SECURITY AND QUALITY OF SERVICE FOR WIRELESS SENSOR NETWORKS

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

BY

EMRAH TOMUR

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

FEBRUARY 2008

Approval of the Graduate School of Informatics

_____
Prof. Dr. Nazife BAYKAL
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

_____
Assoc. Prof. Dr. Yasemin YARDIMCI
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

_____	_____
Assoc. Prof. Dr. Y. Murat ERTEN	Prof. Dr. Semih BİLGEN
Co-Supervisor	Supervisor

Examining Committee Members

Prof. Dr. Nazife BAYKAL	(METU, II)	_____
Prof. Dr. Semih BİLGEN	(METU, EEE)	_____
Assoc. Prof. Dr. Y. Murat ERTEN	(TOBB ETU, CENG)	_____
Assoc. Prof. Dr. Gözde B. AKAR	(METU, EEE)	_____
Dr. Alptekin TEMİZEL	(METU, II)	_____

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 : Emrah TOMUR

Signature :

ABSTRACT

SECURITY AND QUALITY OF SERVICE
FOR WIRELESS SENSOR NETWORKS

Tomur, Emrah
Ph.D., Department of Information Systems
Supervisor: Prof. Dr. Semih Bilgen
Co-Supervisor: Assoc. Prof. Dr. Y. Murat Erten

February 2008, 199 pages

Security and quality of service (QoS) issues in cluster-based wireless sensor networks are investigated. The QoS perspective is mostly at application level consisting of four attributes, which are spatial resolution, coverage, system lifetime and packet loss due to collisions. The addressed security aspects are message integrity and authentication. Under this scope, the interactions between security and service quality are analyzed with particular emphasis on the tradeoff between security and spatial resolution for channel capacity. The optimal security and spatial resolution levels which yield the best tradeoff are determined.

In addition, a control strategy is proposed to achieve the desired quality of service and security levels during the entire operation of a cluster-based sensor network. Compared to the existing studies, the proposed method is simpler and has superior performance.

Keywords: Wireless sensor networks, security, quality of service, spatial resolution, coverage

ÖZ

KABLOSUZ ALGILAYICI AĞLARDA GÜVENLİK VE SERVİS KALİTESİ

Tomur, Emrah
Doktora, Bilişim Sistemleri
Tez Yöneticisi: Prof. Dr. Semih Bilgen
Ortak Tez Yöneticisi: Doç. Dr. Y. Murat Erten

Şubat 2008, 199 sayfa

Kablosuz algılayıcı ağlarda güvenlik ve servis kalitesi konuları incelenmiştir. Servis kalitesi perspektifi daha çok uygulama seviyesinde olup uzamsal çözünürlük, kapsama alanı, sistem ömrü ve çarpışmaya bağlı paket kaybı olmak üzere dört başlıktan oluşmaktadır. Ele alınan güvenlik konuları ise mesaj bütünlüğü ve kimlik doğrulamasıdır. Bu kapsamda, güvenlik ve servis kalitesi arasındaki etkileşimler güvenlik ve uzamsal çözünürlük arasındaki kanal kapasitesi kaynaklı ödünleşime özel bir önem verilerek incelenmiştir. En iyi ödünleşim noktasını sağlayan güvenlik ve uzamsal çözünürlük değerleri tespit edilmiştir.

Ayrıca, küme tabanlı kablosuz algılayıcı ağları tüm operasyon süresi boyunca istenen servis kalitesi ve güvenlik seviyesinde tutmak için bir yöntem önerilmiştir. Bu yöntem literatürdeki diğer çalışmalarla karşılaştırıldığında daha basit ve daha başarılıdır.

Anahtar Kelimeler: Kablosuz algılayıcı ağlar, güvenlik, servis kalitesi, uzamsal çözünürlük, kapsama alanı

To my parents, my wife and my son,

ACKNOWLEDGMENTS

I express sincere appreciation to my supervisor Prof. Dr. Semih Bilgen and my co-supervisor Assoc. Prof. Dr. Y. Murat Erten for their guidance and insight throughout the research. I would also like to thank to the committee members, Prof. Dr. Nazife Baykal, Assoc. Prof. Dr. Gözde Bozdağı Akar and Dr. Alptekin Temizel for their suggestions and comments.

I would like to express my gratitude to my colleagues, particularly Tolga and Rıfat, my manager A. Türkay Varlı, the vice-chairman and the chairman of BDDK for tolerating my absences at work. I would also like to acknowledge the support of TUBITAK during my Ph.D. study.

To my wife, Sevil, I offer sincere thanks for her unshakable faith in me and her willingness to endure with me.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACK	Acknowledgement
AES	Advanced Encryption Standard
AODV	Ad-hoc On-Demand Vector
ATM	Asynchronous Transfer Mode
CBC	Cipher Block Chaining
CBC-CTR	Cipher Block Chaining in Counter mode
CBC-MAC	Cipher Block Chaining Message Authentication Code
CIA	Confidentiality, Integrity, Availability
CSMA	Carrier Sense Multiple Access
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CPU	Central Processing Unit
DES	Data Encryption Standard
DiffServ	Differentiated Services
DoS	Denial of Service
DSR	Dynamic Source Routing
EDF	Earliest Deadline First
FDMA	Frequency Division Multiple Access
GPS	Global Positioning System
HIPAA	Health Insurance Portability and Accountability Act
INSENS	INtrusion-tolerant routing protocol for Sensor Networks
IntServ	Integrated Services
IP	Internet Protocol
LEAP	Lightweight Extensible Authentication Protocol

MAC	Medium Access Control
MIC	Message Integrity Code
MPLS	Multi-Protocol Label Switching
PC	Personal Computer
QoS	Quality of Service
RBAC	Role Based Access Control
SAR	Sequential Assignment Routing
SNEP	Secure Network Encryption Protocol
SNGF	Stateless Non-Deterministic Geographic Forwarding
SPINS	Security Protocols for Sensor Networks
TDMA	Time Division Multiple Access
UWB	Ultra Wide Band
WBAN	Wireless Body Area Network
WLAN	Wireless Local Area Network
WSN	Wireless Sensor Network

CHAPTER 1

INTRODUCTION

This introductory chapter addresses the issues that constitute the background of the thesis study such as the objectives, the scope and the contributions. The outline of the thesis is given at the end of the chapter.

1.1. SERVICE QUALITY AND SECURITY FOR SENSOR NETWORKS

Wireless sensor networks (WSN) provide efficient and reliable means for the observation of some physical phenomena which are otherwise very difficult, if not impossible, to observe, and initiation of right actions based on the collective information received from sensor nodes (Akyildiz, Su, Sankarasubramaniam & Çayırcı, 2002). This feature of WSN has significant impact on several military and civil applications such as disaster management, field surveillance and environmental monitoring (Kuorilehto, Hännikäinen & Hämäläinen, 2005).

Due to strict energy limitations of sensor nodes and their deployment in large numbers, most of the research efforts on WSN focused on communication protocols. These are usually required to be energy aware to maximize network lifetime and scalable to accommodate large quantities of sensors (Shah & Rabaey, 2002; Younis, Youssef & Arisha, 2002). Besides, because medium access is a major consumer of sensor energy, energy-efficient medium access control (MAC)

mechanisms are also explored (Jolly & Younis, 2003). The common feature of these research works is that they address the communication problems of WSN applications that require conventional data communications where main concern is energy-efficiency but they do not consider service quality requirements of sensor networks.

However, the envisioned WSN applications introduce quality of service (QoS) requirements for sensor networks of near future. For instance, real-time WSN applications such as target tracking (Tran & Yang, 2006) call for bounded delay and guaranteed bandwidth. Similarly, surveillance applications like habitat monitoring (Mainwaring, Polastre, Szewczyk, Culler & Anderson, 2002) require a certain level of data precision, a pre-defined coverage guarantee and maximal monitoring time. Therefore, those service quality requirements for WSN span a wide category of attributes ranging from network QoS including latency, jitter, throughput and packet loss (Wang, Liu & Yin, 2006) to application QoS composed of spatial resolution, coverage, exposure and system lifetime (Chen & Varshney, 2004).

For envisioned sensor network applications of near future, another requirement, which is also as important as QoS, is an effective security mechanism (Walters, Liang, Shi & Chaudhary, 2006). Since sensor networks may be in interaction with sensitive data or operate in hostile unattended environments like battlefields, protection of sensor data from adversaries is an inevitable requirement (Law and Havinga, 2005). Similarly, for commercial applications of WSN, the protection of privacy such as personal physiological and psychological information is equally important (Slijepcevic, Potkonjak, Tsiatsis, Zimbeck & Srivastava, 2002).

QoS and security mentioned in the previous two paragraphs are not uncorrelated issues in the context of sensor networks, and hence, it is important to consider them together. The reason is two-fold. First, there are such WSN settings where

both QoS and security are required for successful operation of the sensor network, one example of which is a military target tracking application (Ren, 2006). The second reason is the considerable amount of interactions between these two concepts (Bhattacharya, Hinrichs, Nahrstedt & McHugh, 2000). In other words, adding security to a protocol impacts the level of QoS that can be provided, and similarly, choice of QoS mechanisms might affect the security level of the network (Sakarindr, Ansari, Rojas-Cessa & Papavassiliou, 2005). Therefore, there are both positive and negative impacts of security on QoS and vice versa and, QoS and security are not orthogonal concepts but have remarkable correlations.

Thus, providing both security and QoS for sensor networks in a joint fashion is a challenging task not only due to the limited resources of WSN but also due to the complex interactions between the two. Still, however, this challenge should be taken because of the potential near-future WSN applications which require secure and QoS-provisioned transmission of data.

1.2. THESIS OBJECTIVE

Though there exist research studies which consider QoS for sensor networks and security for sensor networks separately, there are hardly any studies in the literature considering both of these parameters together in the context of WSN. Only a few articles on WSN such as Karlof, Sastry and Wagner (2004); Guimarães, Souto, Kelner and Sadok (2005); and Deng, Han, and Mishra (2003) deal with QoS and security at the same time but all from a constrained viewpoint which only analyze the effect of applied security mechanisms on the performance of the sensor networks. In fact, to the best of the author's knowledge, there is only a single work (Chigan, Ye & Li, 2005) which tries to simultaneously control security and QoS levels of a sensor network where service quality is defined as network performance. Yet, there is lack of analysis in the current literature regarding the simultaneous consideration of security and application level service quality issues.

Therefore, the overall aim in this study is to analyze the interactions between security and quality of service for wireless sensor networks and to propose novel schemes for jointly achieving security and QoS for wireless sensor networks. An application level QoS perspective is taken by including spatial resolution, system lifetime and coverage as the service quality attributes. In addition, a collision-minimizing medium access control (MAC) scheme is adopted to further contribute to the service quality. Considering also security, the main purpose of this research is to design a method which shall simultaneously achieve all of the following five objectives for sensor networks during their entire operation: (1) to keep enough number of sensor nodes active (sending data) to attain a desired spatial resolution level, (2) to have these active sensors communicate at the required security level, (3) to maximize the network lifetime by having active sensors periodically power down and inactive ones power up for a balanced energy dissipation, (4) to provide full coverage by having at least one sensor taking measurements at each part of the operation field and, (5) to minimize packet loss due to collisions. In addition, giving a particular emphasis on the tradeoff between security and spatial resolution for channel capacity, another aim is to determine the optimum spatial resolution and security levels yielding the best combination of the two. Thus, this research study has an overall purpose of presenting a means for satisfying the time-varying QoS and security requirements of sensor networks in an optimal way in the sense that constrained resources are utilized in the most efficient way.

1.3. THESIS SCOPE

Despite the wealth of research studies conducted separately on sensor network QoS and sensor network security, which are surveyed in Chapter 2, joint consideration of those two concepts for WSN in the literature is not so common. The limited number of existing research studies analyzing both security and QoS for sensor networks mostly concentrate on network level service quality and does not consider application QoS attributes as in Chigan et al. (2005) or they only

analyze the effects of encryption on sensor network performance without proposing any methods for achievement of those two concepts together as in Karlof et al. (2004).

In this thesis, a different scope is presumed regarding both quality of service and security concepts. The quality of service perspective used throughout this study is mostly at application level rather than network level by the inclusion of application specific QoS attributes such as spatial resolution, coverage and system lifetime. In fact, application QoS and network QoS are two different perspectives on the broad concept of service quality. While the former addresses the degree of success at which an application can perform its tasks, the latter refers to the assurance by the underlying network to provide a set of measurable service attributes such delay, jitter or guaranteed bandwidth. For the sensor network case, application QoS is related to the precision and accuracy of the collected data and can be provided mostly in data collection process whereas the network QoS should be addressed while transporting this collected data to the information sinks. Consequently, provisioning of network and application level service quality usually requires efforts in two different research domains. In this thesis, application level QoS scope is addressed and network QoS issues are mostly not covered.

The security definition used in this thesis includes only integrity and authentication of data packets sent by sensor nodes and does not cover confidentiality. This aspect of security is intentionally left out of the scope of this study considering the security requirements of target applications that are more biased towards message integrity and authentication rather than confidentiality as detailed in Chapter 2. Moreover, although confidentiality of the aggregated data sent to the WSN sink by cluster heads can be important, the main focus in this thesis is on the communication from sensor nodes to the cluster head and the local data contained in individual sensor readings usually do not reveal too much information.

In this regard, the results of this research includes an analysis on the correlation of security and QoS and a method for simultaneous achievement of both concepts for cluster-based wireless sensor networks where security is taken as integrity and authentication and QoS is at application level consisting of three main attributes, namely, spatial resolution, coverage and system lifetime which are directly controlled and a side attribute, packet loss due to collisions, which is minimized by the adopted MAC scheme. The concrete outcomes of this Ph.D. research study are (i) an enhancement on the existing WSN quality of service control strategies to include both security and additional service quality attributes of coverage and packet loss due to collisions, (ii) a novel QoS and security control method superior to existing strategies, (iii) an analysis for the interactions between security and QoS attributes of sensor networks such as the relationship of security and spatial resolution, coverage and spatial resolution and, security and system lifetime, and finally (iv) an optimization method to determine the best tradeoff between security and spatial resolution.

The expected contribution to the wireless sensor network body of knowledge is a comprehensive assessment for joint achievement of security and quality of service for sensor networks that can be further tested for usefulness and applicability. Future research is recommended to substantiate and improve on the findings of the current study, particularly on an extension to include multi-hop networks and multiple clusters.

1.4. THESIS OUTLINE

The remainder of this thesis is organized as follows. Chapter 2 presents a literature review of a number of approaches related with service quality and security issues of wireless sensor networks. The chapter is subdivided into four sections: (1) quality of service issues for sensor networks at both network and application level, (2) security concepts for wireless sensor networks, (3) research studies analyzing both security and QoS, and based on these, (4) a discussion

presenting a roadmap for joint assessment of security and quality of service for WSN. The chapter provides the background to the research by describing what has been done in prior research and illustrate why this research is unique by documenting the work not covered by previous research studies.

Chapter 3 includes a detailed framework of the research by presenting the assumptions, application scenarios, and a communication and system model. In addition, in this chapter, a clear and exact description of the problem to which the proposed strategy of this thesis offers a solution is given.

Chapter 4 presents an analysis on the interactions between security and some service quality attributes, namely spatial resolution and power consumption. In this chapter, the correlation between security and spatial resolution for cluster-based sensor networks is analyzed and it is shown that there is a tradeoff between those two concepts for channel capacity. An optimization problem is formulated to determine the best tradeoff between security and spatial resolution and a computationally efficient solution for the optimization problem is developed. Finally in this chapter, effect of security on power consumption, which eventually effects system lifetime, is investigated.

Chapter 5 proposes a control strategy to satisfy the security and QoS requirements of cluster-based sensor networks. The proposed method is mainly based on the existing QoS control strategy of Kay and Frolik (2004). Different than Kay and Frolik (2004), however, the proposed control method incorporates security and two additional QoS attributes, namely coverage and packet loss due to collisions. Simulation results assessing the performance of the proposed QoS and security control strategy are included. In the final part of this chapter, an approximate probabilistic analysis on the relationship between coverage and spatial resolution under the proposed method is presented.

Chapter 6 is where a novel strategy to satisfy time-varying QoS and security requirements of cluster-based wireless sensor networks is proposed. This proposed strategy is designed to circumvent the deficiencies of the method presented in Chapter 5, which is based on the ACK strategy of Kay and Frolik (2004). Simulation results are presented to compare the performance of two methods.

Chapter 7 outlines the findings and contributions of this thesis study. This final chapter also addresses the limitations of the thesis and reveals the research directions for future work.

Included in the Appendices is supplementary information on some issues utilized but not detailed in thesis body such as medium access control schemes for WSN, simulation tools for wireless sensor networks, QoS optimization methods and *k*-coverage concept. Also, a role based access control scheme that can be implemented by the proposed control strategy of this thesis is presented in the Appendices.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a survey of the pertinent literature in quality of service and security issues of wireless sensor networks and it is divided into four sections. The first section pertains to service quality issues of sensor networks with a review of existing research on QoS requirements and challenges for WSN in comparison to wired networks. The second is on security requirements and challenges of sensor networks, and also includes information on attacks against WSN and their countermeasures. Section three investigates WSN settings which require both security and QoS at the same time, challenges in jointly providing both concepts for sensor networks and a literature review of studies taking those challenges in simultaneous achievement of security and service quality. Deduced from the literature review of first three sections, the fourth section presents a discussion focusing particularly on the lack of research studies jointly addressing application level QoS and security for sensor networks. Before moving on to the first section, below is brief general background information on sensor networks to provide a basis for more specific issues that follow.

The recent developments in low power wireless communications and Micro Electro-Mechanical Systems have created a new technological direction called wireless sensor networks. Wireless sensor networks are composed of a large number of low-cost, low-power, multifunctional, small sensor nodes. Each sensor

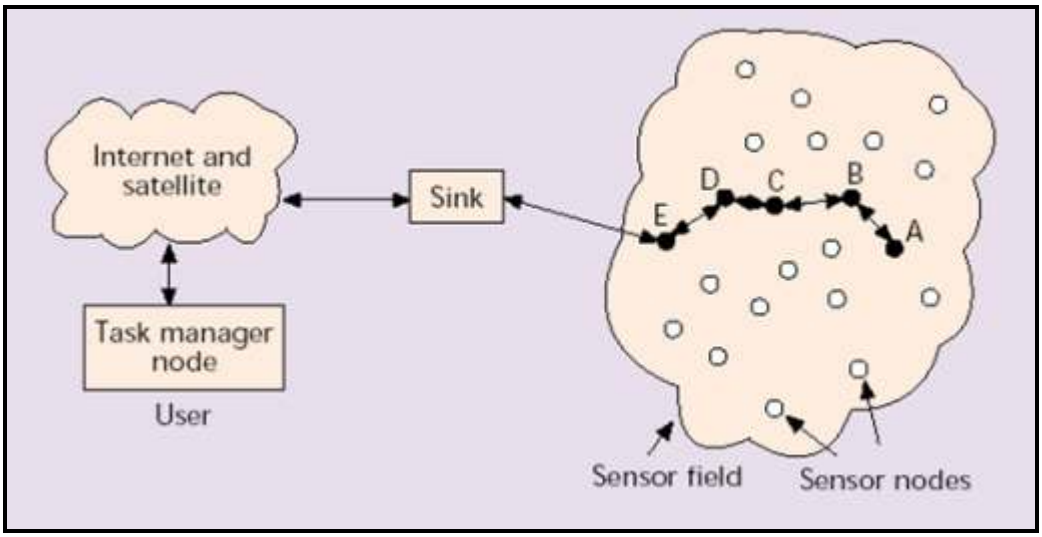


Figure 2.1: A common sensor network architecture (Akyildiz et al., 2002)

node has three main components, namely, a sensing circuitry, a microprocessor and a radio. These sensor nodes, deployed densely inside a physical phenomenon, work in collaboration with each other to perform the application-specific objectives of the sensor network. Sensing circuitry of a sensor node measures the ambient conditions in the vicinity and converts them into electrical signals. Instead of sending this raw measurement data, microprocessor of sensor node processes these signals to carry out some simple computations and then, via its radio, sends this partially processed data to a command center called sink usually through a data fusion center called gateway. A common sensor network architecture is shown in Figure 2.1, redrawn from Akyildiz et al. (2002).

Wireless sensor networks provide efficient and reliable observation of some features of physical phenomena which are otherwise very difficult, if not impossible, to observe, and also initiation of right actions based on collective information from sensor nodes. This feature of WSN has significant impact on several military and civil applications such as target tracking, disaster management, field surveillance and environmental monitoring (Kuorilehto et al., 2005). For instance, sensor networks can be utilized in disaster management

situations like earthquakes to direct emergency response units to affected areas. In military applications, sensors can be used to detect moving targets or presence of dangerous agents such as chemical gases. WSN can also be used in environmental monitoring like tracking of birds or detection of forest fires.

2.1. QUALITY OF SERVICE FOR SENSOR NETWORKS

Before going deeply into the service quality issues for wireless sensor networks, it is useful to define the meaning of the term "quality of service" since different technical communities have various perceptions and interpretations of this term. In fact, application communities use QoS usually to refer to the quality as perceived by the end user/application whereas networking communities take it as a measure of the service quality that the network offers to the applications/users (Ganz, Ganz & Wongthavarawat, 2004). For example, RFC 2386 (Crawley, Nair, Rajagopalan & Sandick, 1998) defines QoS as a set of service requirements to be met when transporting a packet stream from the source to its destination and this corresponds to a networking QoS perspective that means assurance by the underlying network to provide a set of measurable service attributes such delay, jitter, bandwidth and packet loss. Application quality, however, is more difficult to define in a straightforward and generic manner. It is a multi-parameter property linked with the nature of the application and the context of its use (Miras, 2002). In many cases, application quality is synonymous with the human user's degree of satisfaction; in others, it means the degree to which the application is capable of allowing its user to successfully complete a task. Therefore, in a multimedia application, for instance, image resolution, sound and video quality can be considered as application QoS attributes.

These two broad QoS perspectives, namely *Network QoS* and *Application QoS*, can be illustrated by a simple model provided in Ganz et al. (2004) shown in Figure 2.2. In this model, the application/users are not concerned with how the network administers its resources to provide the required service quality. They are

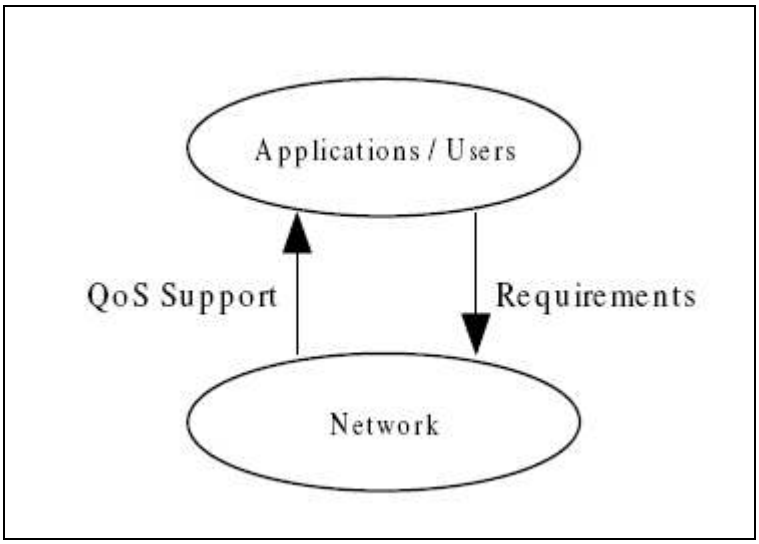


Figure 2.2: A simple QoS model from Ganz et al. (2004)

only interested in the services that networks provide which directly impact the quality of the application. From the network's view, its aim is to provide the required QoS level while maximizing network resource utilization. To achieve this aim, the network is required to analyze the application requirements and deploy various network QoS mechanisms.

QoS issues in traditional data networks have been extensively researched in several studies such as Lee, Hluchyj and Humblet (1995); Wang and Crowcraft (1996); Ma and Steenkiste (1997) and Zhang, Deering, Estrin, Shenker and Zappala (1993), mostly due to the increasing popularity of end-to-end multimedia applications. Nonetheless, so far, there has been little work performed regarding QoS issues in wireless sensor networks. Recently, with the advent of proposed real-time WSN applications, however, the need for providing QoS guarantees in sensor networks has emerged. In this section, after giving a short review of the efforts which aim to support QoS in general communication networks, service quality requirements for wireless sensor networks arising from a wide variety of WNS applications will be investigated. Then, challenges of QoS support in sensor

networks will be presented and current research efforts to overwhelm those challenges will be reviewed.

2.1.1 QoS in traditional wired and general wireless networks

Supporting QoS in wired networks can generally be achieved in two ways. The first one is the over-provisioning of resources and the second one is traffic engineering (Blair, Campbell, Coulson & Hutchison, 1995). In over-provisioning, abundant resources are added to the network so that it can provide satisfactory services to bandwidth-hungry multimedia applications. This method is easy to realize but all the users are served at the same service class. Therefore, the service may become unpredictable during peak traffic. In traffic engineering, users/applications in service classes are classified and each class is assigned a different priority. In the literature, there exist two approaches based on traffic engineering, i.e., reservation-based and reservation-less approaches.

In the reservation-based approach, network resources are reserved during network setup according to an application's QoS request and subject to bandwidth management policy (Patcher, 2006). This is employed in Asynchronous Transfer Mode (ATM), which is a packet-switched network that makes use of virtual circuits. A virtual circuit is established during a setup phase in which the path from sender to receiver is fixed and resources are allocated at each hop. Any type of service guarantee may be made by ATM since all the resources necessary for the connection are reserved for the virtual circuit. Once a virtual circuit is established, ATM is very efficient in terms of the amount of time it takes to forward a packet at a single node. However, ATM is not very efficient in terms of utilization due to the fact the resources are reserved even when no data is flowing in PVC mode. Reservation-based approach is also the approach of the Integrated Services (IntServ) model in the Internet. IntServ uses a call setup stage to reserve the path from sender to receiver and allocate resources at each hop. IntServ reserves resources on a "per-flow" basis where a flow contains all the network

traffic associated with a single application. Unlike ATM, IntServ operates over a heterogeneous network that may have a mixture of IntServ and non-IntServ traffic flowing through each node. As a result IntServ must take measures to guarantee an upper bound on the queuing delays at each hop. IntServ also provides a "controlled-load" service that makes no hard service guarantees, but is designed for real-time multimedia applications.

In the reservation-less approach, no reservation is required. QoS is achieved by some strategies such as admission control, policy managers, traffic classes, and queuing mechanisms. One well-known reservation-less approach is Differentiated Services (DiffServ). In DiffServ, hosts on the edge of the network mark packets with the class of service they should receive. These edge hosts also shape the data they send to ensure that they don't send too much data at once. In the core of the network, routers look at the marking on each packet and forward it according to the per-hop behavior of the packet's class. For example, a packet marked for expedited forwarding will likely spend less time queued than a best-effort packet. The primary advantage of DiffServ over IntServ is that there is much less complexity, and therefore greater efficiency, due to the fact that routers do not need to remember details for multiple flows. However, it can be difficult to implement DiffServ on a heterogeneous network, and it is possible for service guarantees to be violated across the entire network if a single edge host does not mark and shape traffic correctly.

Multi-Protocol Label Switching (MPLS) (Awduche, Malcolm, O'Dell & McManus, 1999) is another QoS architecture that works with IP and intends to bridge the gap between IP and ATM. MPLS has many of the same advantages as ATM and it can handle 1500 byte packets without significant queuing delays. By comparison, an ATM cell is only 53 bytes long and includes 5 bytes of header.

Infrastructure-based wireless networks like Wireless Local Area Networks (WLANs) are the extension of wired networks enabling the connections to be

extended to mobile users. All mobile clients in a communication cell can reach a base station in one hop. QoS challenges in this context arise from the limited bandwidth and user mobility. So, it is intuitive to integrate the QoS architecture deployed in wired networks with wireless MAC protocols. There are several MAC protocols to provide service quality for wireless networks (Ni, Romdhani & Turletti, 2004). Those wireless MAC protocols usually provide data traffic of differentiated classes with corresponding access priorities over the shared wireless medium so that the overall QoS can be supported.

Infrastructure-less wireless networks known as ad hoc networks can be considered as autonomous systems and they have specific individual routing protocols. QoS mechanisms used to support QoS in wired data networks cannot be directly utilized in ad hoc networks due to the bandwidth constraint and dynamic network topology. In this context, it is required to implement complex QoS functionality with limited available resources in a highly dynamic environment. In the literature, QoS support in ad hoc networks includes QoS model, QoS resource reservation signaling, QoS routing, and QoS Medium Access Control (MAC). A QoS model specifies an architecture and impacts the functionality of other QoS components. QoS signaling, whose functionality is determined by the QoS model, acts as a control center in the QoS support system. It coordinates the behavior of QoS routing, QoS MAC, and other components. The QoS routing process searches for a path with enough resources but does not reserve resources, which enhances the chance that resources can be assured when QoS signaling needs to reserve resources. Without QoS routing, QoS signaling can still work but the process of resource reservation may fail. All upper-layer QoS components are dependent on and coordinate with the underlying QoS MAC protocol. A review of these QoS techniques for wireless ad hoc networks is available in Wu and Harms (2001) and Demetrios (2001).

Though sensor networks are also a member of wireless networks family, they have some unique characteristics which do not allow direct use of QoS techniques

mentioned above for generic wireless networks. In addition, because there are several different envisioned sensor network applications, their QoS requirements may be different. For instance, in applications involving event detection or target tracking, the failure to detect or extracting wrong or incorrect information regarding a physical event may arise from many reasons. It may be due to the deployment and network management, i.e., the location where the event occurs may not be covered by any active sensors. Intuitively, coverage (Meguerdichian, Koushanfar, Potkonjak & Srivastava, 2001) or the number of active sensors (Iyer & Kleinrock, 2003) can be defined as parameters to measure the QoS in WSN. In addition, the above failure may be caused by the limited functionality of sensors, e.g., inadequate observation accuracy, low reporting rate of sensors or insufficient observation time. Therefore, parameters such as accuracy, measurement errors or system lifetime can be used to measure QoS for WSN.

Those QoS attributes, i.e., coverage, measurement errors, number of active sensors and system lifetime are directly related to the quality of WSN applications and can be categorized as application level service quality metrics for sensor networks (Wang et al., 2006; Zhou & Mu, 2006). On the other hand, different WSN settings might not be concerned with the quality of applications that are actually carried out but with the performance of data delivery to the sink, e.g., latency between the generation of packets by sensor nodes in case of an event and arrival of those packets to the sink or the total bandwidth that is required to report this event. Those parameters such as packet delay, jitter, bandwidth and throughput are the network level quality of service attributes for WSN (Chen & Varshney, 2004). In the next subsection, some sensor network settings which are in need of both application and network QoS will be examined. Yet, as noted before, the focus of this research is mostly on the application QoS for sensor networks.

2.1.2 Sensor network applications requiring QoS

The first QoS requiring sensor network application of this section is used for environmental monitoring and it has application level service quality needs. The study presented in Mainwaring et al. (2002) is about a habitat monitoring WSN setting where sensors are deployed on the famous Great Duck Island to monitor the microclimates in and around nesting burrows of seabird species called Leach's Storm Petrel. This study intends non-intrusive and non-disruptive monitoring of sensitive wildlife and habitats utilizing the significant advantage of wireless sensor networks over traditional invasive methods of monitoring which involves human presence and infrastructure installation. Deployed sensors of this habitat monitoring application continuously collect data about the temperature, humidity and pressure of the surrounding environment. As pointed out in Mainwaring et al. (2002), in such environmental surveillance applications, the ultimate goal is data collection to derive precise information about the observed phenomenon. Therefore, the application defines a minimum value for the accuracy and spatial precision of the collected data as well as efficient battery consumption of sensor nodes to maximize the monitoring time. Thus, this is an example of a WSN setting with application level QoS requirements such as spatial resolution and system lifetime.

The second sensor network application of this section with QoS needs is real-time target tracking. As mentioned in Younis, Akkaya, Eltoweissy and Wadaa (2004), in a battlefield environment, acoustic sensors may be employed to identify targets and imaging sensors can be used to track them in real time. In such a scenario, once a target is detected and located by contemporary sensors, imaging sensors are immediately turned on to capture images or even video of the target and then to periodically send this multimedia data to the control point. Since this is a military setting, a real-time data exchange between sensor nodes and controller is required to take proper actions in a timely manner. Delivering this kind of time constrained data requires network level QoS guarantees such as minimum

possible delay and certain bandwidth. In Pattem, Poduri and Krishnamachari (2003), a similar target tracking scenario is handled where sensors are scattered over a battlefield first to detect, then identify and finally to track a moving target. In this WSN application, imaging sensors are not employed and sensors are equipped with detectors to realize the presence of a target in their proximity. The aim is to turn on enough number of sensor nodes to take measurements for ensuring adequate accuracy of the target's real-time position. Since activating minimum number of sensor nodes and allowing others to go into the power saving mode will provide battery optimization, the number of active (sensing and data sending) nodes is increased or decreased according to data accuracy needs. In other words, until a target is detected, the smallest possible number of sensors is activated meaning a low spatial resolution requirement. But once the target is detected, QoS, i.e., spatial resolution, requirement is increased for identification and a further rise is needed for real time tracking of the identified target. As seen, despite the similarity of this scenario to the previous target tracking case, QoS requirements are at application level rather than network level. This confirms the observation made previously regarding the diversity of service quality requirements in sensor networks. In the next subsection, challenges brought by this diversity and other characteristics of WSN are investigated.

2.1.3 QoS challenges in sensor networks

The unique characteristics and requirements of sensor networks pose new challenges for QoS support in WSN in addition to the ones inherited from general wireless networks such as link quality and dynamic network environment. Some of these sensor network QoS challenges cited in the literature (Chen & Varshney, 2004; Younis et al., 2004; Gurses & Akan, 2005) such as bandwidth constraints, resource limitations, energy-delay tradeoff, data redundancy, multiple traffic types, unbalanced traffic, scalability, multiple sinks, network dynamics, energy balance and packet criticality are briefly explained below.

Bandwidth constraints: Real-time multimedia applications have high bandwidth requirements that are hard to satisfy even on wire-based networks. Sensor networks, however, have very scarce bandwidth available to them. Furthermore, sensor nodes not only relay their own data but also relay the packets coming from other nodes due to multihop communication strategy and this puts more burden on the available bandwidth. In addition, traffic in a sensor network can be composed of a mix of real-time and non-real-time traffic. Dedication of whole bandwidth to real-time data requiring QoS is not acceptable and a tradeoff in multimedia data quality might be needed to accommodate non-real-time traffic. A solution to overcome those bandwidth limitations might be use of ultra wideband (UWB) technologies.

Resource limitations: Wireless sensor networks have very stringent constraints on resources such as energy, memory, processing capability, transmission power and buffer size. Most important of these limitations is available energy of nodes because it is not feasible to replace or recharge the batteries of sensors once deployed in the field and depletion of energy of a node renders it unusable. Consequently, QoS support mechanisms for WSN should be designed in simplicity and low-complexity avoiding computation intensive algorithms, expensive signaling protocols and overwhelming state information on nodes which increases power consumption.

Energy-Delay tradeoff: Because the transmission power of sensor node radios is limited, use of multi-hop routing is the most common technique in WSN data communication. Although use of multi-hop routing decreases energy consumption of individual nodes during transmission, it comes with a cost, that is, increased latency in end-to-end packet transfer. This increase in accumulated delay is mostly due to packet queuing (not propagation delay) at multiple sensor nodes and therefore it complicates the analysis and handling of QoS constrained traffic. Thus, it may be unavoidable to sacrifice energy efficiency to meet timely delivery requirements when designing QoS methods for sensor networks.

Data redundancy: High redundancy in the generated data is a characteristic of wireless sensor networks. For conventional unconstrained traffic, using aggregation functions to eliminate redundant data is helpful. However, data fusion or aggregation for QoS constrained multimedia traffic is not a trivial task. Comparing video streams or images is a computationally expensive task and consumes energy resources. In addition, these complex computations may also increase latency and therefore complicates QoS design in WSN further. Using a combination of system and sensor level rules might be a solution to make aggregation of QoS traffic computationally feasible. For example, aggregation of imaging data can be selectively performed for data generated by sensor nodes pointing to very close directions.

Multiple traffic types: Since sensor networks usually include heterogeneous sets of sensors, several issues arise regarding support of QoS constrained traffic. For instance, some WSN applications require a mixture of sensor nodes for temperature, pressure and humidity monitoring of the surrounding environment, motion detection using acoustic signatures and capturing image or video of moving targets. Reading of generated data from these sensors can be at different rates, subject to diverse quality of service constraints and following multiple data delivery models. So, this kind of a heterogeneous environment makes QoS support more challenging.

Unbalanced traffic: In majority of sensor network settings, traffic flow is from a large number of sensor nodes to a small set of sink nodes. Therefore, this unbalanced traffic should be taken into account when designing QoS mechanisms for wireless sensor networks.

Scalability: A usual sensor network is composed of hundreds or even thousands of individual sensor nodes densely deployed in the environment. Therefore, QoS schemes designed for WSN should be able to scale up to an enormous number of

nodes. For instance, provided QoS should not degrade quickly when deployed node density increases.

Multiple sinks: In a sensor network, there may exist more than one sink node, which impose different requirements on the network. For example, one sink may ask sensor nodes located in the southeast of the sensor field to send a temperature report every one minute, while another sink node may only be interested in an exceptionally high temperature event in the northwest area. WSNs should be able to support different QoS levels associated with different sinks.

Network dynamics: Network dynamics may arise from node failures, wireless link failures, node mobility, and node state transitions due to the use of power management or energy efficient schemes. Such a highly dynamic network greatly increases the complexity of QoS support.

Energy balance: In order to achieve a long-lived network, energy load must be evenly distributed among all sensor nodes so that the energy at a single sensor node or a small set of sensor nodes will not be drained out very soon. QoS support should take this factor into account.

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

2.1.4 Solutions in the literature to the QoS challenges of sensor networks

As mentioned before, vast majority of WSN research has been focused on energy-efficient methods, and so far little attention has been paid to propose solutions for service quality requirements of sensor networks. A few research attempts have recently been started to address QoS requirements for WSN. In this subsection,

current research efforts addressing QoS challenges of wireless sensor networks are surveyed by summarizing some important published work in the literature. First, studies focusing on network level QoS attributes and then, the ones on application QoS will be surveyed.

Sequential Assignment Routing (SAR) is the first protocol that includes a notion of QoS for sensor networks (Sohrabi, Gao, Ailawadhi & Pottie, 2000). Assuming multiple paths to the sink node, each sensor uses SAR algorithm for path selection. It takes into account the energy and QoS factors on each path, and the priority level of a packet, and accordingly creates trees rooted at one-hop neighbors of the sink. By using the created trees, multiple paths from sink to sensors are formed, only one of which is used and the rest is kept as backup. For each packet routed through the network, a weighted QoS metric is computed as the product of the additive QoS metric and a weight coefficient associated with the priority level of that packet for purposes of performance evaluation. The objective of the SAR algorithm is to minimize the average weighted QoS metric throughout the lifetime of the network. Simulation results show that SAR offers less power consumption than the minimum-energy metric algorithm, which focuses only the energy consumption of each packet without considering its priority. SAR maintains multiple paths from nodes to sink to allow fault-tolerance and easy recovery, but, the protocol suffers from the overhead of maintaining the tables and states at each sensor node, especially when the number of nodes is huge. SAR also does not use redundant routes to split the load and effectively boost the bandwidth.

A QoS routing protocol for sensor networks named SPEED which provides soft real-time end-to-end guarantees is described in He, Stankovic, Lu and Abdelzaher (2003). This protocol requires each node to maintain information about its neighbors and uses location-based routing to find the paths. In addition, SPEED tries to guarantee a certain speed for each packet in the network so that each application, before making the admission decision, can estimate the end-to-end

Figure 2.3: Routing components of SPEED (He et al., 2003)

delay for the packets by considering the distance to the sink and the speed of the packet. Furthermore, SPEED can provide congestion avoidance when the network is overloaded. The routing module in SPEED, which is called Stateless Non-Deterministic Geographic forwarding (SNGF), works with four other modules at the network layer, as shown in Figure 2.3, which is taken from He et al. (2003). The beacon exchange mechanism collects information about the nodes and their location. Delay estimation at each node is basically made by computing the elapsed time when an ACK is received from a neighbor as a response to a transmitted data packet. By checking the delay values, SNGF selects the node that meets the speed requirement. If there is not such a node, the relay ratio of the node is checked. The Neighborhood Feedback Loop module is responsible for calculating the relay ratio, by looking at the miss ratios of the neighbors of a node (the nodes which could not provide the desired speed) and this ratio is fed to the SNGF module. If the relay ratio is below a randomly generated number between 0 and 1, the packet is dropped. Finally, the backpressure-rerouting module is utilized to prevent voids, when a node fails to find a next hop node, and to clear congestion by sending messages back to the source nodes so that they will try to find new routes.

Compared to conventional ad hoc routing protocols such as Dynamic Source Routing (DSR) and Ad-hoc On-Demand Vector routing (AODV), SPEED performs better in terms of end-to-end delay and miss ratio. In addition, the total transmission power is less thanks to the simplicity of the routing algorithm, i.e. control packet overhead is less, and also due to the even traffic distribution. Such

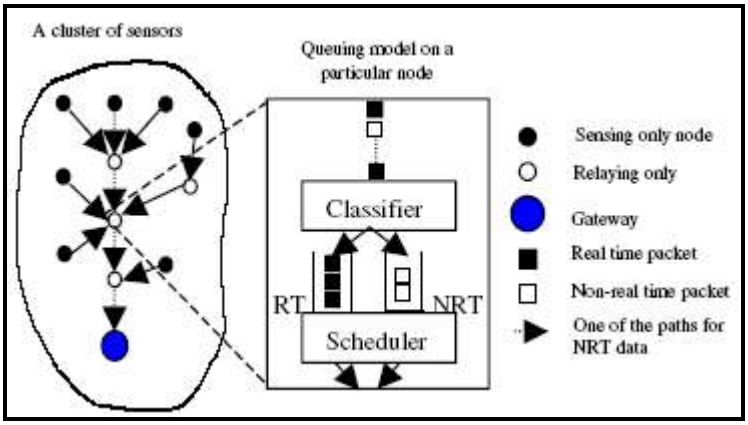


Figure 2.4: Queuing model in Akkaya and Younis (2003)

load balancing is achieved through the SNGF mechanism of dispersing packets into a large relay area. SPEED does not take any energy metric into account in its routing protocol.

A fairly new energy-aware QoS protocol for sensor networks is proposed in Akkaya and Younis (2003). Using an extended version of Dijkstra's algorithm, the proposed protocol tries to find a least cost and energy efficient path that meets certain end-to-end latency during the connection. The link cost function used is composed of the nodes' energy reserve, transmission energy, error rate and other communication parameters. In order to support both best effort and real-time traffic simultaneously, a class-based queuing model is employed. This queuing model allows service sharing for real-time and non-real-time traffic. The bandwidth ratio r, is defined as an initial value set by the gateway and represents the amount of bandwidth to be dedicated both to the real-time and non-real-time traffic on a particular outgoing link in case of a congestion. As a consequence of this, the throughput for non-real-time data does not diminish by properly adjusting this "r" value. The proposed queuing model is depicted in Figure 2.4, which is redrawn from Akkaya and Younis (2003). Simulation results show that the proposed protocol consistently performs well with respect to QoS and energy metrics. However, the same r-value is set initially for all nodes, which does not

provide flexibility in adjusting bandwidth sharing for different links. The protocol is extended in Akkaya and Younis (2005) by assigning a different r-value for each node in order to achieve a better utilization of the links.

