A STUDY OF PRIVACY-PRESERVING MECHANISMS FOR WIRELESS MULTIMEDIA SENSOR NETWORKS IN HEALTHCARE

YASMINE NAGI MOSTAFA SALEH

A thesis submitted in partial fulfilment of the requirement of Staffordshire University for the degree of Doctor of Philosophy

August 2018

Abstract

Although the importance of privacy is well-acknowledged for sensitive data, a significant research effort is still needed to develop robust privacy protection solutions for Wireless Sensor Networks (WSNs) used in the context of healthcare. The focus of this doctoral research is to investigate privacy-preserving mechanisms for Wireless Multimedia Sensor Networks (WMSNs) for use in healthcare, to ensure privacy-aware transmission (from sensors to the base station) of multimedia data captured for healthcare.

Towards achieving the goal stated above, the following research questions are addressed in this thesis: (i) What are the significant privacy threats in a WMSN used in the healthcare domain? (ii) What countermeasures can be deployed to stop privacy attacks that realize these threats? (iii) What is the impact, on the WMSN, of the deployment of the privacy countermeasures, with regards to the enhancement of privacy and to the associated computation, communication and storage overheads?

A threat analysis, conducted in the research reported herein, revealed that linkability, identifiability and location disclosure are significant privacy threats for WMSNs in healthcare. Consequently, privacy countermeasures and the corresponding mechanisms to achieve unlinkability, anonymity / pseudonymity and location privacy are required in a privacy-aware WMSN for healthcare. The AntSensNet routing protocol (Cobo et al., 2010) for WMSNs was adapted in the work reported in this thesis, by adding to it privacy-preserving mechanisms, towards achieving unlinkability, anonymity / pseudonymity and location privacy. The standard AntSensNet routing protocol is vulnerable to privacy threats. Consequently, the following countermeasures were investigated in this thesis: (i) size correlation and encryption of scalar and multimedia data transmitted through a WMSN, and size correlation and encryption of ants, to achieve unlinkability and location privacy; (ii) fake traffic injection, to achieve anonymity, source location and base station location privacy, as well as unlinkability; (iii) pseudonyms, to achieve unlinkability.

To assess the impact of the introduction of the above privacy countermeasures, a quantitative performance analysis was conducted (using the NS2 simulator and a theoretical analysis) to gauge the computation overhead (number of extra operations), communication overhead (number of extra network messages) and storage overhead (number of extra encryption keys) of the privacy countermeasures which were added to the AntSensNet protocol deployed within a WMSN. The performance analysis results show that the messages and memory overheads due to the added privacy countermeasures increase mostly linearly with the number of scalar and multimedia sensors, and the resulting traffic, increases in the network.

Furthermore, a WMSN (with sensors having specifications similar to healthcare sensors, but not deploying the privacy-aware AntSensNet protocol) was simulated using the NS2 simulator, to study the effect of the introduction of fake traffic, towards achieving unlinkability, anonymity and location privacy. Entropy and anonymity set size were adopted to quantify the change in the level of privacy (anonymity, unlinkability and location privacy) as the number of fake sources and the volume of fake traffic increase. The results show that the level of privacy enhancement increases with the number of fake sources and volume of fake traffic, but at the expense of an increased delay in the data delivery and an increased level of multimedia jitter (as a result of the consumption of the available bandwidth by fake traffic). This delay and jitter might not be acceptable in critical situations where rapid medical action is required, such as for a patient who has suffered a stroke or a patient (remotely monitored by cameras) who has fallen and broken a bone.

The novel contributions to knowledge which have arisen from this doctoral research are: (i) the elicitation of privacy threats, through a threat analysis methodology named LINDDUN (Wuyts et al., 2014) ─ applied to WMSNs for healthcare ─ to identify significant threats and hence the privacy enhancement mechanisms required by a privacy-aware WMSN; (ii) the enhancement of the AntSensNet routing protocol for WMSNs, to make it privacy-aware; (iii) the findings from the assessment of the privacy-awareness resulting from the deployed privacy-enhancing countermeasures and findings from the assessment of their associated computation, communication and storage overheads.

Acknowledgments

I would like to express my deepest gratitude and indebtedness to Dr. Claude Chibelushi for his valuable critical advice, additions and improvements throughout my PhD. I am grateful to Dr. Ayman Adel for his valuable support, encouragement and helpful advice. I would also like to express my sincere thanks to Dr. Abdel Hamid Soliman for his help and support.

I am fully indebted to my father and mother for their inspiration and precious encouragement. I am deeply indebted to my husband and children for their priceless support and valuable assistance. Very special thanks to my sister for her priceless help and inspiration.

I would also like to thank the Arab Academy for Science, Technology and Maritime Transport for sponsoring my PhD.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACO	Ant Colony Optimization
BAN	Body Area Network
BANT	Backward Ant
BS	Base Station
CANT	Cluster Ant
CH	Cluster Head
DFD	Data Flow Diagram
ECC	Elliptic Curve Cryptography
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
FANT	Forward Ant
RFID	Radio-Frequency IDentification
HIPAA	Health Insurance Portability and Accountability Act
IOI	Item of Interest
IP	Internet Protocol
LINDDUN	Linkability Identifiability Non-repudiation Detectability information Disclosure content Unawareness policy and consent Non-compliance
PAN	Personal Area Network
PDA	Personal Digital Assistant
QoS	Quality of Service
TTL	Time to Live
URL	Uniform Resource Locator
WBSN	Wireless Body Sensor Network
WMSN	Wireless Multimedia Sensor Network
WSN	Wireless Sensor Network

Chapter 1 Introduction

Physical and mental health are important ingredients of human life. They make a major contribution to the quality of life and economic development of individuals, communities and countries. The provision of quality healthcare, to treat or (ideally) prevent health problems, is thus acknowledged as a worthy priority in modern society. Gaining the trust of individuals, healthcare providers and healthcare organisations participating in the interchange of health information is very important. Lack of trust can affect the disclosure of necessary health information, which might affect the accuracy and completeness of this information and may lead to life threatening risks.

1.1 Motivation of this research

Privacy violations and security risks in healthcare systems may cause leakage of sensitive information about patients' diseases which may be embarrassing or critical; and could cause the patients to lose their jobs, or be unable to obtain insurance, and sometimes lead to risks such as an adversary (or criminal mind) finding the location of a person, with possible life-threatening consequences (Kumar & Lee, 2011). A key factor for the acceptance of the interchange of healthcare information in medical healthcare networks is the consistent and coordinated safeguard of patients privacy and security (US Department of Health and Human Services, 2008). Different countries impose different laws that provide legal foundations for healthcare privacy. Consequently, effective measures against privacy violations are an indispensable prerequisite in most Wireless Sensor Networks (WSN)-based applications, and they are of paramount importance in healthcare applications (Oualha & Olivereau, 2011).

Previous WSN-based healthcare systems may have been experimented on patients or in laboratories using actual sensors (such as CodeBlue (Malan et al., 2004), Mobicare (Chakravorty, 2006) and SATIRE (Ganti et al., 2006)). However, only few healthcare systems, such as ALARM-NET (Wood et al., 2008) and MeDiSN (Ko et al., 2010a), embedded security services in their proposed systems. These systems focus on security issues and view privacy preservation as a by-product of security services. These healthcare systems would be hardly acceptable by patients and by governments who follow privacy legal frameworks. These frameworks refer to laws, regulations and standards concerning the privacy of information relating to the

health of a patient, as stored or exchanged in healthcare systems (Hiller et al., 2011).

Privacy in WSN-based healthcare systems is a complex issue due to the capturing of continuous medical data for potentially long periods of time. The diverse and wide range of data about the medical and the daily routines, and the health information is utilised by different information systems belonging to a wide range of beneficiaries such as insurance companies, life coaches, family, homecare providers, researchers and others (Kotz et al., 2009). In addition, there is the possibility of medical identity theft, where employees having access to the patient records might sell this classified information to third parties or when the identity of a person is forged to illegally receive medication they are not entitled to (Hiller et al., 2011). These issues bring the need for a set of coherent laws and principles to protect the privacy of patients' data.

Privacy in healthcare systems based on Wireless Multimedia Sensor Networks (WMSN) imposes even more challenges due to the nature of the multimedia data (video and audio). Possible privacy risks associated with multimedia content may include the estimation of location information (for example, using multimodal methods which may help estimate the location of a person), estimation of the time of the recording of the multimedia video, identification of the people in the multimedia content (through image-based recognition or voice recognition, for example), detection of valuable objects. In addition, multimedia content can assist in the identification of the sensor types used to record videos or audios (using the unique pixel noises of cameras or frequency responses of microphones), which may threaten the anonymity of the person deploying these sensors (Friedland et al., 2015). The deployment of WMSNs in healthcare systems is constrained by many factors due to the nature and limitations of WMSNs such as power consumption, high bandwidth demand and quality of service (Akyildiz et al., 2007). Time and monetary budget impose a strong constraint when designing a privacy-aware system (Deng et al., 2011). It is important to identify the necessary privacy services for a WMSN-based healthcare system at the design stage to avoid the challenging and the overhead in time and money due to the addition of these services after the system is developed.

"Privacy-by-design" is a privacy engineering methodology that is used to elicit the privacy threats to a system at the design stage to avoid the implementation of the privacy services after the software engineering process (Wuyts et al., 2014).

"Privacy-by-design" refers to privacy protection safeguards that must be taken into consideration during the stages of the engineering process of a system (Danezis et al., 2015). However, privacy is a complex, multifaceted notion that is normally not the main requirement of the system being designed and it might conflict with the system requirements (functional or non-functional). Consequently, the privacy goals must be well defined and evaluated (Danezis et al., 2015). Privacy impact analysis or privacy risk analysis is used to discover the privacy objectives of a system. From a technical point of view, privacy risk analysis is basically about identifying: i) the stakeholders of the system, ii) the risks (putting into consideration the stakeholders of the system) and iii) the possible solutions and recommendations for the risks. This is followed by implementing the solutions and recommendations and finally performing audits and review measures (Danezis et al., 2015). Several methods have been proposed for the identification of the security threats compared to limited research focusing on the identification of the privacy threats (Danezis et al., 2015). One of the few privacy threat analysis methodologies is the LINDDUN methodology (Deng et al., 2011) (Wuyts et al., 2014), which proposes a systematic approach for the identification of privacy threats, using data flow diagrams and threat trees.

When this research was conducted and to the best of the author's knowledge, there has not been a formal privacy threat analysis methodology that was applied to a WMSN-based healthcare sub-system (from the tier of the sensors to the tier of the base station) to assess the significant privacy services that must be present in order to be accepted by patients and governments in real life. Consequently in this research, the LINDDUN privacy threat analysis methodology (Deng et al., 2011) (Wuyts et al., 2014) was applied to a basic WMSN-based healthcare sub-system to discover the significant privacy services that must be present in this healthcare sub-system. The proposed sub-system included medical sensors, environmental sensors, audio and video sensors and the focus of this research is on the part of the healthcare system, from the data capture by the sensors until the data arrives at the base station. The outcomes (elicited privacy services) of the privacy threat analysis were introduced into the healthcare sub-system (used in the privacy threat analysis) to create a privacy-aware WMSN-based healthcare sub-system.

1.2 Contribution to knowledge

The novel contribution to knowledge of this doctoral research is threefold.

The first contribution is the application of a privacy threat analysis method to the WMSN-based healthcare sub-system, to identify the most significant privacy

mechanisms. These privacy mechanisms are used to create a privacy aware WMSN-based healthcare sub-system.

The second contribution is the enhancement of the AntSensNet (Cobo et al., 2010) WMSN-based routing protocol, to make it privacy and security aware (using a key management protocol called LEAP (Zhu et al., 2006)). It is envisaged that the enhancement will increase the domain of deployment of this routing protocol, to include applications requiring privacy and security services. The AntSensNet routing protocol will be used as the underlying routing protocol to create a privacy-aware WMSN-based healthcare sub-system.

The third contribution takes the form of the findings from the assessment of the privacy-awareness resulting from the deployment of the privacy-enhancing countermeasures, and findings from the assessment of their associated computation, communication and storage overhead.

1.3 Aim and objectives

The aim of this doctoral research work is to investigate privacy-preserving mechanisms for healthcare systems based on WMSNs striving to ensure privacy-aware transmission of multimedia-captured data, from the sensors to the base station. To successfully achieve this aim, the following objectives must be fulfilled:

1. Conduct a review of architectures (system models) for WMSN-based healthcare sub-systems (Chapter 2).
2. Conduct a thorough survey to identify the general techniques used in the literature to implement WSN-based privacy mechanisms (Chapter 3).
3. Conduct a systematic investigation and analysis of the reference architecture developed in Objective (1) to define the possible privacy threats. A privacy threat analysis model will be used to identify the potential privacy attacks that can target a WMSN-based healthcare sub-system, and select the corresponding privacy preserving mechanisms (Chapter 4).
4. Perform a systematic survey of the WMSN-based routing protocols to select an appropriate routing protocol for the implementation of the privacy mechanisms identified in Objective (3) (Chapter 5).
5. Evaluate the performance of the WMSN-based healthcare sub-system using exhaustive qualitative and quantitative analysis, which implements the reference architecture and the identified privacy mechanisms for the WMSN sub-system for healthcare applications (Chapter 6, Chapter 7 and Chapter 8).

1.4 Thesis outline

This thesis is made up of nine chapters. The main idea of each chapter is as follows:

- Chapter 1: This chapter is a description of the motivations, contribution to knowledge, aims and objectives of this research work.
- Chapter 2: An overview of the architecture of the WSNs, its applications in the healthcare sector and the most popular WSN-based healthcare systems are discussed.
- Chapter 3: An overview of the basic terminologies related to privacy and a listing of the definitions of the basic privacy terminologies, suggested by (Pfitzmann & Hansen, 2010; Haddad et al., 2011) is presented. Next a survey of the general WSN privacy-preserving techniques for anonymity, pseudonymity, unlinkability, undetectability and unobservability is given, followed by a review of highly cited survey papers, which focused on the privacy of WSNs. Finally, the deployment and the privacy of WMSNs in healthcare systems is discussed.
- Chapter 4: This chapter presents an overview of the privacy threat analysis methodologies that have been reported in the literature, and how the LINDDUN privacy threat analysis methodology was applied to the suggested WMSN-based healthcare sub-system to determine the list of privacy services that need to be considered in this research. The general outline of the LINDDUN privacy threat analysis methodology is discussed followed by the creation of a data flow diagram for a WMSN-based healthcare sub-system and then the data flow elements are mapped to the LINDDUN privacy threats. Finally, a discussion of the mapping results is presented.
- Chapter 5: This chapter outlines a brief overview of a chosen number of application scenarios for the privacy-aware WMSN-based subsystem for the healthcare domain. Next, a brief explanation of the logical layout of the network components of the proposed WMSN-based healthcare sub-system is discussed. This is followed by a short survey of routing protocols in WMSN. Next, a discussion of the choice of the routing protocol, which was adopted in this research, is presented followed by a thorough discussion of the security building block, which was adopted in this research work. Afterwards, a brief privacy assessment of the chosen WMSN-based routing

protocol is presented, followed by an outline of the proposed algorithm and details of the flow of messages among the system components.

- Chapter 6: This chapter presents an overview of the different metrics used to quantify privacy mechanisms, anonymity, unlinkability and location privacy are outlined. The privacy assessment methodology used in this research work is presented; the chapter includes a discussion of the criteria for the choice of privacy metrics, the chosen privacy metrics and the experimental method of this research work.
- Chapter 7: This chapter presents an analysis of the anticipated overhead due to the introduction of privacy measures to the WMSN-based healthcare subsystem. The overhead was analysed using both simulation experiments and theoretical analysis.
- Chapter 8**:** This chapter discusses the assessment of the enhancement of anonymity, unlinkability and location privacy due to the introduction of fake traffic. The chapter includes a discussion of the simulation experiments (such as aims, equipment, experiment design and entropy calculation) and the results of the experiments.
- Chapter 9: A conclusion of the whole thesis, a discussion of the limitations of the work presented in the thesis and possible future work are given in this chapter.

Chapter 2 Wireless Sensor Networks in Healthcare

2.1 Introduction

A WSN is a self-organising multi-hop wireless network of nodes that is made up of tens to thousands of sensor devices that are deployed to collect data from, for example, the surrounding environment or from a human body and wirelessly send the data to a base station. Numerous areas of applications for the deployment of WSN have been in diverse fields such as healthcare; metropolitan, military or environmental monitoring; animal tracking; industrial automation; civil engineering; logistics and transportation; and sports The deployment of WSNs in the healthcare sector promises numerous applications that are expected to enhance the life style of humans in diverse aspects of life ranging from the prediction of foetal diseases to the assistance of elderly people and the disabled. However, the acceptance of these applications is bound by the safety and privacy of the personal and intimate information that can be captured by the WSNs.

The aim of this chapter is to present an overview of the architecture of WSNs and their applications in the healthcare sector. This is followed by a brief outline in which selected WSN-based healthcare systems will be discussed. The basic architecture of WSNs in the healthcare sector will be deployed in this research work in the privacy threat analysis and the investigation of the privacy-preserving mechanisms.

2.2 Basic WSN architecture

A WSN can be deployed to sense the relevant phenomena (such as temperature, motion, light, sound, heart or brain activity, blood pressure or oxygen saturation in the human body). Generally, the sensor devices collect data (from the surrounding environment or from the person the sensors are attached to, for example), then they wirelessly send them to collection devices called base stations (Deif & Gadallah, 2014). Depicted in Figure 2-1 is the basic architecture of WSNs (Sohraby et al., 2007). It is worth mentioning that the diagram in Figure 2-1 does not match the fact that (in the healthcare context) the user could be interacting directly with the sensors (e.g. wearable, implanted sensors for a patient). The diagram only depicts the perspective of healthcare staff or family, who could access the data via the Internet.

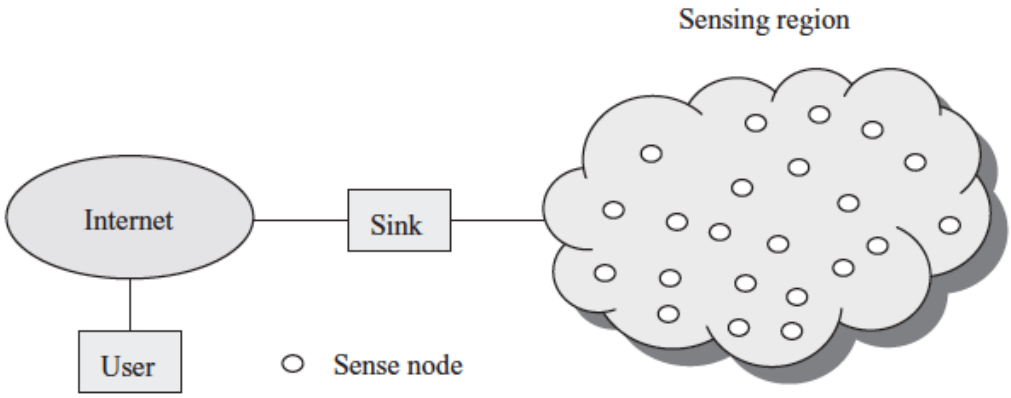


Figure 2-1 General Sensor Network Architecture (Sohraby et al., 2007)

WSNs possess many distinctive features that serve as a guideline for the design and development of protocols and algorithms. Accordingly, WSNs are distributed, data-centric, collaborative, redundant, autonomous, application-specific, hierarchical, and resource constrained. The resource constraint is considered to be a significant design challenge and limitation of WSNs. Besides the limited power resources, constraints are also made on the size, densities, production costs, memory space, communication bandwidth requirements and computation powers of the sensor nodes (Akyildiz, 2007; Wilson, 2004).

2.3 WSNs in healthcare

The deployment of WSNs in the healthcare domain is expected to improve both the quality of healthcare services and the quality of the lives of patients (Tavares et al., 2008). The WSNs deployed in the healthcare sector to monitor the vitals of a patient are viewed as a subset of the WSN called Wireless Body Sensor Networks (WBSN). A WBSN is a group of small, lightweight, intelligent and low power wireless sensor nodes that can be placed on or inside a human body to continuously monitor the health of a human (Ha, 2015). Although WBSNs are, to a significant extent, based on the WSN technology, they impose more challenging design considerations compared to the traditional WSNs particularly in the scale of deployment, important design targets, topology, data rate, power consumption, security level, data/sensor loss tolerance and characteristics of the sensors deployed (Ha, 2015) (Chen et al., 2011) (Honeine et al., 2011).

Figure 2-2 outlines the general deployment of the WSN nodes in the healthcare domain and how they can be further categorised into body sensors (wearable and implanted nodes) and nodes used to sense the surroundings of the patient. The surroundings of the patient can be sensed using environmental sensors to detect,

for example, the pressure, temperature and dust of the patient's environment. Audio and video sensors are used to record audio sounds and videos of the patient and his/her surrounding (Virone et al., 2006).

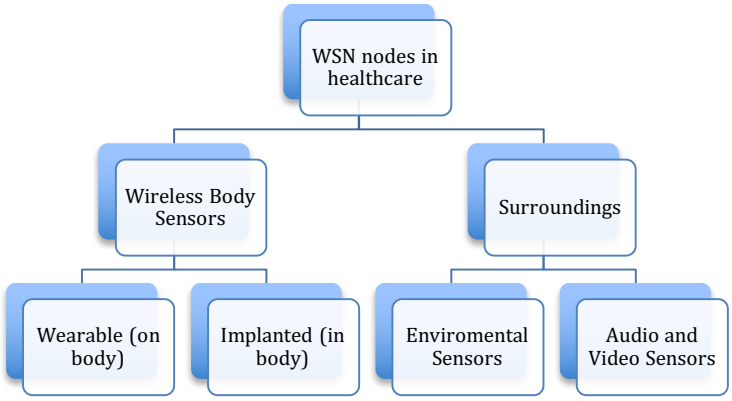


Figure 2-2 Types of sensors deployed in WSN-based healthcare system

The nodes of WBSNs can be sensor nodes and/or actuator nodes (Latré et al., 2011). Sensor nodes are responsible for measuring physiological readings from wearable sensors carried on the human body or/and from implanted sensors placed inside the human body (Latré et al., 2011). Various types of biomedical sensors can be depolyed in a WBSN to wirelessly monitor a patient, depending on the patient's condition. Biomedical sensors can be categorised as wearable and implanted sensors (Crosby et al., 2012). Examples of wearable sensors are pulse oximeters, electrocardiograms (ECG) to monitor heart activity, blood pressure sensors, electromyogram sensor (EMG) to monitor muscle actvity, activity/motion detectors and electroencephalogram (EEG) to monitor the activity of the brain electricity (Milenković et al., 2006) (Crosby et al., 2012). Examples of implanted sensors are glucose monitoring sensors and sensors included in implantable neural stimulators that are deployed to send signals to the human brain in cases of diseases such as Parkinson's disease (Crosby et al., 2012). On the other hand, actuators perform a specific action based on the data collected from the sensors or according to the user's interactions such as in the case of actuators equipped with an insulin pump and reservoir to administer insulin to a diabetic person (Latré et al., 2011).

Several prototypes have been developed for the deployment of WBSNs and they mostly follow a multilayer/multitier architecture (Chen et al., 2011). Figure 2-3 is a general outline of the basic deployment of a WBSN in a typical healthcare system. Tier 1 is the Body Area Network (BAN) and the Personal Area Network (PAN) (Alemdar & Ersoy, 2010). The BAN is made up of the wearable and/or implanted

sensors and/or actuators that are used for the monitoring of the physiological vitals (or the drug administration in the case of actuators). The PAN might consist of the other sensors that can be deployed around the patient such as environmental, audio or video sensors. In tier 1, each sensor is responsible for sensing, sampling and processing the signals captured (Darwish & Hassanien, 2011). Tier 2 is the personal server stage which is made up of a Personal Digital Assistant (PDA), home computer or a cell phone and it is responsible for interfacing with the networks in tier 1 and the medical server(s) in tier 3 (Chen et al., 2011) (Darwish & Hassanien, 2011). The interfacing of the personal server includes the network configuration (sensor node registration, configuration and security setting) and management features (retrieval and processing of data, scheduling, channel sharing and data fusion) (Darwish & Hassanien, 2011). Tier 3 is the medical service centre, which may include caregivers, medical databases and emergency servers (Chen et al., 2011).

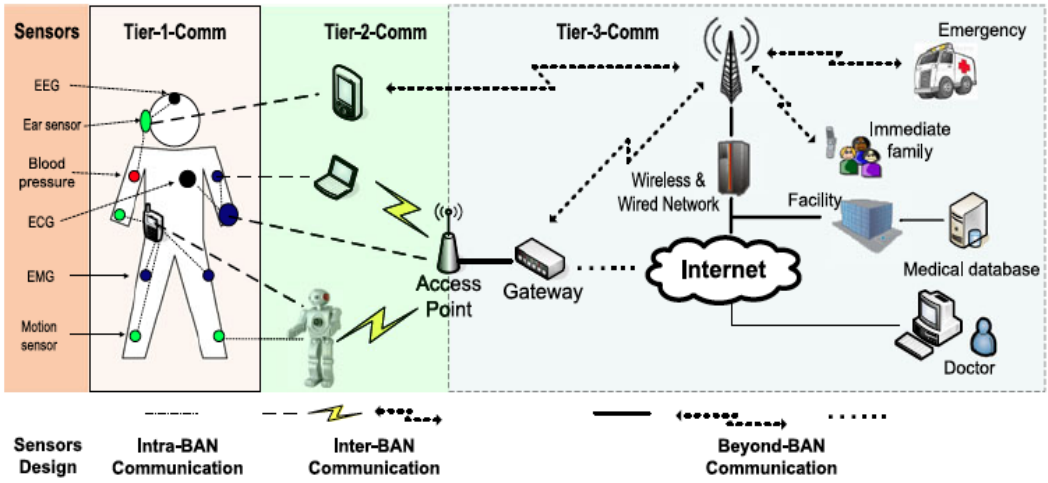


Figure 2-3 Basic WBSN deployment in healthcare systems (Chen et al., 2011)

2.3.1 Applications of WBSNs in healthcare

Many articles have been published that discussed the numerous applications of the WBSNs in different aspects of life, some of them were suggestions for future work and others were proposals or discussions of real projects that were being developed. The applications of the WBSNs can be categorised into healthcare, assisted living, and others (gaming and entertainment, military applications and emergency services) (Chen et al., 2011). Figure 2-4 presents a general outline of the WBSNs applications.

Healthcare applications can be further categorised into chronic disease monitoring, general wellbeing, neonatal healthcare and human activity monitoring. In chronic disease monitoring, WBSNs can be used in the fight against cardiovascular disease where patients are remotely monitored in real time to grant them a healthy lifestyle and provide early prediction of emergencies (Nadeem et al., 2015). According to (Nadeem et al., 2015), the use of WBSNs in general wellbeing can be useful in the development of coaching systems, self-assessment, continuous monitoring and performance evaluation of a normal human being or professionally in the training of athletes, dancers and performers. In neonatal healthcare, a WBSN can be used for the continuous monitoring of newborns and other children, for the detection of infectious diseases, healthy habit monitoring and chronic health issues detection (Nadeem et al., 2015).

Figure 2-4 WBSN applications adapted from (Nadeem et al., 2015)

In assisted living, WBSNs can be deployed within way-finding tools for blind and deaf people, or support for elderly people and rehabilitation. In way finding, WBSNs can be used in the assistance of blind people to move around in familiar and new environments. Furthermore, visually impaired people can have an artificial retina (made up of micro sensors) implanted in the eye that can generate neurological signals based on a camera mounted on eye glasses (Latré et al., 2011). In the context of rehabilitation, continuous remote monitoring of the patients who suffered from, for example, a stroke, joint surgery, or motor dysfunction can be used, without

obstruction, to assist in the rehabilitation and recovery process of these patients. WBSNs can be used for the training and rehabilitation of patients suffering from motor impairment (Nadeem et al., 2015). Sensors can be attached to the legs or the nerves and actuators can be used to stimulate the nerves to assist in the motion (Latré et al., 2011). For elderly support, posture detection applications can be used to detect accidents such as fall of elderly people. WBSNs can be used to assist the independence of elderly people living on their own, by detecting their activities such as walking, lying and even falling (Nadeem et al., 2015).

Other possible applications for WBSNs can be: remote monitoring of soldiers on the battlefield to monitor the fatigue level, the postures and the vitals of the soldiers. In emergency services and extreme situations as with fire fighters and civil protection, WBSNs can be used to detect toxicity in the air and warn them (Latré et al., 2011).

For all previous applications, several attempts have been made, or projects have been developed or are still under development, to support and enhance these applications (Nadeem et al., 2015). Other new application scenarios that are suggested by (Nadeem et al., 2015) are: identifying frostbites for people in very cold weather conditions such as swimmers, soldiers and outside workers, assistance to visually impaired people playing sports such as swimmers and also WBSNs can be used for diabetic people to develop a miniature testing system that will assist, detect and alert in cases of the fluctuation of the sugar level.

2.3.2 Selected WSN-based healthcare systems

This section discusses the early generation of healthcare projects, including highly cited projects in the literature (Minh-Thanh Vo et al., 2015), these are: CodeBlue (Malan et al., 2004), Mobicare (Chakravorty, 2006) and SATIRE (Ganti et al., 2006). With regards to privacy and security, only a few healthcare systems embedded security services such as ALARM-NET (Wood et al., 2008) and MeDiSN (Ko et al., 2010a). However, these systems focus on security issues and view privacy preservation as a by-product of security services. This section presents a quick review of these healthcare projects, along with a table summarising the privacy and security services suggested and implemented by these systems.

The CodeBlue (Malan et al., 2004) architecture has used the Elliptic Curve Cryptography (ECC) technique on its MICA2 motes to ensure security of data transmission. However, the encryption key required 32 seconds to be generated, which is considered unsatisfactory (Egbogah & Fapojuwo, 2011). An extensive

security threat analysis has been conducted on the CodeBlue architecture and it was discovered that it is vulnerable to security attacks such as denial-of-service attacks, snooping attacks, modification attacks, routing loop attacks, grey-hole attacks, Sybil attacks and masquerading attacks (Kumar & Lee, 2011)(Kambourakis et al., 2007). Security attacks on the CodeBlue architecture may have serious impact on privacy. In case of a snooping attack, an adversary might acquire private information about patients by observing the operation of the relevant parts of the physical system. A Sybil attack may result in incorrect decisions based on false health information sent to the sink node. A masquerading attack might have an impact on all privacy services due to gaining access to the whole system using stolen IDs and passwords. Consequently, the CodeBlue suffers from serious privacy vulnerabilities that can prevent it from becoming a WSN-based healthcare system in the real world. Although it was shown to function experimentally, patients may fear to use it when they learn about its serious privacy problems.

The developers of the Mobicare (Chakravorty, 2006) architecture suggested the use of the Wireless Transport Layer Security (WTLS) protocol to provide patient privacy, data integrity and authentication. However, the WTLS was not actually implemented and tested (Kumar & Lee, 2011). This makes the Mobicare architecture vulnerable to all possible privacy attacks, which makes the Mobicare unacceptable by patients in real-life.

SATIRE (Ganti et al., 2006) authors did not implement any of the suggested security and privacy services and considered them as future work (Kumar & Lee, 2011). The developers of SATIRE suggest the use of an access matrix to preserve the privacy of the data in their system. The main idea of this access matrix is to define who can access what. However, this basic security scheme is not enough to stop privacy threats. Limiting the access of the patient's data to authorised personnel or family does not stop adversaries from invading the privacy of the patient. Curious employees or families may access, publish or even sell critical medical information to employers or insurance companies that might cause serious damage to the patient life (caused by the lack of anonymity services). In addition, continuous monitoring might expose the geographical location of the patient, which might not be generally acceptable by many (due to the lack of the location privacy service). Consequently, the SATIRE architecture might be unacceptable to be used by patients in real-life.

The developer of the MeDiSN (Ko et al., 2010a) architecture highlighted the need for encryption to ensure confidentiality and authenticity of the delivered data. However, they did not reveal details about their security authentication or cryptosystems used (Kumar & Lee, 2011).

The Alarm-Net (Wood et al., 2008) architecture uses a secure remote password protocol for user authentication for IP-network security. In addition, sensors used in this healthcare system (such as MicaZ and Telos) have built–in cryptosystems for cryptographic operations. Although Alarm-net offers both authentication and encryption operations, it suffers from major drawbacks. The cryptosystems used are highly platform dependent. In addition, they do not offer decryption options which deny intermediate nodes to access the data during communication (Kumar & Lee, 2011).

Although, the use of both authentication and encryption grants access to lawful personnel and prevents eavesdropping on the traffic, it does not grant location privacy, data delivery privacy, data queries privacy or audio and video privacy. With this long list of privacy services that are not offered by the Alarm-Net architecture, it makes it very hard to be acceptable by patients in real-life applications.

Table 2-1 Summary of the privacy and security mechanisms suggested and implemented by the early generation of WSN-based healthcare systems

	Suggested Mechanisms		**Implemented Mechanisms**	
Project Name	***Security Mechanisms***	***Privacy Mechanism***	***Security Mechanisms***	***Privacy Mechanisms***
CodeBlue	Authentication, cryptography	None	None	None
Mobicare	Shared Keys, Authentication	None	None	None
SATIRE	Authentication	Access Matrix	None	None
ALARM-NET	Authentication, cryptography and key management	Dynamic Privacy	Authentication, encryption	Dynamic authorization to access patient vital information
MeDiSN	Authentication and cryptography	None	Not described	None

Based on Table 2-1, it is evident that the early generation of WSN-based healthcare systems did not focus on the importance of the implementation of privacy services in their systems. None of these healthcare systems thoroughly analysed privacy attacks or studied the related privacy services for defending against those attacks.

Since all the projects mentioned above are not very recent, other recent projects had to be considered to check whether security and privacy services were included or not. More recent WSN-based healthcare systems such as (Nabar et al., 2011), (Rofouei et al., 2011), KNOWME (Mitra et al., 2012) and (Nia et al., 2015) were studied and there was no mention of the implementation of security and privacy services.

In (Nabar et al., 2011), the authors proposed an ECG model-based scheme in which a common ECG model is stored in the sensor and the base station. A senor only transmits data if the recorded data deviates from the pre-stored ECG model. The authors did not mention privacy or security in their proposed work.

In (Rofouei et al., 2011), the authors proposed the use of a soft neck-worn collar with embedded sensors (such as SpO_2, microphone and accelerometer) to monitor and visualise sleep data. The data is collected during the sleep time of a patient and is wirelessly transmitted to a nearby cell phone for storage and processing. This system could be deployed for the detection of sleep disorder. Similar to the previous system, the authors focused on the data collection and sensor management and did not mention privacy or security mechanisms.

KNOWME (Mitra et al., 2012) architecture deployed the three tier architecture depicted in Figure 2-1 and it targeted applications in paediatric obesity. The authors used off-the-shelf medical sensors (such as ECG and SpO_2) and a mobile phone (Nokia N95) to collect long-term data (such as 12 hours per day for several weeks). Their proposed system focused on sensor management and management of buffers collecting data from different sensors. However, the authors did not mention privacy or security measures in their proposed system.

(Nia et al., 2015) proposed an analysis for long term energy-efficient health monitoring. Their proposed analysis focused on the energy and storage requirements of the deployment of eight sensors (heart rate, blood pressure, oxygen saturation, temperature, blood pressure, accelerometer, ECG and EEG). The authors acknowledged the importance of privacy and security in healthcare applications and that was one of the reasons why the authors adopted the BLE communication standard, which supports Advanced Encryption Standard (AES). However, the authors did not mention any details related to the privacy or security of their proposed analysis.

2.4 Summary

This chapter has presented an overview of the WSNs and their general applications, especially in the healthcare sector. The deployment of the WSN in healthcare and a general view of the architecture of the WBSN were depicted. Next, a discussion of the applications of the WSNs in healthcare followed by examples of selected WSN-based healthcare projects was outlined. In addition, a very brief privacy and security assessment of these WSN-based healthcare projects was summarised.

The next chapter will present the state-of-the-art in the privacy of the WMSN.

Chapter 3 State-of-the-art in WMSN healthcare privacy

3.1 Introduction

Although, the use of WSNs in the healthcare sector has undergone some research, most of this research has focused primarily on criteria such as the physical design of the network, system reliability, power consumption, and the cost effectiveness of the prototypes. Comparatively little effort has been dedicated to privacy protection (Minh-Thanh Vo et al., 2015). The importance of privacy and security is well acknowledged for sensitive data, such as data about the health of individuals, but a significant research effort is still needed to develop robust solutions for WSNs used in the context of healthcare. Although technological advances in WSNs for healthcare applications have enhanced the feasibility of continuously monitoring patients or healthy individuals, it is no secret that the success of WSN-based healthcare systems will depend directly on the privacy and security which these systems would be able to provide (Kumar & Lee, 2011).

This chapter starts by presenting an overview of the basic terminologies related to privacy and a review of the definitions of the basic privacy terminologies, suggested by (Pfitzmann & Hansen, 2010; Haddad et al., 2011). The next section presents a survey of the general WSN privacy-preserving techniques for anonymity, pseudonymity, unlinkability, undetectability and unobservability. This is followed by a review of the popular survey papers, which focused on the **privacy** of WSNs. The next section discusses the deployment and the privacy of WMSNs in healthcare systems. Finally, the last section presents an outline of the general privacy-preserving techniques for WMSNs.

3.2 Privacy mechanisms in WSNs

In the survey papers discussing privacy in WSNs, different terminologies were adopted to refer to privacy. Some papers refer to privacy issues as privacy requirements (Li et al., 2010), or goals (Halperin et al., 2008) and others refer to them as privacy problems (Li et al., 2009a). This research work adopts the definitions of services and mechanisms, which are borrowed from information security, and it applies them to privacy protection. According to (Stallings, 2016), a *security mechanism* is defined as “A process (or a device incorporating such a process) that is designed to detect, prevent, or recover from a security attack” and a *security service* is defined as “A processing or communication service that enhances the security of the data processing systems and the information transfers of an organisation. The services are intended to counter security attacks, and they

make use of one or more security mechanisms to provide the service". Security services are subdivided into authentication, access control, data confidentiality, data integrity and non-repudiation (Stallings, 2016).

Following the taxonomy presented by (Li et al., 2009a) and according to the basic information security concepts, in this research work ***privacy aspects*** refer to the main categories of the privacy preserving techniques namely: data privacy (aggregation and query privacy) and contextual privacy (location and temporal privacy). Further, ***privacy services*** will be used to refer to the main privacy goals that ensure the private processing and exchange of information namely: anonymity, pseudonymity, unlinkability, undetectability and unobservability and the privacy goals for the contextual privacy aspects namely location privacy. ***Privacy mechanisms (or techniques)*** will refer to the specific techniques that are used to achieve the privacy aspects of the WSN. Similar to the security services, the privacy services may deploy one or more privacy mechanisms to counter privacy attacks or threats.

To avoid possible misinterpretation of the meaning of the privacy terminology considered by this research, a review of the definitions of the basic privacy terminologies, suggested by (Pfitzmann & Hansen, 2010; Haddad et al., 2011) will first be discussed. The privacy services addressed in this research work are:

Anonymity: "*Anonymity of a subject means that the subject is not identifiable within a set of subjects, the anonymity set*" (Pfitzmann & Hansen, 2010)(Haddad et al., 2011). In other words, anonymity refers to the hiding of the relationship between the identity of a person and the action he/she did or a message he/she sent (Deng et al., 2011).

Pseudonymity: According to (Pfitzmann & Hansen, 2010), "*A pseudonym is an identifier of a subject other than one of the subjects real names. Pseudonymity is the use of pseudonyms as identifiers. A subject is pseudonymous if a pseudonym is used as identifier instead of one of its real names*." In other words, pseudonymity refers to the use of an ID other than the real ID of a person to perform actions (e.g. using pseudonyms to subscribe to online services) (Deng et al., 2011).

Unlinkability: "*Unlinkability of two or more items of interest (IOIs), (e.g., subjects, messages, actions) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not*" (Pfitzmann & Hansen, 2010). In

other words, unlinkability refers to not being able to link two or more messages, actions or identities (e.g. not being able to link two messages from the same source or not being able to link two records in a database to one person) (Deng et al., 2011).

Undetectability: "*The Undetectability of IOIs is the state that whether they exist or not is indistinguishable. In other words, undetectability protects IOIs from being exposed. That is, the message transmission is not discernable from a random noise*" (Haddad et al., 2011) (Pfitzmann & Hansen, 2010).

Unobservability: According to (Haddad et al., 2011)(Pfitzmann & Hansen, 2010), "*unobservabilty can be defined as the undetectability by unrelated subjects together with anonymity (even if an IOIs can be detected)*".

3.3 Privacy preserving techniques for WSNs

Many privacy mechanisms developed for WSNs, in general, can be adapted to the healthcare domain. However, further analysis and experiments need to be applied to these mechanisms to ensure their compatibility with the limited resources such as those applications which require WBSNs. Consequently, the privacy services mentioned in this section are not all explicitly designed for healthcare systems; some are targeting WSNs in general. Following (Li et al., 2009a), the taxonomy tree presented in their work will be adopted in this research work and will be enhanced to fit the privacy for the WMSN-based healthcare systems. The rest of this section is dedicated for the discussion of the privacy services depicted in the taxonomy tree of Figure 3-1.

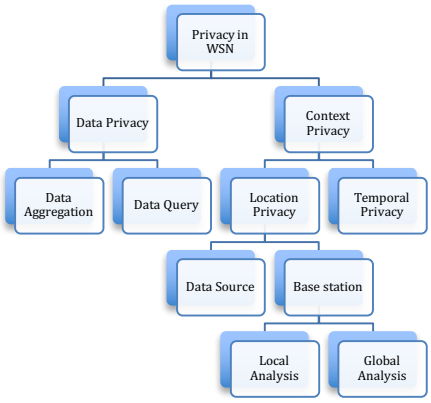


Figure 3-1 Taxonomy of privacy preservation in WSN (Li et al., 2009a)

When attempting to extend the taxonomy tree to the next level to include the privacy services, it is obvious that the tree leaves become interleaving. For example, anonymity will exist underneath both query and location privacy. Consequently,

adopting a tree presentation will not be appropriate to represent the privacy mechanisms. To avoid possible confusion, it is best to base the discussion of the privacy from the point of view of the privacy mechanisms: anonymity, pseudonymity, unlinkability, location privacy, undetectability and unobservability.

3.3.1 Anonymity

In WSNs generally, the anonymity service has been addressed in many papers such as (Ebrahimi & Younis, 2011), (Acharya & Younis, 2010), (Ward & Younis, 2015) and many others. Research papers have focused on specific categories related to anonymity such as

- Base station anonymity
- User anonymity
- Query anonymity
- Source anonymity
- Data collection anonymity
- Communication anonymity

Base station anonymity: Base station anonymity denotes the hiding of the identity, role and location of the base station from external adversaries (Acharya & Younis, 2010). Attacks on the base station can have a debilitating effect on the network because the sink is the central point of all traffic and a critical part of the network. These attacks might cause serious damage to the WSN, which makes the base station very vulnerable (Ebrahimi & Younis, 2011). Different approaches have been adopted to try to protect the anonymity of the base station against malicious attacks. Most of these approaches rely on creating a perception that the base station is a typical sensor node. (Acharya & Younis, 2010) suggest two approaches for base station anonymity. The first approach is based on making the base station transmit messages to random sensors in its neighbourhood. These neighbourhood sensors will later retransmit these messages away from the base station thus deceiving and misdirecting the adversary that the base station is just another sensor node in the network. The second approach is a second line of defence for the long time traffic analysis by adversaries that may eventually reveal the identity of the base station. In this approach, the base station may be relocated when motion is possible. However, the relocation of the base station must be carefully analysed to calculate the threat level and implications of the relocation (Acharya & Younis, 2010). Other means to achieve base station anonymity are suggested by (Deng et al., 2005).

Another approach suggested by (Ebrahimi & Younis, 2011) to increase the anonymity of the base station is to increase the transmission power of the nodes of the network to achieve longer transmission ranges which increases the correlation link between the neighbours and makes traffic analysis very complex. Although this approach has avoided changes in routing protocols and traffic patterns to deceive adversaries, increasing the transmission power may have a serious effect on the network life time and on the interference between the signals (Ebrahimi & Younis, 2011).

Recently, (Ward & Younis, 2015) suggested the use of beamforming to boost the anonymity of the base station while minimising the communication overhead. (Ward & Younis, 2015) explained distributed beamforming as nodes with single antennas cooperating together to form a virtual multi-antenna system to improve the communication range, data rate, energy efficiency, security of the physical layer and decrease the interference with the wireless networks. The deployment of the distributed beamforming is divided into three components: a cross layer relay selection algorithm to determine which nodes will participate in the beamforming, a time synchronisation to construct a common time reference and a carrier synchronisation to create a common reference. According to (Ward & Younis, 2015), their attempt to deploy beamforming in base station anonymity is the first in the literature.

All previous approaches were based on *hiding the identity* of the base station and attempting to make it appear as a sensor node. Other techniques aim to *disguise the location* of the base station. An adversary can use traffic analysis and packet tracing to locate the base station (Li et al., 2009b). Traffic analysis is based on the idea that the traffic volume (number of packets being forwarded) near the base station tends to be bigger than that away from the base station which makes the location of the base station deducible based on the traffic volume (Li et al., 2009b). Packet tracing is deployed by an adversary to learn the hop-by-hop transmission links of the nodes towards the base station (Li et al., 2009b). It is claimed by (Li et al., 2009b) that packet tracing is more efficient that traffic analysis to deduce the base station location. Many authors tend to present countermeasures against both traffic analysis and packet tracing in their work as in (Deng et al., 2005), (Deng et al., 2006). (Li et al., 2009b) and (Chen, 2007) present countermeasures against packet tracing. (Riosa et al., 2015) present their countermeasures against traffic analysis.

(Deng et al., 2005) rely on four techniques based on randomised traffic volumes as a defence against traffic analysis to protect the location of the base station namely: multi-parent routing scheme, random walk, random fake paths and fractal propagation. Multi-parent routing relies on the random selection of one of the parent nodes connected to the node to forward the data to the base station, which makes it hard for an adversary to detect a pattern to lead to the base station. In the random walk technique, the node forwards the packets to its parent nodes based on random forwarding algorithm thus distributing the traffic of the packets and decreasing the effects of rate-monitoring attacks. Random fake paths rely on introducing fake paths on the way from the node to the base station to reduce the effect of time correlation attacks. Finally, fractal propagation is based on the creation and propagation of fake messages into the network to create areas of high activity and randomness in the communication pattern to defend against rate monitoring attacks. In addition to the randomness approach presented by (Deng et al., 2005), the authors suggest more countermeasures aiming at hiding the location of the base station in (Deng et al., 2006). In (Deng et al., 2006), the authors suggest the use of: hidden packet destination address, decorrelating packet sending time and controlling packet sending rates. Hidden packet destination address is done through the encryption of the packet destination address, packet type and content to hide the final destination of the packet (base station). Decorrelating sending time is achieved by introducing a random delay time between the sending and the receiving of the packets between the parent and child nodes to try to stop the adversary from learning the hierarchy tree of the network. Controlling packet sending rates is achieved by creating a uniform traffic volume in the entire network. However, the authors believe that all previous countermeasures which may limit the data sending pattern of the network, are not feasible in cases when urgent data needs to be sent to the base station quickly, as they increase energy loss (due to the use of dummy packets) and increase the overall delay of the network (due to the introduction of random delays). The authors believe that the methods introduced in their previous work (Deng et al., 2005) outperform those in (Deng et al., 2006).

Similarly, other techniques suggested by (Chen, 2007) and (Riosa et al., 2015) aim to provide countermeasures against packet tracing using a location privacy routing protocol combined with fake messages injection. The main idea of (Chen, 2007) is to randomise the routing paths towards the base station and inject fake messages into the network to uniformly distribute the incoming and outgoing traffic at a sensor node. Although their scheme aims at hiding the location of the base station, the

trade-offs between location privacy versus energy consumption and network delays should be analysed. (Riosa et al., 2015) used the injection of fake messages sent towards the base station using a biased random walk combined with a routing table perturbation scheme. The authors claimed that their suggested work is robust against local adversaries.

Another method to hide the location of the base station is suggested by (Li et al., 2009b) who suggests two methodologies to hide the location of the base station during topology discovery and data transmission. During topology discovery, an anonymous topology discovery scheme is used to conceal the location of the base station. In this scheme, a common sensor node, pseudo base station, is randomly chosen by the base station to pretend to be a base station and initiates a topology discovery. During the data transmission phase, the base station location is concealed using an intelligent fake message injection scheme. This scheme is based on the idea that a sensor node will transmit fake messages to the neighbouring nodes at the same time it is forwarding real packets, which causes the adversary to spend more time studying fake paths. This scheme is combined with a simple version of the random walk algorithm to hide the location of the base station.

User anonymity: Allowing users to directly access sensor nodes to obtain real-time data requires considerable security and privacy measures to protect critical data (Ding & Ping, 2014). Accordingly, anonymous user authentication is a crucial mechanism to grant access for rightful users (Ding & Ping, 2014). Smart-card-based password authentication (Two-factor authentication) is deployed in user authentication mechanisms due to their simplicity, portability, and security (Ding & Ping, 2014). Several research works have been conducted in the literature to achieve user anonymity for two-factor authentication in WSN. Some of this research was intended for use in general WSNs and others were intended for WSN-based healthcare systems. (Ding & Ping, 2014) have presented a survey of the two-factor authentication schemes for WSNs with an assessment of previous attempts to design user anonymous two-factor authentication schemes. They discuss different solutions for user anonymity and the complexity of the privacy preservation solutions.

(Nam et al., 2015) present a user anonymity scheme to guarantee the privacy of Smart CArd based user authentication scheme for WSN (SCA-WSN) in which a user holding a special smart card issued by the gateway can gain access to the sensor data after gaining authentication from the gateway. The basic idea of their

scheme is to deploy elliptic curve cryptography that is only used when there is a user-gateway authentication to anonymously authenticate users to access the sensor node data.

The anonymity privacy service has also been a target for the WSN-based healthcare systems. Some papers emphasised the crucial importance of the implementation of anonymity service in healthcare systems as in (Sun et al., 2010), (Meingast et al., 2006). However, they did not attempt to discuss details of anonymisation techniques. To the best of the author's knowledge, the main focus of anonymity for WSN-based healthcare systems was on the user authentication anonymity. For user anonymity, the authors of (Li et al., 2015b) propose a biometric authentication-based protocol for ensuring the patient's privacy. They claim that their protocol enhances the anonymity level compared to others and is computationally more efficient. Other papers aiming for user anonymity are (Agrafioti et al., 2009), (Das & Goswami, 2013). However, there has been little work targeting the rest of the anonymity categories in WSN-based healthcare compared to those in the general WSN.

Query anonymity: Some WSNs are designed to be owned and deployed by the same organisations, while others are designed and deployed by more than one organisation and may extend in more than one country (Carbunar et al., 2010). Clients issuing queries to these WSNs may require the anonymisation of their interests and query patterns, which urges the need for private and efficient queries (Carbunar et al., 2010). In addition, query anonymity can be of critical importance where an adversary can deduce, based on the increasing number of queries to a specific location where a patient dwells, that this patient might have health problems (Li et al., 2009a). Several techniques have been developed to provide query anonymity.

(Carbunar et al., 2010) proposed a protocol to ensure full client query privacy in a network with honest but curious non-cooperative servers. The protocol divides the interaction of the client with the sensor network into two tasks: private naming of each sensor and private accessing of the readings of the sensor nodes. The protocol is made up of four procedures: initialisation, key space generation, query routing and result routing. The main idea of this protocol is that the clients use the key space generation to generate fresh keys shared with the sensor networks and unknown to the servers. The clients utilise these fresh keys during the query and result routing to create packets that will be routed through the network and only

interpreted by the designated nodes. The authors use two privacy metrics to quantify the privacy level: spatial and temporal privacy levels.

(Zhang et al., 2009) present the Distributed Privacy Preserving Access Control (DP^2AC), which ensures anonymous access to the sensor data. The main idea of this scheme is that a client willing to access the data of a sensor network must first buy tokens from the network owner. Once the token is validated, the client is able to access the required data. To ensure client anonymity, the token generated involves blind signatures, which can be validated by the sensor nodes and at the same time cannot identify the token holder. This scheme ensures both the privacy protection of the clients and prevents the unauthorised access of the sensor nodes. Their proposed scheme also has a Distributed Token Reuse Detection (DTRD) scheme to ensure that sold tokens are not reused by malicious users.

For the sake of completeness, it is important to mention that the trade-off between the communication costs versus the query anonymity has been a critical issue that was discussed in many research papers. Further reading about this issue can be found in (Hayawi et al., 2015), (Carbunar et al., 2010).

Source anonymity: Source anonymity is a critical and challenging task where the source node reporting a certain event must be guarded from adversaries to protect it from being captured (like in the case of endangered animals) (Shao et al., 2007). Several solutions have been developed to protect source anonymity, which are based on the utilisation of fake messages and aimed at global adversaries monitoring the whole network traffic as shown below.

(Shao et al., 2007) proposed a scheme called FitProRate to achieve source anonymity. The main idea of their scheme is to use dummy messages such that the sensor nodes maintain a constant traffic pattern in the network. When a real event is detected, the node waits to send the real packet during the same time slots it uses to send the dummy ones. In this way, an attacker can hardly detect the real source of the event. However, this mechanism introduces latency to the whole network. In order to try to decrease the overall latency, the authors suggest the adoption of exponential distribution to determine the time interval to use for sending messages within the network. A similar approach is presented by (Alomair et al., 2010). In addition to the basic source anonymity approach, the authors present further studies to analyse the time intervals for the fake and real packets transmission time to try to defeat the adversary traffic analysis and decrease the overall latency time of the network.

Data collection anonymity: Privacy of data collection in WSNs is achieved through the data aggregation privacy. Data aggregation is concerned with the collection of statistical information about the data collected by the sensor rather than the data itself, to enhance the bandwidth and energy utilisation (He et al., 2008) . Several research papers have been publishing targeting the privacy preservation of data aggregation algorithms.

(Horey et al., 2007) proposed a system called negative survey, which is composed of two protocols: node protocol and base station protocol. The node protocol is used by each node in the network to determine what data will be sent back to the base station. Once the base station receives the data from the sensor nodes, it runs the base station protocol to build the statistical distribution of the data.

Recent work presented by (Li et al., 2015a) proposes a *(α,k)* anonymity based clustering method to ensure data collection privacy.

Communication anonymity: Anonymous communication is concerned with hiding the communication relationship within the traffic flow to make an adversary unable to link two communication parties or link different communications to the same user (Lu et al., 2015). Several research efforts have been presented in the literature to try to protect the anonymity of the communication within a WSN.

(Abuzneid et al., 2015b) presented a scheme called Fortified Anonymous Communication (FAC) to ensure end-end location privacy through deploying temporal privacy and anonymity. The authors claimed that their work is able to ensure sender, receiver and link anonymity, source location privacy, base station privacy and energy preservation. The main idea of their scheme is deploying an anonymity module, which is responsible for the pre-deployment phase, set-up phase and the communication phase where security and privacy measures are considered in all these modules to ensure 100% anonymous communication. This scheme is believed to withstand local, multi-local and global adversaries.

Another protocol for communication anonymity is presented by (Chen et al., 2012) in which the authors propose the Efficient Anonymous Communication (EAC) protocol that guarantees the anonymity of the sender, base station and communication relationship. To ensure anonymous communication, four schemes are deployed: anonymous data sending, anonymous data forwarding, anonymous broadcast and anonymous acknowledgement. Anonymous data sending protects the anonymity of the source node by the deployment of a global anonymous identity

that a source node uses and changes after each message sent. Anonymous data forwarding is concerned with hiding the data forwarding relationship between the neighbouring sensor nodes. Anonymous broadcasting is used to make an adversary unable to distinguish broadcast messages from other messages to hide the location of the base station. Anonymous acknowledgement is used to ensure that there is no loss of messages or transmission error within the communication process anonymously.

Other work on anonymous communication can be found in (Sheu et al., 2008) which presented an anonymous path routing. The protocol uses data encryption and anonyms between the neighbouring sensor nodes and anonyms between the source and destination nodes. Pairwise key data encryption is used to protect the data packets and the anonymous communication protects adversaries from discovering the linkability between the packets.

3.3.2 Pseudonymity

As a privacy service, pseudonymity has not been extensively researched in WSN compared to the other privacy services. Very few papers were published that focus mainly on pseudonymity. To the best of the author's knowledge, pseudonymity in WSN was utilised for user authentication pseudonymity as in (Vaidya et al., 2010) and (He et al., 2015) and location privacy. (Vaidya et al., 2010) proposed an authentication protocol in which the real identity of a user is obfuscated using a hashing-based random pseudonym. The pseudonym is generated using a hashing identifier and XORing it with a random number. The authors claimed that their scheme is resistant to a number of attacks including login replay attacks. (He et al., 2015) suggested the use of the pseudonyms during the authentication and key agreement phase of their temporal-credential-based mutual authentication and key agreement scheme.

In location privacy, pseudonyms can be deployed to as in (Abuzneid et al., 2015a). The authors have proposed the utilisation of disposable pseudonyms to protect the real identity of the sensor nodes and protect their location privacy. The authors claimed that their scheme is able to achieve both source location privacy and sink (base station) location privacy.

3.3.3 Unlinkability

Many papers have addressed unlinkability, in WSN. Although not all of these papers have a primary focus on unlinkability, it was achieved as a by-product from solutions

to other problems. For example, the main target of (Mahmoud, 2012) was to develop a scheme for source node location privacy. This pseudonym-based scheme relies on creating a cloud of fake packets that take different routes and change appearance at each hop, which makes it very hard for an adversary to trace real packets to their source. Privacy protection was achieved by creating an irregularly shaped cloud of fake packets around the source, varying traffic routes and using cryptographic operations to change the appearance of packets at each hop to achieve packet decorrelation (unlinkability).

Another scheme in WSN that indirectly preserved unlinkability was the re-authentication of nodes in a mobile WSN environment (Kim et al., 2011).

Other papers have proposed different techniques for the implementation of the unlinkability service in WSN-based healthcare systems. Some authors emphasise the importance of unlinkability in their work, however they admit their work provides weak unlinkability when adversaries perform traffic analysis. Examples of these papers are (Mare et al., 2011). Other papers rely on random number tags to achieve unlinkability like (Sun et al., 2010). Although their scheme provides unlinkability in cases of medical emergencies, it does not ensure unlinkability in the case of traffic analysis because an adversary can simply monitor all traffic from the patient's Personal Digital Assistant (PDA) and link all messages to this particular patient. Another scheme for unlinkability in emergency call situations is presented by (Liang et al., 2011).

Another category of unlinkability techniques in WSN-based healthcare systems is encryption-based techniques as in (Mare et al., 2011). In their proposed scheme, the authors rely on encrypting the entire packet (the header, payload and Message Authentication Code (MAC)). In addition, their protocols change the header, payload and MAC so that the packets appear pseudorandom and cannot be linked by adversaries to the same sender.

3.3.4 Undetectability

(Pfitzmann & Hansen, 2010) stated that dummy traffic generation can be deployed to achieve undetectability by making the number and length of sent or received messages undetectable by everyone except for the recipients or the senders respectively. The authors of (Shao et al., 2007) have proposed a scheme called Fitted Probabilistic Rate (FitProbRate) in which network-wide dummy packet generation is deployed to achieve privacy. The nodes in the network generate

dummy messages following a predetermined distribution and when a node detects a real event, the real event is transmitted following the same distribution as the dummy traffic. This way, an attacker will hardly distinguish real events from fake ones. Other techniques such as information flooding schemes can also be deployed to achieve undetectability as in the phantom routing proposed by (Ozturk et al., 2004).

In addition to dummy traffic generation, the authors have mentioned that steganography (which can be deployed for audio and video multimedia data), and spread spectrum is popular undetectability techniques.

3.3.5 Unobservability

According to (Shao et al., 2007), unobservability can be achieved by using a mechanism that combines anonymity with dummy traffic such that an adversary cannot tell where the real packets are. Consequently, the mechanisms presented in source anonymity (see Section 3.3.1 which is titled “Anonymity”) that combine fake traffic with anonymity are considered scheme for unobservability. A similar scheme that uses dummy traffic to achieve unobservability can be found in (Yang et al., 2008). In this technique the authors deploy the dummy traffic concept to hide real traffic from adversaries, and thus defeat global adversaries. In addition, the authors suggest the use of sensors that act as proxies to destroy dummy traffic in order to decrease the overall costs of extra traffic. Two schemes are suggested for selecting sensor nodes as proxies: proxy-based filtering scheme and tree-based filtering scheme.

3.4 Survey papers about privacy-enhancing techniques in WSNs

The deployment of privacy and security services in a WSN-based healthcare system should protect all stages of data capturing, communication, processing and storage. In general, privacy services should be applied to the data captured by sensors, data transmitted to the sink, data processed in the sink, data transmitted to the gateway and finally data being processed and stored on remote servers (this stage is out of the scope of this research).

Several survey papers that primarily focused on the general privacy for WSN have been published. Ordered by popularity of citation (by the time this thesis is written), the most popular are: (Li et al., 2010), (Halperin et al., 2008), (Li et al., 2009a), (Al Ameen et al., 2012), (Di & Tsudik, 2010), (Leon et al., 2009), (Islam et al., 2012), (Javadi & Razzaque, 2013), (Oualha & Olivereau, 2011), (Gupta & Chawla, 2012)

and (Wang et al., 2013). Compared to the rest of the papers, as depicted in Figure 3-2, (Li et al., 2010), (Li et al., 2009a) and (Al Ameen et al., 2012) are the most popular survey papers. The papers by (Li et al., 2010), (Halperin et al., 2008), (Al Ameen et al., 2012), (Leon et al., 2009), (Islam et al., 2012), (Javadi & Razzaque, 2013) and (Wang et al., 2013) focus on privacy in healthcare system.

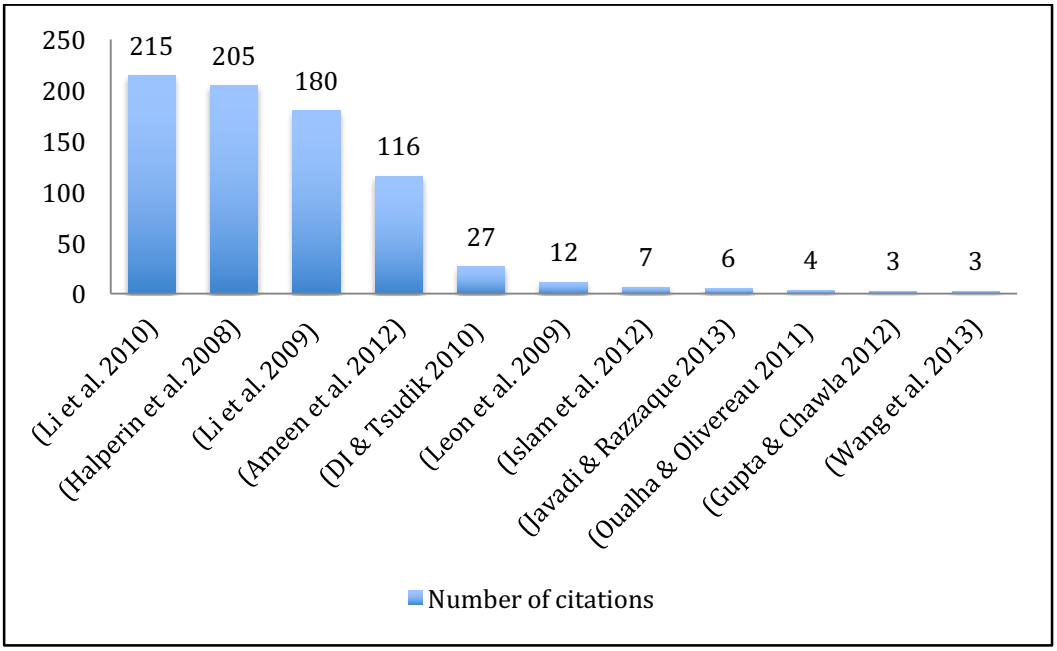


Figure 3-2 Numbers of citations (in May 2015) of survey papers on privacy in WSNs

The most popular survey paper (Li et al., 2010) discusses the security and privacy requirements related to the data storage and transmission in WBAN to ensure the public acceptance of the WBAN technology based on governmental laws. The authors' view of data privacy as an access control problem, where only authorised people should be able to access, view and use the patient-related data. Although their paper is the most popular cited paper, their view of privacy is only limited to access control and discarded other privacy goals.

