Use of Inferential Statistics to Design Effective Communication Protocols for Wireless Sensor Networks

Mohammed Omar M Baz

PhD

University of York
Electronics
September 2014

Abstract

This thesis explores the issues and techniques associated with employing the principles of inferential statistics to design effective Medium Access Control (MAC), routing and duty cycle management strategies for multihop Wireless Sensor Networks (WSNs). The main objective of these protocols are to maximise the throughput of the network, to prolong the lifetime of nodes and to reduce the end-to-end delay of packets over a general network scenario without particular considerations for specific topology configurations, traffic patterns or routing policies.

WSNs represent one of the leading-edge technologies that have received substantial research efforts due to their prominent roles in many applications. However, to design effective communication protocols for WSNs is particularly challenging due to the scarce resources of these networks and the requirement for large-scale deployment. The MAC, routing and duty cycle management protocols are amongst the important strategies that are required to ensure correct operations of WSNs. This thesis makes use of the inferential statistics field to design these protocols; inferential statistics was selected as it provides a rich design space with powerful approaches and methods.

The MAC protocol proposed in this thesis exploits the statistical characteristics of the Gamma distribution to enable each node to adjust its contention parameters dynamically based on its inference for the channel occupancy. This technique reduces the service time of packets and leverages the throughput by improving the channel utilisation. Reducing the service time minimises the energy consumed in contention to access the channel which in turn prolongs the lifetime of nodes. The proposed duty cycle management scheme uses non-parametric Bayesian inference to enable each node to determine the best times and durations for its sleeping durations without posing overheads on the network. Hence the lifetime of node is prolonged by mitigating the amount of energy wasted in overhearing and idle listening. Prolonging the lifetime of nodes increases the throughput of the network and reduces the end-to-end delay as it allows nodes to route their packets over optimal paths for longer periods. The proposed routing protocol uses one of the state-of-the-art inference techniques dubbed spatial reasoning that enables each node to figure out the spatial relationships between nodes without overwhelming the network with control packets. As a result, the end-to-end delay is reduced while the throughput and lifetime are increased.

Besides the proposed protocols, this thesis utilises the analytical aspects of statistics to develop rigorous analytical models that can accurately predict the queuing and medium access delay and energy consumption over multihop networks. Moreover, this thesis provides a broader perspective for design of communication protocols for WSNs by casting the operations of these networks in the domains of the artificial chemistry discipline and the harmony search optimisation algorithm.

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I am extremely grateful to my supervisor Dr Paul Mitchell for his incredibly wise advices and motivations and appreciate his invaluable guidance, patience and contributions throughout my research. Sincere thanks to Dr David Pearce who always had the time to propose new ideas to me and discuss my results. In addition, I highly appreciate Dr David Grace who provided me with considerable and help.

Declaration

All work presented in this thesis as original is so, to the best knowledge of the author. References and acknowledgements to other researchers have been given as appropriate. Some of the research presented in this thesis has resulted in a number of publications, listed in the references [1-8].

Chapter 1

Introduction

1-1 Background

This thesis investigates the techniques and issues associated with designing effective Medium Access Control (MAC), routing and duty cycle management protocols for multihop Wireless Sensor Networks (WSNs) using the principle of inferential statistics. The objective is to achieve three concurrent goals: increase the throughput of networks, prolong the lifetime of nodes and minimise end-to-end delay of packets, independently of topology configurations, traffic patterns or routing policies. Success in designing such protocols is of tremendous importance in realising large scale WSNs need for most envisaged communication systems such as the Internet of Things [9], smart grids [10] and ambient intelligence [11].

The concept of incorporating sensor devices into communication networks emerged from the need to interconnect the physical world with the digital world [9]. Although sensor networks have been used in military and industrial applications for more than seven decades, most of these networks were deployed on a small scale and depend on wired communications technologies [12]. The first major research programme of Wireless Sensor Networks (WSNs) was funded in the 1980s by the Defence Advanced Research Projects Agency (DARPA) dubbed Distributed Sensor Network (DSN) [13]. The main aim of DSN was to build a type of Transmission Control Protocol/Internet Protocol (TCP/IP) network between several autonomous sensor devices that were spatially distributed over the Arpanet (the predecessor of the Internet). The DSN programme attracted a cadre of researchers from different academic institutions to cover different aspects, including researchers from Carnegie Mellon University who focused on the development computer operating systems and communication protocols and researchers from the Massachusetts Institute of Technology who investigated knowledge based signal processing aspects [12-13]. The results of the DSN were demonstrated through a number of testbeds such as vehicle monitoring and low-flying aircraft tracking [12]. Although these applications proved the concept of WSNs, the technologies were not quite ready to adopt their development. Data processing was performed using minicomputers, sensor nodes were very large (shoe box or up) and wireless communication capability was at very early stage [12-13]. Notwithstanding, the results of the DSN programme sparked the initial wave of research in the field of WSNs, and following the DSN programme a large number of research projects were sponsored to investigate WSNs [13].

In the 1990s, advances in Very Large Scale Integration (VLSI), Micro Electro Mechanical Systems (MEMS) and Radio Frequency technologies represented a new paradigm in WSNs. The University of California at Los Angeles in collaboration with the Rockwell Science Centre produced the Wireless Integrated Network Sensor (WINS) [14]. WINS demonstrated the ability to integrate sensor arrays, processor boards, radio transceivers and compact battery cells onto a single chip. WINS supported 100kbps with adjustable power management from 1 to 100mW and were packaged in $3.5 \times 3.5 \times 3$ inch^3 enclosures. The ultimate outcome of the WINS was that it geared research endeavours towards adopting low-power techniques in manufacturing sensor nodes; several sensor nodes were fabricated with the aim of minimising the size of WINS, to reduce their energy consumption or to cut their production cost. To this end, the University of California at Berkeley funded an ambitious project called Smart Dust in 1999 [15]. The aim of this project was to develop an inexpensive and autonomous sensor node whose size was of the order of a grain of sand. Although the complete success of the smart dust project has not been achieved yet, its vision constitutes the main motivation of the current applications of WSNs. Since that time WSNs have found their way into a wide range of applications in many disciplines including medical telemetry, disaster alarming, environmental monitoring, and home and industrial automation [13-15].