The research in Sohrabi et al. (2000), He et al. 2003, Akkaya and Younis (2003) and, Akkaya and Younis (2005) propose WSN routing protocols to satisfy network QoS requirements. Chen and Varshney (2004) discusses that all those methods have some drawbacks. Firstly, they are all based on the concept of end-to-end applications, which may not be necessarily used in sensor networks. Next, the algorithms used in each of those protocols are too complex for sensor network applications which usually have strict resource constraints. Finally, none of them accounts for data-centric routing techniques which are often utilized in sensor networks. There are more recent works on QoS aware WSN routing such as Tang and Li (2006) and Mahapatra, Anand and Agrawal (2006) that attempt to overcome those limitations of previously proposed methods.

Addressing network QoS, some studies proposing WSN medium access control (MAC) protocols involving real-time scheduling techniques also exist. Caccamo, Zhang, Sha and Buttazzo (2002) proposes an implicit prioritized access protocol for sensor networks that utilizes Earliest Deadline First (EDF) scheduling algorithm with the aim of ensuring timeliness for real-time traffic. The main point is to take advantage of the periodic nature of the sensor data traffic to create a schedule instead of using control packets for channel reservation. RAP (Lu, Blum, Abdelzaher, Stankovic & He, 2002) is another work that considers a real-time scheduling policy for wireless sensor networks. RAP is a communication architecture for WSN which proposes velocity-monotonic scheduling in order to minimize deadline miss ratios of packets. Each packet is put to a different FIFO based queue based on their requested velocity, i.e. the deadline and closeness to the gateway. This approach aims to ensure a prioritization in the MAC layer.

One of the initial analyses which investigate application level quality of service issues for wireless sensor networks is Iyer and Kleinrock (2003). It defines sensor network QoS in terms of how many of the deployed sensors are active in sending data to the information sink. In this way, QoS concept is taken to be the same as spatial resolution, which is the amount of useful information that can be constructed by the aggregation of data sent by individual sensor nodes. Therefore, as number of active sensors increases, spatial resolution of the sensor network gets increased.

The main purpose of the research in Iyer and Kleinrock (2003) is to control the sensor network in such a way that the optimal spatial resolution level, which is known a priori, is attained during the sensor network operation period. Besides, in order to maximize the network lifetime, active sensors contributing to the spatial resolution are periodically changed to distribute power usage among all available sensors. Thus, sensors become active by taking turns and at each discrete time interval a different set of sensors send data to the sink. The overall aim is then to have just enough number of sensors in each active set which is sufficient to achieve the desired spatial resolution value at all times and also to have different sensors appear in active sets of different time intervals to conserve power.

The most straightforward way of achieving the aim above is to let a central authority, i.e. cluster head of a cluster-based sensor network, decide which sensors will be active to get the desired spatial resolution value during each time interval and inform the nodes about this. In such an approach, cluster head chooses a different set of sensor nodes for each time and this will continue in a periodic fashion. Unfortunately, however, this approach is not applicable to sensor networks. The reason is that there will be deaths or addition of sensors during the operation of the network and consequently, it is not possible to maintain a list of alive sensor nodes based on their identities.

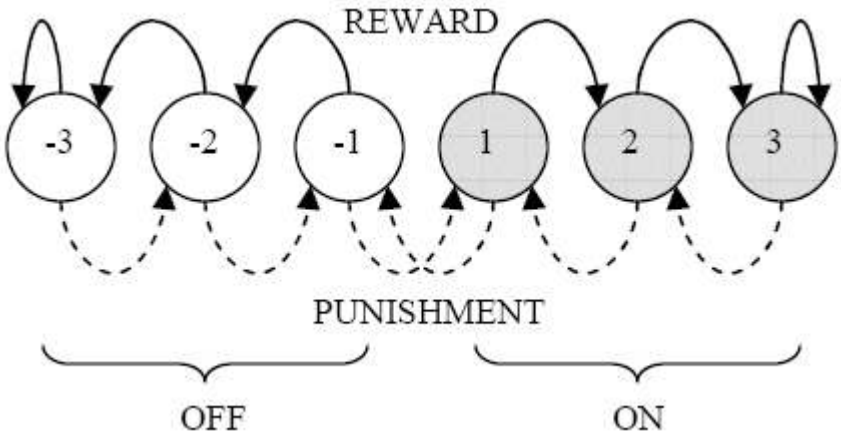


Figure 2.5: Gur Memory of size N=3 (Iyer & Kleinrock, 2003)

Then, one can recommend the use of statistical method having the cluster head broadcast a probability value instead of the sensor's ID list. Making the sensors to switch ON of OFF based on this probability value, one can have the average number of sensors transmitting converge to the desired spatial resolution level. In order to achieve this, the broadcasted probability should be equal to the result of the desired number of active sensors divided by the total number of all non-dead sensors. Yet, this approach is neither applicable because of the fact that one cannot know the number of non-dead/operational sensor nodes due to the reasons previously mentioned.

To accomplish the goal of attaining desiring spatial resolution value and maximizing network lifetime and also taking the above mentioned constraints into consideration, authors of Iyer and Kleinrock (2003) utilize a statistical paradigm called *Gur Game*. In proposed control strategy of Iyer and Kleinrock (2003), cluster head of a cluster-based sensor network, periodically broadcasts a probability at discrete time intervals. Each sensor compares this probability value to its locally generated random number for the current time interval and based on this comparison, jump between the states of a finite state automaton called Gur

Memory shown in Figure 2.5. Based on the state it is in, a sensor either transmits (ON) or does not transmit (STANDBY or OFF) for the current time interval.

The probability broadcasted to all sensor nodes is dependent on the number of ON sensors (not all alive) and computed by using a special reward function whose maxima occurs at statistical mean value of the desired spatial resolution value. For instance, to keep an average of 35 sensors ON throughout the operation of the network, reward probability is computed by p(ON) = 0.2 + 0.8exp(−0.002(ON − 35)^2) whose maxima occurs at ON=35. Therefore, using such a function in conjunction with Gur Game to control the behavior of sensor nodes to switch them between ON and STANDBY states, the desired mean value of the random variable ON can be attained. In other words, number of active sensors can be made to converge to the desired spatial resolution value in the steady state. The disadvantage of this strategy is that it requires all sensors to keep their radio receivers open all the time to get new reward probability value of the current epoch broadcasted by the cluster head. This causes energy inefficiencies.

In Kay and Frolik (2004), authors present an alternative to the Gur Game strategy of Iyer and Kleinrock (2003) to be used in controlling the QoS level of a sensor network. The network topology assumptions and the QoS definition are exactly the same in both studies, i.e., a cluster-based sensor network is considered and QoS is taken to be equal to spatial resolution. The main difference of Kay and Frolik (2004) from Iyer and Kleinrock (2003) is its control scheme which is not dependent on broadcasts by the cluster head. Named as the ACK strategy, proposed control method of Kay and Frolik (2004) relies on cluster head's unicast messages to only transmitting nodes allowing non-transmitting nodes to shut down their radios, thus providing energy efficiency.

In ACK strategy, each sensor is associated with a finite state automaton as illustrated in Figure 2.6. Yet, there are no ON (transmitting) or STANDBY (non-transmitting) states. Instead, all nodes are said to be in varying states of being ON.

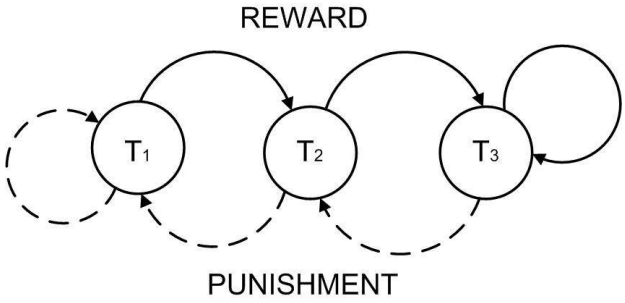


Figure 2.6: Finite state automaton for the ACK strategy of Kay and Frolik (2004)

This means that each state *i* corresponds to a different transmit probability *Ti* such that *Ti*> *Tj* for *i*>*j*. At each discrete time interval (epoch), each node compares its locally generated random number to the transmit probability corresponding to its current state *i*. Then, each node decides whether to transmit or not in the current epoch based on this comparison. At the end of each epoch, cluster head counts the number of packets it received during this epoch and compares this value to the desired spatial resolution value. Cluster head sends the result of this comparison in an ACK packet to only active nodes which transmitted in the epoch. On receiving this acknowledgement packet, transmitting nodes change their states based on the 1-bit information in the ACK packet. They reward themselves if the bit is 1 meaning that the number of transmitting nodes is lower than the desired spatial resolution and they punish otherwise.

As a result, the ACK strategy of Kay and Frolik (2004) causes transmitting nodes to adjust their transmit probability according to the difference between current and desired levels of spatial resolution. So, in the steady state, the sensor network is expected to converge to the desired spatial resolution value. Since non-transmitting nodes can turn off their receivers at the beginning of each epoch, this strategy enhances power conservation. Still, however, all nodes are ON at any moment in the sense that they all generate random numbers and compare these to their transmit probability. Authors have shown that the ACK strategy of Kay and

Frolik (2004) outperforms the Gur Game method of Iyer and Kleinrock (2003) more than five times regarding total network life.

Perillo and Heinzelman (2003) provides application QoS via the joint optimization of sensor scheduling, i.e., selecting active sensor sets, and finding paths for data routing. They address the problem of maximizing lifetime of a wireless sensor network while meeting a minimum level of application reliability. Application reliability is defined as providing enough data for the application so that a reliable description of the environment can be derived and in their simulations, they consider reliability to be the fraction of area covered by the sensor network. It is assumed that for the majority of the network lifetime, the sensors act in a vigilant state, looking for a potential phenomenon in the environment being monitored. In this case, the state of the application in terms of reliability requirements remains constant over time. Yet, in some applications such as object tracking where higher reliability is required in the vicinity of the object and nearby sensors become more critical, QoS requirement can change over time. The method proposed in Perillo and Heinzelman (2003) cannot be directly applied in such cases. It is useful only when a certain level of data reliability requirement is given. By the use of two strategies –turning off redundant sensors and energy efficient routing–, the proposed method tries to extend network lifetime for this QoS (reliability) level. Therefore, the application QoS attributes handled in Perillo and Heinzelman (2003) are system lifetime and application reliability. In some other papers such as Meguerdichian et al. (2001) and Meguerdichian, Koushanfar, Qu and Potkonjak (2001), QoS is also defined as coverage and exposure, respectively. The basic idea here is to determine how to cover the desired area of interest or leave no sensing holes so that sensors can detect unexpected events as quickly as possible and as reliably as possible.

The last paper that will be reviewed in this section is a fairly recent study on WSN application QoS. In Delicato, Protti, Pirmez and de Rezende (2006), an efficient approach for selecting active nodes in WSN is proposed. The primary goal is to

maximize residual energy and application relevance of the selected nodes to extend the network lifetime while meeting application-specific QoS requirements. The authors formalize the problem of node selection as a knapsack problem and adopt a greedy heuristic for solving it. An environmental monitoring application is chosen to derive some specific requirements. Given an application submitting a sensing task to a WSN, the node selection process corresponds to the algorithm that decides which sensors should be active for the execution of that particular task. In order to avoid an early energy depletion of active nodes, the algorithm should alternate between subsets of active nodes during the complete task execution. In the proposed algorithm, execution time is divided into rounds of size t. During each round, the subset of selected nodes and their roles do not change. The decisions made by the algorithm are based upon information contained in interests submitted by the application, which consist of the task descriptor and QoS requirements. The former contains the type of sensor-collected data, the data-sending rate, the geographical area of interest (target area), and the monitoring duration and interval. The QoS requirements are application-dependent but, in the case of environmental monitoring, they may be expressed as minimum values for accuracy and spatial precision of sensor-collected data. The node selection algorithm is executed in the following three cases: (i) initially, when a new application submits its interests to the network; (ii) proactively, for purposes of energy saving or due to changes in the application QoS requirements; or (iii) reactively, whenever some QoS violation is detected. The proposed algorithm seeks to select the best subset of sensors to be activated by using three strategies: (i) minimizing network energy consumption by choosing the smallest possible number of nodes capable of providing the requested level of QoS; (ii) maximizing the sum of the residual energy of selected nodes, so that energy is spent in a uniform way among sensors during task execution time, thus avoiding the premature collapse of excessively used nodes; and (iii) taking into account the potential relevance, from the application point of view, of each individual sensor node.

2.2 SECURITY FOR WIRELESS SENSOR NETWORKS

As the application areas of wireless sensor networks continue to grow, new requirements like QoS, which is detailed in the previous section, has started to appear. For envisioned applications of WSN, another requirement, which is as important as QoS, is effective security mechanisms. Since sensor networks may be in interaction with sensitive data or operate in hostile unattended environments like battlefield, protection of sensor data from adversaries is an inevitable need. Similarly, for commercial applications of WSN, the protection of privacy such as personal physiological and psychological information is equally important. However, because of inherent resource and computing limitations, security challenges posed by sensor networks are quite different than traditional network security challenges. Therefore, there are lots of open research issues waiting to be solved in WSN security. In this section, security requirements for envisioned WSN applications are given first. Then, challenges in designing WSN security schemes are considered in comparison with conventional security methods. Following the subsection describing attacks that can be launched against sensor networks, existing studies in the literature proposing solutions for those attacks and for other WSN security problems are surveyed in the last subsection.

2.2.1 Security requirements of sensor network applications

The fundamental security requirements for typical data networks are confidentiality, integrity and availability, which are also known as CIA triad. These are accompanied by other requirements such as authentication, non-repudiation, accountability, etc. Sensor networks share most of these requirements but also pose unique requirements of their own. These security requirements for sensor networks are data confidentiality, data integrity, data freshness, authentication and availability as given in Walters et al. (2006).

Data confidentiality: In the context of sensor networks, confidentiality relates to the following: (1) Sensor nodes in a sensor network should not leak sensor

readings to non-participating parties. Particularly in military applications, data stored in sensor nodes may be highly sensitive. (2) In many applications, nodes communicate confidential data such as key distribution. Therefore, it is very important to build a secure channel in a wireless sensor network. (3) Public sensor information, such as sensor identities and public keys, should also be encrypted to some extent to protect against traffic analysis attacks.

Data integrity: Providing confidentiality prevents disclosure of data to an adversary. However, this doesn't mean that data is fully safe. The adversary can change the data, with the aim of putting the sensor network into confusion. For instance, a malicious sensor node may add some fragments or modify the data within a packet. This new packet can then be sent to the original receiver leading it into error. Data integrity can be spoiled due to the harsh communication environment even without the presence of a malicious node. Therefore, data integrity providing schemes should be used in WSN to ensure any received data has not been altered in transit.

Data freshness: Even if confidentiality and integrity of data are assured, it is also needed to ensure the freshness of each message since sensor networks stream some forms of time-varying measurements. Data freshness implies that the data is recent, and it ensures that no old messages have been replayed. This requirement is particularly important when shared-key strategies are employed in the design. Shared keys need to be changed over time and it takes time for new shared keys to be propagated to the entire network. In this case, it is easy for the adversary to use a replay attack and to disrupt the normal work of the sensor, especially if the sensor is unaware of the new key change time. To solve this problem, a nonce or another time-related counter can be added into the packet to ensure data freshness.

Authentication: Authentication allows a receiver to verify that the data really is sent by the claimed sender. It is important for several applications in sensor networks. For example, authentication is necessary for many administrative tasks

such as network reprogramming or controlling sensor node duty cycle. In addition, a malicious node can easily inject messages, so the receiver needs to ensure that the data used in any decision-making process comes from the correct source. In the two-party communication case, data authentication can be achieved through a symmetric key mechanism: The sender and the receiver share a common secret key to compute a message integrity code (MIC) which is appended to the data payload. When a message with a correct MIC arrives, the receiver knows that it must have been sent by the sender. This kind of authentication cannot be applied to a broadcast setting unless much stronger trust assumptions are placed on the network nodes. If one sender wants to send authentic data to mutually mistrusted receivers, use of a symmetric MIC is insecure: Any one of the receivers knows the MIC key, and hence could impersonate the sender and forge messages to other receivers. Hence, asymmetric mechanisms are needed to achieve authenticated broadcast.

Availability: Availability refers to the readiness of data for the access of the authorized users when needed. Denial of service (DoS) attacks, which are the most common threat for the availability of traditional networks, threaten also the availability of sensor networks. Most common form of DoS attack are in the form of jamming where an adversary attempts to disrupt the operation of WSN by broadcasting a high energy signal. Or, attackers can induce battery exhaustion in sensor nodes by sending a sustained series of useless communications that the targeted nodes will expend energy processing. Thus, security protocols designed for sensor networks should protect against attacks which menace the availability of the network.

2.2.2 Security challenges in sensor networks

In conventional communication networks, the security mechanisms utilized to support the CIA triad is well known and have been in use for years. For instance, symmetric key encryption algorithms such as DES, 3DES, AES, RC4, etc. and

public key encryption algorithms such as RSA, Elliptic Curve, Knapsack, etc. are in use to provide confidentiality. In order to provide authentication and integrity, message integrity codes, digital signatures, one-way hash functions, etc. are utilized. However, due to the unique challenges posed by sensor networks, traditional security techniques cannot be directly applied to WSN. Challenges in WSN security design can be classified into four broad categories, which are resource constraints, unattended operation, in-network processing and unreliable communication. These WSN security challenges are briefly explained in the following (Walters et al., 2006).

Resource constraints: All security approaches require a certain amount of resources to be implemented such as memory, computational power and energy. However, developed to be compact, sensor nodes are very limited in terms of storage capacity, processing capability and energy sources. For instance, a common sensor type has a 8-bit 4MHz processor with a total of 8K memory and disk space. With such a limitation, the size of the security software developed for a sensor should also be quite small. Besides, it is not feasible to perform computationally complex security algorithms like public key cryptography using very incapable processors of sensor nodes. In a similar way, the limited power capacity of sensors and the inability to replace and recharge batteries once depleted puts strict limitations on the use of energy. Therefore, energy impacts of proposed security schemes for WSN should also be taken into account. Because cryptographic methods cause extra power consumption due to processing functions such as encryption, decryption, verification etc. and also due to transmission of cryptographic overhead like digital signatures, energy efficient security algorithms should be designed for sensor networks.

Unattended operation: Though it depends on the function of the particular sensor network, sensor nodes may be left unattended for long periods of time. There are three main caveats to unattended sensor nodes: (1) Exposure to physical attacks: Sensors may be deployed in an environment open to adversaries, harsh physical

conditions, bad weather, and so on. The possibility of a sensor to suffer a physical attack in such an environment is much higher than a typical network computer, which is located in a secure place and mainly faces attacks from a network. Therefore, attackers may capture sensor nodes, extract cryptographic keys, modify programming codes, or even replace them with malicious nodes under attacker's control. As a result, the challenge is to build secure networks which can operate correctly even when many nodes have been compromised and behave in a malicious way (2) Maintenance difficulties: Since sensor networks usually operate in areas which are far from the control point, it is almost impossible to detect physical tampering (i.e., through tamper-proof seals) and deal with physical maintenance issues (e.g., battery replacement). An example of such a case is a sensor node used for remote reconnaissance missions behind enemy lines. In such a case, the node may not have any physical contact with friendly forces once deployed. Thus, security mechanisms used for WSN should not require any maintenance (3) Lack of central management: A sensor network is a distributed network without a central management point. Although this increases the vitality of the sensor network, it requires that distributed security schemes be used for WSN security.

In-network processing: Mostly, the dominant traffic pattern in sensor networks is many-to-one, with many sensor nodes communicating sensor readings or network events back to a central base station. In-network processing such as aggregation, duplicate elimination, or data compression is used to make this communication pattern in an energy efficient manner. Since in-network processing requires intermediate nodes to access, modify, and possibly suppress the contents of messages, it is highly unlikely that end-to-end security mechanisms between a sensor node and a base station can be used to guarantee integrity, authenticity, and confidentiality of such messages. Instead, link layer security mechanisms can be used. However, in this case, intermediate nodes will have access to any messages routed through it and can tamper with those messages. Therefore, WSN security schemes should protect also against malicious insiders.

Unreliable communication: The poor quality of wireless channel causes high transmission error rates and more packet loss in sensor networks. If the communication protocol used lacks appropriate error handling, critical security packets like cryptographic keys can be lost. Even if the channel is sufficiently reliable, collisions may occur due to the broadcast nature of wireless sensor networks causing critical packet losses. In addition, multi-hop routing, network congestion and processing at nodes may lead to latency in the network and it may be difficult to achieve synchronization among nodes. This synchronization problem can be critical to sensor network security where security mechanism relies on key distribution.

2.2.3 Attacks against sensor networks

Since they are deployed usually in unprotected areas where several security threats exist, sensor networks are vulnerable to several kinds of attacks. These attacks can be performed in a variety of ways ranging from denial of service attacks to physical attacks. Main attack types that can be launched against wireless sensor networks are covered in this subsection. For quick reference, Table 1, taken from Law and Havinga (2005), can be used which illustrates the potential security threats grouped according to application domains.

DoS Attacks: A DoS attack is "any event that diminishes or eliminates a network's capacity to perform its expected function" (Wood & Stankovic, 2002). DoS attacks on sensor networks range from simple jamming of sensor's communication channel to more sophisticated attacks violating 802.11 MAC protocol or any other layer of the protocol stack. DoS attacks can be very dangerous when sensor networks are used in highly critical and sensitive applications. For instance, a sensor network designed to alert building occupants in the event of a fire could be highly susceptible to a denial of service attack. Even worse, such an attack could result in the deaths of building occupants due to the non-operational fire detection network. Another possible use for wireless sensors

Table 2.1: Attacks against WSN. SA=Service Availability, C=Confidentiality, I=Integrity, A=Authenticity (Law & Havinga, 2005)

Application Domain	**Potential Security Threats**	**Properties violated**			
		SA	**C**	**I**	**A**
Military	• Denial-of-service attacks by means of jamming and/or confusing the networking protocols. • Eavesdropping of classified information. • Supply of misleading information, e.g. enemy movements in the East where in fact they are in the West.	X	X	X	X
Disaster detection and relief	• Supply of misleading information, e.g. bogus disaster warnings, by pranksters, causing huge financial loss as a result of unnecessary large-scale evacuation and deployment of relief equipments.				X
Industry	• Eavesdropping of commercial secrets by business rivals. • Intentional disruption of manufacturing processes as a result of misleading sensor readings caused by disgruntled employees or business spies.	X	X	X	
Agriculture	• The agricultural department might want to deploy WSNs to ensure that farmers do not overuse pesticides or other hazardous chemicals on their crops, but unscrupulous farmers might tamper with the sensor nodes.			X	
Environmental Monitoring	• Suppose government-endorsed environmental sensors are installed near a factory to monitor air/water quality to make sure the factory's emission lies beneath the pollution threshold, however by feeding the sensors with wrong information, the factory allows itself to escape detection and let its polluting emission go unchecked.			X	
Intelligent Buildings	• Biometrics-based access control mechanisms are compromisable if the biometric sensors can be bypassed or fooled. • Token-based access control mechanisms are compromisable if the token authentication protocol is insecure.	X		X	X
Health and Medical	• Providing wrong physiological measurements of a patient to the carer or doctor, a miscreant may cause potentially fatal diagnosis and treatment to be performed on the patient.			X	X
Law enforcement	• If criminals are able to eavesdrop the databases of the police departments, or to misguide the detection of gunshots, or to disrupt the network, public safety will be affected.	X	X	X	X
Transportation	• There is no order in the city when traffic information can no longer be trusted because they can easily be spoofed.			X	X
Space exploration	• Space agencies invest billions into space exploration projects, it is only logical they want to ensure all commands executed on their space probes are authorized, and all collected data encrypted and authenticated.	X	X	X	X

is the monitoring of traffic flows which may include the control of traffic lights. A denial of service attack on such a sensor network could be very costly, particularly on major roads. For this reason, researchers have spent lots of time to identify various types of DoS attacks against WSN.

A very common denial-of-service attack specific to sensor networks is battery power exhaustion. Battery life is the critical parameter for the nodes in a sensor network and many techniques are used to maximize it. In one technique, for example, nodes try to spend most of the time in a sleep mode in which they only turn on the radio receiver, or even the processor, once in a while. In this environment, energy exhaustion attacks are a real threat: without sufficient security, a malicious node could prohibit another node to go back to sleep causing the battery to be drained. Although there are several solutions to mitigate DoS for traditional networks, sensor networks cannot afford the computation overhead needed in implementing those methods. New strategies are being developed for WSN to subvert such attacks. Some of these strategies are the subject of Section 2.2.4.

Attacks against privacy: Since sensor networks provide increased data collection capabilities, threats against privacy of collected data are a relevant concern for WSN. Adversaries may use even seemingly insensitive data to derive sensitive information if they correctly correlate multiple sensor inputs. Sensor networks aggravate the privacy problem because they make large volumes of information easily available through remote access. Hence, adversaries need not be physically present to maintain surveillance and can gather information in a low-risk, anonymous manner. Some of the common attacks against sensor privacy are: (1) Monitoring and eavesdropping: This is the most obvious attack to the privacy. Through listening to the data, the adversary could easily discover the communication contents. When the traffic conveys the control information about the sensor network configuration, which contains potentially more detailed information than accessible through the location server, the eavesdropping can act

effectively against the privacy protection. (2) Traffic analysis: Traffic analysis always combines with the monitoring and eavesdropping. An increase in the number of transmitted packets between certain nodes could signal that a specific sensor has registered activity. Through the analysis on the traffic, some sensors with special roles or activities can be effectively identified.

Attacks against authenticity and integrity: Without proper authentication mechanisms, unauthorized people or devices could request services or data of the unprotected sensor nodes. In many cases these services or data may not be public. Malicious users could also try to join the network undetected by impersonating as some other trusted node. As a trusted node, it will now have access to private data or it can disrupt the normal network operations. As important as authenticity of origin (entity authentication) is the authenticity of data (message authentication or integrity). It should be guaranteed that sensor readings are transferred from sensor nodes to the gateways without any modification. Otherwise, wrong data will be processed resulting in incorrect decisions taken in operation centers and this might have disastrous effects such as directing military troops to the wrong side of the battlefield.

Attacks on WSN routing protocols: For the sake of simplicity, almost none of the sensor network routing protocols do not consider security. As a result, WSN routing protocols are susceptible to many kinds of attacks. Most of these network layer attacks against sensor networks are summarized in Table 2.2 (Abd-El-Barr, Al-Otaibi & Youssef, 2005). Since those attacks are particular to sensor networks, they deserve some more explanation, which is given below.

Spoofed, altered or replayed routing information: The most direct attack to a routing protocol is against the routing data exchanged between nodes. By spoofing, altering, or replaying routing information, malicious nodes may be able to create routing loops, attract or repel network traffic, extend or shorten source

Table 2.2: Routing protocol attacks against WSN (Abd-El-Barr et al., 2005)

Threats	**Description**
Selective forwarding	Malicious node blocks the passage of all or selective messages.
Wormholes	Two malicious nodes in different parts of the network colluding to understate their distance from each other to deceive other nodes.
Sybil	Malicious node illegally claims multiple identities
Sinkhole	Fool large number of nodes that compromised node has the high quality route.
Hello Floods	Malicious node with larger enough transmission power, flood Hello packets to far nodes to deceive them to use false route, to cause confusion to the networks.
Acknowledgement spoofing	Spoof Acknowledgement message to sender with reverse information.
Cloning	Malicious node clones the requests, thus inducing an alternative data flow to itself.

routes, generate false error messages, partition the network, increase end-to-end latency, etc.

Selective forwarding: In a selective forwarding attack, malicious nodes may decline to forward certain messages and simply drop them with the aim that they are not propagated any further. The simplest form of this attack is when a malicious node acts like a black hole and refuses to forward any packet. But, in that case, neighboring nodes may conclude that malicious node has failed and decide to seek another route. A more subtle form of this attack is when an adversary selectively forwards packets. A malicious node interested in suppressing or modifying packets originating from a selected few nodes can reliably forward the remaining traffic and hide suspicion of its wrongdoing.

Sinkhole attack: In a sinkhole attack, the adversary's goal is to lure nearly all the traffic from a particular area through a compromised node, creating a metaphorical sinkhole with the adversary at the center. Because nodes on, or near, the path that packets follow have many opportunities to tamper with application data, sinkhole attacks can enable many other attacks like selective forwarding. Sinkhole attacks typically work by making a compromised node look especially attractive to surrounding nodes with respect to the routing algorithm. For instance, an adversary could spoof or replay an advertisement for an extremely high quality route to a base station.

Sybil attack: In a Sybil attack, a single node presents multiple identities to other nodes in the network (Douceur, 2002). The Sybil attack can significantly decrease the effectiveness of fault-tolerant schemes such as distributed storage, dispersity and multipath routing, and topology maintenance. Replicas, storage partitions, or routes believed to be using disjoint nodes could in actuality be using a single malicious node presenting multiple identities. Sybil attacks also pose an important threat to geographic routing protocols which require nodes to exchange location information with their neighbors. It is reasonable to expect a node to accept only a single set of coordinates from each of its neighbors, but by using the Sybil attack, a malicious node can pretend to be in more than one place at simultaneously.

Wormhole attack: In the wormhole attack (Hu, Perrig & Johnson, 2003), a malicious node tunnels packets received in one part of the network over a low-latency link and replays them in a different part. Wormhole attacks usually involve two distant malicious nodes collaborating to understate distance between them by relaying packets along an out-of-band channel. An adversary located close to a base station can totally disrupt routing by creating a well-placed wormhole. The adversary can fool nodes who are normally multiple hops from a base station that they are only one or two hops away via the wormhole. This creates a sinkhole. Since the malicious node on the other side of the wormhole can artificially provide a high quality route to the base station, all traffic in the

surrounding area will be drawn through it if alternative routes are less attractive. This will most likely be the case when the endpoint of the wormhole is relatively far from a base station.

HELLO flood attack: In a HELLO flood attack (Karlof & Wagner, 2003), an attacker broadcasting HELLO packets to announce itself with large enough transmission power could convince every node in the network that the adversary is its neighbor. For instance, a malicious node advertising a very high-quality route to the base station could cause a large number of nodes to attempt to use this route, but those nodes sufficiently far away from the adversary would be sending packets into a nullity. The network can enter into a state of confusion. Even if a node realizes the link to the adversary is false, it has not too many options because all its neighbors might be attempting to forward packets to the adversary as well. Protocols which depend on localized information exchange between neighbor nodes for topology maintenance or flow control are also subject to this attack.

Acknowledgement spoofing: Many WSN routing algorithms rely on implicit or explicit link layer acknowledgements. Because of the inherent broadcast medium, an adversary can spoof link layer acknowledgments for overheard packets addressed to neighbor nodes. By doing this, it may aim to convince the sender that a weak link is strong or that a dead or disabled node is alive. For instance, a routing protocol may select the next hop in a path based on link reliability. Encouraging a weak or dead link is a way of enforcing such an attack. In this case, since packets sent along the weak/dead links are lost, an adversary can launch a selective forwarding attack using acknowledgement spoofing by reinforcing the target node to transmit packets on those links.

Physical attacks: Sensor networks usually operate in hostile outdoor environments. In such settings, the minimality of the sensors, coupled with the unattended and distributed nature of their deployment make them highly susceptible to physical attacks. Different from many other attacks mentioned

above, physical attacks may destroy sensors permanently and the losses can be irreversible. For example, attackers can extract cryptographic secrets, tamper with the associated circuitry, modify programming in the sensors, or replace them with malicious sensors under the control of the attacker. Recent work has shown that today's standard sensor nodes, the MICA2 motes, can be compromised in about only one minute (Hartung, Balasalle & Han, 2004). While these results are not surprising given that the MICA2 lacks tamper resistant hardware protection, they provide a cautionary note about the speed of a well-trained attacker. If an adversary compromises a sensor node, then the code inside the physical node may be modified. Therefore, as stated previously, security schemes for sensor networks should be resilient to node capture.

2.2.4 Solutions in the literature to the security issues of sensor networks

Similar to the QoS, detailed in previous section, security aspects of wireless sensor networks have recently begun to receive attention compared to other aspects like energy efficiency. Those works on WSN security can broadly be categorized into three main classes such as (1) threats against WSN security, (2) WSN security architectures and (3) key management issues. Most of the current studies of the first category have been covered in the previous subsection which mentions attacks against WSN security. Therefore, a brief overview of various key management protocols and security architectures for sensor networks existing in the literature is presented below.

Security Architecture: The Security Protocols for Sensor Networks (SPINS) project (Perrig, Szewczyk, Wen, Culler & Tygar, 2002) is one of the first studies addressing security needs of sensor networks. It consists of two main parts: an encryption protocol for SmartDust motes called Secure Network Encryption Protocol (SNEP) and a broadcast authentication protocol named micro-Timed, Efficient, Streaming, Loss-tolerant Authentication (μTESLA). In SPINS, each sensor node shares a unique master key with the base station. Other keys required

by the SNEP and the μTESLA protocols are produced from this master key. SNEP is based on Cipher Block Chaining implemented in the Counter mode (CBC-CTR), with the assumption that the initial value of the counter in the sender and receiver is the same. Therefore, the sender increments the counter after sending an encrypted message and the receiver increments it after receiving and decrypting it. In order to achieve authenticated broadcasts, μTESLA uses a time-released key chain. The basic idea here is to utilize the uni-directionality of one-way functions. There are two requirements for proper operation of this protocol: (1) the owner of the key release schedule should have enough storage for all the keys in the key chain, (2) every node in the network should at least be loosely time synchronized with minor drifts. The time-released key chain guarantees that messages can be authenticated only after receiving the appropriate key in the correct time slot.

In Karlof et al. (2004), the authors introduce TinySec, the first fully implemented link layer security architecture for wireless sensor networks. Considering that conventional security protocols tend to be conservative in their security guarantees, typically adding 16-32 bytes of overhead, they conclude that sensor networks cannot afford this luxury with small memories, weak processors, limited energy, and 30 byte packets. So they design TinySec to addresses these extreme resource constraints. They explore the tradeoffs among different cryptographic primitives and use the inherent sensor network limitations to their advantage when choosing parameters to find a sweet spot for security, packet overhead, and resource requirements. TinySec is portable to a variety of hardware and radio platforms. Experimental results for TinySec implemented on a 36 node distributed sensor network application demonstrate that this software based link layer protocol is feasible and efficient, adding less than 10% energy, latency, and bandwidth overhead.

Routing security of sensor networks is considered in Karlof and Wagner (2003). This work proposes security goals for sensor networks, presents classes of attacks

and analyzes the security of well-known sensor network routing protocols and energy-conserving topology maintenance algorithms. The authors conclude that all the routing protocols and algorithms for WSN are insecure. The attacks discussed in this work include bogus routing information, selective forwarding, sinkholes, Sybil, wormholes and HELLO flooding. In order to resolve these security problems of WSN, they suggest potential countermeasures.

Communication security in wireless sensor networks is addressed in Slijepcevic et al. (2002). The approach in this work is to classify the different kinds of data that typically exist in sensor networks and then, to identify possible communication security threats according to this classification. The authors propose a scheme in which each kind of data is secured by a corresponding security mechanism. This multi-tiered security architecture proposed where each mechanism has different resource requirements is expected to achieve efficient resource management.

Law, Dulman, Etalle and Havinga (2003) discusses security aspects of the EYES project, which is about self-organizing, collaborative, energy-efficient sensor networks. The contribution of this work is three-fold. The first one is a survey discussing the dominant issues of energy-security trade-off in network protocol and key management design. This survey is used to depict future research directions for the security framework in EYES. Second, the authors propose an assessment framework based on a system problem that enables application classification. Third, some famous cryptographic methods for typical sensor nodes are compared. This work also investigates resource requirements of symmetric key algorithms RC5 and TEA.

Key Management: There are lots of research studies conducted on key management issue for sensor networks. One of the first of those is Carman, Kruus and Matt (2000) which studies various keying protocols applicable to distributed sensor networks. These protocols are classified under pre-deployed keying, arbitrated protocols, self-enforcing autonomous keying protocols and hybrid

approaches. The authors also present detailed comparisons between various keying protocols in terms of energy consumption.

Eschenauer and Gligor (2002) proposes a random key pre-distribution scheme for sensor network security and operation. The scheme involves selective distribution and revocation of keys to sensor nodes as well as node re-keying without substantial computation and communication capabilities. It relies on probabilistic key sharing among the nodes of a random graph and utilizes simple protocols for shared-key discovery and path-key establishment, and also for key revocation, re-keying and incremental addition of nodes. There are other several works based on random key pre-distribution, some of which are Hwang and Kim (2004) and Liu, Ning and Li (2005).

The LEAP protocol described in Zhu, Setia and Jajodia (2003) takes an approach which utilizes multiple keying mechanisms. Their observation is that no single security requirement correctly fits all types of communication in a wireless sensor network. For this reason, four different keys are used depending on whom the sensor node is communicating with. Sensors are preloaded with an initial key from which further keys can be established. As a security precaution, the initial key can be deleted after its use so as to ensure that a compromised sensor cannot cause addition of compromised nodes to the network.

Liu and Ning (2003) proposes an enhancement to the μTESLA system (Perrig et al., 2002) which uses broadcasting of the key chain commitments instead of μTESLA's unicasting approach. The authors present a series of schemes starting with a simple pre-determination of key chains and finally settling on a multi-level key chain technique. The multi-level key chain scheme uses pre-determination and broadcasting to achieve a scalable key distribution technique that is designed to be resistant to some types of wireless sensor network attacks.

Huang, Cukier, Kobayashi, Liu and Zhang (2003) proposes a hybrid key establishment scheme making use of the difference in computational and energy constraints between a sensor node and the base station. It is assumed that an individual sensor node possesses far less computational power and energy than a base station. Under this assumption, they choose to place the major cryptographic burden on the base station where the resources tend to be greater. On the sensor side, symmetric-key techniques are used instead of asymmetric-key alternatives. The sensor and the base station authenticate based on elliptic curve cryptography. Elliptic curve cryptography is usually used in sensor nodes because of the fact that relatively small key lengths are required to achieve a given level of security.

2.3 SECURITY AND QoS FOR WIRELESS SENSOR NETWORKS

In the previous two sections, WSN applications requiring QoS and challenges in providing QoS for those WSN applications, and similarly, security requirements of WSN applications and challenges in providing security for those WSN applications are covered. However, as stated previously, there are some envisioned sensor network applications that require both QoS guarantees and a certain degree of security. The amount of challenge in such settings is more than twice the one in providing only security or QoS because there is an explicit correlation between QoS and security. In this section, first, some examples of sensor network applications needing both security and QoS are given to illustrate that joint achievement of security and QoS for sensor networks, which is the subject of this thesis work, is an actual necessity. Then, challenges in providing security and QoS in a mutual fashion are explained to show that the problem attempted in this thesis is not trivial and it requires a significant amount of work to propose a solution. Finally, the available works in the literature, whose number is very few indeed and which do not present a complete solution to the described problem, are summarized to demonstrate that there is not an explicit and well-known solution to the described research problem and it is worth studying.

2.3.1 Sensor network applications requiring both QoS and security

The first example of a WSN setting where both QoS guarantees and security is desired can be given as the previously mentioned real-time target tracking application in a battle environment (Younis et al., 2004). In such a scenario, lots of sensors are deployed in the battlefield in order to detect, identify, locate and then to track any object belonging to adversary like an enemy tank. Once this object is detected, for instance, by acoustic motion detection sensors, imaging sensors can be used to identify and locate it. After it is identified and located, video sensors can be turned on to track the trajectory of this object while it is moving. Since this object, let us assume that it is a tank, is moving in real-time, it is very important to send the video data about this tank with minimum possible delay so that its trajectory can be observed accurately in real-time. In addition, sending video data of this tank requires a certain amount of bandwidth for an acceptable image quality. These requirements regarding delay and bandwidth indicate that this application needs some kind of service differentiation to guarantee a certain degree of QoS, which, in turn, will ensure the proper operation of the deployed sensor network. Furthermore, in this scenario, protection of the security of the data is vital to the proper operation. For instance, an undesirable situation can occur when the adversary is able to modify the data in transfer to lead the owner of the WSN into confusion. For example, by modifying the sensed video data, the adversary may fool the control center of the WSN into thinking that the tank is moving in the opposite direction. This may result in directing the troops into the wrong direction whose consequences may be disastrous. Therefore, in this military target tracking WSN application, providing both QoS at network level and security is almost an obligation for proper operation of the network.

The second example is from a health monitoring application. Wireless Body Area Networks (WBAN) are composed of a large number of sensor nodes deployed over or inside the human body (Jovanov, Milenkovic, Otto & Groen, 2005). Those sensors are usually implanted tiny medical devices monitoring and sensing signals

from the human body to provide health data in real-time. For example, in a heart-monitoring application mentioned in Perillo and Heinzelman (2003), sensors measuring blood pressure are used. This WBAN application monitors blood pressure of patients with heart attack risk to see if there are blood pressure abnormalities anywhere on the body. Medical data measured by those sensors, i.e., blood pressure as given in the example, is transmitted to the control center of a medical institution usually via a common RF link, i.e., cellular system. Physicians observe this data in real-time for any possible in-body disorder. If any problem is detected, further data can be requested by physicians to make a diagnosis and even some medical actions can be taken remotely through the actuators also deployed inside the body.

In case of abnormal conditions in patient's health detected at the control center, i.e., a heart attack, more intense monitoring of certain vital signs might be required and more sensors may be required to be active in the current vicinity of the monitored object, i.e., heart. Therefore, an increased spatial resolution might be needed compared to normal cases where data sent by smaller number of active sensors may suffice. Spatial resolution requirements may even be more stringent during remote medical actions which are taken through the actuators deployed inside the body. Consequently, this health monitoring scheme has varying QoS requirements at the application level that can change time to time.

This health monitoring WSN application also needs some degree of security because of the privacy of health data. In European Union countries and in the United States, privacy of medical records is protected by laws like HIPAA (US Congress, 1995) and disclosure of medical information of an individual to third parties is strictly prohibited. Therefore, the privacy and integrity of transmitted health data in such applications should be provided through some security measures. The level of the required security may vary based on the environmental conditions. For example, low security may be enough for indoor environment whereas normal security for outdoor environment and high security for military

cases might be needed. As a result, the health monitoring WSN scheme described here has simultaneous QoS and security requirements, which are both time-varying.

The third example of a WSN setting where service quality and security requirements exist is an environmental surveillance application in which sensor networks are utilized. In Trevis and El-Sheimy (2004), authors consider a real time forest fire detection application to identify and precisely locate the fire site as well as define an efficient approach of intervention. A certain amount of sensors are used to measure the temperature throughout the forest to detect any signs of a fire. Once an abnormally high temperature is measured over a certain area, number of sensors making measurements in proximity of this region is increased in order to provide more specific information about the fire such as the location, the direction and speed of its spreading, and this information is immediately relayed to the control center of the sensor network. Therefore, this application needs QoS guarantees at both network level (minimal latency in order to trigger the fire brigade closest to the fire with the least possible delay) and at application level (increased spatial resolution to provide detailed information about the fire).

The security requirements about this WSN scheme are mostly related to the protection of authenticity and integrity. The sensor nodes which alarm the start of a fire should be authenticated to prevent any false alarms initiated by malicious nodes. Again, the transmitted data should not be able to be modified by anyone not to cause misleading of fire-suppressing teams to areas where there is no sign of a fire. Thus, it can be concluded that this environmental monitoring application is another sensor network setting where security and quality of service is needed at the same time.

As can be seen from the three examples of possible WSN applications given in this subsection, there will be several instances where QoS and security should be provided simultaneously for a sensor network deployment. Thus, it is indeed a

necessity to construct schemes which jointly provide security and QoS for wireless sensor networks. In fact, this thesis study aims to investigate the development of such a scheme.