The next most popular paper by (Halperin et al., 2008) is also dedicated for the healthcare system. In their paper, the authors aimed to discuss the challenges of the balance between the security and privacy design goals of the implanted medical devices (such as pacemakers, drug delivery systems and neurostimulators) and the treatment effectiveness and the medical safety. In this report, the authors define the privacy goals as: device-existence privacy, device-type privacy, specific-device ID privacy, measurement and log privacy and bearer privacy. The authors were able to present a profound analysis of the security and privacy goals of the implanted medical devices, and the impact of achieving these goals such as the tension

between security goals against accessibility, device resources and usability. However, all the discussion presented lacked support of technical means such as algorithms and specific techniques to achieve the goals discussed in this paper.

(Li et al., 2009a) is the only paper among the most cited papers that focused on the privacy preservation techniques for the WSN in general (not focusing on healthcare, but with potential applicability to healthcare). The authors categorise the privacy in WSN as two main categories: data privacy and context privacy. Data privacy is concerned with the privacy of the data collection and the queries issued in a WSN. Context privacy focuses on the contextual information as in the location privacy and temporal privacy. The paper presents a taxonomy tree of the privacy categories and their related techniques. The paper concludes with a comparison between the techniques based on the level of privacy, accuracy, delay time and power consumption. Unlike the previous two survey papers, (Li et al., 2010) and (Halperin et al., 2008), this paper (Li et al., 2009a) includes a discussion of privacy goals and the related techniques. However, this paper has failed to mention other popular privacy goals such as unlinkability and pseudonymity which are advocated by (Pfitzmann & Hansen, 2010). Furthermore, its coverage of privacy protection techniques is dominated by location privacy.

Another survey paper is (Al Ameen et al., 2012), which is concerned with the security and privacy issues in WSN-based healthcare systems, presents a summary of selected WSN-based healthcare projects, security threats, attacks and counter-measures. The paper views the privacy problem as one relating to where the data should be stored, who should access that data and who should be responsible for the data maintenance. Their suggested methods of privacy protection measures are: public awareness of privacy; user identification should be on-need basis and encryption of communication in the WSN. The paper mainly focuses on WSN security and lacks a technical discussion of the different possible privacy preservation techniques.

Another survey paper (Di & Tsudik, 2010) presents a discussion of security and privacy in WSNs. The paper includes three main subsections: WSNs, Vehicular Ad Hoc Networks and disruption-tolerant networks. The paper discusses the security and privacy issues in each of these three categories. However, the paper views privacy as a by-product of security; privacy is not covered in its own right.

The paper by (Leon et al., 2009) describes both the underlying technologies for WSNs and a survey of security and privacy related work in general and in the

healthcare applications in particular. Very few of the related work such as (Luh et al., 2007), (Hwang & Yuan, 2007), (Moncrieff et al., 2008) and (Fidaleo et al., 2004) focus on privacy issues. (Luh et al., 2007) proposes a novel distributed approach to protect Visual Sensor Networks against eavesdropping attacks. They consider protection against privacy attacks as a by-product of their technique. Although the title of the paper (Luh et al., 2007) states clearly that they are proposing a novel distributed privacy paradigm for visual sensor networks, there is no focus on privacy and there is no discrete differentiation between the security and privacy terminologies. (Hwang & Yuan, 2007) propose a technique for privacy preservation based on the use of pseudonyms. The use of the pseudonyms ensures the unlinkability of the entities. (Moncrieff et al., 2008) presents a framework for implementing dynamic and flexible privacy measures in a smart house environment. The appropriate privacy measures are determined based on the surrounding contextual environment to identify the possible privacy risks. Data hiding techniques are applied to video, audio and binary sensor data for privacy preservation while preserving enough information for the observer without invading the patient's privacy. (Fidaleo et al., 2004) proposes a network sensor tapestry, named NeSt, which is a software architecture for a test-bed for surveillance systems that allows the control of the privacy of the multimedia data captured by the system. One component of the architecture is the privacy buffering system, which prevents access to private information or transforms private information by removing personal identification information. The identification of the private data that needs to be obfuscated is achieved by the privacy filters. Although the survey by (Leon et al., 2009) presents a number of other research papers that consider privacy, they do not demonstrate a coherent linkage or profound assessment of the work they discuss. The papers mentioned in the survey are just summarised and no clear linkage was established between them. In addition, the title of the survey paper implies that the authors are focusing on both the security and the privacy in the healthcare. However, those papers that focused on the privacy were related to surveillance or smart house applications. Those papers that focus on healthcare are only related to security.

The paper (Islam et al., 2012) is concerned with the security and privacy of WSN in smart homes. This paper adopts the exact same privacy taxonomy presented by (Li et al., 2009a). The paper references different techniques related to the privacy categories mentioned. However, no details or discussion of the privacy techniques are presented.

Following (Li et al., 2010), the paper by (Javadi & Razzaque, 2013) is concerned with security and privacy issues for WBANs. In addition, both papers have the same view of data privacy as being an access control problem where only authorised personnel should be able to access the patient related information. The authors of (Javadi & Razzaque, 2013) clearly view privacy as a by-product of security. They mainly rely on security related solutions such as data confidentiality, data access control, accountability, revocability, non-repudiation, policy requirement and public awareness to ensure the privacy of the data. Their paper clearly lacks focus on privacy solutions.

(Oualha & Olivereau, 2011) focus on the sensor and data privacy for WSN for industrial applications. (Oualha & Olivereau, 2011) adopt a similar outline to their paper as that presented by (Li et al., 2009a). In (Oualha & Olivereau, 2011) the taxonomy of privacy preservation is sub divided into sensor privacy and data privacy. Sensor privacy is concerned with the privacy of the sensor related data such as sensor identity, location, battery, time and CPU, while the data privacy is concerned with the privacy of the data captured by the sensors. The authors conclude their paper with a very brief comparison between the approaches discussed in the data privacy protection schemes. Similar to (Li et al., 2009a), their paper did not consider all privacy goals such as unlinkability and mentioned location privacy very briefly.

Following (Li et al., 2009a), the authors of (Gupta & Chawla, 2012) adopt the same categorisation of privacy preservation techniques. They even present a similar comparison as a conclusion for their paper. In addition, many of the techniques they mentioned were already mentioned in (Li et al., 2009a).

Finally, the paper by (Wang et al., 2013) focuses on the security and privacy issues for patient related data in WSN-based healthcare systems. The paper views the privacy requirements in e-health as anonymity and unlinkability requirements. It does not provide any details of the anonymity and unlinkability techniques that can be used to ensure privacy.

3.5 Deployment of WMSNs in healthcare

The previous sections discussed privacy in WSNs in general. However, the presence of multimedia data captured by audio and video sensors add more challenges to privacy. Consequently, this section will discuss the deployment of

WMSNs in healthcare and the next section will focus on privacy for WSNs with multimedia content (WMSNs).

WMSNs are networks of sensors that collect different types of media such as audio, video, still images and scalar data (Akyildiz et al., 2007). The use of WMSNs have stretched and enhanced the applications of the traditional WSNs (Akyildiz et al., 2007). Many papers have mentioned the deployment of WMSNs in healthcare systems as one of the many possible applications of WMSNs as in (Almalkawi et al., 2010) and (Akyildiz et al., 2007). However, these papers focused on the general WMSN algorithms, architectures, challenges and communication protocols with no in depth analysis for the deployment of WMSNs in the healthcare systems. Only few papers attempted to offer solutions for the deployment of WMSNs in healthcare. For example, (Hamid et al., 2013) focuses on overcoming the challenge of the high bandwidth demand in live telemedicine applications using video and audio streaming.

The in-house healthcare application scenario for WMSNs, which is illustrated in Figure 3-3, is based on scenarios presented in (Alemdar & Ersoy, 2010). They outline how the deployment of different sensors can be used to monitor patients. The figure shows a possible view of an apartment where patients with different needs (for example: chronically ill, handicapped patients or elderly) can live. Each patient will wear or have implanted the appropriate sensors required to monitor his/her health condition. For example, oxygen saturation, heart rate, body temperature and blood pressure sensors can be used to monitor a chronically ill patient (Alemdar & Ersoy, 2010). Other sensors can also be used to monitor the surrounding environment. They may include humidity, pressure sensors, temperature, and Radio-Frequency IDentification (RFID) sensors.

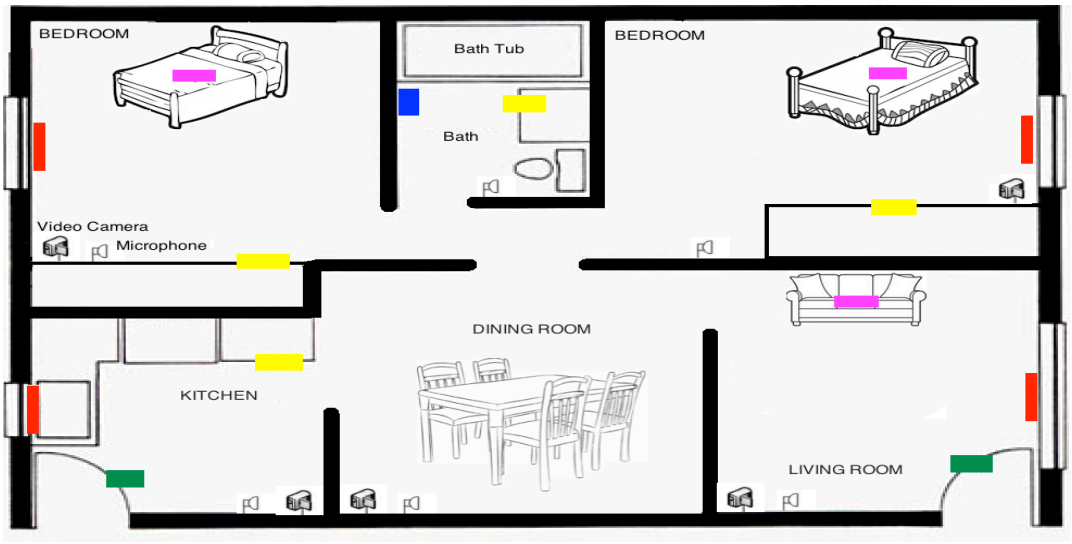


Figure 3-3 A view of a house equipped with a WMSN for healthcare applications. The colour coding is: green for door sensors, red for window sensors, pink for pressure sensors, blue for humidity sensors and yellow for RFID sensors.

As depicted in Figure 3-3, doors and windows sensors can be used to detect when doors and windows are opened or closed. Pressure sensors can be attached to sofas or beds to detect if someone is sitting on them. Humidity sensors can be used in bathrooms to monitor the humidity level. RFID sensors can be attached to the commonly used household items such as medicine cabinet, refrigerator, bedroom closet or any other items to detect when a patient has used them. Microphones and video cameras can be used to monitor the patient, for example to hear their requests for help, or monitor their gait or posture to detect if they stumble into an obstacle or fall down.

A seamless healthcare environment can be built based on a multi-tier WMSN architecture with diverse types of sensors collecting different types of data (scalar, audio and video). Once the sensors capture the data, it is sent to a nearby data manager or gateway (computer or a mobile phone). The data manager sends the data to a remote server (base station) for further analysis or storage, and to the caregiver for the sake of the patient's monitoring. In case of emergency, such as fluctuations in the readings of the sensors or a patient falling down, alerts can be sent to the nearest caregiver for immediate actions or automatic phone calls can be placed for healthcare professionals, ambulances, hospitals or emergency centres.

3.6 Privacy in WMSN-based healthcare systems

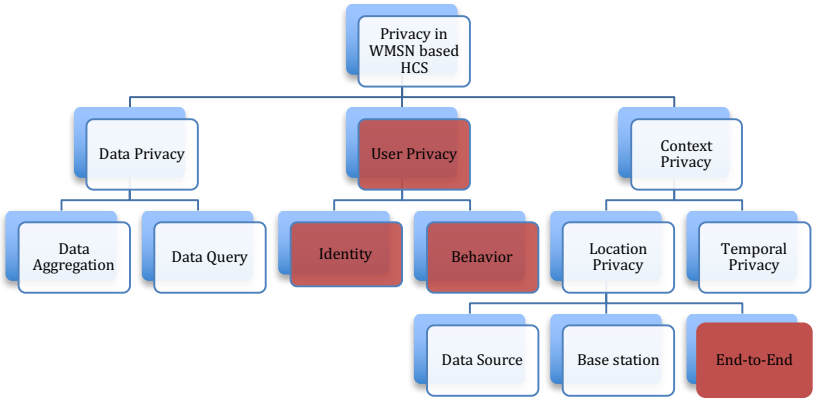


Figure 3-4 Taxonomy for privacy services in a WMSN-based healthcare system (adapted from (Li et al., 2009a))

As mentioned in the previous section, different types of sensors are deployed in WMSN-based healthcare systems including video and audio sensors. Due to the sensitive nature of the images which may reveal critical information about the people being monitored, not only their identities but also their behavioural patterns, additional privacy services are required to maintain the privacy of the patients (Winkler & Rinner, 2014). Accordingly, privacy and security aspects in WSNs with multimedia content need to consider four important aspects: the privacy/security of data, node, network and user identity. The first three aspects are already covered in the taxonomy tree introduced by (Li et al., 2009a). However, since this tree is only concerned with the scalar data, the user privacy (which can be threatened by the use of multimedia data) needs to be considered (Winkler & Rinner, 2014). Figure 3-4 shows an updated tree of (Li et al., 2009a) with extra (red) nodes that need to be put into consideration in WMSN-based healthcare systems.

End-to-end location privacy refers to the protection of the privacy of the location of the user that might be threatened in multimedia data. The video and audio analysis of the multimedia data might reveal information about the location of the user, which might be deployed by adversaries to track or locate a user. End-to-end location privacy services should ensure the protection of the location of both the sender and the receiver of the multimedia content.

The user privacy aspect refers to protecting the privacy of the user whose video and/or audio data is captured and transmitted in the WMSN. The amount of information extracted from the multimedia data can be categorized into two levels: primary level and secondary level (Adams & Sasse, 2001). Primary-level

information refers to the core information captured by the multimedia data such as the medical information depicted in a video-based doctor-patient discussion (Adams & Sasse, 2001). Privacy threats to primary level information may threaten to reveal the identity of the patient. Secondary-level information refer to the interpretative information characteristics (for example social or psychological), that can be derived from the captured multimedia data such as information learned by studying the body language of the patient in captured videos (Adams & Sasse, 2001). In addition, secondary information can lead to the identification of where (location), when (time) and what (actions) (Saini et al., 2014). The analysis of secondary information may threaten the privacy of the user by learning the behaviour of the user and revealing more information about his/her personality and behaviour (behavioural user privacy). In addition, video and/or audio analysis can threaten the recognition of the identity of the patient (identity user privacy).

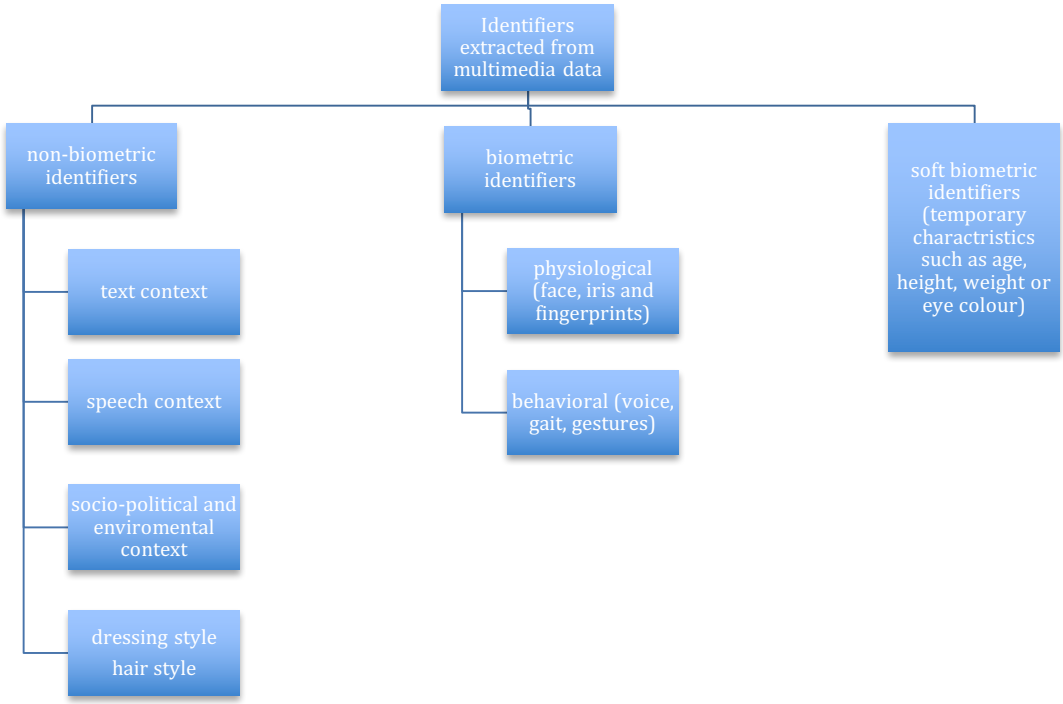


Figure 3-5 Taxonomy of identifiers extracted from multimedia content (Ribaric et al., 2016)

Figure 3-5 depicts the taxonomy of the different identifiers that can be extracted from multimedia content and assist in the identification of the individuals perceptible in the multimedia data. (Ribaric et al., 2016) have suggested different mechanisms depending on the type of identifiers to protect the identity of the individual; for example in biometric identifiers, different techniques such as face region blurring or pixilation.

Different methods have been proposed in the literature to conceal the privacy-sensitive information contained within the multimedia data. Privacy protection for image or video data is often achieved using computer vision techniques to detect sensitive image regions and obfuscate primary identifiers such as a human face (Winkler & Rinner, 2014). (Winkler & Rinner, 2014) states that there are two main approaches for privacy protection of visual data: object-based approaches and global approaches. Object-based approaches are based on the removal or distortion of sensitive regions of the images such as person or faces. Global approaches apply protective operations to the whole image (such as downsampling, blurring or mosaicing). (Winkler & Rinner, 2014) categorized visual privacy techniques into:

1. Detection of sensitive regions: The basic idea of this technique is to detect sensitive image regions (such as a human face or a licence plate number). The privacy of whole video is at risk if the system fails to detect a sensitive region in one of the sequence of images in a video (Winkler & Rinner, 2014). In the networked sensor tapestry (NeST) architecture (Fidaleo et al., 2004), privacy is maintained using three main components: privacy buffers (prevent access to sensitive information or transforms the data to remove sensitive information), privacy filters (determine whether the data is private or not) and privacy grammar (allows end users to generate their own privacy definitions based on low-level and high-level feature of data).
2. Blanking: The basic idea of blanking is removing sensitive regions of the images and leaving blank areas (Winkler & Rinner, 2014). However, blanking affects the usefulness of the data since only the presence and the surroundings of a person can be detected but his/her behaviour is lost (Winkler & Rinner, 2014). For example, in (Wickramasuriya et al., 2004), sensor technology is fused with video surveillance to construct a privacy protecting framework. An object tracker image processing identifies subjects entering the field of view of a camera and the RFID tags warn by the subject determines the level of privacy that will be applied to the image.
3. Obfuscation and scrambling: The main idea of obfuscation is to reduce the level of the sensitive information in the images using different techniques such as blurring, mosaicing or pixilation (Winkler & Rinner, 2014). Obfuscation techniques were adopted by

(Wickramasuriya et al., 2004) and (Gross et al., 2006) for privacy protection.

4. Abstraction: The main idea of abstraction is replacing sensitive regions of images by virtual entities such as avatar, stick figure or silhouettes (Winkler & Rinner, 2014). For example, (Lo et al., 2005) presents UbiSense Distributed Multimedia Sensor Network which is an automated homecare monitoring of the elderly based on movement tracking and posture analysis. UbiSense utilizes video sensors installed in the environment, body sensors and RFIDs. The captured images are immediately processed in the camera to retain abstract information about the monitored scene. However, individuals may still be identified using the gait and posture retained information.
5. Encryption: Encryption can be used to hide sensitive regions of images which can only be viewed using a decryption key (Winkler & Rinner, 2014).
6. Multiple privacy levels: Multiple privacy levels are required in cases when one video stream contains diverse level of information which might require applying different techniques to the sensitive regions of the images (Winkler & Rinner, 2014).

3.7 Summary

This chapter presented an overview of the basic terminologies related to privacy, a review of the definitions of the basic privacy terminologies, a survey of the general WSN privacy-preserving techniques for anonymity, pseudonymity, unlinkability, undetectability and unobservability. In addition, this chapter has reviewed the popular survey papers, which focused on the privacy of WSNs. Finally, the chapter discussed the deployment and privacy of WMSNs in healthcare systems.

The deployment of WMSNs in healthcare systems is constrained by many factors due to the nature and limitation of the WMSN, such as power consumption, high bandwidth demand and quality of service (Akyildiz et al., 2007). Furthermore, time and budget impose a strong constraint when designing a privacy aware system (Deng et al., 2011). Consequently, only important privacy services should be considered and other less important should be discarded or become optional.

Based on the literature survey conducted in this chapter and to the best of the author's knowledge, there has not been a formal privacy threat analysis methodology applied to a WMSN-based healthcare sub-system (from the tier of the

sensors to the tier of the base station) to assess the significant privacy services that should be present in order to be accepted by patients and governments in real life (addressed in Chapter 4 which is titled “Identification of Focal Privacy Threats”). In addition, based on the literature survey and to the best of the author’s knowledge, there is no privacy-aware ant-based routing protocol for WMSN (addressed in Chapter 5, which is entitled “Privacy-Aware Ant Routing Algorithm for WMSNs”). The findings from the assessment of the privacy-awareness resulting from the deployment of the privacy-enhancing countermeasures, and findings from the assessment of their associated computation, communication and storage overhead is another contribution to knowledge claimed by this thesis (addressed in Chapter 6 which is titled “Privacy assessment methodology”, Chapter 7, which is titled “Analysis of Overheads Due to Privacy-Enhancement of the Ant Routing Algorithm” and Chapter 8 which is titled “Assessment of the Enhancement of Anonymity, Unlinkability and Location Privacy Due to Fake Traffic”).

In the next chapter, a privacy threat analysis methodology will be applied to a basic WMSN-based healthcare sub-system to elicit the significant privacy services, which should be included in this healthcare sub-system.

Chapter 4 Identification of Focal Privacy Threats

4.1 Introduction

"Privacy-by-design" is a privacy engineering methodology that is used to elicit the privacy threats to a system during design time to avoid the challenging implementation of the privacy services after the software engineering process (Wuyts et al., 2014). Several methods have been proposed for the identification of security threats compared to little research focusing on the identification of the privacy threats (Danezis et al., 2015). One of the few privacy threat analysis methodologies is the LINDDUN (Deng et al., 2011) (Wuyts et al., 2014) methodology, which proposes a systematic approach for the identification of privacy threats, based on the use of data flow diagrams and threat trees.

The aim of this chapter is to apply a privacy threat analysis methodology, the LINDDUN (Deng et al., 2011) (Wuyts et al., 2014), to a WMSN-based healthcare sub-system to discover the significant privacy services that should be present in this healthcare subsystem. The proposed sub-system will include medical, environmental, audio and video sensors and will focus on the part of the healthcare system, which is from the data capture part from the sensors until the data arrives at the base station

The next section depicts a brief survey of a number of privacy threat methodologies. Next, the general outline of the LINDDUN privacy threat analysis methodology is discussed. This is followed by the creation of a data flow diagram for a WMSN-based healthcare sub-system and then the data flow elements are mapped to the LINDDUN privacy threats. Finally, a discussion of the mapping results and summary of the chapter is presented.

4.2 Privacy threat analysis methodologies

"Privacy-by-design" refers to taking into consideration privacy protection safeguards during the stages of the engineering process of a system, from the early stages until the operation stage (Danezis et al., 2015). However, since privacy is a complex, multifaceted notion that is normally not the main requirement of the system being designed and it might conflict with the system requirements (functional or non-functional), the privacy goals must be well defined and evaluated (Danezis et al., 2015). Privacy impact analysis or privacy risk analysis is used to discover the privacy objectives of a system and from a technical point of view they are basically about identifying the stakeholders of the system, then identifying the

risks (putting into consideration the stakeholders of the system), then identifying the possible solutions and recommendations for the risks, then implementing the solutions and recommendations and finally performing audits and reviews measures (Danezis et al., 2015).

Privacy threat models are used to discover the privacy requirements (or goals) of the system being developed and identify the weaknesses of the architectural design (Wuyts et al., 2014). Several threat analysis methodologies have been developed to systematically elicit the security and privacy requirements. These methodologies differ depending on the approach they adopt to identify the privacy threats. Some methodologies such as those presented in (Luna et al., 2012), (Deng et al., 2011) and (Wuyts et al., 2014) propose systematic steps starting with the documentation of the system under analysis in the form of a data flow diagram. The data flow diagram elements are then mapped to privacy threats (linkability, identifiability, non-repudiation, detectability, disclosure of information, unawareness, and non-compliance in (Wuyts et al., 2014); and linkability, unawareness and intervenability in (Luna et al., 2012). The privacy threats are later analysed using threat trees as in (Wuyts et al., 2014) or attack trees as in (Luna et al., 2012). Unlike (Wuyts et al., 2014), the authors of (Luna et al., 2012) quantify the overall privacy attack to the system based on the quantitative evaluation of the individual attacks in the attack trees developed in earlier steps (risk-based quantification).

Others like (Preneel & Ikonomou, 2014), propose a semi-automated problem-based privacy threat identification methodology. Automated privacy threat graphs are used based on the requirements of the designated system. Their proposed methodology is independent of specific privacy goals. Although their methodology is semi-automated, they use high-level privacy requirements and they do not provide detailed privacy knowledge for detailed analysis unlike (Wuyts et al., 2014) and (Luna et al., 2012). Other methodologies such as (Kalloniatis et al., 2008) view the privacy requirements as an organisational goal and analyse where these goals are best implemented in the system. A mapping between the privacy requirements and the related privacy techniques is used to determine the specific techniques that will be adopted.

4.3 LINDDUN-based privacy threat analysis methodology

The LINDDUN methodology offers a general-purpose systematic set of methods that is not dedicated for a certain field of application, which implies that it can be applied to WMSN-based healthcare systems. The LINDDUN methodology has been

applied to the healthcare field as in the project (Kohlmayer et al., 2014) in which both the LINDDUN privacy threat analysis and the STRIDE security threat analysis were applied to discover the privacy and security threats for the private and secure accessing and sharing of personalised medical data. The analysis conducted by (Wuyts et al., 2012) to a healthcare related patient community case study involves patients, nurses and researchers. However, their analysis is different from that in this research work because theirs is based on the part of the healthcare system involving the processing of the electronic health records.

In this section, the LINDDUN (Deng et al., 2011) (Wuyts et al., 2014) methodology will be applied to a WMSN-based healthcare sub-system in order to determine the privacy services that must be included in the sub-system. First the privacy terminologies and the basic definitions of the privacy services are explained. Next, a general presentation of the main steps of the LINDDUN methodology is outlined. This is followed by a discussion of the data flow diagram representing the WMSN-based healthcare sub-system and a detailed explanation of the different components of the data flow. Finally, the data flow elements are mapped to the LINDDUN privacy threats.

4.3.1 More Privacy terminologies

The authors of the LINDDUN methodology have chosen the privacy properties included in their analysis based on the privacy terminology proposed by (Pfitzmann & Hansen, 2010) namely unlinkability, anonymity/pseudonymity and, undetectability & unobservability (see Section 3.3). In addition, the authors have included plausible deniability privacy property, confidentiality (although it is a security property but the authors believe that it is an important building block to the privacy services anonymity and unlinkability), content awareness and policy and consent compliance.

Besides the definitions of the privacy terminologies given in the previous chapter (see Section 3.3), the extra definitions of the terminologies used by the LINDDUN authors (plausible deniability, confidentiality, content awareness and policy and consent compliance) are explained below.

Plausible deniability: The authors of LINDDUN (Deng et al., 2011) define plausible deniability as the capability of a person to deny performing a certain action that others cannot assure or deny that this action was performed. From the communication point of view, the authors refer to plausible deniability as "*an instance of communication between computer systems leaves behind no*

unequivocal evidence of its having taken place. Features of communications protocols that were seen as defects from the standpoint of non-repudiation can be seen as benefits from the standpoint of this problem, which is called plausible deniability". Possible examples of plausible deniability are a person denying performing an off-the-record conversation, or denying that a certain encrypted file exists (Deng et al., 2011).

Confidentiality: Confidentiality aims at obfuscating the content of the data such as hiding the data content of an encrypted email or protecting the content of a database of sensitive data (Deng et al., 2011). The authors of (Deng et al., 2011) believe that although confidentiality is a security property, it is required as a building block for some privacy services such as anonymity and unlinkability, thus it can be viewed as an important privacy objective.

Content awareness: Similar to confidentiality, content awareness and, policy and consent are not basic privacy properties but they are considered important privacy objectives that must be put into consideration in privacy threat analysis. The aim of content awareness is to ensure that the users of the Web 2.0 technologies who grant the service providers personal information and lose control over their information should be aware that the minimum mandatory personal information should be released for the performance of the application it was released to (Deng et al., 2011). However, in some cases it is important that the information about a person is up to date and correct to avoid serious consequences such as in e-health applications (e.g. a doctor not mentioning in the e-health records that a patient is diabetic can result in life-threatening consequences) (Deng et al., 2011).

Policy and consent: Based on the EU Directive 95/46/EC policy and consent can be defined as: "*Controller shall mean the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data." "The data subject's consent shall mean any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed*" (Parliament, 1995) (Deng et al., 2011). A policy sets the general rules for the protection of the data. Similarity, a consent also specifies rule for the protection of data but these rules are determined and are related to the user him/herself (Deng et al., 2011). The policy and consent property is also related to the legislation concerned with the protection of data such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the Data

Protection Directive 95/46/EC in Europe (Deng et al., 2011). The authors of the LINDDUN (Deng et al., 2011) claim that a small number of technical solutions were developed that have included the policy and consent property in their design.

4.3.2 General outline of the LINDDUN methodology

The LINDDUN privacy threat analysis methodology is analogous to and inspired from the well-established security threat analysis STRIDE developed by Microsoft to discover security threats and build security use cases (Deng et al., 2011). LINDDUN is divided into six main steps:

Step 1: Modelling Data Flow Diagram: A Data Flow Diagram (DFD) is a structured and a graphical diagram used to describe the flow of information in a system among the external entities, data flows, data stores and processing nodes (Wuyts et al., 2014). The DFD phase is a critical part of the analysis because it depicts the flow of the information, where the information will be stored and where it will be processed. A flaw in the representation of the DFD will cause the whole analysis to be inaccurate. In addition, the level of detail of the DFD affects the level of detail of the threats discovered by the privacy threat analysis (Wuyts et al., 2014).

Step 2: Mapping the DFD elements to privacy threats: Each element of the DFD is checked to see which privacy threats can have an effect on this element. The privacy threats (opposite to the privacy properties) that the LINDDUN methodology considers are: **L**inkability (threat to unlinkability), **I**dentifiability (threat to anonymity and pseudonymity), **N**on-repudiation (threat to plausible deniability), **D**etectability (threat to undetectability and unobservability), **D**isclosure of information (threat to confidentiality), **U**nawareness (threat to content awareness) and **N**on-compliance (threat to policy and consent compliance) (thus the acronym LINDDUN)) (Wuyts et al., 2014).

Step 3: Elicitation of privacy threats: This stage is sub-divided into three stages, privacy threats are refined using privacy threat trees, assumptions are documented and threats are documented using threat templates (Wuyts et al., 2014).

Step 4: Threat prioritisation: Limitations of time and budget make it important to select the most significant privacy threats to be considered in the system design. Risk assessment techniques are utilised in this stage to assess the likelihood of the occurrence of a certain attack and the impact of this attack. Threats are prioritised according to their risk assessment values and those threats with high risks are important to consider in the system design (Wuyts et al., 2014).

Step 5: Discovery of privacy requirements: In this stage, the identified privacy threats that must be considered in the system design are mapped to their corresponding privacy requirements (Wuyts et al., 2014).

Step 6: Selection of privacy solutions: In the final stage, privacy mitigation techniques are selected to implement the privacy requirements identified in step 5 (Wuyts et al., 2014).

In this research work, only steps 1 and 2 are applied to the WMSN-based healthcare sub-system and steps 3 to 6 are left for future work. This is because in this research work, the aim of applying the LINDDUN methodology is to elicit the significant privacy threats to the WMSN-based healthcare sub-system and choose some of these threats to focus on, which is carried out in steps 1 and 2 of the LINDDUN methodology. Steps 3 to 6 mainly focus on creating privacy threat trees and misuse cases, stating assumptions, performing risk assessment and discovering privacy requirement and solutions based on the privacy threat trees and misuse cases. Consequently, steps 3 to 6 are left as an extension for this research work in the future.

4.3.3 DFD of WMSN-based healthcare sub-system

According to the LINDDUN methodology, drawing a DFD is considered an important step of the analysis. In order to create the DFD, a survey was conducted to determine the main components and the general flow of data in a typical healthcare sub-system as in the research by (Malan et al., 2004), (Chakravorty, 2006), (Ganti et al., 2006), (Wood et al., 2008), (Ko et al., 2010a), (Mitra et al., 2012), (Rofouei et al., 2011), (Nabar et al., 2011), (Garverick et al., 2011), (Zhang et al., 2011), (Hu et al., 2011), (Yuce, 2010), (López et al., 2010) and (Wood et al., 2008). For the sake of illustration, an abstract level of the DFD (level 0) is depicted in Figure 4-1.

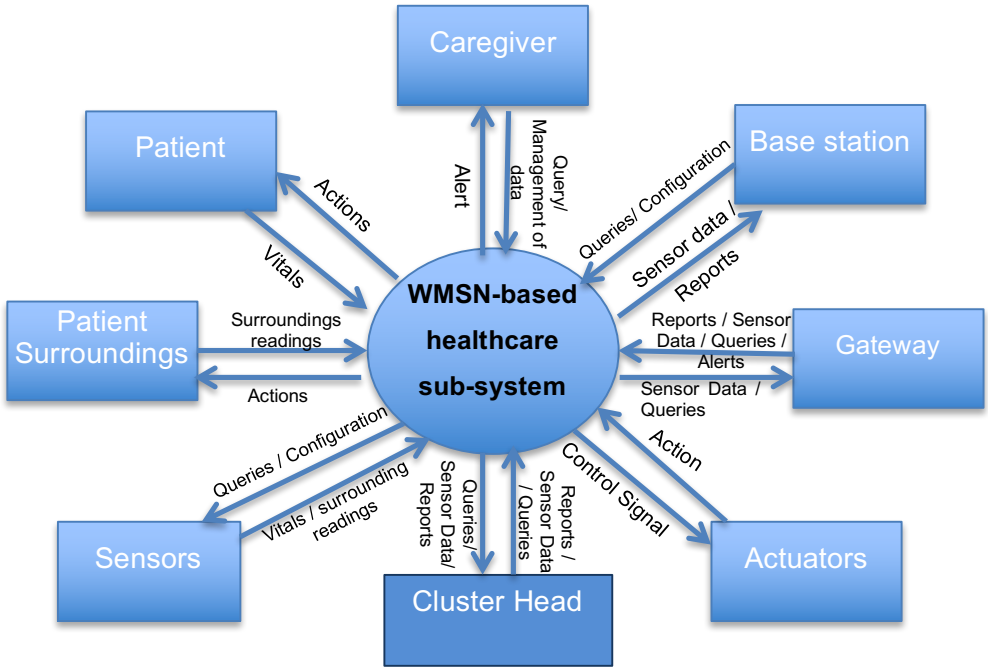


Figure 4-1 Level 0 DFD for the WMSN-based healthcare sub-system

Below is an explanation of the different components of the DFD elements (entities, data flows, data stores and processes)

Entities: Figure 4-1 is an abstract contextual diagram of the WMSN-based healthcare sub-system. This figure shows that there are eight external entities connected to the sub-system namely: patient, patient surroundings, sensors, actuators, gateway, cluster head, base station and caregiver. The role of each entity is as follows:

- Patient: The patient's vitals are captured using medical sensors (wearable on-body sensors or implanted in-body sensors) to capture different vitals depending on the types of sensors and on the medical condition of the patient. In addition, video and audio sensors are used to capture the images and sounds of the patient to assess their gaits or detect falls and hear their calls for help. All these inputs are entered into the WMSN-based healthcare system for processing. In case of abnormal readings, detection of a fall of a patient or call of help, alerts are sent to the caregiver to take appropriate action and alerts are sent to the actuators to take action (e.g. in case of low insulin level, insulin pumps are used to pump insulin into the patient's body). In case of normal readings, the readings are collected and analysed and then sent to the base station.
- Patient's surroundings: The surroundings of the patient are monitored using environmental sensors to detect changes in the surroundings that might

affect the wellbeing of a patient, like changes in temperature or pressure or increase in the level of dust. In case of abnormal readings of the environmental sensors, alerts are sent to the caregivers to take action and control signals are sent to the actuators (e.g. a control signal can be sent to an oxygen pump to increase the oxygen level, or to the air conditioning unit to lower or increase the room temperature). Normal readings are analysed and reports are sent to the base station.

- Sensors: Different types of sensors can be utilised in a WMSN-based healthcare system to monitor the vitals, the status (e.g. walking normally or fallen on the floor) and the surroundings of the patient. The types of sensors will depend on the condition of the patient and what a caregiver or a doctor (or even a trainer for athletes) needs to monitor. Pressure, temperature and humidity sensors (or other types of sensors depending on the surroundings and the patient's health) can be used to monitor the surroundings of the patient. Video and audio sensors can be used to capture the motion or the sounds of the patient to detect abnormal movements or calls for help. Other types of sensors can be used, such as Radio-Frequency Identification (RFID), to monitor how patients interact with their surroundings or detect changes in the behavior of patients (such as changes in the frequency of opening and closing cabinets, or the opening and closing of doors or windows). In general, sensors should capture information (from patients or surroundings) and might receive queries or configuration adjustments. Queries can be issued by caregivers or the base station to collect more information about the heath status of the patient. Configuration adjustments can be received from the gateway to update the settings of the sensors (for example update the sampling rate based on the health status of the patient).
- Actuators: Actuators are devices that receive control signals as an input and change it to a physical action (such as control the temperature or the heat distribution, or dispense medicine through a medical device worn by a patient) (Khan et al., 2012). In this healthcare sub-system, actuators receive control signals based on the analysis of the captured data and output control actions (e.g. turning on heaters or air conditioning units, or instructing an insulin pump).
- Gateway: Gateways or data managers collect data from the different sensors and they can issue queries to sensors, locally process sensor data and send reports to the cluster heads or store them in data stores. In

addition, gateways can also alert caregivers or trigger actuators in case of abnormal sensor readings or analysis.

- Cluster head: Cluster heads collet data and reports from the different gateways and re-route the data to the base station. Cluster heads can also accept queries and re-route them to the designated gateway.
- Base station: The base station receives sensor data aggregated at the cluster heads, or receives reports of data analysis from cluster heads to perform further analysis. It can also send the analysis results or issue queries or configuration data to the cluster heads.
- Caregiver: Caregivers can issue queries for gateways and receive alerts from gateways in case or abnormal readings.

Figure 4-2 depicts the level 1 DFD representing a more detailed view of the WMSN-based healthcare sub-system compared to the abstract view shown in Figure 4-1. Basically, this DFD is made up of eight entities (depicted as round circles labelled a to h): Patient, patient surroundings, sensor, data manager/gateway, cluster head, caregiver, actuator and base station. Processes responsible for the processing of data are depicted as numbered rectangles. Sensors, gateways, cluster heads and the base station have data stores that are responsible for the storage of data related to sensors, data managers, cluster heads and base station respectively. The arrows show the direction of the data movement between the entities, processes and data stores. Each arrow has a label that depicts what data is being moved.

Data Flow: The general flow of the data in the WMSN-based healthcare sub-system is depicted in Figure 4-2. The patient's vitals, video and audio data are captured (by the Sensors entity (labelled c)) and the data about the surrounding environment of the patient is captured (by the Sensors entity (labelled c)). Data collected by a sensor can be stored in the sensor data store. The sensed data is filtered, changed to digital form, processed for analysis and sent to the data manager/gateway. In case of emergency, where the analysis detects improper levels for vital signs such as blood sugar or blood pressure, the gateway sends an alert to the caregiver and control signals to the actuators to take action such as activate the insulin pump. In case of normal data readings, the gateway can store the data captured from the different sensors and can send the data to the base station via the cluster head to perform advanced analysis. The gateway is able to relay queries from the caregiver or issue its own queries to the sensor nodes connected to it. In addition, the gateway is able to send sensor configuration control

signals to change the configurations of the sensors (e.g. change the rate of the data sampling depending on the condition of the patient).

Data Stores: The WMSN-based healthcare sub system, depicted in Figure 4-2 contains four data stores. One data store called “D1 Sensors” to store the data captured by the sensors and store the configuration data of the sensors. Another data store, called “D3 Data Manager”, is responsible for storing the data and reports that belong to the gateway. The data store labelled “D4 cluster head” is used to store data and reports at the cluster head level. The data store labelled “D2 base station” is used to store data and reports at the base station level.

Processes: The processes in the DFD are responsible for performing tasks on the data as it flows around the DFD. The processes in the DFD depicted in Figure 4-2 are numbered 1 to 20. Processes number 1 and 2 are responsible for capturing of the data from the patient and the patient’s surroundings respectively. Processes numbers 3 to 6 are responsible for processing the data captured by the sensors, namely: filtering, performing analogue-to-digital conversion, processing the signals and transmitting the data to a gateway, respectively. Process number 7 is responsible for storing and retrieving the data captured by the sensors. It is also responsible for the storage and retrieval of the configuration data of the sensors (e.g. the sampling rate) that might be updated by the gateway, cluster head or the base station, depending on the medical condition of the patient and depending on how much information needs to be retrieved, to accurately assess the condition of the patient and his/her surrounding environment. Process number 8 is responsible for the management of the queries issued from the base station to the sensors through the gateway and the cluster head. This process is also responsible for the update of the sensor configuration, for example in case a change of sensor sampling rate is required. Process number 9 is responsible for performing data analysis on the data aggregated at the gateway from all the sensors. Data reports generated after the analysis of the data at the gateway are sent to the base station via the cluster head. In case of abnormal results from the data analysis, an alert is issued (process number 14) to alert the caregiver that there is an emergency and instruct the actuator to take action (process number 15) (in case an action can be done by the actuator). Process number 10 is responsible for the storage and retrieval of the data reports to and from the base station data store. Process number 11 is responsible for performing advanced data analysis at the base station (for example advanced data analysis which includes different patients in different geographical locations, for

scientific reasons). Process number 12 is responsible for issuing queries and for the management of sensors by the caregiver, in case a caregiver needs to collect more information about the patient's condition or surrounding environment. Process number 13 is responsible for the storage and the retrieval of the data reports to and from the gateway data store. Process number 16 is responsible for the management of the queries and the sensor configurations at the level of the gateway. Process number 17 is responsible for transmitting data and reports from the gateway to the cluster head. Process number 18 is responsible for sending queries from the base station to the cluster head. Process number 19 is responsible for sending the sensor data and the reports from the cluster head to the base station. Process number 20 is responsible for the storage and retrieval of the data and reports at the cluster head level.

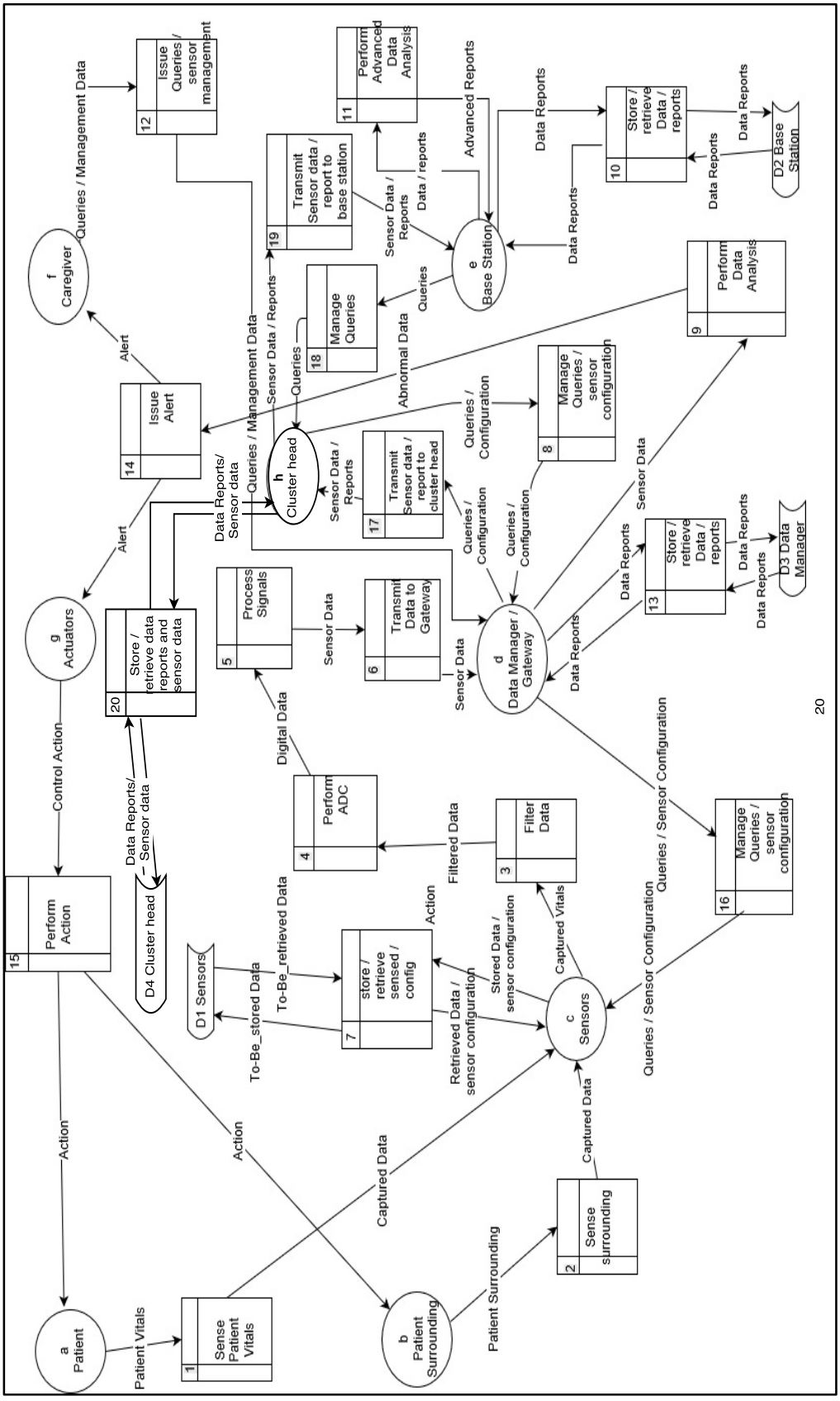


Figure 4-2 Level 1 DFD for a WMSN-based healthcare sub-system

4.3.4 Privacy threats to DFD mapping

According to the LINDDUN privacy threat methodology, after the DFD is created, all elements of the DFD must be mapped against the privacy threats in order to determine which elements are affected by those privacy threats. This mapping is depicted in Table 4-1. This mapping is performed based on the definition of the privacy threats (i.e. depends on the goal of the privacy properties). The mapping of the privacy threats of the LINDDUN is justified in their original work in (Wuyts et al., 2014).

Table 4-1 DFD elements mapped against privacy threats (Wuyts et al., 2014)

Privacy Property	Privacy Threat	Entity	Dataflow	Data Store	Process
Unlinkability	Linkability	X	X	X	X
Anonymity/ Pseudonymity	Identifiability	X	X	X	X
Plausible Deniability	Non-repudiation		X	X	X
Undetectability and Unobservability	Detectability		X	X	X
Confidentiality	Information Disclosure		X	X	X
Content Awareness	Content unawareness	X			
Policy and Consent Compliance	Policy/Consent Noncompliance		X	X	X

4.4 Discussion

Upon examining Table 4-1, it can be noticed that not all threats have direct impact on all DFD elements. For example, non-repudiation affects data flows, data stores and processes; and context unawareness only has a threat on entities. However, other threats may affect all DFD elements. For example, linkability and identifiability impose threats on all DFD elements, which make them expose the healthcare sub-system to a higher risk compared to threats, which affect less DFD elements.

Furthermore, since the privacy of personal information is very crucial and legally protected in many countries (as depicted in the previous chapter), the identifiability of a patient (as opposed to anonymity) is a serious privacy threat that must be considered in healthcare systems. The communication of multimedia data in a WMSN-based healthcare sub-system may disclose information about the identity and the location of a patient (see Section 3.6). However, the LINDDUN authors have not considered the location privacy and the multimedia data privacy in their analysis. A possible reason for not considering location and multimedia data privacy might be because the LINDDUN authors adopted their privacy goals from (Pfitzmann & Hansen, 2010) which discussed the general privacy terminologies without focusing on multimedia-related privacy threats. Consequently, the privacy threat analysis conducted in this research work can be divided into two stages: stage 1: applying the LINDDUN methodology to the WMSN-based healthcare sub-system (see Section 4.3) and stage 2: adding multimedia-related privacy properties (location privacy and multimedia data privacy) to the privacy threats identified in stage 1.

Stage 2: Multimedia-related privacy properties: Location privacy is considered an important property that can protect individuals. Location disclosure may threaten the safety of individuals that can be jeopardized if their location is known as in the case of stalking a person. In addition, the insurance of the privacy of the multimedia data is important due to major consequences, explained later, that might arise due to the leakage of the multimedia content. As a result, location and multimedia data privacy must also be considered in the privacy threat analysis of WMSN-based healthcare sub-system.

Following the same approach as in Section 4.3, first the definitions of location and multimedia data privacy will be discussed. Next, their mapping against the DFD elements of the WMSN-based healthcare sub-system will be presented.

Location privacy*:* According to (Haddad et al., 2011), location privacy is defined as "the ability to prevent other parties from learning one's current and/or past location. In order to get such ability, the concerned (i.e., targeted) node must conceal any relation between its location and the personal identifiable information."

Multimedia data privacy: The nature of the multimedia data (contains audio, images and video, text, …) makes multimedia data privacy a very critical privacy objective that must be put into consideration in the privacy threat analysis. In

multimedia content, diverse identifiers have to be obfuscated to hide the identity of a patient: geographic regions that are smaller than a state, dates related to a person in particular, phone and fax numbers, e-mail addresses, social security numbers, number of medical records, health beneficiary numbers, account numbers certificate or license numbers, vehicle ID numbers and serial numbers (including license-plate numbers), device identifiers and serial numbers, internet universal resource locators (URLs), internet protocol (IP) address numbers, biometric identifiers (such as fingerprints or voiceprints), full-face photographic images and any unique identifying number or characteristic or code (Ribaric et al., 2016). Since diverse numbers and types of identifiers can be easily utilised to identify an individual (see Section 3.6), multimedia data privacy is a very important property that must be included in the privacy threat analysis of WMSN-based healthcare systems.

Table 4-2 DFD elements mapped against multimedia-related privacy threats

Privacy Property	Privacy Threat	Entity	Dataflow	Data Store	Process
Location Privacy	Location Disclosure	X	X	X	X
Multimedia data Privacy	Multimedia Identifiability	X	X	X	X

Similar to Table 4-1, Table 4-2 presents location privacy and the multimedia data privacy mapping against the DFD elements. The justification of the location privacy and the multimedia data privacy is as follows:

- Location privacy: Although the term “location privacy” has not appeared explicitly in the LINDDUN research work, the importance of hiding or protecting the location of an individual has appeared several times in their research work. The threat of revealing an individual’s location appeared in the context of “undetectability and unobservability” privacy property in such a way that this privacy property is important to assist in hiding of whether a person is present in a specific location or not. The authors have also mentioned the importance of the processing and the storage of the location data, as stated in the e-privacy directive (Parliament, 2002), and that the processing of the location data must be anonymous and after the consent of the subscribers of publically available electronic communication services. Location data was referred to in the privacy threat tree analysis of the

linkability of an entity, the linkability of the dataflow and identifiability of a dataflow where the linkability and the identifiability can be threatened based on the disclosure of personal identifiable information based on the behavioural pattern such as biometrics, identifiers or frequency (number of times visiting specific locations) and location. Consequently, location has a direct impact on the entities and the dataflow. In addition, since undetectability and unobservability have a direct impact on the protection of the location data and may bring the possibility of threat on the dataflow, data store and process of the DFD, then the location privacy will also have similar impact on the these DFD elements. As a result, location privacy is thought to have an impact on all DFD elements.

- Multimedia data privacy: The aim of the multimedia data privacy is to hide the identifiers that might reveal the identity of an individual. Similar to the identifiability privacy threat that affects all DFD elements, multimedia data privacy will have impact on all DFD elements.

Focal privacy threats: According to Table 4-1and Table 4-2, it can be concluded that the privacy services: anonymity/pseudonymity, unlinkability, location privacy and multimedia data privacy have a higher potential threat impact as they affect all DFD elements. However, the aim of the multimedia data privacy is to obfuscate all identifiers that can be used to discover the identity of a subject. Privacy protection for images or video data is ***often*** achieved by the anonymisation of the data, which is typically done through a selective protection of image regions by abstracting or obfuscating personally identifiable information. Consequently, in this research work, it will be assumed that anonymity/pseudonymity will refer to both anonymity/pseudonymity and multimedia data privacy. Consequently, in this research work, the **focal privacy threats** will be: identifiability, linkability and location disclosure.

For the sake of completeness, it is important to mention that governments seek to constantly enhance the legal frameworks for healthcare, to guarantee the privacy rights of citizens. However, no direct mapping has been made between those frameworks and the privacy services which have been developed by technologies. Research in privacy requirement engineering has been made to try to extract the privacy requirements from laws, regulations and standards such as the work by the authors of (Breaux & Anton, 2008) who have proposed a methodology to elicit rights and obligations from existing laws and text-based regulations to ensure that the

relevant systems developed are accountable and complaint, and assist software engineers to extract security requirements from regulations. The authors have attempted to extract the software requirements by analysing a whole regulation in healthcare domain (namely HIPAA) to make the system abide with the law regulations. The authors of (Breaux & Antón, 2005) have applied semantic parameterisation to analyse the healthcare privacy legislation HIPPA to extract the requirements needed to be fulfilled by healthcare related industries. The work by (Breaux & Anton, 2008) and (Breaux & Antón, 2005) have extracted the system requirements but has not suggested specific privacy threats or privacy properties requirements. Consequently, a mapping between the functional requirements of technical solutions and legal principles is needed, to ensure that the healthcare systems meet legal governmental requirements.

4.5 Summary

Effective measures against privacy violations are of paramount importance in WMSN-based healthcare sub-systems. “Privacy-by-design” is a privacy engineering methodology that is used to discover the privacy threats to a system during design time to avoid the challenging implementation of the privacy services after the software engineering process. Design and implementation of privacy services in an existing software system might conflict the system requirements, which makes retrofitting privacy services challenging. In this chapter, the privacy threat analysis was conducted in two steps. Step 1: the LINDDUN, a general-purpose systematic privacy threat analysis approach, was deployed to discover the privacy services required for the WMSN-based healthcare sub-system. Data-flow diagrams, at Level 0 and Level 1, were created to model the WMSN-based healthcare sub-system and the elements of the DFD were mapped to the LINDDUN privacy threats. Step 2: Two extra multimedia-related privacy services were considered in the privacy threat analysis: location privacy and multimedia data privacy. Based on this 2-step approach, it was concluded that, in this research, the **focal privacy threats** are: identifiability, linkability and location disclosure. The following chapter will present how the privacy mechanisms will be used to counteract the identified focal privacy threats in a WMSN-based healthcare subsystem.

Chapter 5 Privacy-Aware Ant Routing Algorithm for WMSNs

5.1 Introduction

The aim of this chapter is to discuss the implementation of the privacy mechanisms (anonymity/pseudonymity, unlinkability and location privacy), to counteract the privacy threats identified in the previous chapter, towards creating a privacy-aware WMSN-based healthcare sub-system. Anonymity/pseudonymity is required to make the patient anonymous and hide his/her real ID; unlinkability is required to hide the link between the messages in the network and the senders of these messages; and location privacy is required to hide the real location of the sender of the messages and the base station. There are three main building blocks for the research work presented in this chapter: the routing protocol, the encryption key management technique and the privacy mechanisms. The routing protocol will be utilized by the location privacy mechanism. The key management technique will be used by the privacy mechanisms as explained later in the chapter.

This chapter presents a brief overview of the possible deployment scenarios reported in the literature for WMSN-based healthcare subsystems and a brief explanation of the logical layout of the network components. Next a short survey of routing protocols in WMSNs is presented. This section is concluded by a choice of the routing protocol that will be adopted in this research. This is followed by a thorough discussion of the encryption key management building block that has been adopted in this research work. Next, a brief privacy assessment will be conducted for the chosen WMSN-based routing protocol, followed by an outline of the proposed privacy-enhancing algorithm and details of the flow of messages among the system components.

5.2 Application Scenarios for WMSN in Healthcare

According to the literature, WMSNs can be deployed in diverse scenarios such as the commercial, industrial or healthcare domains (Suh et al., 2008). Possible examples of WMSN applications in the healthcare domain are: studying the behaviour of the elderly, such as by deploying video, audio and wearable sensors (for temperature, pressure and monitoring of other vitals) to remotely assist and monitor the them (Akyildiz et al., 2008); monitoring soldiers in hazardous areas (Rawat et al., 2014); monitoring the injured in mass-causality disasters; monitoring vital signs of patients in hospitals; and provision of monitoring, and motor and sensory decline assistance, at home (Ko et al., 2010b). According to (Sharif et al., 2009), personal and healthcare monitoring for the elderly is considered a key

application for WMSNs in healthcare. From this list of possible applications, three general deployment scenarios (that have been mentioned in diverse research papers) have been chosen to assess the solutions proposed in this thesis for a privacy-aware WMSN in healthcare: in-hospital monitoring, elderly home monitoring, and battlefield monitoring.

5.2.1 Hospital scenario

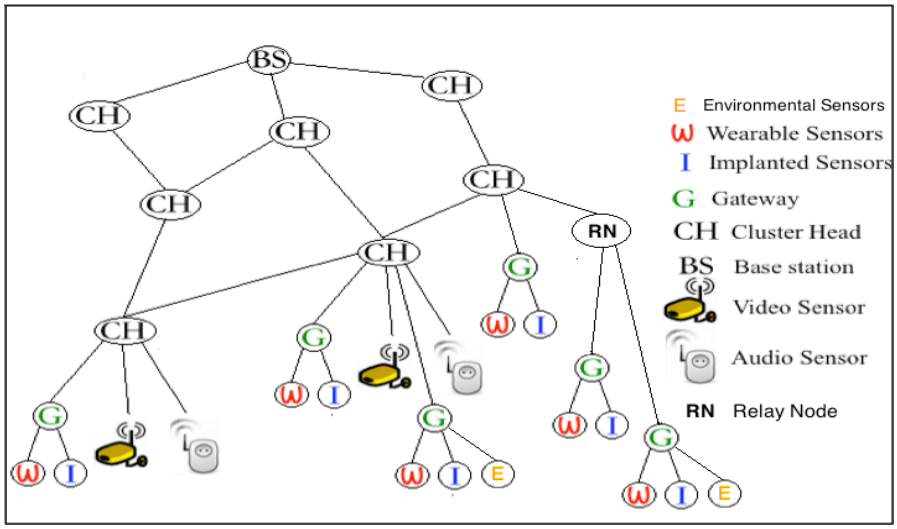


Figure 5-1 A possible layout of the network components in a hospital

Assuming a hospital is made up of *f* floors. Each floor is equipped with *c* cluster heads. The cluster heads are mounted to the walls and dedicated for the collection of the data from the diverse sensors and then routing the data to the base station. Based on this assumption, there is no need for a clustering process since the cluster heads are already identified and their locations are pre-determined. It is also assumed that each room *r* in the hospital is equipped with *v* video sensors, *a* audio sensors and *e* environmental sensors. Each patient has *m* medical sensors attached on, or implanted in, his/her body. A gateway (personal server) that belongs to one patient is used to collect the information from the medical sensors, to send them to the nearest cluster head. Normally a gateway should be directly connected to the nearest cluster head. If a gateway fails to locate a functioning cluster head, it can temporarily connect to a more distant cluster head through a relay node. The function of the relay node is to re-route the data to the nearest cluster head available. If the relay node could not locate a functioning cluster head, it will reroute the data to the next relay node until a cluster head is located. Environmental sensors may be connected to the gateways to collect information about the patient's surroundings, which might be significant in analysing the patient's medical data. A possible layout of the suggested scenario is shown in Figure 5-1. This layout is the

same as that depicted in Figure 5-4 but in Figure 5-1 a relay node is added to re-route data if no cluster head is in range. Audio and video sensors are directly connected to the cluster heads to decrease the load of the routing and processing of the multimedia data on the gateway and save the resources of the gateway for the collection, aggregation and sending of the scalar data collected from the medical and the environmental sensors.

5.2.2 Elderly house scenario

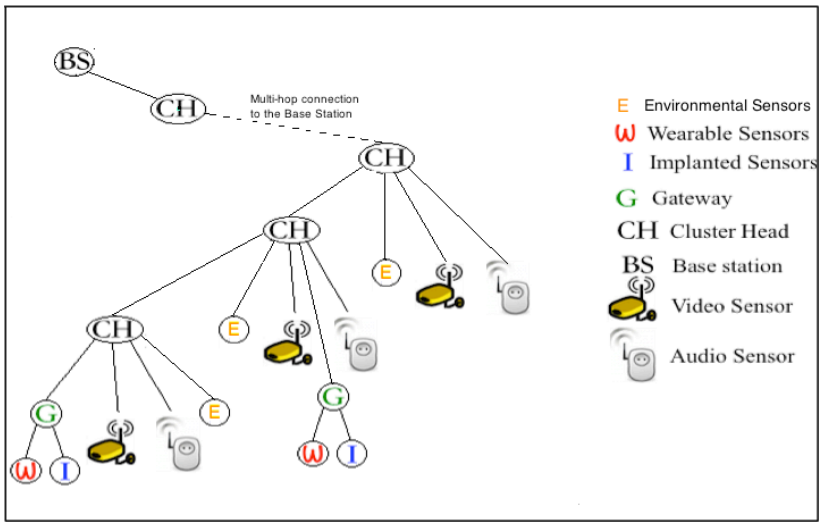


Figure 5-2 A possible layout of the network components in an elderly house

In this scenario, it is assumed that one elderly person or more live in a house that is designed to allow the continuous monitoring of their health and wellbeing without the interference of other family members or strangers. Each room in this house is equipped with environmental sensors (such as pressure, temperature, and RFID sensors), and audio and video sensors. The gateways are deployed to collect data from the wearable and implanted sensors. Cluster heads have a continuous power supply and are installed around the house to provide full coverage to all sensors and gateways. A possible layout of the network components in an elderly house is depicted in Figure 5-2. The communication between the cluster heads inside the house is predetermined before the operation of the network. The cluster heads can be connected to the base station through multi-hop cluster heads (as depicted using the dashed line in the Figure 5-2) or using an Internet connection to send the data outside the house to the base station.

5.2.3 Battlefield scenario

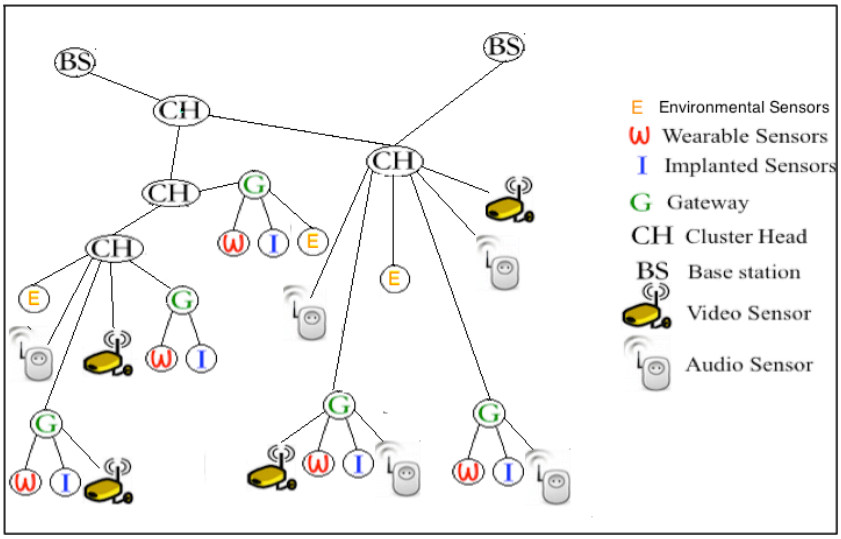


Figure 5-3 A possible layout of the network components in a battlefield scenario

In the battlefield scenario, soldiers can be equipped with implanted and wearable sensors, and other environmental sensors are scattered throughout the battlefield. One or more soldier can be equipped with powerful nodes that act as base stations to collect data from the surrounding cluster heads. These base stations can be connected to remote base stations and remote data collection centres using satellite connections. It is worth mentioning that in real-life battlefields the full coverage of the geographical area of the battle will depend on how scattered the base stations and cluster heads are and the range of their coverage (hardware limitations). In this research, it is assumed that the base stations and cluster heads are scattered all over the geographical area and all gateways have a cluster head in their range.