In the last decade, standards bodies recognised the importance of WSNs and drew up their standards. In 2003, the Institute of Electrical and Electronic Engineers (IEEE) ratified the IEEE 802.15.4 standard [16] that defined the specifications of the physical and MAC layers for low-power personal area networks. This standard not only facilitated the development and proliferation of WSNs commercially but also provided a platform upon which other protocols can be specified. In 2006, the Internet Engineering Task Force (IEFT) released the 6LoWPANs Internet draft [17] that enables transmission of Internet Protocol version 6 (IPv6) datagrams over IEEE 802.15.4 based networks. Integration of IP into WSNs encouraged the International Telecommunications Union (ITU) to propose a paradigm of the Internet of Things (IoT) which aims to incorporate the physical world into business processes by connecting WSNs the Internet [18]. The IoT became the centre of attention at several research institutes, e.g., American National Science Foundation (NFS), European Community of Research and Development Information Service (CORDIS) and the Japanese National Institute of Information and Technology (NICT) [9,18,19]. Extensive research efforts have promoted a plethora of modern approaches, e.g., smart grids [10], Ambient Intelligence (AmI) [11] and cyber-physical systems [20]. The prime aspiration of these approaches is to exploit WSNs to provide a worldwide digital skin over physical environments that make them more informative and convenient for our needs [9,11,19]. From this aspiration emerged a new perspective for WSNs that is global networks of decision makers rather than networks of dummy data collectors.

In the current decade, WSNs face unprecedented challenges that are posed from the need to satisfy two contradictory demands: the large scale deployment driven by potential applications and the miniaturisation of sensor nodes driven by the need to embed sensor nodes into different environments [21]. A typical WSN is an ad-hoc network consists of hundreds or thousands of sensor nodes each of which is equipped with a low-speed processor, a few kilobytes of storage space and a short-range communication transceiver and is powered by a small battery which cannot be replaced in some cases. These characteristics place great constraints on the design space communication protocols of WSNs. Some of these constraints are related to the limited resources of sensor nodes, e.g., the requirements to minimise the computational complexity, storage space and energy expenditure of protocols. Other constraints are associated with the operational conditions of WSNs, e.g., the requirement for scalable and adaptive protocols. Meeting these constraints makes the development of MAC, routing and duty cycle management strategies for WSNs is particularly challenging. The following section highlights these challenges and outlines the main approaches that used to design these protocols.

1-2 Overview of the MAC, routing and duty cycle management protocols for WSNs

The MAC protocol is one of the fundamental mechanisms required to ensure the correct operations of WSNs. Most of the communications of WSNs are carried out over broadcast channels in which a communication between a pair of nodes is susceptible to interference induced by simultaneous transmissions from other nodes that are within the vicinity of the intended receivers [22-23]. The interference has disastrous effects on the performance of the WSNs, as it can reduce throughput and lifetime and magnify end-to-end delay. The key function of the MAC protocols is to regulate the access to the shared channel in such a way that mitigates an effect of interferences. The main challenge associated with devising an effective MAC protocol for multihop WSNs is the need to consider heterogeneous traffic amongst the nodes [4]. In such networks, each sensor node potentially acts as a source, router and final destination simultaneously which in turn makes the traffic load differing amongst nodes. Another main source of heterogeneity of traffic patterns of multihop WSNs is dynamic changes of topology configurations. Most of the WSNs are battery-powered networks [13,22,23], and during operation of these networks some nodes die out due to exhaustion of their batteries. Hence the traffic that was serviced by these inactive nodes is rerouted to other nodes which cause temporal variations of traffic across the WSNs.

Accounting for the heterogeneous traffic characteristics requires an adaptive technique that enables each node to predict the traffic loads and then adapts its behaviour to access to the shared channel dynamically. Some approaches attempt to overcome the aforementioned challenge by allowing nodes to negotiate with each other to derive a schedule for their transmissions [22-23]. In general, the main limitation of employing scheduled based MAC protocols over multihop WSNs is that these protocols consume considerable amount of energy and resources in setting up and maintaining the transmission schedules. An alternative approach used to design MAC protocols is a random access mechanism [22-23]. Basically this mechanism allows a node to compete to access the shared channel based on its traffic intensity and without requiring a prior negotiation, which in turn mitigates the overheads imposed by the scheduled based MAC protocols. However, the key limitation of random access control protocols is that they cannot eliminate interferences completely; but rather reduce the likelihood of their occurrence using a random process [22-23].

Duty cycle management scheme is another important mechanism that is required to prolong the lifetime of sensor nodes by letting them to sleep for a certain period if they have no roles to play during this future period [24-25]. The key challenge of devising an effective duty cycle scheme is to find the effective approach by which a node can determine the best times and durations for its sleeping periods. Letting a node sleep for longer periods can prolong its lifetime but it can reduce the throughput of a network or magnify the end-to-end delay of packets. Conversely, letting a node sleep for shorter periods wastes its energy in doing nothing useful, e.g., overhearing (when a node receives other transmissions that is not destined to it) or idle listening (when a node waits for potential packets). Large-scale deployment of WSNs, their scarce resources and heterogeneous traffic patterns constitute the key challenges of designing effective duty cycle scheme for WSNs.

Most of the duty scheme management schemes presented in the open literature can be classified roughly into three categories: synchronous, asynchronous and geographical [24-25]. Synchronous duty cycle schemes require coordinator nodes (potentially there are more than single coordinators within a network) to broadcast synchronisation packets containing sleeping and waking time information to other nodes. These nodes use this information to manage their duty cycle schemes accordingly. The main limitation of the synchronous approach is that it consumes a substantial amount of energy in periodic transmissions and receptions of synchronisation packets. Asynchronous duty cycle schemes mitigate the overhead associated with synchronisation packets. The underlying approach of asynchronous schemes is to let a node stay asleep most of the time and to be woken only when other nodes wish to communicate with it. Most asynchronous duty cycle schemes assume that a sensor node can be equipped with a radio transceiver for wake up in addition to the radio transceiver that is used to communicate data packets. The main disadvantage of asynchronous schemes is related to the high energy consumption associated with the additional transceiver. Finally, in geographical duty cycle management schemes, nodes exploit their location information to allow a single node within a certain group of nodes to wake whereas other nodes can go to sleep to save their energy. The main limitation of geographical duty cycle management schemes is that they are built on the assumption that sensor nodes can determine their locations; this assumption is not valid in most applications of WSNs as fitting a geographical position system into sensor nodes may increase their production costs, besides the high energy demanded by most geographical systems.

A routing protocol is a mechanism used by nodes to select the optimal routes to the intended final destination(s) [9]. Routing in WSNs is a direct consequence of the small transmission ranges of sensor nodes and the requirement to deploy these networks at a large scale. Fundamentally, the main challenge of designing an effective routing protocol is that determination of the optimal routes requires timely information about the status of a network [13,21-22]. However, obtaining such information can consume a large amount of resources of the tiny resources nodes. In order to overcome this challenge, most of the existing routing protocols trade-off between the freshness of the routing information and the overheads associated with obtaining them. Based on this perspective, routing protocols of WSNs can be classified into five types: flooding, reactive, proactive, geographical and cluster [12,26].

