63
64
65  noFlow(i,j)$(ord(i)=ord(j) or ord(i)=1)..    f(i,j) =e= 0;
66  flowBalance(i)$(ord(i)>1)..           sum(j,f(j,i)) + s(i)*L =e= sum(j,f(i
    ,j));
67  energyConstraint(i)$(ord(i)>1)..
68              e(i) =g= Prx*(sum(j$(ord(j) > 1),f(j,i)))+sum(j,(Ptx(i,j)*f(i,
    j)));
69  transmissonRange(i,j)$(d(i,j)>Rmax)..        f(i,j) =e= 0;
70
71
72
73  * Model determines which constraints are going to be used in the optimizat
    ion.
74
75  Model MaximumLifetime /
76          noFlow
77          flowBalance
78          energyConstraint
79          transmissonRange
80  /;
81
82
83  *-------------------------Files-----------------------------------------
84
85  file Result /c:\Result\LinearTopology-Result.txt/;
86
87  file Coordinate /c:\Result\LinearTopology-Coordinate.txt/;
88
89  *-------------------------------------------------------------------------
    -
90
91
92  * ------------- Coordinates of nodes -----------------------------------
93
94          x('n0')=0;
95          y('n0')=0;
96
97          x('n1')=1;
98          y('n1')=0;
99
100         x('n2')=3;
101         y('n2')=0;
102
103         x('n3')=4;
104         y('n3')=0;
105 *-------------------------------------------------------------------------
    -
106
107
108 * Calculate distances between nodes and store them in parameter d(i,j)
109
110         d(i,j) = sqrt(sqr(x(i)-x(j))+sqr(y(i)-y(j)));
111 *-------------------------------------------------------------------------
    -
112
113
114
115 *-------Energy consumption for transmitting one bit---------
```

```
 116  *-------------of data from node-i to node-j-----------------
 117
 118           Ptx(i,j)$(ord(i)<> ord(j) and ord(i)<>1) = EElec + EAmp*sqr(d(i,j
      ));
 119           e(i)=1e12;
 120
 121  *---------------------------------------------------------------------------
      -
 122
 123  *---- Energy consumption for receiving one bit of data -------------------
 124           Prx = EElec;
 125  *---------------------------------------------------------------------------
      -
 126
 127  *----- Rmax is assigned with a big number.
 128
 129   Rmax=100000;
 130  *---------------------------------------------------------------------------
      -
 131
 132  *------------------------- Data rate of each node -----------------------
 133           s(i)$(ord(i)>1)=1;
 134  *---------------------------------------------------------------------------
      -
 135
 136
 137
 138  *---------Run the model and find solution---------------
 139
 140  Solve MaximumLifetime using lp maximizing L;
 141  *---------------------------------------------------------------------------
      -
 142
 143
 144  * the set element labels are identified using their set
 145  * identifier and the suffix (.tl)
 146  *----Coordinates are written to the "LinearTopology_Coordinate.txt" file--
      -
 147  put Coordinate;
 148  loop(i,
 149                   put i.tl:4:0 x(i):12:0  y(i):12:0/;
 150  );
 151
 152  * (.l) is used to write of value of a variable
 153  *----Solution is written to the "LinearTopology_Result.txt" file---
 154  put Result;
 155  put 'Lifetime = '  L.l:12:8 /;
 156
 157  loop(i,
 158           loop(j,
 159                   put$(f.l(i,j)>0) i.tl:4:0 j.tl:4:0 f.l(i,j):12:3 /;
 160           );
 161  );


COMPILATION TIME     =        0.000 SECONDS      3 Mb  WEX238-238 Apr  3, 2012
GAMS WEX-WEI 23.8.2 x86_64/MS Windows                    10/10/13 17:32:26 Page 2
G e n e r a l   A l g e b r a i c   M o d e l i n g   S y s t e m
Equation Listing    SOLVE MaximumLifetime Using LP From line 140


---- noFlow  =E=  no flow

noFlow(n0,n0)..  f(n0,n0) =E= 0 ; (LHS = 0)
```

```
noFlow(n0,n1)..  f(n0,n1) =E= 0 ; (LHS = 0)

noFlow(n0,n2)..  f(n0,n2) =E= 0 ; (LHS = 0)

noFlow(n0,n3)..  f(n0,n3) =E= 0 ; (LHS = 0)

noFlow(n1,n1)..  f(n1,n1) =E= 0 ; (LHS = 0)

noFlow(n2,n2)..  f(n2,n2) =E= 0 ; (LHS = 0)

noFlow(n3,n3)..  f(n3,n3) =E= 0 ; (LHS = 0)