For simplicity, it is assumed that the soldiers are immobile (mobility will be suggested in the future work). Wearable and implanted sensors are directly connected to gateways. Gateways can only belong to one cluster head. Soldiers may also be equipped with video and/or audio and/or environmental sensors that are connected to their personal gateways. An example of the layout of the sensors is depicted in Figure 5-3. Considering how critical the battlefield application is, more than one base station can be deployed to avoid the failure of the whole sub-system in case only one base station was used and was attacked by an adversary. The base stations can create interleaved networks (using overlapping routes) to collect data from different cluster heads.

5.3 Basic network components

A possible logical layout of the network components for the house design presented in Figure 3-3 is depicted in Figure 5-4. This network can be a logical representation of a room in a healthcare facility where a patient is staying, or a representation of an apartment in a city where a patient is residing.

The network consists of a hierarchal network of nodes in which sensor nodes are in the first level of the hierarchy, gateways are in the second level, cluster heads and a base station are in the third level. Sensor nodes are tightly coupled to one patient (i.e. the sensors ID belong to only one patient from the deployment stage until the retirement, which indicates that if the ID of a sensor is compromised, the ID of the patient will be revealed, causing a possible invasion of the patient's privacy) (Chen et al., 2011) (Darwish & Hassanien, 2011). Sensor nodes are basically medical sensors depending on the condition of the patient and what vitals need to be constantly monitored. Medical sensors can either be implanted inside the patient's body or wearable on the patient's body.

A gateway is mainly responsible for the aggregation of data from the sensor nodes and transmitting the data to the cluster heads through one-hop communication. A gateway can be a mobile device like a smart phone, a laptop or even a server computer. Its main function is to provide a connection between the sensor nodes and the basic infrastructure network (Alemdar & Ersoy, 2010). A gateway can have two modes of operation: online and offline. An online mode refers to the state where the gateway has a direct connection with the cluster head and is able to send data to the base station immediately. An offline mode refers to the case where the gateway cannot have a direct connection with the cluster head (due to patient mobility, for example); in this case, the gateway will have to store the data captured from the sensors until a cluster head is located and a connection established. Gateways are directly connected to cluster heads and not to the base stations, to save their energy for the data aggregation and communication of the data captured by the sensors to be sent to the cluster heads, by avoiding wasting their energy on complicated optimal path finding, routing and complicated services (such as privacy) that need to be conducted by the cluster heads.

A cluster head is responsible for the collection of data from audio and video sensors, and from the gateways underneath it. The cluster head forwards the data to the base station either directly (if there is a direct connection), or through multi-hop routing between the other cluster heads en route to the base station. Each

apartment (or room in the health facility) can have one cluster head. More than one gateway can be connected to the same cluster head (patients visiting or staying in one apartment). In this research work, the video and audio sensors are assumed to be directly connected to cluster heads, to avoid the overhead of the processing and the communication of multimedia data through the gateway, thus saving its limited storage and processing power (in case the gateway is a handheld device such as a mobile phone). Cluster heads are connected together using bidirectional links.

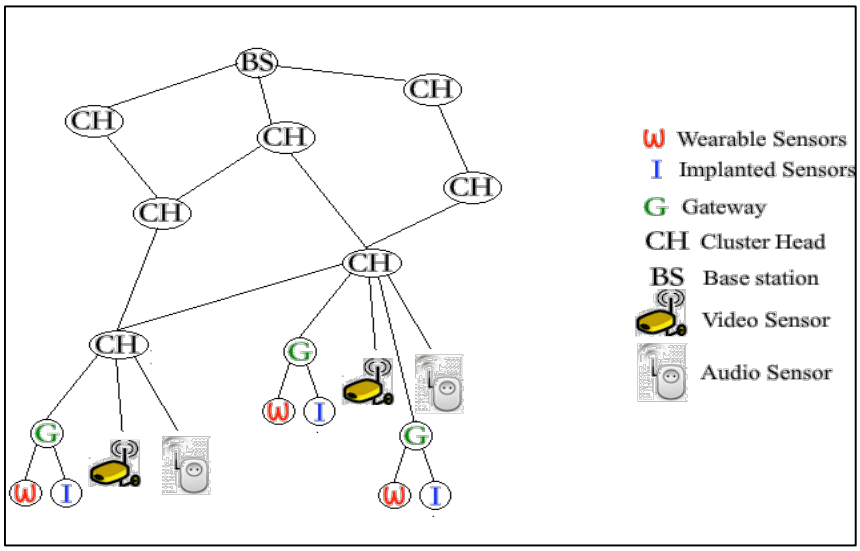


Figure 5-4 A possible logical layout of the logical network components

Figure 5-4 shows the different types of sensors, which might be deployed in a WMSN-based healthcare sub-system, such as video and audio sensors, which may reveal sensitive information about the identity, the health status and the behaviour of the patients being monitored. Consequently, privacy services should be implemented to ensure the acceptance of the WMSN-based healthcare sub-system by both the patients and the governments.

To be able to create a privacy-aware WMSN-based healthcare sub-system, three main building blocks must be included for this research work: the routing protocol, the encryption key management technique and the privacy mechanisms. The routing protocol is an essential building block that will be utilized by the location privacy mechanism, and the key management technique will be used by the privacy services.

5.4 Routing protocol choice for the WMSN-based healthcare sub-system

In general, routing in WSNs is concerned with the transfer of data from the source to the final destination (sink) for data collection and analysis purposes (Akyildiz et

al., 2007). Energy saving is considered a main objective for most WSN routing protocols, while assuming traffic of data with unconstrained delivery requirements (Cobo et al., 2010). On the other hand, WMSN routing imposes several challenging factors that are mostly application-dependent and should be considered, to achieve effective WMSN communication such as quality-of-service (QoS) requirements, energy efficiency, architecture issues, and hole detection and bypassing (Kandris et al., 2011). Routing protocols for WMSNs have been either adopted from traditional WSN, with modifications to meet the stringent QoS requirements, or based on new solutions developed considering the application-based QoS requirements and the network layer metrics (Ehsan & Hamdaoui, 2012) (Almalkawi et al., 2010).

The choice of a routing protocol for this research work is a very critical decision. The addition of both privacy and security services is expected to add more computation, communication and storage overhead to the nodes of the WMSN. Consequently, an energy efficient and QoS assured routing protocol that will allow the efficient use of the sensor resources, adapt to the different types of traffic (scalar, audio and video) to maximise network utilisation and improve the overall performance is crucially required. The survey paper by (Ehsan & Hamdaoui, 2012) presents a comparison between the energy efficient routing techniques, with QoS assurance, developed for WMSNs. The comparison was based on the following criteria: network architecture (flat or hierarchal), location awareness, multipath capabilities, energy efficiency, bounded latency, reliable delivery, data delivery class (query driven or event driven) and hole bypassing. Table 5-1 shows the outcome of this comparison.

In Table 5-1, five columns were chosen as criteria for the choice of a routing protocol for this research work namely: hierarchal architecture, multipath capability, energy efficiency, reliable data delivery and both query and event driven data services. The hierarchal architecture is chosen due to the hierarchal nature of the WBSNs where data is captured and processed to extract the necessary information in a way that makes use of the resources asymmetry, ensures system efficiency and data availability (Darwish & Hassanien, 2011). Multipath capability is required for the efficient delivery of multimedia data. Reliable data delivery is a crucial criterion to be considered in healthcare data where life-critical data is being handled and a case of lost or damaged packets might lead to overlooking an emergency situation, for example (Darwish & Hassanien, 2011). Both query and event driven services are required in healthcare sub-systems as in situation when specific sensors are queried for more details or extra information and, in emergency

situations where emergencies are reported in event driven services. Based on the chosen criteria, only one protocol fulfils most of those requirements: “ASAR: An ant-based service aware routing algorithm for wireless multimedia sensor network” (Sun et al., 2008).

Table 5-1 Comparison between the energy efficient and QoS-aware WMSN-based routing protocols (Ehsan & Hamdaoui, 2012)

Routing protocol	Architecture		Location awareness	Multipath capability	Energy efficiency	Bounded latency	Reliable delivery	Data-delivery model		Hole bypassing
	Flat	Hierarchical						Query-driven	Event-driven	
SAR [40]	✓			✓	✓	✓		✓		
RAP [21]	✓		✓			✓		✓	✓	
EAQoS [41]		✓			✓	✓		✓		
SPEED [42]	✓		✓			✓		✓		
RPAR [45]	✓		✓		✓	✓		✓		
Pothuri et al. [46]		✓			✓	✓		✓		
Yuan et al. [47]	✓		✓		✓	✓		✓		
Khalid et al. [48]	✓			✓	✓	✓		✓		
Ergen et al. [49]	✓				✓	✓		✓		
MMSPEED [22]	✓		✓	✓		✓	✓	✓		
Hamid et al. [50]	✓			✓		✓	✓	✓		
DARA [51]	✓		✓	✓	✓	✓	✓	✓		
RTLD [52]	✓		✓	✓	✓	✓	✓	✓		✓
Sen & Ukil [53]	✓			✓	✓	✓	✓	✓		
Mahapatra et al. [54]	✓			✓	✓	✓	✓	✓		
OEDSR [55]	✓			✓	✓	✓			✓	
DGR [56]	✓			✓	✓	✓	✓			
Poltis et al. [57]		✓		✓	✓	✓	✓			
REAR [58]	✓				✓	✓			✓	
ASAR [59]		✓			✓	✓	✓	✓	✓	
Peng et al. [60]	✓				✓	✓	✓			
Zongwu et al. [61]	✓				✓	✓				
Haiping & Ruchuan [62]		✓	✓		✓		✓			
TPGF [66]	✓		✓	✓	✓	✓	✓			✓
GEAMS [68]	✓		✓	✓	✓	✓	✓			✓

Although the ASAR (Sun et al., 2008) protocol has better convergence and better QoS than traditional routing protocols, it does not support multipath capability (as shown in Table 5-1), which is important for improving the WMSN transmission performance (Cobo et al., 2010). Consequently, another ant-based routing protocol called AntSensNet (Cobo et al., 2010) was chosen as a replacement for ASAR. AntSensNet supports all the criteria that have been chosen, as shown in Table 5-1. Furthermore, AntSensNet has a multipath capability, which further enhances the transmission performance of the routing protocol. In the next section, a general overview of ant-based routing in WMSNs is presented.

5.4.1 Ant-based routing in WMSNs

Ant Colony Optimisation (ACO) is a metaheuristic optimisation method based on the mechanisms adopted in ant colonies in search for food (Saleem et al., 2011). Algorithms based on this concept have been adopted in various applications ranging from optimisation problems to robotics (Saleem et al., 2011). The distributed heuristic nature of the ant-based routing protocols are suitable for WSNs due to: distributed nature of the algorithm (no single point of failure); simple operations carried at the nodes; asynchronous and autonomous algorithm interactions; self-organising nature; adaptation to different traffic and adaptation to topological variation and traffic demand (Cobo et al., 2010).

The basic idea of ant-based routing is to acquire information about the route using a collective learning process for path sampling using concurrent and independent agents (ants) to try out a path to a certain destination (Saleem et al., 2011). Ants can be forward ants which move from the source to the destination to collect information about the quality of the path (e.g. end-to-end delay) or backward ants which move from the destination back to the source to update the route table (pheromone table) in the network nodes with the information collected (Saleem et al., 2011). The pheromone table contains entries for assessing the paths to the destination through the neighbourhood nodes. The entries of the pheromone table are continuously and repeatedly updated (Saleem et al., 2011).

In a recent study conducted by (Nayyar & Singh, 2017), ant-based routing showed the best routing protocol performance (based on less end-to-end delay, less packet overhead, best throughput and less routing overhead) for WSNs, compared to other routing protocols such as Ad hoc On-Demand Distance Vector (AODV), and dynamic destination-sequenced distance-vector (DSDV). According to a recent survey paper (Bhandary et al., 2016), several ant-based routing protocols have been developed for WMSN-based applications, such as AntNet (Caro et al., 1998), M-IAR (Rahman et al., 2008), ASAR (Sun et al., 2008), ACOLBR (Bi et al., 2010), ACOWMSN (Yu et al., 2011), AntSensNet (Cobo et al., 2010) and several others. AntNet (Caro et al., 1998) is an adaptive learning routing protocol in which the network nodes issue forward ants at constant amounts of time, to discover appropriate routes to the base station and the backwards ants are generated by the base station to update the routing tables. Although the AntNet has a better throughput compared to classical routing algorithms, it is considered slow due to high routing overhead (Bhandary et al., 2016). The M-IAR (Rahman et al., 2008) is a flat multi-hop WSN multimedia routing protocol adapted from the IAR (Improved Adaptive Routing) protocol that takes into consideration the end-to-end delay and the jitter QoS requirements. The M-IAR protocol offers reliable solutions, is able to discover shortest paths, can be acknowledgement based and non-acknowledgement based; however, it has high deployment costs and suffers a performance degradation due to the exponential increase of overhead under high network load conditions (Bhandary et al., 2016). The ASAR protocol (Sun et al., 2008) is a hierarchal energy efficient routing protocol that supports different data delivery models like query-driven, data-driven and stream-driven models. Although the ASAR protocol has better convergence and better QoS than traditional routing protocols, it does not support multipath capability and suffers from performance

degradation caused by bottle-necks and continual use of optimal paths (Bhandary et al., 2016) (Saleem et al., 2011). The ACOLBR protocol (Bi et al., 2010) is a hierarchal multipath algorithm that uses a minimum spanning tree to build intra-cluster routing using the cluster head as the root, then ant colony optimisation is adopted to provide an optimal path between the cluster head and the sink. Although the ACOLBR protocol has better end-to-end delay and energy efficiency compared to the M-IAR protocol, and recovers from path failures and has a congestion control capability, it requires complex computations and can cause bottlenecks at the cluster heads (Bhandary et al., 2016). The ACOWMSN protocol (Yu et al., 2011) is a reactive, energy aware and adaptive ant-based routing protocol which is designed to find routes according to specific QoS requirements of the applications. Although this protocol considers QoS parameters such as energy, packet loss and bandwidth requirements, it suffers from slow convergence in large-scale networks (Bhandary et al., 2016). The AntSensNet protocol (Cobo et al., 2010) is a hierarchal routing protocol which is designed especially for WMSNs (see the next section for more details). According to (Bhandary et al., 2016), the AntSensNet protocol can handle congestion control and has better video quality compared to other routing protocols. Consequently, the AntSensNet protocol has been chosen as the underlying routing protocol for this research work.

5.4.2 Privacy-awareness of ant-based routing algorithms

Before deploying the AntSensNet protocol into this research work, an extensive literature survey was conducted to find already existing privacy-aware WMSN routing protocols, which provide privacy services such as those that were considered for adoption in this research work. However, it was noticed that very little research has been published on privacy-aware ant routing protocols in WSNs and none has been published on privacy-aware WMSN routing.

(Dias et al., 2013) proposed the deployment of the ant colony optimisation theory for the private route planning of vehicles without the need to know the source or the destination of the vehicle. Their idea was to distribute the traffic of vehicles over the city and at the same time find the shortest path for individual drivers using the ant colony optimisation. Although the title of their work suggests that the work focuses on privacy, the authors did not discuss any privacy measures except that their algorithm provided general information to the drivers without knowing their source or destination.

Another work (Kalpana & Rengarajan, 2012) has suggested the deployment of anonymous ant based routing with trust. The main idea of their work is that a trusted leader node (a cluster head node) is selected to route the data from the source to the destination. The source leader node uses a hash function to rename the ID of the source, to encrypt it, and then the message is broadcast to the other nodes. The intermediate nodes and the destination nodes can decrypt the ID of the destination and check their routing tables to route the messages towards the destination or keep the message if these nodes are the destination. Although their technique achieves anonymity, a global adversary can still learn the source and destination nodes by tracking the messages flowing in the network.

(Zhou & Wen, 2014) suggested the deployment of the ant colony optimisation scheme to create an energy efficient mechanism to protect the location of a sensor (source location privacy). They used the ant colony optimisation algorithm to try to stop an adversary from tracing a message back to its source location by applying an energy efficient source location privacy mechanism to the evaporation and disposition of the pheromone levels in the routing table and applying random packet forwarding and random delay to route the messages away from the source node. Although their technique has enhanced the location privacy of the source nodes, the communication did not protect the identity of the nodes or the linkability of messages.

5.5 AntSensNet routing protocol

The AntSensNet protocol combines the basics of the ant colony optimisation-based routing with the hierarchal structure of the network, to provide QoS and power efficient multipath video packet scheduling. This routing protocol is both a reactive and proactive protocol. It is reactive due to the fact that the routes are set up when required and then the data packets are sent stochastically over different paths. The protocol is proactive because routes are probed and maintained. AntSensNet is composed of three main parts:

1) Clustering the network nodes into colonies. Clustering is an important step as it allows scalability, saves network resources as resource-rich nodes are selected as cluster heads, forming a backbone of cluster heads and increasing the network lifetime by applying cluster head rotation. In the AntSensNet protocol, the clustering is completely distributed and is derived from the Nature-inspired data

gathering protocol for wireless sensor networks (T-ANT (Selvakennedy et al., 2006)) clustering algorithm.

2) Route discovery between the clusters, which is based on the application requirements.
3) Traffic forwarding is based on the routes discovered in (2).

In the route discovery phase, forward ants (FANTs) leave the source node to the neighbouring nodes to discover the routes surrounding the node towards the base station. As the FANTs move around the network, each node constructs a routing table containing the identification of all surrounding neighbourhood nodes and their corresponding pheromone level.

The data collected by ants is stored in a pheromone table containing the following entries:

1- The cluster head neighbour ID
2- The traffic class (based on the application)
3- QoS parameters
4- Expiration time

The AntSensNet protocol is a three-phase algorithm: the FANTs phase, the backward ants (BANTs) phase and the routing maintenance phase. The FANTs phase is when a cluster head needs to send the data; the pheromone table is checked to find an unexpired route to use. If all paths in the table are expired or the paths are unsatisfactory, FANTs are broadcast from the cluster head to the sink to discover other routes. When a cluster head receives a FANT, it updates the ant. The cluster head adds its ID to the ant nodes stack, increments the hop count field and updates the ant's information field. In the BANTs phase, after the FANT reaches the sink, the sink evaluates the QoS parameters to decide whether these parameters are appropriate for the application requirements. If the parameters in the FANT are not appropriate, then the ant is discarded. If the parameters are appropriate, a BANT is generated and sent back in the same path that the FANT followed. During the return of the BANT, the pheromones are updated in the routing tables of the cluster heads it passes by. In the route maintenance phase, the routes are updated to deal with the congestion and lost link problems.

5.6 Assessment of privacy requirements for the AntSensNet protocol

The authors of the AntSensNet (Cobo et al., 2010) protocol did not mention any privacy or security services supported by their proposed protocol for the protection

of the data, or of the identity of the sender/receiver or the location of the sender/receiver of the data. The lack of privacy and security services can cause serious privacy threats (including the privacy threats which are the focus of this work, as depicted in Table 5-2), which could make it difficult for this protocol to be accepted by users who are concerned about not revealing their identity and safeguarding their data.

Table 5-2 Privacy threats and their corresponding privacy services required for the AntSensNet protocol

	AntSensNet Property	Privacy Threat	Privacy Service Required
1	All FANTs are sent to one sink	A global adversary can monitor the overall traffic of the network and notice all traffic moving towards one sink (base station)	**Location Privacy** of the base station. Fake messages generated at the base station can trick the adversary into thinking the base station is an ordinary sensor node
2	Data captured by the sensors is sent to the cluster head and then routed to the base station	A global adversary can trace the traffic and learn the origin or identity of the sender of the messages	**Location Privacy** of the sender. Fake messages generated randomly at different sensors can hide the location of the real sender of the message. **Anonymity and unlinkability** can be achieved using fake messages, which hide the identity of a sensor, if only one is connected under a cluster head.
3	Every cluster head holds a routing table of the real IDs of cluster heads	A local adversary (captured one or more node) can give away the true identity of the surrounding nodes. A global adversary can learn the true identities of all network nodes	**Pseudonymity** is required to hide the real identities of all the sensor nodes. Sensor nodes can be required to communicate using only their pseudonyms. The use of constantly- changing pseudonyms can stop the adversary from learning the real ID of the node.
4	Every cluster head holds a routing table of the pheromone levels and the paths to the sink	Once a cluster head is captured, an adversary can learn about the network formation and characteristics of the paths	**Pseudonymity** is required to hide the real identities of all the sensor nodes. Sensor nodes can be required to communicate using only their pseudonyms. The pseudonyms can be constantly updated to make it harder for the adversary to relate them to specific sensor nodes. The encryption of the network information and the pheromone levels in the FANTs and BANTs can stop the adversary from learning the characteristics of the network paths.
5	Ants captured during forward phase	A local or global adversary can learn the source node	**Pseudonymity** is required to hide the real identities of all source and the sink nodes. **Unlinkability** through the continuous update of the ants at each cluster head and the encryption of the ant can make it hard for an adversary to link the ants to source nodes. Re-encryption of the FANTs at each cluster head can make the ants look different and hard to trace.
6	Ants captured during backward phase	A local or global adversary can learn the source and sink node	**Pseudonymity** is required to hide the real identities of all source and the sink nodes. **Unlinkability** through the continuous update of the ants at each cluster head and the encryption of the ant can make it hard for an adversary to link the ants to source nodes. Re-encryption BANTs at each cluster head can make the ants look different and hard to trace.
7	An ant is captured by an adversary	It gives away information about all cluster head in the path (ID, hop count, residual energy and other important parameters)	**Pseudonymity** is required to hide the real identities of all source and the sink nodes. **Unlinkability** through the continuous update of the ants at each cluster head and the encryption of the ant can make it hard for an adversary to link the ants to source nodes.

5.7 Anonymity/Pseudonymity, unlinkability and location privacy for a WMSN-based healthcare sub-system

According to the privacy threat analysis presented in Chapter 4 that was applied to a WMSN-based healthcare sub-system, the aim of the required privacy services is to achieve three main privacy goals: anonymity/pseudonymity, unlinkability and location privacy. An extensive literature survey was conducted to study what techniques to use to achieve these privacy goals. Among those techniques, fake (dummy) packet generation and pseudonyms pools were the most popular. In this research work, several techniques have been adopted towards the creation of a privacy-aware WMSN-based healthcare sub-system. Every technique was used to deliver one or more privacy service as follows.

5.7.1 Unlinkability

To apply unlinkability in the WMSN-based healthcare sub-system, size correlation, encryption, pseudonyms and fake traffic were used. Correlation of the size of the ants and of the data packets ensures that an adversary cannot distinguish them. Consequently, an adversary will not be able to tell whether the packet in the network is a FANT discovering the route to the base station, or a BANT arriving from the bases station, or a scalar data packet or one of the packets for a multimedia stream. The encryption and the constant update of the FANTs at each cluster head with different keys makes it hard to link the FANTs together. In addition, encryption is deployed to make the encrypted packets look random. Each FANT should be decrypted at the cluster head, updated with the current information of the cluster head and then re-encrypted with the shared key of the next cluster head or base station, to stop an adversary from linking the same FANT captured at different parts of the network. The same concept applies for the BANTs even though they are not updated on their way back to the source node but they are decrypted and re-encrypted with different keys at each cluster head, which makes them unlinkable. The use of pseudonyms instead of the real IDs of the network nodes will ensure the unlinkability as two different messages captured by an adversary may have two different pseudonyms, which belong to the same sender but the adversary will not be able to determine this because the pseudonyms are regularly updated after each successful transmission of data to the base station. Fake traffic can be used to achieve unlinkability in cases when only one sensor is connected to a cluster head. A global adversary can easily relate all messages generated from this particular cluster head to the only sensor connected to it thus threatening the linkability and identifiability of this sensor.

5.7.2 Anonymity/Pseudonymity

The use of constantly changing pseudonyms instead of the real IDs of the nodes will achieve pseudonymity. The pseudonyms are updated after each successful transmission (when a node receives an acknowledgment from the base station).

The use of fake message generation is important, in some cases, to achieve anonymity. In case when only one gateway is connected to a cluster head, an adversary can easily relate all messages generated from the cluster head to the gateway resulting in zero anonymity. Generating fake traffic in this case can trick an adversary into thinking that there is more than one gateway behind the cluster head and thus increase the level of anonymity of the gateway.

In case of multimedia traffic, encryption of the multimedia content can help obfuscate the characteristics that an adversary can use to identify a patient and thus decrease the level of identifiability of the sub-system.

5.7.3 Location privacy

The size correlation of ants and data packets and the encryption of the data packets and the ants should attempt to stop an adversary (local or global) observing the network from knowing whether the packets are data packets or ants. The size correlation and encryption should ensure unlinkability and the location privacy service because the packets of a multimedia stream divided into smaller packets and routed to the base station among different routes will not be linked to the source node.

The generation of fake packets at both the sensor level and the base station level enhances the source location and base station location privacy, respectively. Generating fake traffic at the level of the source node will stop the adversary from determining the source node. At the base station level, the generation of the fake traffic will trick the adversary into thinking that the base station is just a node like the rest of the network nodes.

5.8 Proposal of a privacy-aware AntSensNet routing protocol

The proposed privacy-aware AntSensNet protocol is based on the integration of privacy-enhancing features together with the Localized Encryption and Authentication protocol (LEAP) key management protocol (Zhu et al., 2006) into the AntSensNet protocol (Cobo et al., 2010). Basically, the operation of the AntSensNet protocol is divided into three main stages: pre-deployment stage, deployment stage

and traffic forwarding stage. The privacy-enhancing features and the key management protocol are integrated into all three stages. Consequently, the work presented in this section is a combination of the AntSensNet protocol with privacy-enhancing features and with the chosen encryption key management protocol to deliver the privacy-aware AntSensNet protocol. A detailed review of all three stages is discussed in this section. The integrated privacy-enhancing features are highlighted (using the orange colour) in Figure 5-5 and Figure 5-6.

5.8.1 Stage 1: Pre-deployment stage

The aim of this stage is to pre-load the (scalar and multimedia) sensors with their hash functions for generating the pseudonyms and the pseudorandom functions for generating the encryption keys, before their deployment, to achieve **pseudonymity and unlinkability**. In order to avoid the compromise of the whole network in case one or more nodes are captured by an adversary, the base station will deploy a pool of hash functions that will be randomly distributed among the sensor nodes. Only the base station will know which hash functions and in what order they belong to every sensor node. Consequently, if one node is compromised, the adversary cannot tell for sure if the rest of the nodes in the neighbourhood are using the same hash functions or not.

The nodes are also pre-loaded with the pseudorandom functions for the operation of the LEAP-based key management protocol. In addition, the individual key, which is only known to the base station, is generated and pre-loaded into the sensor node. For a sensor node *u* with a unique ID, the individual key is generated using pseudorandom function as follows:

$$K_u^m = f_{k^m}(u) \quad (1)$$

where $K_u^m$is the individual key of the sensor node named *u* generated using the master key k^m stored at the base station and only known to it and not to the rest of the sensor nodes (Zhu et al., 2006). This ensures the secure communication between the sensor nodes and the base station through the intermediate nodes (such as gateways and cluster heads). The base station computes K_u^m when communication is required with the node *u*, which does not impose much extra computational effort due to the efficiency of the pseudorandom functions. The same steps are followed for the pairwise key, where a master key is used at the base station to generate a special key K_I loaded into all the sensor nodes before

deployment to generate pairwise keys between the sensor nodes. According to (Zhu et al., 2006), the master key K_u is generated using

$$K_u = f_{K_I}(u) \quad (2)$$

A detailed flowchart of the steps of Stage 1 is depicted in Figure 5-5. In this flowchart, the sensor nodes are loaded with *H* hash functions from which one hash function will be picked in an order only known to both the base station and the sensor node, to generate a pseudonym used to identify the sensor node within its neighbourhood and at the base station. Each time the pseudonym is updated, the sensor node sends encrypted messages, using the pairwise keys between the node and the neighbouring nodes, to update its pseudonym in their routing tables. The part of the flowchart used by the privacy mechanisms is coloured in orange. At the base station, a pool of hash functions is stored in the form of an array-like structure and the base station marks which group of hash functions are preloaded in each sensor node. The number of hash functions depends on the level of privacy required by the users of the healthcare sub-system. The higher the number of hash functions, the more pseudonyms are generated, which makes it harder for an adversary to relate the packets to the source nodes. However, this will impose more storage overhead at both the sensor nodes and the base station.

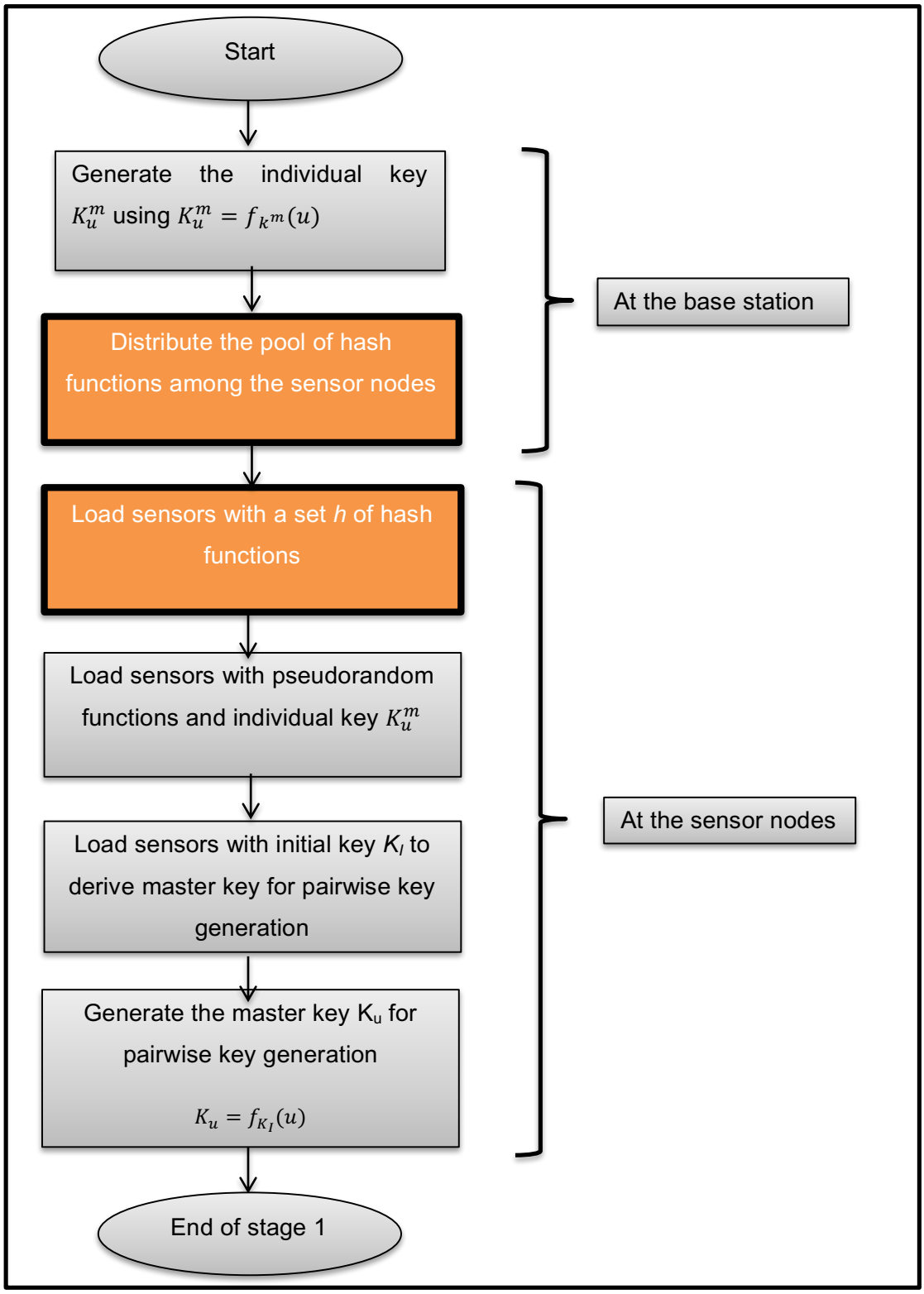


Figure 5-5 Flowchart for Stage 1 (pre-deployment stage)

5.8.2 Stage 2: Deployment and initialization stage

The aim of this stage is to prepare the sensor nodes for the operation stage through the creation of the: pseudonyms for all sensors, clusters of sensors, individual and pairwise encryption keys. Similar to Stage 1, Stage 2 (deployment and initialisation stage) is used to achieve **pseudonymity and unlinkability.** Stage 2 is assumed to be *mostly* carried out during the safety period T_{saf}, which is the time elapsing before an adversary compromises a sensor node. This stage is subdivided into three sub-stages: generation of pseudonyms, generation of encryption keys, and the clustering process. The details of each sub-stage are outlined as follows.

Generation of pseudonyms

Every node is pre-loaded with a set of hash functions that are used to generate pseudonyms for the sensor nodes instead of using the real IDs of the sensor nodes. Every sensor node is expected to use a unique pseudonym for every transmission and then change the pseudonym after the node receives an acknowledgment that the transmission has successfully arrived at the base station. To update the pseudonym at the base station, a sensor node can randomly choose a hash function from the pool of functions that is already pre-loaded in the pre-deployment stage and send an encrypted message to the base station to indicate which hash function will be used in the next transmission. However, this will cause a huge communication overhead in the network. In this research work, the hash functions have been deployed in the same order as that stored at the base station to avoid sending messages to the base station from each cluster head to inform the base station which hash function is used to update the pseudonym. This way the base station will automatically update the pseudonym after sending an acknowledgment that the data has been successfully received. However, only during Stage 2, the sensor node will generate its first pseudonym (p_id_0) and keep it until the end of this stage.

Generation of encryption keys

Encryption of the data being communicated between the different network components is important to guarantee the protection of the data and prevent adversaries from having access to the data being transmitted. Encryption keys are used to encrypt the data at the source before sending it to the destination. Only by using correct key, the destination can decrypt the data and understand the content of the received message. Without the decryption key, an adversary should not be

able to understand the content of a captured message. Only two keys of the LEAP protocol have been adopted into this research work: the individual keys and the pairwise keys. Both the individual key and the master pairwise key are generated and pre-loaded into the sensor nodes during the pre-deployment stage (Stage 1). In Stage 2, the establishment of pairwise shared keys undergo three steps: neighbour discovery, pairwise key establishment and key erasure (Zhu et al., 2006).

Neighbour discovery: According to (Zhu et al., 2006), after deployment and during the T_{saf} time interval, a sensor node *u* tries to communicate with all its next one-hop neighbours using a HELLO broadcast message containing its pseudonym generated in Stage 2 (pseudonym generation). The sensor node *u* awaits the acknowledgment of the neighbours, which is authenticated using the master key K_v generated using $K_v = f_{K_1}(v)$.

Assuming *v* is the sender, *u* verifies the identity of node *v* using:

$$u \longrightarrow *: u$$
$$v \longrightarrow u: v, M\ A\ C(K_v, u|v) \quad (3)$$

where *M A C* is the Message Authentication Code using the symmetric key *k.* In their work, (Zhu et al., 2006) proposed a one-way key chain based authentication scheme. The authentication scheme is a mandatory requirement to avoid the case when an adversary can deplete the energy of a sensor node by inserting false packets into the network. This one-way key authentication is computationally lightweight. The basic idea of this scheme is that each node generates a one-way key chain and sends the first key of the chain (referred to as AUTH key), encrypted using the pairwise key, to each next hop neighbour. When a node is sending a message to another node, the next AUTH key in the chain is added to the message. A neighbouring node verifies the messages using the most recent AUTH key received from the sending node.

Pairwise key establishment: Both sensor nodes *u* and *v* can now compute their pairwise keys K_{uv} using the equation:

$$K_{uv} = f_{K_v}(u) \quad (4)$$

After the pairwise keys are generated, the authentication between sensor node *u* and *v* is no longer required because the messages will be authenticated using K_{uv}.

Key erasure: When T_{saf} elapses, the node u erases the K_l and the master keys of the neighbours that were exchanged during the pairwise key generation. However, only the master key of the node is kept.

Clustering process

Before the clustering phase starts, sensor nodes broadcast HELLO packets containing: node ID, clustering pheromone value and the node state. In order to suppress an adversary listening to the network from acquiring information about the data being communicated and the cluster head election, the node ID is replaced by the pseudonym. In addition, both the clustering pheromone and the node state are encrypted using the LEAP pairwise key established between the neighbouring nodes earlier in this stage.

According to the AntSensNet routing protocol, the clustering algorithm used is divided into rounds and each round is composed of two phases, cluster setup phase and steady phase. In the cluster setup phase, the cluster heads are elected and in the steady phase, the transmission of the data takes place between the sensors and the base station. Cluster Ants (CANTs) are used in the cluster heads elections in such a way that a node in possession of a CANT is elected as a cluster head and the rest of the nodes are required to join the most appropriate cluster in their range. The node in possession of a CANT can be a sensor node or a gateway. However, the election of sensor nodes should be kept to the minimum to save their energy and increase their time span.

Before the clustering phase starts, sensor nodes broadcast HELLO packets containing: node ID, clustering pheromone value and the node state. In order to suppress an adversary listening to the network from learning information about the data being communicated and the cluster head election, the node ID is exchanged with the pseudonym. In addition, both the clustering pheromone and the node state are encrypted using the LEAP pairwise key established between the neighbouring nodes earlier in this stage. At this point, the sensor nodes will be using their initial pseudonyms that were derived without update. The sensor node constructs a neighbour information table to store all the information it received from the HELLO messages. The clustering pheromone is computed using the formula:

$$\phi_c(n) = (ma(n)t)^a . (re(n))^b \quad (5)$$

where *ma(n)* is the available memory, *re(n)* denotes the residual ratio of the energy of the node, and *a* and *b* are tunable values reflecting the importance of the memory and residual energy depending on the application.

After the sensor nodes have constructed the neighbour information table and the information update between the sensor nodes is done, the base station starts releasing the CANTs. The number of released CANTS released (consequently the number of cluster heads) is computed using:

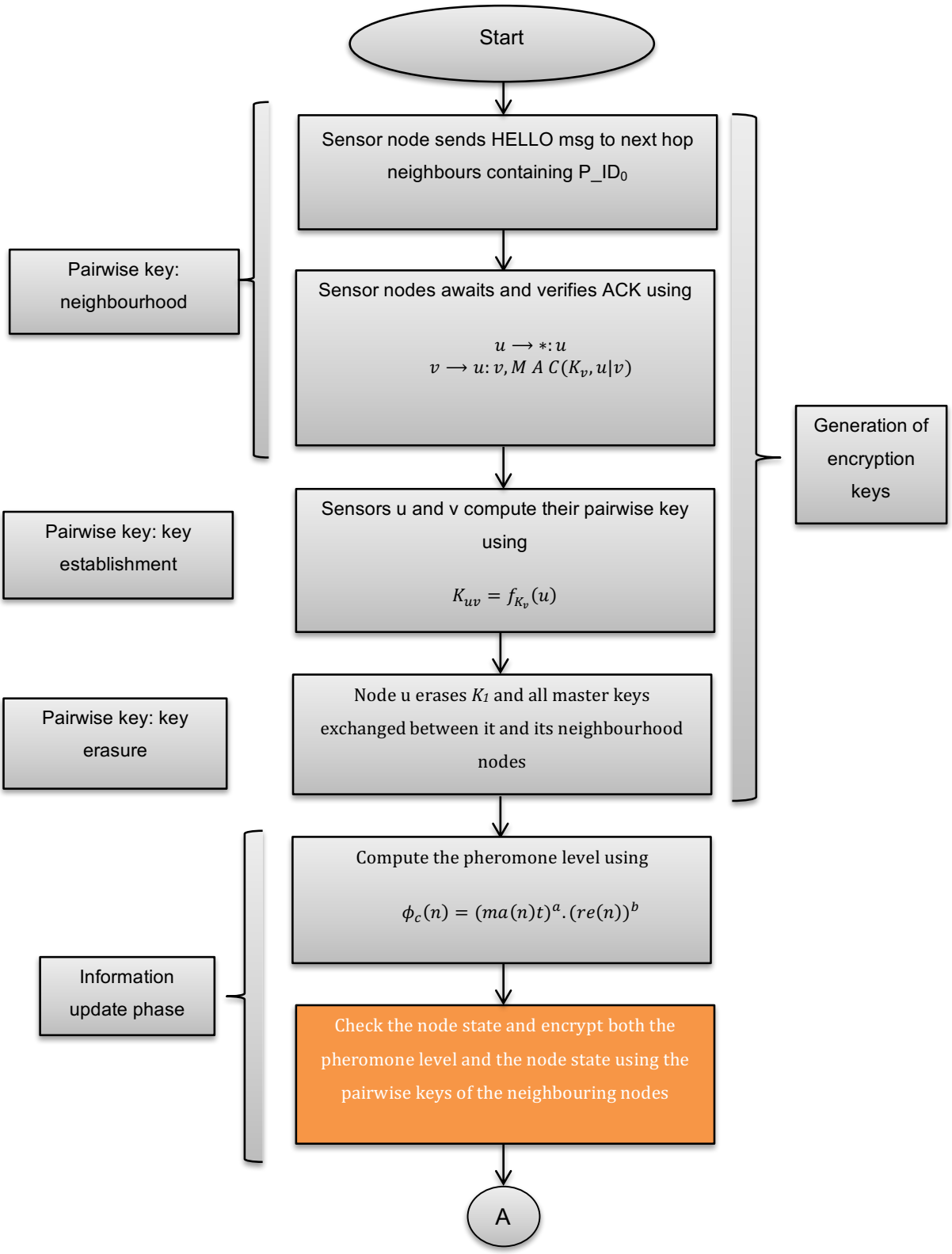
$$\text{Clusters Number} = \left\lceil \frac{M^2}{\pi d^2} \right\rceil \quad (6)$$

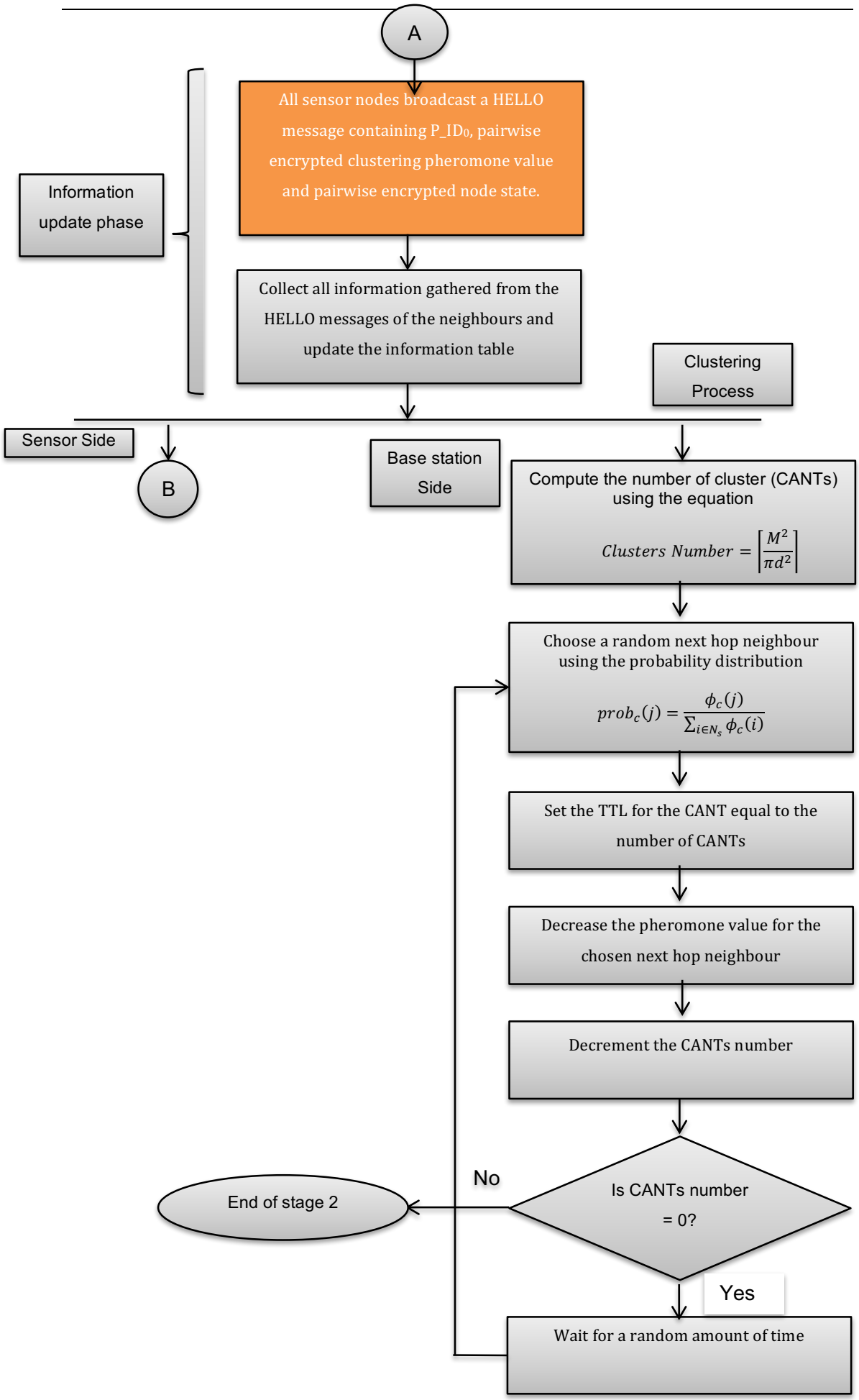
where M represents the network size (M x M) and d is half the cluster radius which is a tunable value denoting the minimum distance between any two clusters. The Time To Live (TTL) of the released CANTS is equal to the cluster number. The base station chooses a random next hop neighbour to send the CANT to, based on the probability distribution:

$$\text{prob}_c(j) = \frac{\phi_c(j)}{\sum_{i \in N_s} \phi_c(i)} \quad (7)$$

where $\phi_c(j)$ is the clustering pheromone of node *j* and N_s refers to all the sensor node neighbours of the base station within an area of cluster radius. After the base station sends a CANT to a neighbour sensor node, it decreases its pheromone value in the information table so that it minimises the chances of choosing the same sensor node again. Before sending the next CANT, the base station would wait for a random amount of time to avoid the interference of the CANTs together.

Figure 5-6 depicts the details of Stage 2. In this flowchart, the ADV_CLUSTER is a message sent by the cluster head to the sensor nodes in their range asking them to join their cluster. The sensor nodes reply with a Join message containing their pseudonym so that the cluster head can store it in its information table. Similar to the previous flowchart in Figure 5-5, all parts in the flowchart highlighted in orange are used in the privacy mechanisms.


Start
Sensor node sends HELLO msg to next hop neighbours containing P_ID0
Pairwise key: neighbourhood
Sensor nodes awaits and verifies ACK using
u ⟶ ∗: u
v ⟶ u: v, M A C(K_v, u|v)
Generation of encryption keys
Pairwise key: key establishment
Sensors u and v compute their pairwise key using
K_uv = f_Kv(u)
Pairwise key: key erasure
Node u erases K1 and all master keys exchanged between it and its neighbourhood nodes
Compute the pheromone level using
φ_c(n) = (ma(n)t)^a . (re(n))^b
Information update phase
Check the node state and encrypt both the pheromone level and the node state using the pairwise keys of the neighbouring nodes
A


A
All sensor nodes broadcast a HELLO message containing P_ID0, pairwise encrypted clustering pheromone value and pairwise encrypted node state.
Information update phase
Collect all information gathered from the HELLO messages of the neighbours and update the information table
Clustering Process
Sensor Side
B
Base station Side
Compute the number of cluster (CANTs) using the equation
$Clusters\ Number = \left\lceil \frac{M^2}{\pi d^2} \right\rceil$
Choose a random next hop neighbour using the probability distribution
$prob_c(j) = \frac{\phi_c(j)}{\sum_{i \in N_s} \phi_c(i)}$
Set the TTL for the CANT equal to the number of CANTs
Decrease the pheromone value for the chosen next hop neighbour
Decrement the CANTs number
Is CANTs number = 0?
No
End of stage 2
Yes
Wait for a random amount of time

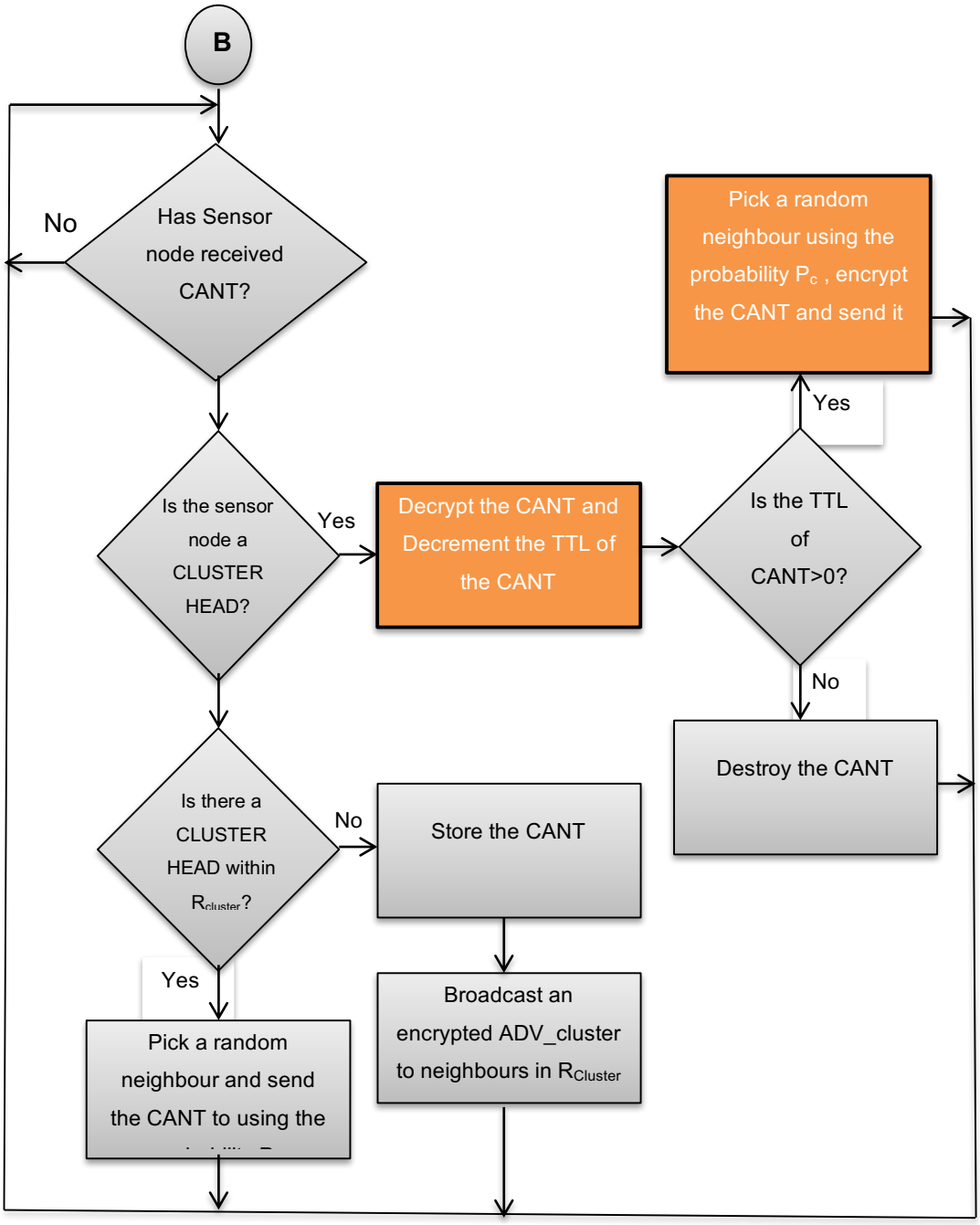


Figure 5-6 Flowchart for stage 2 (deployment stage) adapted from (Zhu et al., 2006)

5.8.3 Stage 3: Traffic forwarding

After Stage 2 is finished, all IDs and keys have been prepared, clusters and cluster heads have been set up and routing tables have been built. The aim of this stage is to deliver the data from the sensors to the base station both privately and securely, thus the title traffic forwarding. In this stage, all three privacy mechanisms, anonymity/pseudonymity, unlinkability and location privacy, are ensured as explained in Stages 1 and 2. To ensure that an adversary monitoring the network

would not relate the pseudonym with its related sensor nodes and threaten the source location privacy, the sensor nodes update their pseudonyms after receiving an AUTH for their transmission to the base station. However, this will create a problem of how to make the rest of the neighbour sensor nodes aware of the pseudonym update. One way to overcome this problem is to encrypt the pseudonym using the LEAP pairwise key and send it to all next hop neighbours in the routing table. However, this will cause a huge computational overhead (an encryption operation per neighbour) and communication overhead (sending the encrypted new pseudonym for each next hop neighbour on the list). Consequently, in this work, the new pseudonym is sent to the cluster head and the cluster head will be responsible for broadcasting the new pseudonym to the sensor nodes neighbours. This decreases the computational and communication overhead (compared to the previous solution) on the sensor nodes. The same idea is applied for the cluster heads where the base station is responsible for broadcasting the new pseudonym to the rest of the cluster heads in the network.

When a sensor node has to report data to the base station, this data should be sent to the cluster head so that it is later forwarded to the base station. However, to ensure the highest privacy and security, the data is encrypted using a LEAP individual key shared between this sensor node and the base station. This will ensure that the data is only accessed and comprehended by the base station and not the intermediate nodes. For multimedia data, the captured video and audio data is processed and either encrypted on-board (at the sensor level) or forwarded to the cluster head to be encrypted using the individual key shared between the cluster head and the base station.

Scalar data is captured by wearable, implanted, or environmental sensors to generate $S\_Sensor_i\ (Data)$ that will be sent to $Gateway_k$ where i denotes the sensor ID and k denotes the gateway ID. The data packet from the sensor to the gateway should contain the identification of the sensor, which generated the data ($Sensor_i$), the identification of the gateway authorized to receive the data ($Gateway_k$) and the captured data ($S\_Sensor_i\ (Data)$). The expected data packet should contain:

Sender	Receiver	Data
$Sensor_i$	$Gateway_k$	$S\_Sensor_i\ (Data)$

In the literature, each multimedia sensor typically processes the data, which it has captured, to detect the presence of a subject of interest, and then apply anonymity procedures (such as blurring the face and body gait in video, or changing the voice tone in audio). However, this is out of the scope of this research. In this work, the privacy preservation for multimedia data is achieved by the encryption of the multimedia data packets before sending them to the base station. The processed data from the multimedia sensors M_Sensor_j *(Data)* is transmitted to the cluster head CH_L where *j* denotes the sensor ID and *L* denotes the cluster head ID. The expected data packet should contain:

Sender	Receiver	Data
$Sensor_j$	CH_L	M_Sensor_j *(Data)*

At the gateway level, there are two possible scenarios that can be deployed. Scenario #1: scalar data is aggregated at the gateway level, then the data is encrypted using the LEAP protocol (Zhu et al., 2006). The data is encrypted using an individual key shared between the gateway $Gateway_k$ and the base station. This ensures the privacy and security of the personal data, as the base station can only interpret it. A compromise of the intermediate nodes will not endanger the transmitted data. The drawback of this scenario is more computational overhead at the gateway due to the encryption process, but the advantage of this scenario is the end-to-end privacy of the data (gateway to base station privacy)

The data packet should contain:

Sender	Receiver	Data
$Gateway_k$	CH_L	$E_{IK\text{-}gateway\text{-}base\ station}[Aggregated(S\_Sensor_{0\text{-}i}\ (Data))]$

where $E_{IK\text{-}gateway\text{-}base\ station}\ [Aggregated(S\_Sensor_{0\text{-}i}\ (Data))]$ denotes the encrypted aggregated data of the scalar sensors under $Gateway_k$ using a pairwise key shared between the gateway and the base station.

Scenario #2: scalar data is aggregated at the gateway $Gateway_k$, then it is sent to the cluster head CH_L without any privacy or security procedures. The drawback of this scenario is the possible privacy and security attacks on the transmitted plain data, which can be intercepted by an adversary while it is on its way to the cluster head CH_L. However, the advantage of this scenario is that it has less computational load on the gateway compared to Scenario #1.

The data packet should contain:

Sender	Receiver	Data
$Gateway_k$	CH_L	$Aggregated(S\_Sensor_{0\text{-}i}\ (Data))$

Scenario 1 is the one that was chosen for this research, as it offers better privacy and security, which is the aim of this research. It is important to note that all the IDs used (sensors, gateways and cluster heads) are pseudonyms that are constantly altered to hide the real ID of the sensors and suppress the identification of the sensors and their linkage to the real identity of the patient.

At the cluster head level, the scalar and multimedia data arrive at the cluster head CH_L to be sent to the base station. The cluster head CH_L should find an appropriate path to route the data to the base station using the AntSensNet (Cobo et al., 2010) routing protocol. The cluster head CH_L checks the pheromone table to route the data from itself to the base station using the intermediate cluster heads CH_M. Depending on the application requirements, the appropriate traffic class is chosen and its related QoS metrics are checked to see whether they are appropriate or not. The QoS metrics are: energy pheromone, delay pheromone, packet loss pheromone and available memory pheromone. The expiration time of the pheromone values must also be checked to determine whether the route is still valid or not. The chosen path is the one with the maximum pheromone for QoS metrics. The pseudocode outlining the general operation of the path finding is depicted in Figure 5-7. The pseudocode outlining the routing table update is presented in Figure 5-8. In Figure 5-8, each data packet is encrypted as one block (each field is not encrypted separately) to avoid the vulnerability to known plaintext attacks.

1: **If** (path found is unsatisfactory or expired)
2: **Then**
3: Update routing table (create FANT and wait for corresponding BANT)
4: **Else**
5: Choose most appropriate path (if more than one exist, choose the one with the highest normalised pheromone value (ψ)
6: **End if**
7: Prepare data packets (pad the packet to make the same size as ANTs)

Sender	Receiver	Next hop node	Data
CH_L	Base station	CH_M	Data packet received from $Gateway_k$

Figure 5-7 Pseudocode for path finding, adopted from (Cobo et al., 2010)

1: **FANT phase**
2: Create FANTS
3: Encrypt the content of each FANT using LEAP protocol using pairwise key between the CH_L issuing the ant and each next hop neighbour CH_M.

The expected data packet should contain:

Ant ID	$E_{PK\_}$ $CH_{L\_}$ CH_M(ant_type)	$E_{PK\_}$ $CH_{L\_}$ CH_M(ant_nodes)	$E_{PK\_}$ $CH_{L\_}$ CH_M(ant_hop_count)	$E_{PK\_}$ $CH_{L\_}$ CH_M(ant_info)

Where:
$E_{PK\_}$ $CH_{L\_}$ CH_M*(ant_type)* is the encrypted ant type using the LEAP pairwise key *PK* between the CH CH_L and the intermediate cluster head CH_M
$E_{PK\_}$ $CH_{L\_}$ CH_M*(ant_nodes)* is the encrypted list of nodes the ants has passed by so far using the LEAP pairwise key *PK* between the CH CH_L and the intermediate cluster head CH_M
$E_{PK\_}$ $CH_{L\_}$ CH_M*(ant_hop_count)* is the encrypted number of intermediate CHs the ant has passed by so far using the LEAP pairwise key *PK* between the CH CH_L and the intermediate cluster head CH_M
$E_{PK\_}$ $CH_{L\_}$ CH_M*(ant_info)* is the encrypted QoS metrics of the path so far using the LEAP pairwise key *PK* between the cluster head CH_L and the intermediate cluster head CH_M

4: Send the FANT to the next hop neighbour CH_M

5: **At each Cluster head**
6: Decrypt the FANT using the pairwise key between the source cluster head and this cluster head
7: Update the FANT fields
8: Re-encrypt the FANT
9: Send to the next hop cluster head neighbour

10: **At the base station**
11: Decrypt the FANT
12: Compare the data collected by the FANT with the application requirements saved at the base station
13: **If** (FANT collected data is appropriate) then
14: Generate a BANT and insert the FANT collected path info
15: Encrypt the BANT using pairwise key of next hop cluster head back to CH_L
16: Send BANT
17: **Else**
18: Discard FANT
19: **End if**

20: **BANT phase**
21: **At each cluster head**
22: Decrypt BANT,
23: Update the pheromone table
24: Re-encrypt using the pairwise of the next hop intermediate cluster head on its way to CH_L

Figure 5-8 Pseudocode for updating routing tables using ants, adopted from (Cobo et al., 2010)

FANTs are encrypted using the LEAP cluster key between the CH_L and all its next hop cluster head neighbours. However, if a cluster head is compromised, the cluster key must be updated for all the other cluster heads and this will impose more communication and computation overhead as opposed to the cancellation of a cluster head ID from the list of next hop cluster head neighbours for CH_L.

For the **transmission of multimedia data,** according to the AntSensNet protocol, the transmission of the video data is based on the Baseline algorithm by (Politis et al., 2008) combined with a multipath approach. When the multimedia data is captured, the multimedia is segmented and multiplexed. The general pseudocode for the transmission of multimedia data is depicted in Figure 5-9 .

```
1:  If (path to sink not existing or unsatisfactory) then
2:      Create VANTs (video ants)
3:      Encrypt VANTs (same procedure as FANTs)
4:      Send VANTS to next hop cluster head CH_L
5:  Else
6:      Determine a set of link-disjoint paths
7:      Apply the baseline packet scheduling to send packets to the sink
8:  End if
```

Figure 5-9 Pseudocode for the transmission of multimedia data, adopted from (Cobo et al., 2010)

A random number generator is used in all cluster heads and in the base station to schedule the generation of fake data packets which are the same size as the ants and have a varying time to live which ranges from 1 to N/2 (where N is the network size). Fake packets are encrypted to make them look the same as real packets. For increased privacy, fake ant packets can be created. However, this will enormously increase traffic overhead on the network, due to the broadcasting nature of communicating the ants between the cluster heads until the ant reaches the base station. Consequently, fake ants were not deployed in this research.

5.9 Fake packet generation in critical scenarios

In critical deployment scenarios, advanced fake packet generation technique may be required to make it harder for an adversary to threaten the privacy of the sub-system. The fake traffic of the whole sub-system can be based on tunable parameters at each level of the deployment starting from the gateway level until the

base station level. The base station is responsible for setting these tunable parameters at each level by sending encrypted messages (using individual keys) to each cluster head instructing it to set a particular fake to real ratio of messages. At the level of the cluster heads right above the gateway level (Level 1 as depicted in Figure 5-10), if the number of gateways is less than two, fake messages must be generated from this cluster head to trick the adversary into thinking that there is more than one gateway connected and thus increase the anonymity level of this gateway. In case there is more than one gateway connected to the cluster head, a ratio can be used to determine the amount of the fake traffic to the amount of the real traffic, for example a 1:2 ratio of fake to real packets respectively can be used. Figure 5-10 shows a possible layout of the gateways, cluster heads and base station. Selected cluster heads (surrounded by a red rectangle) can be set to generate fake traffic to trick an adversary into thinking that there are more gateways connected. Cluster heads generating fake traffic do not have to be directly connected to gateways and/or base stations. However, cluster heads with only one gateway must send fake traffic or else the adversary can easily link all packets to a particular gateway and blow away the anonymity of this gateway. The rest of the cluster heads can either be instructed to generate fake traffic or generate real traffic only.