The flooding approach offers a trivial routing mechanism that overwhelms the network with data packets; such a mechanism does not only waste energy and processing time of sensor nodes but also reduces the throughput of network and increases the collision probability of packets. Due to these reasons, flooding routing mechanisms are inapplicable for most of WSN applications. Other routing protocols overcome the main limitations of the flooding approach by allowing nodes to maintain one or more routing paths to destination(s), and forward data packets over a specific path without flooding the whole network with the same packet. Although this procedure can reduce the overhead of flooding protocols, it requires a node to maintain routing paths to destination(s). Dealing with such requirements has been proposed using various techniques. The proactive routing approach requires a node to always maintain a fresh path to other destination(s); the benefit of this approach is that it enables a node to route packets immediately without delay, however, the disadvantage of this approach is that it requires nodes to exchange their routing information periodically which poses overheads on the limited-resource sensor nodes. The reactive routing approach alleviates this overhead by allowing a node to enquire about the route to destination only when a node has a packet to that destination; however, this advantage comes with the penalty that routing a packet can be delayed until all routing information is gathered, which magnifies the end-to-end delay of packets. The cluster routing approach reduces the overhead imposed by the requirement that each node running proactive or reactive routing protocols has to work out the routing paths to all destination(s) by itself. The cluster approaches divides the network into a number of clusters, and each group of nodes within a cluster elects a leader to represent them; these cluster leaders maintain the global routing information whereas other nodes have to communicate with their leaders only. However, the main limitations of employing the cluster approach to WSNs is that this approach may result in bottleneck as all traffic of nodes within a cluster has to be forwarded to a single node. Finally, the geographical routing approach assumes that a sensor node can determine its geographical position and hence nodes can use this information to route their packets. The obvious limitation of employing such an approach to design WSN routing protocols is that most of the geographical position systems consume a considerable amount of energy which in turn hinders the wide applicability of a geographical routing approach over WSN applications. Besides these four categories, a number of routing protocols based on the weak artificial intelligence discipline [8] (e.g., reinforcement learning, swarm intelligence, genetic algorithm) have been proposed to enhance the performance of classical routing protocols. However, the main limitation of these approaches is that their further developments are restricted by the underlying metaphors made to mimic the behaviour of mindless species.

In conclusion, it can be seen that most of approaches used to design MAC, duty cycle management and routing protocols suffer from one or more shortcomings e.g., the dependence on message passing between nodes, the ignorance of the distinct characteristics of each node or the need to fit nodes with additional equipment. These shortcomings restrict the wide applicability of these approaches and hence there is a strong need for a novel design approach that can account for general cases of WSNs. This thesis proposes such an approach, the fundamental logic of approach which is that WSNs can be considered as a stack of random processes. Thereby, inferential statistics is the natural means to design WSN communication protocols. The next section overviews the inferential statistics and highlights its key advantages in the context of designing communication protocols for WSNs.

1-3 Overview of inferential statistics

Statistics is a branch of science that investigates reliable methods and techniques required to collect and manipulate data, draw conclusions from them and exploit the conclusions to make sound decisions and meaningful predictions [27]. Statistics has a long and colourful history, beginning from the second half of the 17^{th} century when William Petty introduced the political arithmetic approach. This approach aimed to collect data about different aspects related to public charges (e.g., defence forces, population, housing, harbours and banks) and manipulated them in order to enact stable laws based on real data [27-28]. Soon afterward, this approach had received substantial attention from leading researchers e.g., Poisson and Laplace who empowered the political arithmetic with the principles of the calculus of probabilities [29]. During the 18^{th} century, a number of statistical and probability theorems were innovated, such as the law of large numbers, the normal distribution, inductive reasoning and Bayes' theorem. By the end of the 18^{th} century, the term statistics was used to refer to the collections of endeavours that had been made in the area of the political arithmetic. From the beginning of the 19^{th} century onwards, statistics has become a multidisciplinary research field with many theorems and applications [27-29].

In general there are two main subfields of statistics: descriptive statistics and inferential statistics [30]. The descriptive statistics subfield involves procedures and techniques for interpreting, summarising and presenting important characteristics of observations. Descriptive statistics uses numerical methods (e.g., average, mode, median, variance, correlation and kurtosis) to describe the data at hand without consideration for further generalisation beyond the data. Inferential statistics, on the other hand, is defined as moving beyond the data in hand to draw conclusions about the underlying process that generate these observations and to extrapolate the future behaviours of this process. Inferential statistics has deep roots in probability theory and logic reasoning [30, 31]. Probability theory is used to mitigate the effects of uncertainty and variations of the observed data, and logic reasoning facilitates making sensible decisions and accurate predictions. Considering these strong aspects of inferential statistics reveals the reason behind its wide applicability in diverse disciplines such as medical, biological, social sciences, economics, finance, marketing research, manufacturing, management, engineering, computing and even jurisprudence [27-31].

Inferential statistics is suitable for WSNs from several aspects, considering the fact that the ultimate objective of modern WSNs applications is to build smart environments in which each sensor node makes decisions [9-11]. Most of these decisions have to be made based on the incomplete and random information that is gathered from the monitored areas; moreover, this gathered information is contaminated with different levels of uncertainties. From the perspective of inferential statistics, the gathered information constitutes the observed data and actions that are taken by a node constitute the conclusions that are drawn based on this information. Hence, the key role of a sensor node is to infer the correct information amongst uncertain and ambiguous observations and make the sound action. The principles of inferential statistics have been used to design communication protocols for wireless networks. A clear example is the Carrier Sensing Multiple Access (CSMA) mechanism in which a node commences its transmission based on its conclusion about the channel conditions. In particular, a node using carrier sensing to infer whether the channel is idle or busy and based on the conclusion of this assessment a node takes its action either to commence or to refrain transmission [13-22]. Additionally, using inferential statistics to design communication protocols for WSNs can offer many advantages; some of them are summarised in the following:

1. Inferential statistics is built on mathematical theorems that have been investigated and developed over several centuries. These theorems provide a powerful means to reveal the underlying characteristics of random process from different aspects and under different conditions. The importance of this approach in designing WSN communication protocols is that it facilitates revealing the relationships between the underlying processes and behavioural responses of nodes. Hence a node can adapt its responses not only to the instantaneous actions of other nodes but also to the main causes of these actions which in turn accelerates approaching to the optimal interacting procedures.
2. Inferential statistics exploits incomplete information of the observed data to draw conclusions about the unobserved data of interacting systems. Thereby the amount of data required by the inferential statistics protocols is small and does not represent complete characteristics of the system. This aspect is highly desirable in the design of communication protocols for WSNs, as each node can

make a sound decision based on its narrow view for the network and without a need to impose high communication overheads on the network. This in turn can leverage the performance of the network significantly by increasing its channel utilisation and prolonging the lifetime of nodes. Furthermore, considering the fact that the inferential statistics techniques exploit sample observations to make decisions demonstrates the reduction in computational complexity and the savings in storage space that can be gained from using inferential statistics.