2.3.2 Challenges in providing both QoS and security for sensor networks

The challenges to provide QoS for sensor networks and the difficulties to make sensor networks secure were given previously. All of these challenges apply when one tries to mutually achieve QoS and security for wireless sensor networks. In addition to those, some extra challenges exist in simultaneously providing security and QoS for WSN applications due to the remarkable interactions between these two concepts. Particularly, degrading effect of security on some QoS parameters complicate the problem of mutual achievement of QoS and security. In this subsection, these additional difficulties related to the security-QoS correlation will be covered. Table 2.3 summarizes the impacts of both concepts on each other and then, these impacts are detailed in the sequel.

Positive Effects of Security on QoS: Availability of a network is the number one requirement to provide QoS guarantees. Simply put, when a network is unavailable meaning that authorized users cannot access the services when needed, the amount of QoS provided by this network can be said to be zero. Although the availability of a network can be harmed by interruption of communication links or failure of some nodes, DoS attacks are the most important threat to service availability. Jamming or energy deprivation attacks described in previous sections may diminish the performance of the network such that expected services cannot be delivered in a healthy way and users/applications receive unpredictable service quality. With a secure system, however, which is resilient to DoS attacks, availability of the network is sustained even under attack, and therefore, QoS can still be guaranteed. As a consequence, security measures preventing DoS attacks contribute to QoS provisioning in sensor networks by

Table 2.3: QoS-security interactions

	Effects of Security on QoS	**Effects of QoS on Security**
Positive Effects	+ Security confirms availability, which is a vital prerequisite for QoS + Security protects QoS related packet headers	+ Finely tuned QoS policies can detect and prevent unusual network traffic caused by attacks + Delay bounds provided by QoS may deny covert timing channels
Negative Effects	- Security incurs longer packets causing bandwidth consumption and increased delays - Security increases computational load on processors leading more latency	- Unprotected QoS labels can leak information about packets in the network - Poorly configured and excessive QoS reservations can deny service to security critical traffic such as key exchange

providing availability. In addition, security may help the protection of QoS related network traffic so that planned QoS provisioning techniques can be utilized.

Negative Effects of Security on QoS: The standard approach to provide security to any system is to use of cryptographic primitives such as message integrity codes (MIC), digital signatures, one-way hash function, etc. The use of cryptography will mainly have two effects on the performance. The first one is due to the increased overhead in the length of the messages sent and the second one is due to the extra computational demands on the processor. The increased message size causes an increase in packet latency, decrease in throughput and an increased use of available bandwidth. And, the computational overhead results in more latency. These adverse effects are detailed below.

The security methods such as MIC or digital signatures append additional bytes at the end of the data packets to be used in verification at receiver's side. These packet overheads are generally in the range of 8-32 bytes and therefore

inconsequential for conventional data networks. However, for sensor networks, which have already little bandwidth available to use, packet size is usually small, i.e., 30 bytes in Berkeley's MICA motes. Therefore, a 8-byte security overhead is almost 25% of the total packet size and has several effects on the QoS of the sensor network. Firstly, longer packets occupy more bandwidth leaving less available bandwidth for QoS constrained traffic. Moreover, large packets circulating in the network cause an increase in the total traffic load present in the sensor network. This increased amount of packet traffic may cause congestions on intermediate sensor nodes which forward packets. This congestion not only decreases the average throughput of the network but also causes increased overall delays due to the higher queuing times of congested nodes. The increase in latency is contributed also by longer transmission times of longer data packets.

Another element causing degradation of QoS parameters is the computational burden put on the sensor node's processors by cryptographic methods like encryption. Although it varies according to the used cipher algorithm, encryption process usually involves lots of arithmetic and logic operations. These operations take thousands of CPU cycles to be completed by the processor. For Giga-Hertz speed processors used in conventional computers like PCs and laptops, these calculations are not too cumbersome. However, the microprocessors used in generic sensor nodes have much limited capacity. For example, MICA2 motes developed in UC Berkeley possess ATmega128L microprocessors operating at 7.3728 MHz with 128 KB program memory and 4KB data memory. Processing a DES encryption on a 29 bytes payload using these processors take almost one second to complete (Guimarães et al., 2005). This is a considerable amount of time and greatly affects the packet latency throughout the network.

Positive Effects of QoS on Security: Though it is not as intuitive as for the case of security's effect on QoS, employing QoS mechanisms have some contributing impacts on network security. One of these positive effects cited in Bhattacharya et al. (2000) is the prevention of covert timing channels. A covert channel is an

unintended communication channel that may be used to transfer data in a manner that violates the security policy. A potential covert channel is a timing channel if its use involves a process that signals information to another process by modulating its own use of system resources (e.g., CPU time) in such a way that this manipulation affects the real response time observed by the second process. If a QoS policy provides certain bounds for delay or latency, covert timing channels cannot be utilized by malicious nodes, which in turn contributes to the security. Another positive effect of QoS on security mentioned in Sakarindr et al. (2005) is the following. A finely tuned QoS policy that provides certain amounts of bandwidth for some defined types of traffic can detect unusual network traffic caused by some malicious nodes who are launching an attack. If QoS and security systems are in cooperation and can share information, QoS system can alert the security system about the existence of this out-of-profile traffic that is not in accordance with defined QoS policies, and thus help the security system to detect and prevent this attack.

Negative Effects of QoS on Security: If QoS mechanisms are poorly configured, they may have detrimental effects on network security. For instance, if critical traffic that is not QoS sensitive is not taken into account while making QoS reservations for services requiring assured delivery, security critical traffic such as key exchange can be denied. Consequently, system security can considerably be destroyed due to excessive use of QoS mechanisms. Similarly, if the packet headers and other traffic used for negotiating QoS agreements is not properly protected, they may be subject to attacks or at least interference by third parties. In this way, information regarding the importance of packets or other classification levels can leak out. This may provide helpful information for malicious nodes that can be used to launch attacks towards critical points in the network. Thus, unprotected QoS traffic may have negative impacts on network security.

The explanations and examples given in this subsection demonstrate that employing security measures on wireless sensor networks positively or adversely

affect the QoS parameters and vice versa. Therefore, this complex QoS-security relationship puts additional challenges for the simultaneous achievement of security and QoS for WSN, and complicates the solution to joint provisioning of QoS and security.

2.3.3 Studies in the literature jointly addressing QoS and security for WSN

A literature survey for QoS considerations and security-related work in the literature for WSN has been presented previously. These works either address security or QoS, but not both. In fact, to the best of the author's knowledge, there is only a single work which aims to provide QoS guarantees while at the same time achieving a certain degree of security, which is also aimed in this thesis. Besides, a few related studies exist which evaluate the effect of applied security mechanisms to the performance of the sensor network. Below are a summary of these related works and the study having a similar purpose to this Ph.D. research.

As mentioned previously, in Karlof et al. (2004), researchers from UC Berkeley, which is the institution where famous MICA motes and TinyOS operating system for WSN are developed, propose a link layer security architecture named TinySec. In addition to the security issues, which are summarized in literature survey part of the previous section, the authors consider the effect of security on network performance. They, in fact, implement the TinySec on TinyOS of actual MICA motes and observe the effects on packet latency, bandwidth and energy consumption. The observed results indicate that implementation of both authentication and encryption increases packet latency by 8% while the corresponding figure for only authentication case is 1.5% increase compared to the case when there is no security implemented. The results also indicate that 5-bytes overhead in packet sizes due to the appended authentication headers cause a 6% lower throughput. As stated before, this work only considers the effect of security on system performance and does not put forward any solution to provide a certain level of QoS under the proposed security scheme.

Another work evaluating the impact of security mechanisms on sensor nodes and the sensor network as a whole is Guimarães et al. (2005). This evaluation has performed the measurements of power consumption (CPU and radio) and memory occupation by encryption algorithms (RC5, RC6, TEA, SkipJack and DES) which have been implemented in an actual sensor network platform. Measurements have shown that the integrity code length added to application messages using some cryptography algorithms has affected packet throughput in an adverse manner implying a small reduction in packet delivery. The authors comment that depending on the application requirements implemented in the network, this reduction may become critical. Packet latency measurements have shown that implemented security mechanisms cause an increase in all cases, i.e., around 1.5 seconds when Skipjack algorithm is used and around 3 seconds when TEA algorithm is used in a 24 hops network. Under these observations, the authors conclude that applications with loose delay requirements may still use encryption and other security services but state that a balance must be drawn between delay, the number of hops and security requirements. Still, however, they do not propose any method to maintain this balance between QoS and security needs.

A very similar study to the two previous works described above is Deng et al. (2003), which also makes an evaluation of the performance of their proposed routing security method. The authors propose a secure routing protocol for wireless sensor networks named INSENS, an INtrusion-tolerant routing protocol for wireless Sensor Networks. Within the context of INSENS, this study evaluates the performance of implementations of RC5 and AES encryption standards, an RC5-based scheme to generate message integrity codes, and an RC5-based generation of one-way sequence numbers. The authors come up with similar observations regarding the increase in packet delay and increased use of available bandwidth. Again, no attention is paid for providing combined security and quality of service for WSN.

The last paper covered in this subsection, is an interesting work that aims to maintain a balance between security and performance in wireless sensor networks (Chigan et al., 2005). The scheme proposed in this paper sets out a framework which is capable of deploying different combinations of security services to satisfy different security needs at different times for different applications. In this way, a suitable security service is tried to be achieved in an adaptive way to get the maximum overall security services against so called "network-performance-services" during the operation of the sensor network. Before the operation of a sensor network system, the offline optimization modules can suggest different set of security provisioning solutions based on the WSN application profile requiring different possible level of trusts. Thus, each node can be provided with information on all possible combinations of protocols with different security services supported. The success of the controller depends on the accuracy of knowledge-base input, which indicates the combined effects of security and performance provided by the system when different sets of protocols are employed. As a result, the adaptive security-provisioning controller aims at maximizing the overall network security service and network performance service. Moreover, the controller is able to switch the protocol that is under attack to some other protocols while still providing similar degree of security and performance. Among all the studies in the literature summarized so far, the problem addressed by this paper is the closest one to the issue that is addressed by this thesis study. The reason is that authors try to maintain a balance between security and some QoS parameters taking the tradeoff between the two into account.

2.4 DISCUSSION

This Ph.D. thesis study is based on and supports the idea that security and quality of service concepts for wireless sensor networks must be taken up together in order to develop a method for satisfying time-varying QoS and security requirements of envisioned sensor network applications. Related literature has

been reviewed with a focus on studies proposing solutions for WSN security and service quality needs.

During the literature review, it is observed that research on QoS issues of sensor networks is relatively less compared to studies involving conventional data communications in sensor networks with no service quality requirements. Similarly, security considerations for wireless sensor networks have recently begun to get attention of researchers. Most of those research studies on WSN security and service quality considered these two concepts independently and there are hardly any works addressing both security and QoS at the same time for wireless sensor networks.

However, studies on the envisioned WSN applications of near future indicate that applications such as real-time target tracking, health monitoring or fire fighting which require accurate and timely transmission of data to derive precise information about the observed phenomenon introduce new challenges. Such sensor network applications need provisioning of certain quality of service parameters at both application and network level such as delay, bandwidth, spatial resolution, coverage and network lifetime to guarantee performance and accuracy for proper operation of the network. In addition, most of these QoS requiring applications are used in critical settings such as military surveillance or health monitoring, and therefore, also have security and privacy requirements. Providing security and QoS in a joint fashion is not a straightforward task not only because of the severe resource limitations of WSN but also due to the correlations between security and QoS. Therefore, simultaneous achievement of security and service quality for wireless sensor networks is a challenging research problem, which, once solved, will help the realization of WSN applications in need of both concepts.

In the current literature, there is not a completely satisfying answer to the proposed research question of this thesis, that is, a combined investigation of

WSN security and WSN QoS to develop a method for achieving both in the most efficient manner. The study presented in Chigan et al. (2005) puts forward a means for maintaining a balance between security and some QoS parameters taking the correlation of two concepts. Yet, the proposed scheme is a very high level framework that does not suggest any concrete security or QoS provisioning methods. Besides, it takes service quality as the network performance composed of attributes such as delay, throughput, etc. In fact, to the best of the author's knowledge, there are not any studies in the literature which present a solution to maintain a sensor network at desired security and application QoS levels such as spatial resolution, coverage and system lifetime. Current studies on application level service quality for WSN consider only providing required QoS levels for a limited number of attributes, but they do not take security into account.

Based on those observations elaborated above, derived from an extensive survey of the pertinent literature, the author of this thesis claims that there is a lack of research regarding the interactions and simultaneous achievement of security and application QoS for wireless sensor networks. Therefore, in this thesis, the correlations and tradeoffs between some application level service quality attributes and security will be investigated, ways for determining the optimal tradeoffs will be sought, and a control strategy for satisfying the requirements of WSN with varying level security and application QoS needs will be proposed. Before presenting this analysis and providing a solution framework, the next chapter marks out the boundaries in which the proposed analysis and control method of this thesis is valid.

CHAPTER 3

ASSUMPTIONS AND SYSTEM MODEL

In this chapter, details about the scope of this thesis study are presented. In order to specify the boundary of the problem that is addressed by this research, the assumptions on the sensor network topology are given and then, the extent of security and service quality concepts as taken in this study are designated. A communication model which details the employed medium access control (MAC) scheme follows. After the specification of an application model as the target of this research study, finally, it will be attempted to give a clear description of the problem to which a solution is offered in the thesis.

3.1. ASSUMPTIONS AND DEFINITIONS

3.1.1 Network topology assumptions

In this thesis, a clustered sensor network topology similar to the one used in the LEACH architecture (Heinzelman, Chandrakasan & Balakrishnan, 2000) is assumed. In this topology, overall network is divided into non-overlapping clusters. In each cluster, there is a gateway/cluster head located in the communication range of all sensors in this cluster. All sensors can send their data directly (in one hop) to their corresponding cluster head. Each cluster head aggregates the data received from sensors and send this aggregated data to the

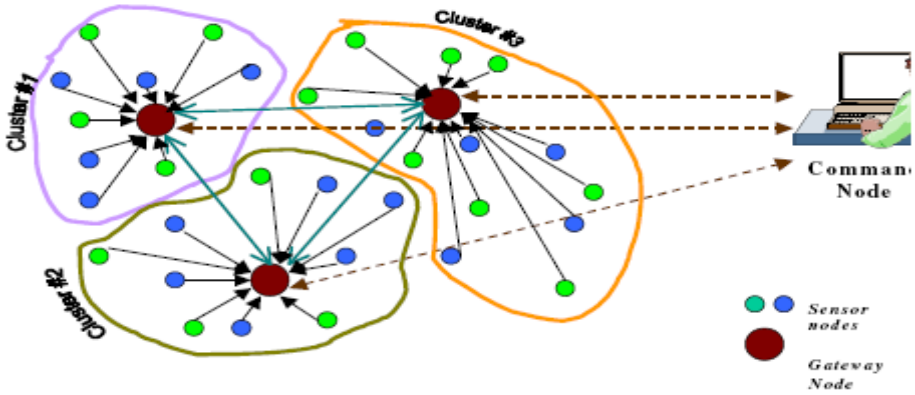


Figure 3.1: Assumed sensor network topology

sink possibly over other cluster heads in a multi-hop fashion. This topology is shown in Figure 3.1.

In this study, only one single cluster of such a network is considered and the analysis and the proposed control methods on security and service quality apply just to this single cluster. Questions such as how clustering is performed, how cluster heads communicate with each other and with the sink, and how data aggregation is performed are all beyond the scope of this thesis.

3.1.2 Quality of service scope

The focus of this research is mostly on application level QoS attributes of wireless sensor networks and network level service quality issues are mostly excluded except for the packet loss due to collisions. In fact, there are four attributes included in the QoS perspective of this thesis study, which are spatial resolution, coverage, network lifetime and packet loss/drop rate due to collisions. The brief definitions of those terms as used throughout this thesis are given in the sequel.

Spatial resolution: The number of sensors that are active in sending data to the cluster head at a specified time duration. Spatial resolution is usually measured by

the number of active sensors per cluster. A higher spatial resolution value usually reflects a higher service quality level since the precision of the useful information that can be constructed by the aggregation of data sent by more sensor nodes is generally higher. Nonetheless, it is better to consider spatial resolution and coverage together as a service quality measure for sensor networks as it is taken in this thesis.

Coverage: Having each and every location in the sensor network field within the sensing range of at least one sensor (*k* sensors for *k-coverage*[1]). Coverage performance is usually expressed by a coverage degree (*k)* and a coverage probability, e.g., *2*-coverage with a *0.9* coverage probability means that every point in the sensor network is covered by at least 2 sensors for 90% of the time during the network operation. High quantities for both *k* and the probability value represent better quality of service and usually keeping the coverage probability above a threshold for a given *k* value is required.

System lifetime: Time duration between the start of operation of a sensor network and the forced end of operation when there remains insufficient number of sensor nodes to collect data due to battery exhaustion of most sensors. A longer system lifetime signifies a better service quality since a certain period of monitoring is often required to correctly capture the temporal variations of the observed phenomenon and insufficient monitoring time might decrease the accuracy and reliability of the collected data.

Packet loss due to collision: This occurs when more than one sensor tries to send data packets to the cluster head at the same time and those packets cannot be received properly by the cluster head. Packet collision is usually measured by the ratio of packets lost due to collision to the successfully received ones during a specified time period. Different from the three QoS attributes given above, packet

[1] More information on coverage and *k*-coverage concepts is given in Appendix A.

collision is generally considered as a network level service quality metric and a lower collision rate reflects better quality.

It is stated above that high spatial resolution, coverage and system lifetime and low packet collision rate values are desirable. However, since those four QoS attributes are in interaction with each other, looking at the levels attained for those attributes individually can be misleading due to the following two reasons. First, some QoS attributes complement each other and having one without the other is useless. This is just the case for spatial resolution and coverage. Though it is true that spatial resolution taken as number of active sensors is a measure of service quality, it cannot not by itself optimally represent the QoS level of the network as assumed in some previous studies such as Kay and Frolik (2004) and Iyer and Kleinrock (2003). Even if the spatial resolution level of a sensor network is sufficiently high, the content of the information that can be produced from the data packets sent by active sensors may not contain enough information to represent the whole network, especially if those data-sending sensors are accumulated in a particular region of the WSN service area. Therefore, in this study, spatial resolution and coverage is considered together and an approximate probabilistic analysis will be presented to depict the mathematical relation between these two QoS attributes.

The second reason for handling the included QoS attributes together is the competitive relationships between some of those attributes with each other and with security, the other WSN requirement addressed by this thesis. Specifically, there are tradeoffs between “security and spatial resolution”, “security and system lifetime”, “spatial resolution and system lifetime” and, “coverage and system lifetime”. In other words, having a higher security level decreases the maximum level of spatial resolution that can be achieved and length of the time that sensor network can remain operational. Similarly, higher spatial resolution and coverage levels cause a drop in the maximum achievable network lifetime. Some of the whys and hows of those inverse relationships will be explained in the proceeding

sections of this thesis. Nevertheless, in this section where the borders of this research study are drawn, it is better to explain the extent of consideration given to those relationships.

Most attention has been paid to the security-spatial resolution correlation in the sense that a mathematical formulation of the tradeoff between those two attributes has been constructed which reveals whether the desired security and spatial resolution values can be achieved under the limited channel capacity. This formulation also allows for the determination of optimal tradeoffs between these two attributes. The relation between security and system lifetime resulting from the increasing power consumption effect of security has also been considered. A new model has not been formulated, as in security-spatial resolution case, to include this interaction between security and power consumption. Instead, existing power consumption models which include the effect of security on battery usage has been employed by this thesis. For the interaction between spatial resolution and network lifetime, another model specifying the effect of increased spatial resolution on lifetime has not been developed. When security level is set to zero, the power consumption model just mentioned includes the effect of having more active sensors on battery usage. In addition, the design feature of the proposed method of this thesis which dynamically selects a reduced number of sensor nodes among all available to remain active is another point where spatial resolution-lifetime correlation has been accounted for. Last but not least, the relationship between coverage and spatial resolution is also considered in the thesis by the inclusion of an approximate probabilistic analysis.

3.1.3 Quality of service assumptions

In this thesis, regarding spatial resolution, it is assumed that there are several spatial resolution levels to meet different requirements. In fact, spatial resolution N of the sensor network cluster can take any positive integer values between $Nmin$ and $Nmax$ which represents the minimum and maximum defined spatial resolution

levels respectively. *Nmin* is the number of active sensors just enough to derive the minimum amount of information required for system functionality and *Nmax* is the number of active sensors needed to derive the best quality information and further increase in *N* does not improve the information quality any further. *Nmin* and *Nmax* are system parameters determined by the sensor network.

As far as the other QoS attribute, coverage, is concerned, area coverage is addressed rather than point or barrier/border coverage and the coverage problem is to keep every point inside a defined geographical region in the sensing range of at least *k* sensor nodes. This coverage definition is in accordance with the requirements of the addressed applications such as environmental monitoring or target tracking that require the monitoring of a certain area. For coverage analysis, sensor network cluster under analysis is divided into *R* virtual sub-regions. Sensors are assumed to be initially deployed in a random but uniform manner over those sub-regions, i.e., initial number of sensors in each sub-region is approximately *Ninitial/R*. Communication range of each sensor spans the whole cluster so that it can send data to the cluster head in a single hop and the sensing range of each sensor spans the sub-region in which the sensor is located. Consequently, full *1*-coverage can be provided when at least one active (data-sending) sensor exists in each of the *R* sub-regions at all times and *k*-coverage requires at least *k* active sensors in each sub-region. The proposed strategy of this paper is not constrained by the shape and size of the sub-regions and assumes that the sensor network cluster is appropriately divided into sub-regions and sensor nodes know in which sub-region they are located, e.g., via GPS. Some sample sensor network clusters with square and hexagonal shapes and their division into sub-regions are shown in Figure 3.2 where it is assumed that the maximum sensing range of each sensor is half of the maximum transmission range.

Regarding system lifetime, it is assumed that the time is divided into equal length discrete time intervals called epochs and total lifetime of the sensor network under analysis is taken as the number of those time epochs until the network dies which

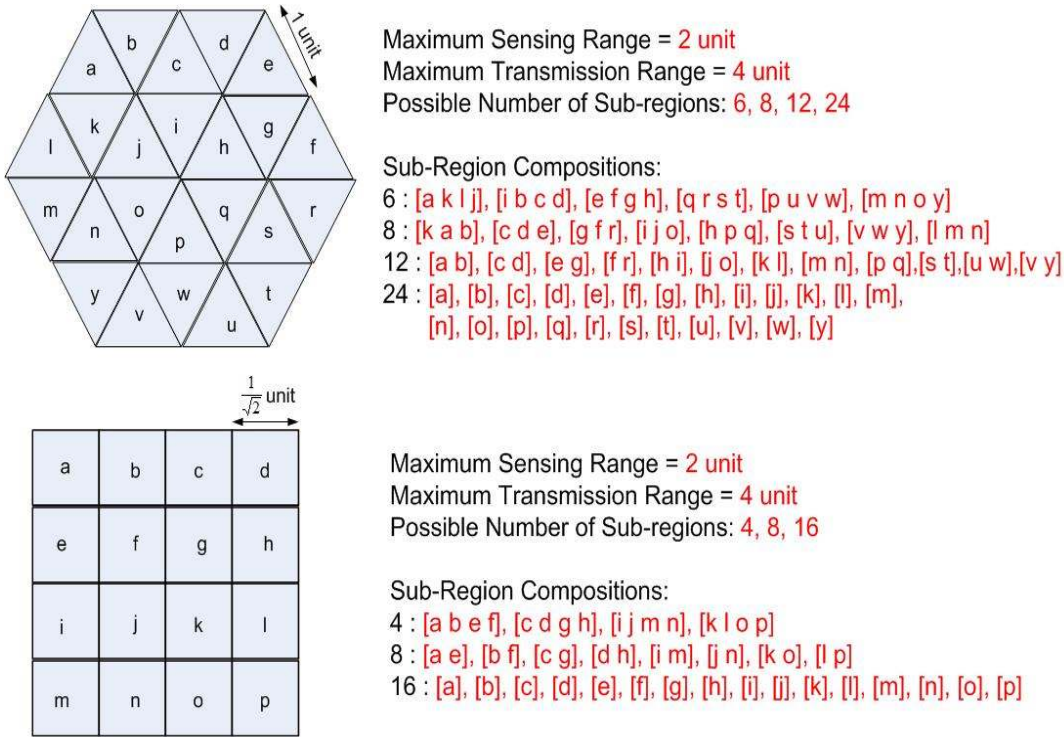


Figure 3.2: Sample WSN service area topologies and their divisions into sub-regions

occurs when the number of operational sensor nodes drop below the pre-defined minimum required spatial resolution value. The only condition for a sensor node to be operational is to have a non-exhausted battery, that is, it is assumed that sensor deaths occur only due to battery exhaustion and other reasons such as physical damage that can make sensor nodes non-operational is excluded. Details of when and at which rate a sensor node consumes its battery are provided by the power consumption model given in Chapter 4.

In order to minimize packet loss due to collisions, the fourth QoS attribute included by this thesis, a centralized Medium Access Control (MAC) scheme is assumed rather than purely contention-based distributed schemes such as ALOHA (Abramson, 1970) or CSMA (Carrier Sense Multiple Access) (IEEE 802.11

Working Group, 2007). Further information on the assumed MAC scheme is presented in the following subsections.

3.1.4 Security assumptions and the threat model

In this study, only the security of sensor to cluster head communication is considered and it is assumed that communications from cluster head to sensors or to the sink are secured by other means. The assumed sensor network environment is such that adversaries constitute a threat mainly on the integrity and authenticity of the packets transmitted rather than the confidentiality. There are several justifications for excluding the possible attacks against confidentiality from the security scope of this thesis study. The first one is that the security requirements of the WSN settings that this research addresses, which will be mentioned in the application model, put more emphasis on the integrity and authentication of packets rather than confidentiality. For instance, in a fire fighting WSN application, in order to prevent false alarms, it is required that only authentic entities can send messages about a possible fire and those messages cannot be modified. Confidentiality of the message which contains information about the location or state of the fire is usually not so much of an importance when the urgency of fire situations and the relative potential loss due to the compromise of confidentiality in such urgent situations are considered. Similarly, for a target tracking application also included in the application model of this thesis, integrity and authentication of messages sent by the sensors to the cluster head are comparably more important than the secrecy because an adversary will try to mislead the WSN application by sending false messages or modifying packets in transit. Besides, in most cases, the target tracked usually belongs to the adversary and since the adversary already has knowledge about the target, a requirement for the protection of the confidentiality is less meaningful compared to the protection of integrity and authenticity.

Nonetheless, there might be cases when confidentiality is of considerable importance in addition to the integrity and authentication. For such cases, security protocols operating on the upper layers than the link layer can be used without any problem. In other words, not turning on the encryption feature of the TinySec link layer security protocol does not mean that confidentiality cannot be provided at all in higher levels of the WSN protocol stack or later when forwarding the consolidated data from the cluster heads to the sink. In fact, the individual packets sent by sensor nodes to the cluster head contain local data which usually does not hold too much confidential information. Since the main communication pattern that is addressed in this thesis is on the transmission of this local data, it can be assumed that encryption can be applied at later stages when cluster heads transfer the aggregated data in each cluster to the sink.

Due to the reasons given in the paragraph above, this thesis study assumes that the preservation of integrity and authenticity of a packet constitutes an acceptable security measure. Message authentication/integrity codes are used to prevent malicious sensor nodes from inserting spoofed data or modifying data in transit. There are multiple security levels defined for the sensor network setting of the thesis. Each security level is associated with a different length message integrity code (MIC). A security level is represented by S and $S=0$ corresponds to the lowest security level where no MIC is used and $S=Smax$ corresponds to the highest security level where longest MIC is used. S can take any positive integer values between 0 and $Smax$.

In order to provide this multi-level security approach adopted in the thesis, TinySec-Auth option of the TinySec (Karlof et al., 2004) sensor network security protocol is utilized. TinySec-Auth appends message integrity codes (MICs) to the end of messages to provide message integrity and authentication. Those MICs are calculated by encryption of the message digest which is the output of the one way hash function applied to the message body. Since this encryption is performed by the use of a symmetric key shared only between the sender and the receiver, the

strength of the authentication and integrity provided by message integrity codes is solely dependent on the secrecy of this shared key. This makes the key distribution an important process for the level of security provided by MICs.

TinySec by itself does not propose any key distribution or management scheme and it just suggests the use of a single pre-distributed key between all senders and receivers of the same group. In the assumed topology of the thesis, this maps to the use of a single key shared by all sensor nodes and the cluster head. Although the management of such a key distribution scheme is fairly simple, for some hostile environments where there is a risk of malicious behavior of even the originally deployed sensor nodes, a more finely grained key management scheme is needed. For example, in the case that all sensor nodes share the same key, a sensor node knowing this single key can send messages to the cluster head spoofing the identity of the any other node in the same cluster and this spoofing attempt cannot be detected because the MIC appended to the message is computed using the correct key. Thus, single key deployment of the TinySec protocol provides protection only against the outside adversaries not knowing the shared key. However, once this key is learnt by an unauthorized third party, due to, for instance, physical capturing of even a single sensor node, this malicious party can inject unauthorized messages into the network or it can alter the messages in transit.

In order to provide resilience against the security weakness mentioned in the paragraph above, this thesis assumes that different keys for each sensor node to be shared with only the cluster head are pre-distributed to all the nodes and cluster head as proposed in Eschenauer and Gligor (2002). In this case, sensor nodes can send messages to the cluster head only on their behalf and cannot spoof the identity of any other node as long as nodes keep their shared keys secret. Thus,

use of separate keys for each sensor node increases the security of the network by providing a stronger authentication compared to the previous case[2].

The final assumption on security is that all sensors communicate at the same security level during a frame duration. It must also be noted that length of the security overhead per packet should either be given or can be computable for all security levels.

3.2 SYSTEM MODEL

In this section, a system model describing the properties of the underlying communication channel and explaining the characteristics of target sensor network applications that can utilize the proposed solution is presented. First, the communication model including the details of the employed MAC scheme is given. Then, the application model is explained.

3.2.1 Communication model

A perfect communication channel is assumed where all transmissions occur without any error or data loss. According to the assumed network topology, i.e., a single cluster, data traffic occurs in a many-to-one and one-hop fashion. Therefore, consideration on layer 3 routing is not needed. Specification of the medium access control scheme used in this thesis suffices to describe the communication model.

As stated previously, in order to utilize the existence of a central entity, i.e., the cluster head, a centralized MAC scheme is adopted rather than contention-based decentralized schemes often used in several kinds of wireless networks. Since fixed-assignment based MAC strategies like pure TDMA may cause channel

[2] Yet, neither of these keying schemes used with TinySec can provide a capability to regulate the access attempts of successfully authenticated sensor nodes. In Appendix E, a role-based access control scheme is proposed that can be implemented by the proposed method of this thesis.

inefficiency due to the empty slots assigned to non-transmitting sensors, the assumed MAC scheme here is a demand-based one. Among demand-based MAC methods, a reservation-based one is preferred in which time is divided into frames each of which is composed of two main parts named as reservation period and data transmission period. Reservation period is where stations requesting to transmit contend for an empty mini slot. Then, in data transmission period which is composed of multiple data slots, stations that have accessed an empty mini slot during the reservation period send their data in their assigned data slot. There are several reservation-based MAC schemes proposed for sensor networks such as DR-TDMA (Frigon, Chan & Leung, 2001) and TRACE[3] (Tavli & Heinzelman, 2003). In this study, a version of TRACE that is modified to suit the specific needs of this research is used. Detailed explanation of this MAC scheme will be presented shortly, but before that some assumptions on the communication model are presented below.

In this thesis, it is assumed that the total channel capacity of the WSN cluster under question is limited and this limit is known in advance as bits per second. Sensors send their data in their assigned slot of the MAC frame. Each sensor is assigned one and only one data slot for each frame and in each data slot, a sensor transmits only one single packet. TinyOS type packets composed of a data part and an overhead part are assumed. Data part has constant length. Overhead part has variable length due to the security overhead which increases as security level increases. This varying length security overhead causes the overall packet to be of varying length. Therefore, the length of the data slot assigned for a sensor's packet should also have non-constant length to accommodate the packets of different security levels. However, total frame time and total data transmission period in each frame has constant duration in accordance with the upper bound of the channel capacity. This means that the number of data slots that can fit in a single frame is upper bounded. This upper bound is equal to the duration of data

[3] Information on MAC schemes suitable for sensor networks and a discussion on why TRACE has been chosen for this thesis can be found in Appendix B.

transmission period in a frame divided by duration of a single data slot. Therefore, the number of active sensors sending data to the cluster head in one frame duration has also the same upper bound. Because the duration of a data slot varies with security level, this limit in number of active sensors is correlated with the security level. The correlation between spatial resolution and security resulting from the capacity limits of the underlying communication channel is an important point taken into consideration in this thesis.

The symbolic representation for the frame format of the TRACE-based MAC scheme used in this thesis is given in Figure 3.3 for two frames. In fact, several slightly modified versions of this MAC scheme are used throughout this study to suit the needs of different methods proposed. Nevertheless, the basics are the same and modified parts will be mentioned when needed. One visible difference of this MAC frame from the original TRACE protocol is the variable length data slot durations. In addition, there is another difference as far as the type of the information sent in the Header part is concerned. The details of this difference will shortly be given but first, the basic operation of this MAC scheme is explained below.

Each frame consists of two sub-frames: a control sub-frame (reservation period) and a data sub-frame (data transmission period). The control sub-frame consists of a beacon message, a contention slot, a header message, and an information summarization (IS) slot. Beacon message is used to synchronize all sensor nodes at the beginning of each frame. Contention slot consists of several mini slots and nodes that have data to send for this frame randomly choose one of these mini slots to transmit their request. If the contention is successful (i.e., no two sensors choose the same mini slot), the contending sensor node is granted a data slot in the data sub-frame. The controller, i.e., cluster head, then transmits the Header, which includes the data transmission schedule for the current frame. Unlike the original TRACE scheme, the Header also includes two more pieces of information, which are the information on the current security level and information regarding the

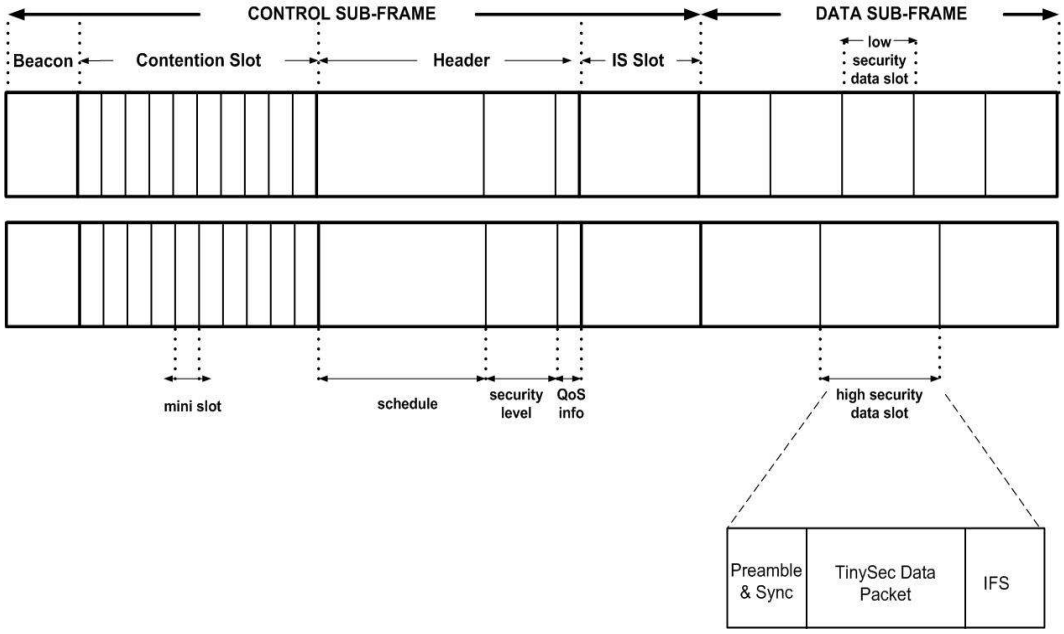


Figure 3.3: The frame format of the MAC scheme (2 frames are shown)

comparison of the current and required QoS (spatial resolution) levels. The IS slot follows the Header slot and is used for partitioning of the network. Control sub-frame ends with the IS slot and the data sub-frame begins.

As stated before, the other difference from the original TRACE protocol is that data sub-frame is broken into variable length data slots. These data slots have variable lengths to accommodate different length packets of different security levels. Actually, the security level of all nodes during a frame duration is assumed to be the same, thus, all data slot lengths of the same frame are equal. However, the security levels of different frames may not be the same and therefore, the data slot lengths of different frames may vary (see Figure 3.3). The data slot lengths required to accommodate a packet at security level S is denoted by Ds. This Ds value is not the same as the packet length PLs at security level S because of the overheads required at each data slot. These overheads are usually due to preamble and synchronization bits and IFS (inter frame space). As in the original TRACE

protocol, a total of 6 bytes overhead for each data slot is assumed (four bytes for packet header and two bytes for the IFS guard band).

3.2.2 Application Model

The target applications to which the proposed method of this thesis can be applied are sensor network settings which have simultaneous security and service quality requirements. Those target applications demand security in the sense that data integrity and authentication are more important than the data confidentiality. If the confidentiality of the data is equally important, this must be provided by other means than the proposed method of this study, e.g., encrypting at higher layers of the protocol stack or encrypting the aggregated data at the cluster heads as explained in Section 3.1.4. Quality of service requirements of the target applications must be at application level rather than the network level. For instance, WSN deployments which need spatial precision of the collected data, a certain level of coverage guarantee and maximal monitoring time of the environment can utilize the proposed method of the thesis. The proposed method can still be applied in sensor network settings that need both application and network QoS but additional strategies must be employed to satisfy the network QoS requirements since the focus of this thesis is on application QoS. In fact, since the proposed QoS control strategy of the thesis operates at the link layer and deals only with intra-cluster communications, solutions aiming to provide network QoS which usually addresses the routing problem and considers the inter-cluster communications can transparently be employed together with the proposed method of this thesis. Some example wireless sensor network applications, QoS and security requirements of which fall inside the scope of this study and therefore that can utilize the proposed control strategy of the thesis are presented in the sequel.

The first target WSN setting is the tracking applications as presented in Pattem et al. (2003). In this scenario, sensors are scattered over a certain region first to detect, then identify and finally to track a moving target. In this sensor network application, sensors are equipped with detectors to realize the presence of a target in their proximity. The aim is to totally cover the application area and to turn on enough number of sensor nodes to take measurements for ensuring adequate accuracy of the target's position. Since activating minimum number of sensor nodes and allowing others to go into the power saving mode will provide battery optimization, the number of active (sensing and data sending) nodes is increased or decreased according to the data accuracy needs. In other words, until a target is detected, the smallest possible number of sensors is activated meaning a low spatial resolution requirement. But once the target is detected, QoS, i.e., spatial resolution, requirement is increased for identification and a further rise is needed for tracking of the identified target. In addition to those possibly time-varying application QoS requirements such as spatial resolution, coverage and system lifetime, this application requires the preservation of the integrity of transmitted messages about the target's identity and position to prevent modification of this data by adversaries which aim to mislead the application. Also, the packets sent by sensor nodes must be authenticated to prohibit any falsifying information inserted into the network by malicious third parties. The level of these security needs can be time-varying as well depending on the conditions, e.g. realizing that there is an ongoing attack to insert false data into the network might require the use of longer message integrity codes. Thus, this target tracking WSN application with security and QoS requirements fitting into the scope of the thesis can benefit from the proposed method. If this application also needs confidentiality and network QoS provisioning such as delay and jitter, which cannot be provided by this study, additional methods can be employed for encryption and delay & jitter minimization while transferring the aggregated data by the cluster head to the sinks.

The second WSN setting where service quality and security requirements are in accordance with the thesis scope is an environmental surveillance application in which sensor networks are utilized. In Trevis and El-Sheimy (2004), authors consider a forest fire detection application to identify and precisely locate the fire site as well as to define an efficient approach of intervention. A certain number of sensors is used to cover and measure the temperature throughout the forest to detect any signs of a fire. Once an abnormally high temperature is measured over a certain area, number of sensors making measurements in proximity of this region is increased in order to provide more specific information about the fire such as the location, the direction and speed of its spreading. Therefore, this application needs QoS guarantees at application level (increased spatial resolution to provide detailed information about the fire). The security requirements about this WSN scheme are mostly related to the protection of authenticity and integrity. The sensor nodes which alarm the start of a fire should be authenticated to prevent any false alarms initiated by malicious nodes. Again, the transmitted data should not be able to be modified by anyone not to cause misleading of fire-suppressing teams to areas where there is no sign of a fire. Thus, it can be concluded that this environmental monitoring application is another sensor network setting which can utilize the control method proposed by this thesis.

In addition to those two applications which are given as the target applications of this thesis, the proposed method can be applied to WSN settings which require application level QoS but no security at all. If the desired security level is set to zero at all times, the proposed strategy of this thesis can satisfy the service quality requirements of such settings at application level. One example of these applications is habitat or wild-life monitoring such as in Mainwaring et al. (2002) where sensors are deployed on the famous Great Duck Island to monitor the microclimates in and around nesting burrows of seabird species called Leach's Storm Petrel. Deployed sensors of this habitat monitoring application continuously collect data about the temperature, humidity and pressure of the surrounding environment. As pointed out in Mainwaring et al. (2002), in such

environmental surveillance applications, the ultimate goal is data collection to derive precise information about the observed phenomenon. Therefore, the application defines a minimum value for the accuracy and spatial precision of the collected data as well as efficient battery consumption of sensor nodes to maximize the monitoring time. Thus, this is another example of a WSN setting with application level QoS requirements that can also utilize the findings of this thesis study. Note that, in such applications, usually there are no network level service quality requirements and no strict confidentiality requirements and therefore, only the proposed method of thesis can be employed to satisfy the requirements of such applications without any need for other additional methods.

3.3 PROBLEM FORMULATION

Under the assumptions and constraints given in this section, the problem to which a solution is presented here is the following: To control a cluster-based sensor network in such a way that time-varying security and QoS requirements are fulfilled during the entire operation. In other words, there are five main objectives: (1) to keep enough number of sensor nodes active (data sending) to attain the desired spatial resolution level, (2) to have these active sensors communicate at the required security level, (3) to maximize network lifetime by having active sensors periodically power down and inactive ones power up for a balanced energy dissipation, (4) to provide full coverage by having at least one (k for k-coverage) sensor taking measurements in each geographical region and, (5) to minimize the packet loss resulting from collisions.

The problem which comprises of only part (1) and (3) above, i.e. controlling spatial resolution and maximizing network life time has already been solved in previous studies Kay and Frolik (2004) and Iyer and Kleinrock (2003). In this thesis, first, security (2) is appended as an additional parameter to this solution and also it is allowed that both desired spatial resolution and security requirements can change in time as needed. Moreover, the QoS concepts used in Kay and Frolik

(2004) and Iyer and Kleinrock (2003) are extended to include coverage (4) and packet drops due to collisions (5). Thus, a QoS and security control strategy to achieve all five objectives above is proposed.

The intended contribution of this study is to propose a novel control strategy to achieve all five security and service quality objectives without having the drawbacks of the ACK-based method proposed in Kay and Frolik (2004). To elaborate, a new QoS and security control strategy for wireless sensor networks will be designed to provide closer values to the required spatial resolution levels, longer network lifetime, better coverage and also to protect message integrity and authentication. Therefore, this thesis aims to propose a novel method to enhance three of the QoS attributes, namely, spatial resolution, network lifetime and coverage, compared to the values achieved in previous studies and to provide security as well.

Finally, this thesis study intends to construct a mathematical model to formulate the relationship between security and spatial resolution in order to depict the tradeoff between these two attributes and to propose a computationally efficient heuristic algorithm to solve an optimization problem whose solution yields the best tradeoff between security and spatial resolution. In addition, an approximate probabilistic analysis is also aimed to be presented to relate coverage and spatial resolution.