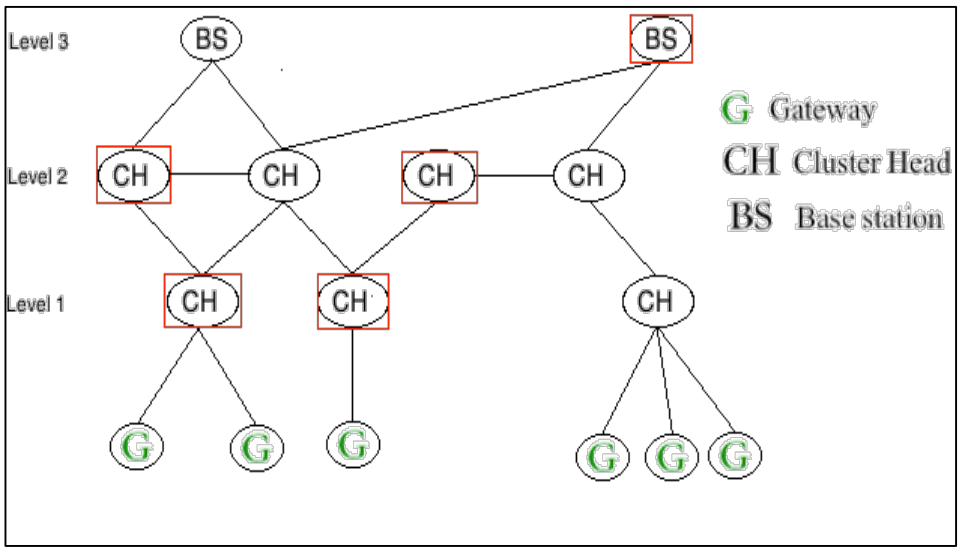


Figure 5-10 Generation of fake packets in critical scenarios

In order to protect the location privacy of the base station, a fake base station can be deployed to decrease the probability of an adversary capturing the real base station. The deployment of a fake base station creates two interleaved networks, a real network and a fake one as shown in Figure 5-11. The fake base station and the

tunable fake to real traffic can be used to even out the total traffic being forwarded to the base station. The fake traffic will be forwarded towards the fake base station and the real traffic will be forwarded towards the real base station. The fake traffic ratio of the cluster heads can be adjusted to make sure that the real traffic forwarded towards the real base station is almost the same amount as fake traffic forwarded towards the fake base station. This way the adversary will not be able to determine the location of the real base station. If the level of privacy of the whole sub-system needs to be increased, the ratio of the fake traffic can be increased and more fake base stations can be allocated.

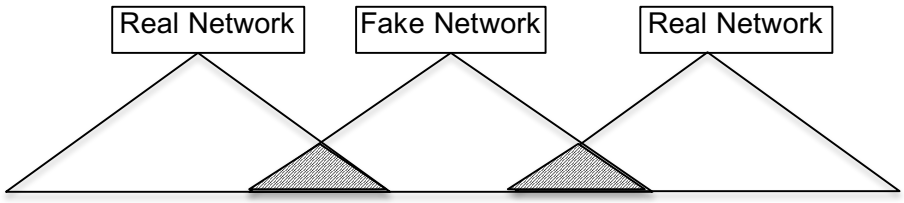


Figure 5-11 Interleaved fake and real network

5.10Summary

This chapter started with a brief overview of the different application scenarios of the WMSN-based healthcare sub-systems in the literature. Next, this chapter briefly outlined possible deployment scenarios of different applications: hospital, elderly house and battlefield. In addition, an overview of the building blocks of the proposed privacy preserving mechanisms for WMSNs in healthcare, which are: the routing protocol, the encryption key management technique, and the privacy mechanisms were presented. A general overview of routing protocols in WMSNs was presented followed by a comparison between a group of energy-efficient and QoS-aware routing protocols, to determine which routing algorithm best fits the requirements of this research work. The choice of the routing algorithm was based on the following criteria: hierarchal architecture, energy aware, reliable data delivery, query-based and event-traffic. The AntSensNet routing protocol is reported in the literature to be a hierarchal routing protocol that is designed especially for WMSNs; it can handle congestion control and delivers better video quality compared to other routing protocols. Thus, the AntSensNet routing protocol was chosen to be the underlying routing protocol for this research work.

The privacy of the AntSensNet routing protocol was assessed and it was concluded that the AntSensNet routing protocol suffered from linkability, location disclosure and identifiability privacy threats due to the absence of privacy and security services. Consequently, the following measures were suggested to defend the routing protocol against the identified privacy threats. To achieve unlinkability: 1) size correlation and encryption of the ants and the data messages is deployed; 2) pseudonyms are used to hide the real identity of the sensors and the cluster heads; 3) fake traffic is used to hide identity information and achieve unlinkability in cases when one gateway is connected to a cluster head. To achieve location privacy, the following privacy enhancing mechanisms are deployed: 1) fake traffic; 2) size correlation and encryption of the ants and the data messages. To achieve anonymity/pseudonymity, the following privacy enhancing mechanisms are used: 1) fake traffic; 2) pseudonyms.

The chapter presented details of how the three building blocks (the AntSensNet protocol, the encryption key management and the privacy mechanisms) were all integrated towards generating a privacy-aware WMSN-based healthcare subsystem. Flowcharts and pseudocode were used to provide a detailed view of the proposed subsystem. Finally, fake packet generation in critical scenarios was briefly discussed.

Although the proposed privacy measures are expected to enhance the level of privacy and withstand privacy threats, assessments should be conducted to determine the level of enhancement of privacy and predict the overhead added to the original routing protocol due to the introduction of the privacy mechanisms.

Chapter 6 Performance Assessment Methodology

6.1 Introduction

The previous chapter discussed how to introduce privacy measures into the AntSensNet protocol, to protect it against the three privacy threats targeted by this thesis: linkability, identifiability and location disclosure. The aim of this chapter is to discuss the assessment of the privacy of the WMSN-based healthcare sub-system after the introduction of the privacy measures.

In general, measuring the quality or effectiveness of privacy-enhancing technology is highly important to: (Danezis, 2013)

- allow the assessment and comparison between different privacy-enhancing mechanisms or system designs, or
- determine what needs to be improved and what impact this improvement will have on characteristics of the system, with regards to performance parameters such as reliability, usability, and overall privacy and security protection.

This chapter starts by a non-exhaustive overview of privacy metrics, which are used to quantify privacy. The word “metric” is used herein in its general sense prevalent in the privacy-enhancement literature, not in its strict mathematical sense (Wagner & Eckhoff, 2015). Next, a discussion of the choice of the particular metrics used in this work is outlined.

6.2 Brief Overview of Privacy Assessment Metrics

Different privacy services require different metrics for their assessment (or measurement). This section presents a brief sketch of privacy metrics for location privacy, unlinkability and anonymity/pseudonymity.

6.2.1 Metrics for location privacy

Metrics such as uncertainty-based, error-based and similarity-based metrics (such as k-anonymity-based metric) have been used in the literature to assess location privacy (Garitano et al., 2015) (Wagner & Eckhoff, 2015). Common uncertainty-based metrics are entropy metric and anonymity set size metric. The entropy metric uses Shannon’s information theory to assess the location privacy, based on the notion that for a network of N nodes, $log_2 N$ bits are needed to identify a node in the

network (in case of an equiprobable distribution of node identities), and that the overall entropy will decrease if the distribution is not equiprobable (Beresford & Stajano, 2003). The anonymity set size metric is frequently used in the literature. It can be viewed as the size of the “crowd” (corresponding to different locations) in which a certain location can blend into (Wagner & Eckhoff, 2015).

An error-based metric measures the difference between the true location of a node and the location that an adversary has estimated. Error-based metrics can be further categorized into clustering error, probability of error and distortion-based metrics (Shokri et al., 2010). Clustering error refers to measuring the success of an adversary to cluster the observed events into partitions and link these partitions to users. Probability of error refers to the probability of an adversary erroneously linking objects to users or identifying users. Distortion-based metrics refer to measuring the distance between the actual location of a user and those trajectories constructed by adversaries to predict the location of the users (Shokri et al., 2010).

k-anonymity refers to the concept in which the location of a subject is considered k-anonymous if and only if the information of the location is indistinguishable from the information of at least k-1 other locations of other clients (Gedik & Liu, 2008). The “k” number in k-anonymity can be used to assess the level of location privacy, such that the higher the k, the stronger the privacy protection (Krumm, 2009). (Reza et al., 2011) proved, quantitatively, that there are situations where the k-anonymity and the entropy metrics may not be appropriate for the quantification of location privacy and proposed a new measurement for location privacy called “location privacy meter” which considers different types of attacks and information disclosure.

6.2.2 Metrics for unlinkability

To measure unlinkability, (Fischer et al., 2008) introduced the expected distance unlinkability metric to estimate the error made by an adversary when linking items of interest to particular senders, based on the uncertainty of the an adversary. Other measures were introduced in the literature, such as: partitioning or equivalence classes (which refers to grouping messages together if they are from the same originator or sender), Mix Zones (a discrete model used to analyse the traceability of scenarios), and k-unlinkability (used to assess the unlinkability of data sets that are partially obscured) (Fischer et al., 2008). In (Nohara et al., 2005), conditional entropy and mutual information were used to measure the level of unlinkability against an attacker.

6.2.3 Metrics for anonymity

Several metrics have been adopted in the literature to measure the level of anonymity, such as the anonymity-set size, crowds-based metrics and entropy-based metrics (Danezis, 2013) (Andersson & Lundin, 2008).

The anonymity-set size refers to the number of senders and receivers of the messages within the network. The larger the anonymity set, the higher the level of the anonymity (Murdoch, 2014). An ideal anonymous network should have equal probability distribution over the senders and receivers of a particular message, which is hard to achieve in real life if the flow of traffic in the network is observed over time by adversaries, which allows them to narrow down the anonymity set and the probabilities of establishing the senders or receivers of messages (Murdoch, 2014). Although an adversary may learn that one participant in the network has a very high probability of being a sender of a certain message, yet the anonymity set will not reflect that because it only depends on the cardinality of the network.

Consequently other measures are often adopted to measure anonymity (Murdoch, 2014). The Crowds-based metric relies on the computation of the probability that a certain action was performed by a specific participant in the network (Danezis, 2013). However, this measure was claimed to be weak (Danezis, 2013). The use of entropy was proposed by (Diaz et al., 2002) and (Serjantov & Danezis, 2003) to quantify the degree of anonymity. It has been shown to be a useful way for quantifying the level of anonymity for systems where the anonymity set metric was not accurate enough to use (Acharya & Younis, 2010) (Danezis, 2013).

6.3 Privacy assessment methodology

6.3.1 Basis for the choice of privacy metrics

According to (Wagner & Eckhoff, 2015), the choice of a particular privacy metric can be determined according to four main considerations: adversary models, data sources, inputs for the metrics and output measures. The taxonomy proposed by (Wagner & Eckhoff, 2015) was adopted in this research, to justify the choice of particular privacy metrics.

Adversary model: Determining the adversary model is an important decision because the strength of the adversary affects the level of privacy (Wagner & Eckhoff, 2015). For example, the degree (level) of anonymity afforded by privacy-enhancing technology can be estimated by the adversary by assigning probabilities to users, or network nodes, as being the originators of the messages under attack

(Diaz et al., 2002). Consequently, the level of anonymity will change if the adversary's model changes, which makes it important to specify the properties of the adversary before estimating the level of the anonymity (Diaz et al., 2002). The adversary can be viewed as having a combination of the following properties (Diaz et al., 2002) (Wagner & Eckhoff, 2015):

- Internal or external adversary: An internal adversary is able to manage a part of the system such as one or more nodes, which makes him/her able to control these nodes and access the data stored in them. On the other hand, an external adversary is only able to listen to, or tamper with, the communication channels.
- Passive or active adversary: A passive adversary eavesdrops on the communication or the information stored in the nodes, without making any alterations. On the other hand, an active adversary jeopardizes messages and internal information.
- Local or global adversary: A local adversary can only attack a part of the network, whereas a global adversary has access to the whole network.
- Static or adaptive adversary: A static adversary follows the same strategy and does not change it throughout his/her attack, whilst an adaptive adversary adapts his/her attack as he/she gains more knowledge of the system under attack.
- Prior knowledge: Adversaries may have previous knowledge (e.g. in the form of probability distributions) of the system that they are attacking.
- Resources: Different adversaries may utilize different resources, such as computational powers.

In this research work, it is assumed that wireless sensor nodes are tightly coupled to a patient (either worn or implanted inside the patient) or mounted to the walls of the hospitals (in case of the video sensors, audio sensors and cluster heads). Consequently, the adversary is assumed to be an external adversary. In addition, an adversary is assumed to passively listen to the traffic between the source and the destination, aiming at identifying the senders and the receivers of the messages. As a result, an adversary is assumed to be passive. An adversary is also assumed to have no prior knowledge of the system such as the number of patients (gateways). In addition, an adversary can either be local, listening to a portion of the network (e.g. listening to messages coming out of one or more cluster heads), or global, monitoring the whole traffic in the network.

Adversary attack models, such as traffic analysis attacks, routing tables inspections and packet tracing attacks, can be deployed by adversaries to discover the location of the base station (Riosa et al., 2015) (Jian et al., 2007). Traffic analysis attacks are based on collecting and extracting information based on the observation of the traffic flow in the network. Detecting the location of the base station can be based on the idea that the nodes near to the base station forward a greater volume of traffic compared to those further away (Jian et al., 2007) (Deng et al., 2006). Adversaries performing traffic analysis attacks are categorized based on the methods used to extract information and on the strength of the eavesdropping capabilities (Riosa et al., 2015). A traffic analysis attack is thought to consume a long time as an adversary should move from one location to another and spend a long time at each location to collect enough information about the traffic rates. Routing table inspections are based on the information retrieved from the captured sensor nodes. An adversary capturing several nodes can passively study their routing tables and learn the location of the base station (Riosa et al., 2015). In packet tracing attacks, the attacker starts at the sender's location and performs hop-by-hop trace analysis until it reaches the receiver's location (Jian et al., 2007). In some cases, adversaries are not passive and attempt to reprogram captured nodes and turn them into malicious nodes (Deng et al., 2006)).

In the solution adopted in this work, the privacy-aware WMSN based healthcare sub-system is expected to be protected against packet tracing attacks due to the deployment of pseudonyms and fake traffic, which should make it hard for an adversary to trace a particular packet to the base station. In addition, in this work, the base station should generate fake traffic and re-forward received packets to nodes further away to trick adversaries into thinking that the base station is just another network node.

Data sources: According to (Wagner & Eckhoff, 2015), data sources specify the sources of the data that need to be protected and how an adversary can capture the data. Data sources can be categorized into published data, observable data, re-purposed data and all other data. Published data is data that has been made public willingly by its owners. Observable data is transient data that an adversary needs to be present to be able to capture such as in situations where adversaries are passively listening to communication channels to detect the senders and receivers. Re-purposed data is data used for a reason different from what it was captured for. All other data refers to the unprotected data that was not made public, not observed

by adversaries, not re-purposed and is not meant to be captured by adversaries such as data obtained by hacking into systems (Wagner & Eckhoff, 2015).

In this research, the data to be protected is transferred from the source (sensors) to the destination (base station) through a wireless sensor network, in which an adversary can eavesdrop the messages to determine their source and/or destination. Consequently, based on the classification of (Wagner & Eckhoff, 2015), the data sources are considered observable.

Input for the metrics: The choice of the privacy metrics depends on the possible input that can be fed into them. In this work, the input to the privacy metrics is an estimate of what would be computed by an adversary. The estimate is in the form of a probability distribution of whether the captured messages originated from a particular sender, or is intended for a particular receiver.

Output measures: The classification of the output measures is useful because a single metric does not assess the overall level of privacy and a thorough assessment can be achieved by utilizing different measures belonging to different output categories (Wagner & Eckhoff, 2015). In general, there are eight categories for the output measures of the privacy metrics: uncertainty, information gain or loss, similarity or diversity, indistinguishability, probability of success by the adversary, error, time and accuracy or precision (Wagner & Eckhoff, 2015).

Consequently, the privacy metrics that best assess the privacy services discussed in this research work (anonymity/pseudonomity, unlinkability and location privacy) should be those for observable data sources, and accepting as input some probability estimates from the perspective of the adversary. Based on these parameters and on the taxonomy proposed by (Wagner & Eckhoff, 2015), the uncertainty-based, information gain or loss, time-based, adversary's success probability, accuracy and indistinguishable-based metrics can be applied to this research work.

The selection of particular privacy metric(s) can be guided based on eight questions (Wagner & Eckhoff, 2015): what privacy aspects need to be quantified? What are the adversary models? Which data sources need protection? Which type of data is available to compute the metric? What are the target audiences (users) of the output of the metrics? Which metrics were used in the related work? What is the quality of the metric used? Can the privacy metrics be implemented? Putting into consideration all previous eight questions and in particular question six (which

metrics were used in the related work?), it is clear that uncertainty-based metrics (namely entropy) have been adopted to assess different privacy services such as unlinkability and anonymity/pseudonymity. Accordingly, the uncertainty-based privacy metrics were chosen to assess the level of unlinkability and anonymity/pseudonymity of this research work. From the uncertainty-based metrics, the normalized entropy and the conditional entropy (Nohara et al., 2005) were chosen to assess the anonymity/pseudonymity and unlinkability respectively. Both metrics depend on the probability of the estimates of the adversaries, which can be computed in this research. Anonymity set size was chosen to assess location privacy. Furthermore, in order to obtain better privacy assessment, another category, the information gain or loss metric, was used to assess the level of privacy.

6.3.2 Further discussion of the chosen privacy metrics

6.3.2.1 Uncertainty-based privacy metrics

The main concept of information entropy is the measure of the information found in a given distribution of probabilities (Diaz et al., 2002). In privacy assessment metrics, entropy is employed to measure the level of privacy of the users of a system associated with the predictions of a certain adversary (Diaz et al., 2002) (Wagner & Eckhoff, 2015). Shannon's entropy has been employed in the literature as a privacy metric for different privacy services such as anonymity (as in (Nezhad et al., 2008)), unlinkability (as in (Deng et al., 2005) (Mahmoud 2012)), and location privacy (as in (Mahmoud, 2012)(Nezhad et al., 2008)). All these papers have deployed fake traffic to provide anonymity and/or unlinkability and/or location privacy and have used entropy to assess the level of their proposed privacy preservation techniques.

The entropy of a system is compared with its maximum entropy to assess the amount of information an adversary might gain after an attack (Diaz et al., 2002). In the context of anonymity, after an attack, the entropy of a system which has an anonymity set *X* can be defined as (Diaz et al., 2002):

$$H(X) = -\sum_{i=1}^{N} p_i log_2(p_i) \quad (8)$$

where p_i refers to the probability that an adversary assigns user *i* as being the true sender of a particular message in a network of *N* members. Thus, Equation (9) calculates the maximum entropy H_M that can be calculated for the *N* honest (uncompromised) senders (Diaz et al., 2002).

$$H_M = log_2(N) \quad (9)$$

The information that an adversary can learn is the difference between the maximum entropy H_M and the entropy of the system after an attack $H(X)$. Consequently, the normalized degree of anonymity can be denoted by (Diaz et al., 2002):

$$d = \frac{H(X)}{H_M} \quad (10)$$

where *d* ranges from 0 ≤ d ≤ 1. If *d* = 0, this means that a particular user is the sender of a message with a probability p_i of 1. On the other hand, *d* = 1 means that all users have an equal probability of being the sender of a message (Diaz et al., 2002).

Extension to the entropy measures, such as joint entropy and conditional entropies are used to measure the level of uncertainty in the joint distribution of two random variables and to measure the level of uncertainty in the conditional distribution of two variables respectively (Bergstrom, 2008). In this research work, conditional entropy will be used to assess unlinkability.

The equation for the joint entropy for two variable X and Y can be expressed as (Bergstrom, 2008)

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y) log_2\, p(x,y) \quad (11)$$

The equation for the conditional entropy for two variables X and Y can be expressed as (Wagner & Eckhoff, 2015)

$$H(X|Y) = -\sum_{x\in X, y\in Y} p(y,x) log_2 p(x|y) \quad (12)$$

6.3.2.2 Information loss or gain privacy metrics

The information gain metrics aim to assess the amount of information that an adversary gains (i.e. the higher the information gain, the less the level of privacy) (Wagner & Eckhoff, 2015). One measure of the information loss or gain privacy metric is the relative entropy. Relative entropy is a measure of the amount of information discovered by an adversary by measuring the distance between the true distribution and the adversary's estimate (Wagner & Eckhoff, 2015). The higher the relative entropy, the better the privacy is. If p denotes the true distribution and q

denotes the adversary's estimate, then the relative entropy can be expressed as (Wagner & Eckhoff, 2015):

$$priv_{RLE} \equiv D_{KL}(p||q) = \sum_{x \in X} p(x) log_2 \frac{p(x)}{q(x)} \quad (13)$$

6.3.2.3 *Anonymity set size*

The anonymity set size will be adopted in this research work to measure location privacy. Anonymity set size refers to the count of potential users that can be a target person *t* (Wagner & Eckhoff, 2015). At each cluster head, the anonymity set size can be expressed by (Wagner & Eckhoff, 2015):

$$priv_{ASS} \equiv |AS_t| \quad (14)$$

Equation (14) represents the number of potential sources of data (number of gateways) under each cluster head. At the base station, the anonymity set size $\bar{S}$ is expressed by (Buttyán et al., 2006):

$$\bar{S} = \sum_{i=0}^{l} \frac{|P_i|^2}{N} \quad (15)$$

where *N* is the total number of members, l is the number of anonymity sets and P_i denotes the number of members belonging to an anonymity set.

6.3.3 Experimental method

The experiments depicted in the Chapter 8 which is titled "Assessment of the Enhancement of Anonymity, Unlinkability and Location Privacy Due to Fake Traffic" were conducted using the NS2 simulator. Each experiment was repeated 10 times and the experiment results were recorded. Each time a random seed was used by the NS2 simulator to generate random traffic and the entropy at each cluster head was recorded together with the amount of time required to receive the same number of messages. At the end of each experiment, the standard deviation, mean, 95% confidence interval and margin of error were computed for all experimental results.

6.4 Summary

In this chapter, a brief overview of privacy assessment metrics and privacy assessment methodologies were discussed. The privacy assessment metrics: anonymity set size, conditional entropy and; information loss or gain and normalized entropy were deployed in this research work to quantity the location privacy,

unlinkability and anonymity respectively. Finally, a brief outline of the experimental method deployed in this research was presented.

The next chapter will present analysis of the overheads due to the introduction of privacy-enhancement mechanisms to the ant routing algorithm.

Chapter 7 Analysis of Overheads Due to Privacy-Enhancement of the Ant Routing Algorithm

7.1 Introduction

Chapter 5 which is entitled “Privacy-Aware Ant Routing Algorithm for WMSNs” presented deployment scenarios reported in the literature for WMSN-based healthcare subsystems and discussed how to introduce privacy measures into the AntSensNet protocol, to protect it against the three privacy threats targeted by this thesis: linkability, identifiability and location disclosure. However, the introduction of the privacy measures into the AntSensNet protocol in the deployment scenarios is expected to add overheads to the WMSN-based healthcare sub-system. Consequently, the aim of this chapter is to investigate the overhead due to the addition of the privacy enhancing mechanisms.

This chapter starts by simulation experiments, using the NS2 simulator, to assess the overhead due to the introduction of the privacy and security measures as discussed in the Chapter 5. Next, this chapter presents a quantitative analysis based on the evaluation of the computation, network messages and storage overhead of the privacy-aware WMSN-based healthcare sub-system in three selected operational scenarios: hospital, elderly house and battlefield (discussed in Chapter 5). The quantitative analysis presented in this chapter is conducted in such a way that it can be applied to any number of sensors (scalar and multimedia) and cluster heads connected to the network and not only to the number of sensors depicted in the figures of results to ensure that this analysis can be applied other scenarios and not just those experimented in this research.

7.2 Simulation of Application Scenarios

In this section, the three application scenarios (hospital scenario, elderly house scenario and battlefield scenario) discussed in Section 5.2 entitled “Application Scenarios for WMSN in Healthcare” are simulated using NS2 network simulator.

The data rates of the sensors deployed in these simulations depend on the type of sensors and where they are deployed (in-body deployment and/or on-body deployment) (Movassaghi et al., 2014). According to (Movassaghi et al., 2014), the sensors of a WBSN have a transmission bit rate ranging from 1 Kbps (kilobits per second) to 10 Mbit/s (megabits per second) depending on the type of sensors. For in-body applications, the data rate can range from few Kbps (as in the glucose sensor) to few Mbps (as in the endoscope capsule). The data rate of the on-body

sensors can range from a few bps to a few Kbps (Movassaghi et al., 2014). Accordingly, in the simulation of the experiments described below, the traffic generated by the gateways is consistent with the average bit rate of the in-body sensors (few Kbps to few Mbps) and on-body sensors (bps to kbps) as presented by (Movassaghi et al., 2014). In the simulation experiments, the average bit rate traffic generated by in-body and on-body sensors is 40Kbps.

The traffic generated by the multimedia sensors is consistent with the multimedia traffic rate in the AntSensNet routing protocol paper (Cobo et al., 2010) which is Constant Bit Rate (CBR) traffic of four packets per seconds. The same rate was deployed in WMSN-based NS2 experiments in the literature such as (Akhlaq & Sheltami, 2012), (Adhyapak & Laturkar, 2018) and (Zhang et al., 2005).

The network components deployed in the three scenarios are as follows: Wearable (W), Implanted (I) and Environmental (E) sensors are defined as Transmission Control Protocol (TCP) traffic generator objects with a File Transfer protocol (FTP) application running on top of them as shown in Figure 7-1. Multimedia sensors (Video (V) and Audio (A) sensors) are defined as User Datagram protocol (UDP) traffic generator objects with CBR application running on top of them as shown in Figure 7-2. The multimedia sensors are set to produce four packets per second. The size of each packet is 1024 bytes.

All network components are wireless nodes. Data is captured by gateways (G) from the sensor nodes and is sent to the cluster heads (CH). The data collected by the cluster heads is sent to the base station (BS).

```
# Set a TCP connection between node_(19) and node_(9)
set tcp_(0) [new Agent/TCP/Newreno]
set sink_(0) [new Agent/TCPSink]
$tcp_(0) set class_ 2
$ns attach-agent $node_(19) $tcp_(0)
$ns attach-agent $node_(9) $sink_(0)
$ns connect $tcp_(0) $sink_(0)
        set ftp_(0) [new Application/FTP]
        $ftp_(0) attach-agent $tcp_(0)
        $ftp_(0) set rate_ 40Kb
        $ns at 0.0 "$ftp_(0) start"

# Set a TCP connection between node_(20) and node_(9)
set tcp_(1) [new Agent/TCP/Newreno]
set sink_(1) [new Agent/TCPSink]
$tcp_(1) set class_ 2
$ns attach-agent $node_(20) $tcp_(1)
$ns attach-agent $node_(9) $sink_(1)
$ns connect $tcp_(1) $sink_(1)
        set ftp_(1) [new Application/FTP]
        $ftp_(1) attach-agent $tcp_(1)
        $ftp_(1) set rate_ 40Kb
        $ns at 0.0 "$ftp_(1) start"
```

Figure 7-1 A sample code for the definition of the wearable, implanted and environmental sensor nodes in NS2

```
set udp0 [new Agent/UDP]
$ns attach-agent $node_(15) $udp0
set null0 [new Agent/Null]
$ns attach-agent $node_(7) $null0
$ns connect $udp0 $null0
$udp0 set fid_ 10
        set cbr0 [new Application/Traffic/CBR]
        $cbr0 attach-agent $udp0
        $cbr0 set packetSize_ 1024
        $cbr0 set interval_ 4
        $ns at 0.0 "$cbr0 start"
        $ns at 500.0 "$cbr0 stop"

set udp1 [new Agent/UDP]
$ns attach-agent $node_(16) $udp1
set null1 [new Agent/Null]
$ns attach-agent $node_(7) $null1
$ns connect $udp1 $null1
$udp1 set fid_ 10
        set cbr1 [new Application/Traffic/CBR]
        $cbr1 attach-agent $udp1
        $cbr1 set packetSize_ 1024
        $cbr1 set interval_ 4
        $ns at 0.0 "$cbr1 start"
        $ns at 500.0 "$cbr1 stop"
```

Figure 7-2 A sample code for the definition of the multimedia sensors in NS2

Description of analysis: The NS2 model of each of the three scenarios is first run for 50 seconds. The generated NS2 trace file is analyzed to calculate average end-to-end delay, throughput, number of generated packets, number of received packets and percentage of Packet Delivery Ratio (PDR). Figure 7-4 and Figure 7-3 depict a sample of the awk code used for the calculation of these analyses. Each

scenario is run 10 different times with random seeds to calculate the mean, standard deviation, margin of error, upper bound and lower bound using 95% confidence interval. The same procedure is repeated for each scenario for different periods of time ranging from 50 seconds to 500 seconds. The results of each run are recorded and plotted in a Microsoft Excel sheet. Refer to Appendices A, B and C for the recorded results of each run for the hospital scenario, elderly house scenario and battlefield scenario respectively.

```
BEGIN {
    seqno = -1;
    droppedPackets = 0;
    receivedPackets = 0;
    count = 0;
    NofSent = 0;
}
{
        if($4 == "AGT" && $1 == "s" && ( $7 == "tcp" || $7 == "cbr" )) {  NofSent++ ;   }

        if($4 == "AGT" && $1 == "s" && seqno < $6) { seqno = $6; }
                else if(($4 == "AGT") && ($1 == "r")) { receivedPackets++;      }
                else if ($1 == "D" && $7 == "tcp" && $8 > 512){ droppedPackets++;  }

    #end-to-end delay

    if($4 == "AGT" && $1 == "s") {

          start_time[$6] = $2;

    } else if(($7 == "tcp") && ($1 == "r")) {

        end_time[$6] = $2;

    } else if($1 == "D" && $7 == "tcp") {

          end_time[$6] = -1;

    }

}
END {

    for(i=0; i<=seqno; i++) {

          if(end_time[i] > 0) {

              delay[i] = end_time[i] - start_time[i];
              count++;

        } else { delay[i] = -1; }

    }
    for(i=0; i<=seqno; i++) { if(delay[i] > 0) { n_to_n_delay = n_to_n_delay + delay[i];}

    }

   n_to_n_delay = n_to_n_delay/count;
print "Total Count =            = " count;

    print "\n";
    print "GeneratedPackets            = " seqno+1;
    print "ReceivedPackets             = " receivedPackets;
    print "Packet Delivery Ratio      = " receivedPackets/(seqno+1)*100 "%";
    print "Total Dropped Packets = " droppedPackets;
    print "Average End-to-End Delay    = " n_to_n_delay * 1000 " ms";
    print "\n";
```

Figure 7-3 AWK code for the calculation of generated packets, received packets, packet delivery ratio, total dropped packets and average end-to-end delay

```
{
  event = $1
  time = $2
  node_id = $3
  pkt_size = $8
  level = $4

  if (level == "AGT" && event == "s") {
    sent++
# Note the change in the next line. This initializes the startTime with the first encountered "time" value.
    if (!startTime || (time < startTime)) {
      startTime = time
    }
  }

  if (level == "AGT" && event == "r") {
    receive++
    if (time > stopTime) {
      stopTime = time
    }
    recvdSize += pkt_size
  }

}

END {
  PDR = (receive/sent) * 100;
  printf("sent_packets\t %d\n",sent);
  printf("received_packets %d\n",receive);
  printf("PDR %.2f \n",PDR);
  printf("Average Throughput[kbps] = %.2f\tStartTime=%.2f\tStopTime = %.2f\n", (recvdSize/(stopTime-
startTime))*(8/1000),startTime , stopTime);
}
```

Figure 7-4 AWK code for the calculation of throughput

7.2.1 Hospital scenario simulation

Simulation model and components: The hospital scenario discussed in Figure 5-1 is modelled and simulated in NS2. Figure 7-5 shows the representation of the hospital scenario in the NS2 Network Animator (NAM).

The network in Figure 7-5 is composed of 1 base station (BS), 7 cluster heads (CH), 1 relay node (RN), 6 gateways (G), 14 wearable (W) / implanted (I) / environmental (E) sensors and 4 video/audio sensors. Following the scenario in Figure 5-1, data generated from wearable and implanted sensors (nodes 19 and 20) is collected by the gateway (node 9). The data collected by the gateway (node 9) and the data generated by the video and audio sensors (nodes 15 and 16) is routed to cluster head (node 7). The data is then sent to the base station (node 0) by multi-hop routing through 2 cluster heads (nodes 4 and 1).

The data collected by gateway (node 12) from wearable and implanted sensors (nodes 28 and 29), the data collected from video and audio sensors (sensors 17 and 18) and the data collected by gateway (node 10) from wearable, implanted and environmental sensors (nodes 21, 22 and 23) is sent to cluster head (node 5). Cluster head (node 5) sends the data to the base station (node 0) through cluster head (node 2).

The data collected by gateway (node 13) from wearable and implanted sensors (nodes 26 and 27) and the data collected by gateway (node 14) from wearable, implanted and environmental sensors (nodes 30, 31 and 32) is sent to relay node (node 8). The data from relay node (node 8) and from gateway (node 11) [collected from wearable and implanted sensors (nodes 24 and 25) is sent to base station (node 0) through cluster heads (nodes 6 and 3).

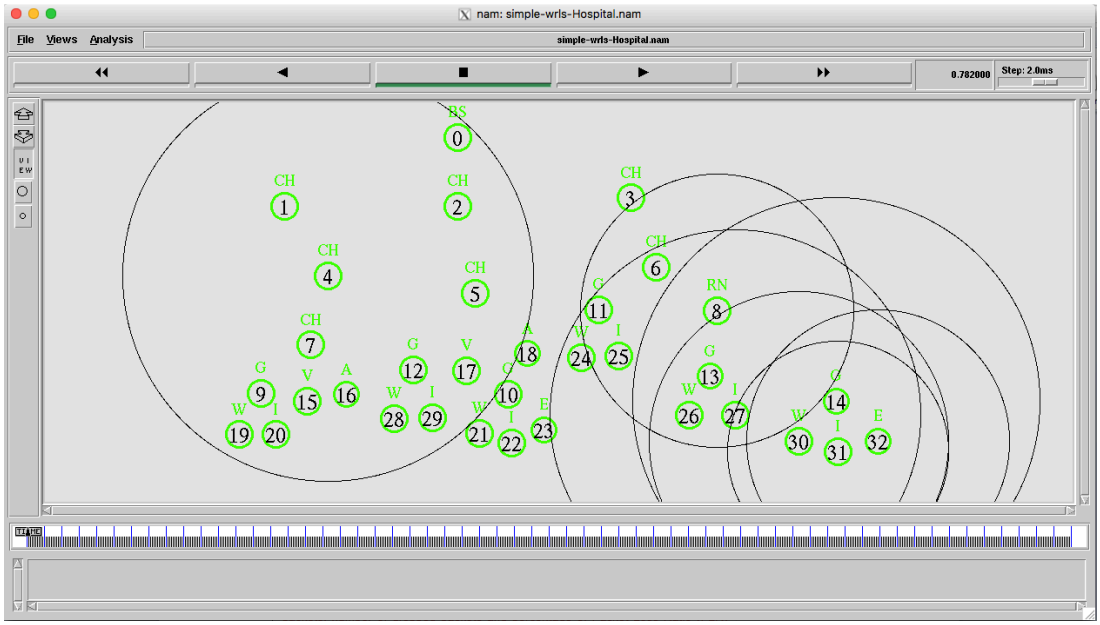


Figure 7-5 NS2 Hospital Scenario on NAM

Results and discussion: The mean values for average end-to-end delay, percentage of PDR, throughput, number of generated packets, number of received packets, number of dropped packets and percentage of Packet Loss Ratio (PLR) are shown in Table 7-1. Figure 7-6, Figure 7-7, Figure 7-8 and Figure 7-9 depict a plot of the results presented in Table 7-1. It is clear from these results, as the simulation time increases and more packets are generated, the throughput increases. In addition, the plot shows that as the throughput increased, the PDR increased.

Table 7-1 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets and percentage of PLR for different simulation times for NS2 hospital scenario simulation

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio
50	1246.423	92.256	599.232	7053	6508	31	7.744
100	1233.455	94.626	613.687	14412	13638	57	5.374
150	1244.398	95.321	614.404	21620	20614	73	4.679
200	1267.437	95.312	607.537	28511	27180	114	4.688
250	1290.177	95.853	612.990	35948	34463	122	4.147
300	1284.695	95.822	613.357	43159	41355	153	4.178
350	1266.950	96.034	620.286	45077	43298	160	3.966
400	1276.281	96.250	627.456	58817	56615	200	3.750
450	1269.052	95.950	628.511	66255	63571	210	4.050
500	1288.700	96.533	628.515	73318	70778	266	3.467

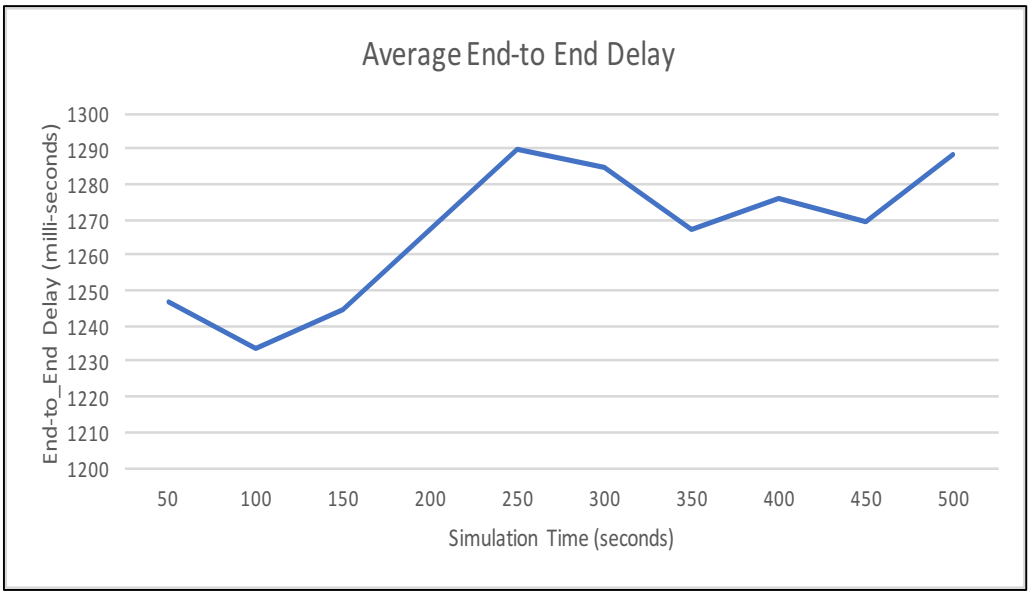


Figure 7-6 Mean Average end-to- end delay for different simulation times for NS2 hospital scenario

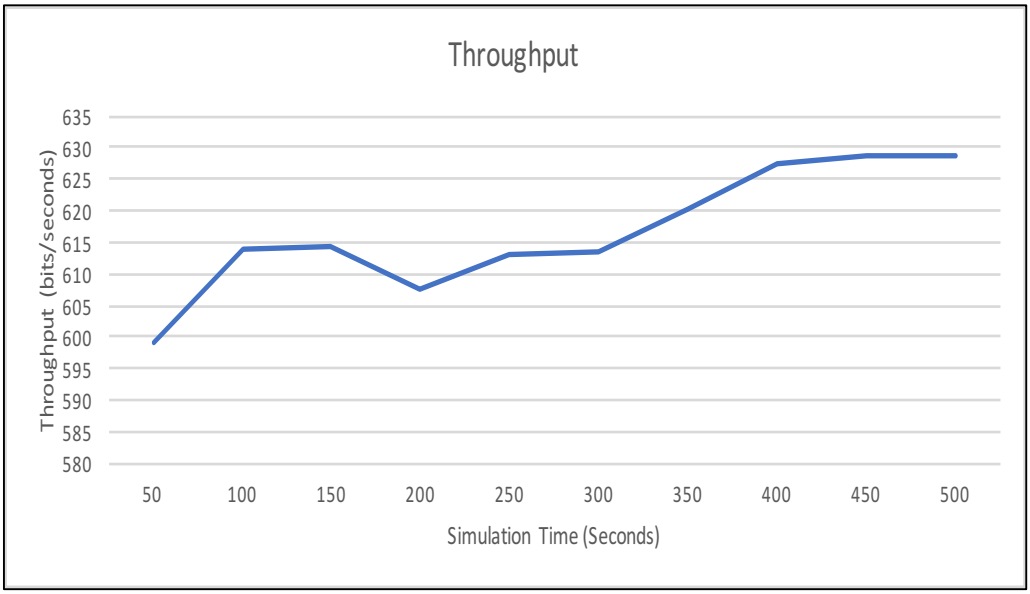


Figure 7-7 Mean throughput for different simulation times for NS2 hospital scenario

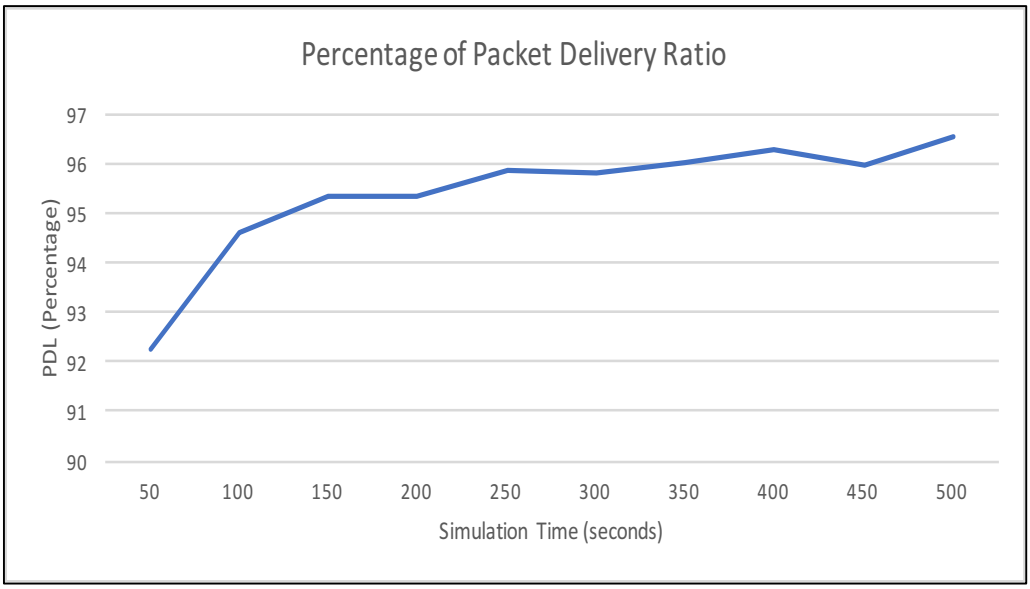


Figure 7-8 Mean percentage of packet delivery ratio for different simulation times for NS2 hospital scenario

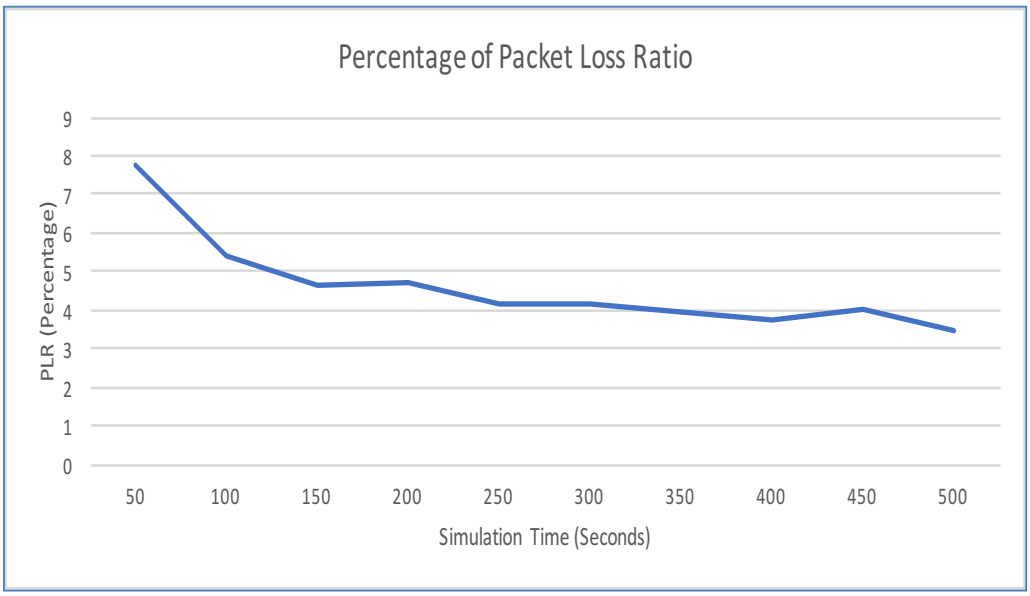


Figure 7-9 Mean percentage of packet loss ratio for different simulation times for NS2 hospital scenario

7.2.2 Elderly house scenario simulation

Simulation model and components: The elderly house scenario discussed in Figure 5-2 is modelled and simulated in NS2. Figure 7-10 shows the representation of the elderly house scenario in the NAM.

The network model in Figure 7-10 is composed of 1 base station (BS), 5 cluster heads (CH), 2 gateways (G), 6 wearable (W) / implanted (I) /environmental (E) sensors and 6 video/audio sensors. Following the scenario presented in Figure 5-2, the network is connected as follows: Gateway (node 6) collects data from wearable and implanted sensors (nodes 7 and 8). The data from video and audio sensors (nodes 9 and 10), environmental sensor (node 11) and gateway (node 6) is sent to cluster head (node 5). The data from cluster head (node 5) is routed to cluster head (node 4), which collects data from gateway (node 12), video and audio sensors (nodes 15 and 16). Gateway (node 12) collects data from wearable and implanted sensors (nodes 13 and 14). The data from cluster heads (nodes 4 and 5) is sent to cluster head (node 3), which collects data from environmental sensor (node 17), video and audio sensors (nodes 18 and 19). The data from cluster heads (nodes 5, 4 and 3) is routed to base station (node 0) through cluster heads (nodes 2 and 1).

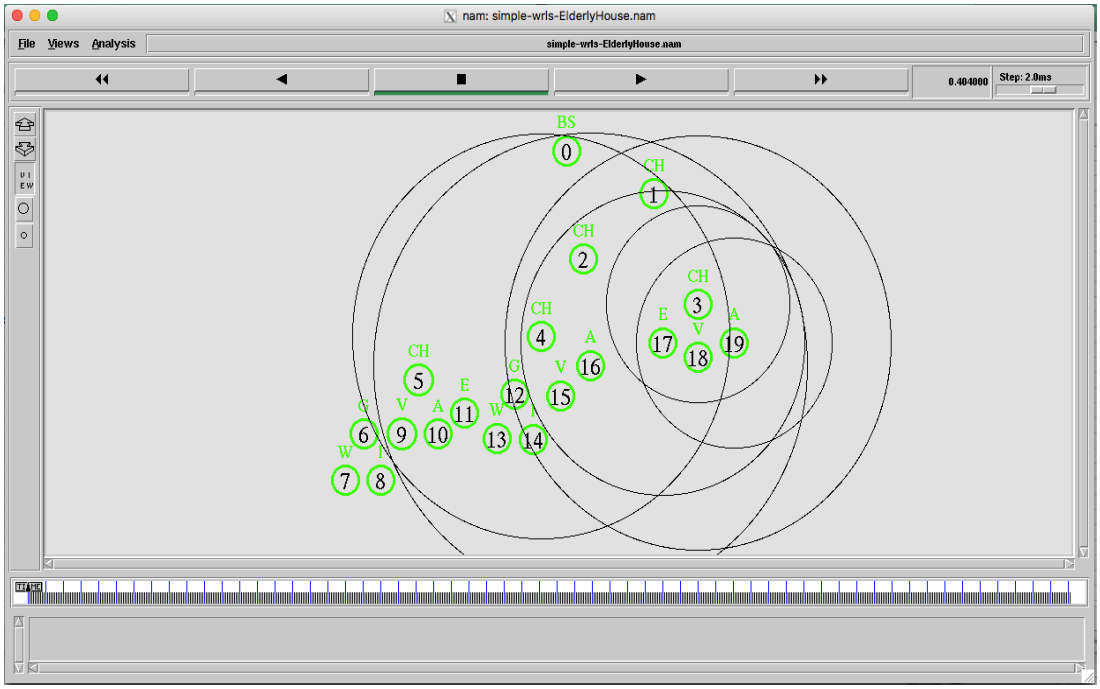


Figure 7-10 NS2 elderly house scenario model on NAM

Results and discussion: The mean values for average end-to-end delay, percentage of PDR, throughput, number of generated packets, number of received packets, number of dropped packets and percentage of PLR are shown in Table 7-2. Figure 7-11, Figure 7-12, Figure 7-13 and Figure 7-14 show the plot of the results in Table 7-2. It is clear from Figure 7-12 that unlike the previous scenario (hospital scenario), throughput is decreasing as more packets are generated. This is because all data generated by all sensors (scalar and multimedia) is collected and sent through one path (from nodes 3 to 2 to 1 to 0). This has caused more packets being dropped in this scenario compared to the hospital scenario, which explains the drop is the mean throughput.

Table 7-2 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets and percentage of PLR for different simulation times for NS2 elderly house scenario simulation

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio
50	865.52	94.82	653.15	7585	7192	57	5.18
100	911.12	96.08	653.89	15244	14646	96	3.92
150	920.53	96.57	649.48	22734	21952	136	3.43
200	935.50	96.81	646.18	30217	29255	196	3.19
250	943.44	96.96	647.77	37873	36722	228	3.04
300	945.35	96.96	648.27	45485	44101	261	3.04
350	925.40	96.94	648.34	53011	51387	289	3.06
400	925.59	96.79	648.24	60707	58756	325	3.21
450	890.94	96.76	646.42	68092	65888	375	3.24
500	914.18	96.76	646.35	75449	73002	400	3.24

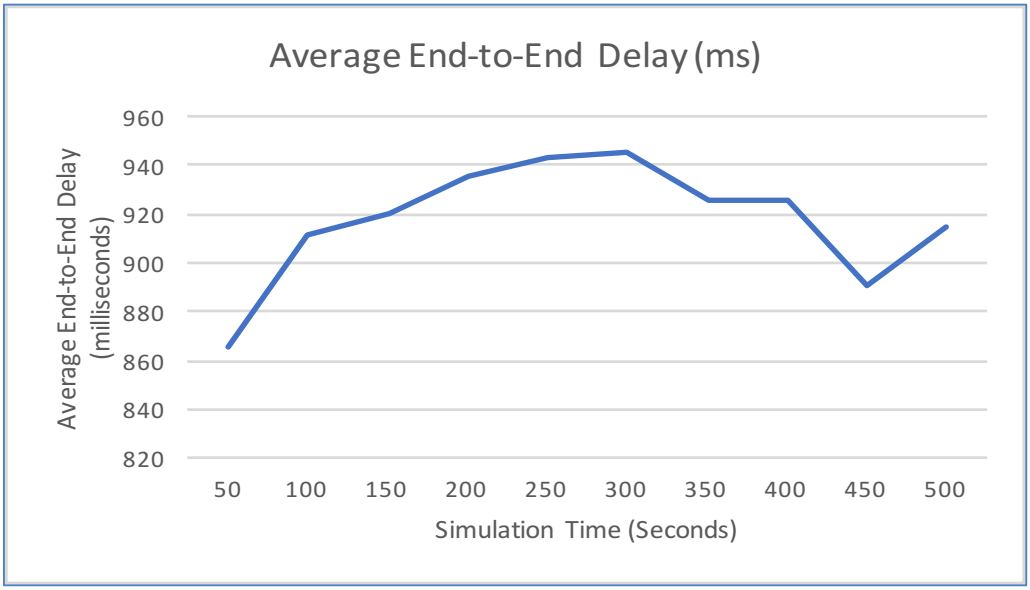


Figure 7-11 Mean Average end-to- end delay for different simulation times for NS2 elderly house scenario

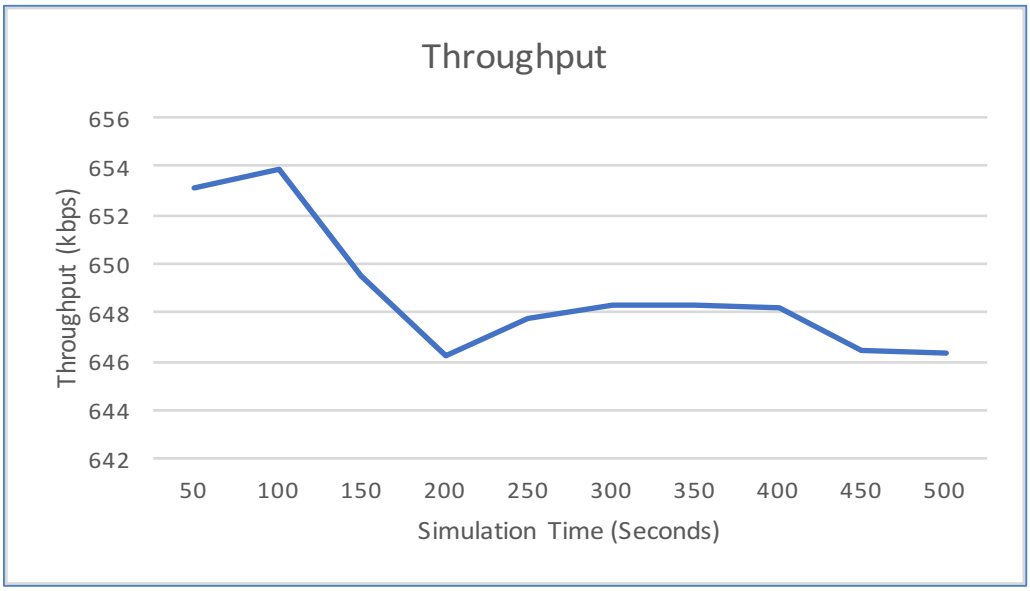


Figure 7-12 Mean throughput for different simulation times for NS2 elderly house scenario

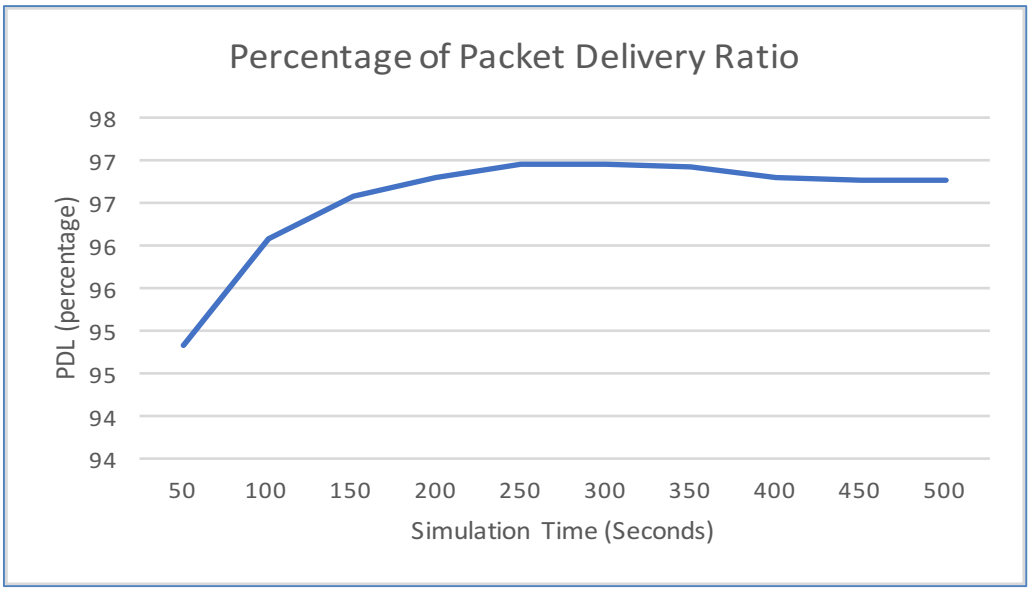


Figure 7-13 Mean percentage of packet delivery ratio for different simulation times for NS2 elderly house scenario

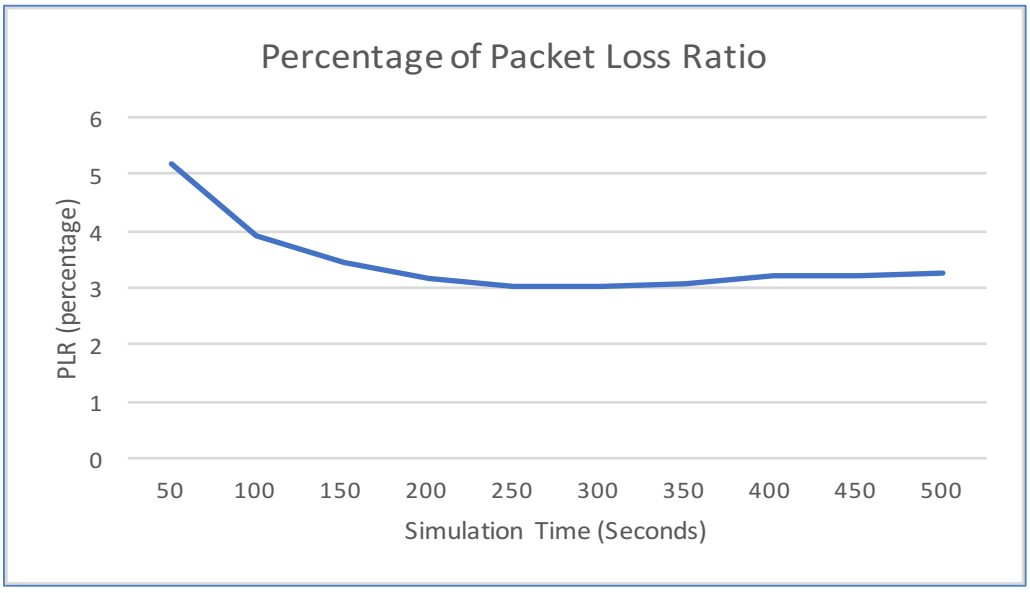


Figure 7-14 Mean percentage of packet loss ratio for different simulation times for NS2 elderly house scenario

7.2.3 Battlefield scenario simulation

Simulation model and components: The battlefield scenario discussed in Figure 5-3 is modelled and simulated in NS2. Figure 7-15 shows the representation of the battlefield scenario in the NS2 NAM.

The network model in Figure 7-15 is composed of 2 base station (BS), 4 cluster heads (CH), 5 gateways (G), 13 wearable (W) / implanted (I) /environmental (E) sensors and 9 video/audio sensors. Following the scenario discussed in Figure 5-3, the network in Figure 7-15 is connected as follows: Gateway (node 14) collects data from wearable and implanted sensors (nodes 15 and 16) and from video sensor (node 17). Gateway (node 13) collects data from wearable and implanted sensors (nodes 18 and 19). The data from gateways (nodes 13 and 14), environmental sensor (node 10), audio and video sensors (nodes 11 and 12) is sent to cluster head (node 5). Gateway (node 6) collects data from wearable, implanted and environmental sensors (nodes 7,8 and 9). The data from cluster head (node 5) and gateway (node 6) is sent to base station (node 0) through cluster heads (node 4 and 2) or to another base station (node 1) through cluster heads (nodes 4, 2 and 3).

Gateway (node 24) collects data from video, audio sensors (nodes 25 and 28), wearable and implanted sensors (nodes 26 and 27). Gateway (node 29) collects data from wearable, implanted sensors (nodes 30 and 31) and audio sensors (node 32). The data collected from gateways (nodes 24 and 29), audio, video sensors (nodes 20, 22 and 23) and environmental sensors (node 21). The data collected by cluster head is sent to base station (node 1) or to base station (node 0) through cluster head (node 2).

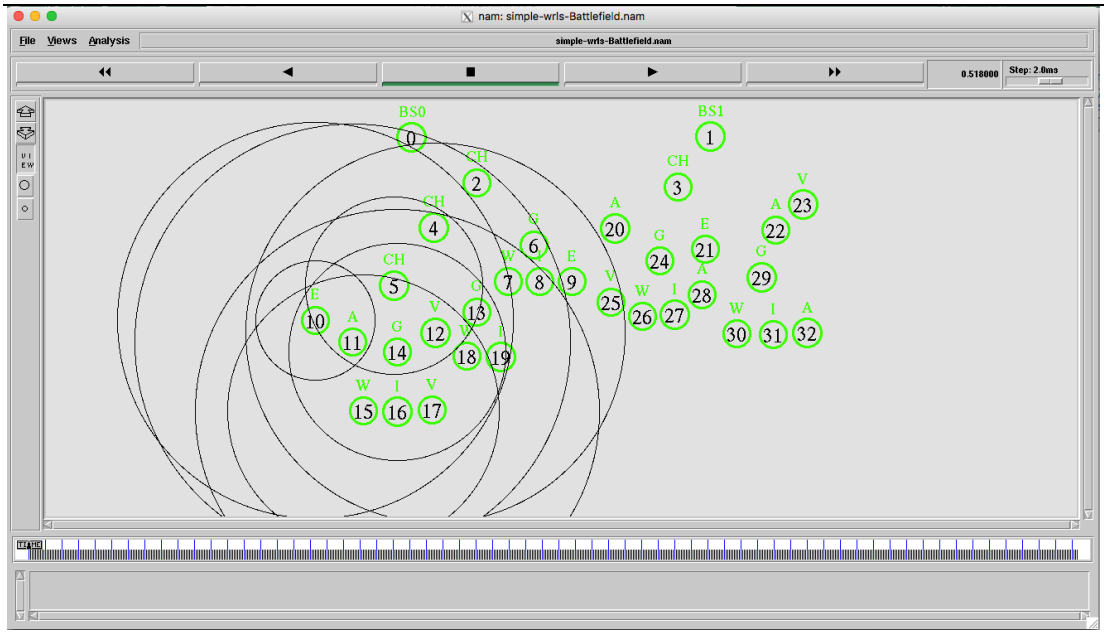


Figure 7-15 NS2 battlefield scenario model on NAM

Results and discussion: The mean values for average end-to-end delay, percentage of PDR, throughput, number of generated packets, number of received packets, number of dropped packets and percentage of PLR for the battlefield scenario are shown in Table 7-3. Figure 7-16, Figure 7-17, Figure 7-18 and Figure 7-19 show the plot of the results in Table 7-3. Similar to the elderly house scenario, as more packets are generated, the throughput decreases. However, the percentage decrease in throughput in the battlefield scenario (0.068%) is less than that in the elderly house scenario (1.04%). Although in the battlefield scenario the data is not sent through a single path similar to the elderly house scenario, the throughput decreased. The decrease in the throughput is due to the increased number of multimedia messages due to the high number of multimedia sensor nodes present in the battlefield scenario which cause a high number of packet drop thus the decrease in throughput.