3. Inferential statistics accounts for uncertainties in the observational data and mitigates their effects on decision making. This is a very important aspect in designing WSN communication protocols since most of the data and interacting processes of WSNs are contaminated with uncertainties. Hence inferential statistics enables sensor nodes to come to sound decisions regardless of the associated uncertainties.
4. Inferential statistics is a multidisciplinary discipline which facilitates exploiting other disciplines as an inspirational source for developing communication protocols.

1-4 Main contributions of the thesis

The main thrust of this thesis is to propose effective design principles and methodologies that can be used to devise communication protocols for multihop WSNs based on the inferential statistic. These design principles and methodologies are employed to develop MAC, duty cycle management and routing protocols for WSNs that can increase the throughput of the network, prolong the lifetime of nodes and reduce the end-to-end delay of packets. It is worth noting that although these proposed protocols can improve the three aforementioned performance metrics compared to some widely used protocols, this does not mean that the proposed protocol is always able to optimise all three metrics simultaneously. These protocols exploit the dependency between these metrics in order to improve them; for instance, the proposed MAC protocols reduce the collision probability which in turn enables nodes to save retransmission energy (thereby, prolonging the lifetime of nodes) and improves the channel utilisation (thereby, improving the network throughput). The proposed protocols consider a general multihop network scenario without particular consideration for specific topology configurations, traffic patterns or routing policies. Another main contribution of this thesis is that it exploits probability and statistical theorems to evaluate the statistical characteristics of queuing and medium access delay of WSNs as well as their energy consumption. Moreover, this thesis proposes a number of multidisciplinary metaphors for WSNs and uses them in developing WSN protocols.

The first main contribution of this thesis is to develop a rigorous analytical framework that can predict the statistical characteristics of queuing and medium access delay and to use this framework to evaluate the performance of the IEEE 802.15.4 CSMA-CA (Carrier Sensing Multiple Access with Collision Avoidance) protocol over multihop networks [6]. This contribution is motivated by the fact that WSNs are large scale networks with low data rate, hence the performance metrics of the network (e.g., end-to-end delays, throughput and energy) are dominated by the queuing and medium access delay at each node. Another primary motivation of using statistical theorems in the proposed analytical framework is emerged from the ability of these theorems to derive the complete probability distributions of the inter-arrival, inter-departure, service and queuing times of packets. This in turn facilitates obtaining meaningful insights into the behaviours of medium access delay under different operational conditions. A justification for using IEEE 802.15.4 CSMA-CA protocol to exemplify the proposed analytical framework can be acquired by considering that this protocol is the first official standard that was specified for WSNs and has been used widely in both industrial and academic disciplines.

Design an effective MAC protocol for WSNs is the second main contribution of this thesis [4]. This contribution is motivated by the performance limitations of the IEEE 802.15.4 CSMA-CA protocol and the need to overcome them. In particular, the results of the analytical framework as well as the simulation outcomes demonstrate that this protocol provides a poor performance over multihop networks. Furthermore, a thorough investigation shows that this performance degradation is due mainly to the poor design space of the mechanisms specified in this protocol, e.g., Binary Exponential Back-off (BEB) algorithm and collision resolution policy. Instigated by this limitation the proposed CSMA-CA protocol replaces the BEB with a more efficient mechanism based on Gamma distribution and devises an intelligent collision resolution policy. The Gamma distribution has been selected due to its appealing statistical characteristics, as this distribution is the conjugate prior distribution for most of the probability distributions that are used to model the traffic in communication networks. Fundamentally, the proposed protocol develops an inferential statistics algorithm that allows each node to adjust the parameters of the Gamma distribution according to its prediction for the contention behaviour of other nodes. The objective of this adjustment is to achieve the three objectives stated in this thesis, i.e., reduce the end-to-end delay of packets, increase the throughput of the networks and prolong the lifetime of nodes. The proposed protocol reduces the end-to-end delay by minimising the service time of packets and mitigating the high level of variation of the BEB. Minimising the service time leverages the channel utilisation and reduces the energy consumed in contention to access the channel.

The third main contribution is to derive an effective duty cycle management scheme for WSNs that can achieve the three goals of this thesis [7]. The main motivation behind this contribution is the lack of a framework that can account for the dynamic interaction between nodes while being able to provide a means of measuring the level of optimisations. This thesis overcomes this limitation by exploiting the metaphors between wireless communication networks and chemical reaction networks to draw the analogies between these disciplines. Based on these analogies, the analytical aspects of the Artificial Chemistry (AC) field is used to predict the energy consumption of WSNs accurately. Moreover, the inferential statistics aspects of AC are used to devise the required duty cycle management scheme. The underlying approach of the proposed scheme is to enable each node to determine its sleeping time and duration using one of the state-of-the-art inferential statistics algorithms dubbed non-parametric Bayesian inference. The key benefits of this approach are that it is not limited to specific network configurations and it considers the distinct requirements of each node without posing overheads on the network. Thereby, lifetime of nodes can be prolonged without degradation of throughput and end-to-end delay.

Developing an effective routing protocol for WSNs is the final main contribution of this thesis [8]. This contribution is motivated by the need to overcome the main limitations associated with pertinent approaches that are used to develop routing protocols. This thesis draws analogies between WSNs and the improvisation process conducted by a musician ensemble during their endeavours to produce the most pleasing harmony. Based on these metaphors, this thesis utilises the Harmony Search (HS) optimisation algorithm to devise the routing protocol. Basically, the proposed protocol employs one of the most powerful inference techniques called spatial reasoning to enable a node to work out the spatial relationships between nodes and the divergent thinking to enables nodes to infer the routing metrics. Hence a node can find the optimal or near optimal route without overwhelming the network with control packets which in turn achieves the three goals of this thesis, i.e., increase the throughput of network, prolong the lifetime of nodes and reduce the end-to-end delay of packets.

1-5 Structure of the thesis

The main body of the thesis is divided into eight chapters whose contents are summarised in this section. Figure 1.1 illustrates the structure of the thesis as well as the relationships between different chapters.