CHAPTER 4

SECURITY AND QoS RELATIONSHIP

In the previous chapter, it was stated that the main problem of this thesis is to devise a method for satisfying the time-varying security and QoS requirements of wireless sensor networks. However, there are some correlations between security and some service quality attributes which should be taken into consideration while designing the QoS-security control strategy proposed by this research study.

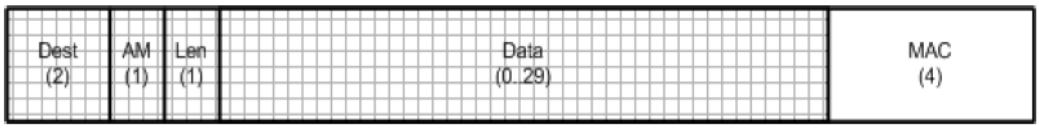
One such correlation is between security and spatial resolution resulting from the fact that they both use up the same scarce resource, which is channel capacity. Therefore, there might be cases where the requested security and spatial resolution levels exceed the available channel capacity, hence, cannot be supported by the network. In the first section of this chapter, this tradeoff between those two concepts is analyzed. Then, in the following section, an optimization problem is formulated to determine the best tradeoff between security and spatial resolution and a computationally efficient solution for the optimization problem is developed. The third section discusses the interaction of security with another QoS parameter, network lifetime. This final section presents a model describing the effect of security on power consumption of individual sensor nodes, which eventually effects the overall network lifetime since deaths of sensors due to battery exhaustion finally cause the sensor network become non-operational.

4.1. CORRELATION OF SECURITY AND SPATIAL RESOLUTION

As mentioned in the previous sections, security adds some overhead bits due to the message integrity code appended to the end of the packet transmitted by sensor nodes. This increase in packet size causes an increase in the data slot duration of the MAC frame required to transmit this packet. Since the duration of total data transmission period is fixed, increase in durations of individual data slots result in a decrease in number of data slots that can be accommodated in the data transmission period of a single MAC frame. Because each sensor can transmit during only one data slot of each frame, number of non-empty data slots in a frame is equal to the number of sensors that transmit in that frame, which is the spatial resolution definition. Thus, this fact proves that an increase in the employed security level during a specified time duration result in a decrease in the spatial resolution for that time duration.

In order to mathematically formulate this inverse relationship between security and spatial resolution, it is needed to specify the effect of security on packet length and data slot duration by determining the data slot durations corresponding to each security level. Then, it can be computed that how many of these data slots can be accommodated in a MAC frame for each security level. It is shown in the sequel that how this computation is performed.

In Chapter 3, it was stated that TinyOS packet format is assumed. This type of packet has a total length of 36 bytes with 29 bytes of data, 5 bytes of communication overhead and 2 bytes of CRC. It is further assumed that the security concept of this thesis includes preservation of integrity and authenticity of packets by message integrity codes (MIC). A security method suitable for such a security definition is the authentication only (TinySec-Auth) option of TinySec sensor network security protocol proposed in Karlof et al. (2004). According to the format of the TinySec-Auth packet given in the cited work, total length of a packet with 4-byte MIC appended is 37 bytes because *Grp* (1 byte) and



(b) TinySec-Auth packet format

(c) TinyOS packet format

Figure 4.1: TinyOS and TinySec packet formats (Karlof et al., 2004)

CRC (2bytes) fields are no more needed. Thus, security adds up only a 1 byte (4-2-1=1) overhead to the data packet when TinySec-Auth is used as shown in Figure 4.1.

TinySec protocol assumes only a single option for the length of MIC used as 4 bytes and this is not in accordance with the multi-level security perspective of this thesis study. Yet, there is no reason for not to extend the TinySec-Auth to allow multiple MIC length selections. In fact, the *security suite* feature of IEEE 802.15.4 specification (IEEE 802.15 Working Group, 2003) is an example of this multi-mode security approach. IEEE 802.15.4 describes wireless and media access protocols for personal area networking devices and these protocols are commonly used by sensor network community. This specification defines eight security suites by the properties they offer: no security, encryption only (AES-CTR), authentication only (AES-CBC-MAC) and both encryption and authentication. Among these eight security suite options of IEEE 802.15.4, there are three authentication-only options with MIC sizes of 4, 8 and 16 bytes. Adopting this approach in this thesis, four security levels are considered, one of which is no security and others include 4, 8 and 16 byte MICs. Knowing that a TinyOS packet with no MIC is 36 bytes and a TinySec-Auth packet with 4-byte MIC is 37 bytes,

Table 4.1: Packet lengths corresponding to different security levels

Security level *S*	**Description**	**Packet length *PLs***
0	No security	36 bytes
1	4 bytes MIC	37 bytes
2	8 bytes MIC	41 bytes
3	16 bytes MIC	49 bytes

then the relationship between security level *S* and the corresponding packet length *PLs* is as given in Table 4.1.

Before proceeding, an important point about the generality of this research study should be noted. The proposed QoS and security control strategy of this thesis is neither coupled to any of the above mentioned security methods nor supports only 4 security levels. As long as the packet lengths *PLs* corresponding to each security level *S* is known, the proposed strategy is applicable.

As previously mentioned in the communication model, data slot length required to accommodate a packet at security level *S* is denoted by *Ds* and this *Ds* value is not the same as the packet length *PLs* due to the overheads required at each data slot. These overheads usually result from preamble and synchronization bits and IFS (inter frame space). As in the original TRACE protocol, a total of 6 bytes overhead for each data slot is assumed (4 bytes for packet header and 2 bytes for IFS guard band). So, *Ds* values corresponding to the *PLs* values of Table 4.1 are D_0*=42 bytes,* D_1*=43 bytes,* D_2*=47 bytes* and D_3*=55 bytes*.

Now, what remains is to compute how many of these data slots can be accommodated in the fixed data sub-frame duration of a single frame.

Representing the constant data sub-frame length with *DSF* and the maximum number of data slots that can fit into a frame at security level *S* with $N_{s,max}$, the following inequality should hold not to exceed the channel capacity: $N_{s,max} \leq DSF/Ds$. Since the constant value *DSF* is known and all *Ds* values are computed, maximum spatial resolution that can be supported, $N_{s,max}$, can be computed for any security level *S*. And, the inequality $N_{s,max} \leq DSF/Ds$ is the mathematical relationship[4] between security and spatial resolution that is sought. Representing the security level requirement by S^* and spatial resolution requirement by N^*, it can easily be checked whether a required security-spatial resolution pair (S^*, N^*) is supported by substituting these values into above inequality. If $N^* \leq DSF/Ds^*$, the required levels are supported, otherwise they are not.

If network parameters of Table 4.1 is used together with a *DSF* value of 1050 bytes (coming from an assumption of 25 data slots can fit in a frame at no security, i.e., $N_{0,max} = 25$ and *DSF*=25x42=1050), then remaining $N_{s,max}$ values can be determined as follows: $N_{1,max} \leq 1050/43 \approx 24$, $N_{2,max} \leq 1050/47 \approx 22$, $N_{3,max} \leq 1050/55 \approx 19$ (and it is already known that $N_{0,max} = 25$). Then, for example, during the operation of the sensor network, cluster head can check that requirements of $(S^*, N^*) = (1, 23), (3, 15)$ and $(2,20)$ are supported whereas $(1, 25), (3, 20)$ and $(2,23)$ are not. In Figure 4.2, the concept of these supported and unsupported *(S,N)* levels are visually illustrated for a network where the defined security range is [0,3] and spatial resolution range is [15,30]. In the figure, supported *(S,N)* tuples are represented by squares and unsupported ones with circles.

For such cases where the required security-resolution pairs are not supported, the supported values that are closest to the requirements should be determined. This is not a straightforward task because there usually exist more than one supported

[4] Although this relationship was determined for the specific TRACE-based MAC scheme, similar relationships can be determined for other MAC strategies intuitively in a similar fashion. For instance, for FDMA-like methods, $N_{s,max}$ should be less than the total available bandwidth dedicated to data transmission divided by the bandwidth assigned to each sensor.

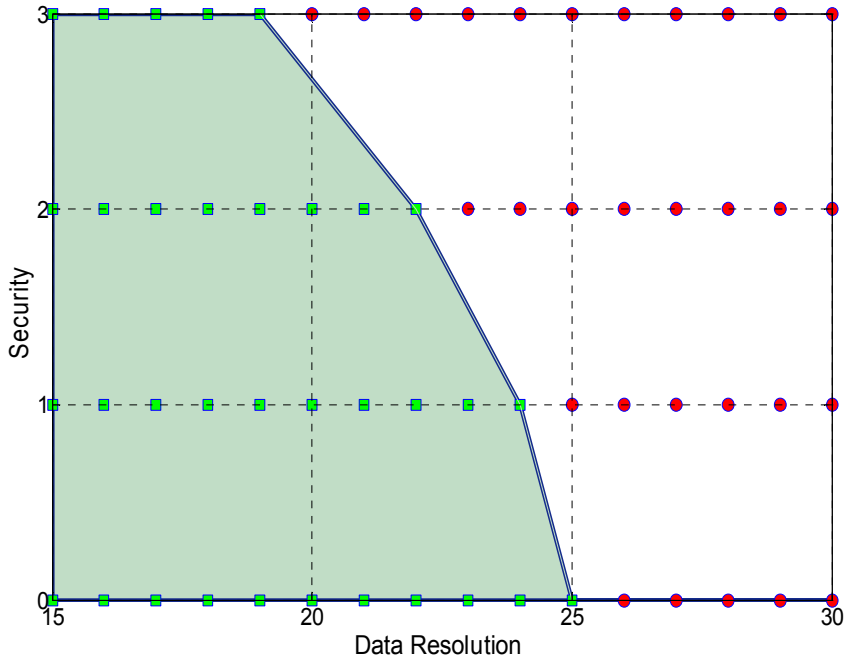


Figure 4.2: Security versus spatial resolution

security-spatial resolution pair. Taking the example case for an unsupported requirement of *(S*, N*) = (3, 25),* should one sacrifice security by choosing the supported pair of *(0, 25)*, or sacrifice spatial resolution and choose *(3, 19)*, or else sacrifice from both sides and choose *(2, 22)*? Next section explains how this thesis will deal with such unsupported *(S*, N*)* requirements.

4.2. BEST TRADEOFF FOR SECURITY AND SPATIAL RESOLUTION

Determination of the best tradeoff for unsupported *(S*, N*)* requirements is a resource allocation problem where the scarce resource is channel capacity and competing factors are security and spatial resolution. Such resource allocation problems are optimization problems which are studied in several works in the literature. One of such studies is Lee, Lehoczky, Rajkumar and Siewiorek (1999) whose problem modeling fits into the scope of this thesis. In Lee et al. (1999), authors present a framework for optimally allocating finite resources to satisfy the

QoS requirements of multiple applications along multiple QoS dimensions. As an example problem, they mention allocation of bandwidth among several QoS dimensions such as cryptographic security, packet loss, video picture color depth, audio sampling rate, etc. of various applications such as web, ftp, video conferencing, etc. Their proposed solution is based on the maximization of an aggregate system utility function[5] which is composed of the utility/benefit brought by all QoS dimensions of all applications.

The main approach of Lee et al. (1999) based on finding the values which maximize an aggregate utility function will be employed by this thesis to determine the best security-resolution tradeoff. Two individual utility functions for security and spatial resolution represented as $Us(S)$ and $U_N(N)$, respectively are assumed in this thesis. These functions map the security and spatial resolution values in their defined range to a positive utility value representing the benefit provided by the corresponding security or spatial resolution value. Then, the aggregate utility function to be maximized, which is a weighted sum of the individual utility functions reflecting the marginal benefits of each factor competing for the scarce resource is the overall utility function to be maximized, which is equal to $U = Ws.Us(S) + W_N.U_N(N)$.

Thus, in order to determine the optimal supported security and spatial resolution values when requirements cannot be satisfied, the optimization problem given in Formula 1 below should be solved for S and N.

$$\begin{aligned} &\text{Maximize} && U = Ws.Us(S) + W_N.U_N(N) \\ &\text{Subject to} && N \leq DSF/Ds, \\ & && Nmin \leq N \leq N^*, \\ & && Smin \leq S \leq S^* \end{aligned} \qquad \textbf{(1)}$$

[5] More information on utility functions and utility maximization approach of Lee et al. (1999) to solve resource allocation problems is given in Appendix C.

Here, *Nmin* and *Smin* stands for the minimum required levels for spatial resolution and security. These are different than actual requirements represented by *N** and *S** and used for preventing the sensor network from operating at undesirably low security and QoS levels. If the minimum requirements *Nmin* and *Smin* cannot be supported, the sensor network ceases its operation until the minimum requirements can be satisfied.

Since there are only two unknowns (*N* and *S*) and the possible values for these unknowns are both upper and lower bounded, the above optimization problem can be solved by enumeration of the whole solution space. So, given an unsupported *(S*,N*)* pair, one can find the optimal supported pair *(S',N')* by trying all possible *(S,N)* combinations in the range $Nmin \leq N \leq N^*$ and $Smin \leq S \leq S^*$ and pick the one yielding the maximal *U* value which also satisfies the condition $N \leq DSF/Ds$. Yet, for cases where this brute force approach is not feasible, this thesis proposes a computationally efficient heuristic for the solution of the problem in Equation 1.

Before presenting the details of this heuristic, however, an analysis depicting the effect of weights W_S and W_N on the optimal solution is given below.

The individual security and spatial resolution utility functions shown in Figure 4.3 are taken for this analysis. As seen from the figure, both functions are non-decreasing indicating a utility gain for increased levels of security and resolution. Specifically, for security utility function $U_S(S)$, decreasing levels of increments are taken as security level increases. The reason is that the marginal advantage of using a higher security level usually gets lower for higher security levels. For instance, security benefit increase for using a 128 bit encryption instead of 64 is usually higher than the security benefit increase of using 256 bit instead of 128 bit. For the spatial resolution utility function $U_N(N)$, a linear utility increase is assumed from the minimum defined resolution level *Nmin*=15 to the maximum defined resolution level *Nmax*=30. But, after *Nmax*, increasing the spatial resolution further does not contribute to the service quality of the network.

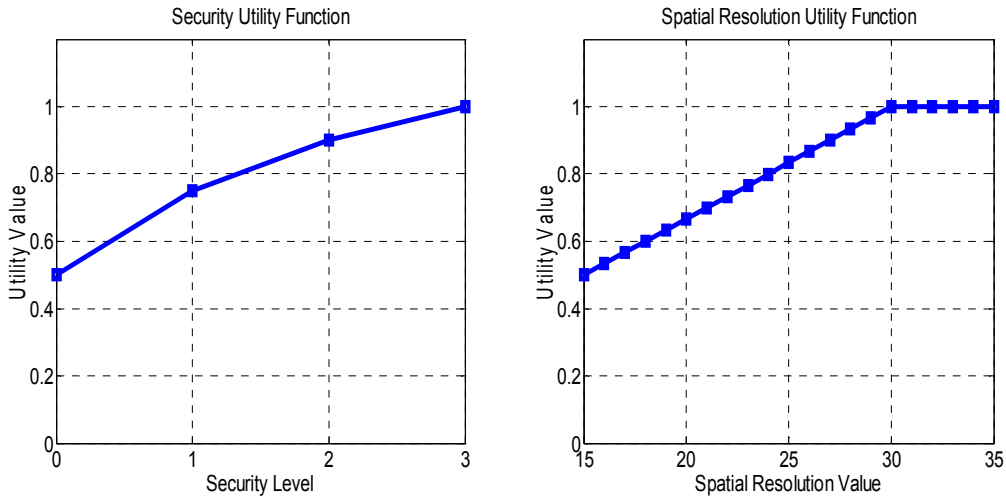


Figure 4.3: Security and spatial resolution utility functions

Now, using the above utility functions, the optimization problem of Equation 1 is solved by brute force for (S^*,N^*)=(3,25) with several different weights W_S and W_N to find the corresponding optimal (S,N) pairs. The results are given in Table 4.2.

As seen from the table, for unity values of W_S and W_N, security is preferred over resolution and the optimal supported tuple is determined as (3,19). As the weight for spatial resolution increased from 1 to 2, 5 and 100, resolution starts to be preferred over security and higher spatial resolution values are selected corresponding to higher W_N values. While keeping W_N at unity, increasing the weight of security from 1 to 2 or 5 does not change the preference of security over resolution since security is selected already for W_S =1. But, decreasing the W_S value below 1 causes resolution to be preferred.

So far, the results of the solution to the optimization problem to determine the best security-resolution tradeoff have been given. Yet, how the optimization problem can be solved has not been detailed. The problem given in Equation 1 is a combinatorial optimization problem in which a scarce resource, namely, channel capacity is aimed to be optimally allocated among two competing factors, which are security and spatial resolution. In the literature, these problems are either

Table 4.2: Best tradeoff for *(S*,N*)=(3,25)*

Ws	W_N	Optimal *(S,N)* pair
1	1	(3,19)
1	2	(2,22)
1	5	(1,24)
1	10	(1,24)
1	100	(0,25)
2	1	(3,19)
5	1	(3,19)
0.5	1	(2,22)
0.2	1	(1,24)
0.1	1	(1,24)
0.099	1	(0,25)

solved by brute force exploring the usually-large solution space or by specifically designed algorithms which reduce the effective size of the space and explore this reduced space efficiently. In accordance with those two approaches, two methods are given for solving the optimization problem to determine the best security-spatial resolution tradeoff for cases where requirements exceed the channel capacity. The first method follows the brute force approach and the second one is based on a heuristic algorithm reducing the space in which the optimal solution is searched for.

As mentioned previously, the brute force method is performed in the following way: Given the unsupported *(S*,N*)* pair, all possible *(S,N)* combinations in the range $Nmin \leq N \leq N^*$ and $Smin \leq S \leq S^*$ are tried and the pair yielding the maximal U value which also satisfies the condition $N \leq DSF/Ds$ is selected. The algorithm for the brute force method is given below:

BruteForce(*N*,S**)

1. *Uold := 0*
2. *Unew:=0*
3. **for** *S = Smin* **to** *S** **do**
4. **for** *N = Nmin* **to** *N** **do**
5. *Unew := Ws.Us(S-Smin) + WN.UN(N-Nmin)*
6. **if** *(Unew>Uold* **and** *N<=Nsmax(S-Smin))* **then**
7. *Uold := Unew*
8. *Noptimum := N*
9. *Soptimum := S*
10. **return** *Noptimum, Soptimum*

At the worst case where $N^*=Nmax$ and $S^*=Smax$, the BruteForce algorithm searches the whole solution domain defined by $[Smin, Smax] \times [Nmin, Nmax]$. If the differences of $|Smax-Smin|$ and $|Nmax-Nmin|$ are not too large, this approach may not cause too much problem regarding the computational efficiency. However, large values of these differences may put excessive load on the processor of the sensor node that will perform the optimization computations because the operational complexity of the brute force method increases by the product of $|Smax-Smin+1| \times |Nmax-Nmin+1|$. This may both increase the latency of the network and decrease the network lifetime due to excessive battery drain.

Alternative to the BruteForce algorithm, this thesis proposes a heuristic with a much more endurable computational complexity. The basis of this heuristic is the fact that the overall utility function U is a non-decreasing function of both S and N since individual utility functions Us and UN hold this property. This fact makes it unnecessary to check all (S,N) combinations in the range $[Smin, S^*]$ and $[Nmin, N^*]$. In fact, if one can find a tuple (S',N') satisfying the all three constraints defined by three inequalities of Equation 1, then it is unnecessary to check also $(S'-k, N'-m)$ for any $k \geq 1$ and $m \geq 1$ because the utility value $U(S',N')$ is always greater than or equal to the utility value $U(S'-k, N'-m)$ due to the non-decreasing feature of utility functions mentioned before.

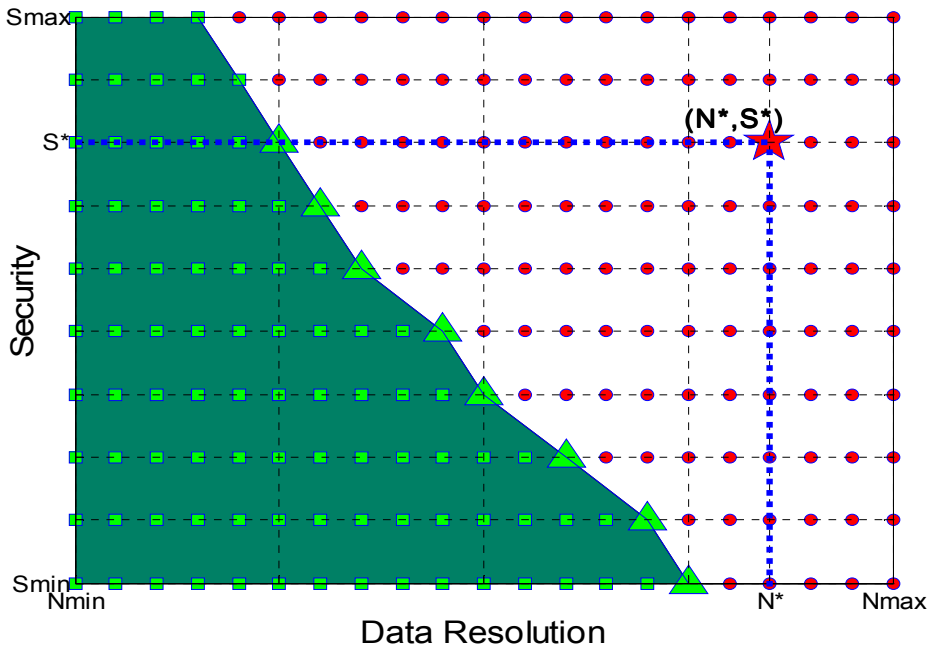


Figure 4.4: Security vs spatial resolution to explain the heuristic

This fact becomes comprehendible if one looks into the graph given in Figure 4.4 where the concept of supported and unsupported *(S,N)* levels are visually illustrated for a network where defined security range is *[Smin,Smax]* and spatial resolution range is *[Nmin,Nmax]*. In the figure, supported *(S,N)* tuples are represented by squares and unsupported ones with circles. All supported security-spatial resolution pairs lie in the shaded region. From this figure and the observation of the previous paragraph, one can deduce the following fact: In order to determine the optimal supported security-spatial resolution values for the unsupported requirement of *(S*,N*)* shown with a big star in the graph, it suffices to check only the *(S,N)* points located on the boundary between the shaded and unshaded regions of the graph. Those boundary points are indicated as big triangles in the figure.

More specifically, one should compare the utility values for only those *(S,N)* tuples residing on the boundary of the graph and only those of which are located

inside the rectangular area defined by *[Smin, S*] × [Nmin, N*]* shown as dashed lines in the figure. Thus, instead of checking all the points located inside the marked rectangular area of the graph as brute force method suggests, it is sufficient to compute the utility values of a few points shown with big triangles and pick the one with the highest utility value as the optimal supported security-spatial resolution pair.

Those boundary points represented as big triangles in the graph correspond to the previously mentioned *(S',N')* tuples for which it is certain that $U(S',N') \geq U(S'-k, N'-m)$ for any $k \geq 1$ and $m \geq 1$. Therefore, if one can find a method to determine those boundary points, then it is easy to compute the optimal supported security-spatial resolution levels with much less computational complexity compared to the brute force approach. The heuristic algorithm developed in this thesis is solely based on this fact. It is given below.

Heuristic(*N*,S**)
1. *Uold := 0*
2. *Unew:=0*
3. *S=S*;*
4. **while (***N* >= Nsmax(S-Smin)* **and** *S>Smin)* **do**
5. *N := Nsmax(S-Smin)*
6. *Unew := Ws.Us(S-Smin) + W_N.U_N(N-Nmin)*
7. **if** *(Unew>Uold)* **then**
8. *Uold := Unew*
9. *Noptimum := N*
10. *Soptimum := S*
11. *S :=S-1*
12. **if** *(N* >Nsmax(S-Smin))* **then**
13. *N := Nsmax(S-Smin)*
14. **else**
15. *N=N**
16. *Unew := Ws.Us(S-Smin) + W_N.U_N(N-Nmin)*
17. **if** *(Unew>Uold)* **then**
18. *Noptimum := N*
19. *Soptimum := S*

return *Noptimum, Soptimum*

At the worst case with $N^*=Nmax$ and $S^*=Smax$ where the solution space to be explored is the largest, the number of *(S,N)* tuples whose corresponding utility values will be compared by the heuristic algorithm is only $|Smax-Smin+1|$. This number, which represents the operational complexity of the heuristic, is always smaller than the corresponding merit of the brute force algorithm. In fact, the number of operations that will be performed in the heuristic algorithm linearly increases with $|Smax-Smin+1|$ as opposed to the multiplicative dependence of the brute force approach given as $|Smax-Smin+1| \times |Nmax-Nmin+1|$. This is again visible from the graph of Figure 4.4 as only the points on the boundary line are included for the search of the heuristic algorithm where all the points in the rectangular area are included in the brute force algorithm. Therefore, the heuristic should provide considerable efficiency regarding the computational time.

In this study, both the brute force and heuristic algorithms just mentioned have been implemented in C language and the corresponding codes have been executed on a hardware platform called Gumstix (*Gumstix*, n.d.). Gumstix is the brand name of tiny 200 or 400 Mhz single board computers based on the Intel XScale processors. All Gumstix computers and motherboards come preloaded with the Linux operating system. The motherboards are in the size of 80 mm x 20 mm x 6.3 mm, which is comparable to a stick of chewing gum. A range of daughtercards is available that can extend the I/O function of the system in a wide range of possible ways such as serial, USB, Ethernet, Bluetooth and Wi-Fi wireless interfaces.

The 400 MHz processor speed of Gumstix boards is considerably higher compared to 4-8 MHz Atmel Atmega 128L processor of MICA2 motes which are extensively used in sensor network research. However, Gumstix is also a suitable platform to be used as the cluster head of a cluster-based sensor network that will compute the optimal security-spatial resolution levels as the system modeling of this thesis suggests. The reason is that the cluster head is usually assumed to have more powerful hardware capabilities compared to the standard sensor nodes.

Thus, one may consider that the Gumstix computer is the cluster head performing the optimization computations and MICA2 motes are the surrounding sensor nodes in this cluster.

The executables of corresponding C codes of the brute force and heuristic approaches have been run on a 400 MHz Gumstix computer and the CPU times were recorded for 1000 consecutive computations for the requirement of (S^*, N^*) = (*9, 115*) where *Smin=0, Smax=9, Nmin=15* and *Nmax=115*. Both |*Smax-Smin+1*| and |*Nmax-Nmin+1*| values have been designated large on purpose to better see the performance difference of two approaches. The time that took the processor to perform 1000 consecutive computations is 2.03 seconds for brute force and 0.02 seconds for the heuristic. So, the Gumstix processor can perform a single best tradeoff computation in 2.03 milliseconds using brute force and 0.02 milliseconds using the heuristic algorithm. As expected, the heuristic approach has much better computational efficiency. In fact, for this specific case, the heuristic approach operates almost 100 times faster and this result is in full agreement with the presented computations of the operational complexity of two algorithms. Heuristic requires the order of |9-0+1|=10 operations while the brute force requires |9-0+1| × |115-15+1|=1010 operations, which is almost 100 times more. Knowing that a single frame duration of a MAC frame in TRACE protocol is 25 milliseconds during which the computation of the optimal supported security-spatial resolution value should be completed, the advantage of the heuristic algorithm on computation time may turn out to be critical, especially for cases where the range for spatial resolution *[Nmin, Nmax]* is large.

As the final research finding of this section, to determine the effect of utility functions over the optimal supported security-spatial resolution levels, the shapes of some individual utility functions U_N and U_S have been changed by modifying the utility values at satisfaction knee points. The utility functions that are used are $U_S(S)= 1-\exp(a*S+b)$ and $U_N(N)=c*N+d$,. Those utility functions drawn for

Table 4.3: Effect of utility functions on the optimal solution

$U_S(0)$	$U_N(15)$	Optimal *(S,N)* pair
0.5	*0.5*	*(2,22)*
0.5	*0.99*	*(3,19)*
0.5	*0.4*	*(1,24)*
0.5	*0*	*(1,24)*
0.6	*0.5*	*(1,24)*
0.97	*0.5*	*(0,25)*
0.2	*0.5*	*(2,22)*
0	*0.5*	*(2,22)*
0	*0*	*(1,24)*
0	*0.95*	*(3,19)*
0.95	*0*	*(0,25)*
0.95	*0.95*	*(2.22)*

sample parameters *Smin=0, Smax=3, Nmin=15, Nmax=35,* $U_S(0)=0.5$, $U_S(3)=1$, $U_N(15)=0.5$ and $U_N(35)=0.999$ are shown in Figure 4.5.

The optimal solution for several different $U_S(Smin)$ and $U_N(Nmin)$ values have been computed while keeping the weight values W_N and W_S constant at 1. Specifically, both $U_S(0)$ and $U_N(15)$ have been decreased and increased to several different values above and below 0.5, which was the value assigned to both previously. This modification causes a change in the slope/concavity of security and spatial resolution utility functions shown in Figure 4.5. The numerical results computed for several different shaped utility functions are presented in Table 4.3.
If the Table 4.3 is analyzed, it is seen that increasing the utility value at satisfaction knee point of *N=15* from 0.5 to 0.99 causes the optimal solution to shift from *(2,22)* to *(3,19)*, that is, preference moves towards security. At first sight, this may appear as counterintuitive since increasing utility values for spatial

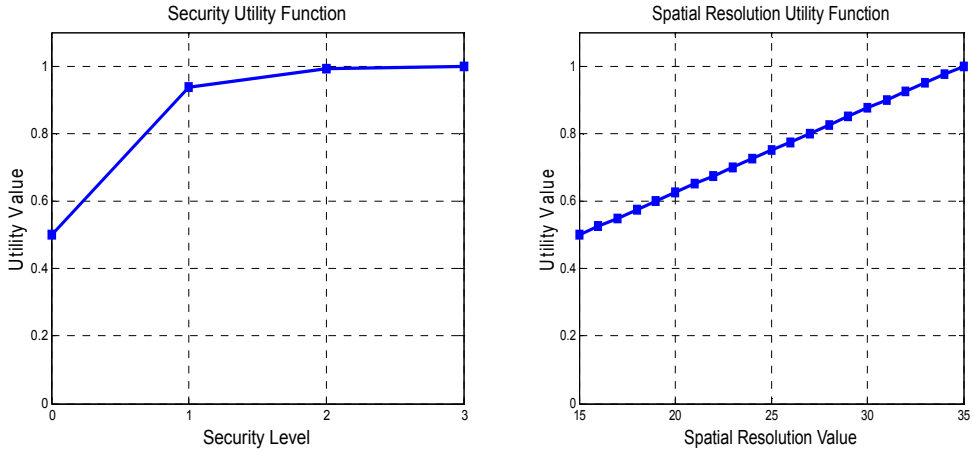


Figure 4.5: Second set of security and spatial resolution utility functions

resolution makes the optimal solution biased in favor of security. Yet, this should be expected because increasing $U_N(Nmin)$ decreases the slope of U_N graph and this causes the utility value differences between high and low levels of spatial resolution to become very little. Thus, this causes the optimal solution to prefer higher security levels to increase the overall utility value U by having increased security utility at the expense of very little loss of spatial resolution utility. Similarly, decreasing $U_N(15)$ below 0.5 changes the optimal solution in favor of spatial resolution, i.e., from *(2,22)* to *(1,24)* because the slope of resolution utility function is now increased making the utility increase for higher resolution values much more profitable. However, this effect is not able to shift the optimal solution from *(1,24)* to *(0,25)* even when $U_N(15)$ is decreased down to zero.

The same behavior can be observed when the security utility function value is modified at satisfaction knee point of $S=0$. Increasing $Us(0)$ makes the optimal solution to shift towards the spatial resolution and decreasing it has the reverse effect. This is seen from the change of optimal (S,N) pair from *(2,22)* to *(1,24)* for $Us(0)=0.6$ and to *(0,25)* for $Us(0)=0.97$. In last four rows of the Table 4.3, the optimal supported security-spatial resolution levels are included for some extreme utility function values at *Nmin* and *Smin*. As seen, for the highest $Us(0)$ and the

lowest $U_N(15)$ values, the optimal solution is at the highest spatial resolution and the lowest security level, that is, *(0,25)* and for the opposite case of the lowest $U_S(0)$ and the highest $U_N(15)$ values, the optimal solution is at the highest security and the lowest spatial resolution level of *(3,19)*.

4.3. EFFECT OF SECURITY ON POWER CONSUMPTION

In Chapter 3 where assumptions and definitions of the thesis were given, system lifetime is described as the time duration between the start of operation of the sensor network and the forced end of operation when there remains insufficient number of sensor nodes due to deaths of most sensors. This forced end of operation of the sensor network occurs when the number of operational sensor nodes drop below the pre-defined minimum required spatial resolution value. It is also assumed that the only condition for a sensor node to be operational is to have a non-exhausted battery, that is, this thesis presumes that sensor deaths occur only due to battery exhaustion and other reasons such as physical damage that can make sensor nodes non-operational is excluded. Therefore, under those assumptions, the only factor determining the lifetime of the network is power consumption/battery usage of the sensor nodes.

On the other hand, of the three main tasks, sensing, computing and communicating, which cause a sensor to drain its battery, the last two are affected by the security features employed in the wireless sensor network under consideration. In fact, a higher level of security usually increases the power consumption during both computing operations performed by the processor and communication performed by the radio of the sensor node. Since a multi-level security approach is taken by this thesis which assumes several security levels provided by different length message integrity codes, the correlation between security and power consumption, which in turn affects another QoS attribute, system lifetime, needs to be investigated. In this section, a model is constructed from the existing studies in literature to depict the mathematical relationship between the required security level S and the power usage of sensor nodes.

Table 4.4: Electrical properties of a MICA2 node

Component	Current Drawn	Operation State
CPU	8 mA	Active
	15 uA	Sleep
RADIO	27 mA	Transmit at 10 dBm
	10 mA	Receive at 10 dBm
	1 uA	Sleep

A parameter called *Ps* is defined to represent the amount of energy consumed by a node which transmits a single packet at security level *S.* To be able to determine the values of *Ps* for each security level *S,* the well-known formula $P=I \times \Delta t$ will be used where *P* is the power spent by an electronic device which draws *I* Amperes current during Δt seconds. This formula has been already used in several research studies about power consumption in sensor networks such as Guimarães et al. (2005).

Now, a closer look will be taken at the two main factors of the formula which are *I* and Δt**.** There are two main components that draw current from the battery of a sensor node. These are radio and CPU. Radio draws current during transmit, receive operations and also in sleep. CPU draws current mostly during cipher operation. The values for these current draws are presented in Table 4.4 reconstructed from the data sheet of MICA2 motes (*MICA2 Specifications*, n.d.).

The current drawn in sleep states of both CPU and radio will be ignored since they have negligible values. Then the power consumption of a transmitting node during a time epoch is $Ps = I\_cpu \times \Delta t\_cpu + I\_tran \times \Delta t\_tran + I\_rec \times \Delta t\_rec$. The

values for I_cpu, I_tran and I_rec are already known from the table above and it is only needed to determine Δt values to compute Ps.

Starting with the receive duration Δt_rec, it is known from the assumed MAC scheme that the only time when a node receives information during an epoch is the Header slot. All nodes that have an intent to transmit in an epoch has to listen during the all Header slot duration to learn the schedule and other QoS and security related information. Therefore, Δt_rec is equal to the duration of a Header slot. In TRACE on which the MAC scheme of this thesis is built, duration of a Header slot is given as 5 Byte durations plus 2 Byte durations for each scheduled node. Then, one can find, $\Delta t\_rec=(5+2N(t))$ Byte durations. In the specifications of a MICA2 mote, a Byte duration is given as 0.42 msec and it is already known that I_rec is 10 mA. Therefore, the power consumption of a sensor node regarding the receive function is $P\_rec=I\_rec \times \Delta t\_rec=10mA \times (5+2N(t))$ *Byte* $\times$ *0.42 msec/Byte* $\times$ *(1 msec/3600.1e3 hour).* It results in $\mathbf{P\_rec=1.1667 \times 1e\text{-}6 \times (5+2N(t))\ mAh.}$ Note that the receive power consumption does not depend on the security level and also the units are converted into mAh since battery capacities are usually given as mAh (milli Ampere Hours).

Now, it is needed to find Δt_tran to determine the transmit power $P\_tran= I\_tran \times \Delta t\_tran$. Again from the proposed MAC scheme, it is known that a transmitting node of a single epoch makes three transmittals in that epoch. The first one of these is in the contention slot, then in the IS slot and finally in the data slot. Lengths of both IS slot and contention mini slot are equal as 5 Byte durations as given in TRACE. Length of the data slot is dependent on the security level S and equals to Ds as mentioned previously. Then, $\Delta t\_trans=(10+Ds)$ Byte durations and $P\_tran= I\_tran \times \Delta t\_tran=27mA \times (10+Ds)$ *Byte* $\times$ *0.42 msec/Byte* $\times$ *(1 msec/3600.1e3 hour)* gives $\mathbf{P\_tran=3.15 \times 1e\text{-}6\ x\ (10+Ds)\ mAh.}$

What remains is to determine Δt_cpu. Here, it will be assumed that most of the CPU activity of a node during an epoch is due to the cipher operation that it

performs to compute the message integrity code (MIC) over TinyOS packet. Therefore, it is needed to determine the length of the time period it takes to compute MIC over a 29+4=33 Bytes TinyOS packet. Actually, this information is given in Karlof et al. (2004) for SkipJack algorithm and a 4-byte long MIC. But, since several MIC lengths are assumed in this thesis study, information on CPU time needed to compute only a 4-byte MIC is not sufficient and, the values for 8 and 16 Bytes MICs are also required.

For this reason, details on how a message integrity code is computed in TinySec protocol will be investigated. TinySec uses a CBC-MAC (Cipher Block Chaining Message Authentication Code) mode to generate a MIC using a block cipher algorithm such as SkipJack, DES or AES. In a CBC-MAC operation process as shown in Figure 4.6, message to be hashed is divided into equal length blocks and the output (MIC) has the same length as the block size.

Taking message size as M and block size as B, a MIC of length also B is computed in $\lceil M/B \rceil + 1$ cipher operations. TinySec uses a block size of 8 Bytes to compute MIC using SkipJack in CBC-MAC mode. This produces an 8-Byte output and only 4 leftmost bytes of this output is taken as the message integrity code. But, since in the multi-level MIC approach 16-Byte MICs are needed, this 8-Byte block of TinySec is not enough as it was previously stated. If AES is used as the block cipher, however, and 16 Bytes as the block size, then it will work out. This can be done without any problem since in Karlof et al. (2004) authors state that AES algorithm instead of SkipJack can be used with TinySec. Therefore, it can safely be assumed that AES in CBC-MAC mode with 16-Byte block size is used to compute MIC and only the required number of leftmost bits of the output will be used as the MIC as exactly done in Xiao, Chen, Sun,Wang and Sethi (2006) for IEEE 802.15.4.

Consequently, for every possible length of MIC such as 4, 8 or 16 Bytes, the same number of cipher operations are performed since 16-Byte block size is used for all

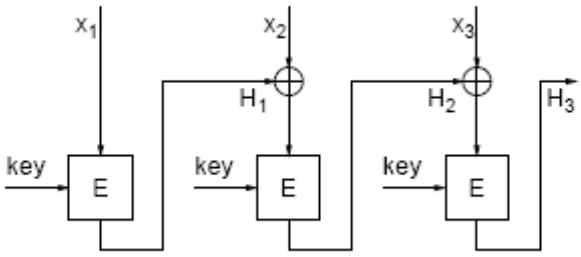


Figure 4.6: CBC-MAC operation process

cases. Therefore, CPU time required for all these MIC computations are the same. In Kaps and Sunar (2006), authors present their results on power consumption of TinySec protocol when implemented in AES CBC-MAC. According to their results, using AES to compute MIC over a 33 Byte TinyOS packet takes about 3.2 msec. So, it can be taken that Δt_cpu is equal to this value for all security levels except for no security case in which Δt_cpu is zero. Knowing that I_cpu is 8 mA, one can determine $P\_cpu = I\_cpu \times \Delta t\_cpu$ as ***P_cpu=7.111 × 1e-6 mAh for s≠0***

Finally, one can compute the power consumption *Ps* of a node during a single epoch when it transmits at security level *S*. It is equal to ***Ps = P_rec + P_tran + P_cpu*** where all these *P_rec, P_tran, P_cpu* values are computed above. Therefore, the mathematical relationship between security level *S* and power consumption *Ps* of a sensor node at security level *S* is

Ps *= 1.1667 × 1e-6 × (5+2N(t)) + 3.15 × 1e-6 × (10+Ds) + 7.111 × 1e-6 mAh (for s≠0)*

CHAPTER 5

ACKNOWLEDGEMENT BASED QoS AND SECURITY CONTROL FOR WIRELESS SENSOR NETWORKS

In the first chapter where the objective of this research study was given, it was stated that this thesis had an overall aim of proposing a method to satisfy the time-varying service quality and security requirements of cluster-based wireless sensor networks. The same chapter defined the quality of service attributes addressed by the thesis as spatial resolution, system lifetime, coverage and packet loss due to collision, and the security scope was stated to be composed of message integrity and authentication. While surveying the relevant work in the literature in Chapter 2, it was mentioned that Kay and Frolik (2004) has a similar QoS scope to this thesis' in the sense that authors propose a strategy named *ACK method* to provide application level QoS for cluster-based wireless sensor networks where only spatial resolution and system lifetime included as service quality parameters.

In this chapter, a QoS and security control strategy is proposed based on the ACK method of Kay and Frolik (2004). The proposed method enhances the study in Kay and Frolik (2004) by incorporating security and also adding two more QoS attributes, namely coverage and packet loss due to collision. Besides, it allows both security and spatial resolution requirements to be time-varying as another

advantage over the original ACK method. Thus, the control strategy proposed in this chapter meets all of the five objectives stated in the problem statement section, which are security, spatial resolution, system lifetime, coverage and packet loss due to collision. Simulation results assessing the performance of the proposed QoS and security control strategy are also included in this chapter. In the last part of this chapter, an approximate probabilistic analysis on the relationship between coverage and spatial resolution under the proposed ACK-based method is presented.

5.1. QoS AND SECURITY CONTROL METHOD BASED ON THE ACK STRATEGY OF KAY AND FROLIK (2004)

The ACK strategy of Kay and Frolik (2004) utilizes the finite state automaton shown in Figure 2.6 of Chapter 2 to keep the sensor network at the required spatial resolution level. Single bit information included in the acknowledgement packets which are unicasted to only transmitting sensor nodes causes each of those nodes to jump between the states of the automaton. Since the states of those finite state automaton correspond to different data transmit probabilities, the number of active (data-sending) sensors are adjusted according to the difference between the desired and current spatial resolution levels with an expectation that the active sensor number will finally converge to the desired spatial resolution level. In addition, the ACK method causes only the transmitting nodes to keep their radio open to receive the acknowledgment packets and also different sensor groups are become activated in different turns. Both of these two features help power conservation, which will lengthen the overall system lifetime. Simulation results given in Kay and Frolik (2004) show that those design principles indeed work well and the ACK method has a fair performance in providing the required spatial resolution level in a lifetime extending manner.

Therefore, the first QoS and security control method of this thesis given in this chapter (the second one is given in Chapter 6) is based on the ACK method to inherit its success in satisfying two of the included QoS attributes of this study, which are spatial resolution and system lifetime. In order to control also the remaining two QoS attributes, packet collisions and coverage, and the other parameter addressed by this research, security, additional mechanisms are incorporated into the ACK method. Those mechanisms are explained in the sequel.