Table 7-3 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets and percentage of PLR for different simulation times for NS2 battlefield scenario simulation

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio
50	1194.62	91.78	566.28	6657	6111	55	8.22
100	1241.62	94.30	567.70	13348	12586	90	5.70
150	1296.89	94.19	577.39	20326	19146	168	5.81
200	1266.89	95.55	565.37	26557	25375	168	4.45
250	1267.46	95.72	564.75	33177	31758	217	4.28
300	1247.16	96.05	567.41	39968	38391	251	3.95
350	1265.68	95.89	562.67	46261	44359	281	4.11
400	1329.67	96.20	564.97	53049	51035	313	3.80
450	1283.90	96.26	566.66	59907	57665	360	3.74
500	1268.70	96.50	565.89	66388	64067	375	3.50

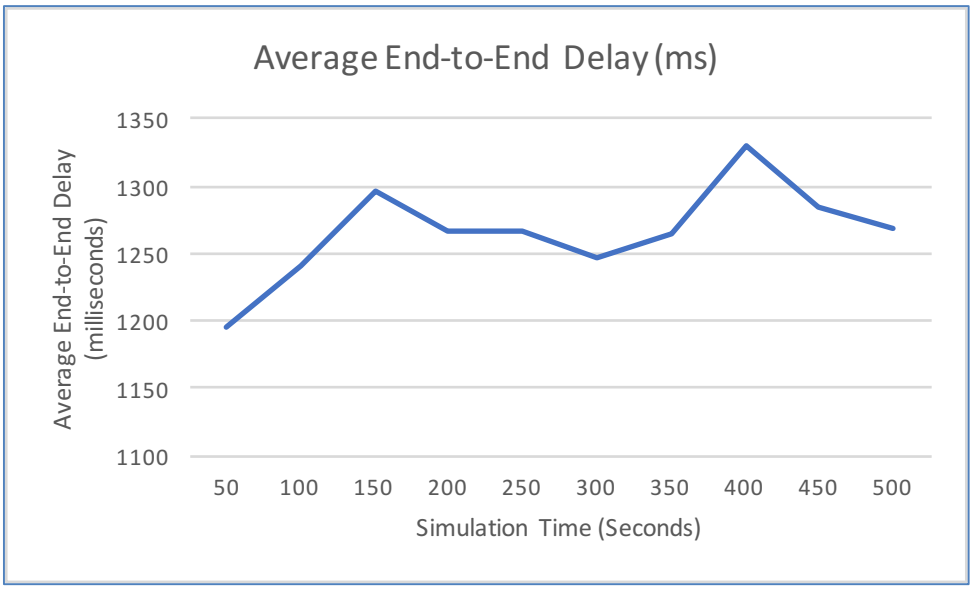


Figure 7-16 Mean Average end-to- end delay for different simulation times for NS2 battlefield scenario

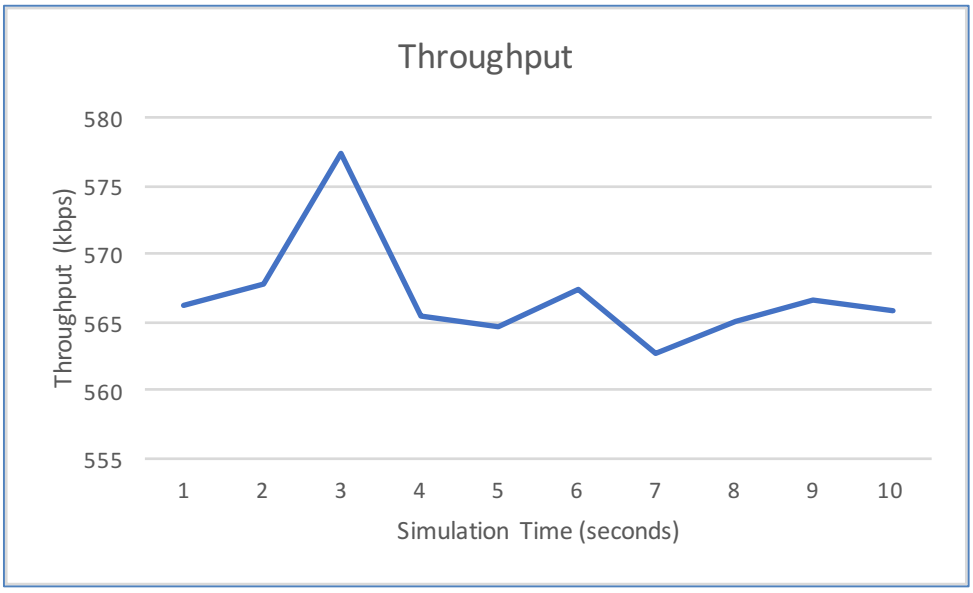


Figure 7-17 Mean throughput for different simulation times for NS2 battlefield scenario

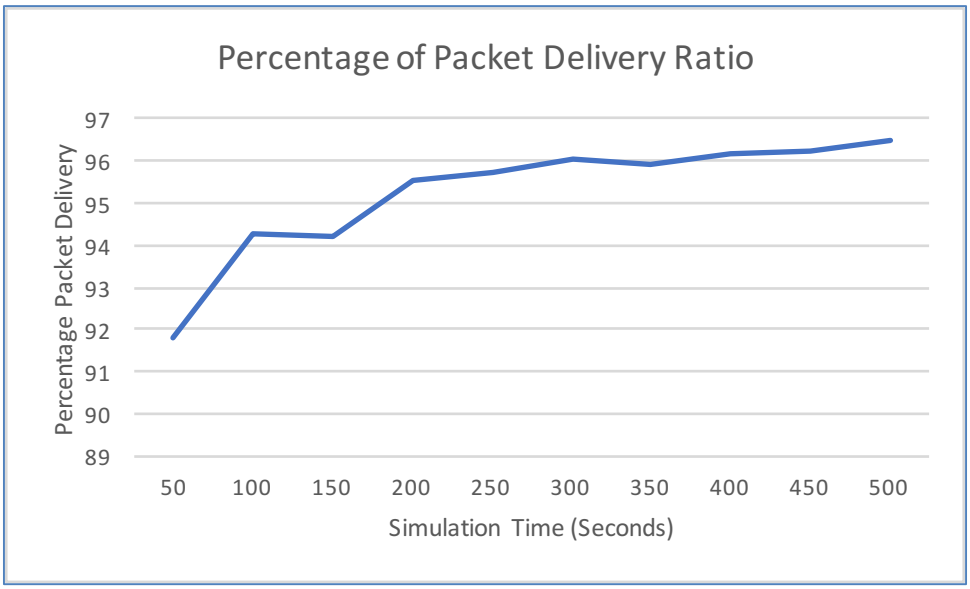


Figure 7-18 Mean percentage of packet delivery ratio for different simulation times for NS2 battlefield scenario

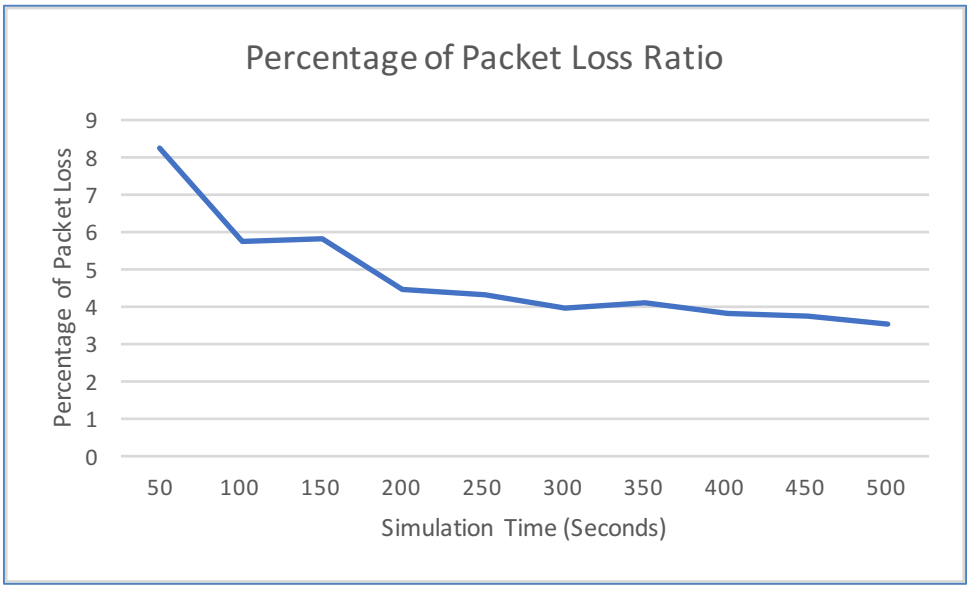


Figure 7-19 Mean percentage of packet loss ratio for different simulation times for NS2 battlefield scenario

7.3 Analysis of overheads for a privacy-aware WMSN-based healthcare sub-system

Different methodologies can be adopted to evaluate a network design. Implementations and testbeds of systems being analysed offer high accuracy but suffer from hardware limitations and high costs (Tan et al., 2011). Simulation might be relatively accurate (provided that the models for simulation are realistic) but it is sometimes slow and in some cases does not provide intuition for the output results (Caesar, 2010). Finally, analytical results may give insights into the system being analysed but sometimes they are inapplicable (Caesar, 2010).

In this research work, both a network simulator tool and theoretical analysis have been used to assess the overhead due to the introduction of the privacy and security measures into the WMSN-based healthcare sub-system.

7.3.1 Simulation-based analysis of overheads

Aim: The aim of this experiment is assess the overhead due to the introduction of the privacy and security measures into the WMSN-based healthcare sub-system.

Method: To assess the overhead due to the introduction of the privacy measures into a WMSN-based healthcare sub-system, the NS2 simulation tool was used to simulate a sub-system configured according to the hospital scenario depicted in Figure 5-1. The simulation model, components, parameters and analysis method deployed in this experiment are the same as those deployed in Section 7.2. The network model in Figure 7-5 is used in this experiment. First, only one scalar sensor (for example a wearable sensor as in Figure 5-1) was allowed to generate traffic under each gateway. The network, which was based on the ant routing protocol, was allowed to run for varying times ranging from 50 seconds to 500 seconds. Each time, each experiment was run 10 times to be able to calculate the mean, standard deviation, margin of error, upper bound and lower bound for a 95% confidence interval. Average end-to-end delay, percentage of packet delivery ratio, throughput, number of generated packets, number of received packets, number of dropped packets, percentage of packet loss ratio and average number of simulation clock ticks required to run the whole network was recorded. Next, the same experiment was repeated but the data generated from the scalar sensor was encrypted using LEAP protocol before sending it to the base station and the average number of clock ticks was recorded. Finally, the experiment based on ant routing and encrypted data was run and fake traffic was introduced. The fake traffic

was generated with the same rate as the traffic generated by the scalar sensor (a rate of 40 kbps). The same approach was repeated for combinations of different numbers of scalar (as depicted by a *w* or *i* in Figure 5-1) and multimedia sensors (as video sensor or audio sensor in Figure 5-1) per cluster head and the results were recorded.

Results: A summary of the mean results recorded for the experiments are shown in Table 7-4, Table 7-5, Table 7-6, Table 7-7, Table 7-8, Table 7-9, Table 7-10, Table 7-11, Table 7-12, Table 7-13, Table 7-14 and Table 7-15. The plots of the results in these tables are depicted in Figure 7-20, Figure 7-21, Figure 7-22, Figure 7-23, Figure 7-24, Figure 7-25, Figure 7-26 and Figure 7-27. Samples of the recorded results of the 10 times run of each experiment with the calculated mean, standard deviation, margin of error, upper and lower bounds are presented in Appendix D.

According to the tables (Table 7-4, Table 7-5, Table 7-6, Table 7-7, Table 7-8, Table 7-9, Table 7-10, Table 7-11, Table 7-12, Table 7-13, Table 7-14 and Table 7-15), it can be noticed when only security is applied, average end-to-end delay, packet delivery ratio, throughput , packet loss ratio, number of generated, received and lost packets are not altered. This is because security is applied at the level of the node and the network packets are only encrypted (no extra packets are injected into the network). However, in case of applying privacy (injecting fake traffic into the network), the number of packets are changed which affects average end-to-end delay, packet delivery ratio, throughput, packet loss ratio, number of generated, received and lost packets.

Based on Figure 7-20, it is clear that the shortest average end-to-end delay was achieved when only one scalar sensor was deployed in the network. The highest average end-to-end delay longest was achieved when two scalar and two multimedia sensors were deployed in the network. This shows that as more sensors are deployed in the network, the average end-to-end delay increases because the bandwidth is more congested and more packets are competing to reach their destination.

From Figure 7-21, it is clear that as the simulation time increased, the mean percentage of packet delivery ratio has increased. The minimum recorded mean percentage of packet delivery ratio was 92.76% and the maximum was 97.51%.

Figure 7-22 depicts a plot of the mean throughput for different simulation times. Overall, the least throughput was in the case of two scalar sensors and two multimedia sensors with privacy and security applied and the highest throughput was in the case of only one scalar sensor deployed under each gateway. In the first case (lowest throughput), less packets were generated and less packets were received (please refer to Figure 7-23 and Figure 7-24) as more sensors shared the bandwidth causing less packets to be generated leading to a lower throughput. In the second case (highest throughput), the bandwidth is available for one sensor to send as many packets as possible thus the increased throughput. In addition, as more data sources are introduced and fake traffic is injected into the network, the throughput significantly decreases.

The mean number of clock ticks reported in Figure 7-26 show that when privacy and security are applied, the mean number of clock ticks has significantly increased especially in the presence of multimedia data. When security was introduced in the presence of one scalar sensor, the mean number of clocks ticks increased (on average) 517%. When both privacy and security were applied, the number of clock ticks increased (on average) 558%.

When a multimedia sensor was added (along with the scalar sensor), the number of clock ticks increased (on average) by 17.22% compared to when only one scalar sensor was deployed. The addition of encryption caused the number of clock ticks (on average) to be 13 times the number of clock ticks compared to when no fake traffic or encryption were added (a relative increase of 1211.165%). The introduction of both fake traffic and encryption caused (on average) the number of clock ticks to be 13.4 times (a relative increase of 1239%), compared to when no protection mechanisms were added. When two multimedia sensors were deployed along with one scalar sensor, the addition of security caused the number of clock ticks to be 12.5 times (on average) the number of clock ticks when no security was applied (a relative increase of 1146.86%). The introduction of privacy and security caused the number of clock ticks to be (on average) 13 times the number of clock ticks when no privacy and security were applied (a relative increase of 1199.97%). Finally, in the case of two scalar sensors and two multimedia sensors, the application of security caused a relative increase (on average) of 970.8% compared to when no security was applied. In case of applying both privacy and security, a relative increase (on average) of 980.6% was recorded.

Discussion: It can be concluded from the previous tables and figures that as more sensors are introduced in the network (especially multimedia sensors) with the application of privacy and security, increased average end-to-end delay and decreased throughput are reported. In addition, it is clear from Figure 7-26 that the application of fake traffic or encryption to multimedia data adds a significant overhead (presented in the form of simulation clock ticks) to the overall network. The significant increase in the average number of clock ticks denotes that there is a significant increase in the total time required to enhance the privacy of multimedia data. This implies that in cases when real-time or very quick collection of healthcare data is required, multimedia data should be cut to the minimum possible, to decrease the overhead due to the processing of the multimedia data.

The results depicted in the previous figures and tables presented in this section did not provide insights into the privacy-aware WMSN-based healthcare sub-system. The results did not show what the causes for the reported overhead are and what the delay components at each stage of the routing protocol are. The NS2 simulator provided analysis at the network level and did not provide insight into the processing inside the network nodes. Consequently, simulation results are not enough to study the overhead analysis, and thus theoretical analysis was required to study the causes of the overhead.

Related Work: The previous chapter presented a literature survey on privacy-aware ant routing (see Section 5.3.2). This literature survey discussed the work by (Dias et al., 2013), (Kalpana & Rengarajan, 2012) and (Zhou & Wen, 2014). The algorithm suggested by (Dias et al., 2013) provided general information to the drivers without knowing their source or destination which, from their point of view, was the privacy measure they provided. The results presented by (Kalpana & Rengarajan, 2012) showed an overhead increase by 100%. However, their technique addressed privacy using encryption only with an overhead of 100% compared to the work presented in this thesis, which addressed privacy using both encryption and fake traffic with an overhead of almost 101%. The technique presented in this work is resistant to identifiability, linkability and location disclosure attacks whereas their technique is only resistant to identifiability. (Zhou & Wen, 2014) provided location privacy by routing modification and energy-preservation in ant-based colony optimization. Consequently, their approach is different from that presented in this work and cannot be compared to it.

Table 7-4 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor deployed under each gateway

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1102.33	93.43	597.45	7023.10	6561.50	38.50	6.57	240807.66
100	1136.94	95.61	615.37	14397.20	13765.60	52.20	4.39	241650.68
150	1143.94	95.94	621.70	21783.40	20901.40	70.50	4.06	243287.70
200	1167.05	96.66	634.40	29627.60	28637.50	103.40	3.34	243708.60
250	1155.60	96.93	641.06	37403.90	36257.50	129.80	3.07	247117.70
300	1169.62	96.95	648.25	45338.00	43955.10	139.60	3.05	251173.16
350	1133.00	97.06	646.22	52741.60	51189.70	154.80	2.94	256860.62
400	1137.44	97.33	652.70	60840.30	59215.00	166.80	2.67	258698.04
450	1177.91	97.46	654.34	68630.30	66885.20	208.40	2.54	259345.76
500	1143.22	97.38	655.10	76286.80	74293.00	203.50	2.62	260365.46

Table 7-5 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor deployed under each gateway with security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1102.33	93.43	597.45	7023.10	6561.50	38.50	6.57	1512323.67
100	1136.94	95.61	615.37	14397.20	13765.60	52.20	4.39	1518668.34
150	1143.94	95.94	621.70	21783.40	20901.40	70.50	4.06	1529383.04
200	1167.05	96.66	634.40	29627.60	28637.50	103.40	3.34	1534029.04
250	1155.60	96.93	641.06	37403.90	36257.50	129.80	3.07	1542054.34
300	1169.62	96.95	648.25	45338.00	43955.10	139.60	3.05	1549876.30
350	1133.00	97.06	646.22	52741.60	51189.70	154.80	2.94	1551312.75
400	1137.44	97.33	652.70	60840.30	59215.00	166.80	2.67	1558118.61
450	1177.91	97.46	654.34	68630.30	66885.20	208.40	2.54	1565110.63
500	1143.22	97.38	655.10	76286.80	74293.00	203.50	2.62	1573832.50

Table 7-6 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor deployed under each gateway with privacy and security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1233.36	92.76	618.13	7190.78	6670.56	37.56	7.24	1579678.48
100	1233.59	94.90	619.50	14514.00	13857.00	58.00	4.53	1585210.32
150	1267.14	95.58	632.36	22203.40	21223.20	95.30	4.42	1590915.20
200	1295.64	96.34	636.72	29756.70	28667.00	106.40	3.66	1595009.93
250	1332.44	96.70	639.03	37324.10	36092.40	147.20	3.30	1610060.61
300	1307.96	96.58	641.81	44982.40	43442.30	162.50	3.42	1660539.35
350	1339.39	97.03	639.22	52222.20	50671.80	179.90	2.97	1703257.12
400	1278.78	96.80	638.96	59671.10	57762.40	199.70	3.20	1709224.20
450	1322.22	96.68	639.13	67129.50	64902.20	220.20	3.32	1715757.00
500	1293.89	96.88	635.75	74173.00	71859.17	223.33	3.12	1720770.30

Table 7-7 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor and one multimedia sensor

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1106.88	93.44	598.12	7030.40	6569.10	39.80	6.56	276524.90
100	1163.83	95.40	615.17	14406.30	13744.80	61.60	4.60	277873.60
150	1183.54	96.05	620.19	21771.40	20911.50	85.40	3.95	286343.60
200	1155.58	96.06	624.28	29238.90	28087.30	114.70	3.94	291377.30
250	1193.66	96.78	630.96	36898.60	35709.80	128.40	3.22	293132.60
300	1198.58	96.82	638.87	44784.50	43361.40	144.70	3.18	295216.50
350	1218.12	96.99	640.48	52400.33	50824.00	179.33	3.01	299285.60
400	1233.67	97.04	639.47	57981.80	53856.64	170.80	3.03	301059.30
450	1174.83	96.99	642.94	67625.40	65595.50	205.80	3.01	303430.60
500	1201.18	96.94	641.50	75009.80	72716.20	228.00	3.06	310143.00

Table 7-8 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor and one multimedia sensor with security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1106.88	93.44	598.12	7030.40	6569.10	39.80	6.56	3646521.10
100	1163.83	95.40	615.17	14406.30	13744.80	61.60	4.60	3724686.40
150	1183.54	96.05	620.19	21771.40	20911.50	85.40	3.95	3735480.60
200	1155.58	96.06	624.28	29238.90	28087.30	114.70	3.94	3762862.10
250	1193.66	96.78	630.96	36898.60	35709.80	128.40	3.22	3803479.80
300	1198.58	96.82	638.87	44784.50	43361.40	144.70	3.18	3846331.50
350	1218.12	96.99	640.48	52400.33	50824.00	179.33	3.01	3903069.30
400	1233.67	97.04	639.47	57981.80	53856.64	170.80	3.03	3978752.50
450	1174.83	96.99	642.94	67625.40	65595.50	205.80	3.01	4005928.90
500	1201.18	96.94	641.50	75009.80	72716.20	228.00	3.06	4063629.40

Table 7-9 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor and one multimedia sensor with privacy and security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1106.35	93.70	582.34	6859.50	6427.70	30.60	6.30	3729050.90
100	1147.73	95.51	590.44	13848.00	13225.90	55.70	4.49	3765037.60
150	1231.18	96.25	605.09	21234.60	20437.70	74.40	3.75	3830002.10
200	1268.07	96.24	609.82	28555.50	27481.80	103.40	3.76	3860661.30
250	1233.90	96.87	611.26	35745.10	34627.90	129.00	3.13	3929760.00
300	1237.52	96.97	614.08	43051.00	41746.38	128.00	3.03	3764883.85
350	1247.50	97.10	620.96	50798.00	49323.50	149.75	2.90	4044556.60
400	1254.02	96.76	626.73	58642.83	56744.83	212.33	3.24	4075714.10
450	1288.60	97.31	627.22	65953.00	64179.00	240.00	2.69	4136474.30
500	1289.32	97.51	628.07	73527.00	71695.00	192.00	2.49	4157218.70

Table 7-10 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor and two multimedia sensors

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1130.99	93.92	590.07	6955.30	6532.60	23.40	6.08	300949.20
100	1147.67	95.72	599.55	14072.17	13469.50	45.00	4.28	302330.80
150	1207.66	96.11	611.02	21482.00	20646.00	77.14	3.89	312812.80
200	1224.97	96.81	625.08	29257.00	28323.00	92.33	3.19	319612.50
250	1217.31	96.60	620.12	36306.20	35071.00	123.60	3.40	320372.50
300	1245.52	97.07	632.23	44382.00	43080.00	206.00	2.93	322819.00
350	1198.52	96.96	626.32	51289.67	49733.00	180.33	3.04	327130.80
400	1217.55	96.75	612.42	57474.00	55606.00	138.00	3.25	330958.30
450	1204.25	96.51	636.97	67028.25	64688.75	214.50	3.49	332480.40
500	1230.58	97.17	636.07	74396.33	72292.67	260.00	2.83	338177.60

Table 7-11 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor and two multimedia sensors with security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1130.99	93.92	590.07	6955.30	6532.60	23.40	6.08	3873706.60
100	1147.67	95.72	599.55	14072.17	13469.50	45.00	4.28	3885396.10
150	1207.66	96.11	611.02	21482.00	20646.00	77.14	3.89	3918875.10
200	1224.97	96.81	625.08	29257.00	28323.00	92.33	3.19	3935556.10
250	1217.31	96.60	620.12	36306.20	35071.00	123.60	3.40	3975664.70
300	1245.52	97.07	632.23	44382.00	43080.00	206.00	2.93	4006329.80
350	1198.52	96.96	626.32	51289.67	49733.00	180.33	3.04	4043284.60
400	1217.55	96.75	612.42	57474.00	55606.00	138.00	3.25	4060945.90
450	1204.25	96.51	636.97	67028.25	64688.75	214.50	3.49	4151628.10
500	1230.58	97.17	636.07	74396.33	72292.67	260.00	2.83	4120672.70

Table 7-12 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for one scalar sensor and two multimedia sensors with privacy and security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1160.11	93.61	570.19	6725.67	6297.00	20.33	6.39	3944423.50
100	1174.89	95.77	586.68	13784.50	13201.25	47.25	4.23	4014438.30
150	1179.14	96.07	594.16	20888.50	20066.75	56.25	3.93	4077411.50
200	1192.80	96.36	594.71	27898.56	26885.44	94.44	3.64	4142158.30
250	1200.93	96.30	599.73	35135.00	33835.75	133.00	3.70	4184836.90
300	1303.70	96.05	582.92	41098.50	39483.00	131.50	3.95	4107734.00
350	1217.02	96.03	611.36	50143.25	48153.75	197.25	3.97	4277014.10
400	1295.13	95.88	576.83	54286.70	52051.30	183.40	4.12	4159182.30
450	1220.05	97.41	616.22	64877.00	63196.50	216.00	2.59	4370637.60
500	1223.91	96.72	625.43	73209.00	70811.00	247.00	3.28	4410978.60

Table 7-13 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for two scalar sensors and two multimedia sensors

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1274.44	94.68	593.63	13919.14	13179.86	59.29	5.32	356850.60
100	1274.44	94.68	593.63	13919.14	13179.86	59.29	5.32	356850.60
150	1332.94	95.13	594.86	20935.50	19919.00	106.63	4.87	362406.20
200	1281.55	95.99	598.26	28007.14	26882.29	94.57	4.01	366800.60
250	1374.21	96.42	603.69	35357.20	34092.00	146.40	3.58	373310.60
300	1371.92	96.06	609.69	42812.50	41127.33	170.50	3.94	378320.30
350	1370.76	96.51	610.82	50082.89	48336.44	188.56	3.49	386266.10
400	1335.56	96.50	614.03	57501.86	55495.00	199.57	3.50	391867.30
450	1381.19	96.46	607.93	64028.89	61761.22	225.00	3.54	395464.50
500	1297.46	96.42	613.08	71808.11	69241.78	249.56	3.58	399427.30

Table 7-14 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for two scalar sensors and two multimedia sensors with security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1274.44	94.68	593.63	13919.14	13179.86	59.29	5.32	3882011.60
100	1274.44	94.68	593.63	13919.14	13179.86	59.29	5.32	3895584.70
150	1332.94	95.13	594.86	20935.50	19919.00	106.63	4.87	3952379.30
200	1281.55	95.99	598.26	28007.14	26882.29	94.57	4.01	3998720.70
250	1374.21	96.42	603.69	35357.20	34092.00	146.40	3.58	4024818.90
300	1371.92	96.06	609.69	42812.50	41127.33	170.50	3.94	4038350.70
350	1370.76	96.51	610.82	50082.89	48336.44	188.56	3.49	4074519.90
400	1335.56	96.50	614.03	57501.86	55495.00	199.57	3.50	4106184.10
450	1381.19	96.46	607.93	64028.89	61761.22	225.00	3.54	4151628.10
500	1297.46	96.42	613.08	71808.11	69241.78	249.56	3.58	4194154.40

Table 7-15 Mean values for average end-to-end delay, percentage of PDR, throughput, generated packets, received packets, dropped packets, percentage of PLR and clock ticks for two scalar sensors and two multimedia sensors with privacy and security applied

Simulation Time (seconds)	Average End-to-end Delay (milliseconds)	Percentage of Packet Delivery Ratio	Throughput (kbps)	Generated Packets	Received Packets	Dropped Packets	Percentage of Packet Loss Ratio	Clock Ticks
50	1205.77	93.24	580.90	6831.60	6370.60	16.90	6.76	3891161.70
100	1328.89	94.85	577.82	13577.00	12878.10	49.60	5.15	3920779.60
150	1301.51	94.99	570.35	20137.40	19128.00	90.60	5.01	3976679.70
200	1330.07	95.75	576.85	27106.60	25956.40	97.70	4.25	4038877.80
250	1343.58	95.87	581.11	34136.38	32728.88	113.38	4.13	4067277.90
300	1303.70	96.05	582.92	41098.50	39483.00	131.50	3.95	4107734.00
350	1329.36	96.22	592.47	48737.33	46898.11	188.22	3.78	4130544.60
400	1295.13	95.88	576.83	54286.70	52051.30	183.40	4.12	4159182.30
450	1335.98	95.91	574.62	60820.44	58335.78	198.67	4.09	4170678.80
500	1334.65	96.03	583.47	68575.80	65854.20	234.10	3.97	4225267.80

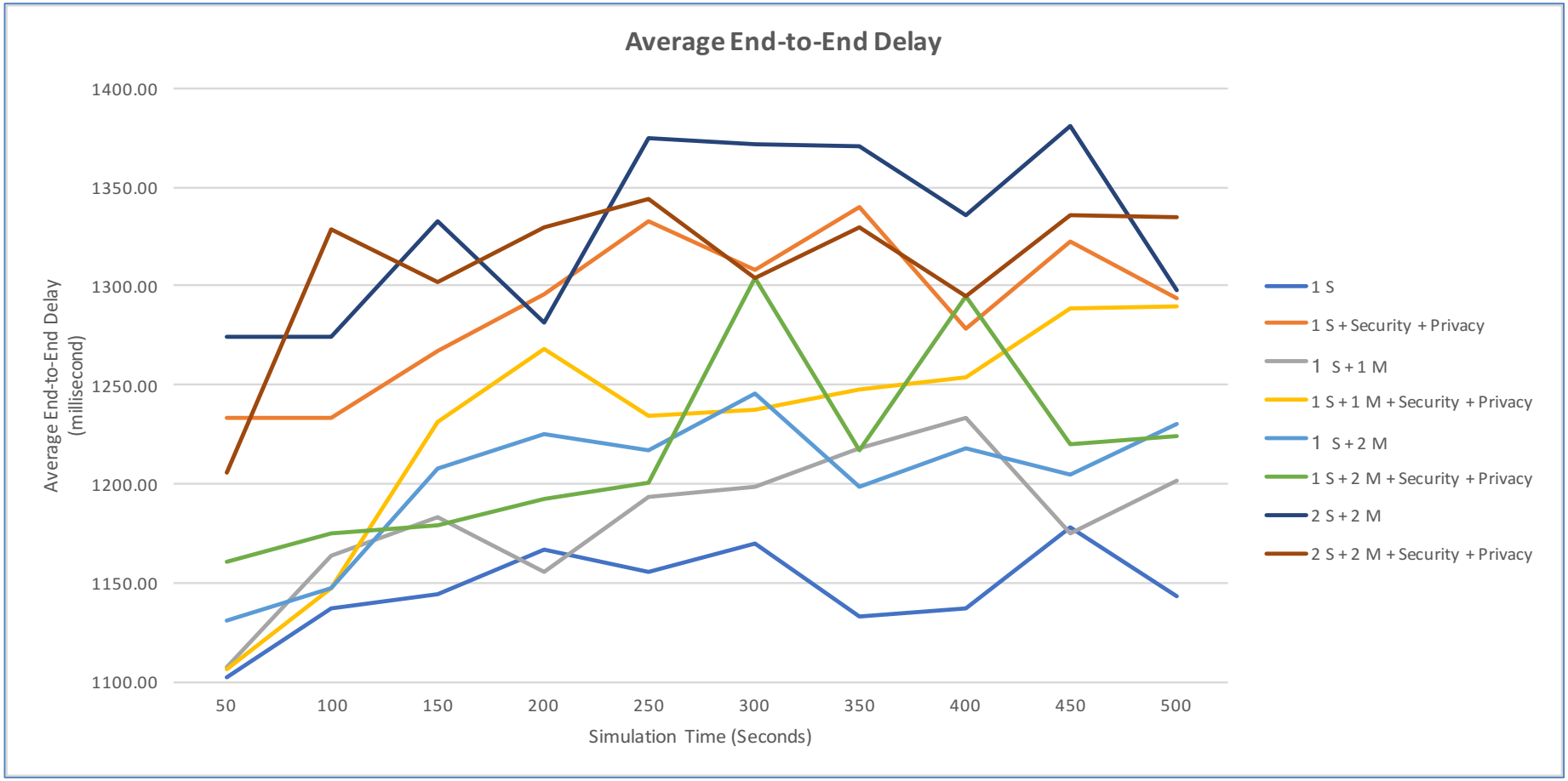


Figure 7-20 Mean Average end-to- end delay for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensor

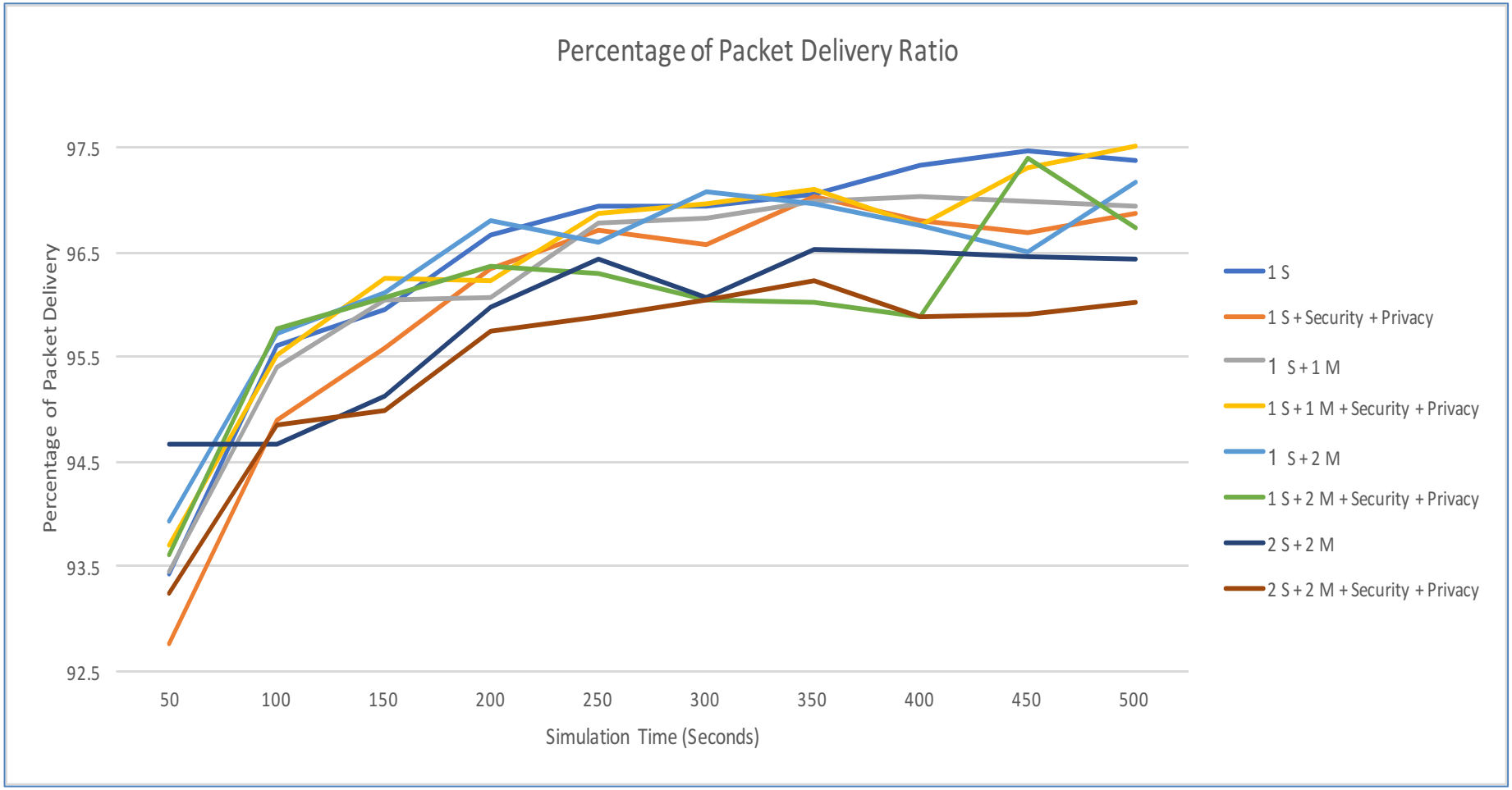


Figure 7-21 Mean percentage of packet delivery ratio for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensor

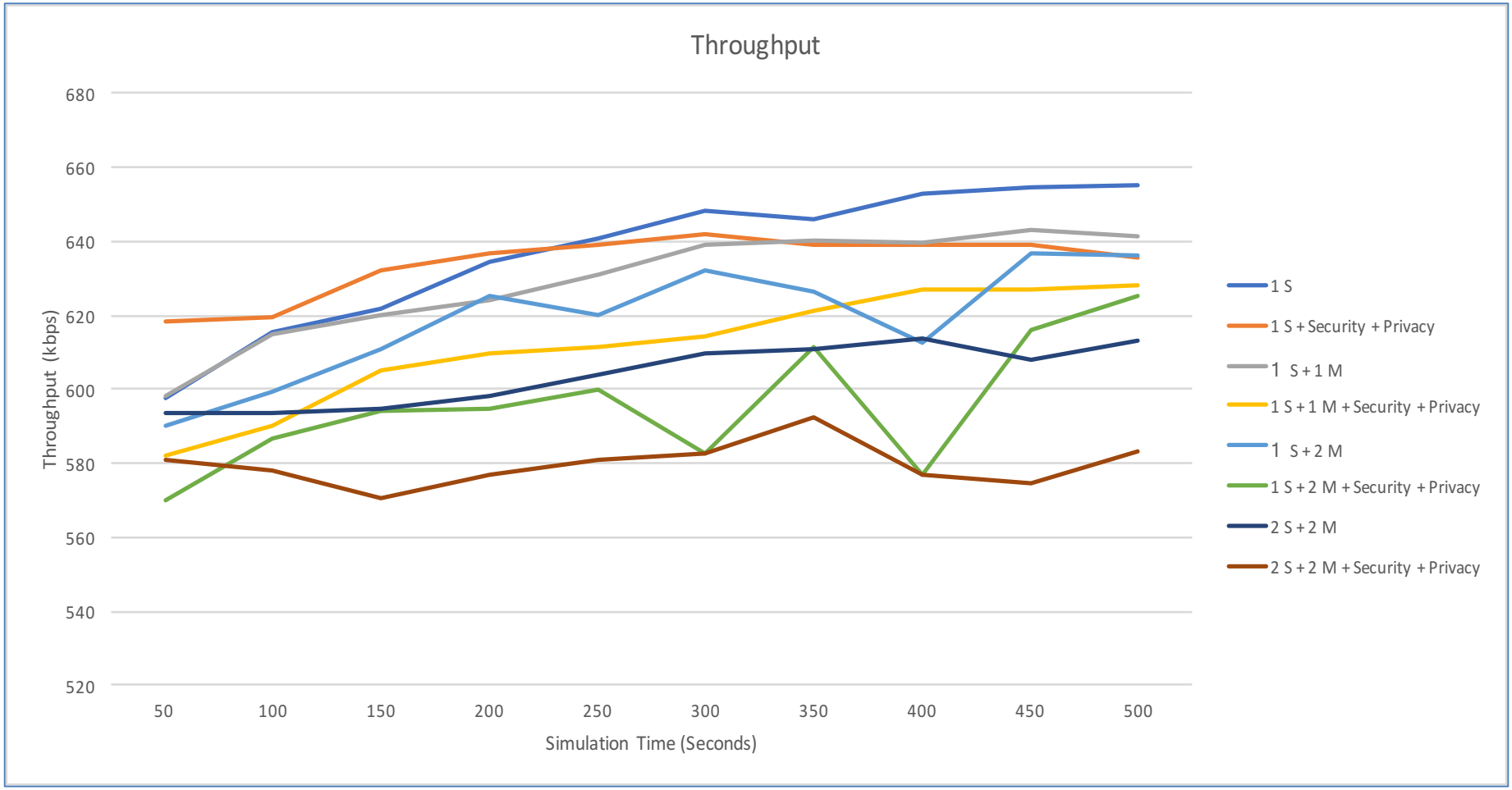


Figure 7-22 Mean throughput for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensor

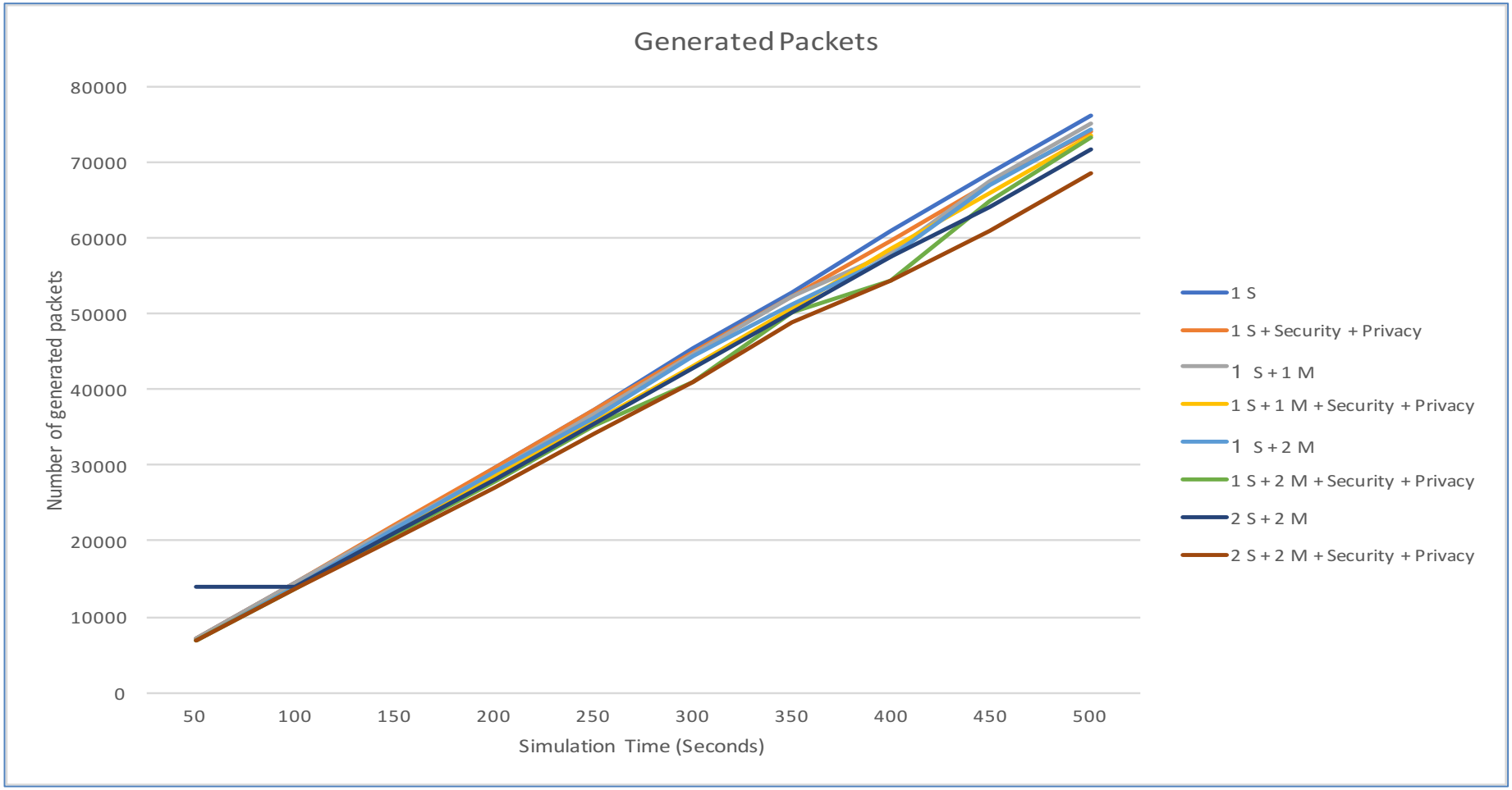


Figure 7-23 Mean number of generated packets for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensor

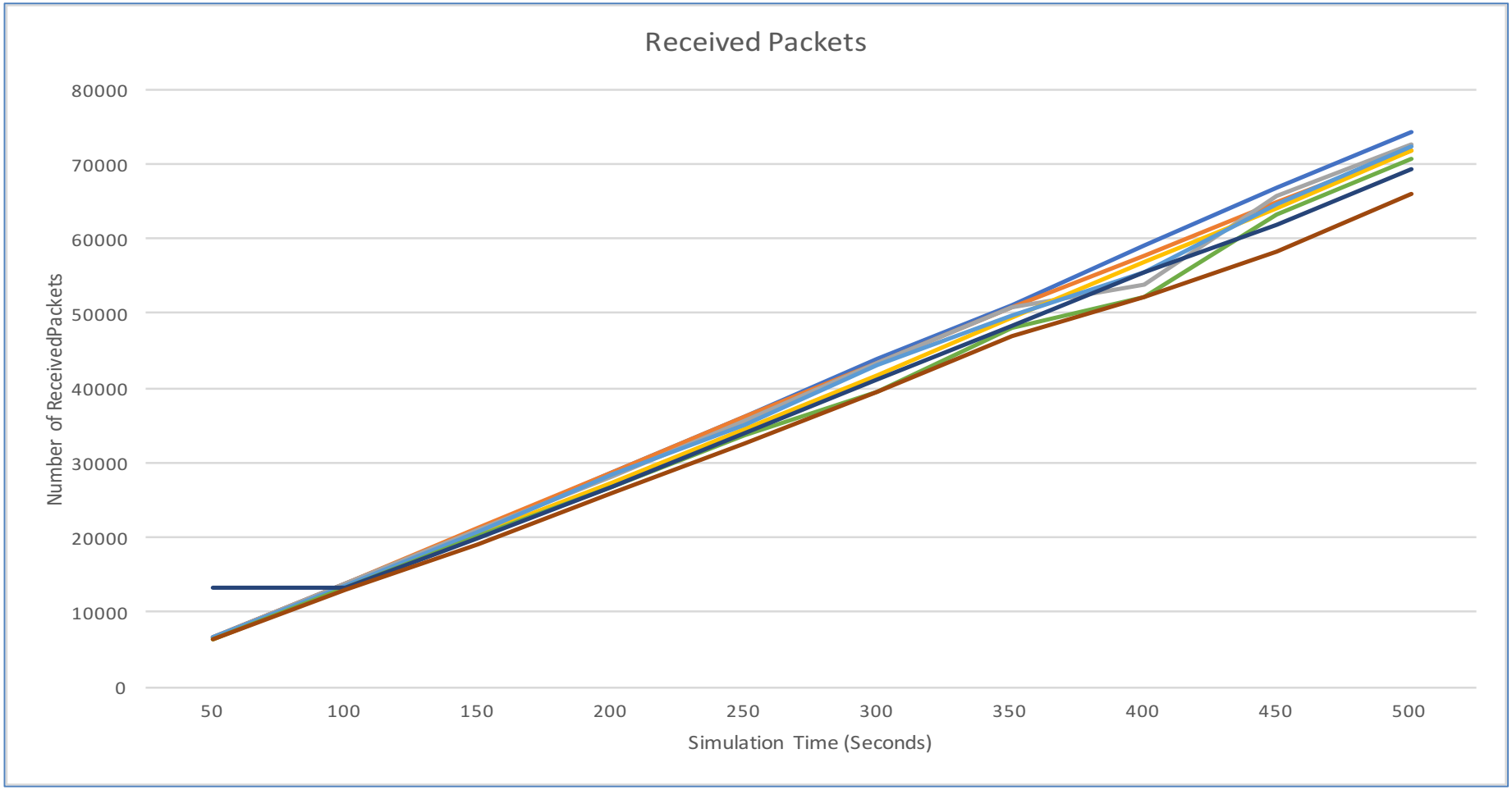


Figure 7-24 Mean number of received packets for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensor

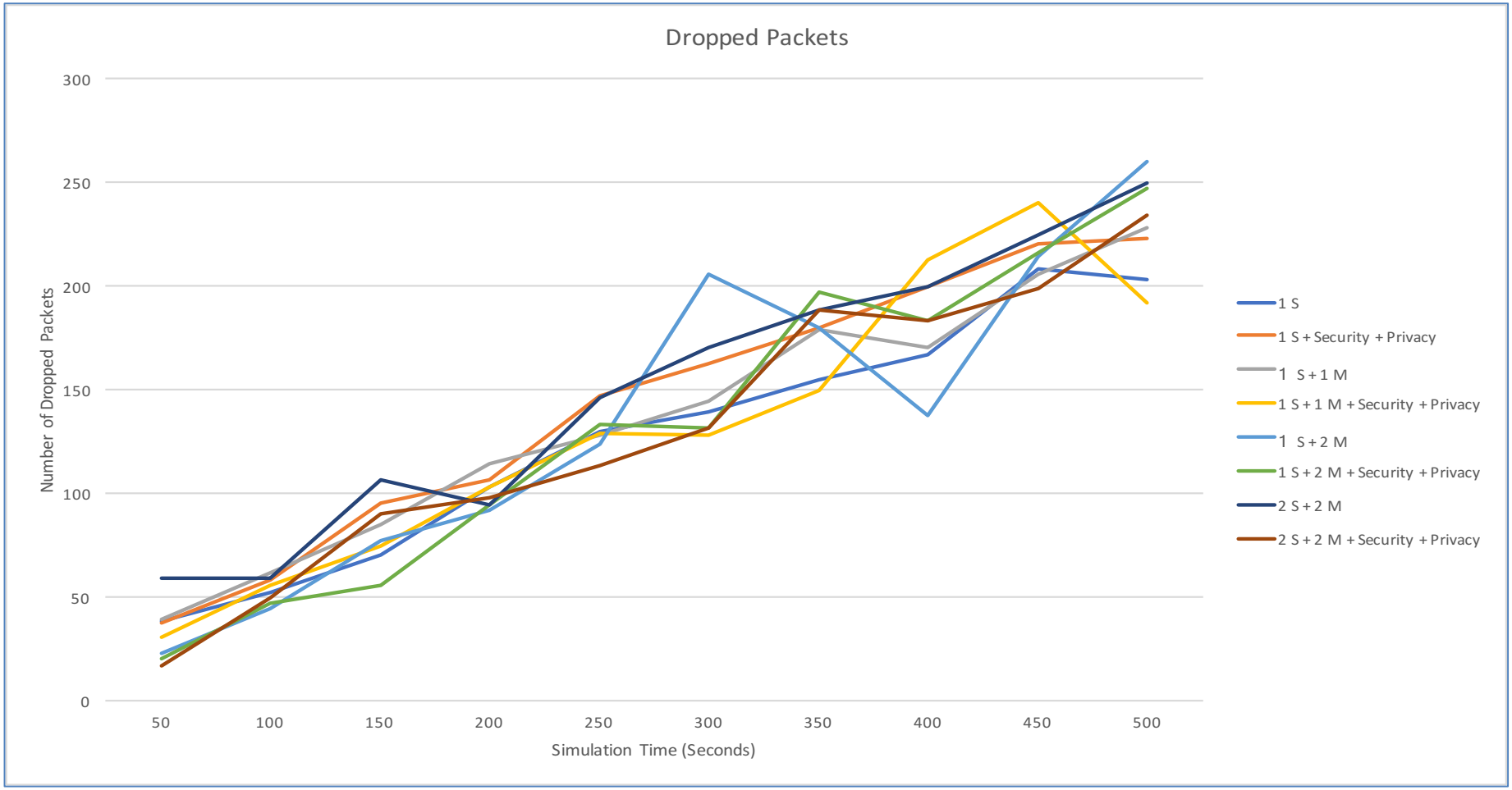


Figure 7-25 Mean number of dropped packets for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensor

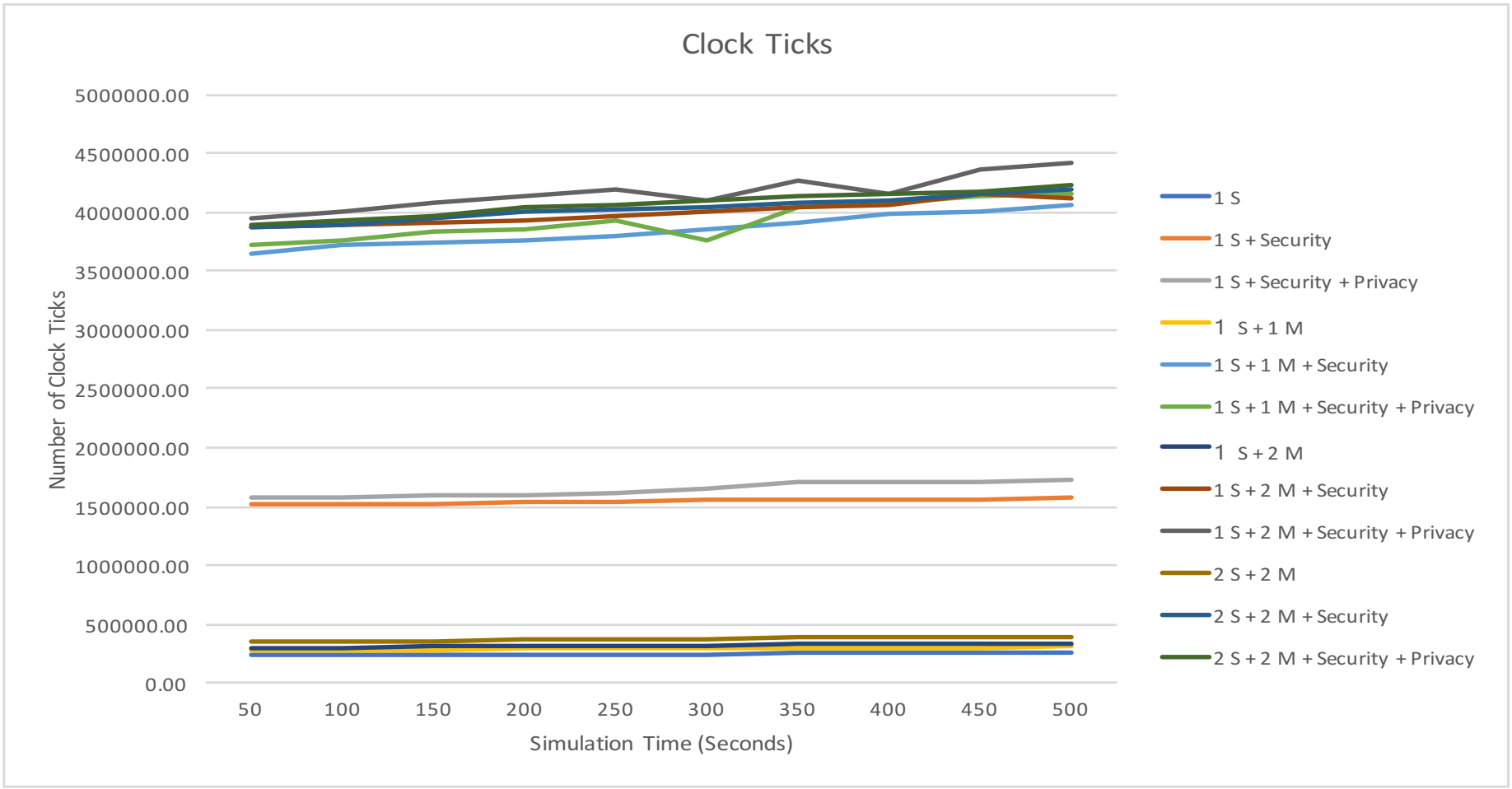


Figure 7-26 Mean number of clock ticks for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensor

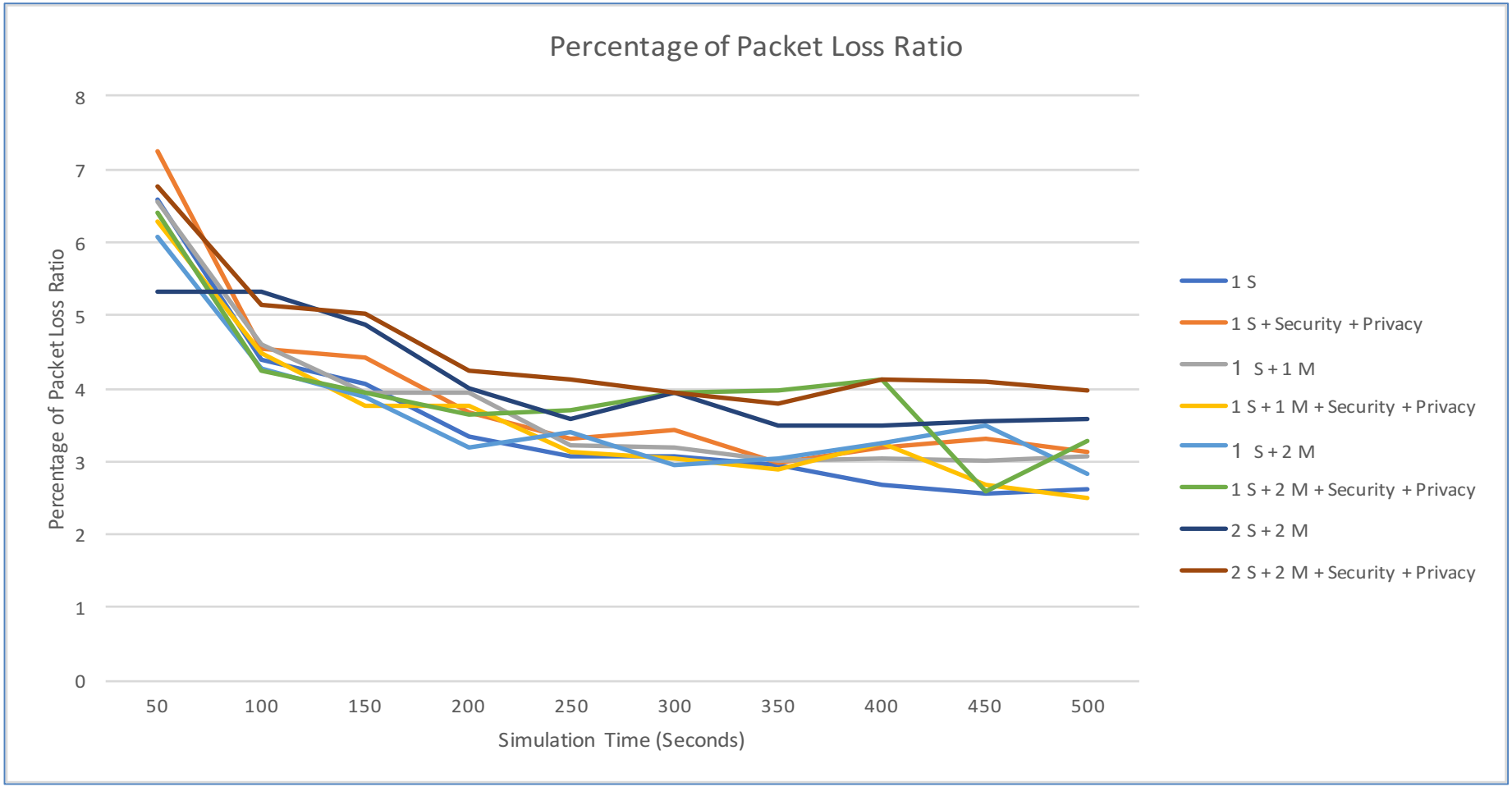


Figure 7-27 Mean percentage of packet loss ratio for different simulation times with different types and numbers of sensors deployed where S denotes scalar sensor and M denotes multimedia sensors

7.3.2 Theoretical analysis of overheads

The representation of the overhead, which is described in this section, was adopted from the analysis of algorithms (as in (Ganesan et al., 2003)) and operating systems in which the number of bits, computer cycles and memory usage is computed to assess the performance of these algorithms and systems.

The aim of this section is to evaluate the computation, network messages and storage overhead of the proposed privacy-aware WMSN-based healthcare sub-system, in three different operational scenarios: hospital, elderly house and battlefield. These operational scenarios are among the most popular application scenarios in the healthcare field. In each scenario, mathematical equations will be used to conduct a *quantitative analysis* of the *worst-case* overhead due to the introduction of the three privacy services (anonymity/pseudonymity, unlinkability and location privacy as identified in Chapter 4). The overhead is calculated for the adapted AntSensNet routing protocol. The representation of the overhead is in the form of equations, which makes this assessment methodology a generalised one, as an assessor can easily estimate the overhead when (s)he varies the number and the length of encryption keys, pseudorandom functions, pseudonyms, hash functions and other privacy and security related terms (as in Table 7-16). Using these equations, the operators and users of the healthcare sub-system can easily compare the level of privacy against the most feasible overhead depending on the application requirements.

The equations presented in this section only consider the overhead caused directly by the introduction of the privacy mechanisms. Other potentially underlying factors that may be affected by the introduction of the privacy mechanisms (such as change in the queue delays at the different network components) are out of the scope of this research. Based on (Kurose & Ross, 2017), these factors are expected to add little overhead (in terms of microseconds or milliseconds) compared to the overhead added due to the introduction of the privacy measures. Consequently, these factors will neither affect the overall findings nor the validity of the theoretical analysis. These factors in the equations are similar to the simplified models adopted by simulators to simulate a scenario. However, the theoretical analysis will allow an insight into the overheads, which arise because of the introduction of the privacy measures; unlike the simulation (presented in the previous section), which provides an overall estimation of the overhead with no details of where and how the overhead was introduced.

The theoretical analysis presented in this section provides an estimation of the overhead in storage, processing and communication messages. The estimated overhead does not depend on a specific sensor hardware technology. Even if the estimated overheads would put considerable strain on today's hardware technology (especially with the presence of multimedia traffic), it could be acceptable in near future technology.

Table 7-16 Definitions of the symbols used in the mathematical representation of overheads

Symbol	Description
c	# of cluster heads
m	# of medical and implanted sensors
cam	# of video sensors
aud	# of audio sensors
e	# of environmental sensors
g	# of gateways
h	# of hash functions
i_k	# of bits of the individual key
m_k	# of bits of the master key
pr_m_k	# of bits of the master key used in the pairwise key generation
pseud_f	# of bits of the pseudorandom function
hash_f	# of bits of the hash function
p_id	# of bits of the pseudonym
AUTH	# of bits in an AUTH received from neighbouring node
L	# of bits (length) of the one-way key chain of AUTH for next hop neighbours for message authentication
pr_id	# of bits of the pairwise key
pherm_vlaue	# of bits of pheromone of the AntSensNet information table

7.3.2.1 Analysis of overheads for the hospital scenario

7.2.2.1.1 Theoretical analysis

The aim of this analysis is to compute the overheads in the memory, computation and network messages due to the introduction of the privacy mechanisms. The overhead is computed in comparison to the original *AntSensNet* routing algorithm presented in (Cobo et al., 2010). The overhead will be computed after each stage based on the algorithm flowcharts and the stages presented in the previous section. It is important to note that the overhead is calculated based on the worst-case scenario when all sensor nodes connected to a gateway are sending data at the same time, and all video sensors, audio sensors and gateways connected to a

cluster head are all sending data to the cluster head at the same time. A summary of the overhead of each stage is given in Table 7-17, for the hospital scenario.

Stage 1: Pre-deployment stage

Each sensor involved in the communication of data will be assigned: a pseudorandom function of length *pseud_f* bits, an individual key of length *i_k* bits, a master key of length *m_k* bits, another master key of length *pr_m_k* bits for the pairwise communication between nodes and *h* hash functions, each of length *hash_f* bits. The keys and functions are preloaded into the sensor nodes prior to their deployment to avoid computation and network messages overhead in the communication channel. It is assumed that the data channel between the medical sensors, environmental sensors, and the gateway are secured. If the channel is not secure, the medical and environmental sensors will be assigned their own keys like the rest of the sensor nodes. At this stage, all gateways and cluster heads will have the following overhead:

Memory Overhead = $i\_k + m\_k + pr\_m\_k + pseud\_f + h * hash\_f$ (16)

Computation Overhead = none (17)

Network messages Overhead = none (18)

Based on the previous analysis for Stage 1, it is clear that the overhead in Stage 1 for the addition of the privacy mechanisms is a memory overhead due to the addition of the keys and pseudorandom and hash functions.

Stage 2: Deployment and initialisation stage

In the suggested hospital scenario, it is assumed that there will be no need for the clustering process as explained previously. When computing the memory overhead, the size of the LEAP information table depends on the number of the individual keys, pairwise keys and AUTHs received from the next hop neighbouring nodes. The term *pr_id* refers to the pairwise keys shared between the gateway *g* and/or other cluster heads *c* and/or medical sensors *m* and/or environmental sensors *e* (in case of an insecure communication channel between the sensors and the gateways).

In general, the computation overhead at this stage is caused by the verification of the AUTHs received from all next hop neighbours, the computation of the pairwise keys of all neighbours, the computation of the pheromone level, the encryption of the ants (pheromone level and state of the node) to send to the neighbouring nodes, the encryption of the HELLO messages using pairwise keys and the updating of the information table.

At the gateway level, there are two possible scenarios for the overhead computation, one where the channel between the sensors and the gateway is assumed to be secured and another where the channel is insecure.

If the communication channel between gateways and sensors is secured, then there will be no need to encrypt the data being moved between the sensors and the gateway and there will be no need to overload the memory at the sensor level and the gateway with the computation of the pairwise keys. In addition, there will be no need to include the AUTH key chain in the LEAP information table because the gateway has to only secure the communication with the cluster head using the pairwise key and no local broadcasting will be required. Consequently, at the gateway level:

$$\begin{aligned}\textbf{Memory Overhead} &= p\_id + LEAP\ information\ table \\ &= p\_id + i\_k + pr\_id\ of\ CH \\ &= \textbf{\textit{3 extra Keys}}\end{aligned} \tag{19}$$

$$\begin{aligned}\textbf{Computation Overhead} &= compute\ (p\_id) + compute\ (pr\_id) + \\ &compute\ M\ A\ C\ of\ pr\_id + erase\ pairwise\ key\ initial\ key \\ &= \textbf{\textit{4 extra operations}}\end{aligned} \tag{20}$$

$$\begin{aligned}\textbf{Network Messages Overhead} &= send\ HELLO\ message\ to\ CH \\ &+ receive\ ACK\ from\ CH \\ &= \textbf{\textit{2 extra messages}}\end{aligned} \tag{21}$$

Based on the previous equations, the memory overhead at the gateway level will be due to: the pseudonym (*p_id*), the individual key shared between the gateway and the base station (*i_k*) and the pairwise key shared between the gateway and the cluster head (*pr_id*). The computation overhead will be caused by the computation of the pseudonym using the pre-stored hash functions, the computation of the pairwise key between the gateway and the cluster head, the authentication of this pairwise key and the erasure of the master key used for the pairwise key generation.

Finally, the network overhead is caused by sending the HELLO messages to the cluster head and receiving an ACK from the cluster head. It can be noticed that at the gateway level the overhead will always remain constant since the gateway is only connected to one cluster head. The memory overhead will be three extra keys, the computation overhead will be only four extra operations and the network messages overhead will be two extra messages.