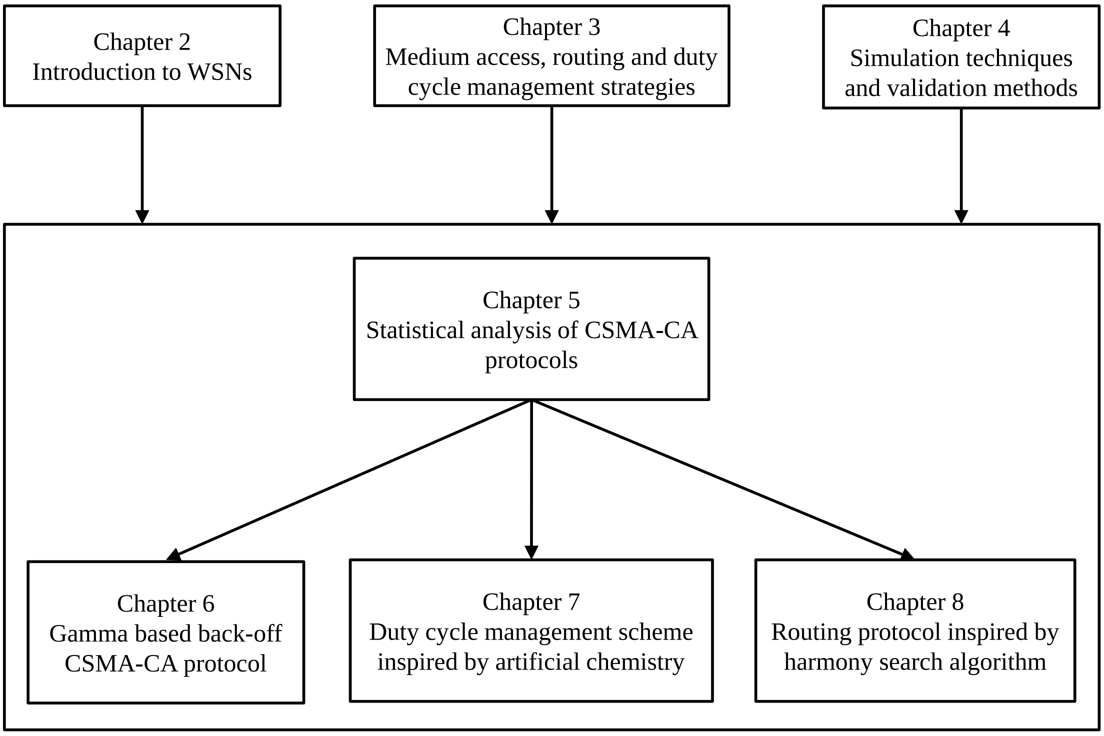


Figure 1.1 Structure of the main body of the thesis

Chapter 2 introduces wireless sensor networks and their fundamental aspects including hardware and software platforms and then overviews the basic architecture of WSNs. Some current and envisaged applications of WSNs are also considered in this chapter. The main objective of this chapter is to highlight the limited resources of WSNs and their impact on the design space of WSN protocols.

Chapter 3 provides an extensive literature review on MAC, duty cycle and routing as well as their cross layer protocols with the aim to identify their key roles in WSNs. This chapter exemplifies these roles by discussing the main contributions and limitations of some protocols. Based on this discussion, chapter 3 highlights the motivations for using inferential statistics to design WSNs protocols.

Chapter 4 presents the simulation techniques and validation methodologies employed in this thesis. This chapter overviews some of the simulation platforms used to assess the performance of WSNs and then provides a detailed review for the chosen simulator, NS3. Furthermore, chapter 4 presents the general simulation parameters that are used to assess the performance of the proposed protocols.

Chapter 5 presents the first main contribution of this thesis which is a novel analytical framework that can analyse the statistical characteristics of queuing and medium access delay of WSNs over multihop networks [6]. This chapter provides a detailed description for the proposed framework and then uses it to evaluate the performance of the IEEE 802.15.4 CSMA-CA protocols. Moreover, this chapter assesses the integrity of the proposed framework by comparing its outcomes with simulation results under different scenarios. This chapter then exploits these results to highlight the main limitation of this protocol and to provide design guidelines for the proposed MAC protocol.

Chapter 6 provides the second main contribution of this thesis, a novel CSMA-CA protocol based on the Gamma distribution [4]. This chapter overviews the basic mechanisms of the existing CSMA-CA protocols, outlines their main shortcomings and reviews the statistical characterises of the Gamma distribution. Based on this discussion, this chapter presents the design principles of the proposed protocol and provides its pseudo-code. Comprehensive assessments for the performance of the proposed protocol compared to the IEEE 802.15.4 CSMA-CA protocol are also given in this chapter.

Chapter 7 provides the third main contribution of this thesis which is a novel duty cycle management scheme for WSNs inspired by the principle of Artificial Chemistry (AC) [7]. This chapter starts with an overview of the fundamental principles of chemical processes and then draws analogies between the reactions of chemical species and the communications of wireless nodes. Based on this metaphor, this chapter utilises the modelling aspect of AC, in particular, molecular statistical mechanics to derive the time evolution of the energy consumption in multihop networks. It then employs the computational algorithm aspect of AC to design the effective duty cycle management scheme. The pseudo-code for the proposed protocol is given in this chapter. Assessment for the integrity of the modelling approach as well as the benefits of the proposed protocol is presented in this chapter.

Chapter 8 provides the fourth main contribution of this thesis by developing an effective routing protocol for WSNs inspired by an improvisation process [8]. This chapter conceptualises a routing protocol as a solution of a multi-objective optimisation problem and then exploits this concept to discuss the fundamental logic of well-known approaches used to design routing protocols for WSNs. Chapter 8 then exploits this discussion to draw the future trends in designing routing protocols and highlights the main benefits of using the optimisation algorithms that are based on meta-heuristics approaches to design routing protocols. Based on these findings, this chapter introduces the Harmony Search (HS) algorithm to derive the proposed routing protocol. The pseudo-code of the proposed protocol as well as an assessment for its performance compared to the IEEE 802.15.5 and an ant colony based protocol is given in this chapter.

Chapter 9 provides some ideas for the future work that can be carried out based on the contributions of this thesis and finally chapter 10 concludes this thesis.