As previously mentioned, this thesis study proposed to secure the sensor-to-cluster head communication by the use of variable length message integrity codes, the length of which are determined by the required/supported security level S of the current interval. Therefore, sensor nodes deciding to transmit based on the transmit probability of the current state of the automaton need to be informed about the current desired security level. The method presented in this chapter proposes to send this security information during the Header slot of the MAC frame, details of which are given in Figure 3.3. Thus, the cluster head learns the security requirement of the current time interval from the control center of the sensor network, checks whether this security level can be supported under the current spatial resolution requirement and computes the supported level as explained in the last chapter, and then announces this current security level requirement to the sensor nodes in the Header slot of the MAC frame. All sensors should send their data with a message integrity code corresponding to this announced security level. This will provide the security of sensor to cluster head communications.

In order to minimize the packet drops due to collisions, another QoS attribute uncovered by the original ACK method, the proposed method of this chapter relies on the slotted MAC scheme employed. As detailed in Chapter 3, since this slotted MAC algorithm based on TRACE (Tavli & Heinzelman, 2003) allows transmission of only the selected nodes in their corresponding data slots, there is

no risk of collision during the data sub-frame. Therefore, the probability of data packet loss due to collisions during the data transmission period is zero. In addition, if the number of contention mini slots is designated to be sufficiently higher than the number of data slots, this will further reduce the collisions that can occur during the control sub-frame where nodes state their intend to transmit. In fact, the number of contention mini slots should ideally be set to *e* times the number of data slots, because the maximal throughput of a slotted ALOHA system is *1/e*. Although the sensor network considered in this study is not a generic ALOHA network, sensor nodes independently decide whether to send data or not through a statistical mechanism, and access the mini slots if they have any data to transmit. In this manner, it is similar to the slotted ALOHA system and designing the contention period *e* times longer than the data transmission period will greatly reduce the collision probability during the control sub-frame.

Besides those mentioned features of the utilized MAC scheme of this thesis which will result in zero data packet collisions and very little control packet collisions, the intrinsic purpose of the proposed control strategy which regulates spatial resolution, i.e., the number of data sending sensor nodes, is another factor limiting packet collisions because of the following reason. The probability of collisions in the contention period (control sub-frame) gets higher when number of sensor nodes that intend to transmit increases. In fact, there will certainly be collisions if the number of nodes trying to access the contention mini-slots is higher than the number of those mini slots. The proposed control method of this thesis, however, regulates the transmit rates of individual nodes when the current resolution is higher than the required level and therefore also regulates the number of sensors that try to access the contention mini slots to indicate their intention to transmit. Thus, both the adopted MAC scheme and the control method itself help minimizing the packet drops due to collisions.

For coverage, the third attribute not considered by the ACK method, the method proposed in this chapter will not involve any direct control mechanisms in the first

place. The reason is that the stochastic nature of the proposed strategy causing each sensor to make transmissions based on an independent probability determined by the automaton state is expected to result in a geographically balanced distribution of active nodes. If this actually causes sensors to evenly spread over the service area making at least one active sensor taking measurements over each sub-region, it will be an indication of fair coverage. Therefore, the ACK-based control strategy of this chapter does not have a dedicated means for ensuring a certain coverage level. Yet, it includes coverage as a QoS attribute in addition to spatial resolution since those two parameters together constitute a more meaningful service quality level than considering only spatial resolution as done in the original ACK method of Kay and Frolik (2004). The reason is that the effective sensing ranges of sensors are usually shorter than the communication ranges and, though there might be enough number of active sensors sending data to the cluster head, measurements taken by those sensors might contain information regarding only a specific region of the WSN area but exclude some other regions. Besides this necessity of addressing those two concepts simultaneously, coverage and spatial resolution has some correlation that will be investigated in Section 5.3 of this chapter. But before this, the acknowledgement based security-QoS strategy proposed by the thesis is presented below.

This strategy, utilizing the ACK method of Kay and Frolik (2004) to provide spatial resolution and lifetime extension and involving the principles explained in the last three paragraphs above to provide security, coverage and minimal collisions respectively, is composed of several steps that occur periodically in each epoch, i.e., discrete interval. Duration of each epoch is equal to one frame duration of the proposed MAC scheme and epochs are synchronized with frames. The steps of the proposed method are presented below:

1. Cluster head (CH) starts transmitting the beacon message.
2. CH checks whether there is a change in the required security and spatial resolution levels *(S*,N*)* which are announced by the control center of the sensor network. If there is a change in either *S** or *N** with respect to the previous epoch, CH proceeds to step 3, otherwise it goes to step 6.
3. CH checks whether new security and spatial resolution requirements are supported by using the method given in Section 4.1. If they are supported, it goes to step 6. If the required levels *(S*,N*)* are not supported, it computes the optimal supported levels *(S', N')* by the method of Section 4.2 and then proceeds to step 6.
4. Before the beacon period ends, each and every nodes decides whether to transmit or not during the current epoch. Nodes make this decision in the same way as in the ACK strategy, i.e., by comparing their locally generated random number to the transmit probability of their current state. Nodes which decide to transmit open their radio, synchronize with the beacon and proceed to step 5. Others switch to the sleep mode.
5. After the beacon period ends, nodes deciding to transmit in the previous step contend for a mini slot in the contention slot.
6. Before the transmission of the Header packet, CH should have finished the calculation of the optimal security and spatial resolution values *(S', N')*. Also, in this step, CH determines the number of sensors that request to be active for the current epoch by counting the number of accessed contention mini slots. This number *Nt* represents the current (expected) level of the spatial resolution of the current epoch *t* and will be compared to the desired value of *N** as done in ACK strategy. The difference of the proposed method than the original ACK strategy is that *Nt* is computed before nodes transmit their actual data.
7. During the Header period, CH first unicasts the schedule of data transmissions for the current frame. This schedule is an ordered list of sensor nodes prepared according to the order in contention mini slot access and it also includes the slot duration *Ds** (or *Ds'*) corresponding to the security level of the current epoch. Before announcing the schedule, CH should check that the number of sensors

desiring to transmit *Nt* do not exceed the supported (or desired) number of active sensors. If *Nt* exceeds *N'* (or *N**), CH chooses only *N'* (or *N**) of sensor nodes randomly and include only those sensors in the announced schedule.

8. After the schedule is announced, CH informs nodes that intend to transmit about two more issues during the Header period. The first one is the desired security level of the current epoch (*S** or *S'*) and the second one is the information on whether the current spatial resolution *Nt* is above or below the desired level *N** (or *N'*). Information on the security level occupies 2 bits since 4 security levels are assumed. Yet, it may be increased if more security levels exist (i.e., 3 bits for 8 levels). The second piece of information is only 1-bit. It is 1 if *Nt*<*N**(or *N'*) indicating that the current spatial resolution level is less than the desired value and, it is 0 otherwise.
9. All of the *Nt* nodes receiving the 1-bit information regarding the comparison of the current and the desired spatial resolution levels change their state in the automaton shown in Figure 2.6. If this value is 1, they reward themselves, otherwise, they punish.
10. Of *Nt* nodes which changed state in the previous step, the ones which are not listed in the announced transmission schedule go into the sleep mode. Only the nodes which find their name in the schedule transmit their packets at the security level announced and in the data slot assigned to them.
11. After the data sub-frame ends, all sensor nodes return to step 4 and CH returns to step 1.

5.2. SIMULATIONS FOR THE ACK-BASED METHOD

The verification method of this thesis study which will be used for assessing the performance of the proposed QoS and security control strategy is chosen to be simulation. While one reason for this choice is the difficulty of acquiring sufficient number of sensor nodes to implement an actual sensor network cluster, another important reason is that the proposed control method is composed of steps which can easily be converted to simulation code and simulation provides much

more control on the verification process due to the possibility of changing every control parameter to model various settings.

For the simulation platform, MATLAB (*MATLAB, The Language of Technical Computing,* n.d.) has been used where the technical programming environment it provides is sufficient to code the discrete time simulator required by the scope and assumptions of the thesis. Despite the existence of several network simulators[6] such as NS2 (*The Network Simulator, NS2*, n.d.) or OPNET (*The OPNET Modeler*, n.d.) which provide very powerful simulation environments for wireless sensor networks, abstraction of the details of the radio channel in the problem definition has resulted in little need of the capabilities provided by such simulation tools and has provided encouragement for implementing a specific simulator in MATLAB for this thesis. Assumptions and parameters regarding those MATLAB simulations as well as the simulation results are presented in the following subsections.

5.2.1 Simulation assumptions

For all simulations, the assumptions and system model described in Chapter 3 are used, i.e., a single WSN cluster where sensors send data to the designated cluster head in one hop under the TRACE-based MAC scheme. Initially 100 sensors are randomly distributed over a square-shaped cluster area the size of which is 100 m × 100 m. MICA2 type sensors are assumed with a maximum transmission range of 150 m (500 ft) as given in *MICA2 Specifications* (n.d.) allowing one-hop communication between the farthest points inside the square with a diagonal length of $100\sqrt{2} \approx 140$ m . Half of this maximum communication range, i.e., 75 m is taken as the maximal sensing range and this supports several possibilities to divide the cluster into sub-regions for coverage analysis such as 4 (50 m × 50 m square sub-regions with diagonal $50\sqrt{2} \approx 70$ m which is smaller than the maximum sensing range of 75 m), 8 (50 m × 25 m rectangular sub-regions) and

[6] Information on sensor network simulation platforms is provided in the Appendix D.

16 (25 m × 25 m square sub-regions). A perfect communication channel where all transmissions occur without any error or data loss is assumed. Only the packet drops due to collisions occurring in the contention period are taken into account. In simulations, a pre-determined and fixed node is designated as the cluster head. It served as the cluster head during the whole simulation duration and therefore there was no need for cluster head election. However, this does not put a constraint on the performance of the proposed method of thesis since it can run for cases where cluster-head is dynamically elected as long as the elected cluster head runs the software of the proposed algorithm.

In the simulations, deaths of sensor nodes due to battery exhaustion are modeled in a different way than the previous works such as Kay and Frolik (2004) and Iyer and Kleinrock (2003) with a purpose of including the effect of security on power consumption. These cited works consider the lifetime of sensor nodes as an exponentially distributed random variable and their simulations make sensor nodes die randomly due to battery exhaustion at some specified exponential rate. This approach is fairly simplistic since there is no need to track the battery level of individual sensor nodes. In the approach of this thesis, a simulation variable storing the battery level of each and every sensor will be kept and at each epoch during which a node transmits, some specified amount of power from this variable will be extracted. Since the amount of consumed power at each packet transmit (Ps) depends on the current security level S as detailed in Section 4.3, the effect of security on power consumption will be included in simulations.

In order to simulate the power consumption of sensors, one more parameter is needed to be set, which is the initial battery capacity of sensor nodes. From the specifications of MICA2 motes (*MICA2 Specifications*, n.d.), it is known that they work with 2xAA size batteries. It is also known that AA size batteries can have capacities ranging from 600 mAh to 3000 mAh depending on their type (Alcaline, NiCd, etc.) (*List of Battery Sizes*, n.d.). Considering the average current drawn from a MICA2 mote for transmit and receive operations as (27+10)/2=18.5 mAh

and assuming that they use 2x1500 mAh batteries, then the average lifetime of a MICA2 mote is 3000/18.5 ≈ 162 hours. The battery life test presented in *MICA2 AA Battery Pack Service Life Test* (n.d.) yields a similar result by measuring the total lifetime of a MICA2 mote as 172 hours. Regarding the simulations of this study, if sensors with full batteries are assumed, then the simulation would have to be run for about billions of discrete time intervals since the epoch duration which is equal to one MAC frame duration is in the range of milliseconds. Since it is not feasible to perform such long simulations, an average lifetime ranging between 400-500 epochs for each sensor will be taken depending on the simulation settings. In fact, previous studies make similar assumption on the lifetime of a sensor node to keep the simulation duration in feasible length. For instance, Kay and Frolik (2004) assumes an average lifetime of 250 epoch and Iyer and Kleinrock (2003) takes this value as 101.

Using the *Ps* formula given in Section 4.3 with a mean *N* value of 25 and averaging the values that the formula yield for each security level *S*=0 to *S*=4 with corresponding *Ds* values, the average consumption of a node at each transmission is computed to be 0.264 x 1e-3 mAh. Therefore, for each simulation, the initial battery capacity *BC* of each sensor is set to *LT* x 0.264 x 1e-3 mAh where *LT* represents the average lifetime assumption of a sensor node for the corresponding simulation, which was just said to be ranging between 400 to 500. Then, for each transmitting sensor, *Ps* mAh power will be extracted from the battery level of it starting from an initial value of *BC*. When the battery capacity of a node drops to zero, it will be assumed that the node has died due to battery exhaustion. This is the way this thesis include the deaths of sensors in the simulations.

Regarding the finite state automata adopted from Kay and Frolik (2004), some values for the simulations are also needed to be set, which are the number of automata states and corresponding transmit probabilities. In Kay and Frolik (2004), authors use a 3-state automaton with transmit probability values of 0.05,

0.1 and 1. Yet, they mention that these simulation parameters cause a large variance in the attained spatial resolution value and for less variance they recommend use of a 4-state automaton with probabilities of 0.001, 0.5, 1 and 1. In the simulations, this 4-state automaton with the given transmit probabilities are used due to the stated better variance property.

Final simulation assumption that will be given in this subsection is on the utility function parameters. For the simulations of this chapter, the utility functions shown in Figure 4.3 are used with unity weighting constants for both Ws and W_N.

5.2.2 Inputs, outputs and simulation parameters

Since the overall aim of the proposed control strategy of this chapter is to keep the network at desired security and spatial resolution levels (S^*, N^*) while at the same time providing sufficient coverage and minimizing the packet loss due to collisions, the simulation results should demonstrate the performance of the proposed method by illustrating how close the attained security-resolution values (S, N) to the desired ones, whether the active sensors cover the whole service area and what the packet loss ratio due to collisions is.

Therefore, given the requirements of $S^*(t)$ and $N^*(t)$ for all simulation duration t=0 to $Tsim$, the simulations will produce the attained values $S(t)$ and $N(t)$ as well as showing the geographical distribution of data sending sensors over the field and computing the number of collisions occurring during the contention period. Then, the main simulation inputs are $S^*(t)$ and $N^*(t)$ and main outputs are $S(t)$, $N(t)$, $Ni(t)$ and L, where Ni represents the number of active sensor over the sub-region i and L represents the ratio of lost packets due to collisions to the successfully transmitted ones during the whole simulation interval. As previously explained, there might be times when requirements $S^*(t)$ and $N^*(t)$ cannot be supported and for such cases supported values $S'(t)$ and $N'(t)$ are computed by the cluster head. In such cases, the proposed control algorithm tries to attain these supported levels

instead of the required values and therefore, the real performance of the method is reflected by the closeness of the attained values *S(t)* and *N(t)* to the supported values *S'(t)* and *N'(t)* rather than the requirements *S*(t)* and *N*(t)*. Other simulation parameters and pseudo code of the actual simulation code written in MATLAB are included in Appendix D.

Before presenting the simulation results in the next section, it must be noted that the simulation parameters used in the simulations, particularly the security and spatial resolution requirements might not reflect the real numerical values representing the requirements of an actual sensor network deployment since there were no opportunities during the research period neither to set up such a network nor to learn the real values of an already deployed network. Nonetheless, all the simulation parameters are selected to be comparable to the values used in simulations of similar studies in the literature and care has been taken not to be in contradiction with technical specifications of commercially available sensor nodes. In fact, the values given in the data sheets of such devices are assumed as much as possible.

5.2.3 Simulation results for the acknowledgement-based method

Using the above assumptions and the system model given in the previous sections, several simulations have been performed each starting with an initial deployment of 100 sensors. In the first simulation whose results are illustrated in Figure 5.1, it is set as $S^*=0$ and $N^*=35$ for all simulation duration to benchmark the proposed strategy against the ACK strategy of Kay and Frolik (2004) which aim to maintain 35 active sensors throughout the network operation. As seen from Figure 5.1, the proposed control strategy is able to keep the spatial resolution level at around 35 till most of the initially deployed 100 sensors die at around time epoch 130. This result is very similar to the simulation outputs presented in Kay and Frolik (2004) where it is shown that the ACK strategy outperforms the Gur Strategy of Iyer and Kleinrock (2003) regarding the overall network life since it dies before time epoch

30. Therefore, the results of this first simulation show that the proposed control method of this chapter performs well in both maintaining spatial resolution at the desired level and also in maximizing the WSN lifetime.

In order to see the performance of the proposed strategy in controlling both security and spatial resolution levels simultaneously, another simulation is performed. This time time-varying security and spatial resolution requirements are used as shown in the first subplots of Figure 5.2 and Figure 5.3 respectively. Illustrated in those first subplots are also the supported levels S' and N' which are computed using the optimization problem for cases when S^* and N^* are not attainable due to the limited channel capacity. In the second subplots of Figure 5.2 and Figure 5.3, the actual attained security and spatial resolution values S and N that the proposed method produce are shown against the supported values. As can be seen, the achieved security level S exactly traces the supported security value S' since the proposed strategy forces all active sensors to transmit at the required or supported security level. Similarly, except the transient times when the spatial resolution requirement N^* changes, the attained level N is able to track the supported value N'.

As the final simulation plot, in Figure 5.4, the results regarding the coverage performance of the proposed strategy are presented. In this case, the sensor network cluster is divided into four geographic sub-regions over which sensors are initially deployed in a random but uniform way. Then, this setup is simulated with the same parameters/requirements of the previous case and the geographic distribution of active sensors contributing to spatial resolution over those four sub-regions is observed. As shown in Figure 5.4, owing to the statistical nature of the proposed method, active sensors are almost evenly distributed and in each sub-region there are usually more than one active sensors almost at all times. This is an indication of fair coverage in the sensor network cluster since there are active sensors taking measurements in all of the geographic regions. Yet, this does not provide a hard guarantee such that full coverage is ensured at all times.

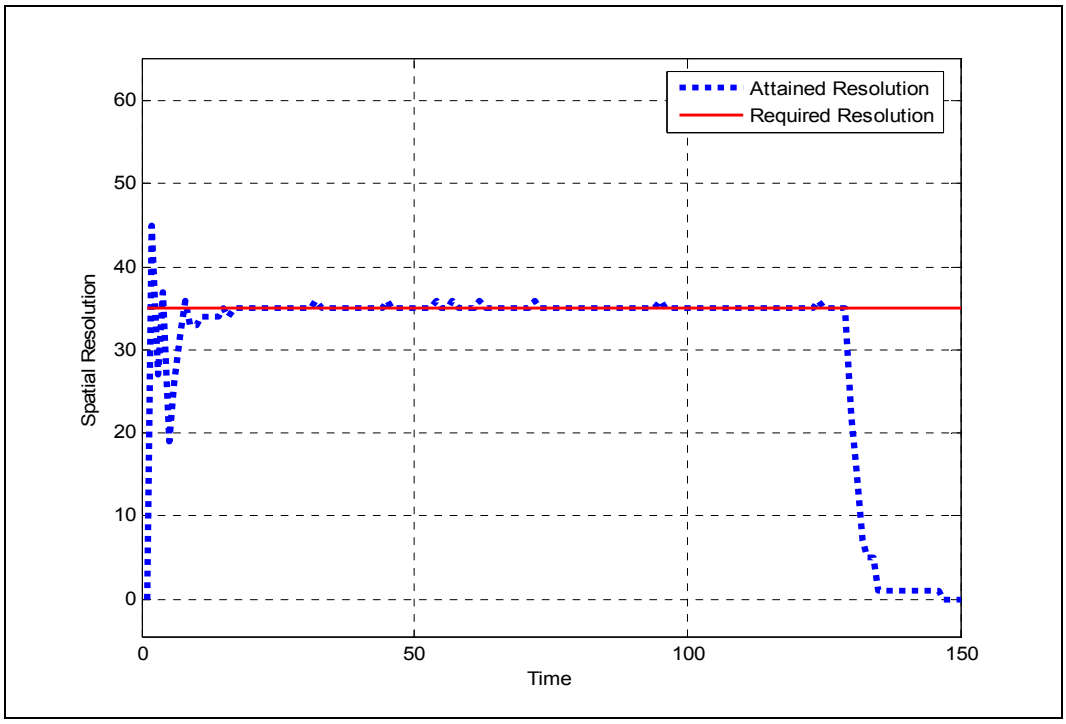


Figure 5.1: Spatial resolution vs time for zero level security

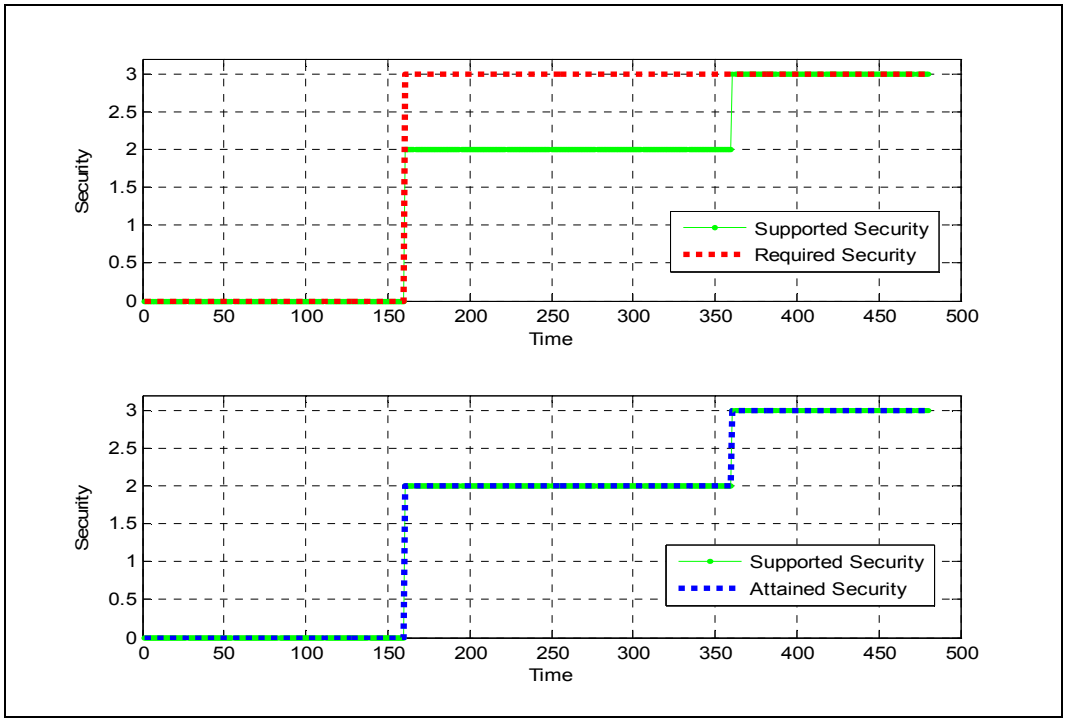


Figure 5.2: Supported & required (top) and supported & attained (bottom) security levels

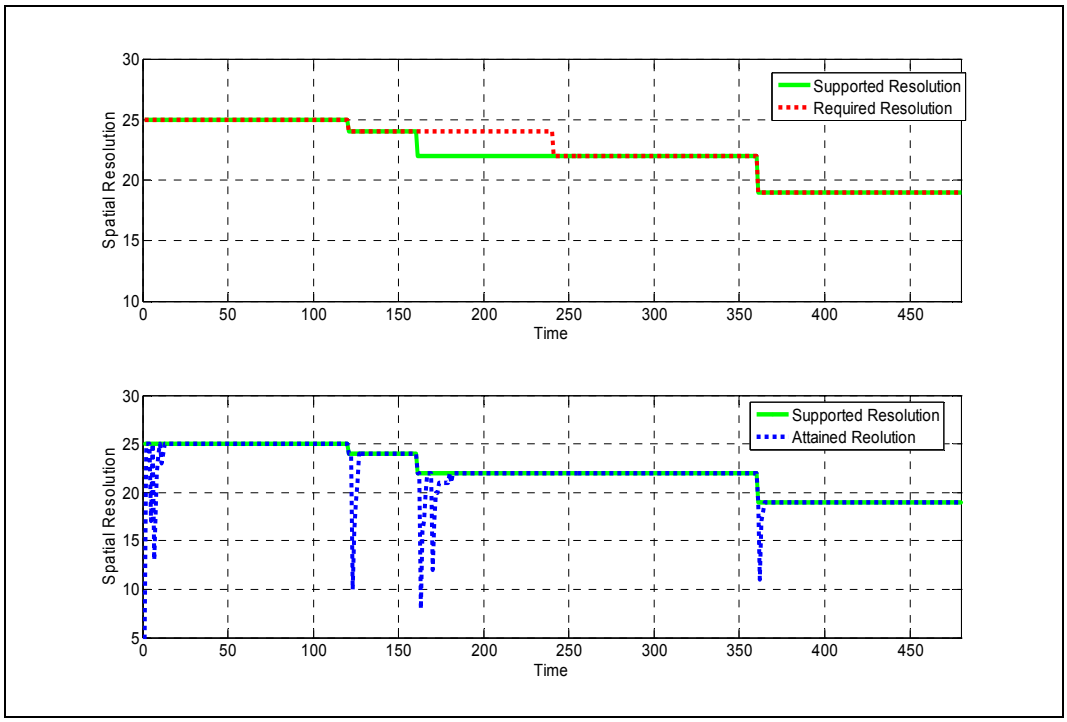


Figure 5.3: Supported & required (top) and supported & attained (bottom) spatial resolution

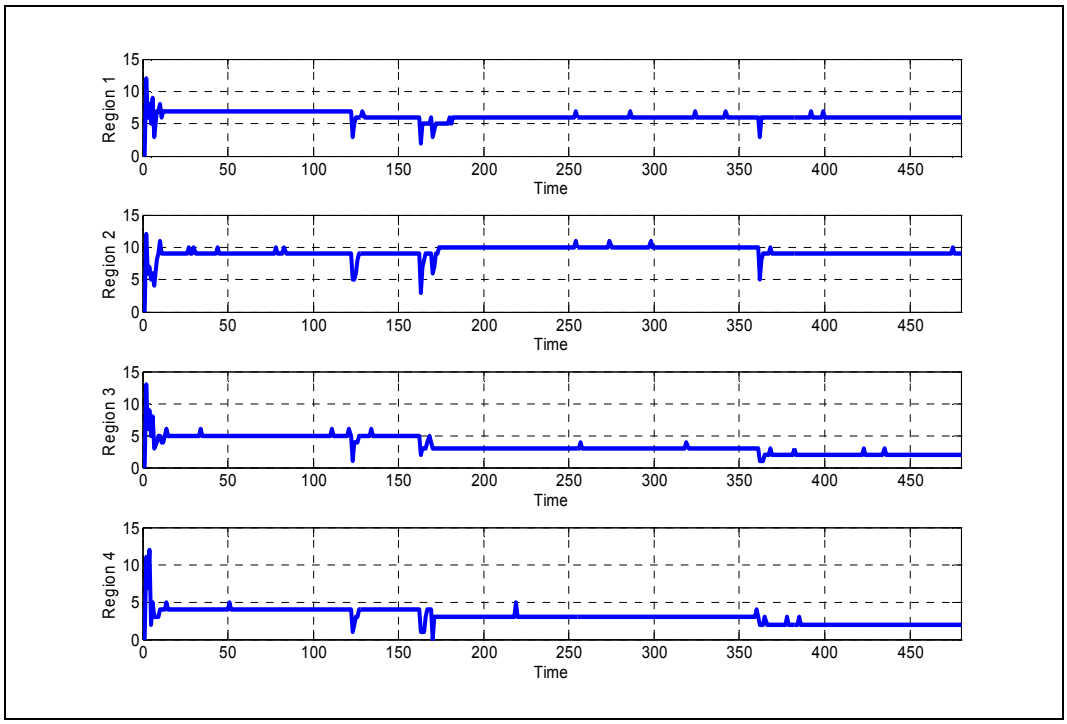


Figure 5.4: Distribution of active sensors over sub-regions

During the simulations, also the number of packets dropped due to collisions and the total number of successfully transmitted packets are recorded to determine the packet loss ratio due to collisions, which is *L*. As explained previously, collisions occur only in the contention period and it is assumed that the data packets of both of the colliding nodes accessing the mini-slots in the contention period are dropped. Note that this assumption is too pessimistic because although the colliding nodes cannot transmit their data, other nodes that are scheduled to send in the same frame can provide the required spatial resolution level rendering the packets that cannot be sent due to collisions unimportant. Even under that assumption, simulation results show that only 12 packets out of 10640 are dropped due to collisions, which yields an *L* value of 0.001.

Thus, through simulations, it is shown that the proposed strategy of this chapter has a fair performance in proving all of the five design aims, namely maintaining security and spatial resolution levels (Fig.5.2 and 5.3), extending the network lifetime (Fig.5.1), providing coverage (Fig.5.4) and finally minimizing the packet loss due to collisions.

5.3 APPROXIMATE PROBABILITY ANALYSIS FOR COVERAGE AND SPATIAL RESOLUTION

In this section, a deeper investigation will be presented about the relationship between coverage and spatial resolution for cluster-based wireless sensor networks and the scope in Section 5.1 will be extended to include *k*-coverage. First, an approximate probabilistic analysis is provided on the correlation of those two concepts and this analysis is verified by simulation. In addition, a modified QoS and security control strategy will be proposed to provide better *k*-coverage performance for the acknowledgement-based strategy of this chapter compared to the simulation results produced by the original control strategy of Section 5.2.

Table 5.1: Definitions for the probability analysis

Symbol	Definition
R	number of sub-regions
$N_{initial}$	initial number of deployed sensors
$N_{total}(t)$	total number of alive sensors at time *t*
$N_i(t)$	number of alive sensors in sub-region *i* at time *t*
$N^*(t)$	required spatial resolution value at time *t*
$P_{on}(t)$	probability that a sensor is active at time *t*
$P_{k-i-on}(t)$	probability that exactly *k* sensors active in sub-region *i* at time *t*
$P_{k-i-\mathrm{cov}}(t)$	probability that at least k sensors active in sub-region *i* at time *t*
$P_{k-\mathrm{cov}}(t)$	probability that k-coverage is achieved at time t

The same cluster-based network topology is considered for this probability analysis where a cluster head is located in the communication range of all sensors in each cluster. All sensors can send their data directly (in one hop) to their corresponding cluster head. The WSN cluster under analysis is divided into *R* virtual sub-regions to investigate coverage. Sensors are initially deployed in a random but uniform manner over those sub-regions, i.e., the initial number of sensors in each sub-region is equal to $N_{initial}/R$. Sensing range of each sensor spans the sub-region in which the sensor is located and the communication range of each sensor spans the whole cluster. Consequently, full coverage can be provided when at least one active (data-sending) sensor exists in each sub-region and k-coverage requires at least k active sensors in all of the R sub-regions. Before introducing the analysis on k-coverage and spatial resolution, some definitions are given in Table 5.1.

The exact probability of being active at time *t,* that is, $P_{on}(t)$, is very difficult to compute since it was shown in Kay and Frolik (2004) that the number of states in Markovian modeling of the system is equal to $2^{Ntotal} \times G^{Ntotal}$ where *G* is the number of automaton states for each sensor. Therefore, the value of $P_{on}(t)$ will be approximated based on the following intuition. The control method based on the ACK method of Kay and Frolik (2004) tries to maintain *N*(t)* number of active sensors by adjusting transmit probabilities of each sensor. Simulation results show that the proposed method is able to attain *N*(t)* active sensors out of $N_{total}(t)$ alive sensors almost during the whole life of the sensor network. Therefore, on the average, probability of a sensor's being active is approximately equal to $N^*(t)/N_{total}(t)$. So, it will be assumed that the approximate value of $P_{on}(t) \approx N^*(t)/N_{total}(t)$.

Now, coverage probability can be computed by determining the probability of having at least one active sensor in all of the sub-regions. Probability of having at least one active sensor in any of the sub-regions and probability of having exactly zero active sensor in that sub-region must sum up to one. So, $P_{1-i-\mathrm{cov}}(t) = 1 - P_{0-i-on}(t)$. Having exactly zero active sensor in a sub-region *i* equals to the case that all of the N_i alive sensors are non-transmitting. So, $P_{0-i-on} = (1-P_{on})^{Ni}$ and $P_{1-i-\mathrm{cov}} = 1-(1-P_{on})^{Ni}$. An approximate value for P_{on} is already computed and it is needed to determine the value of N_i. Because initially sensors are deployed uniformly over the sub-regions and sensors transmit independent of each other, it may be assumed that deaths of sensors due to battery exhaustion will occur in a uniform fashion for each sub-region and this will preserve the uniform distribution of sensors on the sub-regions over time. Therefore, the following approximation can be made: $N_i(t) \approx N_{total}(t)/R$. This

leads $P_{1-i-\text{cov}} = 1-(1-\frac{N^*}{N_{total}})^{Ntotal/R}$. This expression does not yet give the coverage probability since it is the probability that at least one active sensor exists in only one of the sub-regions. In order to determine the coverage probability $P_{1-\text{cov}}$, it is needed to compute the probability of having at least one transmitting sensor in all of the R sub-regions. Therefore, $P_{1-\text{cov}} = (P_{1-i-\text{cov}})^R = \left(1-(1-\frac{N^*}{N_{total}})^{Ntotal/R}\right)^R$ **(2)**.

Having determined the 1-coverage probability, it will now be tried to verify the fair 1-coverage performance of the proposed control strategy illustrated in Figure 5.4. In fact, coverage probability $P_{1-\text{cov}}$ for this simulation setting must be very close to 1 since there is one active sensor in all of the four sub-regions most of the time as seen in the figure. For that simulation, the time average of the spatial resolution value was 24 starting with a number of $N_{initial} = 100$ deployed sensors. During initial time epochs where all of the 100 sensors are alive, $P_{on} = 24/100 = 0.24$ and $N_i = 100/4 = 25$, and therefore this yields $P_{1-i-\text{cov}}(t_{small}) = 1-(1-0.24)^{25} = 0.9989..$ Towards the end of the network lifetime where only 40 of 100 sensors are alive, for example, $P_{on} = 24/40 = 0.6$ and $N_i = 40/4 = 10$ and $P_{1-i-\text{cov}}(t_{large}) = 1-(1-0.6)^{10} = 0.9998..$ Then, it can be computed that the coverage probabilities at these times as follows: $P_{1-\text{cov}}(t_{small}) = (0.9989)^4 = 0.9956$ and $P_{1-\text{cov}}(t_{large}) = 0.9992$. These probability values which are very close to one, explains how the proposed method maintains at least one active in each sub-region during most of the time.

Actually, 1-coverage probability $P_{1-\text{cov}}$ will usually be large independent of the network parameters due to the following: When $N_{total}(t)$ is large at early times of the network operation while most sensors are alive, the exponent N_{total}/R in the expression for $P_{1-i-\text{cov}}$ will be large. So, even if $1-N^*/N_{total}$ is close to 1, the overall exponential value $(1-N^*/N_{total})^{Ntotal/R}$ will be very small making $P_{1-i-\text{cov}}$ close to 1 which in turn causes the coverage probability to be large. Similarly, when $N_{total}(t)$ is small towards the end of the network life where most sensors have died, N^*/N_{total} will be large causing $1-N^*/N_{total}$ to approach to zero. Therefore, $P_{1-i-\text{cov}}$ becomes close to 1 again making a large coverage probability.

So far, an approximate probabilistic analysis is provided to relate coverage and spatial resolution which explains the fair coverage performance of the proposed control method. Now, this analysis will be extended to include *k*-coverage. The concept of *k*-coverage is generally defined as having every point in the sensor network to be in coverage range of at least *k* sensors. In this setting, this maps to having at least *k* active sensors in each sub-region, i.e., $P_{k-\text{cov}}(t) = (P_{k-i-\text{cov}}(t))^R$. Probability of having at least *k* active sensors in a sub-region is equal to the sum of the probabilities for having exactly *k, k+1,..., Ni* sensors. This yields the following equation: $P_{k-i-\text{cov}} = \sum_{m=k}^{N_i} P_{m-i-on} = 1 - \sum_{m=0}^{k-1} P_{m-i-on}$

In order to determine the probability of having exactly *k* sensors in a single sub-cluster, Binomial distribution is used as follows.

$P_{k-i-on} = \binom{N_i}{k} \times (P_{on})^k \times (1-P_{on})^{N_i-k}$. This will lead to $P_{k-i-\mathrm{cov}} = 1 - \sum_{m=0}^{k-1} \binom{N_i}{m} \times (P_{on})^m \times (1-P_{on})^{N_i-m}$ And, at the end, this results in

$$P_{k-\mathrm{cov}} = (1 - \sum_{m=0}^{k-1} \binom{N_i}{m} \times (P_{on})^m \times (1-P_{on})^{N_i-m})^R \quad \textbf{(3)}$$

where $P_{on} = \frac{N^*}{N_{total}}$ and $N_i = \frac{N_{total}}{R}$.

The expression in Equation 3 is the approximate probabilistic relationship between *k*-coverage and spatial resolution that is sought. Please note that this expression reduces to the one for 1-coverage given in Equation 2 when one substitutes *k*=1 in the equation.

As it has been done for 1-coverage case in the previous paragraphs, some numerical observations will be made regarding *k*-coverage. For this, the curves in Figure 5.5 will be used where the plot $P_{k-\mathrm{cov}}$ is drawn for *k*=1, *k*=2 and *k*=3 against N_{total} ranging from $N_{total}(t_{l\arg e}) = 24$ to $N_{total}(t_{small}) = 240$ for a constant resolution value *N**=24 and *R*=4 sub-regions. As seen from the figure, *k*-coverage probability decreases for all values of N_{total} as *k* increases.

This is intuitive since a more stringent coverage requirement is harder to achieve. And, this behavior is also in accordance with Equation 3 since one more subtraction is performed for the expression inside the parenthesis for each increasing *k* value. What is not intuitive about the behavior of these plots is the decrease of k-coverage probability for increasing values of number of alive sensors N_{total}. As seen, *k*-coverage probability is less during the initial times of the network operation, i.e., at t_{small} where N_{total} is large and it gets increased as

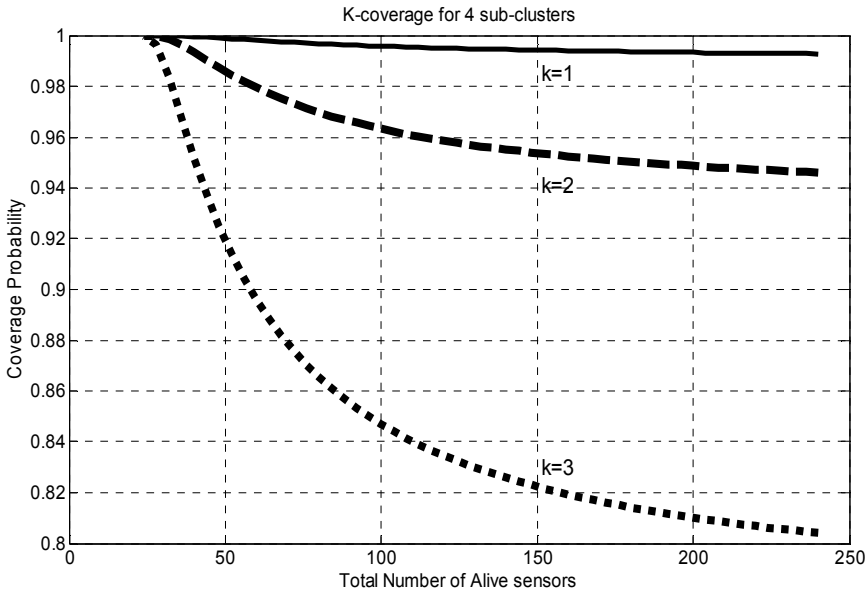


Figure 5.5: *k*-coverage probability versus number of alive sensors

the number of alive sensors drops towards the end of network life at t_{large}. This behavior can be explained as follows: In order to provide the desired spatial resolution value, *N** number of sensors must be active. While the total number of alive sensors is very large at network start, these *N** transmitting sensors selected from N_{total} sensors may have accumulated over some sub-region of the cluster whereas some sub-regions might not have any active sensors due to N_{total}'s being comparably larger than *N**. However, during the end of network operation, only a small number of alive sensors are available and almost all of those N_{total} sensor nodes must become active to attain the desired resolution level of *N**, i.e., $N_{total} \approx N^*$. Since almost all of the available sensors are transmitting, probability of having active sensors in all of the sub-regions is higher which means a higher *k*-coverage probability.

Now, it will be tried to verify the results produced from the approximate probability analysis by simulation. For this, a sensor network cluster is simulated

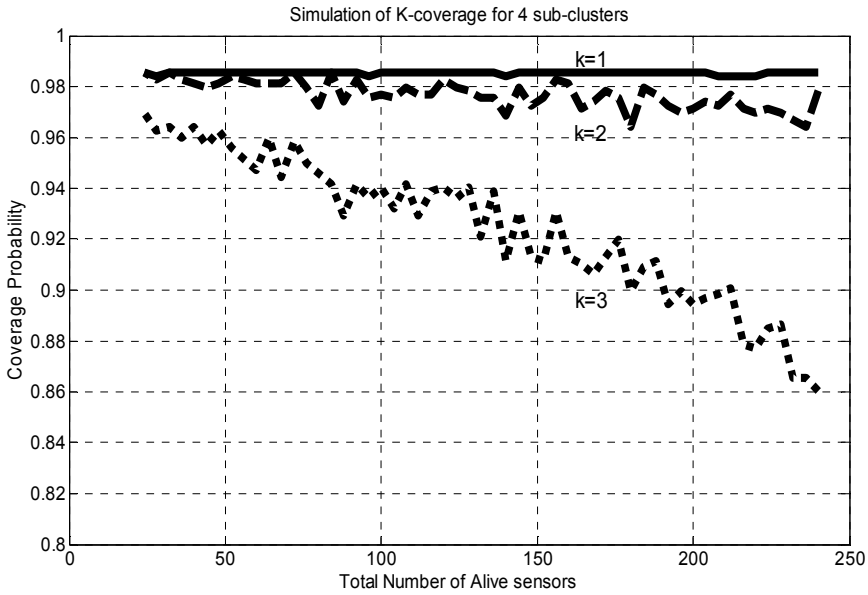


Figure 5.6: Simulation results for *k*-coverage probability

starting with an initial deployment of 240 sensors until there remain 24 sensor nodes due to deaths caused by battery exhaustion. For each *k* value (*k=1,2,3*) and *N*=24* with *R=4* sub-regions, 50 simulations have been performed and the number of instances when the required coverage is attained has been recorded. Results are shown in Figure 5.6. Same behavior regarding the change of *k*-coverage probability with respect to both *k* and N_{total} is observed in simulation plots. Yet, the coverage probabilities produced by simulations are slightly higher than the values computed by the approximate probabilistic expressions of Equation 2 and 3, particularly for *k*>1 and at large values of N_{total}. This is most probably because of the assumption on the preservation of fair distribution of alive sensors over sub-regions together with the multiplication of *k*-coverage probabilities of each sub-region $P_{k-i-\mathrm{cov}}$ to get $P_{k-\mathrm{cov}}$ which both cause overall coverage probability to artificially decrease.

The final comment on the *k*-coverage probability based on above plots is that although the proposed method based on ACK method of Kay and Frolik (2004) seems to provide sufficiently high (greater than 0.98), 1-coverage probabilities during the entire life of a sensor network, coverage probabilities for $k>1$ are not satisfactory enough, e.g., it drops below 0.9 for $k=3$ case. Therefore, the probabilistic assurance of the proposed method of Section 5.1 not enforcing any direct control on *k*-coverage might not be sufficient for some networks where requirements regarding coverage are tighter. In the next paragraph, a modification on the original ACK-based QoS and security control strategy will be presented to improve *k*-coverage performance in a considerable amount.

To be able to provide an assurance on *k*-coverage, it is needed to control the number of data transmitting sensors in each sub-region of the WSN cluster. If more than k active sensors can be kept in each of the sub-regions, *k*-coverage can be achieved. Since the originally proposed method already provides a means for controlling the number of active sensors in the whole cluster, what is needed is to change this strategy for controlling each individual sub-region instead of the overall cluster.