On the other hand, if the communication channel between the gateway and the sensors connected to it (medical and environmental sensors) is not secured, then pairwise keys will need to be shared between the gateway and the sensors connected to it, to encrypt the data being communicated between them. In addition, the one-way authentication scheme will have to be applied to verify the broadcasting that a gateway might have to do for the sensors connected to it. A possible reason for broadcasting will be to announce the updated pseudonym of the gateway that is regularly changed after each successful transmission to the base station. To guarantee the authentication of the messages being sent and received between the different network nodes, an authentication key chain L is generated in each node using the LEAP mechanism. The authentication key chain is generated in each node and then the first key of the chain (AUTH Key) is encrypted, using the neighbour's pairwise key, and sent to the neighbour where it is stored (each node should keep the most recent AUTH keys received from all its next hop neighbours). AUTH keys are attached to messages sent to neighbours to verify the authenticity of these messages. Consequently, the following overhead will be added to the gateway in case of an insecure communication channel:

Memory Overhead = *p_id* + *LEAP information table*
= *p_id* + *i_k* + *pr_id* of CH + $\sum_{i=0}^{m-1} pr\_id(i) of\ medical\ sensor$ + $\sum_{i=0}^{e-1} pr\_id(i)\ of\ enviromental\ sensor$ + (*m*+*e*)*AUTH + L
= **3 + (2m + 2e)*AUTH + L extra keys** (22)

It is assumed that all keys and AUTH keys are the same length and that there is a minimum of one medical sensor and one environmental sensor. Up to this point, no sensor data has been forwarded yet, only keys are exchanged. Also, assuming that the key chain L is the same length as the number of sensors connected to the cluster head, Figure 7-28 presents the number of keys that need to be stored as the total number of medical and environmental sensors increase. It is clear from the graph that the increase is linear. The graph starts with a minimum of two sensors,

which require 9 keys for storage (2 pairwise keys for each sensor, p_id, i_k, pr_id of cluster head, the AUTH keys for the two sensors plus the L AUTH key chain, in this case a chain that is two keys long). The assumption that the AUTH key chain L is the same length as the number of neighbours is just an example for illustration but longer key chains can be used and their overhead can be directly calculated using the previous equations.

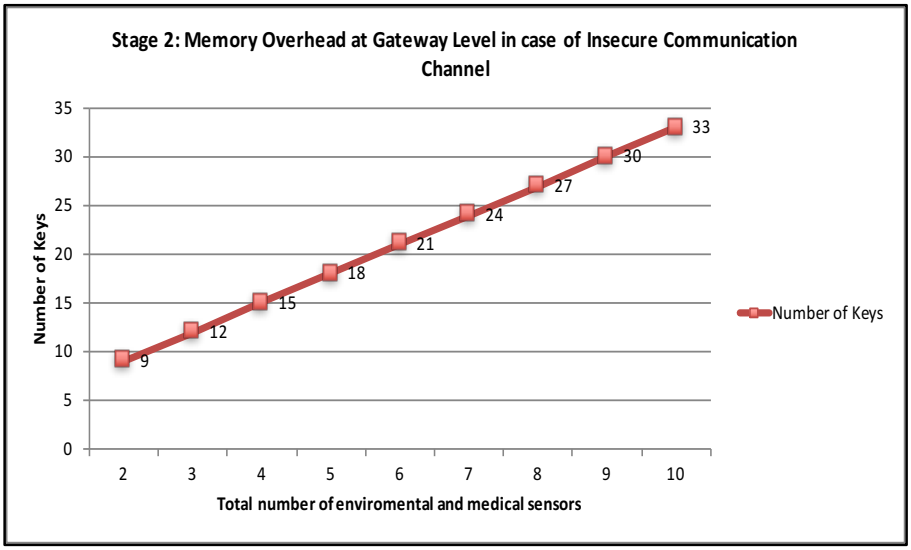


Figure 7-28 Memory overhead at the gateway level in Stage 2 (deployment and initialization Stage)

Computation Overhead = compute (p_id) + compute M A C * (*1+m+e*)+ $\sum_{i=0}^{m+e-1} compute\ pr\_id(i)$ + erase master pairwise keys + generate L key chain
= 4+2m+2e extra operations (23)

Figure 7-29 depicts the computation overhead at the gateway level as the number of sensors increases. The overhead is calculated starting with a minimum of two environmental and medical sensors. The overhead of the pairwise key computation, computation for the M A C for the cluster head and the pairwise master key erasure will remain constant, no matter what number of environmental and medical sensors is used, which accounts for three operations. As the number of sensors increases, the computation overhead will increase linearly because two operations are dedicated for every extra sensor: pairwise key computation and the authentication of the M A C and the generation for the key chain L.

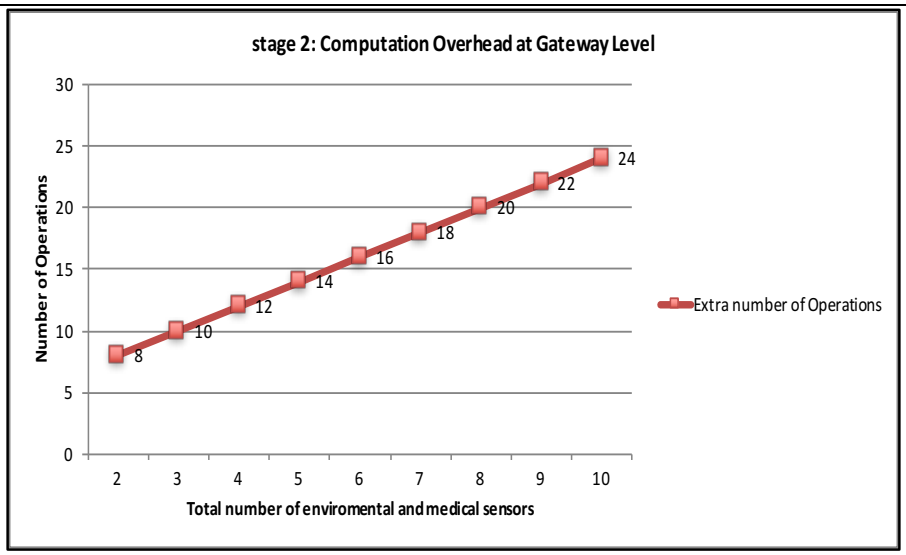


Figure 7-29 Computation overhead at the gateway level in Stage 2 (deployment and initialization Stage)

Network Messages Overhead =send HELLO msg to CH + receive ACK from CH + $\sum_{i=0}^{m+e} broadcast\ hello\ msg\ to\ compute\ pr\_id(i)$ + $\sum_{i=0}^{m+e} receive\ ACK\ to\ compute\ pr\_id(i)$

= 2 + 2m + 2e extra message (24)

Figure 7-30 depicts the network messages overhead at the gateway level as the number of sensors increases. Two basic overhead messages (HELLO message and ACK message of the cluster head) are always included no matter what the number of the environmental and medical sensors is. Two extra messages are added to the network for every sensor involved, one for the HELLO message and one for the ACK. This explains why the graph is a linear graph with constant increase as the number of sensors increases.

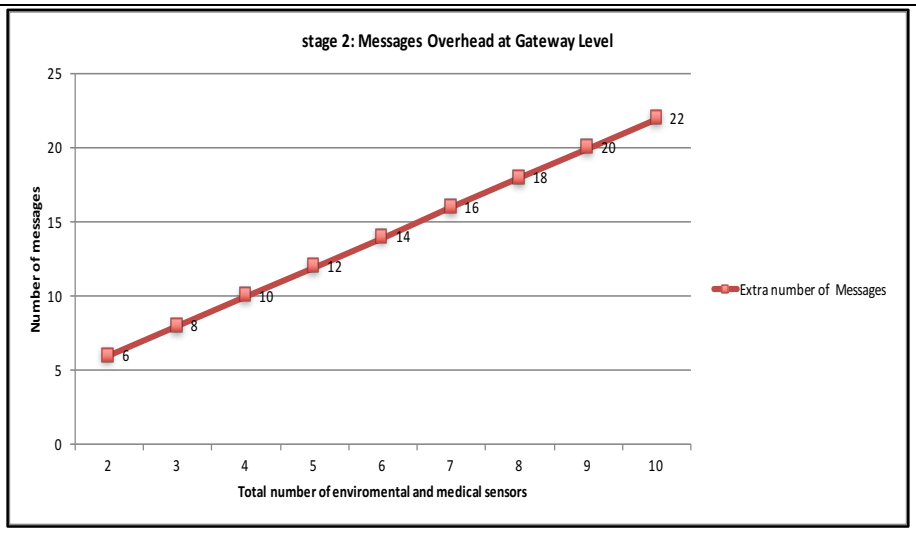


Figure 7-30 Network messages overhead at the gateway level in Stage 2 (deployment and initialization Stage)

At the cluster head level, there will be an AntSensNet information table which is a data structure that the routing protocol uses to store the pheromone information of the route from the current cluster head to the base station. This data structure consists of columns referring to the different application–dependent queuing models (traffic classes) and the rows referring to the neighbours. In the suggested hospital scenario, only two traffic classes will be considered, the *real-time, loss tolerant multimedia class* and the *real-time, loss intolerant data stream class*. The choice of real-time traffic classes is due to the immediate reporting of any emergencies or abnormal sensor reading, or sudden fall of a patient, to the monitoring station. Accordingly, the data structure will only have two columns for the chosen traffic classes. Each column contains five values: energy pheromone, delay pheromone, packet loss pheromone, available memory pheromone and the expiration time of each class.

Since the aim of this analysis is to compute the overhead, as compared to the basic underlying routing protocol (AntSensNet), then the memory needed for the storage of the AntSensNet information table will not be considered as a memory overhead because it is part of the AntSensNet basic operation. At the cluster head level, the expected overhead:

Memory Overhead = *p_id* + [*LEAP information table*]
= *p_id* + [*i_k* + $\sum_{i=0}^{c-1} pr\_id(i)$ of neighboring CH + $\sum_{i=0}^{g-1} pr\_id(i)$ of gateways + $\sum_{i=0}^{cam-1} pr\_id(i)\, of video\ sensors$ + $\sum_{i=0}^{aud-1} pr\_id(i) of\ audio\ sensor$ + (c +g+cam+aud)*AUTH] + L (25)

= 2 + 2c + 2g+ 2cam + L extra keys

As explained before and depicted in Figure 6-1, the video and audio sensors will be connected directly to the cluster heads and not the gateways to save the energy of the gateways. Figure 7-31 is a graph that presents the memory overhead at the cluster head level. In Figure 7-31, it is assumed that all keys and all AUTHs keys are the same size, and that as a minimum there will be one gateway, one video sensor and one audio sensor and that the cluster head is directly connected to the base station. Based on Equation (25), it is clear that two basic keys are used no matter what the total number of sensors are, (p_id and i_k) and two extra keys are introduced for every cluster head and/or gateway and/or video and/or audio sensor introduced. Again the length of the AUTH key chain will be assumed to be equal to the number of sensors connected to the cluster head.

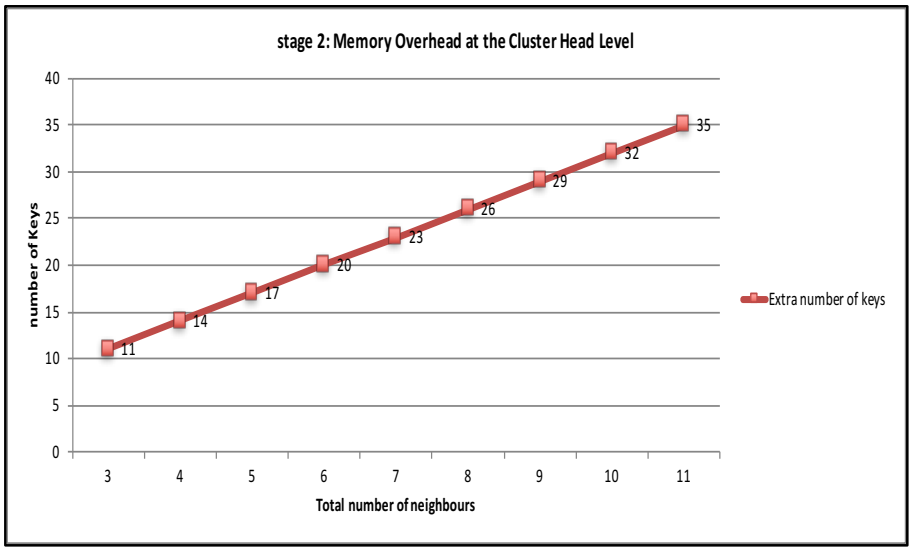


Figure 7-31 Memory overhead at the cluster head level in Stage 2 (deployment and initialization stage)

To estimate the computation overhead, the basic operations already existing in the AntSensNet protocol are not considered and only the extra operations due to the introduction of the security and privacy mechanisms are included. The computation overhead can be calculated as follows:

Computation Overhead = compute (p_id) + compute M A C * (c+g+cam+aud)+ $\sum_{i=0}^{c+g+cam+aud-1} compute\ pr\_id(i)$ + $\sum_{i=0}^{c-1} encrypt\ (phermone\ value + node\ state\ using\ pr\_id(i))$ + *erase master pairwise keys + generate L key chain*
= (3+2c+2g+2cam+2audio) computations + c encryptions extra operations (26)

Since the clustering process will not be used in the hospital scenario because the cluster heads are already mounted to the wall and their location is predetermined, the broadcasting of the pheromone and node state will be used to construct the AntSensNet routing information table, for data routing. For each cluster head, there is a minimum of one gateway, one video sensor and one audio. The chart in Figure 7-32 depicts the percentage of the number of the encryption operations to the number of key generation operations. Since the two operations require different processing powers, this chart will show the percentage for different number of sensors. As the number of sensors increases, the overall percentage of encryption to key generation operations will decrease because, based on Equation (26), the introduction of an extra sensor will cost two extra operations compared to c constant encryptions (related to the number of the cluster heads) no matter what the number of sensors are.

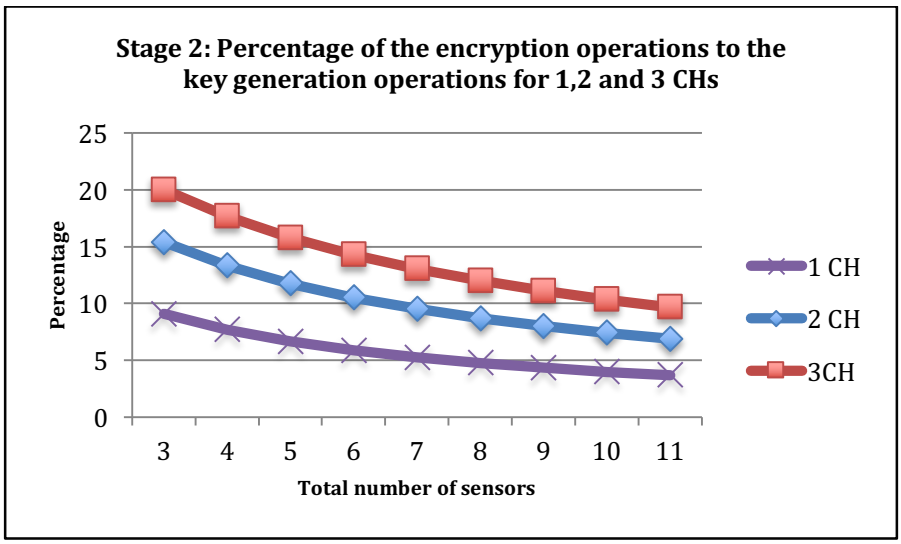


Figure 7-32 Percentage of encryption operations to key generation operations in Stage 2 (deployment and initialization stage)

$$\textbf{Network Overhead} = \textbf{Messages} \sum_{i=0}^{c+g+cam+aud-1} broadcast\ hello\ msg\ to\ compute\ pr\_id(i) + \sum_{i=0}^{c+g+cam+aud-1} receive\ ACK\ to\ compute\ pr\_id(i) + \sum_{i=0}^{c} send\ encrypted\ phermone\ value\ and\ state$$
$$\textbf{= 3c + 2g+ 2cam+ 2aud extra message} \quad (27)$$

Figure 7-33 is a chart which presents the message overhead at the cluster head level when using 1, 2 and 3 cluster heads, assuming that as a minimum there will be one gateway, one video sensor and one audio sensor and the cluster heads are directly connected to the base station. Based on Equation (27), for every cluster

head; the increase of the number of messages will be two extra messages per sensor, which explains the linear increase of the graphs.

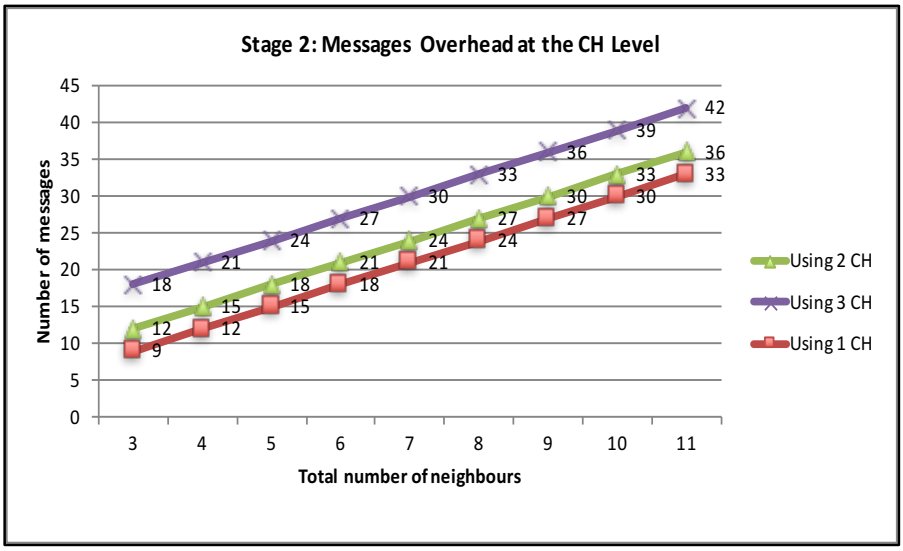


Figure 7-33 Messages overhead at the cluster head level in Stage 2 (deployment and initialization stage)

Stage 3: Traffic forwarding

In the traffic forwarding stage, the normal operation of the gateway should be the gathering of the data from the medical and environmental sensors connected to it and sending the data to the cluster head. The expected overhead compared to the underlying AntSensNet routing protocol is as follows:

At the gateway level

Memory Overhead = none (28)

Computation Overhead = $\sum_{i=0}^{m+e-1} aggregate\ data$ + encrypt data using p_id of CH
= (m+e) aggregation operations and 1 encryption operations (29)

Figure 7-34 shows the percentage of aggregation operations to the encryption operations. Normally, a gateway should be connected to one cluster head (a gateway belongs to one cluster). Consequently, the gateway will need to encrypt once to send the data to the cluster head, but it will need to aggregate the data from m+e sensors. Consequently, as the number of sensors increases, the encryption will always remain one, which explains why the percentage curve decreases.

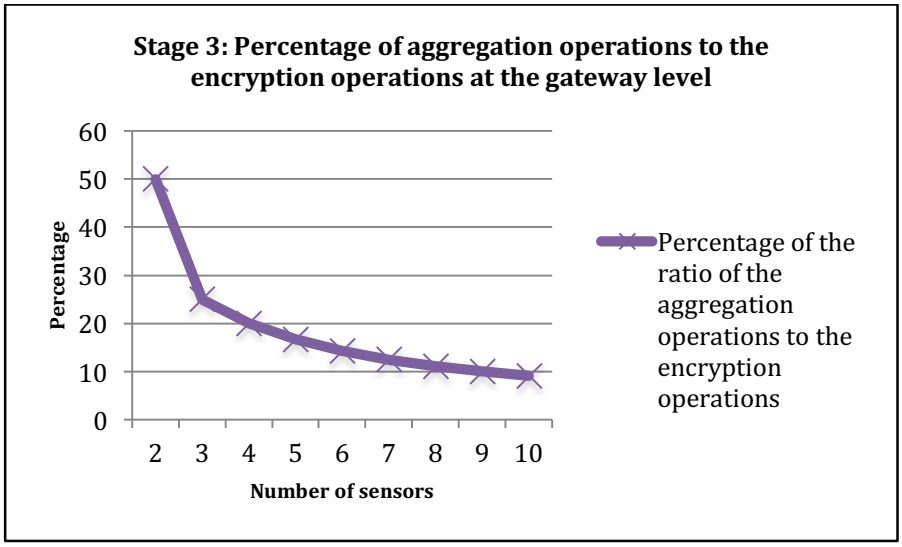


Figure 7-34 The percentage of aggregation operations to the encryption operation at the gateway level in Stage 3 (traffic forwarding)

Network Messages Overhead = none (30)

At the cluster head sending scalar data: The fake packet generation rate is dynamic depending on the size of the network. Generally, the closer the cluster head is to the base station, the more traffic it will route towards the base station. This will cause more communication overhead at the cluster heads near the base station than those away from the base station. If all nodes are instructed to generate fake messages, this will add even more overheads to the cluster heads near the base station. Consequently, the base station can estimate the size of the network and instruct the cluster heads away from the base station, which have lower traffic, to generate fake packets to even out the traffic all over the network. Again, since the aim of this analysis is to compute the worst-case overhead, the following equations compute the overhead at the far away nodes that will be adding fake traffic.

Memory Overhead = none (31)

Computation Overhead = $\sum_{i=0}^{c-1} encrypt\ FANT\ with\ p\_id(i)$ + [processing the BANT] + pad data packets to look like ants + fake packet generation

= $\sum_{i=0}^{c-1} encrypt\ FANT\ with\ p\_id(i)$ + decrypt the BANT + pad data packets to look like ants + fake packet generation

= (c+1) encryptions and decryptions + 2 packet related extra operations.

(32)

In addition to the basic AntSensNet related operations that a cluster head has to perform such as generating FANTs, receiving BANTs and other tasks, the privacy mechanisms have added more tasks to the cluster head as depicted in the computation overhead equation. Among these tasks is the size correlation between the data packets and the ants packets to be the same so that an adversary cannot distinguish the routing packets (ant packets) from the data packets. Tracking an ant packet can easily tell an adversary that the source node generating the ant packet needs to send data to the base station, which is why size correlation of the ant and data packets may serve to maintain better source location privacy for the nodes. Unfortunately, the authors of the AntSensNet routing protocol did not discuss the size of the ants in their proposed protocol. They only presented the main structure of the ant but did not specify or suggest an appropriate size for the ants even though in the experimental results the authors have clearly stated the size of the data packets (32 bytes) and the multimedia packets (1024 bytes) they used in their simulation. The authors have only mentioned that one field of the ant structure is a stack that stores the nodes visited by the FANT from the source node to the base station which implies that the FANTs grow as the number of nodes they pass by increases. The authors of (Caro et al., 1998) have mentioned that the ant based algorithm they proposed considers the size of the ants as parameters that must be set. For example the AntNet algorithm set the ant size as 24 bytes + 8 bytes * N_h where N_h is the number of hops made by the FANT (Caro et al., 1998). Following similar settings for the FANT size as (Caro et al., 1998), an average size for the FANT can be estimated depending on the location of the cluster head from the base station and the data packets can be padded to be the same size as the average size of the ants. In addition, the FANTS originating from the nodes can be the same size as the data packets and even if they need to be increased to add more nodes to the stack, an adversary cannot tell if the increase happened at the source node or not. Figure 7-35 depicts a plot of the percentage of encryption/decryption operations to packet operations at the cluster head level.

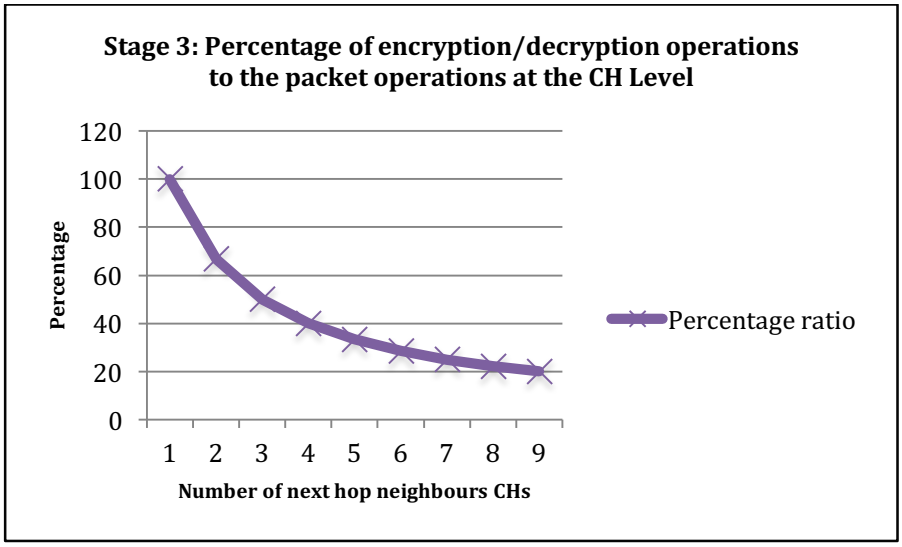


Figure 7-35 The percentage of encryption/decryption operations to aggregation operations at the cluster head level in Stage 3 (traffic forwarding)

Network Messages Overhead = Fake messages generation overhead (33)

The network communication at the cluster head involves the broadcasting of the FANTS to the neighbouring cluster heads, routing data to the cluster heads and receiving BANTs. However, these will not be considered in the overhead calculations because they are already included in the original AntSensNet protocol. Consequently, the network messages overhead will be the fake messages only. The percentage of the privacy related operations (fake messages generation) to the basic AntSensNet operations ($\sum_{i=0}^{c-1} broadcast\ encrypted\ FANT$ + receive BANT + send date to next CH) would be 1 operation to c+2 operations.

When an FANT is received at an intermediate cluster head on the way to the base station, the cluster head must update the FANT information. The cluster head decrypts the FANT using the shared p_id between the current cluster head and the one that sent it. If this FANT has not been received before (Loop FANT), the cluster head will update the information (energy, delay, packet loss and memory) of the FANT. After the update, the cluster head computes the normalised pheromone value for all the cluster heads on the way to the base station that this FANT have not passed by. The pheromone values for all the cluster heads are compared and the next hop cluster head is selected. The overhead due to the privacy mechanisms will be as follows:

Memory Overhead = none (34)

Computation Overhead = [processing FANT] + [processing BANT] + pad data packets to look like ants
= [decrypt FANT info + $\sum_{i=0}^{c-1}$ encrypted FANT using p_id] + [decrypt the BANT + encrypt the BANT using the p_id of the next CH] + pad data packets to look like ants + fake message generation
= (3+c) encryption/decryptions operations + 2 packet operations (35)

Network Messages Overhead = **fake message generation** (36)

Pseudonym Update: To ensure that a global adversary studying the network does not link the messages to particular gateways (i.e. to achieve unlinkability), the pseudonyms are constantly updated. The pseudonym is updated after each successful transmission of data packets. A hash function is picked by the sensor and the pseudonym is updated. Once the new pseudonym is generated, the gateway notifies the cluster head using encrypted messages that its pseudonym has been updated. The same procedure is applied at the level of the cluster head where the cluster head updates its pseudonym and notifies the neighbouring nodes with this update.

In case of an insecure communication channel, the expected overhead due to the pseudonym update at the sensors level will be:

Memory Overhead = none (37)

Computation Overhead = apply the hash function + encrypt new pseudonym
= 1 operations + 1 encryption (38)

Network Messages Overhead = send 1 message to gateway
= 1 extra message (39)

At the gateway level

Memory Overhead = none (40)

Computation Overhead = apply the hash function + encrypt new pseudonym
= 1 operation + 1 encryption (41)

Network Messages Overhead = send 1 message to CH
= 1 extra message (42)

At the cluster head Level

Memory Overhead = none (43)

Computation Overhead = apply the hash function + $\sum_{i=0}^{c-1}$ encrypt new pseudonym
= 1 operation + c encryption (44)

Network Messages Overhead = $\sum_{i=0}^{c-1}$ send new pseudonym message to next hop CHs
= c extra messages (45)

In the hospital scenario, it is assumed that multimedia data is required to be of relatively good quality. Consequently, special video ants (called VANTS) will be deployed to detect multiple link-disjoint paths to send multiple multimedia packets at a time. Following the same scenario as scalar data, forward VANTS called VFANTs and backward VANTs called VBANTs are deployed to determine the paths to the sink. In this case, the extra overhead that will be introduced at the cluster head due to the sending of the video data will be as follows:

Memory Overhead = none (46)

Computation Overhead = $\sum_{i=0}^{c-1} encrypt\ VFANT\ with\ p\_id(i) +$ decrypt the VBANT + fake message generation
= (c+1) encryptions and decryptions + 2 packet related operations (47)

The computation overhead due to the introduction of the privacy mechanisms in case of sending multimedia traffic, as depicted in the equation, will have a percentage of two packet operations to (c+1) encryptions and decryptions.

Network Messages Overhead = fake messages (48)

Table 7-17 Summary of overheads at each stage of the AntSensNet protocol, for the hospital scenario

Overhead / Stage	At gateway			At cluster head		
	Memory Overhead	**Computation Overhead**	**Network messages Overhead**	**Memory Overhead**	**Computation Overhead**	**Network messages Overhead**
Pre-deployment Stage	i_k + m_k + pr_m_k + pseud_f + h* hash_f	None	None	i_k + m_k + pr_m_k + pseud_f + h* hash_f	None	None
Deployment and Initialization stage	Insecure Communication 3 extra Keys Secure Communication 3 + (2m + 2e)*AUTH + L extra keys	Insecure Communication 4 extra operations Secure Communication 4+2m+2e extra operations	Insecure Communication 2 extra message Secure Communication 2 + 2m + 2e extra message	2 + 2c + 2g+ 2cam + L extra keys	(3+2c+2g+2cam+2audio) computations + c encryptions extra operations	3c + 2g+ 2cam+ 2aud extra message
Traffic Forwarding	None Pseudonym Update None	(m+e) aggregation operations and 1 encryption operations Pseudonym Update 1 operations + 1 encryption	None Pseudonym Update 1 extra message	Scalar Data None At intermediate cluster head None Pseudonym Update None Video Overhead None	Scalar Data (c+1) encryptions and decryptions + 2 packet related extra operations At intermediate cluster head (3+c) encryption/decryptions operations + 2 packet operations Pseudonym Update 1 operation + c encryption Video Overhead (c+1) encryptions and decryptions + 2 packet related operations	Scalar Data Fake messages generation overhead At intermediate cluster head fake message generation Pseudonym Update c extra message Video Overhead fake messages

7.2.2.1.2 Comparison between theoretical and simulation analysis

The results discussed in Section 7.3.1, represent the analysis using simulation of the privacy-aware ant routing, whereas the results discussed in Section 7.3.2.1 represent the overhead using theoretical analysis. Based on the theoretical analysis summarized in Table 7-17, the total computational overhead at the level of the cluster heads is [(3+2c+2g+2cam+2audio) computations + c encryptions extra operations] + [(c+1) encryptions and decryptions + 2 packet related extra operations] + [1 operation + c encryption] + [(c+1) encryptions and decryptions + 2 packet related operations]. Based on Figure 5-1 (upon which the simulation experiment was based), the total number of cluster heads *c* is 7, the total number of gateways *g* is 6 and a total of 4 multimedia sensors (video and audio).

Consequently, the computation overhead at the cluster head for 1 scalar sensor under each gateway (similar to the parameters set for the simulation experiment in Section 7.3.1) equals [(3 + 2*7 + 2*6) computations + 7 encryptions] + [8 encryption + 2 packet operations] + [1 operation + 7 encryptions]. Referring to Table 7-4 and Table 7-6, at time 50 seconds, the mean number of clock ticks when no privacy or security is applied is 240807.66 and the mean number of clock ticks when both privacy and security are applied is 1579678.48. Dividing the second number by the first number, it can be assumed that the encryption operations require $\frac{1579678.48}{240807.66} = 6.55$ times the computation power of other packet operations. According to (Corporation, 2017), the latency of the arithmetic operations (such as addition, subtraction, multiplication or division) is equal (in case of addition and subtraction) or more (in case of multiplication and division) than shift operations (used in packet operations (such as padding)). Consequently, assuming, as a worst-case scenario, that both packet operations and computation operations require the same number of clock ticks. As a result, for 1 scalar sensor, the theoretical-based computation overhead is [3 + 2*7 + 2*6 + 7*6.55] + [8*6.55 + 2] +[1 + 7*6.55] ≅177 operations. Another example, at time 250 seconds, the encryption operations require $\frac{1610060.61}{247117.70} = 6.51$ times the computation power of other packet operations. Consequently, for 1 scalar sensor, the theoretical-based computation overhead is [3 + 2*7 + 2*6 + 7*6.51] + [8*6.51 + 2] +[1 + 7*6.51] ≅ 176 operations.

Similarly, the theoretical-based computation overhead for 1 scalar sensor and 1 multimedia sensor (similar to the parameters set for the simulation experiment in Section 7.3.1) equals [(3 + 2*7 + 2*6 + 2*1) computations +7 encryptions] + [8

encryptions + 2 packet operations] + [1 operation + 7 encryptions] + [8 encryptions + 2 packet operations]. Referring to Table 7-7 and Table 7-9, at time 50 seconds, the mean number of clock ticks when no privacy or security is applied is 276524.90 and the mean number of clock ticks when both privacy and security are applied is 3729050.90. Dividing the second number by the first number, it can be assumed that the encryption operations require $\frac{3729050.90}{276524.90} = 13.49$ times the computation power of other packet operations. In addition, assuming that both packet operation and computation operations require the same number of clock ticks. As a result, for 1 scalar sensor and 1 multimedia sensor, the theoretical-based computation overhead is [(3 + 2*7 + 2*6 + 2*1) computations +7*13.49] + [8*13.49 + 2 packet operations] + [1 operation + 7*13.49] + [8 *13.49 + 2 packet operations] ≅ 441 operations. Another example, at time 250 seconds, the encryption operations require $\frac{4044556.60}{299285.60} = 13.51$ times the computation power of other packet operations. As a result, for 1 scalar sensor and 1 multimedia sensor, the theoretical-based computation overhead is [(3 + 2*7 + 2*6 + 2*1) computations +7*13.51] + [8*13.51 + 2 packet operations] + [1 operation + 7*13.51] + [8 *13.51 + 2 packet operations] ≅ 442 operations.

Dividing the number of operations of 1 scalar sensor and 1 multimedia sensor by the number of operations of 1 scalar sensor equals almost 2.5 times compared to 2.38 times (based on the simulation results for the experiment in Section 7.3.1). This shows that the results of both the theoretical-based analysis and simulation analysis are close with a 4.8% difference between them.

Applying the same concept on 1 scalar and 2 multimedia sensors, at time 50 seconds, encryption operations are 13.1 times the computation operation; the theoretical-based computation overhead is almost 429 operations. [Note the encryption overhead decreased because less packets are generated due to the competition between the sensors to send packets on the shared bandwidth]. Comparing the ratio of the number of operations between 1 scalar sensor and 2 multimedia sensors with 1 scalar sensor and 1 multimedia sensor with that of the results of the simulation experiment, the difference is almost 8.5%. For 2 scalar and 2 multimedia sensors and assuming the encryption operations is 10.8 times the computation operation; the computation overhead is 360 operations. Comparing the ratio of the number of operations between 1 scalar sensor and 2 multimedia sensors with 2 scalar sensors and 2 multimedia sensor in the theoretical-based analysis with that of the results of the simulation experiment, the difference of almost is 14%.

7.3.2.2 Analysis of overheads for the elderly house scenario

The same analysis for the hospital scenario (see Section 7.3.2) will apply to this scenario but all operations related to the AntSensNet protocol will be discarded for all cluster heads inside the house except for the last cluster head (before the dashed line) that routes the data outside the house.

7.3.2.3 Analysis of overheads for the battlefield scenario

In a battlefield scenario, it is very hard to predict the geographical location of the cluster heads and the base stations. Consequently, in the analysis of this scenario, the AntSensNet clustering algorithm will be used to elect the cluster heads, and the gateways will have to connect to the closest cluster head in range. Environmental, video and audio sensors can be scattered on the battlefield and they can directly connect to the nearest cluster heads in their range.

Stage 1: Pre-deployment stage

Similar to the previous scenarios, the hospital and the elderly house, each sensor involved in the communication of data will be assigned: a pseudorandom function, an individual key, a master key and *h* hash functions. All sensors in the network will be assigned a pseudorandom function of length *pseud_f* bits, an individual key of length *i_k* bits, a master key of length *m_k* bits, another master key of length *pr_m_k* bits for the pairwise communications and *h* hash functions, each is of length *hash_f* bits. Although military-grade security would be deployed for all communication on a battlefield, in this research work, all communications are considered insecure and all sensors will be assigned keys for data encryption. This is to present the worst-case (maximum) overhead in the following analysis.

Even though more than one base station is being deployed; the keys and hash functions will be identical. Later on, each base station will determine the update rate of the pseudonyms and the hash functions of the cluster heads belonging to it. For example, one base station might require the update of the pseudonyms using the same sequence of the hash functions preloaded into the sensor nodes and another base station may instruct the nodes to use one hash function and then skip one or more functions in the preloaded sequence.

Stage 2: Deployment and initialisation stage

Contrary to the previous two scenarios (hospital and elderly house), clustering must be considered in this scenario because the layout of the cluster heads can never be

predicted. Consequently, extra overhead will be added (compared to the previous scenarios) due to the consideration of the clustering operations. Since the multimedia sensors (cameras and audio sensors) should be more powerful (computation, memory and energy) compared to scalar sensors, only multimedia sensors will be involved in the clustering process. Other scalar sensors (wearable, implanted and environmental), and gateways will not participate due to the limited computation and energy power of the scalar sensors and to protect the energy of the gateways from depleting due to the overhead of the clustering operations.

Although the clustering operation is part of the original AntSensNet protocol, it was not deployed in the previous scenarios due to the predefined layout of the cluster heads that is designed to cover all the areas in the hospital or the elderly house. However, since the battlefield scenario cannot have a predefined layout for the cluster heads, the clustering operation must be done in order to select the appropriate cluster heads. Consequently, the whole clustering operation will be considered an overhead.

The clustering operation will commence after the establishment of the pairwise keys. Since the aim of this analysis is to present the maximum (worst case scenario) overhead, then the analysis for the cluster head will be one of three cases: the CANT arrives at a sensor node that is already a cluster head, the CANT arrives at a node that is not a cluster head but has a cluster head in range, or the CANT arrives at a node that is not a cluster head and does not have a cluster head in range. Similar to the ants in the previous scenarios, the CANTs will be encrypted to suppress an adversary from learning the content of the CANT if it is captured.

Case 1: If the CANT arrives at a cluster head, then this cluster head will decrypt the CANT, decrement the TTL and re-encrypt the CANT. If the TTL is zero, then the CANT will be destroyed else the CANT will be routed to a randomly picked sensor node. Assuming the TTL is not zero, then two encryption operations (one decrypting the CANT and another re-encrypting the CANT) and 2 other operations (decrementing of TTL and P_c calculation) (refer to flowchart in Figure 5-6) are required and 1 communication operation (forward the CANT to the next hop neighbour).

Case 2: If the CANT arrives at a sensor node that is not cluster head and has a cluster head in range, then the sensor node will re-route the CANT based on the

random choice of another sensor node using the probability P_c. In this case, 1 computation and 1 communication operation are required.

Case 3: The CANT arrives at a sensor node that is not a cluster head and has no cluster head in its range, then the sensor node will decrypt the CANT, store it and broadcast to all neighbours within $R_{cluster}$ that this sensor node has become a cluster head. In this case the computation overhead will be multiple encryption/decryption operations (1 for the decryption of the CANT and multiple encryption for the ADV_cluster for the nodes in the neighbourhood), the memory overhead will be the storage of the CANT and the communication overhead will be the broadcasting of the ADV_cluster to all sensor nodes within the neighbourhood. Since the broadcasting of the ADV_cluster requires the encryption of this message using the pairwise key of each sensor node in range, case 3 is considered the worst-case scenario due to the large number of encryption operations and communication operations.

Clustering overhead at the audio and video sensors

Memory Overhead = CANT (49)

Computation Overhead = Decrypt the CANT + $\sum_{i=0}^{g-1} pr_{id(i)} of\ gateways\ to\ encrypt\ ADV\_cluster$ + $\sum_{i=0}^{aud+cam+e-1} pr\_id(i)\ of\ sensors\ to\ encrypt\ ADV\_cluster$
=1+g+cam+aud+e encryptions (50)

Network Messages Overhead= (51)
$\sum_{i=0}^{g+\ aud+cam+e-1} broadcast\ ADV\_cluster$
= g+aud+cam+e messages

At the gateway level, the expected overhead will be more than that in the hospital and the elderly house scenario due to the extra number of keys for all sensor nodes connected to the gateway. The gateway will store a pairwise key for every sensor node (either medical, environmental, video or audio). In addition, the gateway will have to store the most recent AUTH keys received from all its next hop neighbours and the L AUTH key chain. Accordingly, the memory overhead will basically be four keys at each gateway (one pseudonym, one individual key, one pairwise key for the cluster head to which the gateway is connected and one recent authentication key for the communication with the cluster head). For every extra sensor connected to the gateway, two extra keys will be allocated (one for the pairwise key between the sensor and the gateway and one authentication key) and the L key chain.

Memory Overhead = *p_id* + *LEAP information table*
= *p_id* + *i_k* + *pr_id* of CH + $\sum_{i=0}^{m-1} pr\_id(i) of\ medical\ sensors$ + $\sum_{i=0}^{e-1} pr\_id(i)\ of\ enviromental\ sensors$ + $\sum_{i=0}^{cam-1} pr\_id(i) of\ video\ sensors$ + $\sum_{i=0}^{aud-1} pr\_id(i) of\ audio\ sensor$ + (*1+m+e+cam+aud*)*AUTH + L
= **4 + 2m + 2e +2cam + 2aud + L** (Assuming that the keys and AUTHs have the same size) (52)

The computational overhead at the gateway will be due to the computation of the pairwise key between the cluster head and gateway, computing the M A C for the cluster head, erasing the pairwise master key and generating the AUTH chain L. In addition, there will be two extra operations for every sensor connected to the gateway (one operation for the computation of the pairwise key and one for the MAC).

Computation Overhead = compute (p_id) of CH + compute M A C * (*1+m+e+cam+aud*)+ $\sum_{i=0}^{m+e+cam+aud} compute\ pr\_id(i)$ + erase master pairwise keys + generate L key chain
= 4+2m+2e+2cam+2aud extra operations (53)

The network overhead is similar to the previous two scenarios. However, in the previous two scenarios, video and audio sensors were not directly connected to the gateway. For every extra sensor connected to the gateway, two extra messages will be generated (one HELLO message and one AUTH message).

Network Messages Overhead =send Hello msg to CH + receive ACK from CH+ $\sum_{i=0}^{m+e+cam+aud} broadcast\ hello\ msg\ to\ compute\ pr\_id(i)$ + $\sum_{i=0}^{m+e+cam+aud} receive\ AUTH\ to\ compute\ pr\_id(i)$
= 2 + 2m + 2e + 2cam+ 2aud extra messages (54)

As stated in the hospital scenario, the AntSensNet information table consists of columns referring to the different application–dependent queuing models (traffic classes) and the rows referring to the neighbours. In the battlefield scenario, only two traffic classes will be considered which are real-time, loss tolerant multimedia and data stream and *real-time, loss tolerant data stream class*. Accordingly, the data structure will only have two columns for the chosen traffic classes to be used to send both multimedia and data packets. Unlike the hospital and the elderly house scenario, the high quality video transmission will not be considered in this scenario to decrease the overhead. Consequently, the VFANTs and VBANTs will not be considered in the analysis of this scenario.

The rest of the analysis for the computation of the overhead at the cluster head level and the traffic forwarding will be the same as in the previous two scenarios. It is important to note that extra network messages will be generated at the base stations. These network messages are generated to instruct the cluster heads to adjust their pseudonyms using a specific order for the hash functions deployment; and to instruct each cluster head to change the percentage of the fake messages generation.

7.4 Summary

The aim of this chapter is to investigate the overhead due to the addition of the privacy enhancing mechanisms: anonymity/pseudonomity, unlinkability and location privacy (identified in Chapter 4). In this chapter, the three scenarios introduced in Chapter 5 (Hospital, elderly house and battlefield scenarios) were simulated on NS2. For each scenario, a plot of average end-to-end delay, throughput, percentage of packet delivery ratio and percentage of packet loss ratio were presented. Next, a simulation of one of the deployment scenarios (hospital scenario) to assess the overhead due to the introduction of the privacy measures was presented. The simulation showed almost seven times overhead due to the introduction of privacy measures for scalar data compared to almost fourteen times overhead due to the application of the privacy measures for multimedia data. This indicates that in critical medical cases when quick intervention of medical help is required, the communication of the multimedia data should be kept to the minimum. However, the simulation showed the overall overhead and no details of the causes of the overhead were available. Consequently, a theoretical analysis was used to assess the memory, computation, and network messages overhead due to the introduction of the privacy mechanisms. The overhead is represented in the form of equations outlining the extra keys (memory overhead calculation), the number of extra operations (computation overhead) and the number of extra messages (network messages overhead) compared to the original AntSensNet routing protocol. This representation of the overhead in the form of equations makes it easier to calculate the overhead for diverse scenarios and different network configurations with varying key sizes and other privacy related factors (such as the rate of fake packet generation) unlike other assessment methodologies such as simulators in which a certain network topology or a particular scenario is tested using a specific number of sensors. Both the results of the simulation-based analysis and the theoretical-based analysis were compared. Both analyses showed a nearly linear increase in the computation overhead due to the addition of privacy. The comparison between the results of simulation and

theoretical analysis showed a maximum difference of 14.5% and a minimum difference of 4.8%.

It is clear from the theoretical analysis that the introduction of privacy protection mechanisms added memory, computation and network messages overhead on the three operational stages of the routing protocol (except in the pre-deployment stage in which only memory overhead was added). However, the added overhead is mainly at the levels of the gateways and the cluster heads. Overhead is added at the level of the sensors only in case of unsecured communication channels between the sensors and the gateways. The absence of overhead at the sensors levels will not cause any extra energy depletion compared to the normal operation of the sensors in systems where no privacy measures are added. In an unsecured communication channel, the memory, computation, network messages overhead cause more energy consumption by the sensors and decrease the lifetime of the sensors. However, in critical situations, overhead can be ignored where privacy of a patient is a very high priority. The trade-off between the privacy and the lifetime of the sensors should be critically assessed.

For today's technology the privacy overhead might take up the limited memory space of healthcare sensors, or add more effort on the limited computation power of these sensors, or add more delay due to the increased number of network messages. However, with the frequent advances in the hardware technology of healthcare sensors and the network connections, the overhead arising from privacy protection mechanisms could be negligible for tomorrow's technology.

The following chapter focuses on the assessment of the enhancement of the privacy achieved after the introduction of the privacy mechanisms into the WMSN-based healthcare sub-system.

Chapter 8 Assessment of the Enhancement of Anonymity, Unlinkability and Location Privacy Due to Fake Traffic

8.1 Introduction

The previous chapter discussed how the overhead due to the introduction of the privacy measures into a WMSN-based healthcare sub-system was estimated using both NS2 simulation and theoretical analysis. The aim of this chapter is to discuss the assessment of the privacy of the WMSN-based healthcare sub-system after the introduction of the privacy measures. This chapter presents methods, results and discussion of the simulation experiments that were conducted to study the effect of the introduction of selected privacy countermeasures, as discussed in the previous chapters, on the level of privacy of a proposed healthcare sub-system.

8.2 Simulation experiments

8.2.1 Aims

The systematic experiments described in this section aim to analyse the enhancement of the overall level of privacy after the introduction of the privacy mechanisms to enhance anonymity, unlinkability and location privacy. In specific, these experiments focus on how the injection of fake traffic enhances the level of privacy of the system used for illustration, and the drawback of the concept of fake traffic. Although the idea of fake traffic has been studied and discussed in the literature, to the best of the author's knowledge the study of the effect of fake traffic in WMSN has not been extensively researched.

All data sent by sensors (wearable, implanted, audio and video) were size correlated and encrypted using the LEAP key management protocol (see Section 5.7) and constantly changing pseudonyms were used (to hide the identity of the sender). The introduction of fake traffic is expected to enhance unlinkability, location privacy and anonymity. It is envisaged that if an adversary captures a message of the network, he/she cannot:

- tell whether this message is real or fake.
- link the captured message to a previous one or a particular sender (due to the constantly changing pseudonyms and due to the fake traffic which decrease the probability that messages belong to a particular sender).

- relate the message to a particular sender (as the fake traffic increases, the location privacy is enhanced because the real messages are hidden within fake messages from different fake generators. An adversary capturing a message would not be able to tell whether it is a real or fake one. In addition, if there is only one sender, fake traffic will trick the adversary into thinking there is more than one sender).
- link the captured message to a particular source (sources are tightly coupled to patients, the introduction of fake traffic decreases the probability that a particular source is the origin of the message and thus increase the level of the anonymity).
- determine the exact healthcare sensor(s) worn by the patient, because the frequency of the message generation, which can be tied to a particular sensor, is altered due to the introduction of fake traffic.

8.2.2 Method

8.2.2.1 Equipment

All experiments were conducted using simulations deployed in the NS2 network simulator, on a personal computer. This computer has the following specifications: MacBook Pro, processor 2.66 GHz Intel Core i7, memory 4 GB 1067 MHz DDR3 and operating system macOS Sierra.

8.2.2.2 Experiment design

Based on the literature (as depicted in (Milenković et al., 2006)), healthcare scenarios follow a multi-tier hierarchal architecture (see Figure 2-3) in which sensors (in Tier 1) collect, sample and process data to send to personal servers or cluster heads (in Tier 2). Finally, Tier 2 sends the data to the medical servers or base station (in Tier 3). In this research work, a similar multi-tier hierarchal healthcare scenario is simulated in NS2 to assess the enhancement in the level of privacy after the introduction of the privacy countermeasures.

No multi-hop cluster head communication is deployed in the experiments to ensure that the analyses are not dependent on a particular routing strategy and that the results of the analysis are general and not valid for ant routing or any routing protocol in particular. The experiments were conducted on a topology designed for a small hospital or health facility, as depicted in Figure 8-1. This hospital is made up of four floors. There are two cluster heads on each floor. Each floor contains four rooms, and it is assumed that one patient is staying in each room. In Figure 8-1, cluster

heads CH_1 and CH_2 belong to the first floor, cluster heads CH_3 and CH_4 belong to the second floor, cluster heads CH_5 and CH_6 belong to the third floor and, cluster heads CH_7 and CH_8 belong to the fourth floor. The gateways collect data from sensors and route them to the cluster heads. Each gateway is responsible for collecting the readings of the sensors of one patient (i.e. there is a one-to-one link between each gateway and the patient who uses this gateway; hence, if an adversary discovers the identity of the gateway, it will directly be related to the patient).

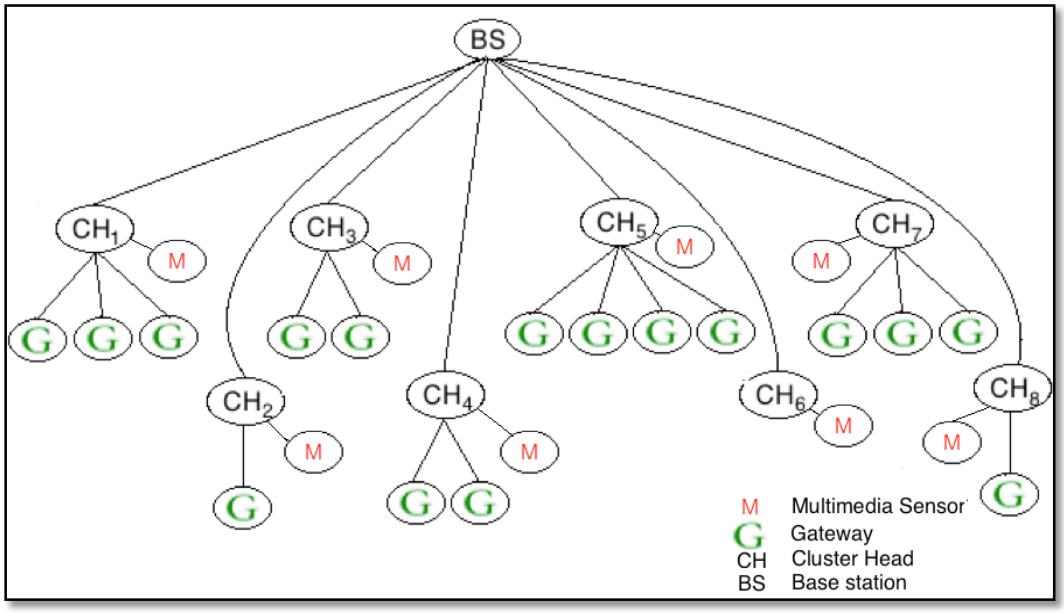


Figure 8-1 Layout of the network components for the hospital scenario

The sensors connected to the gateways depend on the medical condition of the corresponding patient. Multimedia sensors are directly connected to the cluster heads. The multimedia sensors are deployed depending on the medical case of the patients. It is assumed that one multimedia sensor is connected to each cluster head.

8.2.2.3 *Attacker and observability of data sources*

The attacker is a global attacker who can analyse the traffic throughout the network. Probabilities can be estimated about the sender of a message, such as $p(s) = \frac{n_i}{M}$ where n_i is the number of messages reaching a given receiver from a given sender i (i.e. a gateway or a cluster head) and M is the total number of messages reaching the receiver (i.e. cluster head, or base station). The probabilities calculated will be used in the calculations of the entropy (to measure anonymity) and conditional entropy (to measure unlinkability).

8.2.2.4 Entropy calculation

Let us define the following random variables:

- G_c represents a gateway, in a subset $\{g_{1c}, \cdots, g_{Pc}\}$, which is connected to a given cluster head indexed by c,
- C_b represents a cluster head, in a subset $\{c_{1b}, \cdots, c_{Qb}\}$, which is connected to a given base station indexed by b,
- B represents a base station, in a set $\{b_1, \cdots, b_R\}$.

8.2.2.4.1 Entropy at a cluster head

The entropy calculated at each cluster head represents the uncertainty of the local adversary that might be listening to the traffic between a cluster head C_b and the base station B. Consequently, as the number of messages generated by each real gateway and the fake gateways change, the overall entropy at the cluster heads will change. For assessing the anonymity of the sender of a message, which reaches the c^{th} cluster head, a suitable probability measure is the probability $p(G_c = g_{ic})$ that the message is from a given gateway (i.e. patient in the scenario shown in Figure 8-1). Here, g_{ic} is the i^{th} gateway connected to the cluster head. This probability can be calculated as

$$p(g_{ic}) = \frac{n_{ic}}{M_c} \tag{55}$$

where n_{ic} is the total number of messages reaching the cluster head from the gateway g_{ic}, and $M_c = \sum_{i=1}^{N} n_{ic}$ is the total number of messages reaching the cluster head from all gateways connected to it. Hence, the entropy of the outcome of a random variable G_c observed at a cluster head can be computed as

$$H(G_c) = -\sum_{i=1}^{P} p(g_{ic}) \log_2 p(g_{ic}) \tag{56}$$

This represents the level of uncertainty about what gateway sent a message.

8.2.2.4.2 Entropy at a base station

As the adversary is a global one, (s)he would be able to also use information from analysing the traffic upstream of the cluster heads. Hence, his/her estimate of probabilities will include information about the data rates from the gateways (which are an indicator of the likely identities of the senders (i.e. gateway or equivalently patient as gateways are tightly coupled to one patient)). Hence, the probability calculations at the base station should pool the identity-related probabilities about the

sending gateways. The general formula of the entropy (Equation (8)) does not reflect the knowledge possessed by the attacker, because the formula is about the possible identities of the sending cluster heads only. The global attacker could be able to gather probabilities about what cluster head might have sent a message to the base station, but also what gateway might have sent a message to a cluster head. Consequently, to calculate the overall entropy at the base station, two alternatives for the entropy calculation can be used:

- Joint entropy: to estimate the level of uncertainty about both the sending cluster head and the sending gateway, for a message received at a base station.
- Conditional entropy: to estimate the level of uncertainty about the sending gateway, for a message received at a base station, if the attacker can determine the sending cluster head. Conditional entropy is used to represent the average information to link a message to the sending gateway.

(i) Joint entropy

To measure the uncertainty associated with both the random variables G_c and C_b (i.e. uncertainty about what gateway and what cluster head sent a message which reaches a base station), the joint Shannon entropy of the two variables can be computed as

$$H(G_c, C_b) = -\sum_{j=1}^{Q}\sum_{i=1}^{P} p(g_{ic}, c_{jb}) \log_2 p(g_{ic}, c_{jb}) \quad (57)$$

(ii) Conditional entropy

The probability that a message reaching a base station b is from a cluster head c_{jb} is $P(c_{jb}) = \frac{n_{jb}}{M_b}$ where n_{jb} is the total number of messages reaching base station b from a given cluster head, and $M_b = \sum_{i=1}^{Q} n_{jb}$ is the total number of messages reaching base station b from all cluster heads connected to the base station.

The level of unlinkability (observed at a base station b) that a message reaching a base station b was sent by (linked to) a gateway g_{ic}, given that the base station has received the message through a known cluster head c_{jb}, is defined as the entropy of G_c conditioned on C_b: The conditional entropy can be written as:

$$H(G_c|C_b) = -\sum_{j=1}^{Q} P(c_{jb}) \sum_{i=1}^{P} P(g_{ic}|c_{jb}) \log_2 P(g_{ic}|c_{jb}) \quad (58)$$

where $P(g_{ic}|c_{jb})$ is a conditional probability.

8.2.2.5 Procedure

Experiment 1 - Effect of the injection of fake traffic on privacy, for a network without multimedia sensors

Based on the literature survey conducted in Chapter 3, it was concluded that the deployment of fake traffic and fake sources of data could be used to implement different privacy services such as anonymity, location privacy and unlinkability. In this experiment, fake sources and fake volumes of traffic were deployed to attempt to prevent an adversary from determining whether messages generated by gateways are real or not (to achieve unlinkability to previous messages or senders of messages) and whether they belong to one or more patients (to achieve anonymity). In addition, fake sources and fake volumes of traffic are expected to hide the location of the senders and receivers of data as it is harder for an adversary to relate the message to a particular sender as the real messages are hidden within the fake messages from different fake generators thus achieving location privacy. Consequently, the analysis and the measurements of the effect of fake sources and volumes of traffic can be mapped to the effect on anonymity, unlinkability and location privacy.

The aim of this experiment is to study the effect of the number of sources of fake traffic and the volume of fake traffic on level of *privacy (anonymity, unlinkability and location privacy)* resulting from the introduction of the corresponding mechanisms. The implications of the results of this experiment will assist a designer of a WMSN for healthcare to study how well fake traffic would protect the privacy of data when multimedia sensors are turned off.

Fake sources were deployed at the same level as the gateways (i.e. fake sources act like fake gateways) and were connected to the cluster heads. It is assumed that all messages in the network are encrypted and constantly changing pseudonyms are used. This way, if an adversary captures any messages, he/she will not be able to comprehend the contents of the encrypted message. In addition, if an adversary captures two messages, he/she will not be able to tell whether these two messages belong to the same source or not due to the constantly changing pseudonyms. In this

experiment, only scalar sources are considered (i.e. no multimedia sources are taken into consideration).

The methodology followed in this experiment was as follows: The network depicted in Figure 8-1 was defined in the NS2 simulator. All gateways and cluster heads are connected wirelessly to the base station. Within the network, it is assumed that data has already arrived at the gateways and that all nodes (gateways, cluster heads, and base station) communicate over a Transmission Control Protocol (TCP), to ensure reliable communication between the different nodes. Reliability is an important issue due to the safety-critical nature of many healthcare applications, which often require correct and complete delivery of data from source to destination.

To study the effect of injection of fake messages into the network on the level of privacy, one fake source is added to each cluster head (as if a new fake gateway was added to each cluster head) and the new calculated entropy is recorded. In case of one fake source added, the ratio of the real to fake sources is G:1 where G denotes the number of gateways under the cluster head. Consequently, the communication channel between the gateways and their cluster heads is now shared between the real gateways and one fake source which means that more time is needed to send the same number of messages when no fake sources where present.

The level of uncertainty (from the perspective of an adversary) about the sender of the messages was calculated at each cluster head and the overall uncertainly was calculated at the base station. The calculation of the level of uncertainly is based on the entropy metric. In addition to uncertainty, anonymity set size was used to assess location privacy.

Experiment 2 - Effect of the injection of fake traffic on privacy, for a network which includes multimedia sensors

Similar to Experiment 1, in this experiment it is assumed that all messages in the network are encrypted and that constantly changing pseudonyms are used. The methodology followed in this experiment is as described below.

The network depicted in Figure 8-1 was defined in the NS2 simulator. All gateways and cluster heads are connected wirelessly to the base station. Within the network, it is assumed that data has already arrived at the gateways and that all nodes (gateways, cluster heads, and base station) communicate over a Transmission Control Protocol (TCP), to ensure reliable communication between the different

nodes. Multimedia nodes (representing video and audio sensors) are connected directly to the cluster heads instead of the gateways to save the energy of the gateways. The multimedia nodes wirelessly communicate with the cluster heads over User Datagram Protocol (UDP) channel.

To study the effect of the injection of fake messages into the network on the level of uncertainty, similar to Experiment 1, one fake source is added to each cluster head (as if a new fake gateway was added to each cluster head) and the new calculated entropy is recorded. The ratio of the traffic generated by the fake source is calculated as G:1 where G is the number of real gateways underneath the cluster head. In this experiment, the number of multimedia messages is considered as real messages and the number of fake messages is a ratio of both the real messages generated by the gateways and the multimedia sources. Each cluster head is then allowed to run long enough to receive the same number of real messages and multimedia messages that it received when no fake sources were present. In addition, each cluster head should also run until it has received the predetermined number of fake messages. However, in this experiment the number of fake messages is significantly higher compared to the number of fake messages in Experiment 1, due to the presence of multimedia traffic. Each time an extra fake source is added, the experiment was run 10 times. In this experiment, it was noticed that the total time to send all real and fake messages increased substantially as the number of fake sources increased. Consequently, Experiment 2 was conducted for 4 fake sources only. Each time a random seed was used by the NS2 simulator to generate random traffic, and the entropy at each cluster head was recorded together with the amount of time required for each cluster head to receive the same number of messages that it received when there was no fake traffic. In addition, jitter was calculated to assess the effect of the injection of fake traffic on the multimedia stream.

The same procedure is repeated. However, this time all traffic is simultaneous. All cluster heads are running at the same time instead of allowing each cluster head to run until it has received the predetermined number of messages. The total transmission completion time is recorded for each cluster head. In addition, the same procedure is repeated but with varying number of multimedia sensors under each cluster head. The total transmission completion time is recorded for each cluster head.

8.2.3 Results and discussion

The network depicted in Figure 8-1 was simulated on NS2. Figure 8-2 shows the simulation of the hospital network on NAM and the wireless ranges of the sensors. The experiments in this section used the default NS2 wireless ranges (250 meters). In Figure 8-2, cluster heads 1, 2, 3, 4, 5, 6, 7 and 8 are presented as CH_1, CH_2, CH_3, CH_4, CH_5, CH_6, CH_7 and CH_8 respectively.