Chapter 2

Introduction to Wireless Sensor Networks

2-1 Introduction

Wireless Sensor Networks (WSNs) have enabled interconnections between the physical worlds and the digital worlds. This interconnection has found wide range of applications in vast areas and it is expected that this proliferation will increase [9-11]. The ultimate goal of WSNs is to build a worldwide “digital skin” over the physical environment; this goal requires fabrication of tiny sensor nodes that can be embedded in almost every item of our everyday life. While the advances in related technologies (e.g., Very Large Scale Integration (VLSI), Micro Electro Mechanical Systems (MEMS) and Radio Frequency technologies) have led to miniaturisation of sensor nodes, this miniaturisation and the requirement to build large scale networks imposes great challenges on the design space of WSNs [21].

The main objective of this thesis is to develop medium access control, duty cycle management and routing strategies for WSNs using the principle of inferential statistics. The first step towards achieving this objective is to identify the fundamental characteristics of WSNs and to discuss the design space of protocols in light of these characteristics. This chapter is devoted to this purpose; it starts by providing a conceptual structure of a typical sensor node which demonstrates its elementary components and the main functions of each component. This demonstration is then utilised to review some state-of-the-art sensor nodes and to provide a feeling for the amount of resources that can be accumulated within sensor nodes. Applications of WSNs are the second important characteristics that need to be considered before discussion of the design space of communication protocols, and this chapter explores some of current and envisaged applications of WSNs. These examples illustrate the wide range of applications that can be constructed based on WSNs. In the light of the limited resources of sensor nodes and the large-scale demands, this chapter discusses the design space of WSNs communication protocols. This discussion identifies the main factors that should be taken into account in the design phase and highlights the random and interacting processes that dominate communications within WSNs.

The remainder of this chapter is organised as follows. Section 2-1 introduces the conceptual structure of both of hardware and software platforms of a typical sensor node. Section 2-2 overviews some examples for the state-of-the-art sensor nodes and section 2-3 summarises some of the modern and envisaged applications for WSNs. In section 2-4, the design space of WSNs communication protocols is discussed and section 2-5 concludes this chapter.

2-2 Overview of sensor node hardware and software platforms

This section overviews the conceptual structure of the hardware and software of a typical sensor node. It starts by considering the hardware platform in section 2-2-1 and then the software platforms in section 2-2-2.

2-2-1 Hardware platforms of sensor nodes

A typical wireless sensor node consists of five main components: processor, sensing devices, transceiver, memory and power supply. A conceptual diagram for the architecture of wireless sensor node is illustrated in figure 2.1.

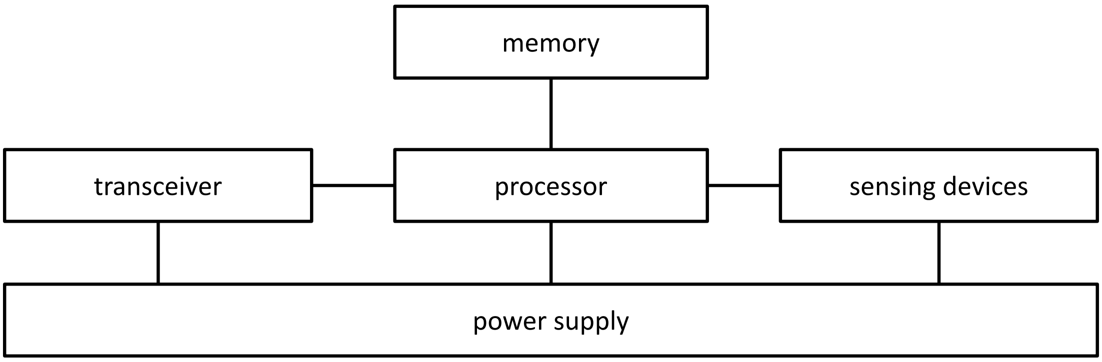


Figure 2.1 Conceptual diagram for the architecture of wireless sensor node

The main function of the sensing component is to gather information about phenomena of interest from the physical environment and to convert them into a form that can be maintained by the processor of the sensor node [13,22]. Basically, the sensing component consists of one or more transducers which convert one form of energy into electrical energy. Since the typical outputs of transducers are analogue signals, Analogue to Digital Converters (ADCs) are used to convert the analogue signals into their digital equivalents [13,22]. Most current sensor nodes are equipped with a sensor array that is able to sample different phenomena such as temperature, sound, vibration, humidity and/or pressure.

The second main subsystem of a sensor node is the processing component. The main aim of the processing unit is to control the functionality of other components and process their data. Different types of processor have been used in sensor nodes including microcontrollers, Digital Signal Processors (DSPs), Field Programmable Gate Arrays (FPGAs) and Application Specific Integrated Circuits (ASICs) [32]. Most existing sensor nodes use microcontrollers as this class of processors provides flexible programming platforms with built-in memory at low production cost. However, microcontrollers are not as powerful as DSPs, ASICs or FPGAs either in terms of energy saving capability or computation complexity. DSPs, on the other hand, can perform complex data processing with high efficiency; this characteristic makes it suitable for some WSN applications such as motion detection. However, the main limitations of DSPs are high energy consumption, large size and high production cost. FPGAs and ASICs offer trade-offs between the advantages and disadvantages of the microcontroller and DSP processors. FPGAs alleviate both the low speed of the microcontrollers and the high energy consumption of DSPs. However FPGA processors require reconfiguration in the field whenever software updating is required which in turn consumes time and effort. ASICs provide the best energy saving profile compared to other processors; however, this advantage is associated with high development costs [13-32].

The third main component of a sensor node is the storage subsystem. A typical sensor node stores a large volume of data including the operating system, communication protocol codes, sensory data and the packets that are either generated by the node itself or received from other nodes [32-33]. Storing and processing this data within a tiny sensor node requires an effective data management subsystem. Generally, most of the processing subsystems used in sensor nodes provide limited storage space of the order of a few kilobytes of RAM to store data and of the order of a few tens of kilobytes of ROM to store programme instructions. The main reason for the limited storage capacity offered by the processing component is that increasing the storage space inside the processing unit consumes most of its chip area which in turn reduces the capability of the processor and increases its energy consumption. A remediation for the limitation of on-chip storage space is carried out by embedding a flash memory into a sensor node. Flash memory is a non-volatile Electrically Erasable Programmable Read Only Memory (EEPROM) which can access n-byte tuples in a single operation. Although the size and prices of flash memory are suitable for sensor nodes, the time required to access the contents of these memories as well as the energy budgets may result in reducing the lifetime of nodes significantly [33].