---- flowBalance  =E=  flow balance

flowBalance(n1)..  L + f(n0,n1) - f(n1,n0) - f(n1,n2) - f(n1,n3) + f(n2,n1)

      + f(n3,n1) =E= 0 ; (LHS = 0)

flowBalance(n2)..  L + f(n0,n2) + f(n1,n2) - f(n2,n0) - f(n2,n1) - f(n2,n3)

      + f(n3,n2) =E= 0 ; (LHS = 0)

flowBalance(n3)..  L + f(n0,n3) + f(n1,n3) + f(n2,n3) - f(n3,n0) - f(n3,n1)

      - f(n3,n2) =E= 0 ; (LHS = 0)


---- energyConstraint  =G=  energy constraint

energyConstraint(n1)..  - 50100*f(n1,n0) - 50000*f(n1,n1) - 50400*f(n1,n2)

      - 50900*f(n1,n3) - 50000*f(n2,n1) - 50000*f(n3,n1) =G= -1000000000000 ;

      (LHS = 0)

energyConstraint(n2)..  - 50000*f(n1,n2) - 50900*f(n2,n0) - 50400*f(n2,n1)

      - 50000*f(n2,n2) - 50100*f(n2,n3) - 50000*f(n3,n2) =G= -1000000000000 ;

      (LHS = 0)

energyConstraint(n3)..  - 50000*f(n1,n3) - 50000*f(n2,n3) - 51600*f(n3,n0)

      - 50900*f(n3,n1) - 50100*f(n3,n2) - 50000*f(n3,n3) =G= -1000000000000 ;

      (LHS = 0)


---- transmissonRange  =E=  R_max

               NONE

GAMS  WEX-WEI 23.8.2 x86_64/MS Windows                  10/10/13 17:32:26 Page 3
G e n e r a l   A l g e b r a i c   M o d e l i n g   S y s t e m
Column Listing      SOLVE MaximumLifetime Using LP From line 140


---- L  lifetime

L
               (.LO, .L, .UP, .M = -INF, 0, +INF, 0)
        1      flowBalance(n1)
```

```
        1       flowBalance(n2)
        1       flowBalance(n3)


---- f  flow

f(n0,n0)
                (.LO, .L, .UP, .M = 0, 0, +INF, 0)
        1       noFlow(n0,n0)

f(n0,n1)
                (.LO, .L, .UP, .M = 0, 0, +INF, 0)
        1       noFlow(n0,n1)
        1       flowBalance(n1)

f(n0,n2)
                (.LO, .L, .UP, .M = 0, 0, +INF, 0)
        1       noFlow(n0,n2)
        1       flowBalance(n2)

REMAINING 13 ENTRIES SKIPPED
GAMS  Rev 238  WEX-WEI 23.8.2 x86_64/MS Windows          10/10/13 17:32:26 Page 4
G e n e r a l   A l g e b r a i c   M o d e l i n g   S y s t e m
Model Statistics    SOLVE MaximumLifetime Using LP From line 140


MODEL STATISTICS

BLOCKS OF EQUATIONS           4     SINGLE EQUATIONS           13
BLOCKS OF VARIABLES           2     SINGLE VARIABLES           17
NON ZERO ELEMENTS            46


GENERATION TIME      =        0.062 SECONDS      4 Mb  WEX238-238 Apr  3, 2012


EXECUTION TIME       =        0.062 SECONDS      4 Mb  WEX238-238 Apr  3, 2012
GAMS WEX-WEI 23.8.2 x86_64/MS Windows                   10/10/13 17:32:26 Page 5
G e n e r a l   A l g e b r a i c   M o d e l i n g   S y s t e m
Solution Report     SOLVE MaximumLifetime Using LP From line 140


               S O L V E      S U M M A R Y

     MODEL   MaximumLifetime     OBJECTIVE  L
     TYPE    LP                  DIRECTION  MAXIMIZE
     SOLVER  CPLEX               FROM LINE  140

**** SOLVER STATUS     1 Normal Completion
**** MODEL STATUS      1 Optimal
**** OBJECTIVE VALUE         19387528.1551

 RESOURCE USAGE, LIMIT          0.015      1000.000
 ITERATION COUNT, LIMIT         3    2000000000

IBM ILOG CPLEX   Mar 17, 2012 23.8.2 WEX 31442.32372 WEI x86_64/MS Windows
Cplex 12.4.0.0

LP status(1): optimal
Optimal solution found.
Objective :    19387528.155052


---- EQU noFlow  no flow
```

```
         LOWER     LEVEL     UPPER    MARGINAL

n0.n0      .         .         .         EPS
n0.n1      .         .         .         .
n0.n2      .         .         .         .
n0.n3      .         .         .         .
n1.n1      .         .         .         .
n2.n2      .         .         .         .
n3.n3      .         .         .         .