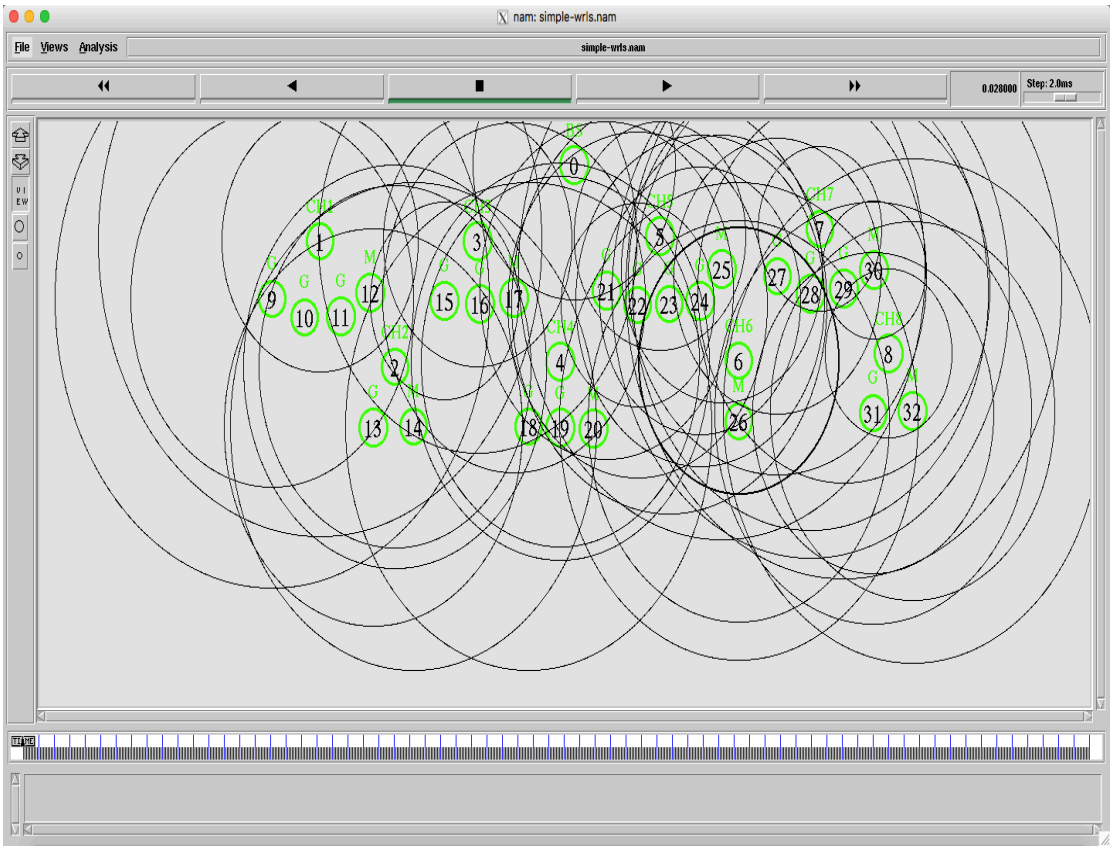


Figure 8-2 NAM simulation of the hospital scenario depicted in Figure 8-1. BS is Base Station, CH is Cluster Head, G is Gateway and M is Multimedia sensor

8.2.3.1 Experiment 1 - Effect of the injection of fake traffic on anonymity (network without multimedia sensors)

8.2.3.1.1 Experimental results with no fake sources

8.2.3.1.1.1 Entropy at the cluster heads

The simulation was run with no fake traffic introduced into the network. Table 8-1 presents the average of the total number of messages received at each cluster head and the entropy calculated at each cluster head using Equation (56). The value of entropy at cluster head 6 (CH_6) was directly placed as a zero because there is no

traffic generated at this cluster head due to the absence of any gateways. From the perspective of an adversary, this cluster head is not sending any messages and therefore it has entropy of 0.

Table 8-1 Recorded average number of messages and entropy for each cluster head in the case of no fake traffic

Cluster Head	Average number of messages received by the cluster head	Entropy at the cluster head
CH_1	8042	1.58
CH_2	8118	0
CH_3	8076	0.99
CH_4	7988	0.99
CH_5	8070	1.99
CH_6	0	0
CH_7	7676	1.57
CH_8	8308	0

8.2.3.1.1.2 Anonymity set size at the cluster heads

Table 8-2 Anonymity set size at each cluster head in case of no fake traffic

Cluster Head	Anonymity set size
CH_1	3
CH_2	1
CH_3	2
CH_4	2
CH_5	4
CH_6	0
CH_7	3
CH_8	1

Table 8-2 depicts the anonymity set size at each cluster head in case of no fake traffic. It is clear from the table that the anonymity set size equal 0 at CH_6 because there are no gateways under this cluster head.

8.2.3.1.1.3 Entropy at the base station

The values of the entropy depicted in Table 8-1 present the level of privacy at each cluster head. To assess the level of privacy at the base station, two alternatives for the entropy calculations were used (see Section 6.3.2.1): joint entropy and conditional entropy (to measure unlinkability). Using Equation (57), the joint entropy calculated at the base station equals 3.83. Using Equation (58), the conditional entropy calculated at the base station equals 1.01.

8.2.3.1.2 Experimental results with 1 fake source

8.2.3.1.2.1 Entropy at the cluster heads

The simulation was run with one fake gateway added under each cluster head. The fake gateways are responsible for generating fake traffic under each cluster head. The results recorded for one fake source are illustrated in Table 8-4, Table 8-5, Table 8-6 and Table 8-7. Table 8-4 and Table 8-5 give the entropy calculated at each cluster head for each one of the ten different times the experiment was run with random seeds. The transmission completion time, recorded next to each calculated entropy value, is the time when all messages (fake and real) have just been sent.

The “Start time” column indicates when the gateways under the cluster head started sending messages. For example, the gateways under cluster head 2 started sending messages at time (start time) 500 seconds. In the first trial, cluster head 2 received all messages at time 912. 95 seconds, in the second trial, cluster head 2 received all messages at time 796.69 seconds and so on. Table 8-6 and Table 8-7 show the analysis conducted at each cluster head. The analysis included the calculation of the mean, standard deviation, margin of error, upper bound and lower bound. The sample size used for the analysis is 10 and the confidence interval is at the 95% confidence level.

Table 8-8 summarizes the entropy, ratio of real to fake and delay (extra time required for the transmission completion) recorded at each cluster head. The “mean value” depicted in Table 8-8 refers to the mean of all the entropy values recorded in the ten times that the one fake source experiment was run. The results depicted in Table 8-8 show that the entropy has increased at each cluster head but at the expense of the extra time required to send the messages.

8.2.3.1.2.2 Anonymity set size at the cluster head

Table 8-3 shows the anonymity set size and the relative percentage increase in the anonymity set size after the introduction of one fake source (gateway) under each cluster head. It is clear from the table that the relative increase is higher when the original number of gateways is small. This implies that the location privacy is significantly improved when fake sources of traffic are added especially in cases when the original number of sources is low.

Table 8-3 Anonymity set size and relative percentage increase in anonymity set size after the introduction of 1 fake source

Cluster Head	Anonymity set size	Relative percentage increase in the anonymity set size
CH_1	4	33.3%
CH_2	2	100%
CH_3	3	50%
CH_4	3	50%
CH_5	5	25%
CH_6	1	100%
CH_7	4	33.3%
CH_8	2	100%

8.2.3.1.2.3 Entropy at the base station

After one fake source has been added under each cluster head, the results show that the entropy has increased at the level of each cluster head. At the base station, the total number of messages received (including the fake messages) is 92525 messages compared to 56278 messages when no fake traffic was deployed. To assess the level of anonymity at the base station, the joint entropy at the base station after introducing the fake traffic equals 4.29, which indicates that the joint entropy has increased by about 11%. To assess the level of unlinkability, the conditional entropy calculated at the base station is 1.47, which indicates that the conditional entropy has increased by about 31%.

Table 8-4 Entropy and total transmission completion time to send all messages for the cluster heads 1, 2, 3 and 4

		Start time (Seconds)		Start time (Seconds)		Start time (Seconds)		Start time (Seconds)
		0		500		1000		1500
Trial	**CH1 entropy**	**Transmission completion time (Seconds)**	**CH2 entropy**	**Transmission completion time (Seconds)**	**CH3 entropy**	**Transmission completion time (Seconds)**	**CH4 entropy**	**Transmission completion time (Seconds)**
1	1.99	141.85	1	913	1.58	1418.73	1.58	1803.8
2	1.99	136.57	1	796.7	1.58	1372.07	1.58	1812.4
3	1.98	135.06	1	726.8	1.58	1405.9	1.58	1859.2
4	1.99	132.86	1	806.5	1.58	1375.36	1.58	1876.2
5	1.99	139.56	1	867.6	1.58	1382.75	1.58	1820.1
6	1.99	136.36	1	748.9	1.58	1406.86	1.58	1883.5
7	1.99	190.95	1	858.7	1.58	1380.38	1.58	1830.6
8	1.98	139.3	1	926	1.58	1381.43	1.58	1860.3
9	1.99	150.57	1	864.1	1.58	1399.43	1.58	1885.3
10	1.99	132.77	1	941.1	1.58	1390.49	1.58	1893.6

Table 8-5 Entropy and total transmission completion time to send all messages for the cluster heads 5, 6, 7 and 8

		Start time **(Seconds)**		Start time **(Seconds)**		Start time **(Seconds)**		Start time **(Seconds)**
		2700		5500		6000		6500
Trial	**CH5 entropy**	**Transmission completion time (Seconds)**	**CH6 entropy**	**Transmission completion time (Seconds)**	**CH7 entropy**	**Transmission completion time (Seconds)**	**CH8 entropy**	**Transmission completion time (Seconds)**
1	2.32	3346.47	0	5631	1.99	6242.7	1	6795
2	2.32	3340.66	0	5597.4	1.99	6246.47	1	6799
3	2.32	2997.16	0	5621.5	1.99	6260.25	1	6814.1
4	2.32	3330.53	0	5633.5	1.99	6285.72	1	6790.3
5	2.31	2507.67	0	5630.2	1.999	6245.32	1	6774.3
6	2.32	3223.62	0	5627.3	1.999	6269.93	1	6839.2
7	2.32	3260.99	0	5610.2	1.99	6271.56	1	6788.8
8	2.32	2479.2	0	5599	1.99	6272.53	1	6822.7
9	2.29	2883.46	0	5582.1	1.99	6207.8	1	6816.2
10	2.3	2911.6	0	5613.7	1.99	6271.6	1	6823.4

Table 8-6 Analysis for the results recorded for 1 fake source for cluster heads 1, 2, 3 and 4

	CH1		CH2		CH3		CH4	
	Entropy	Transmission completion time (Seconds)	Entropy	Transmission completion time (Seconds)	Entropy	Transmission completion time (Seconds)	Entropy	Transmission completion time (Seconds)
Mean	1.99	143.6	1	844.9	1.58	1391.3	1.58	1852.5
Standard deviation	$6x10^{-3}$		0		$1.7\ x10^{-3}$		$1.7\ x10^{-4}$	
Margin of error	$3.7x10^{-3}$		0		$1.1\ x10^{-3}$		$1.1\ x10^{-4}$	
Upper bound	$199.72x10^{-2}$		1		$158.5\ x10^{-2}$		$158.48\ x10^{-2}$	
Lower bound	$198.96x10^{-2}$		1		$158.3\ x10^{-2}$		$158.46\ x10^{-2}$	

Table 8-7 Analysis recorded for the results for 1 fake sources for cluster heads 5, 6, 7 and 8

	CH5		CH6		CH7		CH8	
	Entropy	Transmission completion time (Seconds)	Entropy	Transmission completion time (Seconds)	Entropy	Transmission completion time (Seconds)	Entropy	Transmission completion time (Seconds)
Mean	2.31	3028.1	0	5614.6	1.99	6257.4	1	6806.3
Standard deviation	$1x10^{-2}$		0		$3.9x10^{-3}$		0	
Margin of error	$6.2x10^{-3}$		0		$2.4x10^{-3}$		0	
Upper bound	$232x10^{-2}$		0		$199.8x10^{-2}$		1	
Lower bound	$231x10^{-2}$		0		$199.4x10^{-2}$		1	

Table 8-8 Calculated mean entropy, real-to-fake ratio, extra delay (seconds) and percentage increase in entropy for each cluster head

Cluster Head	Mean entropy	Ratio of real sources to fake sources	Extra delay recorded (Seconds)	Relative Percentage increase in entropy
CH_1	1.99	3:1	43.59	20.49%
CH_2	1	1:1	244.92	100%
CH_3	1.58	2:1	291.34	36.86%
CH_4	1.58	2:1	252.49	36.9%
CH_5	2.31	4:1	228.14	13.6%
CH_6	0		0	0
CH_7	1.99	3:1	157.39	21.3%
CH_8	1	1:1	206.3	100%

Table 8-8 summarises the results of introducing one fake source under each cluster head. The table shows the calculated mean entropy at each cluster head, the ratio of the number of the real gateways to the number of fake gateways, the time (delay) needed to send all messages (including the fake messages) and the relative percentage increase in the entropy compared to the experiment which was run with no fake sources. Table 8-8 shows that the entropy at all cluster heads, except cluster head 6, has increased with a minimum of 13.5% (at cluster head 5) and a maximum of 100% (at cluster heads 2 and 8). Before introducing fake sources of traffic, the entropy at cluster heads 2 and 8 were 0. After the introduction of fake sources, the entropy has increased to 1 resulting in 100% relative percentage increase in entropy. The entropy at cluster head 6 remained equal to zero because even though a fake gateway has been added, all messages originate from one source causing all these messages to have a probability of 1 that belongs to this sources, thus it has an entropy of 0.

8.2.3.1.3 Experimental results with more than 1 fake source

8.2.3.1.3.1 Entropy at the cluster heads

The previous experiment is repeated for a varying number of fake sources (2, 3, 4, 5, 6, 7 and 8). The overall entropy, maximum entropy, normalized entropy and total transmission time for all messages are recorded at each cluster head. Table 8-10, Table 8-11, Table 8-12 and Table 8-13 give the results of these experiments. The maximum entropy is calculated according to Equation (9). For example, at cluster head 1, when there is no fake traffic, only three cluster heads are the possible senders of messages, accordingly, the maximum possible entropy equals $log_2(3)$. When one fake gateway is added, the senders of the messages increased by one and thus the maximum entropy is $log_2(4)$

8.2.3.1.3.2 Anonymity set size at the cluster heads

Table 8-9 depicts the relative percentage increase in the anonymity set size at each cluster head after the introduction of the different numbers of fake sources of traffic. It is clear from the table that the highest relative increase in location privacy is when one fake source (gateway) is introduced under each cluster head.

Table 8-9 Anonymity set size and percentage increase of anonymity set size at each cluster head

	Relative percentage increase in anonymity set size							
Number of fake sources	CH_1	CH_2	CH_3	CH_4	CH_5	CH_6	CH_7	CH_8
1	33.3%	100%	50%	50%	25%	100%	33.3%	100%
2	25%	50%	33.3%	33.3%	20%	100%	25%	50%
3	20%	33.3%	25%	25%	16.7%	50%	20%	33.3%
4	16.7%	25%	20%	20%	14.3%	33.3%	16.7%	25%
5	14.3%	20%	16.7%	16.7%	12.5%	25%	14.3%	20%
6	12.5%	16.7%	14.3%	14.3%	11.1%	20%	12.5%	16.7%
7	11.1%	14.3%	12.5%	12.5%	10%	16.7%	11.1%	14.3%
8	10%	12.5%	11.1%	11.1%	9.1%	14.3%	10%	2.5%

Table 8-10 Entropy, maximum entropy, normalized entropy and total transmission completion time (in seconds) for cluster heads 1 and 2

	CH1				CH2			
# of fake gateways	CH1 Entropy	Maximum Entropy	CH1 Norm Entropy	CH1 Transmission completion Time	CH2 Entropy	Maximum Entropy	CH2 Norm Entropy	CH2 Transmission completion Time
0	1.58	1.58	0.99	100	0	0	0	100
1	1.99	2	0.99	143.6	1	1	1	344.9
2	2.31	2.32	0.99	170.5	1.58	1.58	0.99	597.2
3	2.46	2.58	0.95	219.7	1.99	2	0.99	771.4
4	2.65	2.81	0.94	251.6	2.31	2.32	0.99	945
5	2.7	3	0.9	300.8	2.58	2.58	0.99	1173.6
6	2.85	3.17	0.9	362.3	2.75	2.8	0.98	1336
7	2.79	3.32	0.84	447.8	2.93	3	0.97	1659.5
8	2.95	3.46	0.85	448.1	3.05	3.17	0.96	1862.2

Table 8-11 Entropy, maximum entropy, normalized entropy and total transmission completion time (in seconds) for cluster heads 3 and 4

	CH3				CH4			
# of fake gateways	CH3 Entropy	Maximum Entropy	CH3 Norm Entropy	CH3 Transmission completion Time	CH4 Entropy	Maximum Entropy	CH4 Norm Entropy	CH4 Transmission completion Time
0	0.99	1	0.99	100	0.99	1	0.99	100
1	1.58	1.58	0.99	391.3	1.58	1.58	0.99	352.5
2	1.99	2	0.99	513.8	1.99	2	0.99	454.9
3	2.31	2.32	0.99	634.6	2.32	2.32	0.99	567.4
4	2.52	2.58	0.97	773.6	2.54	2.58	0.98	737.3
5	2.74	2.8	0.97	931.8	2.73	2.8	0.97	866.5
6	2.88	3	0.96	1126.5	2.8	3	0.94	1028.4
7	2.99	3.17	0.94	1212.6	2.95	3.17	0.93	1213.8
8	3.04	3.32	0.91	1311.8	3	3.32	0.9	1351.2

Table 8-12 Entropy, maximum entropy, normalized entropy and total transmission completion time (in seconds) for cluster heads 5 and 6

	CH5				CH6			
# of fake gateways	CH5 Entropy	Maximum Entropy	CH5 Norm Entropy	CH5 Transmission completion Time	CH6 Entropy	Maximum Entropy	CH6 Norm Entropy	CH6 Transmission completion Time
0	1.99	2	0.99	100	0	0	0	100
1	2.31	2.32	0.99	328.1	0	0	0	114.6
2	2.42	2.58	0.94	425.4	0.99	1	0.99	278.3
3	2.61	2.8	0.93	471.4	1.58	1.58	0.99	409.6
4	2.74	3	0.91	549.6	1.99	2	0.99	551.3
5	2.83	3.17	0.89	610.2	2.31	2.32	0.99	665.9
6	2.99	3.32	0.89	684.4	2.56	2.58	0.99	810.4
7	3.03	3.46	0.88	798.3	2.74	2.8	0.98	939.6
8	3.09	3.58	0.86	897.2	2.89	3	0.97	1070.2

Table 8-13 Entropy, maximum entropy, normalized entropy and total transmission completion time (in seconds) for cluster heads 7 and 8

	CH7				CH8			
# of fake gateways	CH7 Entropy	Maximum Entropy	CH7 Norm Entropy	CH7 Transmission completion Time	CH8 Entropy	Maximum Entropy	CH8 Norm Entropy	CH8 Transmission completion Time
0	1.57	1.58	0.99	100	0	0	0	100
1	1.99	2	0.99	257.4	1	1	1	306.3
2	2.3	2.32	0.99	313.6	1.58	1.58	0.99	509.8
3	2.51	2.58	0.97	397	1.99	2	0.99	683.4
4	2.71	2.8	0.97	466.2	2.32	2.32	0.99	908.3
5	2.83	3	0.94	513	2.49	2.58	0.96	1103.6
6	2.91	3.17	0.92	657.5	2.7	2.81	0.96	1267.5
7	2.98	3.32	0.9	745.1	2.82	3	0.94	1465.5
8	3.05	3.46	0.88	999.7	2.93	3.17	0.93	1762.3

A plot of the results of Table 8-10, Table 8-11, Table 8-12 and Table 8-13 are presented in Figure 8-3, Figure 8-4 and Figure 8-5. It is clear from the Figure 8-3 that the entropy at the cluster heads increases as the number of fake gateways increases. In addition, in Figure 8-5, the total transmission completion time required to send all messages increases as the number of fake gateways increases. This is due to the fake messages that have utilized the bandwidth and resulted in more time required to send all real messages. However, this is not the case for the normalized entropy. In Figure 8-4 the normalized entropy is at its highest for one fake source (except for cluster heads 2, 6 and 8) and then starts decreasing gradually. This is because the relative increase in the entropy (in the numerator of Equation (10)), as the number of fake sources increase, is less than the increase in the maximum entropy (in the denominator of Equation (10)).

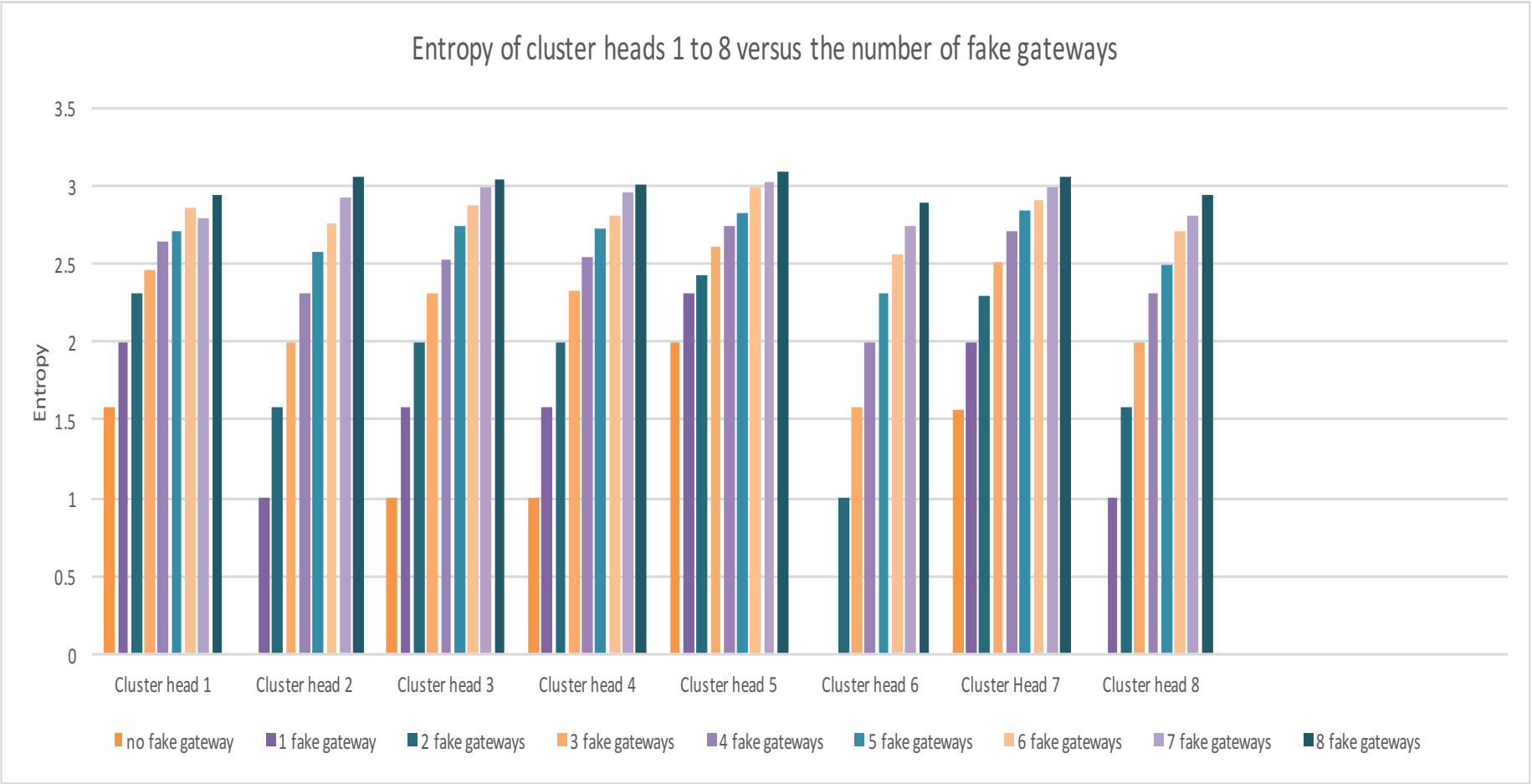


Figure 8-3 Entropy of each cluster head versus the number of fake gateways

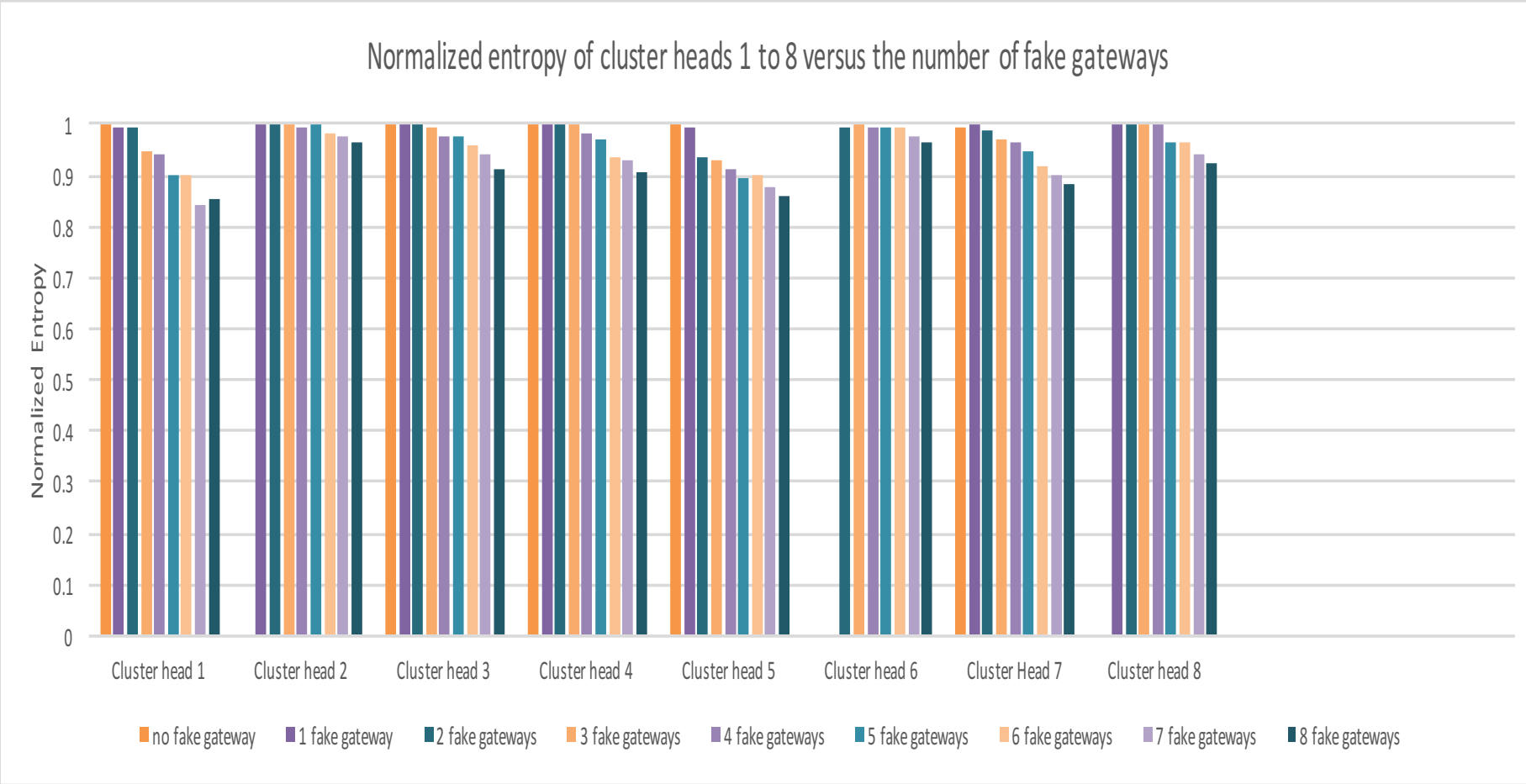


Figure 8-4 Normalized entropy of each cluster head versus the number of fake gateways

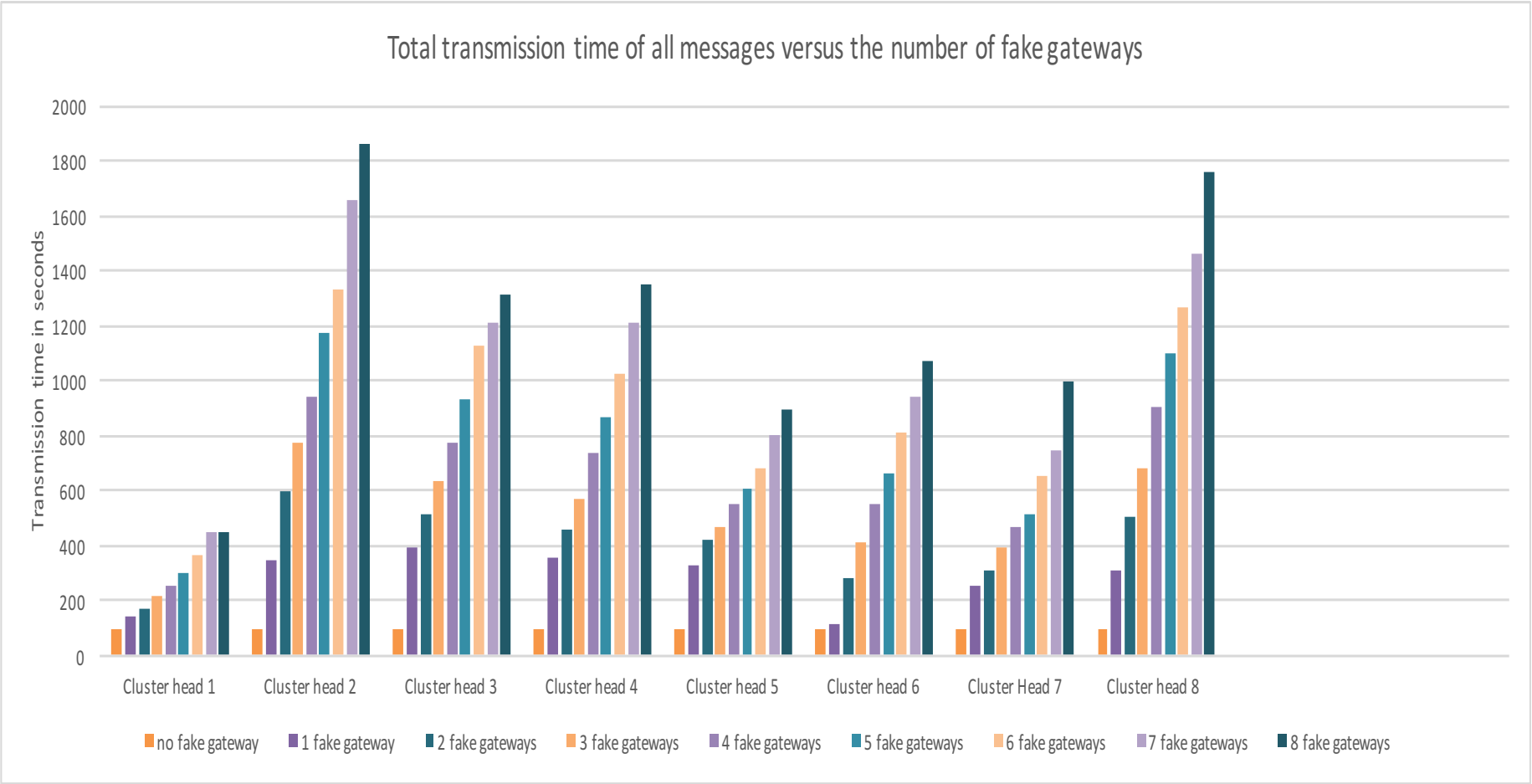


Figure 8-5 Total transmission completion times for each cluster head of all messages (real and fake) versus the number of fake gateways

8.2.3.1.3.3 Entropy at the base station

The maximum entropy, joint entropy and conditional entropy are given in Table 8-14. Plots of the results of Table 8-14 are depicted in Figure 8-6 and Figure 8-7. It is clear from the figures that the overall joint entropy and conditional entropy increase as the number of fake source increases. Table 8-9 depicts the anonymity set size and the percentage increase in the anonymity set size after the introduction of the fake sources of traffic.

Table 8-14 Maximum entropy, joint entropy and conditional entropy at the base station

Number of fake sources	Maximum Entropy	Joint Entropy	Conditional entropy
0	4	3.83	1.01
1	4.58	4.29	1.47
2	5	4.72	1.87
3	5.32	5.04	2.19
4	5.58	5.33	2.44
5	5.81	5.43	2.62
6	6	5.73	2.78
7	6.17	5.81	2.9
8	6.32	5.78	2.99

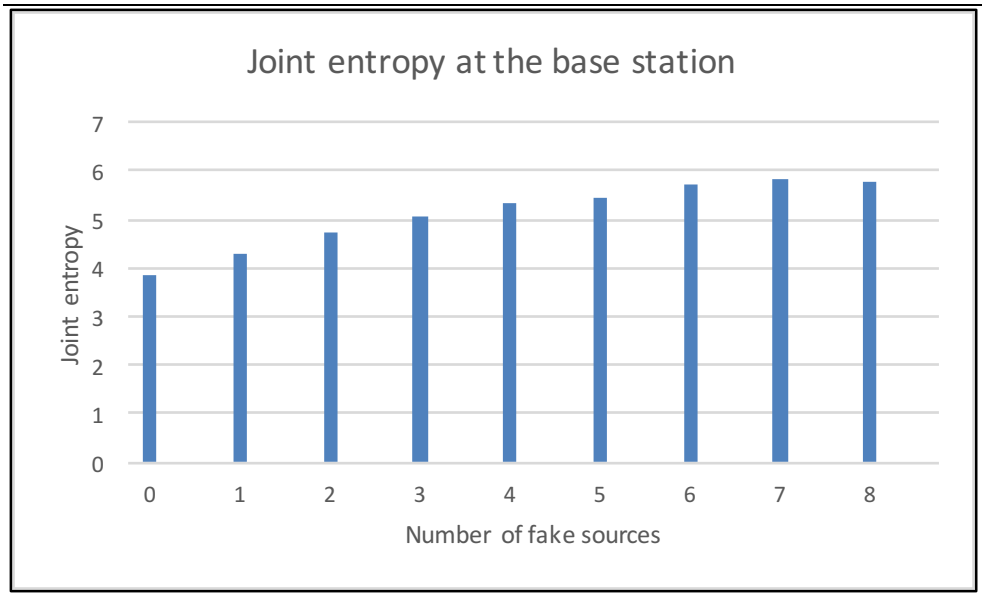


Figure 8-6 Joint entropy at the base station

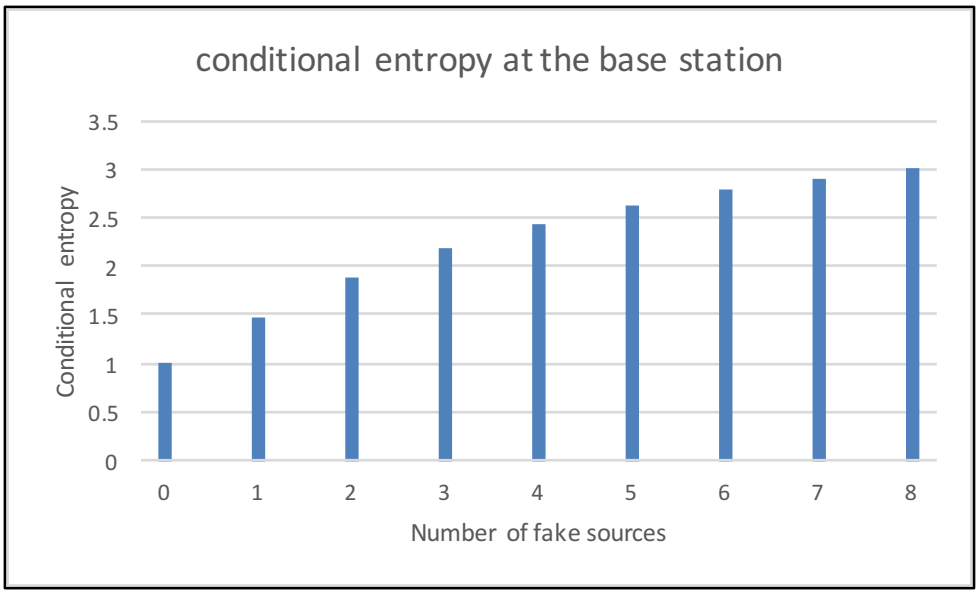


Figure 8-7 Conditional entropy at the base station

8.2.3.1.3.4 Relative entropy at the base station

According to Equation (13), relative entropy can be used to assess the amount of information an adversary can gain. Table 8-15 depicts the relative entropy versus the number of fake sources. It is clear from the table that as the number of fake sources increases, the value of the relative entropy increases, which indicates a higher level of privacy.

Table 8-15 Relative entropy versus the number of fake sources

Number of fake sources	Relative entropy
1	0.096
2	0.16
3	0.2
4	0.22
5	0.24
6	0.26
7	0.27
8	0.28

8.2.3.1.3.5 Information gain or loss at the base station

The information gain or loss metric represented by Equation (13) was used to assess the level of privacy at the station.

Table 8-16 Information gain or loss versus the number of fake sources

Number of fake sources	Information gain	Relative percentage increase in information gain or loss
1	0.096	-
2	0.16	62.36%
3	0.196	25.21%
4	0.22	14.13%
5	0.24	9.16%
6	0.26	6.46%
7	0.27	4.8%
8	0.28	3.73%

The results given in Table 8-16 show that the information gain or loss metric increases as the number of fake sources increases, which implies that the level of privacy increases. However, the relative increase in the information gain or loss metric reaches its peak at two fake sources and then starts decreasing again.

8.2.3.1.3.6 Anonymity set size at the base station

Table 8-17 shows the anonymity set size based on Equation (15) and the relative percentage increase in the anonymity set size after the introduction of each fake source of traffic under each cluster head. The results show that the maximum increase was after the addition of the first fake source under each cluster head.

Table 8-17 Anonymity set size and percentage of relative increase of anonymity set size at the base station

Number of fake sources	Anonymity set size	Relative percentage increase in anonymity set size
1	3.5	27.3%
2	4.375	25%
3	5.3	21.1%
4	6.25	17.9%
5	7.2	15.4%
6	8.2	13.5%
7	9.2	12%
8	10.15	10.7%

8.2.3.1.4 Discussion of the results of experiment 1

The aim of experiment 1 was to analyse the enhancement of the overall level of privacy after the introduction of fake sources (gateways) and varying volumes of fake traffic. Different metrics were used to assess privacy at the cluster heads and the base station: entropy was used to assess anonymity, conditional entropy was used to assess unlinkability and anonymity set size was used to assess location privacy.

It is clear from the previous tables and figures that as the number of fake sources increases and more fake messages are introduced into the network, the entropy at the cluster heads and at the base station increases but at the expense of the extra time (delay) required to send all messages. The results show that, on average, a trebling of the volume of traffic compared to the real traffic can (on average) double the level of privacy (anonymity / pseudonymity, unlinkability and location privacy) (as indicated by entropy estimated at the cluster heads), at the expense of an eight-fold increase of the transmission delay for real message packets. A trebling of the overall volume of traffic at the base station has caused a relative increase of almost 26% in the joint entropy and a relative increase of almost 58% in conditional entropy.

Although the introduction of fake traffic has enhanced the level of privacy (anonymity / pseudonymity, unlinkability and location privacy), the high increase of the transmission delay may not be acceptable in critical medical situations where real-time or quick medical help is required. Consequently, when applying these privacy measures, the trade-off between the level of privacy and the transmission delay should be studied depending on how critical the medical condition of the patient is and whether fast medical intervention is needed.

In addition, cluster heads such as cluster head 1 achieved the highest normalized entropy at 1 fake source whereas cluster head 2 achieved its highest normalized entropy at two fake sources. This implies that analysis should be conducted at each cluster head in real systems to determine the appropriate number of fake sources with regard to the number of real sources while considering the acceptable amount of delay. Moreover, different numbers of fake sources under each cluster head and a varying number of fake sources increase entropy (estimated as an indicator of privacy (anonymity, unlinkability and location privacy). Thus, they make it harder for an adversary to identify the real sources, link different messages to their source or discover the location of sources, if the adversary uses traffic analysis.

Traffic analysis attacks may assist an adversary in discovering the type of medical sensors deployed on a patient by studying the rate of packet generation of the sensors. Body sensors tend to have limited storage space, thus these sensors send their measurements in a periodic matter. The rates of these measurements can be related to specific body sensors (see Section 8.2.2.2), which may reveal information about the health problem of a patient (Buttyan & Holczer, 2012). For example, the data rate of an electro-encephalogram (EEG) sensor, used to measure the electric activity of a brain in medical cases such as epilepsy, is 43.2 kbps (Latré et al., 2011). Assuming the multimedia sensors are turned off, for a communication channel of 50 Mbps, an fake traffic generation up to a rate of approximately 1000 times the rate of the EEG sensor can be introduced into the channel. This amount of fake traffic will hide the actual data rate of the EEG sensor but at the expense of added delay and network channel utilization.

In addition, sensors which have low data rates allow the injection of higher fake traffic rates compared to sensors with higher data rates, for the same communication channel. For example, consider the case of two patients are wearing one sensor: a glucose sensor having a data rate of 1600 bps on the first patient, and an electromyography (EMG) sensor having a data rate of 320 kbps on the second

patient. For a given level of maximum network traffic (i.e. network capacity), about a 1000 times more fake traffic could be injected for the glucose sensor than for the EMG sensor; hence the anonymity of the first patient could be protected more than that of the second patient.

The setup of the sensors in this experiment (Experiment 1) can be used to study the effect of the number of fake traffic sources and the amount of fake traffic that is needed to achieve the required level of privacy versus an acceptable amount of delay. The sensors deployed in this experiment can be tuned to generate traffic at a rate of any commercial sensor and the appropriate amount of fake traffic can be introduced to hide the actual rate of the sensor and achieve the required level of privacy. In addition, the hierarchal setup of the gateways, cluster heads and base station complies with the basic hierarchal WSN architecture for healthcare (see Section 2.3), which makes the results presented in this experiment *not* specific to a special setup unique to this research work.

8.2.3.2 Experiment 2 - Varying numbers of fake sources in the presence of multimedia sources

8.2.3.2.1 Experimental results with no fake sources

Table 8-18 shows a summary of the results for running the experiment with no fake traffic. The table presents the average number of messages sent by the scalar sensors and the multimedia sensors at each cluster head. The table also depicts the calculated entropy at each cluster head.

Table 8-18 Recorded average number of messages received and entropy calculated at each cluster head with no fake source

Cluster Head	Total number of messages received at the cluster head		Entropy at the cluster head
	Scalar Sensors	Multimedia sensors	
CH_1	5244	60466	0.53
CH_2	5231	52464	0.44
CH_3	5002	63720	0.44
CH_4	5017	58386	0.48
CH_5	5667	66834	0.55
CH_6	0	65520	0
CH_7	5201	61802	0.52
CH_8	5014	58666	0.4

8.2.3.2.2 Experimental results with 1 or more fake sources

Table 8-19, Table 8-20, Table 8-21 and Table 8-22 present the results of experiment 2. The tables summarize the calculated entropy, maximum possible entropy, normalized entropy and total time to transmit all messages for the cluster heads. Figure 8-8, Figure 8-9 and Figure 8-10 depict plots of the results presented in the tables. Finally, to measure the effect of the fake traffic on the multimedia stream, the average jitter was plotted versus the number of fake sources; it is presented in equation (59). Average jitter is used to calculate the average delay between two consecutive multimedia frames (Inan et al., 2006). Average jitter is calculated for two consecutive frames *i* and *j* using the equation

$$jitter = \frac{(receivetime_j - sendtime_j) - (receivetime_i - sendtime_i)}{j - i}, \quad j > i \tag{59}$$

Although the overshoots of the average jitter plots may seem high for small numbers of fake gateways, the overall average jitter for no fake gateways is less than for one fake gateway which is less than for three gateways, and so on. The overshoots are due to the domination of the bandwidth by the fake gateways and scalar sensors.

Table 8-19 Entropy, maximum entropy, normalized entropy and total transmission time at cluster heads 1 and 2

	CH1				CH2			
# of fake gateways	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data
0	0.53	2	0.26	1960.7	0.44	1	0.44	1728.3
1	1.21	2.32	0.52	2030.4	1.22	1.58	0.77	1718.7
2	1.69	2.58	0.65	1986.8	1.73	2	0.87	1730.5
3	2.06	2.81	0.73	1974.2	2.11	2.32	0.91	1705.3
4	2.35	3	0.78	2008.4	2.41	2.58	0.93	1712.7

Table 8-20 Entropy, maximum entropy, normalized entropy and total transmission time at cluster heads 3 and 4

	CH3				CH4			
# of fake gateways	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data
0	0.44	1.58	0.28	2014.9	0.48	1.58	0.3	1912.2
1	1.22	2	0.6	2085.2	1.24	2	0.62	1918.5
2	1.72	2.32	0.74	2087.8	1.74	2.32	0.75	1903.4
3	2.1	2.58	0.81	2121.6	2.11	2.58	0.82	1902.7
4	2.4	2.8	0.86	2066.5	2.41	2.81	0.86	1908.11

Table 8-21 Entropy, maximum entropy, normalized entropy and total transmission time at cluster heads 5 and 6

	CH5				CH6			
# of fake gateways	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data
0	0.55	2.32	0.24	2155.3	0	1	0	2163.8
1	1.16	2.58	0.45	2185	1	1.58	0.63	2151.4
2	1.62	2.81	0.58	2166.5	1.58	2	0.79	2142.4
3	1.97	3	0.66	2188.7	1.99	2.32	0.86	2154
4	2.27	3.17	0.72	2186.2	2.32	2.58	0.89	2104.2

Table 8-22 Entropy, maximum entropy, normalized entropy and total transmission time at cluster heads 7 and 8

	CH7				CH8			
# of fake gateways	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data	Entropy	Maximum Entropy	Normalized Entropy	Total Time to Transmit all data
0	0.52	2	0.26	2024	0.4	1	0.4	1890.8
1	1.19	2.32	0.52	2040.4	1.19	1.58	0.76	1929.1
2	1.68	2.58	0.65	2038.6	1.72	2	0.86	1912.7
3	2.05	2.81	0.73	2029.8	2.1	2.32	0.9	1925.2
4	2.35	3	0.78	2014.7	2.4	2.58	0.93	1935.8

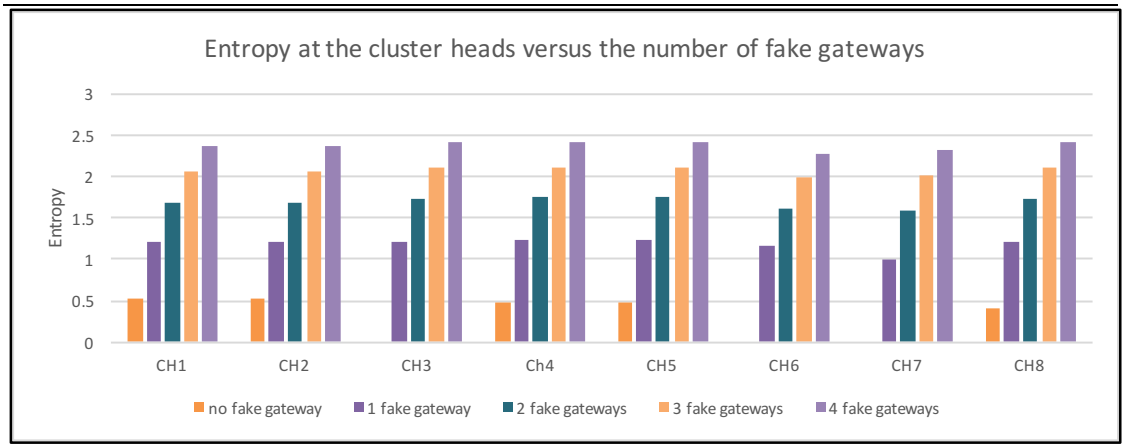


Figure 8-8 Entropy at the cluster heads versus the number of fake gateways

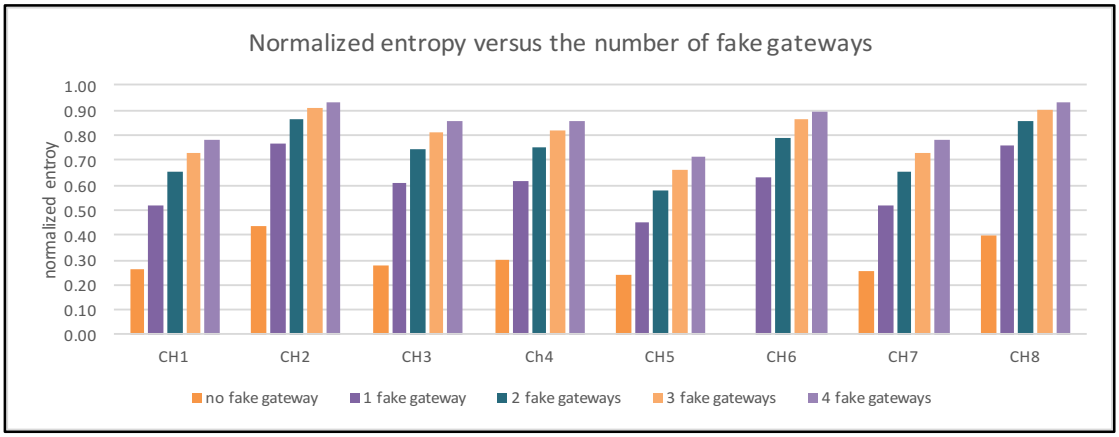


Figure 8-9 Normalized entropy versus the number of fake gateways

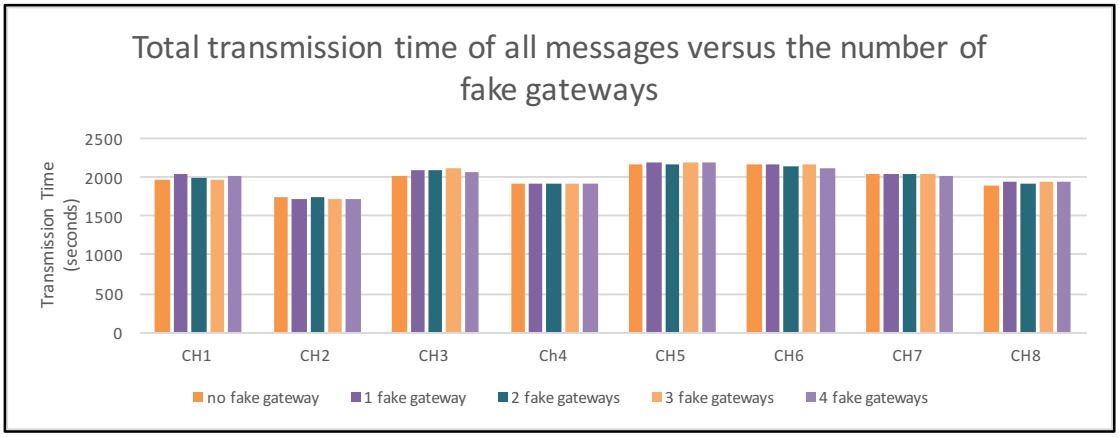


Figure 8-10 Total transmission times of all messages versus the number of fake gateways

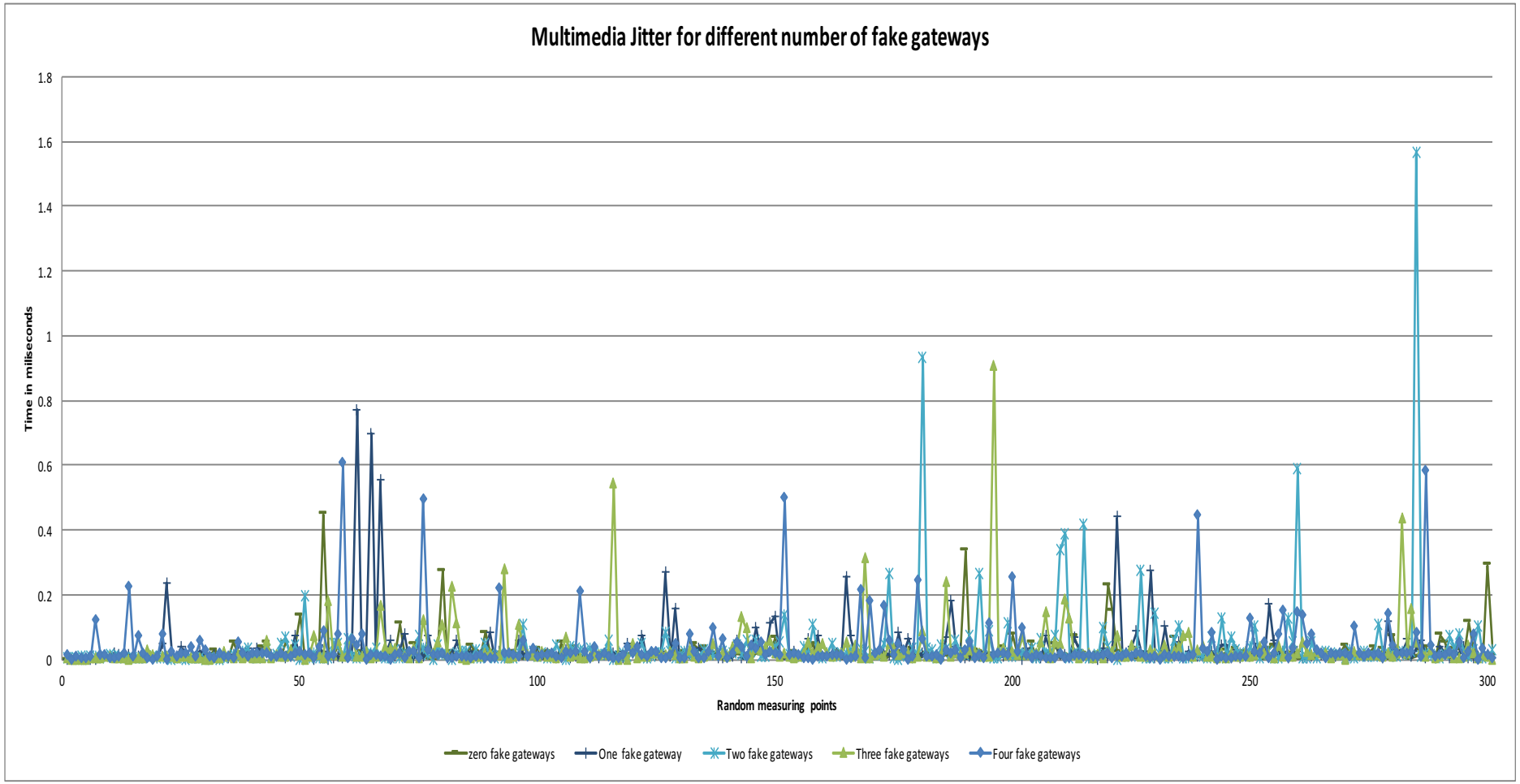


Figure 8-11 Average jitter versus the number of fake gateways

8.2.3.2.2.1 Entropy at the base station

After one fake source has been added under each cluster head, the results show that the entropy increases at the level of each cluster head. At the base station, the total number of messages received (including the fake messages) is 839557 compared to only 524234 when no fake traffic was deployed. To assess the level of privacy at the base station, the joint entropy at the base station after introducing the fake traffic equals 4.16, which indicates that the joint entropy has increased by about 18%. The conditional entropy calculated at the base station is 0.49 (see Table 8-23), which indicates that the conditional entropy has increased by about 93%.

Table 8-23 Maximum entropy, joint entropy and conditional entropy for the network with multimedia sources

Number of fake sources	Maximum Entropy	Joint Entropy	Conditional entropy
0	4.09	3.41	0.033
1	4.64	4.16	0.49
2	5.04	4.64	0.97
3	5.36	5	1.38
4	5.61	5.3	1.72

8.2.3.2.2.2 Information gain or loss at the base station

The information gain or loss metric represented by Equation (13) was used to assess the level of privacy at the station. Table 8-24 depicts the information gain or loss and the relative increase in the information gain or loss at the base station. The table shows that the information gain or loss increases as the number of fake sources increases. The highest relative increase is achieved at two fake sources.

Table 8-24 Information gain or loss and relative increase at the base station

Number of fake sources	Information gain	Relative increase in information gain or loss
1	0.066	-
2	0.11	67.68
3	0.14	27.28
4	0.16	15.29

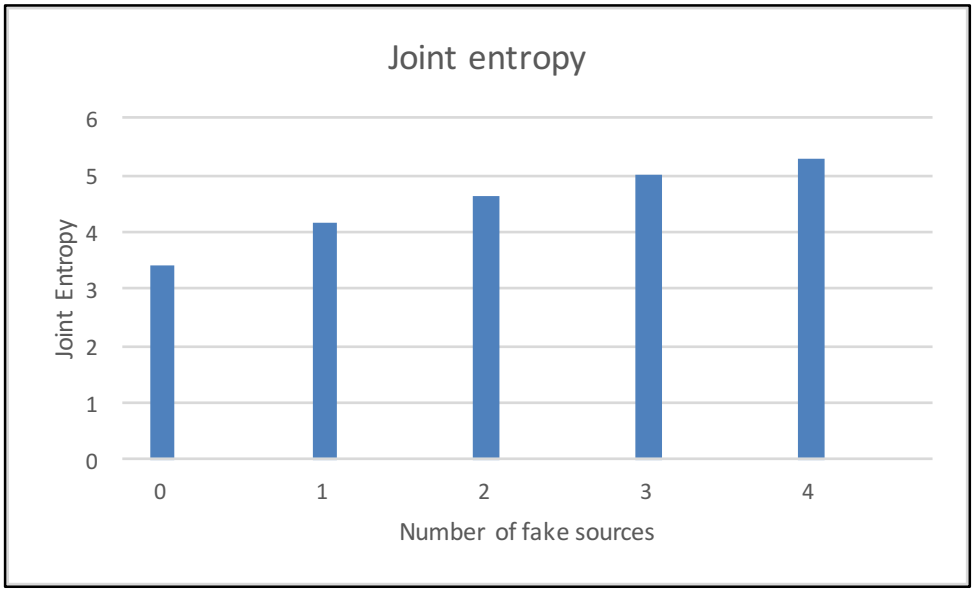


Figure 8-12 Joint entropy at the base station

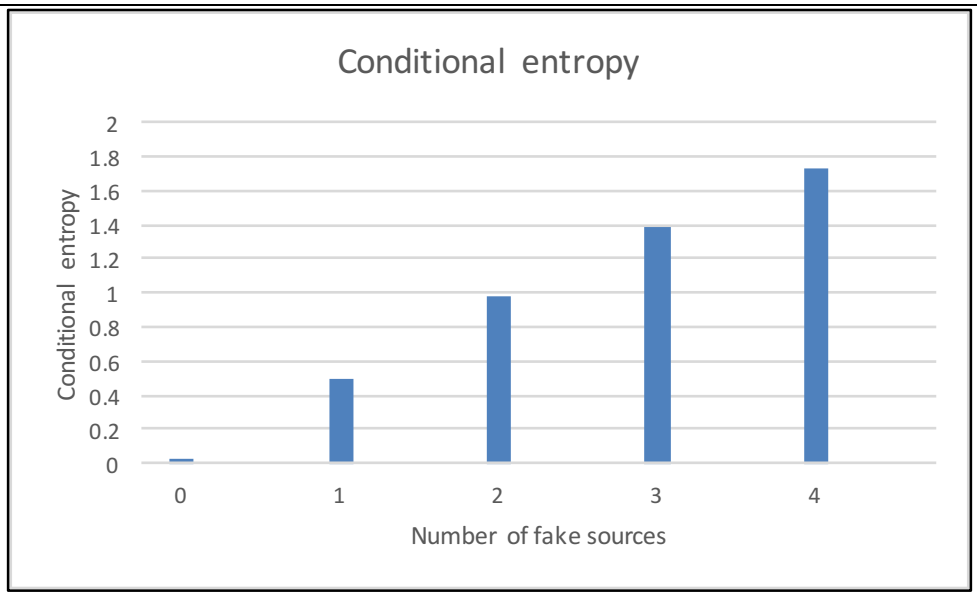


Figure 8-13 Conditional entropy at the base station

8.2.3.2.3 Experimental results with simultaneous traffic

In this experiment, all gateways, multimedia sensors and cluster heads were run simultaneously at the same time. The network was experimented with zero, one and two fake sources of packets. For each number of fake sources, the network is run 5 times using random seeds. Entropy and total transmission completion time for each cluster head were recorded as depicted in Table 8-25, Table 8-26, Table 8-27, Table 8-28, Table 8-28 and Table 8-29. The numbers presented in the tables were rounded to 2 decimal places. However, the original calculations were based on 9 decimal places. Comparing these tables with Table 8-19, Table 8-20, Table 8-21 and Table 8-22, it can be deduced that the total transmission completion time had changed. The absolute percentage relative difference between the previous experiment (non-simultaneous traffic) and this experiment (simultaneous traffic) has been calculated and recorded in Table 8-25 to Table 8-29.

According to Table 8-25 and Table 8-26, in the case of zero fake sources; the least percentage relative difference in transmission completion time was 0.28% at cluster head CH7 and the highest percentage difference was 3.32% at cluster head CH3. Based on Table 8-27 and Table 8-28, in the case of one fake sources; the least percentage relative difference in transmission completion time was 0.03% at cluster head CH2 and the highest percentage difference was 1.63% at cluster head CH1. From Table 8-29 and Table 8-30, in the case of two fake sources; the least percentage relative difference in transmission completion time was 4.45×10^{-3}% at cluster head CH3 and the highest

percentage difference was 2.48% at cluster head CH5. For each trial, the results show different transmission completion times at the same cluster heads. This difference is due to the random seed, which instructs NS2 to generate random number of packets at each source. The random packets arrive at the base station at different times, which justifies the different transmission completion times. To sum up, there was no significant difference between transmission completion times (maximum difference was 3.32%) of simultaneous traffic compared to non-simultaneous traffic. However, this was different for average jitter.

Figure 8-14 depicts a plot of average jitter for simultaneous traffic in case of zero, one and two fake sources. Comparing Figure 8-14 with Figure 8-11, it can be noticed that the average jitter has significantly increased in the case of simultaneous traffic. Moreover, as more fake sources are introduced, the average jitter increases. This is due to the increased traffic and increased competition between the cluster heads for the bandwidth. As more cluster heads are simultaneously utilizing the bandwidth and more sources of network packets are introduced to the network, average jitter increases.