For this, slight modifications will be made in the proposed control method. Before presenting the modifications in the control strategy, it is needed to mention two changes in the originally proposed MAC frame illustrated in Figure 3.3 of Chapter 3. First, in addition to the source ID information which nodes announce during the contention period of the MAC frame while they show their intent to transmit, each node will also include its location information, i.e, ID of the sub-region it resides in, as seen in Figure 5.7. And secondly, cluster head should be able to send reward/punish information separately for each sub-region so that it can control the number of active sensors in those sub-regions. Therefore, Header sub-slot of the MAC frame is modified to include an R-bit reward/punish information which was 1-bit in the original protocol. Under this MAC scheme, the steps of the modified ACK-based QoS and security control method to enhance k-coverage are given

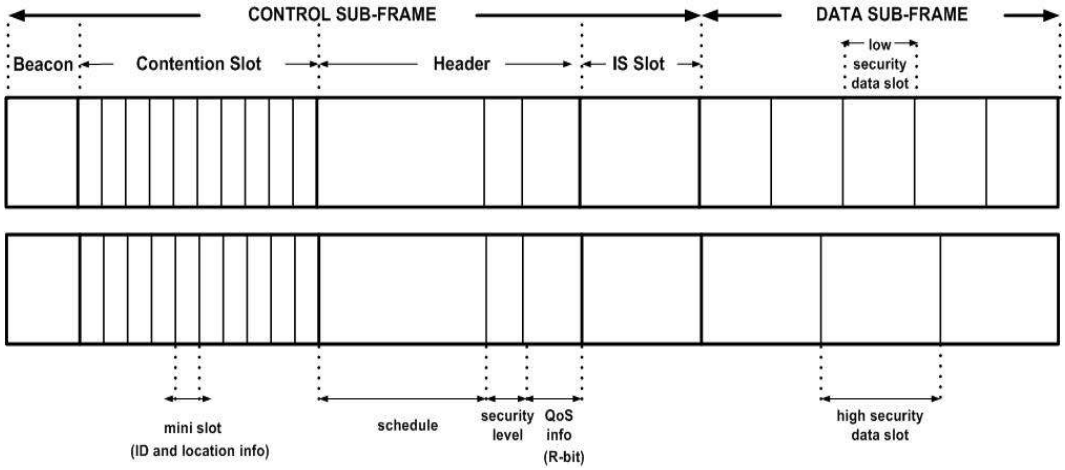


Figure 5.7: MAC Frame format of the coverage-enhanced ACK-based method (2 frames are shown)

below. Please note that those steps periodically occur at each time epoch and the first five steps are exactly the same as the original control method.

1. Cluster head (CH) starts transmitting the beacon message.
2. CH checks whether there is a change in the required security and spatial resolution levels *(S*,N*)* which are announced by the control center of the sensor network. If there is a change in either *S** or *N** with respect to previous epoch, CH proceeds to step 3, otherwise it goes to step 6.
3. CH checks whether new security and spatial resolution requirements are supported. If they are supported it goes to step 6. If required levels *(S*,N*)* are not supported, it computes the optimal supported levels *(S', N')* and then proceeds to step 6.
4. Before the beacon period ends, each and every nodes decides whether to transmit or not during the current epoch. Nodes make this decision in the same way as in the ACK strategy, i.e., by comparing their locally generated random number to the transmit probability of their current state. Nodes which decide to transmit open their radio, synchronize with the beacon and proceed to step 5. Others switch to the sleep state.

5. After the beacon period ends, nodes deciding to transmit in the previous step contend for a mini slot in the contention slot.
6. Before the transmission of Header packet, CH should have finished the calculation of optimal security and spatial resolution values *(S', N')*. Also, in this step, CH determines the number of sensors in each sub-region *i* that intend to be active for the current epoch by counting the number of accessed contention mini slots. Number of sensors desiring to transmit in each sub-region is represented as N_i and sum of them, that is, $\sum_{i=1}^{R} N_i = Nt$ represents the current (expected) level of the spatial resolution of the current epoch t and will be compared to the desired value of N^*.
7. During the Header period, CH first unicasts the schedule of data transmissions for the current frame. This schedule is an ordered list of sensor nodes prepared according to the order in contention mini slot access and it also includes the slot duration *Ds* corresponding to the desired security level of current epoch. Before announcing the schedule, CH should check that the number of sensors desiring to transmit *Nt* do not exceed the supported (or desired) number of active sensors *N'* (or N^*). If *Nt* exceeds *N'* (or N^*), CH chooses only *N'* (or N^*) of sensor nodes in a round-robin fashion, that is, one sensor from each sub-region at each turn until *N'* (or N^*) is reached to provide uniform distribution of active sensors over each sub-region. CH includes only those *N'* (or N^*) selected sensors in the announced schedule.
8. After the schedule is announced, CH informs nodes that intend to transmit about two more issues during the Header period. The first one is the desired security level of the current epoch (S^* or *S'*) and the second one is the information on whether the current spatial resolution *Nt* is above or below the desired level N^* (or *N'*). The second piece of information is R-bit, 1-bit for each sub-region. If $Nt \leq N^*$, corresponding bit for sub-region *i* is set to 1 if $N_i \leq N^*/R$ or else set to 0. If $Nt > N^*$, corresponding bit for sub-region *i* is set to 1 if $N_i \leq$ k and, 0 otherwise.

9. All of the *Nt* nodes receiving the 1-bit information regarding the comparison of current and desired spatial resolution levels for their corresponding sub-region, change their state in the finite state automaton. If this value is 1, they reward themselves by jumping to a state with higher transmission probability, otherwise, they punish by jumping to a lower-transmit probability state.
10. Of *Nt* nodes which changed state in the previous step, the ones which are not listed in the announced transmission schedule go into sleep. Only the nodes which find their name in the schedule transmit their packets at the security level announced and in the data slot assigned to them.
11. After the data sub-frame ends, all sensor nodes return to step 4 and CH returns to step 1.

Using this modified strategy, a simulation for the same sensor network setting of the previous case is performed. The achieved spatial resolution levels in each sub-region are shown in Figure 5.8. As the plots indicate, the number of active sensors in each sub-region is very balanced and uniform throughout the simulation period. In fact, each sensor is able to attain $N^*/R=24/4=6$ active sensors almost at all times which will provide a very high *k*-coverage probability even for the high values of *k* such as 4, 5 and 6. Therefore, this proves that the modified strategy of this section has a superior *k*-coverage performance compared to the previous method given in Section 5.1. This *k*-coverage performance is one of the motivations to propose a novel QoS and control strategy for improving the method presented in this chapter, which was mainly built on the ACK-based strategy of Kay and Frolik (2004). This novel strategy is presented in the next chapter.

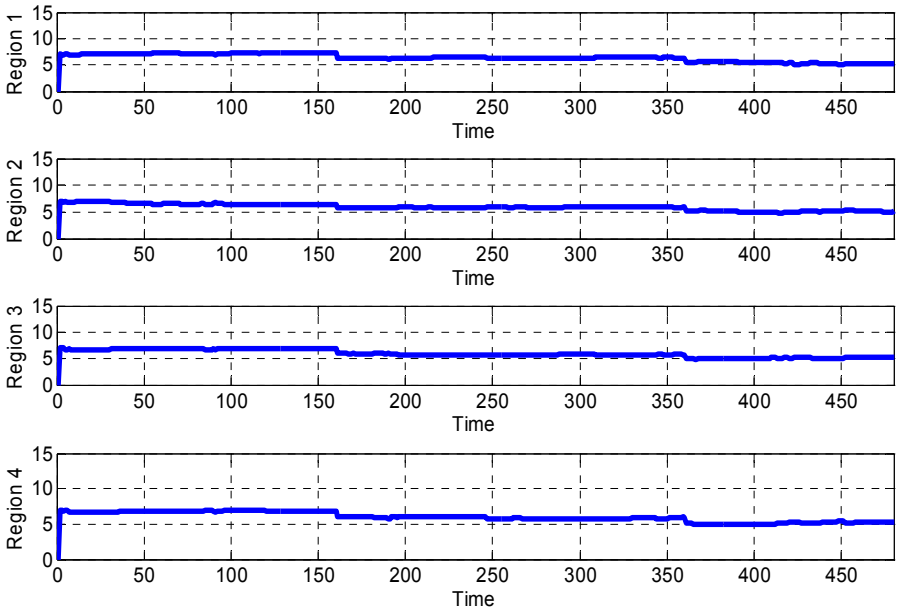


Figure 5.8: Distribution of active sensors on sub-regions for the coverage-enhanced ACK-based method

CHAPTER 6

THE NOVEL QoS AND SECURITY CONTROL METHOD

In the previous chapter, a method is proposed, based on a previous work, i.e., Kay and Frolik (2004), to provide a solution to the following problem: How a cluster-based sensor network can be controlled in such a way that the time-varying security and QoS requirements are fulfilled during the entire operation. There were five main objectives: (1) to keep enough number of sensor nodes active (ON) to attain the desired spatial resolution level, (2) to have these active sensors communicate at the required security level, (3) to maximize the network lifetime by having active sensors periodically power down and inactive ones power up for a balanced energy dissipation, (4) to provide full coverage by having at least one sensor taking measurements on each geographical region and, (5) to minimize the packet loss resulting from collisions. In this chapter, another method will be proposed to achieve all these five security and service quality objectives. This novel strategy aims not to have the drawbacks of the previously proposed method based on the ACK strategy of Kay and Frolik (2004). To elaborate, it is sought to design a new QoS and security control strategy for wireless sensor networks which will provide closer values to the required spatial resolution levels, longer network lifetime, and better *k*-coverage performance. Thus, the method to be presented in this chapter, which will be designated as the proposed method of the

thesis, aims to enhance three of the included QoS attributes, namely, spatial resolution, network lifetime and coverage, compared to the values achieved in the previous acknowledgement based method. Besides, it will preserve the security providing and packet collision minimizing features of the previous method.

As just mentioned in the paragraph above and also as illustrated by the simulation results presented in the last chapter, there are some problematic issues in the ACK-based control algorithm. One of those problems is the following: Though it is able to have all packets transmitted at the required security level, the control method of the previous chapter cannot make the network attain exactly the desired spatial resolution levels, i.e., at several times, the achieved resolution value is different than the required one, as can be seen from the simulation results given in Section 5.2.3. This is due to the large non-zero spatial resolution variance values of the Markovian modeling of the ACK-based network detailed in Kay and Frolik (2004). Another issue in the ACK-based control method is the unequal participation of available sensor nodes in the data transmission process. In other words, for some periods, some sensors transmit packets more frequently while others hardly ever transmit. This results in unbalanced battery dissipation among the nodes and causes some nodes to die sooner. Although previously less active nodes start transmitting instead of the dead nodes after a while, it takes some time for the network to reach back to the desired resolution value. What is worse, there is a tradeoff between variance and equal participation of nodes, diversity as called in Kay and Frolik (2004), and this fact makes it very difficult to provide the required spatial resolution and balanced power usage for lifetime maximization at the same time for the ACK-based control method. On top of those deficiencies, the original acknowledgement based method of Kay and Frolik (2004) does not consider coverage at all. Though the statistical nature of the ACK strategy helps a balanced distribution of active sensor nodes over the region as shown in Figure 5.4, thus resulting in a fair 1-coverage performance, this is not enough to provide an assurance on the *k*-coverage.

The mentioned drawbacks of the previously proposed method are all due to the utilization of the ACK-based strategy of Kay and Frolik (2004). Therefore, in this chapter, a novel QoS and security control strategy not based on the ACK method is proposed. This new method is designed to be free from the limitations of Kay and Frolik (2004), that is, to provide the desired resolution at all times, to have even power consumption among sensors for system lifetime extension and to assure a certain *k*-coverage level.

6.1. FUNDAMENTALS OF THE NOVEL QoS AND SECURITY CONTROL METHOD

As stated previously, the control strategy proposed in this thesis aims to provide five attributes for cluster-based wireless sensor networks, which are security, spatial resolution, maximal network lifetime, minimal packet loss and sufficient coverage. To provide best performance for the achievement of those attributes, the following principles are applied while designing the new control strategy presented in this chapter:

- If the method can cause enough (more than *N*)* number of sensors to show their intent to transmit at each MAC frame, then the cluster head can choose a certain number (exactly *N**) of sensors to transmit. This will provide the desired *spatial resolution* value at each frame duration. Thus, in this new method, each sensor node *i* will independently decide to transmit or not for each frame duration by comparing its locally generated random number to a probability value *Pi*. To make more than *N** sensors intend to transmit, each node will update its probability value at each frame using the following rules.
    - If a node has decided to transmit and its name is included in the data transmission schedule, the probability value *Pi* for this node will not be changed. (Since this case will usually occur when the number of nodes intending to transmit is just fine to provide the required spatial resolution level, there is no need to change transmit probabilities.)

- If a node has decided to transmit but its name is not included in the data transmission schedule, the probability value *Pi* for this node will be decreased. (Since this case will usually occur when the number of nodes intending to transmit is above the required spatial resolution level, transmit probabilities must be decreased.)
- If a node has not decided to transmit, the probability value *Pi* for this node will be increased. (This is to prevent nodes from remaining passive for long periods of time and aims to provide equal power consumption of available sensors.)

- In order to further contribute to the balanced energy dissipation, the new method will use the battery level information of sensor nodes which they will state during the contention period. This battery level information will be just two bits for each node, which makes up four battery levels such as very low, low, high and very high. Thus, among the nodes accessing contention mini slots, the ones having the highest battery levels will be selected to transmit. This way, battery consumption of nodes will be evenly distributed in time providing *longer lifetime*.

- In order to provide an assurance on *k-coverage*, it is needed to control the number of data transmitting sensors in each sub-region of the WSN cluster. If one can keep more than *k* active sensors in each of the sub-regions, then *k*-coverage can be achieved. For this, in addition to the source ID and battery level information which nodes will announce during the contention period of the MAC frame, each node will also include its location information, i.e, ID of the sub-region it resides in. This location information will contain enough bits to make up *R* sub-regions, e.g., 3 bits for 8 sub-regions. In order to provide the desired *k*-coverage level, the new control algorithm will select at least k sensor nodes from each sub-region among the nodes that request to transmit. Details on how battery and location information used together in the selection of active

nodes is explained in the operational steps of the proposed strategy given in the next subsection.

- As before, in the Header period of the MAC frame, the desired security level of the current period is announced to the nodes that will transmit data for that frame. All sensors should send their data with a message integrity code corresponding to the announced security level. This will provide *security* of sensor to cluster head communications.

- Since the slotted MAC algorithm that is used allows transmission of only the selected nodes in their corresponding data slots, there is no risk of collision during the data sub-frame. Also, as in the previous ACK-based method, the number of contention slots is designated to be sufficiently, i.e., e times higher than the number of data slots and this will further reduce the collisions that can occur during control sub-frame. Therefore, the proposed control method will *minimize packet loss due to collision*.

Since this new control strategy will utilize the location and battery information of sensor nodes, it is required to make sensors to announce this information during the MAC frame to let the cluster head be aware of the sub-region and battery level of each node. In addition, it is not needed to send any QoS information, i.e., punish/reward bit, as in the previous control scheme of Chapter 5, which was based on the acknowledgement method of Kay and Frolik (2004). Therefore, some modifications are required on the originally proposed MAC scheme of Chapter 3, which was based on TRACE (Tavli & Heinzelman, 2003).

The frame format of the modified MAC scheme of the new control method of this chapter is given in Figure 6.1. Different from the MAC scheme of the acknowledgement based method illustrated in Figure 3.3 where nodes transmit only their source ID during the contention mini slots, in the modified MAC scheme, nodes give two more pieces of information during this period, which are

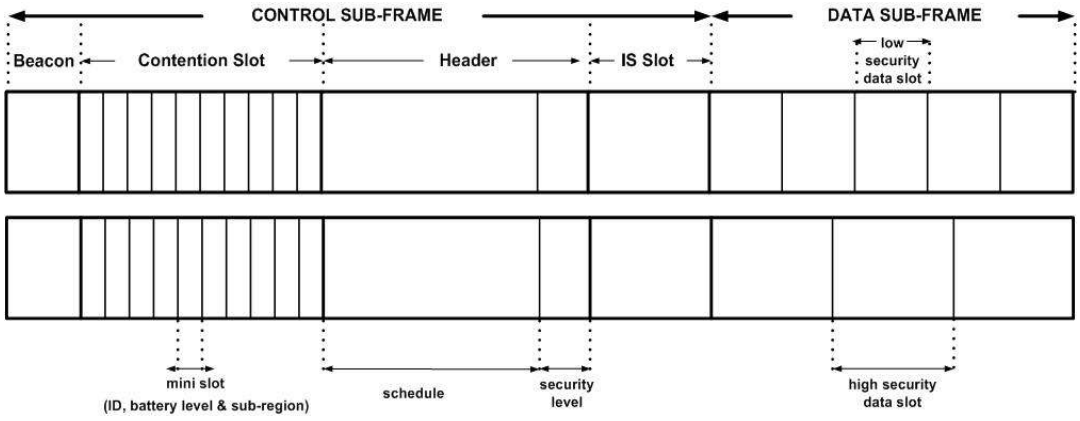


Figure 6.1: Frame format for the new method's MAC frame

their remaining battery level and geographical sub-region in which they reside. As mentioned before, the battery level information is just two bits for each node, which makes up four battery levels such as very low, low, high and very high. Besides, it is assumed that the overall sensor network cluster is divided into R sub-regions and nodes know in which sub-region they reside, e.g., via GPS (Global Positioning System). In fact, for settings where the locations of sensor nodes do not change during the operation, cluster head can use a pre-loaded table which includes the location information of nodes corresponding to their source ID's. In such a case, nodes do not need to send their sub-region information at each frame.

If a node succeeds to win in the contention period (i.e., no other sensor chooses the same mini slot) and if it finds its ID in the schedule announced in the Header period, the contending sensor node can transmit its packet in the data transmission period without any collision risk. Following the contention period, the controller, i.e., cluster head, transmits the Header, which includes the data transmission schedule for the current frame. The nodes that will be included in the schedule are selected from the successfully contending nodes based on their battery level and sub-region information. In fact, cluster head will select k highest battery level nodes from each sub-region and announce these selected nodes in the Header. The

IS slot follows the Header slot and is used for partitioning of the network. Control sub-frame ends with the IS slot and data sub-frame begins.

6.2. OPERATIONAL STEPS OF THE NOVEL QoS AND SECURITY CONTROL METHOD

Before proceeding with the operational steps of the proposed quality of service and security control strategy of this chapter, here are some final remarks about the setting. The initial values for the probability value *Pi* for each node *i* is set to *Nmax/Ninitial* where *Nmax* is the maximum defined spatial resolution value and *Ninitial* is the total number of sensors initially deployed. The increment *inc* that will be used to update *Pi* yields better simulation results when set to any value between 0.05 and 0.25. Time is divided into discrete intervals named as epochs. Duration of each epoch is equal to one frame duration of the MAC scheme and epochs are synchronized with frames. During each epoch, the following events occur in the given order.

1. Cluster head (CH) starts transmitting the beacon message.
2. CH checks whether there is a change in the required security and spatial resolution levels *(S*,N*)* which are announced by the control center of the sensor network. If there is a change in either *S** or *N** with respect to previous epoch, CH proceeds to step 3, otherwise it goes to step 6.
3. CH checks whether the new security and spatial resolution requirements are supported. If they are supported it goes to step 6. If required levels *(S*,N*)* are not supported, it computes the optimal supported levels *(S', N')* and then proceeds to step 6.
4. Before the beacon period ends, each and every nodes decides whether to transmit or not during the current epoch. Nodes make this decision by comparing their locally generated random number to the current value of probability *Pi*. Nodes which decide to transmit open their radio, synchronize

with the beacon and proceed to step 5. Others switch to the sleep (standby) mode.

5. After the beacon period ends, nodes deciding to transmit in the previous step contend for a mini slot in the contention slot by sending their ID number, 2-bit battery level and sub-region information during a mini slot.
6. Before the transmission of the Header packet, CH should have finished the calculation of optimal security and spatial resolution values *(S', N')*. Also, in this step, CH determines the source IDs, battery levels and sub-regions of sensors that request to be active for the current epoch by checking the accessed mini slots of the contention period. Number of nodes desiring to transmit will be represented as *Nt* for epoch *t*.
7. During the Header period, CH unicasts the schedule of data transmissions for the current frame. This schedule is an ordered list of sensor nodes with corresponding node ID's and it also includes the slot duration *Ds'* (or *Ds**) corresponding to the supported (or desired) security level of the current epoch. CH determines the sensor nodes to be included in the data transmission schedule in the following way. If *Nt* is smaller than both *N'* (or *N**), the supported (or desired) number of active sensors, CH includes all of the *Nt* nodes in the data transmission schedule regardless of their sub-regions and battery levels. If *Nt* exceeds *N'* (or *N**), then CH chooses *min(k,N'/R)* (or *min(k,N*/R))* highest battery-level sensor nodes from each sub-region. The "minimum" operator is used to prevent the unbalanced distribution of active nodes over sub-regions for cases when the coverage requirement *k* is not in agreement with the current spatial resolution requirement, i.e., above *N'/R* (or *N*/R*). If the total number of selected nodes is less than *N'* (or *N**), then CH also includes the unselected highest-battery sensor nodes regardless of their location until the total number of the selected sensor nodes sum up to *N'* (or *N**). If battery levels of two or more nodes are the same, CH makes a random selection among them.

8. After the schedule is announced, CH informs nodes that request to transmit about one more issue during the Header period, which is the desired security level of current epoch (*S** or *S'*).
9. Each alive sensor node *i* updates its probability value *Pi* in the following way. If it has not desired to transmit for this epoch, then it sets $Pi = \min(Pi + inc, 1)$. If it has desired to transmit but its name was not announced in the data transmission schedule, then it sets $Pi = \max(Pi - 2 \times inc, 0)$. Otherwise, sensor does not modify *Pi*.
10. All the sensor nodes which are not listed in the announced transmission schedule go to sleep. Only the nodes which find their name in the schedule transmit their packets at the security level announced and in the data slot assigned to them.
11. After the data sub-frame ends, all sensor nodes return to step 4 and CH returns to step 1.

6.3. SIMULATIONS FOR THE NOVEL QoS AND SECURITY CONTROL METHOD

In order to see the performance of the proposed QoS and security control method of this chapter whose operational steps are just given above, some simulations have been performed again using the software code written in MATLAB. The simulation assumptions given in Section 5.2.1 are still valid for the simulations performed in this chapter except for the simulation duration and the finite state automaton parameters which were specific to the ACK-based method of the previous chapter. The simulation duration, and therefore the initial battery capacities of sensor nodes are taken to be higher than the ones in the previous chapter in order to better illustrate the increase in overall system lifetime that the new method presented in this chapter provides. The simulation duration is 9600 time epochs and the other simulation parameters for the method of this chapter are initial probability value *Pi(t=0)* and the probability value increment *inc*. The values for those parameters as well as some other are given in the following.

Ninitial=100, *Smin=0, Smax=3, Nmin=15, Nmax=35, Pi(t=0)=0.35, inc=0.05, N0,max = 25, N1,max = 24, N2,max = 22, N3,max =19, Ws=*W_N*=1, Us(S)= 1-(exp(-2.07*S-0.69))* and $U_N(N)=0.025*N+0.125$.

The results of simulations are shown below in Figures 6.2 to 6.5. Simulation results for spatial resolution, lifetime, security, and coverage are included in those plots, respectively. In all the plots, the results produced under the previous ACK-based method are shown in the upper part of the figure whereas the results belonging to the newly proposed method of this chapter are given in the lower part.

Figure 6.2 illustrates the performance of the proposed strategy in controlling spatial resolution and Figure 6.3 in controlling security, under time-varying security and spatial resolution requirements. In Figure 6.2 and Figure 6.3, the dashed lines represent the required levels (*S** and *N*),* dotted lines marked with squares represent the supported levels (*S'* and *N')* computed using the heuristic given in Chapter 4 and continuous lines represent the attained levels (*S* and *N).*

As can be seen from the lower part of Figure 6.2, the newly proposed control method is able to exactly attain the supported spatial resolution level from the beginning until the death of the network when the number of alive sensors is less than the minimum supported resolution level *Nmin*. However, there are several spike-like parts in the attained resolution graph of the ACK-based method. This is mostly due to the previously mentioned non-zero variance and unequal participation of nodes properties of the ACK strategy (Kay & Frolik, 2004). Since the new strategy does not utilize the ACK method and relies on the simple idea of making a little more number of sensors intend to transmit than the required spatial resolution value and then select just the required number of them, it is able to provide a spike-less, smooth appearance for the attained spatial resolution graph. Therefore, the proposed method is much better than the previous ACK-based method in providing spatial resolution. In fact, the proposed method is able to

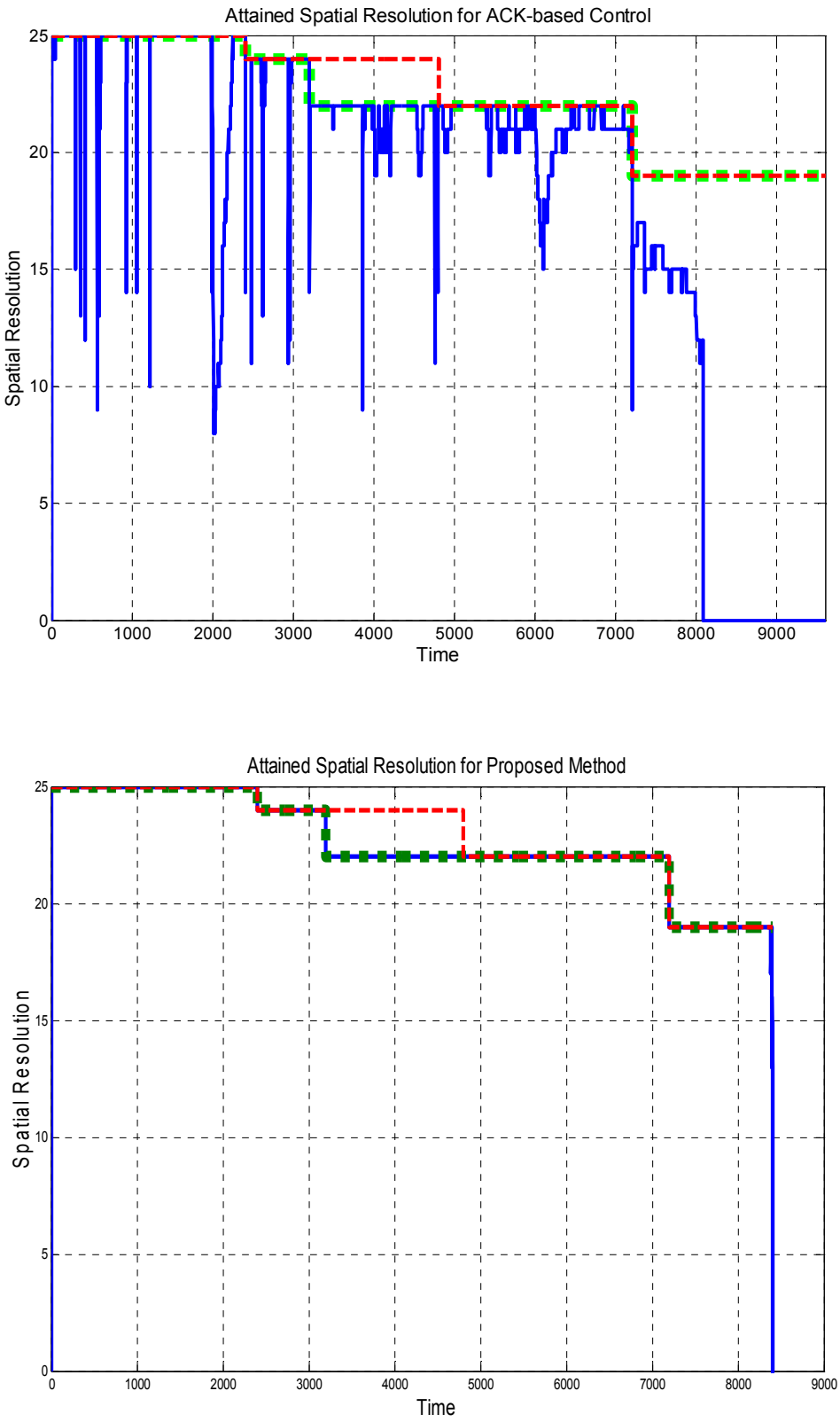
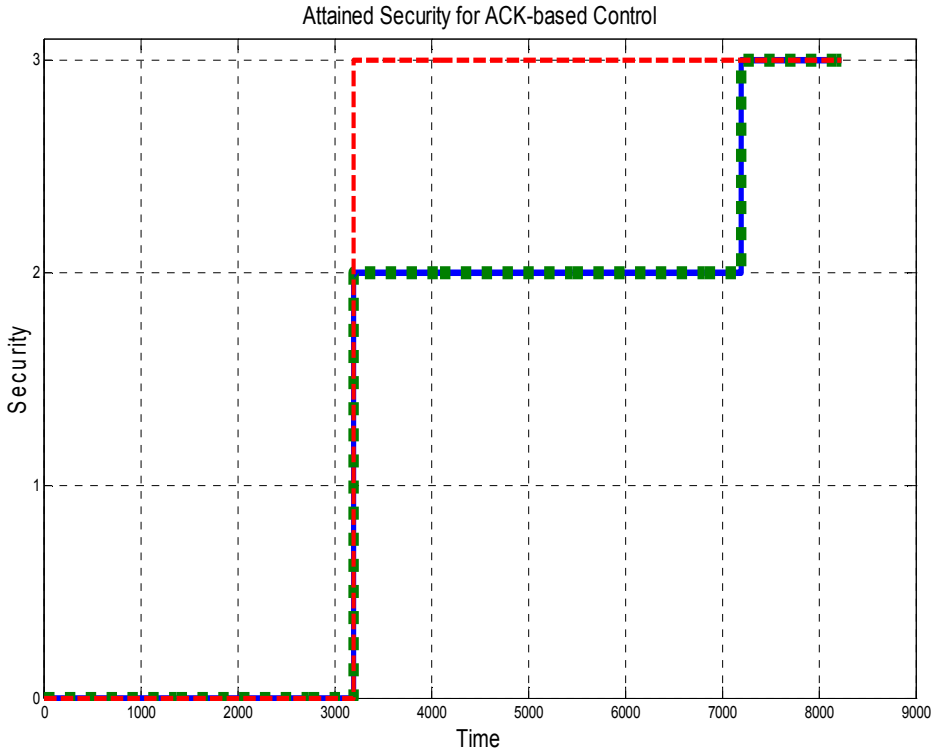


Figure 6.2: Spatial resolution vs time for both methods



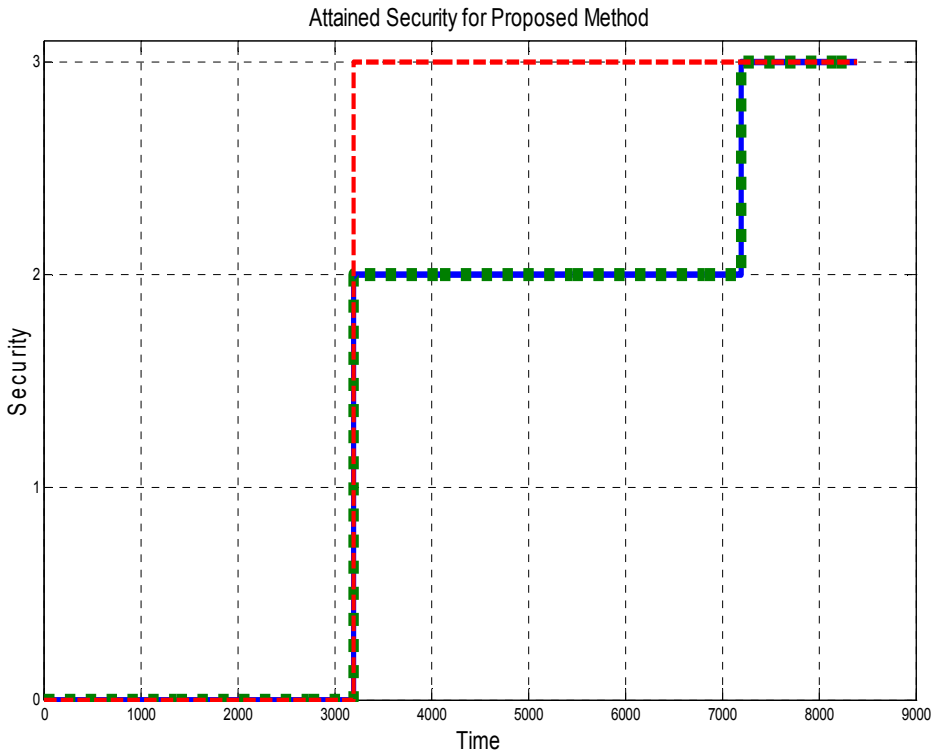


Figure 6.3: Security vs time for both methods

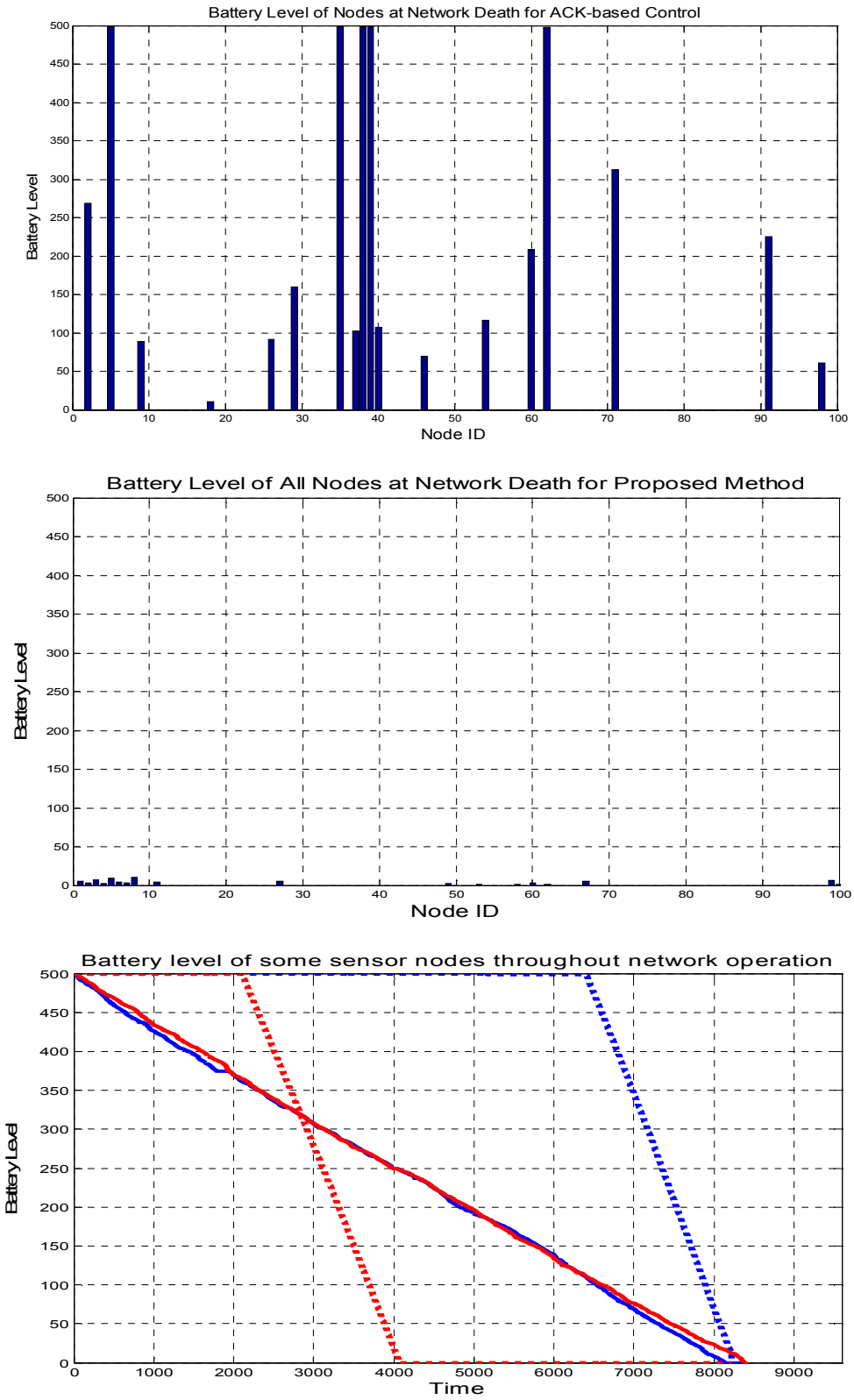


Figure 6.4: Battery level of nodes for both methods

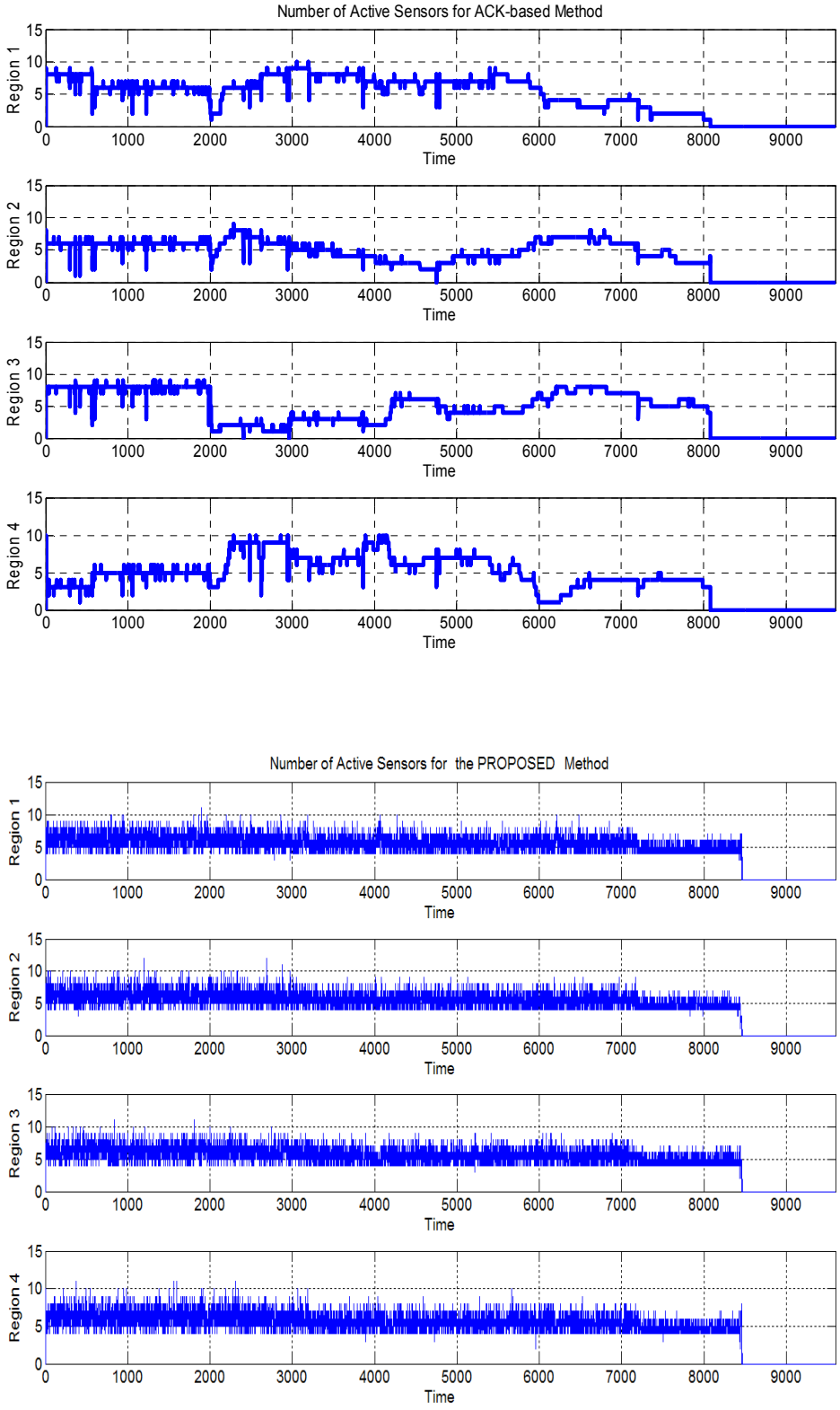


Figure 6.5: Coverage performance of both methods

provide the desired (required/supported) spatial resolution value for 8389 of the 8401 epochs where the network is operational whereas the corresponding figure is 5769 out of 8226 for the ACK-based approach. Another important point to note from those numbers is that the total lifetime of the network achieved with proposed method is longer than the one for the previous method, i.e., 8401 versus 8226 epochs. Before illustrating the main reason behind this lifetime extension, the security levels provided by both methods will be compared. As Figure 6.3 shows, performance of both methods regarding security is the same, that is, the attained level *S* exactly traces the supported security value *S',* because both strategies force all active sensors to transmit at the required or supported security level.

Returning back to the network lifetime, Figure 6.4 gives information about the battery consumption of sensor nodes for the proposed method of this chapter and the ACK-based method of the previous chapter. The top plot of Figure 6.4 shows the battery level of sensors when network controlled by the ACK-based method dies and the middle plot illustrates the same case for the proposed method of this chapter. As can be observed, battery dissipation of sensors under the control of the ACK-based strategy is very unbalanced since living nodes have lots of unused battery. This is due to the previously mentioned low diversity problem of the ACK strategy. Yet, for the new method, at the end of the life of the network, almost all nodes have run out of battery indicating a balanced power dissipation among all sensors. This even distribution of battery levels indicating an equal contribution of sensor nodes to the spatial resolution becomes more obvious in the last plot given in the bottom part of Figure 6.4 which illustrates battery levels of two nodes throughout their lifetime for both methods. Represented by solid lines, batteries of sensors under the control of the proposed method are consumed in a very balanced fashion. Starting with full batteries at startup, both sensors are able to distribute usage of their battery almost until the network dies, indicated by diagonal-like shape of the plots. However, for the ACK-based strategy, some of

the sensors do not use their batteries until a specific time, i.e., does not transmit at all, and thereafter they quickly consume up their battery, most probably due to continuous data transmission for some period. This is illustrated by the two dotted lines in the plot. Therefore, Figure 6.4 indicates that the proposed method of this chapter performs well at providing balanced battery dissipation of all nodes and this results in a longer network life.

Observing the coverage plots given in Figure 6.5, it can be stated that the new method provides considerable improvement in *k*-coverage performance. In fact, it is able to achieve at least 4 active sensors in all of the sub-regions throughout the entire lifetime of the sensor network, providing a *k*-coverage assurance for k=4, as it is set at the beginning of the simulation. However, as seen from the plots, the ACK-based method without location awareness have several times when there is not any active sensor in some sub-regions meaning that it cannot guarantee even *1*-coverage.

During the simulation, the number of packets dropped due to collisions and also the total number of successfully transmitted packets are recorded to determine the packet loss ratio due to collisions, which is *L.* As explained previously, collisions occur only in the contention period and it is assumed that the data packets of both of the colliding nodes accessing the mini-slots in the contention period are dropped. Simulation results show that only 183 packets out of 170346 are dropped due to collisions, which yields an L value of 0.001.

As a consequence, it can be concluded that the QoS and security control method proposed in this chapter is the most successful strategy in satisfying requirements for all five of the design aims of this thesis, which are security, spatial resolution, system lifetime, coverage, and packet loss due to collision. That is why it is designated as “the proposed QoS and security control strategy of this thesis”.

6.4. UTILIZATION OF THE PROPOSED CONTROL STRATEGY FOR SETTINGS NOT REQUIRING SECURITY

As mentioned previously in Section 3.2.2 while explaining the target applications of this thesis study, the proposed control method here can be utilized for WSN settings which are in need of application-level service quality but not requiring security. One example of such settings is habitat monitoring applications which require continuous monitoring of a phenomenon in an environment where certain coverage and data precision requirements exist. Such applications usually do not have strict security requirements owing to the nature of the applications which are mostly academic and research-oriented works. They are neither too much time-critical but they often require spatial precision of the collected data and maximal monitoring time of the environment. Such an application can require the completion of a task such as reporting the temperature and relative humidity for at least one week with a spatial resolution of 25 active sensors over the target area providing a *k*-coverage assurance of degree 2.