---- EQU flowBalance  flow balance

      LOWER     LEVEL     UPPER    MARGINAL

n1      .         .         .        0.007
n2      .         .         .        0.014
n3      .         .         .        0.979

---- EQU energyConstraint  energy constraint

      LOWER     LEVEL     UPPER    MARGINAL

n1 -1.00E+12 -1.00E+12     +INF  -1.327E-7
n2 -1.00E+12 -1.00E+12     +INF  -2.821E-7
n3 -1.00E+12 -1.00E+12     +INF  -1.897E-5

---- EQU transmissonRange  R_max

               NONE


                      LOWER     LEVEL     UPPER    MARGINAL

---- VAR L              -INF  1.9388E+7     +INF       .

  L  lifetime

---- VAR f  flow

         LOWER     LEVEL     UPPER    MARGINAL

n0.n0      .         .        +INF       .
n0.n1      .         .        +INF     -0.007
n0.n2      .         .        +INF     -0.014
n0.n3      .         .        +INF     -0.979
n1.n0      .    1.9674E+7     +INF       .
n1.n1      .         .        +INF     -0.007
n1.n2      .         .        +INF     -0.028
n1.n3      .         .        +INF     -1.928
n2.n0      .    1.9518E+7     +INF       .
n2.n1      .         .        +INF     -0.013
n2.n2      .         .        +INF     -0.014
n2.n3      .         .        +INF     -1.927
n3.n0      .    1.8970E+7     +INF       .
n3.n1      .    2.8656E+5     +INF       .
n3.n2      .    1.3057E+5     +INF       .
n3.n3      .         .        +INF     -0.949


**** REPORT SUMMARY :        0     NONOPT
                             0 INFEASIBLE
                             0  UNBOUNDED
GAMS  Rev 238  WEX-WEI 23.8.2 x86_64/MS Windows          10/10/13 17:32:26 Page 6
```

```
G e n e r a l   A l g e b r a i c   M o d e l i n g   S y s t e m
E x e c u t i o n


**** REPORT FILE SUMMARY

Result c:\Result\LinearTopology-Result.txt
Coordinate c:\Result\LinearTopology-Coordinate.txt