Table 8-25 Entropy, total transmission completion time and analysis for the results of zero fake sources for cluster heads 1, 2, 3 and 4

Trial	CH1 entropy	Transmission completion time (Seconds)	CH2 entropy	Transmission completion time (Seconds)	CH3 entropy	Transmission completion time (Seconds)	CH4 entropy	Transmission completion time (Seconds)
1	0.53	1931.86	0.44	1768.09	0.45	2039.27	0.47	1905.23
2	0.53	2017.54	0.44	1709.29	0.45	2044.07	0.48	1936.44
3	0.53	1994.94	0.44	1712.75	0.45	2127.31	0.48	1950.82
4	0.53	1942.30	0.44	1720.74	0.45	2080.85	0.48	1937.54
5	0.53	2021.56	0.44	1759.64	0.45	2117.29	0.48	1916.91
Mean	0.53	1981.64	0.44	1734.10	0.45	2081.76	0.48	1929.39
Standard deviation	5.22×10^{-4}		0		9.62×10^{-4}		1.65×10^{-3}	
Margin of error	4.58×10^{-4}		0		8.44×10^{-4}		1.44×10^{-3}	
Upper bound	0.528		0.44		0.449		0.478	
Lower bound	0.527		0.44		0.448		0.476	
Previous mean results	0.53	1960.68	0.44	1728.26	0.44	2014.94	0.48	1912.25
Percentage relative difference	0.06%	1.07%	0%	0.34%	2.58%	3.32%	0.25%	0.90

Table 8-26 Entropy, total transmission completion time and analysis for the results of zero fake sources for cluster heads 5, 6, 7 and 8

Trial	CH5 entropy	Transmission completion time (Seconds)	CH6 entropy	Transmission completion time (Seconds)	CH7 entropy	Transmission completion time (Seconds)	CH8 entropy	Transmission completion time (Seconds)
1	0.54	2177.02	0	2140.48	0.52	2019.11	0.40	1958.24
2	0.55	2205.85	0	2119.90	0.52	2041.39	0.40	1960.84
3	0.55	2188.58	0	2152.58	0.52	1984.28	0.40	1984.28
4	0.53	2155.59	0	2116.66	0.52	2071.94	0.40	1894.09
5	0.55	2177.00	0	2126.35	0.51	2031.47	0.40	1935.20
Mean	0.54	2180.81	0	2131.19	0.52	2029.64	0.40	1946.53
Standard deviation	9.31×10^{-3}		0		2.2×10^{-3}		0	
Margin of error	8.16×10^{-3}		0		1.93×10^{-3}		0	
Upper bound	0.55		0		0.517		0.3978	
Lower bound	0.53		0		0.514		0.398	
Previous mean results	0.55	2155.31	0	2163.81	0.52	2024.00	0.40	1890.83
Percentage relative difference	1.27%	1.18%	0%	1.51%	0.24%	0.28%	0%	2.95%

Table 8-27 Entropy, total transmission completion time and analysis for the results of one fake source for cluster heads 1,2,3 and 4

Trial	CH1 entropy	Transmission completion time (Seconds)	CH2 entropy	Transmission completion time (Seconds)	CH3 entropy	Transmission completion time (Seconds)	CH4 entropy	Transmission completion time (Seconds)
1	1.21	2007.85	1.22	1755.35	1.22	2126.76	1.24	1892.88
2	1.21	2004.56	1.22	1666.31	1.22	2038.11	1.24	1865.99
3	1.20	1948.42	1.22	1720.47	1.22	2049.60	1.24	1886.08
4	1.20	2016.82	1.22	1736.65	1.22	2094.51	1.24	1913.57
5	1.20	2009.01	1.22	1711.81	1.22	2061.79	1.24	1946.99
Mean	1.21	1997.33	1.22	1718.12	1.22	2074.15	1.24	1901.10
Standard deviation	1.1×10^{-3}		0		1.99×10^{-4}		2.58×10^{-4}	
Margin of error	9.6×10^{-4}		0		1.74×10^{-4}		2.26×10^{-4}	
Upper bound	1.21		1.22		1.218		1.237	
Lower bound	1.20		1.22		1.217		1.24	
Previous mean results	1.21	2030.35	1.22	1718.69	1.22	2085.19	1.24	1918.53
Percentage relative difference	0.15%	1.63%	0%	0.03%	0.01%	0.53%	0.01%	0.91%

Table 8-28 Entropy, total transmission completion time and analysis for the results of one fake sources for cluster heads 5, 6, 7 and 8

Trial	CH5 entropy	Transmission completion time (Seconds)	CH6 entropy	Transmission completion time (Seconds)	CH7 entropy	Transmission completion time (Seconds)	CH8 entropy	Transmission completion time (Seconds)
1	1.16	2205.84	1	2083.46	1.20	2036.23	1.20	1954.83
2	1.16	2128.43	1	2166.75	1.20	2038.90	1.20	1916.88
3	1.16	2165.14	1	2194.91	1.20	1983.06	1.20	1910.19
4	1.16	2226.64	1	2200.35	1.19	2057.73	1.20	1951.08
5	1.16	2253.69	1	2115.91	1.20	2007.97	1.20	1899.53
Mean	1.16	2195.95	1	2152.27	1.20	2024.78	1.20	1926.50
Standard deviation	5.51×10^{-4}		0		1.82×10^{-3}		0	
Margin of error	4.83×10^{-4}		0		1.59×10^{-3}		0	
Upper bound	1.163		1		1.197		1.198	
Lower bound	1.162		1		1.194		1.198	
Previous mean results	1.16	2184.99	1	2151.44	1.20	2040.37	1.20	1929.11
Percentage relative difference	0.04%	0.50%	0%	0.04%	0.24%	0.76%	0%	0.14%

Table 8-29 Entropy, total transmission completion time and analysis for the results of two fake source for cluster heads 1,2,3 and 4

Trial	CH1 entropy	Transmission completion time (Seconds)	CH2 entropy	Transmission completion time (Seconds)	CH3 entropy	Transmission completion time (Seconds)	CH4 entropy	Transmission completion time (Seconds)
1	1.69	1962.72	1.73	1766.80	1.72	2134.35	1.74	1897.90
2	1.68	2033.78	1.73	1712.01	1.72	2083.40	1.74	1962.24
3	1.68	2003.43	1.73	1673.30	1.72	2036.32	1.74	1926.93
4	1.69	2020.89	1.73	1672.91	1.72	2082.72	1.74	1935.01
5	1.68	1994.75	1.73	1728.74	1.72	2101.64	1.74	1932.26
Mean	1.68	2003.11	1.73	1710.75	1.72	2087.69	1.74	1930.87
Standard deviation	1.2×10^{-3}		1.11×10^{-4}		2.01×10^{-4}		1.82×10^{-4}	
Margin of error	1.1×10^{-3}		9.73×10^{-5}		1.76×10^{-4}		1.6×10^{-4}	
Upper bound	1.69		1.73		1.724		1.739	
Lower bound	1.68		1.731		1.723		1.738	
Previous mean results	1.69	1986.81	1.73	1730.47	1.72	2087.77	1.74	1903.44
Percentage relative difference	0.12%	0.82%	0.02%	1.14%	0.01%	4.45×10^{-3}%	0.01%	1.44%

Table 8-30 Entropy, total transmission completion time and analysis for the results of two fake source for cluster heads 5, 6, 7 and 8

Trial	CH5 entropy	Transmission completion time (Seconds)	CH6 entropy	Transmission completion time (Seconds)	CH7 entropy	Transmission completion time (Seconds)	CH8 entropy	Transmission completion time (Seconds)
1	1.62	2220.98	1.58	2132.70	1.68	2055.93	1.72	1960.33
2	1.61	2193.26	1.58	2171.46	1.67	2039.18	1.72	1864.97
3	1.62	2257.89	1.58	2113.97	1.68	2036.34	1.72	1894.97
4	1.61	2200.10	1.58	2191.56	1.67	1952.01	1.72	1906.55
5	1.62	2229.39	1.58	2182.05	1.67	2052.55	1.72	1928.83
Mean	1.62	2220.33	1.58	2158.35	1.68	2027.20	1.72	1911.13
Standard deviation	2.33×10^{-3}		2.49×10^{-4}		2.65×10^{-3}		1.42×10^{-4}	
Margin of error	2.05×10^{-3}		2.2×10^{-4}		2.32×10^{-3}		1.24×10^{-4}	
Upper bound	1.62		1.585		1.678		1.718	
Lower bound	1.61		1.584		1.673		1.717	
Previous mean results	1.62	2166.54	1.58	2142.41	1.68	2038.56	1.72	1912.72
Percentage relative difference	0.18%	2.48%	0.01%	0.74%	0.29%	0.56%	0%	0.08%

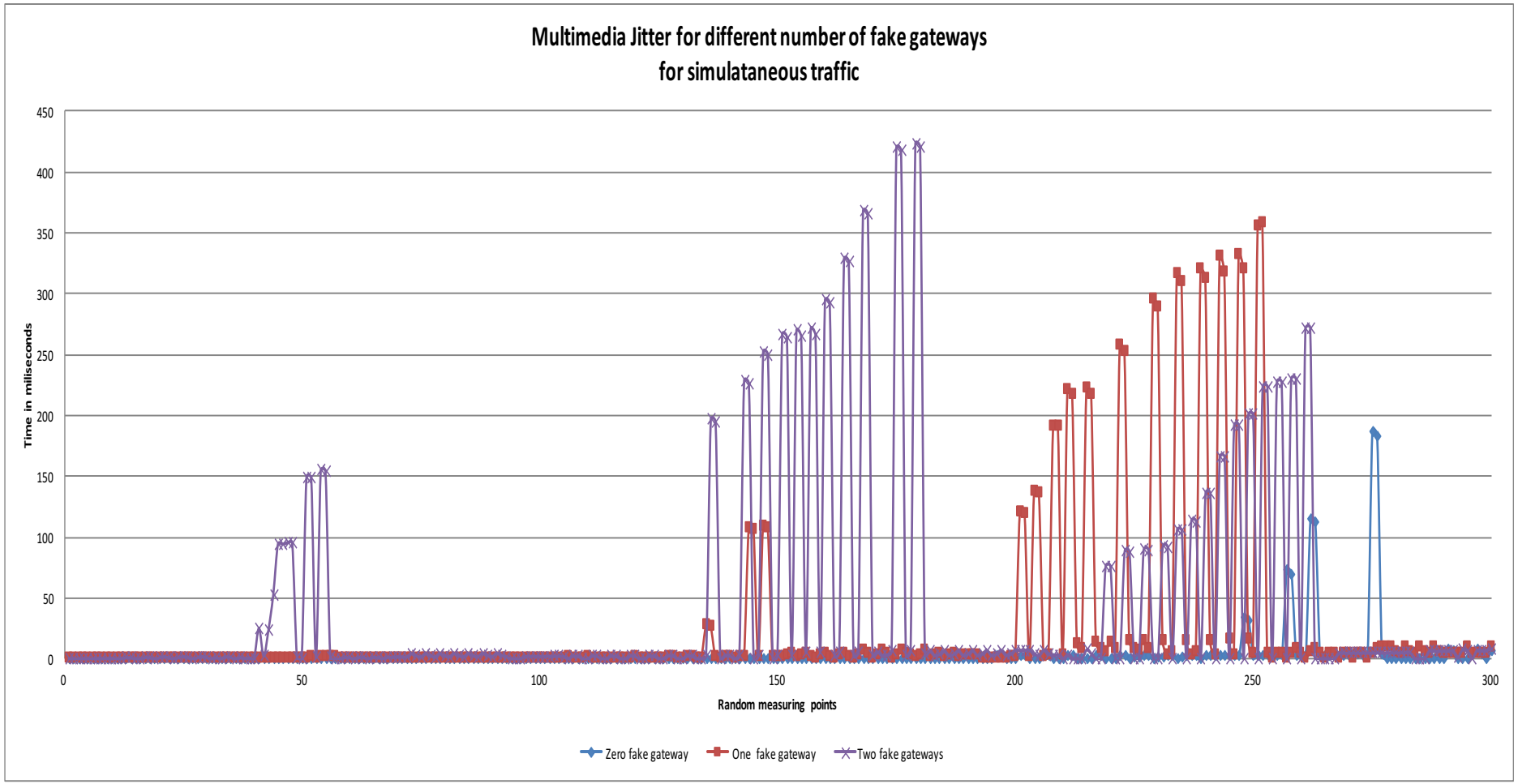


Figure 8-14 Average jitter for simultaneous traffic

8.2.3.2.4 Experimental results with varying number of multimedia sensors

In this experiment, varying numbers of multimedia sensors were randomly deployed under each cluster head as depicted in Table 8-31. The aim of this experiment it to measure the entropy under each cluster head and determine the total transmission completion time for varying number of multimedia sensors. With varying number of multimedia sensors, the longest transmission completion time was 2198.084 seconds (refer to cluster head CH_6 in Table 8-30) compared to 2180.81 seconds in case of zero fake sources (see cluster head CH_5 in Table 8-25 and Table 8-26), and compared to 2195.95 seconds in case of one fake source (see cluster head CH_5 in Table 8-27 and Table 8-28) and compared to 2220.33 seconds in case of two fake sources (see cluster head CH_5 in Table 8-29 and Table 8-30). The longest transmission completion time in Table 8-30 is at cluster head CH_6 because it has the highest number of multimedia sensors connected to it. In Table 8-25 to Table 8-30, the longest transmission completion time is at cluster head CH_5 because it has the highest number of gateways connected to it.

For varying number of multimedia sensors, the joint entropy at the base station is 2.55 and the conditional entropy is 0.061. The joint entropy is less than the joint entropy of zero fake sources in Table 8-23. The conditional entropy is less than the conditional entropy of one fake source in Table 8-23. This is due to the increased number of packets received at the cluster heads and base station due to a higher number of multimedia sensors deployed in the network. The higher number of packets at the cluster heads and base station causes the denominator of equations (57) and (58) to increase thus lower joint and conditional entropies.

Table 8-31 Number of multimedia sensors, entropy and transmission completion time for cluster heads 1 to 8

Cluster Head CH_1			Cluster Head CH_2			Cluster Head CH_3		
# of multimedia sensors	Entropy	Transmission completion time (Seconds)	# of multimedia sensors	Entropy	Transmission completion time (Seconds)	# of multimedia sensors	Entropy	Transmission completion time (Seconds)
2	1.193	1996.304	3	1.730	1709.996	1	0.449	2126.626
Cluster Head CH_4			**Cluster Head CH_5**			**Cluster Head CH_6**		
# of multimedia sensors	Entropy	Transmission completion time (Seconds)	# of multimedia sensors	Entropy	Transmission completion time (Seconds)	# of multimedia sensors	Entropy	Transmission completion time (Seconds)
2	1.237	1929.935	1	0.552	2174.393	4	2	2198.084
Cluster Head CH_7			**Cluster Head CH_8**					
# of multimedia sensors	Entropy	Transmission completion time (Seconds)	# of multimedia sensors	Entropy	Transmission completion time (Seconds)			
1	0.514	2035.629	3	1.718	1911.611			

8.2.3.2.5 Discussion of the results of Experiment 2

The aim of experiment 2 was to analyse the enhancement of the overall level of privacy after the introduction of fake sources (gateways) and varying volumes of fake traffic, in the presence of multimedia traffic. It is clear from the previous tables and figures that as the number of fake sources increases and more fake messages are introduced into the network, the entropy at the cluster heads and at the base station increases but at the expense of the extra time (delay) required to send all messages and at the expense of extra multimedia jitter.

At the level of the cluster heads, the introduction of one fake source under each cluster head has increased the entropy (level of privacy (anonymity / pseudonymity, unlinkability and location privacy)) by an average of 75% among all cluster heads. Although the delay after the introduction of the fake traffic has only increased by an average of 1.5 seconds, the average jitter has increased by 21%. Compared to experiment 1, the introduction of fake traffic has not added a significant amount of delay to send all messages (fake and real) because the number of messages is already huge due to the multimedia sensors. However, the fake traffic has added a considerable amount of jitter of 21%.

At the level of the base station, the introduction of fake traffic has increased the level of privacy (anonymity / pseudonymity, unlinkability and location privacy). The most significant increase in both the joint entropy and conditional entropy is due to the addition of one fake source (caused an average of 18% increase in joint entropy and an average of 93% increase in conditional entropy), compared to the addition of more fake sources which did not significantly increase the entropy (level of privacy (anonymity / pseudonymity, unlinkability and location privacy)).

Although the introduction of fake traffic has enhanced the level of privacy (anonymity / pseudonymity, unlinkability and location privacy), more fake traffic has increased the multimedia jitter. However, fake traffic can be used to conceal the actual data rates of the medical sensors to hide their type; which can be tied to a specific illness. In addition, fake traffic can be deployed to conceal the data rate of the multimedia feed that can be tied to a specific camera or microphone model; which can be tied to a specific apartment or a certain buyer, thus revealing the real identity and/or the source location of the data. The continuous generation of fake traffic can even the overall traffic generated at each source, and maintains it at a specific level. With the increase of data feed from the sensors due to the illness of a patient, fake traffic can be lowered and the real traffic can be increased.

This way the overall traffic will be maintained to almost the same level and an adversary will not be able to detect the illness of the patient.

In medical cases when patients are monitored for the sake of obtaining data (scalar and multimedia) for research and analysis, multimedia feed can be analysed to assess the level of recovery. This would apply, for example, in the case of an elderly person who had undergone orthopaedic surgery, and remote monitoring of his recovery is required. Cameras can be installed in his/her apartment and the gait can be analysed towards ensuring the correct recovery from the surgery. In critical cases when quick analyses are required (for example, a patient falls down or suffers from sudden medical problems), the trade-off between the level of privacy and the fake traffic must be carefully assessed to obtain an acceptable level of jitter.

Similar to Experiment 1, the setup of the sensors in this experiment (Experiment 2) can be used to study the effect of the number of fake traffic sources and the amount of fake traffic that is needed to achieve the required level of privacy versus an acceptable amount of delay and jitter. The results and findings of both Experiment 1 and Experiment 2 can be expanded to simulate different health facilities (of different sizes and capacities) allowing the analysis of privacy, transmission delay and jitter.

8.2.3.3 Comparing the results of Experiment 1 and Experiment 2

Before the introduction of fake traffic: At the cluster head level, (based on Table 8-1 and Table 8-18) it is clear that the introduction of multimedia traffic has decreased the overall entropy at the level of each cluster head. In experiment 1, the average entropy at the cluster heads was 0.89 compared to entropy average of 0.42 in experiment 2. At the level of the base station, although the maximum entropy is higher for the experiment 1 compared to experiment 2, both the joint entropy and the conditional entropy are less in experiment 2 compared to experiment 1.

After the introduction of fake traffic: At the level of the cluster head, given the same number of fake sources, the maximum entropy in the presence of multimedia traffic is always more compared to the maximum entropy in experiment 1. However, the calculated entropy and normalized entropy at each cluster head in experiment 2 is less than that in experiment1. In addition, with the presence of multimedia traffic, extra transmission delay has been recorded. When 1 fake source was introduced, an average of 1881 extra seconds were required to send both all traffic. However, the average difference tends to decrease as more fake sources are added in both experiments because the average transmission time to send multimedia traffic does not significantly change compared to a

notable increase in the transmission time recorded in experiment 1. Although the entropy, joint entropy and conditional entropy is less in experiment 2 compared to experiment 1, the relative increase in the information loss or gain of experiment 2 is more than that recorded in experiment 1. This implies that the introduction of the multimedia traffic has revealed less information to the adversary.

8.2.3.4 Comparison with Related Work

Although fake traffic injection is one of the popular countermeasures against privacy attacks, this technique has not been widely applied in wireless sensor networks in the healthcare domain. Searching the literature for fake traffic injection in WMSNs or WSNs in the healthcare domain yielded a long list of survey papers (such as (Alemdar & Ersoy, 2010) (Al Ameen et al., 2012) (Li et al., 2009a)) that emphasize the importance of privacy in sensor networks and highlight the major privacy threats. In the literature, to the best of the author's knowledge, the research presented by (Buttyan & Holczer, 2012) is the closest to the research work presented in this thesis. (Buttyan & Holczer, 2012) proposed the introduction of fake traffic to WBSNs to prevent traffic analysis attacks. Their results showed that the introduction of fake traffic increased the entropy from 0.0264 to 1.6078. However, (Buttyan & Holczer, 2012) focused on the analysis of traffic signals between the sensors and the gateways and not overall traffic messages at the cluster heads and the base station. Other research papers have deployed the fake data injection mechanism but not for the healthcare domain in particular such as (Kamat et al., 2005) and (Deng et al., 2005).

(Kamat et al., 2005) proposed the deployment of fake sources of messages, which injected fake messages that are encrypted and look like real ones so that an adversary cannot tell the difference between the fake and the real messages. (Kamat et al., 2005) suggested that fake sources generate fake traffic only when an event is detected to save the energy of the sensors. (Kamat et al., 2005) suggested two strategies for fake traffic generation: short-lived fake source routing strategy and persistent fake source routing strategy. In the short-lived fake source routing strategy, a sensor sends out a fake message to its next hop neighbour and the next hop neighbour ignores the message. This strategy is energy efficient but an adversary can easily identify these fake paths. In the persistent fake source routing strategy, the locations of the fake sources are carefully chosen and their distances from the sink are comparable with the real sources. (Kamat et al., 2005) focused on the analysis of different routing techniques and did not provide an analysis of the enhancement of privacy. (Deng et al., 2005) proposed the creation and propagation of fake messages to create randomness in the network. Fake messages

were created in neighbouring nodes of sensors sending out real messages to the base station. Each fake message had a random time-to-live parameter that was decreased at each node until it reached zero. (Deng et al., 2005) used the entropy metric to assess the level of randomness. However, (Deng et al., 2005) combined fake traffic injection with random walks, fractal propagation and hotspots which make their work different and cannot be compared to this research work.

8.3 Summary

NS2-based simulation experiments were presented in this chapter to assess the enhancement in the level of privacy (anonymity, unlinkability and location privacy) after the introduction of fake sources of traffic and varying volumes of fake traffic.

Experiment 1 studied the effect of the injection of fake traffic on privacy for a network without multimedia sensors. The aim of this experiment is to study how fake traffic enhances the level of privacy of a WMSN for healthcare when multimedia sensors are turned off. Multimedia sensors in WMSN-based healthcare sub-system maybe turned off in situations when the monitoring of a patient's vitals is more important than monitoring the patient gaits (such as when patients are too sick to move or are in coma). In this situation, the network traffic needs to be kept to the minimum for speedy sending of information. The results showed that the entropy (anonymity measure) increased as more fake sources and volumes of fake traffic are introduced under each cluster head. However, the results showed that the highest normalized entropy is achieved when one fake source is added under each cluster head. This showed that adding more fake sources, and volumes of traffic, added more congestion and delay to the network and did not necessarily improve the level of privacy. In addition, the results showed that cluster heads such as cluster head 1 achieved the highest normalized entropy with 1 fake source whereas cluster head 2 achieved its highest normalized entropy with two fake sources. This showed that analysis should be conducted at each cluster head in real systems to determine the appropriate number of fake sources with regard to the number of real sources while considering the acceptable amount of delay. The results of Experiment 1 also showed that the anonymity set size (location privacy measure) increased as the number of fake sources increased. At the level of the base station, the results showed that the conditional entropy (unlinkability measure), and the information gain or loss (anonymity measure) increased as the number of fake sources and volumes of fake traffic increased. However, the highest relative percentage increase in information gain or loss was achieved when two fake sources were added, which proved that the improvement of

the level of privacy is not significantly proportional to the increase of the number of fake sources and fake volume of traffic.

In Experiment 2, the effect of fake traffic was assessed in the presence of multimedia traffic. Similar to Experiment 1, the entropy increased as more fake sources were introduced. However, unlike Experiment 1, there was no significant increase in the amount of delay due to the introduction of more fake sources and volumes of traffic. This is because the network is already handling relatively large volumes of multimedia traffic. The addition of extra fake traffic did not cause significant delay compared to the increased delay patterns when fake sources and volumes of traffic were introduced into the network in Experiment 1 (absence of multimedia traffic). Similar to Experiment 1, the highest relative increase in information gain or loss was achieved when two fake sources were added under each cluster head. This showed that adding more fake source and volumes of traffic does not necessarily enhance privacy. However, the increase of fake traffic had a serious effect on multimedia jitter. Although the results showed a significant increase in the entropy at the level of the cluster heads and the base station, that was at the expense of extra delay and the extra jitter in the multimedia network.

Consequently, based on the experiments presented in this chapter, it can be concluded that fake sources and volumes of traffic enhance privacy (anonymity / pseudonymity, unlinkability and location privacy) but at the expense of extra delay and multimedia jitter. The sources and the volumes of fake traffic must be carefully studied depending on the health condition of a patient to determine what vitals need to be monitored, whether video and audio monitoring is required or not versus the required level of privacy. In cases when the privacy of a patient is the highest priority and the collection of health related information does not have to be sent in real-time, a higher number of fake sources can be deployed.

Chapter 9 Conclusion and Future Work

9.1 Conclusion

Physical health and mental health are important ingredients of human life. They make a major contribution to the quality of life and to the economic development of individuals, communities and countries. The provision of quality healthcare, to treat or (ideally) prevent health problems, is thus acknowledged as a worthy priority in modern society. In healthcare, many researchers have highlighted the importance of privacy and security issues, however, security issues have been more extensively researched compared to little effort directed at research into privacy issues. In addition, many researchers view privacy as a by-product of applying security measures.

Consequently, the focus of this research work was to study privacy-preserving mechanisms for WMSN-based healthcare, towards ensuring the privacy-aware transmission of the data (scalar and multimedia) captured from the sensors to the base station. The novel contribution to knowledge that resulted from this research work is: (i) the identification of privacy threats, based on a threat analysis methodology named LINDDUN (Wuyts et al., 2014) which was applied to WMSNs for healthcare, to identify significant threats and hence the privacy enhancement mechanisms required by a privacy-aware WMSN; (ii) the enhancement of the AntSensNet (Cobo et al., 2010) WMSN-based routing protocol to make it privacy-aware, which widens the domain of deployment of this routing protocol to include privacy-aware applications; (iii) the results and conclusions from the experiments and analysis used for the assessment of the privacy-awareness resulting from the deployed privacy-enhancing countermeasures and from the assessment of their associated computation, communication and storage overheads.

To sum up, in this research work, countermeasures against the privacy threats identifiability, linkability and location disclosure were deployed to ensure the safe transmission of data within a WMSN-based healthcare sub-system from the sensors to the base station. The choice of these particular privacy countermeasures was based on a privacy threat analysis methodology, named LINDDUN (Wuyts et al., 2014). The methodology was applied to a WMSN-based healthcare sub-system and it was concluded that the three privacy services: anonymity/pseudonymity, unlinkability and location privacy need to be present in a privacy-aware WMSN-based healthcare sub-system. To create a privacy-aware WMSN-based healthcare sub-system, three main building blocks were investigated: the routing protocol, the security key management technique and the privacy mechanisms. The AntSensNet (Cobo et al., 2010) was chosen as the underlying routing

protocol and the LEAP key management protocol (Zhu et al., 2006) was used to implement the underlying management of security keys . However, the original AntSensNet protocol suffers from serious potential privacy threats. Consequently, the AntSensNet protocol was enhanced by adding privacy countermeasures. The enhancement was based upon an analysis, which determined the required privacy countermeasures, to be able to safely transmit data within a privacy-aware WMSN-based healthcare sub-system.

To assess the overhead due to the deployment of privacy countermeasures, the NS2 simulator was used to simulate a privacy-aware WMSN-based healthcare subsystem built upon the AntSensNet routing protocol, using a hospital scenario. The NS2 simulation showed a twofold overhead due to the introduction of privacy measures for scalar data compared to a seven-fold overhead due to the application of the privacy measures for multimedia data. This indicated that in critical medical cases, when quick intervention of medical help is required, the communication of the multimedia data should be kept to the minimum to reduce the overhead and send critical healthcare data more quickly. However, the NS2 simulation showed only the overall overhead and no details of the causes of the overhead were available.

Consequently, a theoretical analysis was used to assess the memory, computation, and network messages overhead due to the introduction of the privacy mechanisms. The overhead is represented in the form of equations outlining the extra keys (memory overhead calculation), the number of extra operations (computation overhead) and the number of extra messages (network messages overhead), compared to the original AntSensNet routing protocol. The analysis results show that the messages and memory overhead due to the added privacy mechanisms grow mostly linearly as the number of scalar and multimedia sensors increases in the network.

Finally, privacy assessments based on entropy and anonymity set size were deployed to quantify the change in the level of privacy (anonymity, unlinkability and location privacy) as the number of fake sources and the volume of traffic increase. The results show that the level of privacy increased as the number and volume of fake traffic increased. Although better privacy enhances the acceptability by patients concerned for their privacy, it is at the expense of an increased delay in the data delivery and increased level of multimedia jitter (as a result of the consumption of part of the available bandwidth by the fake traffic). High delay and multimedia jitter might not be acceptable in critical situations where quick medical help is required such as patient suffering a stroke or patient (remotely monitored by cameras) falling down and breaking a bone.

9.2 Limitations of this research work

The applied privacy countermeasures focus on the high layers of the network stack (application layer) and do not offer complete protection against potential privacy attacks at the lower layers of the network stack (such as at the MAC layer and physical layer). In addition, the privacy-aware WMSN-based healthcare sub-system proposed in this research work does not protect against other privacy threats such as unobservability and undetectability. Privacy techniques to achieve unobservability and undetectability should be included, and a patient should be able to choose what privacy services s(he) wants to apply to the healthcare and multimedia data.

Other limitations relate to the overhead performance analysis presented in Chapter 6. The analysis of the overhead was based on both simulation experiment and theoretical overhead analysis. Unfortunately, the results generated from a simulator may not be as accurate as those generated from the real implementation of the healthcare sub-system, due to use of simplified models to mimic real life scenarios. In addition the results generated from theoretical overhead analysis may not be as accurate as those generated from the real implementation of the healthcare sub-system due to ignoring network related parameters (such as queuing delays) and the use of assumptions to ease the analysis.

Finally, there might be a limited generalizability of the findings due to the limited coverage of WMSN parameters and topologies considered in this research work. The experiments presented in Chapters 6 and 7 covered a limited range of sensors (scalar and multimedia), with a limited set of data rates. In addition the hierarchal network topology upon which the experiments are based may not be appropriate for general applications (other than healthcare) such as random deployment of sensors in hazardous environments, for example: volcanoes, deep sea or dangerous zones where sensors are randomly scattered and no prior knowledge of the network topology can be determined.

9.3 Future research directions

A possible direction for the future work is to perform real implementation of the proposed healthcare sub-system and test it using functioning health, environmental, video and audio sensors. In addition, privacy mechanisms such as unobservability should be added to the privacy mechanisms investigated in this thesis, to add more choices for healthcare users who are concerned about their privacy.

After conducting a thorough literature survey in the field of privacy in WSN-based healthcare systems, it was discovered that this field needs a lot more research effort to

specifically target the privacy in the field of healthcare. Many knowledge gaps have been discovered that need attention by researchers in the future such as:

- Implementation of the identified privacy services (based on the privacy threat analysis methodology) in all layers of the network stack (from the application layer to the physical layer).
- A generic layered architecture for the WMSN-based healthcare systems stating the mechanisms that exist in each layer and on each component (sensor, gateway, end system) for all tiers (sensors tier until the processing and analysis phase at the server end).
- A formal mapping between the privacy services and the Open System Interconnection (OSI) model to enhance the development and the implementation of privacy services.
- Applying the privacy threat methodologies and risk assessment to the WMSN-based healthcare systems to create a priority list of the privacy and security services for these systems.
- Creating attack trees or threat trees to document the details of possible attacks for further analysis and consideration in systems under construction.
- Although governments are constantly seeking to enhance the legal healthcare frameworks and laws to guarantee the privacy rights of citizens, there has not yet been a direct technical mapping between the privacy services and techniques and those legal frameworks. There has to be a well-established direct mapping between technical and legal approaches to improve the possibility that governments and patients would accept the developed healthcare systems.

References

Abuzneid, A., Sobh, T. & Faezipour, M. (2015a). An enhanced communication protocol for anonymity and location privacy in WSN. In: *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNCW)*. 2015, New Orleans, LA, pp. 91–96.

Abuzneid, A.S., Sobh, T., Faezipour, M., Mahmood, A. & James, J. (2015b). Fortified anonymous communication protocol for location privacy in WSN: A modular approach. *Sensors*. 15 (3). p.pp. 5820–5864.

Acharya, U. & Younis, M. (2010). Increasing base-station anonymity in wireless sensor networks. *Ad Hoc Networks*. 8 (8). p.pp. 791–809.

Adams, A. & Sasse, M.A. (2001). Privacy in multimedia communications: Protecting users, not just data. In: *People and Computers XV—Interaction without Frontiers*. 2001, Springer, London, pp. 49–64.

Adhyapak, D.S.B. & Laturkar, A.P. (2018). Swarm Based Cross Layer Optimization Protocol for WMSN. *Indonesian Journal of Electrical Engineering and Computer Science*. 10 (1). p.pp. 302–308.

Agrafioti, F., Bui, F.M. & Hatzinakos, D. (2009). On supporting anonymity in a BAN biometric framework. In: *16th International Conference on Digital Signal Processing*. 2009, pp. 1–6.

Akhlaq, M. & Sheltami, T.R. (2012). Performance comparison of video compression and streaming over wireless ad hoc and sensor networks using MPEG-4 and H. 264. In: *International Conference on Networked Digital Technologies*. 2012, Springer, Berlin, Heidelberg, pp. 368–377.

Akyildiz, I., Melodia, T. & Chowdhury, K. (2007). A survey on wireless multimedia sensor networks. *Computer Networks*. 51 (4). p.pp. 921–960.

Akyildiz, I.F., Melodia, T. & Chowdhury, K.R. (2008). Wireless multimedia sensor networks: Applications and testbeds. *Proceedings of the IEEE*. 96 (10). p.pp. 1588–1605.

Alemdar, H. & Ersoy, C. (2010). Wireless sensor networks for healthcare: A survey. *Computer Networks*. 54 (15). p.pp. 2688–2710.

Almalkawi, I.T., Zapata, M.G., Al-Karaki, J.N. & Morillo-Pozo, J. (2010). Wireless multimedia sensor networks: current trends and future directions. *Sensors*. 10 (7). p.pp. 6662–6717.

Alomair, B., Clark, A., Cuellar, J. & Poovendran, R. (2010). Statistical framework for source anonymity in sensor networks. In: *Proceedings of IEEE Global Telecommunications (GLOBECOM)*. 2010, Miami, Florida, USA, pp. 1–6.

Al Ameen, M., Liu, J. & Kwak, K. (2012). Security and privacy issues in wireless sensor networks for healthcare applications. *Journal of Medical Systems*. 36 (1). p.pp. 93–101.

Andersson, C. & Lundin, R. (2008). On the fundamentals of anonymity metrics. In: *TFischer-Hübner S., Duquenoy P., Zuccato A., Martucci L. (eds) The Future of Identity in the Information Society. IFIP — The International Federation for Information Processing*. 2008, Boston, MA: Springer US, pp. 325–341.

Beresford, A.R. & Stajano, F. (2003). Location privacy in pervasive computing. *IEEE Pervasive computing*. 2 (1). p.pp. 46–55.

Bergstrom, C. (2008). *Lecture 3: Joint entropy, conditional entropy, relative entropy, and mutual information*. [Online]. 2008. Available from: http://octavia.zoology.washington.edu/teaching/429/lecturenotes/lecture3.pdf. [Accessed: 15 January 2017].

Bhandary, V., Malik, A. & Kumar, S. (2016). Routing in wireless multimedia sensor networks: A survey of existing protocols and open research issues. *Journal of Engineering*. 2016. p.p. 27.

Bi, J., Li, Z. & Wang, R. (2010). An ant colony optimization-based load balancing routing algorithm for wireless multimedia sensor networks. In: *Proceeding of 12th IEEE International Conference on Communication Technology (ICCT)*. 2010, pp. 584– 587.

Breaux, T.D. & Anton, A.I. (2008). Analyzing regulatory rules for privacy and security requirements. *IEEE transactions on software engineering*. 34 (1). p.pp. 5–20.

Breaux, T.D. & Antón, A.I. (2005). Mining rule semantics to understand legislative compliance. In: *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. 2005, Alexandria, Virginia, USA, pp. 51–54.

Buttyan, L. & Holczer, T. (2012). Traffic analysis attacks and countermeasures in wireless body area sensor networks. In: *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 2012, pp. 1–6.

Buttyán, L., Holczer, T. & Vajda, I. (2006). Optimal key-trees for tree-based private authentication. In: *Privacy Enhancing Technologies*. 2006, Springer Berlin/Heidelberg, pp. 332–350.

Caesar, P.M. (2010). *Lecture 14: Performance Analysis*. [Online]. 2010. University of illinois. Available from: https://courses.engr.illinois.edu/cs438/sp2010/slides/lec15_perf.pdf. [Accessed: 20 January 2017].

Carbunar, B., Yu, Y., Shi, W., Pearce, M. & Vasudevan, V. (2010). Query privacy in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*. 6 (2). p.p. 14.

Caro, D., Gianni & Dorigo, M. (1998). AntNet: Distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research*. 9. p.pp. 317–365.

Chakravorty, R. (2006). A programmable service architecture for mobile medical care. In: *Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, (PerCom)*. 2006, Pisa, Italy, pp. 13–17.

Chen, J., Du, X. & Fang, B. (2012). An efficient anonymous communication protocol for wireless sensor networks. *Wireless Communications and Mobile Computing*. 12 (14). p.pp. 1302–1312.

Chen, M., Gonzalez, S., Vasilakos, A., Cao, H. & Leung, V.C.M. (2011). Body area networks: A survey. *Mobile Networks and Applications*. 16 (2). p.pp. 171–193.

Chen, S. (2007). Protecting Receiver-Location Privacy in Wireless Sensor Networks. In: *26th IEEE International Conference on Computer Communications INFOCOM 2007.* 2007, pp. 1955–1963.

Cobo, L., Quintero, A. & Pierre, S. (2010). Ant-based routing for wireless multimedia sensor networks using multiple QoS metrics. *Computer Networks*. 54 (17). p.pp. 2991–3010.

Corporation, I. (2017). *Intel® 64 and IA-32 architectures optimization reference manual*. [Online]. 2017. Available from: https://software.intel.com/sites/default/files/managed/9e/bc/64-ia-32-architectures-optimization-manual.pdf#page740. [Accessed: 30 November 2017].

Crosby, G. V, Ghosh, T., Murimi, R. & Chin, C. a (2012). Wireless body area networks for healthcare : A survey. *International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC)*. 3 (3). p.pp. 1–26.

Danezis, G. (2013). Measuring anonymity: a few thoughts and a differentially private bound. In: *Proceedings of the DIMACS Workshop on Measuring Anonymity.* 2013, p. 10.

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Metayer, D. Le, Tirtea, R. & Schiffner, S. (2015). *Privacy and data protection by design– from policy to engineering*.

Darwish, A. & Hassanien, A.E. (2011). Wearable and Implantable Wireless Sensor Network Solutions for Healthcare Monitoring. *Sensors*. 11 (6). p.pp. 5561–5595.

Das, A.K. & Goswami, A. (2013). A secure and efficient uniqueness-and-anonymity-preserving remote user authentication scheme for connected health care. *Journal of medical systems*. 37 (3). p.p. 9948.

Deif, D.S. & Gadallah, Y. (2014). Classification of wireless sensor networks deployment techniques. *IEEE Communications Surveys & Tutorials*. 16 (2). p.pp. 834–855.

Deng, J., Han, R. & Mishra, S. (2005). Countermeasures against traffic analysis attacks in wireless sensor networks. In: *First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SecureComm)*. 2005, pp. 113–126.

Deng, J., Han, R. & Mishra, S. (2006). Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks. *Pervasive and Mobile Computing*. 2 (2). p.pp. 159–186.

Deng, M., Wuyts, K., Scandariato, R., Preneel, B. & Joosen, W. (2011). A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*. 16 (1). p.pp. 3–32.

Di & Tsudik, G. (2010). Security and privacy in emerging wireless networks. *IEEE Wireless Communications*. 17 (5). p.pp. 12–21.

Dias, J.A.C., Machado, P. & Pereira, F.C. (2013). Privacy-aware ant colony optimization agorithm for real time route planning. In: *Proceedings of the World Conference on Transport Research*. 2013, p. 9.

Diaz, C., Seys, S., Claessens, J. & Preneel, B. (2002). Towards measuring anonymity. In: *International Workshop on Privacy Enhancing Technologies*. 2002, Springer Berlin Heidelberg, pp. 54–68.

Ding, W. & Ping, W. (2014). On the anonymity of two-factor authentication schemes for wireless sensor networks: Attacks, principle and solutions. *Computer Networks*. 73. p.pp. 41–57.

Ebrahimi, Y. & Younis, M. (2011). Increasing transmission power for higher base-station anonymity in wireless sensor network. In: *IEEE International Conference on Communications (ICC)*. 2011, pp. 1–5.

Egbogah, E.E. & Fapojuwo, A.O. (2011). A survey of system architecture requirements for health care-based wireless sensor networks. *Sensors*. 11 (5). p.pp. 4875–4898.

Ehsan, S. & Hamdaoui, B. (2012). A survey on energy-efficient routing techniques with QoS assurances for wireless multimedia sensor networks. *IEEE Communications Surveys and Tutorials*. 14 (2). p.pp. 265–278.

Fidaleo, D. a, Nguyen, H.-A. & Trivedi, M. (2004). The networked sensor tapestry (NeST): A privacy enhanced software architecture for interactive analysis of data in video-sensor networks. In: *Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*. 2004, pp. 46–53.

Fischer, L., Katzenbeisser, S. & Eckert, C. (2008). Measuring unlinkability revisited. In: *Proceedings of the 7th ACM workshop on Privacy in the electronic society*. 2008, pp. 105–110.

Friedland, G., Janin, A., Lei, H., Choi, J. & Sommer, R. (2015). Content-based privacy for consumer-produced multimedia. In: *Multimedia Data Mining and Analytics*. Springer International Publishing, pp. pp. 157–173.

Ganesan, P., Venugopalan, R.P.P., Dean, A., Mueller, F. & Sichitiu, M. (2003).

Analyzing and modeling encryption overhead for sensor network nodes. In: *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications*. 2003, pp. 151–159.

Ganti, R.K., Jayachandran, P., Abdelzaher, T.F. & Stankovic, J. a. (2006). Satire: A Software Archiecture for Smart AtTIRE. In: *Proceedings of the 4th international conference on Mobile systems, applications and services (MobiSys)*. pp. 110–123.

Garitano, I., Fayyad, S. & Noll, J. (2015). Multi-metrics approach for security, privacy and dependability in embedded systems. *Wireless Personal Communications*. 81 (4). p.pp. 1359–1376.

Garverick, S.L., Ghasemzadeh, H., Zurcher, M., Roham, M. & Saldivar, E. (2011). Wireless fetal monitoring device with provisions for multiple births. In: *2011 International Conference on Body Sensor Networks (BSN)*. 2011, pp. 113–118.

Gedik, B. & Liu, L. (2008). Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*. 7 (1). p.pp. 1–18.

Gross, R., Sweeney, L. & Fernando De la Torre, S.B. (2006). Model-based face de-identification. In: *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06).* 2006, pp. 161–168.

Gupta, P. & Chawla, M. (2012). Privacy preservation for WSN: A Survey. *International Journal of Computer Applications*. 48 (3). p.pp. 11–16.

Ha, I. (2015). Technologies and research trends in wireless body area networks for healthcare : A systematic literature review. *International Journal of Distributed Sensor Networks*. 11 (6). p.pp. 1–14.

Haddad, W., Nordmark, E., Haddad, E.N. & Nordmark, E. (2011). *Privacy aspects terminology*. [Online]. 2011. Network Working Group. Available from: https://www.ietf.org/archive/id/draft-haddad-alien-privacy-terminology-06.txt. [Accessed: 29 November 2017].

Halperin, D., Heydt-Benjamin, T.S., Fu, K., Kohno, T. & Maisel, W.H. (2008). Security and privacy for implantable medical devices. *IEEE Pervasive Computing*. 7 (1). p.pp. 30–39.

Hamid, N.I. Bin, Harouna, M.T., Salele, N. & Muhammad, R. (2013). Comparative analysis of various wireless multimedia sensor networks for telemedicine. *International Journal of Computer Applications*. 73 (16). p.pp. 39–44.

Hayawi, K., Mortezaei, A. & Tripunitara, M. V. (2015). The limits of the trade-off between query-anonymity and communication-cost in wireless sensor networks. In: *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*. 2015, pp. 337–348.

He, D., Kumar, N. & Chilamkurti, N. (2015). A secure temporal-credential-based mutual authentication and key agreement scheme with pseudo identity for wireless sensor networks. *Information Sciences*. 321. p.pp. 263–277.

He, W., Nguyen, H., Liuy, X., Nahrstedt, K. & Abdelzaher, T. (2008). iPDA: An integrity-protecting private data aggregation scheme for wireless sensor networks. In: *Military Communications Conference IEEE (MILCOM)*. 2008, Piscataway, NJ, USA, pp. 1–7.

Hiller, J., McMullen, M.S., Chumney, W.M. & Baumer, D.L. (2011). Privacy and security in the implementation of health information technology (electronic health records): US and EU compared. *BUJ Sci. & Tech. L.* 17. p.pp. 1–39.

Honeine, P., Mourad, F., Kallas, M., Snoussi, H., Amoud, H. & Francis, C. (2011). Wireless sensor networks in biomedical: body area networks. In: *Proceedings of 2011 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*. 2011, Tipaza, Algeria, pp. 388–391.

Horey, J., Groat, M.M., Forrest, S. & Esponda, F. (2007). Anonymous Data Collection in Sensor Networks. In: *Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services (MobiQuitous)*. 2007, Philadelphia, PA USA, pp. 1–8.

Hu, S., Shao, Z. & Tan, J. (2011). A real-time cardiac arrhythmia classification system with wearable electrocardiogram. In: *2011 International Conference on Body Sensor Networks (BSN)*. 2011, Dallas, TX, USA, pp. 119–124.

Hwang, Y. & Yuan, S. (2007). A Privacy-Aware identity design for exploring ubiquitous collaborative wisdom. In: *Computational science*. 2007, Springer-Verlag, Berlin, Heidelberg, pp. 433–440.

Inan, I., Keceli, F. & Ayanoglu, E. (2006). An adaptive multimedia QoS scheduler for

802.11 e wireless LANs. In: *IEEE International Conference on Communications (ICC'06)*. 2006, pp. 5263–5270.

Islam, K., Shen, W. & Wang, X. (2012). Security and privacy considerations for Wireless Sensor Networks in smart home environments. In: *IEEE 16th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 2012, pp. 626–633.

Javadi, S.S. & Razzaque, M. a. (2013). Security and privacy in wireless body area networks for health care applications. In: *Wireless Networks and Security*. 2013, Springer, Berlin, Heidelberg., pp. 165–187.

Jian, Y., Chen, S., Zhang, Z. & Zhang, L. (2007). Protecting receiver-location privacy in wireless sensor networks. In: *26th IEEE International Conference on Computer Communications (INFOCOM)*. 2007, pp. 1955–1963.

Kalloniatis, C., Kavakli, E. & Gritzalis, S. (2008). Addressing privacy requirements in system design: The PriS method. *Requirements Engineering*. 13 (3). p.pp. 241–255.

Kalpana, R. & Rengarajan, N. (2012). Mobile anonymous trust based routing using ant colony optimization. *American Journal of Applied Sciences*. 9 (8). p.pp. 1283–1289.

Kamat, P., Zhang, Y., Trappe, W. & Ozturk, C. (2005). Enhancing source-location privacy in sensor network routing. In: *Proceedings of 25th IEEE International Conference on Distributed Computing Systems (ICDCS)*. 2005, pp. 599–608.

Kambourakis, G., Klaoudatou, E. & Gritzalis, S. (2007). Securing medical sensor Environments: The CodeBlue framework case. In: *The Second International Conference on Availability, Reliability and Security (ARES)*. 2007, pp. 637–643.

Kandris, D., Tsagkaropoulos, M., Politis, I., Tzes, A. & Kotsopoulos, S. (2011). Energy efficient and perceived QoS aware video routing over Wireless Multimedia Sensor Networks. *Ad Hoc Networks*. 9 (4). p.pp. 591–607.

Khan, I.M., Jabeur, N., Khan, M.Z. & Mokhtar, H. (2012). An overview of the impact of wireless sensor networks in medical health care. In: *Proceedings of the 1st International Conference on Computing and Information Technology (ICCT)*. 2012, Al-Madinah Al-Munawwarah, Saudi Arabia, pp. 12–14.

Kim, J., Baek, J. & Shon, T. (2011). An efficient and scalable re-authentication protocol over wireless sensor network. *IEEE Transactions on Consumer Electronics*. 57 (2). p.pp. 516–522.

Ko, J., Dutton, R.P., Lim, J.H., Chen, Y., Musvaloiu-E, R., Terzis, A., Masson, G.M., Gao, T., Destler, W. & Selavo, L. (2010a). MEDiSN : Medical Emergency Detection in Sensor Networks. *ACM Transactions on Embedded Computing Systems*. 10 (1). p.pp. 1–29.

Ko, J., Lu, C., Srivastava, M. & Stankovic, J. (2010b). Wireless sensor networks for healthcare. *Proceedings of the IEEE*. 98 (11). p.pp. 1947–1960.

Kohlmayer, F., Bild, R., Västrik, I., Kuhn, K., Rodriguez, B., Brunner, S., Lamichhane, A. & Ohmann, C. (2014). *Building data bridges between biological and medical infrastructures in Europe*. [Online]. 2014. BioMedBridges. Available from: http://www.biomedbridges.eu/sites/biomedbridges.eu/files/documents/deliverables/d8-1_process_specification_for_secure_sharing_of_and_access_to_pm_data.pdf. [Accessed: 15 August 2017].

Kotz, D., Avancha, S. & Baxi, A. (2009). A privacy framework for mobile health and home-care systems. In: *Proceedings of the first ACM workshop on Security and privacy in medical and home-care systems*. 2009, pp. 1–12.

Krumm, J. (2009). A survey of computational location privacy. *Personal and Ubiquitous Computing*. 13 (6). p.pp. 391–399.

Kumar, P. & Lee, H.-J. (2011). Security Issues in healthcare applications using wireless medical sensor networks: A survey. *Sensors*. 12 (1). p.pp. 55–91.

Kurose, J.F. & Ross, K.W. (2017). *Computer networking : a top-down approach*. 7th Ed. Boston : Pearson.

Latré, B., Braem, B., Moerman, I. & Blondia, C. (2011). A Survey on wireless body area networks. *Wireless Networks*. 17 (1). p.pp. 1–18.

Leon, M.D.L.A.C., Hipolito, J.I.N. & Garcia, J.L. (2009). A security and privacy survey for WSN in e-health applications. In: *Proceedings of the Conference on Electronics, Robotics and Automotive Mechanics (CERMA'09)*. September 2009, Cuernavaca, Morelos, Mexico, pp. 125–130.

Li, H., Ma, J. & Fu, S. (2015a). A privacy-preserving data collection model for digital community. *Science China Information Sciences*. 53 (3). p.pp. 1–16.

Li, M., Lou, W. & Ren, K. (2010). Data security and privacy in wireless body area networks. *IEEE Wireless Communications*. 17 (1). p.pp. 51–58.

Li, N., Zhang, N., Das, S.K. & Thuraisingham, B. (2009a). Privacy preservation in wireless sensor networks: A state-of-the-art survey. *Ad Hoc Networks*. 7 (8). p.pp. 1501–1514.

Li, X., Niu, J., Kumari, S., Liao, J., Liang, W., Khan, M.K. & Khan, M.K. (2015b). A new authentication protocol for healthcareapplications using wireless medical sensornetworks with user anonymity. *Security and Communication Networks*. 9 (15). p.pp. 2643–2655.

Li, X., Wang, X., Zheng, N., Wan, Z. & Gu, M. (2009b). Enhanced location privacy protection of base station in wireless sensor networks. In: *5th International Conference on Mobile Ad-hoc and Sensor Networks (MSN'09)*. 2009, pp. 457–464.

Liang, X., Lu, R., Chen, L., Lin, X. & Shen, X. (2011). PEC: A privacy-preserving emergency call scheme for mobile healthcare social networks. *Journal of Communications and Networks*. 13 (2). p.pp. 102–112.

Lo, B.P., Wang, J.L. & Yang, G.-Z. (2005). From imaging networks to behavior profiling: Ubiquitous sensing for managed homecare of the elderly. In: *Adjunct Proceedings of the 3rd International Conference on Pervasive Computing (PERVASIVE)*. 2005, Munich, pp. 101–104.

López, G., Custodio, V. & Moreno, J.I. (2010). LOBIN: E-textile and wireless-sensor-network-based platform for healthcare monitoring in future hospital environments. *IEEE Transactions on Information Technology in Biomedicine*. 14 (6). p.pp. 1446–1458.

Lu, T., Yao, P., Zhao, L., Li, Y., Xie, F. & Xia, Y. (2015). Towards Attacks and Defenses of Anonymous Communication Systems. *International Journal of Security and Its Applications*. 9 (1). p.pp. 313–328.

Luh, W., Kundur, D. & Zourntos, T. (2007). A novel distributed privacy paradigm for visual sensor networks based on sharing dynamical systems. *EURASIP Journal on Applied Signal Processing*. 2007 (1). p.pp. 218–218.

Luna, J., Suri, N. & Krontiris, I. (2012). Privacy-by-design based on quantitative threat modeling. In: *7th International Conference on Risk and Security of Internet and Systems (CRiSIS)*. 2012, pp. 1–8.

Mahmoud, M.E. (2012). Secure and efficient source location privacy-preserving scheme for wireless sensor networks. In: *IEEE International Conference on Communications (IEEE ICC'12)*. 2012, Ottawa Canada, pp. 1123–1127.

Malan, D., Fulford-Jones, T., Welsh, M. & Moulton, S. (2004). Codeblue: An ad hoc sensor network infrastructure for emergency medical care. In: *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*. 2004, pp. 3–6.

Mare, S., Sorber, J., Shin, M., Cornelius, C., Kotz, D. & Korea, S. (2011). Hide-n-Sense : Privacy-aware secure mHealth sensing. *Technical Report TR2011-702, Dartmouth College, Computer Science, Hanover, NH*. p.pp. 1–18.

Meingast, M., Roosta, T. & Sastry, S. (2006). Security and privacy issues with health care information technology. In: *Proceedings of the 8th Annual International Conference of the IEEE Engineering in Medicine and Biology*. 2006, pp. 5453–5458.

Milenković, A., Otto, C. & Jovanov, E. (2006). Wireless sensor networks for personal health monitoring: Issues and an implementation. *Computer Communications*. 29 (13). p.pp. 2521–2533.

Minh-Thanh Vo, T.T., Nghi, T., Tran, V.-S., Mai, L. & Le, C.-T. (2015). Wireless sensor network for real time healthcare monitoring: Network design and performance evaluation simulation. In: *5th International Conference on Biomedical Engineering in Vietnam*. 2015, Berlin: Springer International Publishing, pp. 87–91.

Mitra, U., Emken, B.A., Lee, S., Li, M. & California, S. (2012). Commuincation in ubiquitous healthcrae Knowme: A case study in wireless body area sensor network design. *IEEE communications magazine*. 50 (5). p.pp. 116–125.

Moncrieff, S., Venkatesh, S. & West, G. (2008). Dynamic privacy assessment in a smart house environment using multimodal sensing. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*. 5 (2). p.pp. 10:1–10:29.

Movassaghi, S., Abolhasan, M., Lipman, J., Smith, D. & Jamalipour, A. (2014). Wireless body area networks: A survey. *IEEE Communications Surveys & Tutorials*. 16 (3). p.pp. 1658–1686.

Murdoch, S.J. (2014). Quantifying and measuring anonymity. In: *Data Privacy Management and Autonomous Spontaneous Security*. 2014, Springer Berlin Heidelberg, pp. 3–13.

Nabar, S., Banerjee, A., Gupta, S.K.S. & Poovendran, R. (2011). GeM-REM: Generative model-driven resource efficient ECG monitoring in body sensor networks. In: *Proceedings - 2011 International Conference on Body Sensor Networks (BSN)*. 2011, pp. 1–6.

Nadeem, A., Hussain, M.A., Owais, O., Salam, A., Iqbal, S. & Ahsan, K. (2015). Application specific study, analysis and classification of body area wireless sensor network applications. *Computer Networks*. 83. p.pp. 363–380.

Nam, J., Choo, K.-K.R., Han, S., Kim, M., Paik, J. & Won, D. (2015). Efficient and anonymous two-factor user authentication in wireless sensor networks: Achieving user anonymity with lightweight sensor computation. *Plos One*. 10 (4). p.p. e0116709.

Nayyar, A. & Singh, R. (2017). Simulation and performance comparison of Ant Colony Optimization (ACO) routing protocol with AODV, DSDV, DSR routing protocols of wireless sensor networks using NS-2 simulator. *American Journal of Intelligent Systems*. 7 (1). p.pp. 19–30.

Nezhad, A.A., Miri, A. & Makrakis, D. (2008). Location privacy and anonymity preserving routing for wireless sensor networks. *Computer Networks*. 52 (18). p.pp. 3433–3452.

Nia, A.M., Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A. & Jha, N.K. (2015). Energy-efficient long-term continuous personal health monitoring. *IEEE Transactions on multiscale computing systems*. 1 (2). p.pp. 85–98.

Nohara, Y., Inoue, S., Baba, K. & Yasuura, H. (2005). Quantitative evaluation of unlinkable ID matching schemes. In: *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. 2005, pp. 55–60.

Oualha, N. & Olivereau, A. (2011). Sensor and data privacy in industrial wireless sensor networks. In: *Conference on Network and Information Systems Security*

(SAR-SSI). 2011, pp. 1–8.

Ozturk, C., Zhang, Y. & Trappe, W. (2004). Source-location privacy in energy-constrained sensor network routing. In: *Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*. 2004, pp. 88–93.

Parliament, E. (2002). *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)*. [Online]. 2002. Official Journal of the European Communities. Available from: http://europa.eu.int/eur-lex/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf. [Accessed: 21 January 2016].

Parliament, E. (1995). *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data*. [Online]. 1995. Official Journal of the European Communities. Available from: http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf. [Accessed: 21 January 2016].

Pfitzmann, A. & Hansen, M. (2010). *A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management*. [Online]. 2010. Technical University Dresden. Available from: http://dud.inf.tu-dresden.de/ Anon_Terminology.shtml. [Accessed: 25 January 2017].

Politis, I., Tsagkaropoulos, M., Dagiuklas, T. & Kotsopoulos, S. (2008). Power efficient video multipath transmission over wireless multimedia sensor networks. *Mobile Networks and Applications*. 13 (3-4). p.pp. 274–284.

Preneel, B. & Ikonomou, D. (2014). Privacy technologies and policy. In: *First Annual Privacy Forum, APF 2012, Limassol, Cyprus, October 10-11, 2012, Revised Selected Papers*. Springer-Verlag Berlin Heidelberg, p. 215.

Rahman, A., Ghasemaghaeil, R., Saddikl, A. El & Gueaieb, W. (2008). M-IAR : Biologically inspired routing protocol for wireless multimedia sensor networks. In: *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (IMTC)*. 2008, pp. 1823–1827.

Rawat, P., Singh, K.D., Chaouchi, H. & Bonnin, J.M. (2014). Wireless sensor networks: a survey on recent developments and potential synergies. *The Journal of supercomputing*. 68 (1). p.pp. 1–48.

Reza, S., G, T., JY, L.B. & JP, H. (2011). Quantifying location privacy. In: *2011 IEEE Symposium on Security and Privacy*. 2011, pp. 247–262.

Ribaric, S., Ariyaeeinia, A. & Pavesic, N. (2016). De-identification for privacy protection in multimedia content: A survey. *Signal Processing: Image Communication*. 47. p.pp. 131–151.

Riosa, R., Cuellarb, J. & Lopez, J. (2015). Probabilistic receiver-location privacy protection in wireless sensor network. *Information Sciences*. 321. p.pp. 205 – 223.

Rofouei, M., Sinclair, M., Bittner, R., Blank, T., Saw, N., DeJean, G. & Heffron, J. (2011). A Non-invasive wearable neck-cuff system for real-time sleep monitoring. In: *2011 International Conference on Body Sensor Networks*. 2011, pp. 156–161.

Saini, M., Atrey, P.K., Mehrotra, S. & Kankanhalli, M. (2014). W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. *Multimedia Tools and Applications*. 68 (1). p.pp. 135–158.

Saleem, M., Di Caro, G. & Farooq, M. (2011). Swarm intelligence based routing protocol for wireless sensor networks: Survey and future directions. *Information Sciences*. 181 (20). p.pp. 4597–4624.

Selvakennedy, S., Sinnappan, S. & Shang, Y. (2006). T-ANT: a nature-inspired data gathering protocol for wireless sensor networks. *Journal of Communications*. 1 (2). p.p. 2006.

Serjantov, A. & Danezis, G. (2003). Towards an information theoretic metric for anonymity. In: *Proc. 2nd International Workshop on Privacy Enhancing Technologies (PET)*. 2003, pp. 41–53.

Shao, M., Yang, Y., Zhu, S. & Cao, G. (2007). Towards statistically strong source anonymity for sensor networks. In: *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM)*. 2007, pp. 1298–1306.

Sharif, A., Potdar, V. & Chang, E. (2009). Wireless multimedia sensor network

technology: A survey. In: *7th IEEE international conference on Industrial informatics (INDIN)*. 2009, pp. 606–613.

Sheu, J.P., Jiang, J.R. & Tu, C. (2008). Anonymous path routing in wireless sensor networks. In: *IEEE International Conference on Communications (ICC)*. 2008, Piscataway, USA, pp. 2728–2734.

Shokri, R., Freudiger, J. & Hubaux, J.-P. (2010). A unified framework for location privacy. In: *3rd Hot Topics in Privacy Enhancing Technologies (HotPETs)*. 2010, pp. 1–19.

Sohraby, K., Minoli, D., Znati, T. & Sohraby K, Minoli D, Z.T. (2007). Wireless sensor networks: technology, protocols, and applications. In: *John Wiley & Sons*. Wiley-Interscience, p. 328.

Stallings, W. (2016). Cryptography and network security: principles and practices. In: *Cryptography and Network Security*. Pearson Education India, p. 768.

Suh, C., Mir, Z.H. & Ko, Y.-B. (2008). Design and implementation of enhanced IEEE 802.15. 4 for supporting multimedia service in Wireless Sensor Networks. *Computer Networks*. 52 (13). p.pp. 2568–2581.

Sun, J., Zhu, X. & Fang, Y. (2010). Preserving privacy in emergency response based on wireless body sensor networks. In: *IEEE Global Telecommunications Conference (GLOBECOM)*. 2010, pp. 1–6.

Sun, Y., Ma, H., Liu, L. & Zheng, Y. (2008). ASAR: An ant-based service-aware routing algorithm for multimedia sensor networks. *Frontiers of Electrical and Electronic Engineering in China*. 3 (1). p.pp. 25–33.

Tan, K., Wu, D., Chan, A.J. & Mohapatra, P. (2011). Comparing simulation tools and experimental testbeds for wireless mesh networks. *Pervasive and Mobile Computing*. 7 (4). p.pp. 434–448.

Tavares, J., Velez, F.J. & Ferro, J.M. (2008). Application of wireless sensor networks to automobiles. *Measurement Science Review*. 8 (3). p.pp. 65–70.

US Department of Health and Human Services (2008). *Nationwide privacy and security framework for electronic exchange of individually identifiable health information*. [Online]. 2008. Available from: https://www.healthit.gov/sites/default/files/nationwide-ps-framework-5.pdf.

[Accessed: 1 March 2016].

Vaidya, B., Rodrigues, J.J. & Park, J.H. (2010). User authentication schemes with pseudonymity for ubiquitous sensor network in NGN. *International Journal of Communication Systems*. 23 (9-10). p.pp. 1201–1222.

Virone, G., Wood, A., Selavo, L., Cao, Q., Fang, L., Doan, T., He, Z., Stoleru, R., Lin, S. & Stankovic, J. a (2006). An Advanced wireless sensor network for health monitoring. In: *Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare (D2H2)*. 2006, pp. 2–5.

Wagner, I. & Eckhoff, D. (2015). *Technical privacy metrics: a systematic survey*. [Online]. 2015. Available from: http://arxiv.org/abs/1512.00327. [Accessed: 21 September 2017].

Wang, J., Zhang, Z., Xu, K., Yin, Y. & Guo, P. (2013). A research on security and privacy issues for patient related data in medical organization system. *International Journal of Security and its Applications*. 7 (4). p.pp. 287–298.

Ward, J.R. & Younis, M. (2015). Increasing base station anonymity using distributed beamforming. *Ad Hoc Networks*. 32. p.pp. 53–80.

Wickramasuriya, J., Datt, M., Mehrotra, S. & Venkatasubramanian, N. (2004). Privacy protecting data collection in media spaces. In: *Proceedings of the 12th annual ACM international conference on Multimedia*. 2004, pp. 48–55.

Winkler, T. & Rinner, B. (2014). Security and privacy protection in visual sensor networks : A survey. *ACM Computing Surveys (CSUR)*. 47 (1). p.pp. 1–42.

Wood, A.D., Stankovic, J.A., Virone, G., Selavo, L., He, Z., Cao, Q., Doan, T., Wu, Y., Fang, L. & Stoleru, R. (2008). Context-aware wireless sensor networks for assisted living and residential monitoring. *IEEE Network*. 22 (4). p.pp. 26–33.

Wuyts, K., Griet Verhenneman, Riccardo Scandariato, W.J. & Dumortier, J. (2012). What electronic health records don't know just yet. A privacy analysis for patient communities and health records interaction. *Health and Technology*. 2 (3). p.pp. 159–183.

Wuyts, K., Scandariato, R. & Joosen, W. (2014). Empirical evaluation of a privacy-focused threat modeling methodology. *Journal of Systems and Software*. 96. p.pp. 122–138.

Yang, Y., Shao, M., Zhu, S., Urgaonkar, B. & Cao, G. (2008). Towards event source unobservability with minimum network traffic in sensor networks. In: *Proceedings of the first ACM conference on Wireless network security (WiSec)*. 2008, pp. 77–88.

Yu, X., Luo, J. & Huang, J. (2011). An ant colony optimization- based QoS routing algorithm for wireless multimedia sensor networks. In: *Proceedings of the 3rd International Conference on Communication Software and Networks (ICCSN)*. 2011, pp. 37–41.

Yuce, M.R. (2010). Implementation of wireless body area networks for healthcare systems. *Sensors and Actuators A: Physical*. 162 (1). p.pp. 116–129.

Zhang, L., Shen, D., Shan, X. & Li, V.O. (2005). An ant-based multicasting protocol in mobile ad-hoc network. *International Journal of Computational Intelligence and Applications*. 5 (2). p.pp. 185–199.

Zhang, R., Zhang, Y. & Ren, K. (2009). DP^2AC: Distributed Privacy-Preserving Access Control in Sensor Networks. In: *28th Conference on Computer Communications (INFOCOM)*. 2009, pp. 1251–1259.

Zhang, Z.-Q., Pansiot, J., Lo, B. & Yang, G.-Z. (2011). Human back movement analysis using BSN. In: *2011 IEEE International Conference on Body Sensor Networks*. 2011, pp. 13–18.

Zhou, L. & Wen, Q. (2014). Energy efficient source location privacy protecting scheme in wireless sensor networks using ant colony optimization. *International Journal of Distributed Sensor Networks*. 10 (3). p.p. 14.

Zhu, S., Setia, S. & Jajodia, S. (2006). {LEAP}: Efficient security mechanisms for large-scale distributed sensor networks. *ACM Transactions on Sensor Networks (TOSN)*. 2 (4). p.pp. 500–528.