The proposed QoS and security strategy of this thesis, operational steps of which were given in Section 6.2, can easily be adapted to satisfy the requirements of such applications requiring only service quality. Indeed, if the security requirement is set to zero, that is, $S^*=0$, and the supported spatial resolution level is equated to the desired resolution level at all times, i.e., $N'=N^*$, the proposed strategy presented in Section 6.2 will satisfy the application QoS requirements by maintaining the desired spatial resolution and *k*-coverage levels during the operation and also extending the overall lifetime of the system. In this case where security is not a requirement, the proposed control algorithm will even be simpler since there is no need to check whether the requirements can be supported by the underlying channel capacity and therefore it is not necessary to compute the optimal supported security-spatial resolution pairs. Consequently, in the MAC frame, it is not required to include any information about the current desired security level.

In fact, for WSN application settings which do not have any security requirements, the proposed control strategy could be improved to better satisfy the service quality requirements. One such enhancement can be achieved if the application level QoS requirement is modified to control the number of sub-regions and the number of active sensor nodes in each sub-region instead of controlling the total number of active sensor nodes in the whole cluster. Such a modification in the definition of the service quality provides a fine grained control over the QoS level that can be achieved. In other words, a requirement statement that is expressed as having an R^* number of sub-regions with k active sensor nodes in each of those sub-regions provides the sensor network with the ability to better distinguish the different features of the physical phenomenon in closely spaced geographical regions since the distribution of active sensors over the WSN field is more strictly controlled. Actually, the previous approach of counting the total number of active sensors in the cluster as the attained spatial resolution level and then making sure that those active sensors cover the entire cluster due to the existence of sufficient number of sensors on each geographical part of the cluster already ensured the fulfillment of both spatial resolution and coverage requirements. Nonetheless, it is apt to handle spatial resolution and coverage as separate entities (a spatial resolution level of 24 and coverage degree of 2) rather than combining them into a single QoS metric (8 sub-regions with 3 active sensors in each) as the latter approach does. Such an approach is also in better agreement with the classical definition of spatial resolution as the ability to distinguish between two closely spaced objects (Sabins, 1997) since the size of the sub-regions each having at least k active sensors (cluster area size / number of sub-regions) can now be specified.

The operational steps of the modified control strategy that allows for the specification of the number of sub-regions R^* and the number of active sensors k in each sub-region are presented below. Note that there are not any items involved

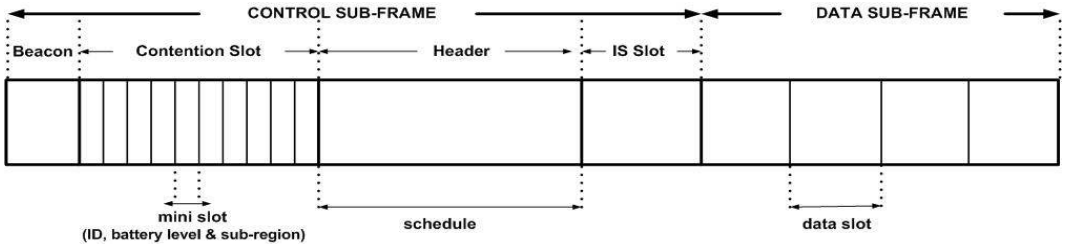


Figure 6.6: Frame format of the MAC scheme for the modified method

with security either in the control strategy or in the MAC frame shown in Figure 6.6 since this modified method addresses the sensor network cases where only application QoS is required. Also note that the cluster head chooses exactly *k* sensors from each sub-region and include no more sensors in the transmission schedule as opposed to the original control method where additional sensors with highest battery levels regardless of their location are included until the total active sensor number reaches the desired spatial resolution level. Consequently, this more detailed specification of the QoS requirements might have a negative effect on the overall lifetime, particularly when R^* is high because sensors with lower battery levels should be selected to transmit in order to include sensors from each of the sub-regions[7].

1. The cluster head (CH) learns the current spatial resolution and coverage requirements $R^*(t)$ and $k(t)$ from the control center and starts transmitting the beacon message.

2. Before the beacon period ends, each and every node decides whether to transmit or not during the current epoch. Nodes make this decision by comparing their locally generated random number to the current value of probability *Pi*.

[7] An example case might occur when a spatial resolution requirement of $N^*=24$ and coverage degree of *2* is desired for the original method and $R^*=8$ and $k=3$ is required for the modified method. For the former method, the algorithm will select 2 nodes from 4 sub-regions and then select the remaining 16 sensors among highest battery nodes. But for the latter method, 3 nodes from each of 8 sub-regions must be selected increasing the possibility of inclusion of nodes with lower battery levels.

Nodes which decide to transmit turn their radio on, synchronize with the beacon and proceed to step 3. Others switch to the standby mode.

3. After the beacon period ends, nodes which decided to transmit in the previous step contend for a mini slot in the contention slot by sending their ID number, 2-bit battery level and sub-region information during the mini slot.

4. Before the transmission of the Header packet, CH determines the source ID's, battery levels and sub-regions of sensors that request to be active for the current epoch by checking the accessed mini slots of the contention period.

5. During the Header period, CH unicasts the schedule of data transmissions for the current frame. This schedule is an ordered list of sensor nodes with corresponding node ID's. CH determines the sensor nodes to be included in the data transmission schedule in the following way: Cluster head chooses the k highest battery-level sensor nodes from each of the R^* sub-regions. If the battery levels of two or more nodes are the same, CH makes a random selection among them.

6. Each alive sensor node i updates its probability value Pi in the following way: If it has not attempted to transmit in this epoch, then it sets $Pi = \min(Pi + inc, 1)$. If it has requested to transmit but its name was not announced in the data transmission schedule, then it sets $Pi = \max(Pi - 2 \times inc, 0)$. Otherwise, the sensor does not modify Pi.

7. All the sensor nodes which are not listed in the announced transmission schedule switch to the standby state. Only the nodes which find their name in the schedule transmit their packets in the data slot assigned to them.

8. After the data sub-frame ends, all nodes return to step 2 and CH returns to step 1.

Simulations are performed for the same topology in Section 6.3 to test this modified method. Two scenarios with two different QoS requirements are tested, the first one being $R^*=4$ and $k=3$ and the second one being $R^*=8$ and $k=2$. The results of these simulations are shown in Figures 6.7 to 6.9. In Figures 6.7 and 6.8, time variation of the number of active sensors in each sub-region is plotted for the first and second QoS requirements respectively. In Figure 6.9, in order to show the overall network lifetime, the total number of active sensors in all sub-regions is plotted for the method of this section and the method of Kay and Frolik (2004) both for the first and the second requirements.

As can be seen from the lower parts of both Figure 6.7 and Figure 6.8, the modified control method of this section is able to attain the required QoS level from the beginning until the network dies when the number of alive sensors is not enough to support the required service quality level. However, the acknowledgement based method of Kay and Frolik (2004) has a lot of variations in time and most of the time the attained number of active sensors is not close to the requirement. Plots shown in Figure 6.9 indicates another important feature of the proposed method, that is, the total lifetime of the network achieved with the proposed method is longer than the one for method of Kay and Frolik (2004), for both cases. In fact, the overall system lifetime when the WSN start not attaining the desired QoS level is 9454 versus 8638 epochs for the first QoS requirement and 7715 versus 7085 for the second one indicating 9.5 % and 8.8 % increases respectively.

So far in this section, it is shown that, when modified properly, the strategy proposed in this thesis is able to provide the QoS requirements of WSN application not requiring security. Simulation results also indicate that it performs much better in providing the required application QoS levels and extending the overall system lifetime when compared to the similar study Kay and Frolik (2004). In fact, in the literature, there exist other studies such as Delicato et al. (2006) and Perillo and Heinzelman (2003a) which provide better performance in

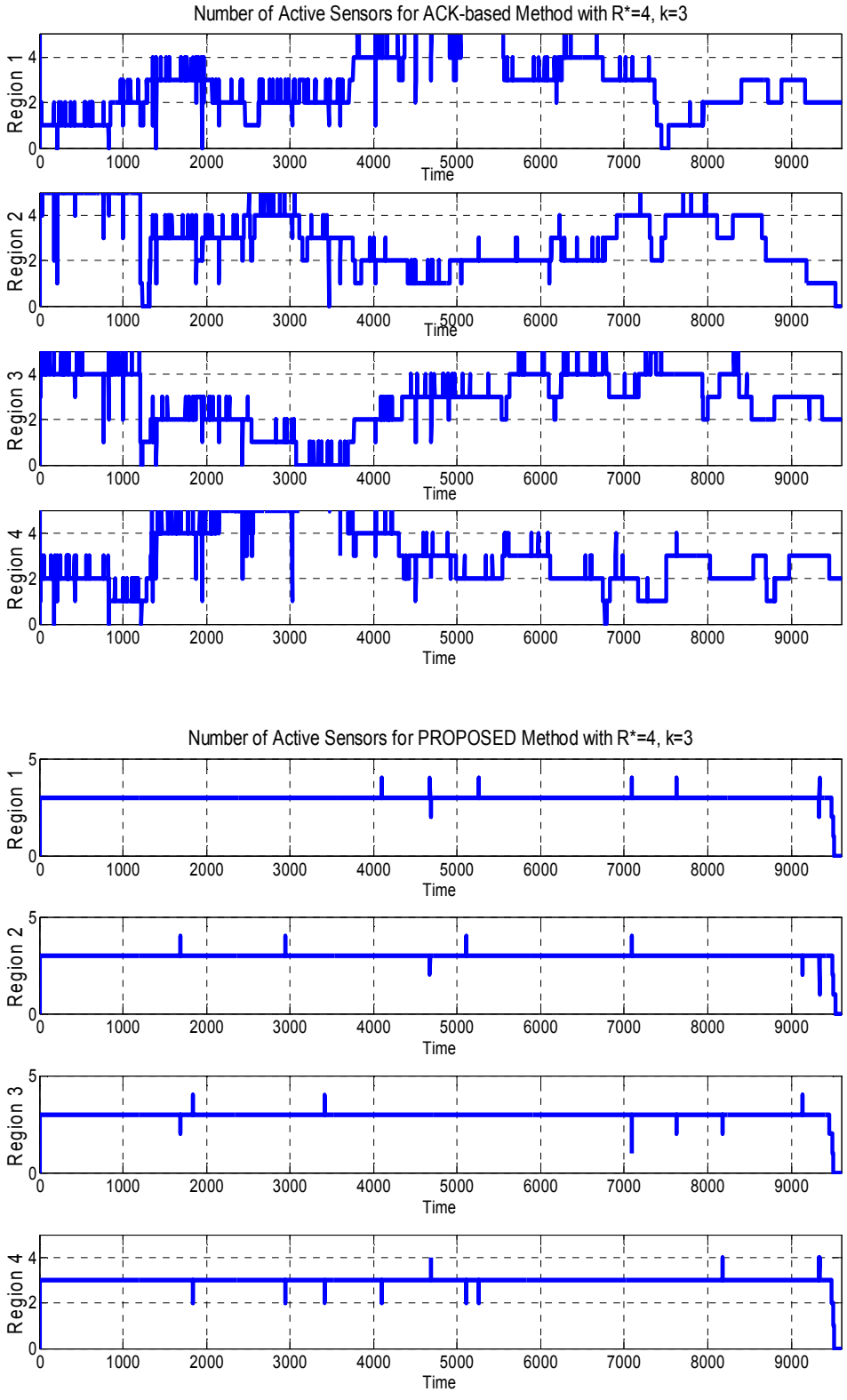


Figure 6.7: Performance of both methods for R*=4 and k=3

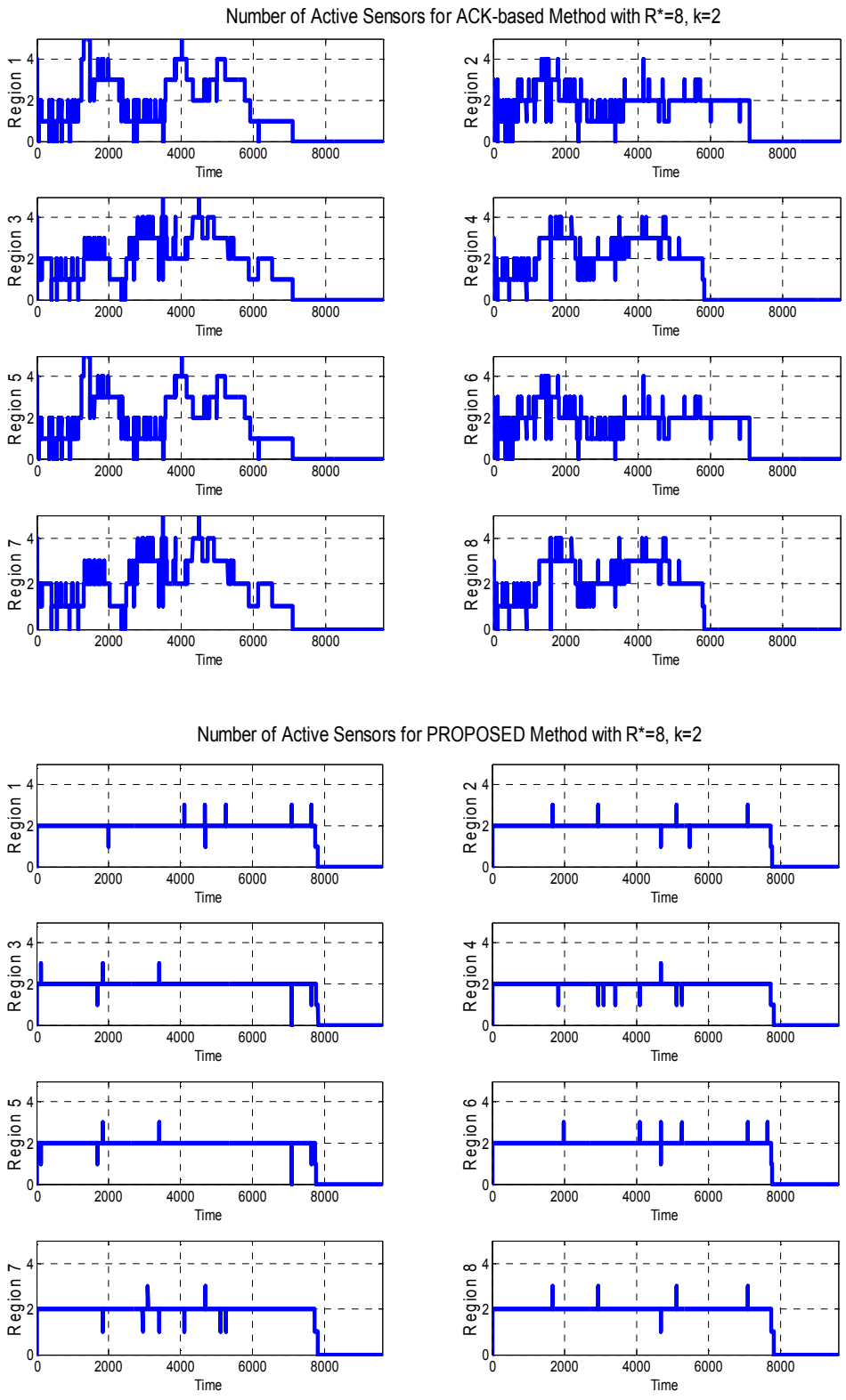


Figure 6.8: Performance of both methods for R*=8 and k=2

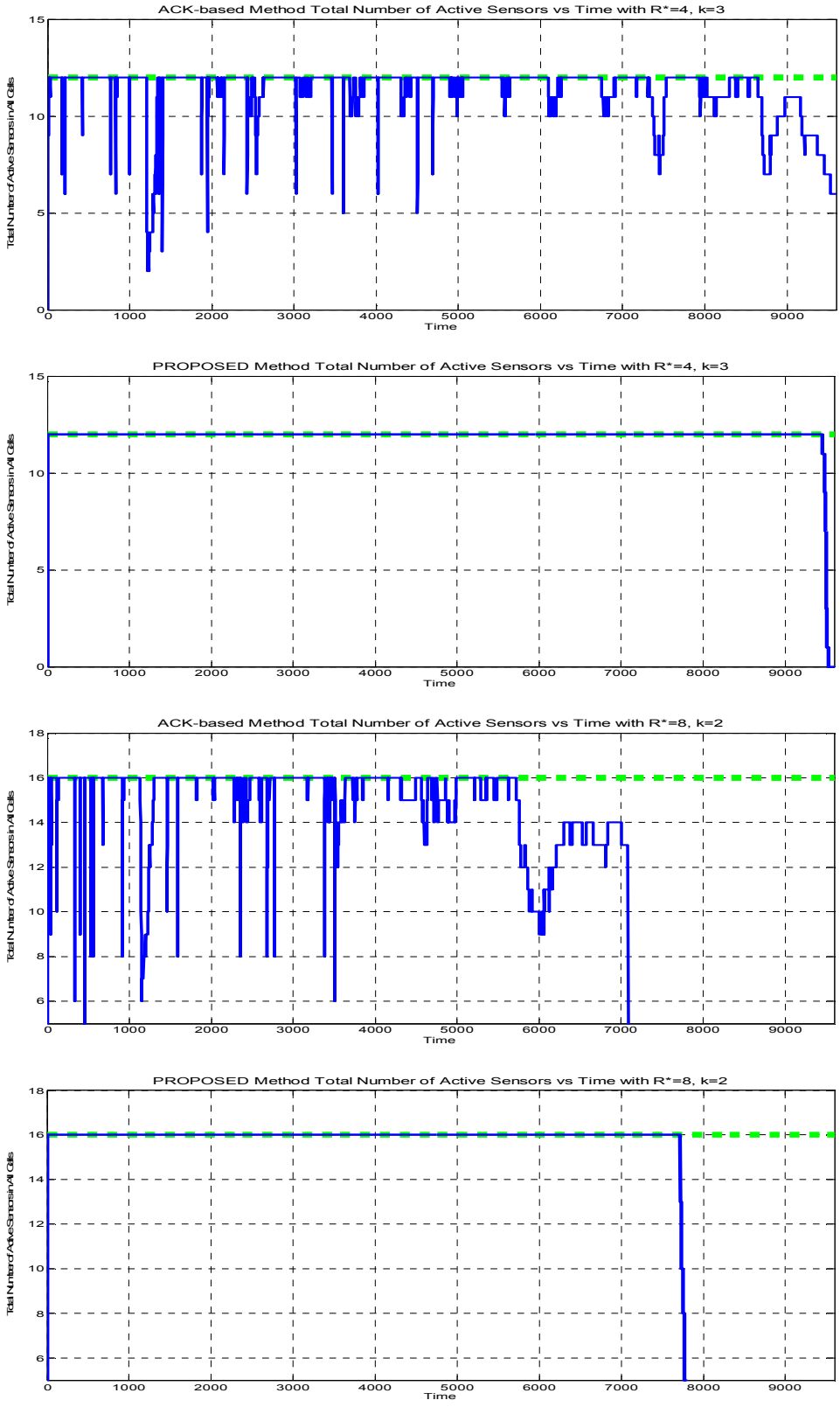


Figure 6.9: Overall Network Lifetime

attaining the required application QoS levels. Yet, both of those methods rely on the solution of fairly complicated mathematical problems in order to determine the best schedule maximizing the network lifetime. In fact, the method of Perillo and Heinzelman (2003a) requires the solution of a linear program and the method of Delicato et al. (2006) solves a knapsack problem at each round to determine an optimal transmission schedule. Therefore, both of those methods are totally centralized approaches and the proposed algorithms, particularly the one in Perillo and Heinzelman (2003a), must be run offline by a computationally strong platform before starting the network operation. However, when the concern is only service quality as in almost all of the previous studies Kay and Frolik (2004), Delicato et al. (2006) and Perillo and Heinzelman (2003a), the proposed method of this thesis does not require the solution of any mathematical problems during the operation and therefore can run online. The proposed strategy presented in this section is a simple semi-distributed method where each node indicates is intent to transmit but the cluster head makes the final decision centrally. Consequently, it provides a reasonable compromise between simple distributed techniques (Kay & Frolik, 2004; Iyer & Kleinrock, 2003) that yield poor performance in satisfying QoS requirements and more complicated centralized methods (Delicato et al., 2006; Perillo & Heinzelman, 2003a) that can achieve desired QoS levels by the use of computationally intensive algorithms that are difficult to implement in resource constrained sensor networks.

CHAPTER 7

CONCLUSIONS

This last chapter begins with an overview of the research studies performed during the thesis period. Afterwards, the contributions of the research are presented. Then, limitations of this study are given and finally, further work that can be conducted in the future by other researchers to overcome those limitations are recommended.

7.1 SUMMARY OF WORK DONE

This thesis has investigated the subject of security and quality of service for wireless sensor networks. The joint provisioning of those two concepts has become very important for envisioned WSN applications of near future which require security and service quality at the same time. The constrained resources of sensor networks and the interactions and tradeoffs between security and QoS, however, complicate the problem.

From the extensive literature survey presented in Chapter 2, it can be observed that the existing studies do not offer a comprehensive solution on how to provide security and application level service quality to wireless sensor networks. This Ph.D. research study first analyzed the correlation between those two concepts and then proposed two different methods to control the security and service

quality levels of a cluster-based wireless sensor network to attain the required values where possible or otherwise the optimal security and QoS levels yielding the best tradeoff under the limited channel capacity.

The proposed optimization method to determine this best tradeoff was implemented on a hardware platform called Gumstix which can emulate the cluster head of a sensor network due to its comparable processing capabilities and this implementation showed that the proposed algorithm could run in real sensor network settings without a problem. Then, the proposed security and QoS methods were simulated to test their performance. The corresponding simulation results presented in Chapter 5 and Chapter 6 for the proposed two methods, respectively, showed that both strategies were able to satisfy the security and QoS requirements of the sensor network under analysis with quite a fair performance. Those results also indicated that the method presented in Chapter 6, which was originally developed during this Ph.D. research study, had both comparably superior performance with respect to existing studies and also was simpler and less complicated concordant to the limitations existent in sensor networks.

7.2 RESEARCH CONTRIBUTION

The current Ph.D. thesis study provided a detailed consideration of the security and quality of service issues of wireless sensor networks. It presented both theoretical analysis and also provided results obtained from computer simulations and partial real-world implementations. Therefore, both the academic community and practitioners can benefit from the contributions of this research.

The concrete outcomes of this Ph.D. research study are (i) an extension to the scope of the existing studies on sensor network QoS by inclusion of security and additional service quality attributes such as coverage and packet loss due to collisions, (ii) an analysis for the interactions between security and QoS attributes of sensor networks such as correlation of security and spatial resolution, coverage

and spatial resolution and, security and system lifetime, (iii) an optimization method to determine the best tradeoff between security and spatial resolution, and finally (iv) a novel QoS and security control method superior in both performance and simplicity to existing strategies. Research contributions brought in by those outcomes are summarized in the following paragraphs.

The control methods proposed in the thesis has comparably a more comprehensive scope which allows the provisioning of two additional QoS attributes and also another important requirement, security. In particular, inclusion of coverage in addition to the QoS parameters already considered in existing studies such as spatial resolution and system lifetime provides a more realistic service quality definition helping the achievement of better QoS levels. In fact, measuring the application QoS level by counting only the number of active sensors without taking the geographical distribution of those sensors into account causes a very important sensor network component to be neglected, that is, coverage. Therefore, in this study, spatial resolution and coverage are considered together as well as two more service quality attributes, namely, packet loss due to collision and network lifetime.

The second point making the scope of this study more comprehensive than previous studies is that the proposed method here employs a control on security. To expand on, in addition to the four service quality attributes mentioned above, it is possible to control the security level of packets sent from sensor nodes to the cluster head in a cluster-based sensor network setting. Furthermore, in this thesis study, the interactions between security and one of the QoS attributes, i.e., spatial resolution were analyzed and it was shown that there existed a tradeoff between the attainable security and resolution levels under the limited capacity of the communication channel. A heuristic algorithm was also developed for solving the optimization problem to determine the best tradeoff between security and spatial resolution.

Thus, the fundamental contribution of this thesis study to the wireless sensor network body of knowledge is a comprehensive assessment for joint achievement of security and quality of service for sensor networks. Simulation results showed that the proposed security and control method developed in this thesis could determine the optimal supported QoS and security levels of a cluster-based sensor network pertaining to the limits of the channel capacity and also could keep the network at those QoS and security levels during its entire lifetime. Therefore, together with the future studies which will improve the findings of this research, envisioned sensor network applications that simultaneously require certain levels of security and service quality can be realized in near future.

Another considerable contribution of this thesis is the applicability of the proposed QoS and security control strategy to satisfy the requirements of sensor network setting which need application level service quality but no security. In fact, due to its consideration of the interactions between QoS attributes and also owing to the attempt to refrain from the deficiencies of previous studies, the proposed application level QoS control strategy provides better performance compared to several previous studies. Through simulations, it was shown that the proposed method achieves superior performance in attaining all four of the considered QoS attributes by providing better coverage and spatial resolution, longer system lifetime and less packet collisions. In addition, when compared to some other similar studies in sensor network QoS field, the proposed sensor network QoS control method of the thesis is less complicated and simpler. As opposed to those other studies, QoS control steps of this thesis' strategy does not involve the solution of any linear programs or optimization problems and therefore, it is more likely to be implemented in real sensor platforms which have constrained computational and communication resources. In fact, due to its semi-distributed nature where some part of the algorithm is run by sensor nodes and the other part is run in the cluster head, the overall method can run online during the operation and therefore, presents a reasonable compromise between simple distributed techniques that yield poor performance in satisfying QoS requirements

and more complicated centralized methods that can achieve better performance by the use of computationally intensive algorithms but hard to implement in sensor networks.

7.3 LIMITATIONS AND FURTHER RESEARCH

This research study focused mainly on the application level quality of service issues, which is only one of the two broad categories of the QoS concept. As it was explained previously, there were two reasons for mostly excluding the other major QoS classification, that is, the network level service quality. The first reason was that both categories are rather broad including several attributes to be considered and accordingly, in order to perform a deep analysis as aimed by this thesis, focusing on only one of those QoS perspectives was more reasonable. The second reason was that the solution domains for these two QoS categories were different in the sense that provisioning network QoS requires methods that operate in the network layer for multi-hop topologies but the focus of this research was one-hop topologies and the link layer. Due to the limitation brought by these reasons, the proposed control method of this thesis can only be utilized for settings which require application level service quality. If a WSN application is in need of both application and network QoS, then additional methods providing network QoS must be employed together with the method proposed in the thesis.

A similar limitation exists because of the presumed security scope of the thesis which consists of integrity and authentication. The reason for exclusion of the third fundamental security principle, confidentiality, was that the security requirements of the sensor network applications addressed by this thesis were more biased towards message integrity and authenticity and also the confidentiality of the packets holding local data inside the cluster is less critical. Therefore, the method proposed in this research study does not provide confidentiality by itself but it can still be utilized for cases where confidentiality is a requirement if encryption is performed by other methods at higher layers of the

protocol stack or later when transferring the aggregated data of the cluster to the sink.

Another limitation of the proposed control method of this thesis study is that it is applicable only to a single cluster of a cluster-based sensor network where communication from the sensors to the cluster head occurs in one hop. In fact, the proposed solution can easily be extended to work in multiple clusters by applying the method independently to each cluster. However, it is not possible to utilize the suggested method in sensor network topologies where sensor nodes cannot send their packets to the cluster head in one hop and therefore multi-hop transfer is required.

The mentioned limitations of this study can be overcome by further research in the same area. In the short term, this research can be enriched with adoption of the proposed QoS and security control strategy within multi-hop networks and multiple clusters. This would increase the applicability of proposed method by rendering it possible to be utilized in many more sensor network applications and topologies.

Furthermore, a full real-world implementation of the proposed strategy with commercial sensor network hardware to control the QoS and security levels of a real sensor network cluster including tens of sensor nodes and a cluster head would be an important practical research study that could prove the usability of the findings of this research in sensor network applications planned to be realized in near future.

To conclude, this study was only a step in the ever growing field of wireless sensor network research and future studies will certainly substantiate and improve on the findings of this thesis.

REFERENCES

Abd-El-Barr, M. I., Al-Otaibi, M. M. & Youssef, M. A. (2005). Wireless Sensor Networks - Part II: Routing Protocols and Security Issues. *Proceedings of the 18th IEEE Annual Canadian Conference on Electrical and Computer Engineering* (pp. 69-72).

Abramson, N. (1970). The ALOHA System - Another alternative for computer communications. *Procedings of AFIPS Conference* (pp. 295-298).

Akkaya, K. & Younis, M. (2003). An Energy-Aware QoS Routing Protocol for Wireless Sensor Networks. *Proceedings of the 23rd IEEE Workshop on Mobile and Wireless Networks* (pp. 710-715).

Akkaya, K. & Younis, M. (2005). Energy and QoS aware routing in wireless sensor networks. *Journal of Cluster Computing on Ad Hoc Networks*, *8*(2-3)*,* 179-188.

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. & Çayırcı, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, *38*, 393-422.

Awduche, D.O., Malcolm, J., O'Dell, M. & McManus, J. (1999). Requirements for Traffic Engineering Over MPLS, RFC2702. Retrieved September 11, 2007, from http://www.ietf.org/rfc/rfc702.txt

Bhattacharya, P., Hinrichs, S., Nahrstedt K. & McHugh, J. (2000). Security and quality of service interactions. National Information Systems Security Conference Program, Program RD2. Retrieved April 11, 2006, from http://csrc.nist.gov/nissc/program/rd2.htm

Blair, G., Campbell A.T., Coulson G. & Hutchison, D. (1995). Quality of Service Management in Distributed Systems. In M. Sloman (Ed.), *Network and Distributed Systems Management.* United States: Addison Wesley.

Caccamo, M., Zhang, L. Y., Sha, L. & Buttazzo, G. (2002). An Implicit Prioritized Access Protocol for Wireless Sensor Network. *Proceedings of the 23rd IEEE Real-Time Systems Symposium* (pp. 39-45).

Cardei, M. & Wu, J. (2006). Energy-efficient coverage problems in wireless ad-hoc sensor networks. *Computer Communications, 29*(4), 413-420.

Carman, D. W., Kruus, P. S. & Matt, B. J. (2000). *Constraints and Approaches for Distributed Sensor Network Security (Final)* (Technical Report 00-010). Glenwood, MD: NAI Labs, The Security Research Division Network Associates, Inc.

Chen, D. & Varshney, P. K. (2004). QoS Support in Wireless Sensor Networks: A Survey. *Proceedings of International Conference on Wireless Networks 2004* (pp. 227-233).

Chigan, C., Ye, Y. & Li, L. (2005). Balancing Security Against Performance in Wireless Ad Hoc and Sensor Networks. *Proceedings of 2005 IEEE Vehicular Technology Conference* (pp. 4735-4739).

Crawley, E., Nair, R., Rajagopalan, B. & Sandick, H. (1998). A framework for QoS-Based Routing in the Internet, RFC 2386. Retrieved August 11, 2006, from http://www.ietf.org(rfc/rfc2386.txt

Delicato, F., Protti, F., Pirmez, L. & de Rezende, J. F. (2006). An efficient heuristic for selecting active nodes in wireless sensor networks. *Computer Networks*, *50*(18), 3701 – 3720.

Demetrios, Z. (2001). *A Glance at Quality of Services in Mobile Ad-Hoc Networks.* Retrieved April 15, 2006, from http://www.cs.ucr.edu/~csyiazti/courses/cs260/manetqos.pdf .

Deng, J., Han, R. & Mishra, S. (2003). A Performance Evaluation of Intrusion-Tolerant Routing in Wireless Sensor Networks. *Proceedings of 2nd IEEE International Workshop on Information Processing in Sensor Networks.*

Douceur, J.R. (2002). The Sybil attack. *Proceedings of the 1st International Workshop on Peer-to-Peer Systems* (pp.251-260).

Eschenauer, L. & Gligor, V. (2002). A Key Management Scheme for Distributed Sensor Networks. *Proceedings of the 9th ACM conference on Computer and communications security* (pp. 41-47).

Frigon, J.-F., Chan, H.C.B. & Leung, V.C.M. (2001). Dynamic reservation TDMA protocol for wireless ATM networks. *IEEE Journal on Selected Areas in Communications*, *19*(2), 370-383.

Ganz, A., Ganz, Z. & Wongthavarawat, K. (2004). *Multimedia Wireless Networks: Technologies, Standards and QoS*, Upple Saddle River, NJ: Prentice Hall.

Guimarães, G., Souto, E., Kelner, J. & Sadok, D. (2005). Evaluation of Security Mechanisms in Wireless Sensor Networks. *Proceedings of International Conference on Sensor Networks* (pp. 428-433).

Gumstix (n.d.). Retrieved December 12, 2007, from http://www.gumstix.org

Gurses, E. & Akan, O.B. (2005). Multimedia Communication in Wireless Sensor Networks. *Annals of Telecommunications, 60*(7-8), 799-827.

Hartung, C., Balasalle, J. & Han, R. (2004). *Node compromise in sensor networks: The need for secure systems* (Technical Report CU-CS-988-04). Colorado, US: University of Colorado at Boulder, Department of Computer Science.

He, T., Stankovic, J.A., Lu, C. & Abdelzaher, T. (2003). SPEED: A Stateless Protocol for Real-time Communication in Sensor Networks. *Proceedings of the 2003 International conference on Distributed Computing Systems.*

Heinzelman, W.R., Chandrakasan, A. & Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences* (pp. 10-16).

Hu, Y.-C., Perrig, A. & Johnson, D.B. (2003). Packet leashes: a defense against wormhole attacks in wireless networks. *Proceedings of the IEEE Infocom, 2003* (pp. 1976-1986).

Huang, Q., Cukier, J., Kobayashi, H., Liu, B. & Zhang, J. (2003). Fast authenticated key establishment protocols for self-organizing sensor networks. *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications* (pp. 141–150).

Hwang, J. & Kim, Y. (2004). Revisiting random key pre-distribution schemes for wireless sensor networks. *Proceedings of the 2nd ACM workshop on Security of Ad hoc and Sensor Networks* (pp. 43–52).

IEEE 802.11 Working Group (2007). *IEEE 802.11-2007: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*.

IEEE 802.15 Working Group (2003). *IEEE 802.15.4-2003: Wireless medium access control and physical layer specifications for low-rate wireless personal area networks*.

Ilyas, M. & Mahgoub, I. (2005). *Hand Book of Sensor Networks: Compact Wireless and Wired Sensing Systems*, Boca Raton, FL: CRC Press LLC.

Iyer, R. & Kleinrock, L. (2003). QoS Control for Sensor Networks. *Proceedings of the 2003 IEEE International Conference on Communications* (pp. 517-521).

Jolly, G. & Younis, M. (2003). Energy Efficient Arbitration of Medium Access in Sensor Networks. *Proceedings of the 2003 IASTED Conference on Wireless and Optical Communications.*

Jovanov, E., Milenkovic, A., Otto, C. & Groen, P.C. (2005). A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, *2*(6).

Kaps, J. P. & Sunar, B. (2006). Energy Comparison of AES and SHA-1 for Ubiquitous Computing. *Proceedings of the Workshop on Emerging Directions in Embedded and Ubiquitous Computing* (pp. 372-381).

Karlof, C. & Wagner, D. (2003). Secure routing in wireless sensor networks: Attacks and countermeasures. *Proceedings of First IEEE International Workshop on Sensor Network Protocols and Applications* (pp. 113-127).

Karlof, C., Sastry, N. & Wagner, D. (2004). TinySec: A Link Layer Security Architecture for Wireless Sensor Networks. *Proceedings of the Second ACM Conference on Embedded Networked Sensor Systems* (pp. 162-175).

Kay, J. & Frolik, J. (2004). Quality of Service Analysis and Control for Wireless Sensor Networks. *Proceedings of 1st International Conference on Mobile Ad-Hoc and Sensor Systems* (pp. 359-368).

Kuorilehto, M., Hännikäinen, M. & Hämäläinen, T.D. (2005). A Survey of Application Distribution in Wireless Sensor Networks. *EURASIP Journal on Wireless Communications and Networking*, *2005*(5), 774-788.

Law, Y. W. & Havinga, P. J. M. (2005). How to Secure a Wireless Sensor Network. *Proceedings of the 2005 Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) Conference* (pp. 89-95).

Law, Y.W., Dulman, S., Etalle, S. & Havinga, P. (2003). Assessing Security-Critical Energy-Efficient Sensor Networks. *Proceedings of 18th IFIP International Information Security Conference* (pp. 459-463).

Lee, C., Lehoczky, J., Rajkumar, R. & Siewiorek, D. (1999). On Quality of Service Optimization with Discrete QoS Options. *Proceedings of the Fifth IEEE Real-Time Technology and Applications Symposium* (pp. 276-286).

Lee, W.C., Hluchyj, M.G. & Humblet, P.A. (1995). Routing Subject to Quality of Service Constraints Integrated Communication Networks. *IEEE Network*, *9*(4), 46-55.

List of Battery Sizes (n.d.). Retrieved October 2, 2006, from http://en.wikipedia.org/wiki/List_of_battery_sizes

Liu, D. & Ning, P. (2003). Efficient distribution of key chain commitments for broadcast authentication in distributed sensor networks. *Proceedings of the 10th Annual Network and Distributed System Security Symposium* (pp. 263–276).

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

Lu, C., Blum, B. M., Abdelzaher, T. F., Stankovic, J. A. & He, T. (2002). RAP: A Real-Time Communication Architecture for Large-Scale Wireless Sensor Networks. *Proceedings of the Eighth IEEE Real-Time and Embedded Technology and Applications Symposium* (pp. 55-66).

Ma, Q. & Steenkiste, P. (1997). Quality-of-Service routing with Performance Guarantees. *Proceedings of the 4th IFIP Workshop on Quality of Service*.

Mahapatra, A., Anand, K. & Agrawal, D.P. (2006). QoS and Energy Aware Routing for Real Time Traffic in Wireless Sensor Networks. *Computer Communications, 29*(4), 437-445.

Mainwaring, A., Polastre, J., Szewczyk, R., Culler, D. & Anderson, J. (2002). Wireless sensor networks for habitat monitoring. *Proceedings of the First*

ACM International Workshop on Wireless Sensor Networks and Applications (pp. 88-97).

MATLAB, The Language of Technical Computing (n.d.). Retrieved October 2, 2006, from http://www.mathworks.com/products/matlab/

Meguerdichian, S., Koushanfar, F., Potkonjak, M. & Srivastava, M.B. (2001). Coverage Problems in Wireless Ad-hoc Sensor Networks. *Proceedings of IEEE Infocom 2001* (pp. 1380-1387).

Meguerdichian, S., Koushanfar, F., Qu, G. & Potkonjak, M. (2001). Exposure in Wireless Ad-hoc Sensor Networks. *Proceedings of the 7th annual international conference on Mobile computing and networking* (pp. 139-150).

MICA2 AA Battery Pack Service Life Test (n.d.). Retrieved October 2, 2006, from http://xbow.com/support-pdf-files/MICA2-BatteryLifeTest.pdf

MICA2 Specifications (n.d.). Retrieved September 23, 2006, from http://www.xbow.com

Miras, D. (2002). *A Survey of Network QoS Needs of Advanced Internet Applications* (Working Paper). London: University College, Internet2 QoS Working Group.

Ni ,Q., Romdhani, L. & Turletti, T. (2004). Survey of QoS enhancements for IEEE 802.11 wireless LAN. *Wiley Wireless Communications and Mobile Computing, 4*(5)*,* 547-566.

Patcher, A. (2006). *QoS in Wireless Data Networks*. Retrieved August 23, 2007, from http://www.cse.wustl.edu/~jain/cse574-06/ftp/wireless_qos/index.html

Pattem, S., Poduri, S. & Krishnamachari, B. (2003). Energy-Quality Tradeoffs for Target Tracking in Wireless Sensor Networks. *Proceedings of Information Processing in Sensor Networks.*

Perillo, M. & Heinzelman, W. (2003). Providing Application QoS Through Intelligent Sensor Management. *Proceedings of the 1st IEEE International Workshop on Sensor Network Protocols and Applications* (pp.93-101).

Perillo, M.A. & Heinzelman, W.B. (2003a). Optimal sensor management under energy and reliability constraints. *Proceedings of the IEEE Wireless Communication and Networking Conference* (pp.1621-1626).

Perrig, A., Szewczyk, R., Wen, V., Culler, D. & Tygar, J.D. (2002). SPINS: Security Protocols for Sensor Networks. *Wireless Networks Journal, 8*(5), 521-534.

Ren, X. (2006). Security Methods for Wireless Sensor Networks. *Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation* (pp. 1925-1930).

Sabins, F. (1997). *Remote Sensing, Principles and Interpretation*. New York: W.H. Freeman and Company.

Sakarindr, P., Ansari, N., Rojas-Cessa, R. & Papavassiliou, S. (2005). Security-enhanced Quality of Service (SQoS): Design and Architecture. *Proceedings of IEEE 2005 Sarnoff Symposium on Advances in Wired and Wireless Communications* (pp. 129-132).

Shah, R. & Rabaey, J. (2002). Energy Aware Routing for Low Energy Ad Hoc Sensor Networks. *Proceedings of IEEE Wireless Communications and Networking Conference 2002* (pp. 350-355).

Slijepcevic, S., Potkonjak, M., Tsiatsis, V., Zimbeck, S. & Srivastava, M. B. (2002). On Communication Security in Wireless Ad-Hoc Sensor Networks. *Proceedings of 11th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises* (pp.139-144).

Sohrabi, K., Gao, J., Ailawadhi, V. & Pottie, G. (2000). Protocols for Self-Organization of a Wireless Sensor Network. *IEEE Personal Communications*, *7*(5)*,* 16-27.

Tang, S. & Li, W. (2006). QoS Supporting and Optimal Energy Allocation for a Cluster-based Wireless Sensor Network. *Computer Communications*, *29*(13-14)*,* 2569-2577.

Tavli, B. & Heinzelman, W. (2003). TRACE: Time Reservation Using Adaptive Control For Energy Efficiency. *IEEE Journal on Selected Areas in Communications*, *21*(10), 1506-1515.

The Network Simulator, NS2 (n.d.). Retrieved October 2, 2006, from http://www.isi.edu/nsnam/ns/

The OPNET Modeler (n.d.). Retrieved October 2, 2006, from http://www.opnet.com/solutions/network_rd/modeler.html

Tomur, E. & Erten, Y.M. (2006). Application of temporal and spatial role based access control in 802.11 wireless networks. *Computers & Security*, *25*(6), 452-458.

Tran, S.P. & Yang, T. A. (2006). Evaluations of target tracking in wireless sensor networks. *Proceedings of the 37th SIGCSE Technical Symposium on Computer Science Education* (pp. 97-101).

Trevis, L. & El-Sheimy, N. (2004). The Development of a Real-time Forest Fire Monitoring and Management System. *Proceedings of the 20th Congress of International Society for Photogrammetry and Remote Sensing* (pp. 65-71).

US Congress. (1995). *Health Insurance Portability and Accountability Act*. Washington, DC: US Government Printing Office.

Walters J.P., Liang, Z., Shi, W. & Chaudhary, V. (2006). Wireless sensor network security: a survey. In Y. Xiao (Ed.), *Security in distributed, grid, and pervasive computing* (pp. 367-411). United States, FL: Auerbach Publications.

Wang, Y., Liu, X. & Yin, J. (2006). Requirements of Quality of Service in Wireless Sensor Networks. *Proceedings of the 2006 International Conference on Networking, Systems, Mobile Communications and Learning Technologies* (pp. 116-121).