The main responsibility of the power supply subsystem is to feed the whole sensor node with power, hence the power supply is the paramount component in a node. Most of the current sensor nodes are powered by batteries and in some situations replacing these batteries is impractical, therefore the lifetime of a WSN depends on a limited and non-renewable resource. The main challenge of using batteries to power sensor nodes is that most of the existing batteries provide light energy densities, In particular, the energy densities of Alkaline, Lithium and Zinc-air batteries are 1.2, 2.8 and 3.7kilojoule/cm^3 respectively [22]. Comparing these parameters with the size of current sensor nodes (of the order of few cubic millimetres) demonstrates the difficulty of deriving abundant power for sensor nodes from batteries. Other possibilities to supply sensor nodes with power have been proposed e.g., micro-fuel cells, energy harvesting and super capacitor [13-22]. However these technologies are in early stages of developments.

The last main component of a sensor node is the transceiver which provides communication capabilities for sensor nodes. A transceiver is a circuit that comprises both a transmitter and a receiver. The transceiver is the most power consuming component in the sensor node. Compared to the power consumption of the processing component, transmitting one bit expends about as much power as executing several hundred instructions [13,22]. Due to this large power consumption, a node saves energy by switching the transceiver between different states. In general there are four states: transmit, receive, idle and shutdown. The transmit state represents the case when a node is transmitting its packets; similarly the receive state represents the case when a node is receiving signals. In the idle state, some part of the transceiver is functioning, however, it is not in transmit or receive states. Therefore, a node can switchover to receive state as soon as it hears a signal whose power is greater than or equal to its receiving sensitivity threshold. Finally, in the shutdown state, the transceiver is turned off to save energy. It is worth noting that most of the current transceivers provide various transmission power levels which facilities controlling the range of receivers.

2-2-2 Software platform of sensor nodes

From a software perspective, the core of a wireless sensor node is the Operating System (OS). The main functions of an OS are to manage the hardware resources (e.g., processor management, memory allocations and multithreading scheduling) and to provide unified application programming interfaces [10, 34]. Traditionally, there are two design approaches for the OS of WSNs: event-driven and multithreaded. An event-driven OS provides a set of event handlers each of which is fired in response to an event e.g., reception of a signal or availability of sensory data. The event-driven system is a single task OS; the processor cannot proceed to the next handler unless the previous one has been completed. A well-known example of the event-driven OS is TinyOS [35]. A multithreaded OS, on the other hand, allows a processor to execute multiple tasks (threads) concurrently. In a multithreaded-driven OS, each thread is allocated a portion of processing time a so-called quantum, and a quantum of the current thread is elapsed a processor saves the current context and executes the next thread after retrieving its context. Some examples for the multithreaded OS are RETOS [36] and Mantis [37]. The multithreaded OS can improve the responsiveness of a sensor node by allowing multiple threads to be executed in parallel, e.g., a node can receive a packet from the communication subsystem while instructing the sensing subsystem to sample the physical environment. However, this advantage is traded-off against the energy and resources required to store and retrieve the contexts of each thread.

2-3 Sensor nodes examples

Since the introduction of the concept of the WSNs, a large number of development nodes have been fabricated to suit the vast and diverse applications of these networks. In general, there are two main techniques to fabricate sensor nodes: Commercial Off-The-Shelf (COTS) and System on Chip (SoC) [38]. The main approach of the COTS technique is to assemble a sensor node using commercially available components. COTS provides many advantages such as shortening design-to-production cycles and supports backward compatible with legacy products. Moreover, the COTS technique increases the responsiveness of nodes to inevitable change and reduces the maintenance bill. However, the main limitation of COTS design approach is that the further development of sensor nodes depends on market trends. On the other hand, the SoC approach attempts to integrate all components of sensor nodes (e.g., sensor, processing and transceiver units) into a single chip. The key design objectives of the SoC are to miniaturise the size of sensor nodes by exploiting unused space of different components and to reduce the production cost by minimising the post-assembly requirements. Therefore, the SoC technique provides more reliable products as the integrations between different subsystems are considered carefully at the design phase. However, the main drawback of SoC is that it requires long development time. This section demonstrates the COTS and SoC design approaches by providing some examples of state-of-the-art sensor nodes in the following two subsections [38,39].

2-3-1 Examples of commercial off-the-shelf sensor nodes

This section considers two examples for COTS sensor nodes: the Mica [40] and Medusa [41] sensor families.

The Mica sensor nodes are a family of sensor nodes also known as Mica motes. This family started in the late of 1990s as a spin-off of research projects at the University of California at Berkeley [40]. The Mica family comprises a number of motes, e.g., WeC, Mica, Mica2 and Telos. The WeC [42] mote was produced in 1998. This mote featured an Atmel AT90LS8535 microcontroller [43] with 4MHz clock speed. This microcontroller provided two main operational states: an active state which consumed 15mW and a sleep state which consumed 45µW. The WeC mote used the TR1000 transceiver [44] that supported a 10kbps data rate with 36mW transmission power at 0dBm, with 9mW receiving power. The storage component of WeC comprised 0.5kbytes RAM, 8kbytes ROM and 32kbytes flash memory.

Although the WeC motes provided powerful processing and communication capabilities, the research efforts that had been conducted yielded even smaller and lower power nodes. In 2001, the second generation of the Mica motes was introduced. This generation included a number of versions (e.g., Mica, Mica2, Mica2Dot and MicaZ) [40]. The Mica motes enhanced the capabilities of processing and communication units of the WeC motes. The Mica motes replaced the Atmel AT90LS8535 processor of the WeC node with more powerful processing unit, the Atmel Atmega103L microcontroller that increased the RAM to 4kbytes and the ROM capacity to 128kbytes. Moreover this new processor reduced the active power by more than a 50% compared to the microcontroller used in the WeC mote. In terms of the communication capabilities, the Mica motes increased the data rate by four times compared to the WeC mote.

In 2002, the Mica2 motes were fabricated [45]. These motes were equipped with a more powerful processor, the Atmel Atmega128L [46] which provided up to 7MHz clock speed instead of 4MHz processor of the Mica motes. The Mica2 motes used the Chipcon CC1000 [47] transceiver module from Texas Instruments Inc. The key advantage of this transceiver is that it features a frequency selective circuit that provides a range of frequencies.

One year later, the MicaZ [45] mote was produced. This mote supports the IEEE 802.15.4 standard and provides a data rate up 250kbps with on chip data encryption and authentication policies. In 2004, the Telos mote [48] has been realised. This mote employs ultralow-power and a powerful microcontroller TIMSP430 which operates at 8MHz with 3mW active power and 15µW sleep power. The Telos also incorporates powerful transceiver modules, CC2420 that supports an enhanced modulation scheme O-QPSK with a data rate of 250kbps. Furthermore, the Telos mote provides a number of enhancements compared to the previous versions e.g., built-in printed circuit board antenna which reduces the cost of production of sensor nodes and a built-in USB port to extend interfacing between motes and PCs. Figure 2.2 illustrates some examples of the Mica mote family.