EXECUTION TIME       =        0.110 SECONDS      3 Mb  WEX238-238 Apr  3, 2012




**** FILE SUMMARY

Input      Tutorial.gms

Output     Tutorial.lst
```

APPENDIX B

CURRICULUM VITEA

DAVUT İNCEBACAK

Information Systems Department
Middle East Technical University
Informatics Institute, Universiteler Mahallesi, Dumlupınar Bulvarı, No:1, 06800, Ankara, Turkey
TEL: +90 312 210 7720 FAX: +90 312 210 3745
EMAIL: idavut@metu.edu.tr

AREAS OF INTEREST

- Computer and communications networks
- Ad hoc and sensor networks
- Cellular networks
- Mathematical Programming
- Data and network security
- Distance Learning

EDUCATION

Middle East Technical University, Ankara, Turkey
Doctor of Philosophy in Information Systems 2007-2013
Dissertation Topic: Analysis and Modeling of Routing and Security Problems in Wireless Sensor Networks with Mathematical Programming
Advisor: Prof. Dr. Nazife Baykal, Assoc. Prof. Dr. Kemal Bıçakcı

Middle East Technical University, Ankara, Turkey
Master of Science in Information Systems 2005-2007
Thesis Title: Design and Implementation of a Secure and Searchable Audit Logging System
Advisor: Assoc. Prof. Dr. Kemal Bıçakcı

Sakarya University, Sakarya, Turkey
Bachelor of Science in Computer Engineering 1999-2003

PROJECT EXPERIENCE

Web Based Examination 2006-2007
Intelligent Web-based Multimodal Learning Environment
Role: Developer
Funded by METU

Avicenna 2005-2006
Network of Universities for Open Distance Learning.
Responsible for producing an online computer security and cryptography course.
Role: Co-Tutor
Funded by UNESCO

TEACHING ASSISTANTSHIPS

- Computer Networking for Information Systems, Assistant
- Information Security Management, Assistant
- Web Services and Service Oriented Architecture, Assistant
- Regulatory and Legal Aspects of Information Systems, Assistant
- Introduction to Information Technology and Applications, Developer and System Administrator

PAPERS

1- D. Incebacak, K. Bicakci and B. Tavli *"*Evaluating Energy Cost of Route Diversity in Wireless Sensor Networks with Linear Programming" *Ad Hoc Networks* (Submitted).

2- E. Uzun, B. Tavli, K. Bicakci and D. Incebacak "The Impact of Scalable Routing on Lifetime of Smart Grid Communication Networks" *Ad Hoc Networks* (Submitted).

3- H. U. Yildiz, B. S. Ciftler, B. Tavli, K. Bicakci, and D. Incebacak "The Impact of Incomplete Secure Connectivity on the Lifetime of Wireless Sensor Networks" *IEEE Communications Letters* (Submitted).

4- D. Incebacak, B. Tavli, K. Bicakci and A. Altın "Optimal Number of Routing Paths in Multi-path Routing to Minimize Energy Consumption in Wireless Sensor Networks" *EURASIP Journal on Wireless Communications and Networking, 2013.*

5- A. U. Batmaz, B. Tavli, D. Incebacak and K. Bicakci "The Impact of Link Unidirectionality and Reverse Path Length on Wireless Sensor Network Lifetime" *IEEE International Conference on Communications (ICC) - Ad-hoc and Sensor Networking Symposium, 2013.*

6- E. Uzun, A. Aksac, O. Ozturk, H. E. Kiziloz, D. Incebacak, B. Tavli, K. Bicakci "Network Lifetime Maximization and Localized Routing Tradeoff in Wireless Sensor Networks" *Signal Processing and Communications Applications Conference (SIU), 2013.*

7- A. U. Batmaz, B. Tavli, D. Incebacak, K. Bicakci "Effects of Handshake Hop Length of Unidirectional Links on the Lifetime of Wireless Sensor Networks" *Signal Processing and Communications Applications Conference (SIU), 2013.*

8- D. Incebacak, K. Bicakci and B. Tavli, *"*Energy Cost of Mitigating Physical Attacks in Wireless Sensor Networks" *Fifth IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2012.*

9- D. Incebacak, K. Bicakci, and B. Tavli "Investigating the Tradeoffs between Spatial Granularity and Energy Requirements in Wireless Sensor Networks" *European Modeling Symposium, 2010.*

10- B. Bakır, S. Özkan, R. Köseler, T. Taşkaya Temizel, D. Incebacak, and M.Kaya "The Attitudes of Students with Diverse Backgrounds on Computer and Information Literacy Subjects: Evidence from a First Year Course." *39th Annual Frontiers in Education (FIE) Conference, 2009.*

11- R. Köseler, T. Taşkaya Temizel, B. Bakır, D. Incebacak, M. Kaya, and S. Özkan "Work in Progress: Iterative Curriculum Development for an Interdisciplinary Online-Taught IT Course" *39th Annual Frontiers in Education (FIE) Conference, 2009.*

12- Y. Uzunay, D. Incebacak and K. Bicakci, "Towards Trustable Digital Evidence with PKIDEV: PKI based Digital Evidence Verification Model" *2nd European Conference on Computer Network Defense, University of Glamorgan, United Kingdom, December 2006.*

ON GOING STUDIES

1- Investigating Trade-offs between Sum Rate Maximization and Fairness in Relay-Enhanced OFDMA-Based Cellular Networks
2- Energy Efficient Compromise Tolerant Routing in Wireless Sensor Networks
3- Energy Efficient Path Restoration in Wireless Sensor Networks
4- Impact of Key Distribution on the Lifetime of Wireless Sensor Networks

PROFESSIONAL ACTIVITIES

- Technical Committee member and Chair of the 4th ACM S3 workshop held in conjunction with: 18th ACM International Conference on Mobile Computing and Networking (MobiCom) 2012.

- Technical Committee member of IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2013.

TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü [ ]

Sosyal Bilimler Enstitüsü [ ]

Uygulamalı Matematik Enstitüsü [ ]

Enformatik Enstitüsü [X]

Deniz Bilimleri Enstitüsü [ ]

YAZARIN

Soyadı : İNCEBACAK
Adı : DAVUT
Bölümü : BİLİŞİM SİSTEMLERİ

TEZİN ADI (İngilizce) : ANALYSIS AND MODELING OF ROUTING AND SECURITY PROBLEMS IN WIRELESS SENSOR NETWORKS WITH MATHEMATICAL PROGRAMMING

TEZİN TÜRÜ : Yüksek Lisans [ ] Doktora [X]

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın. [X]
2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullancılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) [ ]
3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) [X]

Yazarın imzası ..... Tarih .....