Wang, Z. & Crowcraft, J. (1996). QoS-based Routing for Supporting Resource Reservation. *IEEE Journal on Selected Area of Communications*, *1996*(14), 1228-1234.

Wood, A. D. & Stankovic, J. A. (2002). Denial of service in sensor networks. *Computer, 35*(10), 54–62.

Wu, K. & Harms, J. (2001). QoS Support in Mobile Ad Hoc Networks. *Crossing Boundaries – an interdisciplinary Journal*, *1*(1), 92-107.

Xiao, Y., Chen, H., Sun, B., Wang, R. & Sethi, S. (2006). MAC Security and Overhead Analysis in the IEEE 802.15.4 Wireless Sensor Networks. *EURASIP Journal on Wireless Communications and Networking, 2006*(2), 81-93.

Younis, M., Akkaya, K., Eltoweissy, M. & Wadaa, A. (2004). On Handling QoS Traffic in Wireless Sensor Networks. *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences* (pp. 10-16).

Younis, M., Youssef, M. & Arisha, K. (2002). Energy-Aware Routing in Cluster-Based Sensor Networks. *Proceedings of the 10th IEEE/ACM Sym. on Modeling, Analysis and Simulation of Computer and Telecom. Systems* (pp. 129-136).

Zhang, H. & Gburzynski, P. (2002). A Variable Slot Length TDMA Protocol for Personal Communication Systems. *Wireless Personal Communications: An International Journal*, *22*(3)*,* 409-432.

Zhang, L., Deering, S., Estrin, D., Shenker, S. & Zappala, D. (1993). RSVP: A New Resource ReServation Protocol. *IEEE Network*, *7*(5)*,* 8-18.

Zhou, J. & Mu, C. (2006). A Kind of Application-Specific QoS Control in Wireless Sensor Networks. *Proceedings of IEEE International Conference on Information Acquisition* (pp. 456-461).

Zhu, S., Setia, S. & Jajodia, S. (2003). Leap: efficient security mechanisms for large-scale distributed sensor networks. *Proceedings of the 10th ACM conference on Computer and communications security* (pp. 62–72).

APPENDICES

APPENDIX A. Coverage and K-coverage Concepts

An important problem addressed in literature is the sensor coverage problem. This problem is centered around a fundamental question: ``How well do the sensors observe the physical space ?" (Cardei & Wu, 2006). The coverage concept is a measure of the quality of service (QoS) of the sensing function and is subject to a wide range of interpretations due to a large variety of sensors and applications. The goal is to have each location in the physical space of interest within the sensing range of at least one sensor.

Coverage problem is usually classified as area coverage, point coverage and barrier coverage as show in Figure A1 below, redrawn from Cardei and Wu (2006). The most studied coverage problem is the area coverage problem, where the main objective of the sensor network is to cover (monitor) an area (also referred sometimes as region). In the point coverage problem, the objective is to cover a set of points. The barrier coverage is defined as the coverage with the goal of minimizing the probability of undetected penetration through the barrier (sensor network).

In the literature, this problem has been formulated in various ways. For example, the Art Gallery Problem is to determine the number of observers necessary to

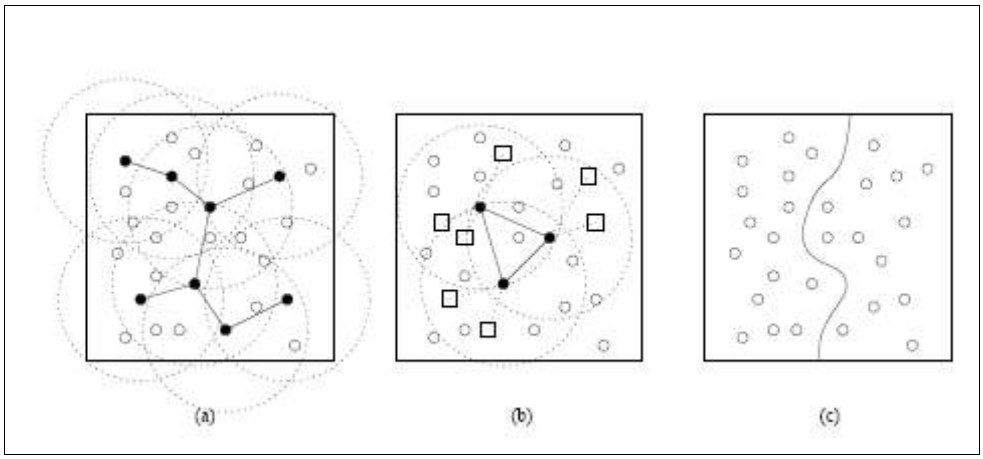


Figure A1: Area coverage (a), Point coverage (b) and Barrier coverage (c) (Cardei & Wu, 2006)

cover an art gallery (i.e., the service area of the sensor network) such that every point in the art gallery is monitored by at least one observer. This problem can be solved optimally in a 2D plane, but is shown to be NP-hard when extended to a 3D space. Some other studies define a sensor coverage metric called surveillance that can be used as a measurement of quality of service provided by a particular sensor network, and centralized optimum algorithms that take polynomial time are proposed to evaluate paths that are best and least monitored in the sensor network. Another work further investigates the problem of how well a target can be monitored over a time period while it moves along an arbitrary path with an arbitrary velocity in a sensor network. Localized exposure-based coverage and location discovery algorithms are proposed in several papers.

On the other hand, some works are targeted at particular applications, but the central idea is still related to the coverage issue. For example, sensors' on-duty time should be properly scheduled to conserve energy. Since sensors are arbitrarily distributed, if some nodes share the common sensing region and task, then one can turn off some of them to conserve energy and thus extend the lifetime of the network. This is feasible if turning off some nodes still provide the same "coverage" (i.e., the provided coverage is not affected). Some papers

propose a heuristic to select mutually exclusive sets of sensor nodes such that each set of sensors can provide a complete coverage of the monitored area. Also targeted at turning off some redundant nodes, another paper proposes a probe-based density control algorithm to put some nodes in a sensor-dense area to a doze mode to ensure a long-lived, robust sensing coverage. A coverage preserving node scheduling scheme is presented by some authors to determine when a node can be turned off and when it should be rescheduled to become active again.

Recently, a more general sensor coverage problem is started to be considered. Given a set of sensors deployed in a target area, one wants to determine if the area is sufficiently k-covered, in the sense that every point in the target area is covered by at least k sensors, where k is a predefined constant. As a result, the aforementioned previous problem can be regarded as a special case of this problem with $k = 1$. Applications requiring $k > 1$ may occur in situations where the stronger environmental monitoring is necessary, such as military applications. It also happens when multiple sensors are required to detect an event. For example, the triangulation-based positioning protocols require at least three sensors (i.e., $k \geq 3$) at any moment to monitor a moving object. Enforcing $k \geq 2$ is also necessary for fault-tolerant purpose.

APPENDIX B. Multiple Access Control (MAC) Schemes for Wireless Sensor Networks

In the literature, either contention based distributed random access MAC schemes such as ALOHA or CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) or TDMA/FDMA-like centralized methods are proposed for wireless sensor networks. Since in this thesis some fairness is sought in the average number of transmissions of sensor nodes for power saving purposes and also since a centralized entity exists (cluster head), it was preferred to employ a centralized MAC scheme to which the setting fits better[8]. But, since only a portion (not all) of the sensors will be sending information at the same time, use of the fixed assignment-based methods such as pure TDMA of FDMA may cause channel inefficiencies due to the empty slots or frequency channels unassigned.

Therefore, the best choice for a MAC scheme suiting the requirements of this thesis would be a demand-based one in which a central controller assign channels to the clients who inform the controller about their access requests. Since a central authority arbitrates channel access requests, a fair distribution of channel capacity without any wastage can be achieved by demand assignment MAC schemes.

The most straightforward form of demand assignment MAC methods is *polling* where the controller asks client devices one by one whether they need to access

[8] In fact, there are studies recommending use of contention-based methods such as ALOHA or CSMA for similar settings aiming QoS control. Such studies justify their choice by stating that slotted systems use communication channel inefficiently under low load. However, it is proposed here that the hybrid MAC schemes combining contention-based and slotted methods are best suited for QoS control mechanisms which aim to be power efficient. The first reason for this is that such hybrid methods have better throughput under heavy load. They are also power-efficient since nodes can shut down their receivers in unassigned data slots. What is more, due to their centralized nature, they are useful for controlling purposes such as deciding which sensors will become active. The last but not least advantage is they allow to inform nodes about the current security level before they transmit their data. Those were the reasons for employing a reservation-based MAC scheme in this thesis.

channel. The controller then assigns channel access to the requesting devices. Best example of polling based MAC for wireless networks is IEEE 802.15.1 Bluetooth. However, polling technique is not suitable for networks such as the one in this thesis where there are a large number of clients and these clients infrequently need channel access. Specifically in this study, it is aimed to distribute the active sensors in time to save power. Thus, polling inactive sensors very frequently not only will increase network latency but also cause energy drain due to wake-ups of inactive sensors.

A reservation-based scheme among demand-assignment type MAC methods is more appropriate for this study because in such schemes, clients desiring to transmit should reserve a channel beforehand. In other words, during a determined time frame called reservation period, clients inform the controller about their transmission request and only then clients can send data according to the channel access allocation schedule determined by the controller.

In reservation-based MAC schemes, time is divided into frames which are composed of two main parts: reservation period and data transmission period. Data transmission period is mostly based on a TDMA-like structure where clients who are assigned channels transmit in their allocated time-slots. Reservation period can also be based on a TDMA-type structure in which each client in the network is assigned a mini slot in the reservation period and if a client has a data to send, it transmits a specific codeword in its assigned mini slot to indicate its need. But, such kind of a reservation period is not well suited to the setting of this thesis because the number of sensors deployed in an area is non-constant due to deaths or re-deployment of sensor nodes. Therefore, a reservation-based MAC scheme for the assumed network topology here had better a contention-based reservation period. In other words, stations do not have fixed mini slots in the reservation period to announce their transmission need but contend for a free mini slot in the reservation period. Only the stations that find a mini slot not accessed by another station can transmit in the data transmission period. These type of

MAC schemes are sometimes called Hybrid MAC protocols which combine contention-based and centralized methods.

There are several studies in the literature proposing such hybrid MAC protocols such as DR-TDMA (Frigon et al., 2001), TRACE (Tavli & Heinzelman, 2003) and variable slot-length TDMA (Zhang & Gburzynski, 2002). Though some of these protocols are designed for wireless ATM, they can be used in non-ATM wireless networks as stated in Ilyas and Mahgoub (2005). Summary of some of these methods and the details of TRACE are given below.

Dynamic Reservation Time Division Multiple Access (DR-TDMA) (Frigon et al., 2001)*:* The fixed length DR-TDMA MAC frame is time-duplexed into an uplink and downlink channel and the boundary between these two parts is dynamically adjusted as a function of the traffic load. Downlink and uplink channels are further dynamically divided into control and data transmission periods. Slots assigned for control purpose are divided into control mini-slots used to transmit the control packets. The base station has absolute control in determining the number of slots in each frame period and which mobile will receive or send information during the data slots. The modem preamble is used by radio physical layer functions while the frame header announces the frame periods boundaries

Variable slot-length TDMA (Zhang & Gburzynski, 2002): This is another reservation-based TDMA scheme designed for wireless networks. Reservation period is contention-based as in TRACE. The difference of Zhang and Gburzynski (2002) lies in its data transmission part of the time frame. Data transmission slots are not equal length and data transmission period of each frame is dynamically partitioned into variable length transmission slots according to the current bandwidth needs of stations. In this way, granularity of the bandwidth assignment is not constrained by one or multiples of slot size. Another advantage of such a scheme over other MAC strategies which assign multiple fixed length slots to

stations requiring more bandwidth is that no wastage is done for multiple data slot boundaries.

The MAC scheme that is assumed in this study is a combination of Tavli and Heinzelman (2003) and Zhang and Gburzynski (2002) in the sense that a very similar frame structure is used to the one in TRACE but the assumed MAC scheme here utilizes variable length data slot sizes as in Zhang and Gburzynski (2002). This variable-duration data slot approach increases the granularity of the MAC scheme because only one data slot is assigned for each sensor node which is just enough to transmit one packet. If a fixed-duration data slot approach were used, then multiple slots would have to be assigned to the sensors with packets which are longer due to the increased security overhead. This would cause channel inefficiency because of worse granularity. For instance, if the slot duration were fixed to 0-level security packet length PL_0, then for sensor nodes transmitting at security level S, it would be required to assign $\llcorner PL_S/PL_0 \lrcorner$ (nearest integer to PL_S/PL_0) data slots. If PL_S/PL_0 is non-integer, the slot would go under-utilized.

To solve this granularity problem, one might recommend the use of unity length slots (1 bit or byte duration) and $PL_0 \times PL_1 \times PL_2 \times \ldots \times PL_{smax}$ number of slots in each frame. In this case, PL_S slots will be assigned to a sensor with security level S and $(PL_0 \times \ldots \times PL_{smax}) / PL_S$ number of active sensors would transmit in a frame. But, this kind of a MAC scheme is not feasible in real life because it requires too short slot duration but too long frame duration. Also, it cannot solve the problem of more slot boundary overhead problem. As a conclusion, variable length data slot approach that is taken in the MAC strategy of the thesis is the most suitable one for the assumed topology.

Details of the TRACE (Tavli & Heinzelman, 2003) Protocol

A. Overview

TRACE (Time Reservation Using Adaptive Control for Energy Efficiency) is an energy-efficient dynamic time-division multiple-access (TDMA) protocol designed for real-time data communications. In TRACE, data transmission takes place according to a dynamically updated transmission schedule. A controller in the network is responsible for creating the TDMA schedule based on which nodes have successfully contended for data slots in the current frame. The controller transmits this schedule to the rest of the nodes in the network at the beginning of the data subframe. Whenever the energy of the controller drops below the energy level of the other nodes in the network by more than a set amount, it assigns another radio with higher energy than itself as the next controller. Controller handover takes place during the TDMA schedule transmission by specifying the ID of the new controller. Finally, if the number of transmissions in a frame exceeds a predetermined threshold, each node listens only to data from certain nodes. Each node determines which transmitters to listen to based on information obtained from all the nodes during the information summarization (IS) slot. The following sections describe these ideas in more detail.

B. Basic Operation

TRACE is organized around time frames with duration matched to the periodic rate of packets. The frame format is presented in Figure B1. Each frame consists of two subframes: a control subframe and a data subframe. The control subframe consists of a beacon message, a contention slot, a header message, and an IS slot. At the beginning of every frame, the controller node transmits a beacon message. This is used to synchronize all the nodes and to signal the start of a new frame. The contention slot, which immediately follows the beacon message, consists of *Nc* subslots. Upon hearing the beacon nodes that have data to send but did not reserve data slots in the previous frame, randomly choose subslots to transmit their requests. If the contention is successful (i.e., no collisions), the controller

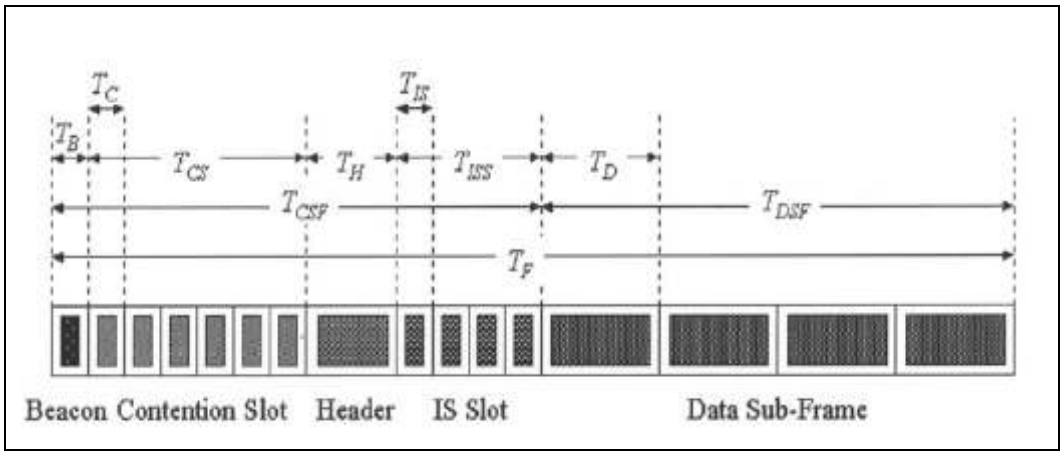


Figure B1**:** Frame format of TRACE protocol (Tavli & Heinzelman, 2003)

grants a data slot to the contending node. The controller then sends the header, which includes the data transmission schedule of the current frame.

The transmission schedule is a list of nodes that have been granted data slots in the current frame along with their data slot numbers. A contending node that does not hear its ID in the schedule understands that its contention was unsuccessful (i.e., a collision occurred or all the data slots are already in use) and contends again in the following frame. If the waiting time for a packet during contention for channel access exceeds the threshold *Tdrop*, it is dropped. The header also includes the ID of the controller for the next frame, which is determined by the current controller according to the node energy levels.

The IS slot begins just after the header slot and consists of *Nd* subslots. Nodes that are scheduled to transmit in the data subframe transmit a short IS message exactly in the same order as specified by the data transmission schedule. An IS message includes the energy level of the transmitting node, enabling the controller node to monitor the energy level of the entire network, and an end-of-stream bit, which is set to one if the node has no data to send. Each receiving node records the received power level of the transmitting node and inserts this information into its IS table. The information in the IS table is used as a proximity metric for the

nodes (i.e., the higher the received power the shorter the distance between transmitter and receiver nodes). Using the receive signal strength to estimate the relative distance of the transmitter to the receiver is a method employed in previous studies. If the number of transmissions in a particular frame is higher than a predetermined number of transmissions, *Nmax*, each node schedules itself to wake up for the top *Nmax* transmissions that are the closest transmitters to the node. Hence, the network is softly partitioned into many virtual clusters based on the receivers; this is fundamentally different from transmitter based network partitioning.

The data subframe is broken into constant length data slots. Nodes listed in the schedule in the header transmit their data packets at their reserved data slots. Each node listens to at most *Nmax* data transmissions in a single frame, therefore, each node is on for at most *Nmax* data slots. All nodes are in the sleep mode after the last reserved data slot until the beginning of the next frame.

If the power level of the controller node is lower than any other node by a predetermined threshold, then in the next frame controller handover takes place. The controller node assigns another node (any other node in the network with energy level higher than that of the controller) as the controller, effective with the reception of the header packet. Upon receiving the header packet, the node assigned to be the controller assumes the controller duties. A node keeps a data slot once it is scheduled for transmission as long as it has data to send. A node that sets its end-of-stream bit to one because it has no more data to send will not be granted channel access in the next frame (i.e., it should contend to get a data slot once it has new data to send).

C. Initial Startup

At the initial startup stage, a node listens to the medium to detect any ongoing transmissions for one frame time , because it is possible that there might already be an operational network. If no transmission is detected, then the node picks a

random time, smaller than the contention slot duration, at which to transmit its own beacon signal, and the node listens to the channel until its contention timer expires. If a beacon is heard in this period, then the node stops its timer and starts normal operation. Otherwise, when the timer expires, the node sends a beacon and assumes the controller position. In case there is a beacon collision, none of the colliding nodes will know it, but the other nodes hear the collision, so the initial setup continues. All the previously collided nodes, and the nodes that could not detect the collision(s) because of capture, will learn of the collisions with the first successful beacon transmission.

D. Prioritization

TRACE supports an optional prioritized operation mode. In this mode, the nodes have three preassigned priority levels, of which priority level-1 (PL1) is the highest priority and PL3 is the lowest priority. The highest level has the highest quality-of-service (QoS) and the lowest level has the lowest QoS. Prioritization is incorporated into the basic protocol operation at three points: contention, scheduling, and receiver based soft clustering.

The number of contention slots per node is higher for the higher priority levels, which results in less contention for higher priority nodes. In scheduling, PL1 and PL2 nodes are always given channel access, even if all the data slots are reserved. If all the data slots are reserved, then reservations of PL3 nodes are canceled starting from the latest reservation and granted to the higher priority nodes. All the nodes should listen to data from PL1 nodes, whether or not they are close to the nodes. Prioritization does not affect the general protocol operation, because it is assumed that the number of PL1 and PL2 nodes is much less than the number of PL3 nodes.

E. Receiver-Based Soft Cluster Creation

Each node creates its receiver-based listening cluster, which has a maximum of *Nmax* members, by choosing the closest nodes based on the proximity information

obtained from the received power from the transmissions in the IS slot. Priority has precedence over proximity; therefore, transmissions by PL1 nodes are always included in the listening cluster by removing the furthest node in the cluster.

F. Reliability

In case the controller node fails, the rest of the network should be able to compensate for this situation and should be able to continue normal operation as fast as possible. Failure of the controller manifests itself at two possible points within a frame: beacon transmission and header transmission. A backup controller, assigned by the controller, could listen for the beacon and header and become the controller whenever the controller fails. However, if both the backup controller and the controller die simultaneously, then the network is left dead. Instead of assigning a backup controller, there is a more natural and complete way of backing up the network: the transmission schedule is a perfect list of backup controllers in a hierarchical manner. The first node in the schedule is the first backup controller, the second node is the second controller, and the *N*th node is the *N*th backup controller.

The backup nodes listen to the beacon, which is a part of normal network operation. If the first backup controller does not hear the beacon for interframe space (IFS) time, then the controller is assumed dead and the first node transmits the beacon. If the beacon is not transmitted for two IFS time, then the second backup controller understands that both the controller and the first backup controller are dead, and transmits the beacon. The backup procedure works in the same way for all the nodes listed in the transmission schedule in the previous frame. If after IFS time no beacon is transmitted, then the rest of the nodes understand that the controller and all the backup nodes are dead, and they restart the network. Restartup is the same as the initial network startup, but in this case nodes do not listen for an existing controller for *Tf*; instead they start right away, because they know the controller is dead and there is no need for waiting.

The response of the network to the controller failure in header transmission is very similar to that of beacon failure. The succeeding backup node transmits the transmission schedule of the previous frame by updating it with the information in the IS slot of the previous frame denoting nodes with reservations that no longer have data to transmit. However, none of the nodes, including the backup nodes, listen to the contention slot, so the transmission schedule cannot be updated for the contending nodes. Since controller node failure is not a frequent event, it is better not to dissipate extra energy on controller backup. If all the backup nodes die simultaneously during header transmission, then the rest of the nodes begin restartup. Also, if there were no transmissions in the previous frame, then in case of a controller failure, nodes just enter restartup (i.e., there are no backup nodes).

APPENDIX C. QoS Optimization and Utility Functions

In Lee et al. (1999), authors present a framework for optimally allocating finite resources to satisfy QoS requirements of multiple applications along multiple QoS dimensions. As an example problem, they mention allocation of bandwidth among several QoS dimensions such as cryptographic security, packet loss, video picture color depth, audio sampling rate, etc. of various applications such as web, ftp, video conferencing, etc. Their proposed solution is based on the maximization of an aggregate system utility function which is composed of the utility/benefit brought by all QoS dimensions of all applications. Their problem formulation is as follows:

- There are n tasks/applications named as *T1, T2, T3, ..., Tn*
- QoS dimensions for task *Ti* is represented by a vector *Qi=(Q1, Q2, ..., Qn)*
- For each QoS dimension *j* of each task *i*, a utility function *Ui(Qij)* is defined which quantifies the benefit added by *Qij* to the system.
- There are m shared resources represented by *Ri*.
- Application utility function of task *i* is $Ui(Qi)=\sum_{j=1}^{di} w_{ij}U_i(Q_{ij})$ and overall system utility function is $U=\sum_{i=1}^{n} w_i U_i(Q_i)$ where w_i and w_{ij} are constant weights.
- Therefore, the optimization problem is as below:

$$\text{Maximize} \quad U(Q1, Q2, \ldots Qn) = \sum_{i=1}^{n} w_i U_i(Q_i)$$

$$\text{Subject to} \quad Qi \geq Qi,min \quad \text{for i=1,2,…..,n} \quad \text{(QoS constraint)}$$

$$\sum_{i=1}^{n} R_{ij} \leq R_j^{\max} \quad \text{for j=1,2,….,m} \quad \text{(Resource constraint)} \quad \textbf{(C1)}$$

$$R_i \Xi_i Qi \quad \text{for i=1,2,….,n} \quad \text{(Resource profiles)}$$

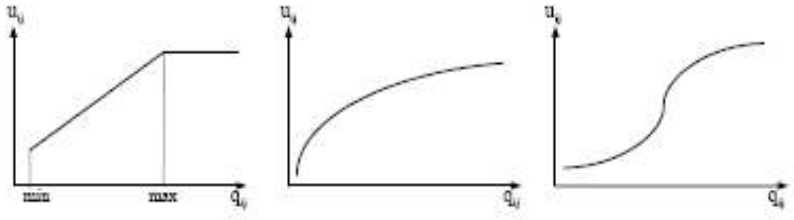


Figure C1: Some typical utility functions

Regarding the utility functions *Ui(Qij)* mentioned above, authors mention that some common properties associated with dimensional quality utility functions are observed including: non-decreasing, often quasi-continuous and piecewise concave. In Figure C1, some typical utility function shapes are depicted.

In economics, utility is a measure of the relative satisfaction gained by consuming different bundles of goods and services. And, the utility maximization problem is the problem consumers face: "how should I spend my money in order to maximize my utility?". Since utility is directly related to consumer preferences, it is convenient to represent preferences with a utility function and then deal with utility functions instead of preferences in utility maximization problem. For X representing the consumption set defined as the set of all mutually-exclusive packages that consumer could consume, the consumer's utility function $U: X \rightarrow R$ ranks each package in the consumption set. If $U(x) \geq U(y)$, then the consumer strictly prefers x to y or he/she is indifferent between them. Most utility functions used in modeling or theory are well-behaved in the sense that they usually exhibit monotonicity, convexity and quasi-continuity.

Another point needing to be clarified about utility functions is how they are determined. In fact, since they reflect consumer preferences, they should best be determined by consumers. Yet, even if it is assumed that an ordinary consumer is able to express his/her preferences by choosing a certain shaped utility function, it

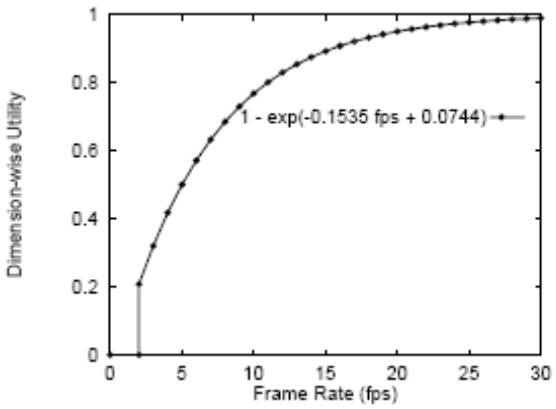


Figure C2: Utility function for a video application

would be overoptimistic to assume that he/she could also determine the parameters of this function fitting best to his/her preferences. At this point, *Satisfaction Knee Point* approach mentioned in Lee et al. (1999) may help to abstract the mathematical complexities regarding utility functions. In Lee et al. (1999), through the user interface of a computer program, authors propose to let the consumer graphically specify some points on the utility function graph which represent the level of satisfaction he/she will get. For instance, regarding the use of a video conference system, they ask users to specify the frame rates that will give them 50% and 95% satisfaction. Assuming an exponential utility curve of $U(f)=1-exp(a*f+b)$ where f represents frame rate, if a user chooses *(5 fps, 0.50)* and *(20 fps, 0.95)* points, the corresponding utility function $U(f)$ can completely be determined for all frame rates since one can compute a and b values as $a=-0.1535$ and $b=0.0744$.

Such a dimensional utility function $U(f)$ for a video application Ti's frame rate is given in Figure C2. As can be shown from the figure, for this utility function, a frame rate of 5 fps corresponds to a utility value of 0.50 while the utility increases to 0.95 for 20 fps. More details regarding utility functions can be found in Lee et al. (1999).

Returning back to the optimization problem formulation given in Formula C1, it is seen that this formulation can be utilized to set up the security-spatial resolution optimization problem of this study. The corresponding case here is much simpler since there is only one application (the network under consideration has a single operational purpose and only a single type of traffic is considered, so n=1), only one resource (the limited resource is channel capacity, so m=1) and there are only two QoS dimensions which are security and spatial resolution. So, the general problem reduces to the one below.

$$\text{Maximize} \quad U(Q1) = \sum_{j=1}^{d1} w_{1j} U_{1j}(Q_{1j}) = w_{11} U_{11}(Q_{11}) + w_{12} U_{12}(Q_{12})$$

$$\text{Subject to} \quad Q1 \geq Q1,min \quad \equiv \quad Q_{11} \geq Q_{11},\min \text{ and } Q_{12} \geq Q_{12},\min$$

$$R_{11} \leq R_1^{\max}$$

$$R_1 \Xi_1 Q1$$

In the setting of this thesis, $w_{11} = w_N$ (spatial resolution weighting constant), Q_{11} = N (spatial resolution), $Q_{11},\min = Nmin$, $U_{11} = U_N$ (spatial resolution utility function) and similarly $w_{12} = w_S$ (security weighting constant), Q_{12} = S (security), $U_{12} = U_S$ (security utility function), $Q_{12},\min$ = $Smin$, R_{11} is the used channel capacity, i.e., $R_{11} = Ds \times N$ and $R_1^{\max}$ is the total channel capacity, i.e., $DSF = D_0 \times N_{0,max}$. Also, resource profile constraint is not applicable because there cannot be more than one resource allocation scheme to achieve the same quality point due to the fact that there is only one application and one resource in the current case. Consequently, the final version of the optimization problem to determine the best tradeoff between security and spatial resolution, also given in Chapter 4, is below.

$$\text{Maximize} \quad U = Ws.Us(S) + W_N.U_N(N)$$

$$\text{Subject to} \quad N \leq DSF/Ds,$$

$$Nmin \leq N \leq N^*,$$

$$Smin \leq S \leq S^*$$

APPENDIX D. Supplementary Information On Simulations

Simulation Tools: NS2 may be the most famous discrete event simulator that provides extensive support for simulating TCP/IP, routing and multicast protocols over wired and wireless networks. Radio propagation model based on two ray ground reflection approximation and a shared media model in the physical layer, an IEEE 802.11 MAC protocol in the link layer and an implementation of dynamic source routing for the network layer were developed in the Monarch project. This facilitates the simulation of wireless networks by NS2.

SensorSim builds on NS2 and claims to include models for energy and the sensor channel. At each node, energy consumers are said to operate in multiple modes and consume different amounts of energy in each mode. The sensor channel models the dynamic interaction between the physical environment and the sensor nodes. This simulator is no longer being developed and is not available.

Objective Modular Network Test-bed in C++ (OMNeT++) is a public-source, component-based, modular simulation framework. It has been used to simulate communication networks and other distributed systems. The OMNeT++ model is a collection of hierarchically nested modules. OMNeT++ offers an extensive simulation library that includes support for input/output, statistics, data collection, graphical presentation of simulation data, random number generators and data structures. OMNeT++ simulation kernel uses C++ which makes it possible to be embedded in larger applications

OPNET Modeler is a commercial platform for simulating communication networks. Conceptually, OPNET model comprises processes that are based on finite state machines and these processes communicate as specified in the top-level model. The wireless model is based on a pipelined architecture to determine

connectivity and propagation among nodes. Users can specify frequency, bandwidth, and power among other characteristics including antenna gain patterns and terrain models.

J-Sim is another object-oriented, component-based, discrete event, network simulation framework written in Java. Modules can be added and deleted in a plug-and-play manner and J-Sim is useful both for network simulation and emulation by incorporating one or more real sensor devices. This framework provides support for target, sensor and sink nodes, sensor channels and wireless communication channels, physical media such as seismic channels, power models and energy models.

GlomoSim is a collection of library modules, each of which simulated a specific wireless communication protocol in the protocol stack. It is used to simulate Ad-hoc and Mobile wireless networks.

Simulation parameters used in the thesis:

Variable Name	**Type**	**Explanation**	**Simulation Value**
Tsim	Integer constant	Simulation duration in epochs	10000
Ninitial	Integer constant	Initial number of sensors deployed	100
Nmin	Integer constant	Minimum defined spatial resolution	15
Nmax	Integer constant	Maximum defined spatial resolution	35
Smin	Integer constant	Minimum defined security	0
Smax	Integer constant	Maximum defined security	3

N*min	Integer constant	Minimum spatial resolution requirement	15
S*min	Integer constant	Minimum security requirement	0
N*	Vector of size Tsim	Spatial resolution requirement (input)	-
S*	Vector of size Tsim	Security requirement (input)	-
Ds	Vector of size Smax-Smin+1	Data slot length for each security level	[42 43 47 55]
Nsmax	Vector of size Smax-Smin+1	Maximum spatial resolution supported at each security level	[25 24 22 19]
Us	Vector of size Smax-Smin+1	Values of security utility function for each security level	[0.5 0.75 0.9 1]
U_N	Vector of size Nmax-Nmin+1	Values of spatial resolution utility function for each security level	[0.525 0.55 … 0.975 1]
Ws	Integer constant	Weight for security utility function	1
W_N	Integer constant	Weight for s. resolution utility function	1
BC	Integer Constant	Initial battery capacity	0.066
A	Integer Constant	Number of automata states	3 or 4
P_A	Vector of size A	Transmit probabilities of each automata state	[0.05 0.1 1] or [0.001 0.5 1 1]
T	Integer Variable	Current time epoch	1 to Tsim
N’	Vector of size Tsim	Supported spatial resolution values	-

S'	Vector of size Tsim	Supported security values	-
N	Vector of size Tsim	Attained spatial resolution (output)	-
S	Vector of size Tsim	Attained security (output)	-
Ntotal	Integer Variable	Total number of alive sensors	-
Nt	Integer Variable	Expected current spatial resolution	-
IsReward	Boolean Variable	Whether nodes will be punished or rewarded	-
Nodes	Matrix of size 4 x Ninitial	State matrix for all sensor nodes. Four columns keep info on (1) automata state, (2) current battery level, (3) whether node is dead, (4) whether it wants to transmit. This matrix, an example of it is below, is updated at each epoch.	-

	Node1	.....	Ninitial
State	1		3
Battery	0.0054		0.0061
IsDead	0		1
IsActive	1		0

Pseudocode of the simulation algorithm used in the thesis:

The implemented simulation code in MATLAB reads all constant values and vectors given above from an input file. The integer and vector variables are initialized as follows: Nt=0, Ntotal=Ninitial, t=1, N=0, S=0, N'=N*, S'=S*. All

rows of the **Nodes** matrix will be initialized as state= ⌐A/2¬ , Battery=BC, IsDead=0, IsActive=0. Then, the following main loop will be executed.

WHILE (number of alive sensors is enough to support minimum spatial resolution requirement) and (simulation time not finished)

IF either security or resolution requirement changed with respect to previous epoch

IF new requirements cannot be supported

Compute supported security and resolution levels S' and N'

END

END

*IF computed supported values suffice minimum requirements (S'>N*min, N'> S*min)*

SET Nt = 0;

FOR all of the Ninitial number of nodes

IF (node is alive) and (transmit probability greater than generated random num.)

SET node's IsActive attribute

INCREASE Nt by 1

ELSE

UNSET node's IsActive attribute

END

END

IF Nt is less than Nsmax

SET N(t)=Nt

ELSE

SET N(t)=Nsmax

END

SET S(t)=S' (t);

IF Nt is less than N'(t)

SET IsReward=1;

ELSE

IsReward=0;

END

SET Ntotal=Ninitial;

FOR all of the Ninitial nodes

IF node is active

CHANGE the node's state according to IsReward

IF node actually transmits,.ie., its name is in the scheduled nodes list

DECREASE the node's power level by Ps units

END

IF node's battery level is below zero

SET node's IsDead attribute ;

END

END

SET Ntotal=Ntotal-nodes's IsDead attribute;

END

END

SET simulation time t=t+1;

END

APPENDIX E. Implementation of a Temporal Role Based Access Control Scheme

Although access control is more vital for enterprise level networks where there are several objects (resources) to which several subjects (users) require access, it is important also for wireless sensor networks not due to resource diversity but resource limitations. Therefore, protecting these resources by preventing the unintended use of them enhances the operation of sensor networks. One of such resources that is dealt with in this thesis is the channel capacity. Under a certain channel capacity limit, it is tried to optimize the number of sensor nodes that can send data at a certain security level. Therefore, better utilization of the communication channel by regulating who can access it helps the achievement of the overall aim of this thesis, that is, providing certain QoS and security levels under a certain communication bandwidth.

Based on those observations, in the sequel, it will be shown that the proposed control method of this thesis can implement a role based access control scheme to prohibit the unauthorized use of the channel by mediating access attempts of sensor nodes. Note that the originally proposed method in the main chapters has already included access control by the way of cluster head's exclusive selection of nodes to transmit data. Yet, it did not involve any formal access control method and did not mention how to implement access control. Below, the details of the access control involved in the proposed method are given. It is based on the temporal and spatial RBAC (TS-RBAC) scheme proposed in a previous journal paper (Tomur & Erten, 2006).

Three roles are defined as *active_sensors, all_sensors* and *cluster_head*. Members and privileges of those roles are shown in the table below. Membership for those roles is static during the operation of the sensor network except for the *active_sensors* role whose members are dynamically determined by the cluster

Table E1: Roles and access privileges

Role Name	**Members**	**Access Privileges**
active_sensors	Sensors transmitting during data transmission period	Access the assigned data slot announced by cluster head
all_sensors	Originally deployed authentic sensors	Access a random mini slot of the contention subframe
cluster_head	Cluster Head	Access the Header slot

head at each frame. Thus, role membership is temporal for the *active_sensors* role. In fact, there is no limitation to make the other two roles temporal depending on the requirements. For instance, if new sensors are deployed, *all_sensors* role can be temporal or if the cluster head duty is rotated among available sensors, then the *cluster_head* role can be temporal as well. Yet, under the scope of this thesis, it is considered that only the *active_sensors* role is temporal.

Member list of *active_sensors* and *all_sensors* are stored in the cluster head whereas member list of the *cluster_head* is kept by all the nodes. Cluster head updates the member list of the *active_sensors* role before the header period of each MAC frame. Those members are selected using the battery level and location criteria explained in Chapter 6. Therefore, the role membership for the *active_sensors* is temporal depending on the location and available power information of sensor nodes, which makes up a temporal and spatial RBAC scheme as in Tomur and Erten (2006) additionally having battery awareness. This RBAC scheme operates as follows when implemented in the proposed QoS and security control strategy.

- In step 5 of the proposed strategy given in Chapter 6, nodes contend for a mini slot during the contention period. Cluster head checks node ID's of contending nodes against the *all_sensors* membership list and includes

only the ones which are members of the role in the selection process. Mini slot access attempts of contending nodes which are not members of *all_sensors* are discarded and those nodes are blacklisted.

- Using the selection algorithm given in step 6 of the proposed strategy, cluster head selects the nodes to transmit during this frame according to their battery level and location, and updates the member list of *active_sensors* role.
- When cluster head announces the list of sensor nodes to transmit in step 9, sensor nodes receiving this announcement check that the sender of this list is member of the *cluster_head* role. If it is not, they discard the message.
- When the sensor nodes send their data during the data transmission period in step 10, cluster head checks that node IDs of the transmitting nodes are included in the current *active_sensors* role membership table. If there is any node which transmits but it is not a member of this role, data sent by this node is discarded and the node is blacklisted. The cluster head further checks the authenticity of the active nodes which are members of the role by comparing the appended MIC value to the one calculated by the cluster head itself. If any mismatch is detected, the node is marked as malicious.

This spatio-temporal role based access control scheme explained above increases the security level existent in the sensor network as formally proven in Tomur and Erten (2006) and can be employed in hostile environments with stringent security requirements. One limitation of the above scheme is that it only authenticates the identities of sensor nodes via MICs computed by shared keys but it relies on the location and battery level information declared by sensor nodes. So, it is assumed that authentic nodes do not lie about their location and battery levels.

Most of the wireless sensor network security protocols suffer from loosing their effectiveness when sensor nodes are physically compromised by adversaries and cryptographic keys are extracted. This provides the intruder with an ability to send data falsifying the identity of the captured node by using the extracted key. If separate keys are used for different nodes as the proposed scheme of the thesis

suggests, the detrimental effect is less compared to the case when the single key shared among all sensors are compromised by the malicious third parties. Nonetheless, even in the case that sensors possess different keys, node compromise may soil the information quality of the network due to false data sent by the intruder node. By making a small addition to the presented access control scheme above, it can detect whether a node is captured and if it spoofs an authentic sensor node. This modification and the reasoning behind it are given in the sequel.

If a sensor node is the member of the *active_sensors* role for more than T consecutive time intervals and if the total number of nodes that show intent to transmit is above the required spatial resolution level during that time interval, this node is considered to be a malicious one spoofing the identity of an authentic node. The reasoning behind this rule is the following. From simulation results, it is known that the proposed QoS and security control scheme of the thesis causes an even activation of nodes over the time, e.g., any node hardly transmits for two successive intervals. If a node tries to continually transmit despite the discouragement of the employed strategy to do so, this node is most probably a malicious one aiming to insert false messages into the network. Since the overall control strategy is known only by the cluster head and it is assumed that the cluster head has physical resilience against capture, it is not possible for an intruder to find out the selection pattern of the algorithm and change its behavior accordingly. Therefore, if such a traffic pattern is detected conforming to the rule given in the beginning of the paragraph, the node generating this traffic is considered to be malicious. Thus, the role based access control scheme that can be implemented by the proposed strategy of this thesis can also be enhanced to have an ability for detection of malicious nodes that use the keys of authentic nodes to mislead the network.

VITA

PERSONAL INFORMATION

Surname, Name:	Tomur, Emrah
Nationality:	Turkish (TR)
Date and Place of Birth:	August 15, 1977, İZMİR
Marital Status:	Married
Phone:	+ 90 312 455 6726
Fax:	+ 90 312 424 0877
E-mail:	e140795@metu.edu.tr
	emrah.tomur@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	Bilkent University, Electrical & Electronics Engineering	2001
BS	Bilkent University, Electrical & Electronics Engineering	1999

WORK EXPERIENCE

Year	Place	Enrollment
2001 - Present	BDDK, Information Management Dept.	IT Specialist
2000 - 2001	Bilkent Univ., Electrical Engineering Dept.	Teaching Assistant
1999 - 2000	MİKES A.Ş.	Systems Engineer

PUBLICATIONS

1. **E. Tomur** and Y.M. Erten, Y.M., “An Analysis for the Correlation of Coverage and Spatial Resolution for Wireless Sensor Networks”, International e-Conference of Computer Science, 2007

2. **E. Tomur** and Y.M. Erten, “Security and Service Quality Analysis for Cluster-Based Wireless Sensor Networks”, Fifth International Conference on Wired / Wireless Internet Communications (WWIC 2007), May 2007, Coimbra, Portugal

3. **E. Tomur** and Y.M. Erten, “Tradeoff Analysis and Optimization of Security and Spatial Resolution for Sensor Networks”, 41st Annual Conference on Information Sciences and Systems (CISS’07), March 2007, Baltimore, MD, United States

4. **E. Tomur,** R. Deregözü, T. Genç, “A Wireless Secure Remote Access Architecture Implementing Role Based Access Control: WiSeR”, XVIII. International Conference on Computer and Information Science and Engineering (CISE’06), December 2006, Vienna, Austria

5. **E. Tomur** and Y.M. Erten, "Application of temporal and spatial role based access control in 802.11 wireless networks", Computers & Security, Volume 25, Issue 6, Sep. 2006, pp 452-458.

6. **E. Tomur** and Y.M. Erten, “A Layered Security Architecture for Corporate 802.11 Wireless Networks”, 2nd Wireless Telecommunications Symposium (WTS 2004), May 2004, Pomona, CA, United States

RESEARCH INTERESTS

Wireless Networks, Information Systems Security, Computer Networks