WeC Mica2 Telos

Figure 2.2 Some examples for the mote family adopted from [48]

Another example for the COTS sensor family is dubbed Medusa. This family has been developed by the University of California at Los Angeles since 2002 [41]. The design philosophy of this family is to provide a powerful sensor node that is suitable for high computation and large storage space applications. Amongst the existing family, Medusa MK-2 has been widely used either as a front-end sensor or base station [49]. The Medusa MK-2 node is equipped with two processors: the first one is 8bit, 4MHz Atmel ATmega128L microcontroller with 4kbytes of RAM and 32kbytes of flash memory. The main function of this process is to manage the activities of sensor and radio transceiver modules. The second processor is 16/32bit, 40MHz, ARM7TDMI microcontroller with 36kB of RAM and 1MB flash memory package. The main function of the latter processor is to perform sophisticated operations such as signal processing. Interaction between these two processors is carried out over a pair of interrupt lines and a serial peripheral interface bus. The communication component of Medusa MK-2 is based on the TR3000 transceiver module, which provides two distinct modulation schemes: On-OFF Keying (OOK) and Amplitude Shift Keying (ASK). Moreover, this transceiver provides a wide range of data rates ranging from 2.4kbps to 115kbps. This node provides MEMS accelerometer with a rich set of interfaces including eight 10bits ADC, serial ports and general purpose Input/Output ports. Moreover, this node is equipped with 40kHz ultrasonic transducers that are used to estimate distances between nodes within the network. Figure 2.3 depicts the Medusa MK-2 sensor node.

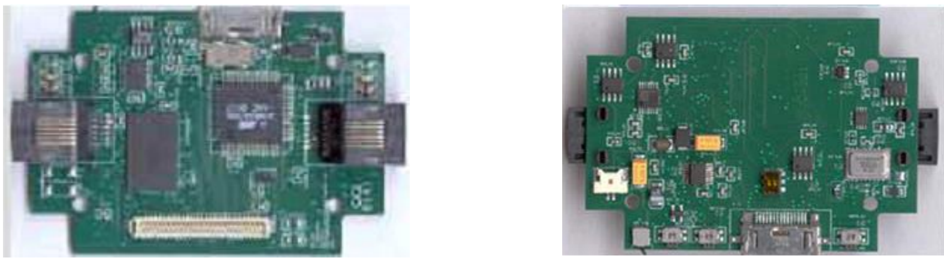


Figure 2.3 Medusa MK-2 sensor node adopted from [49]

2-3-2 Examples of the system on chip sensor nodes

This section considers two examples for System on Chip (SoC) sensor nodes: the Spec [50] and Pico [51] sensor node family.

The Spec mote is a SoC node that has been produced at the University of California at Berkeley [50]. This mote integrated a RISC based processor, 3kbytes RAM, micro-radio, an analogue-to-digital convertor, a temperature sensor as well as the TinyOS operating system onto a tiny chip measured a mere $5mm^2$. In order to enable this mote to perform sophisticated operations demanded by the applications of the smart dust project, the Spec node provided hardware accelerators that facilitated running the kernel of TinyOS more efficiently over the stringent resources of Spec. Furthermore, these accelerators allowed hardware data encryption which is much more efficient than software encryption used in other platforms. Another main advantage of the Spec mote is its low fabrication cost which is about $0.30 per a piece. Indeed, the design philosophy of Spec was to reduce the production cost by reducing the requirement of the post-assembly. Figure 2.4 illustrates the Spec mote.

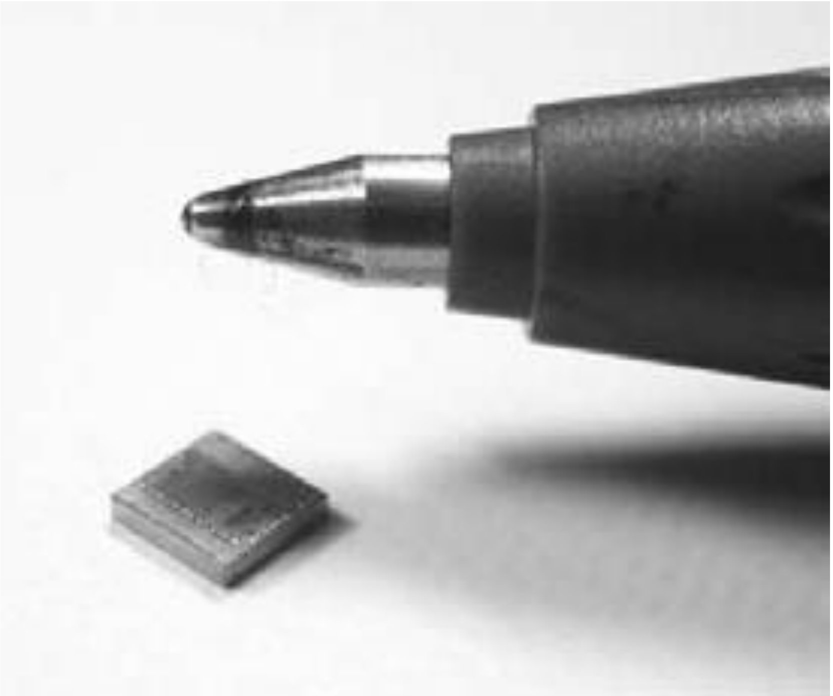
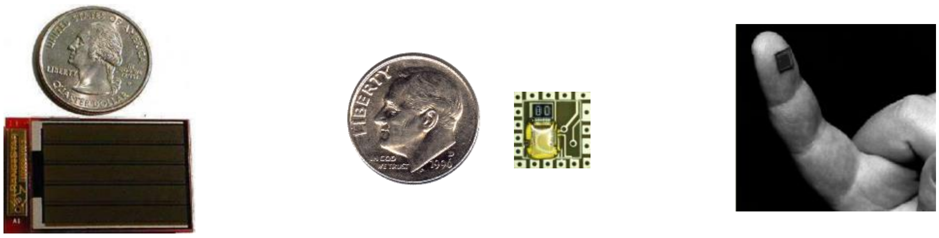


Figure 2.4 Spec mote adopted from [50]

Another example for the SoC based sensor node is Pico [51] sensor node family; this family was developed at the Berkeley Wireless Research Centre with the aim of producing a sensor node that is powered by energy scavenging techniques instead of batteries. The PicoBeacon [13] was the first transceiver that used solar and vibration energy sources instead of battery. This transceiver module was able to operate at unity duty cycle using lighting sources and at 0.26 using ambient vibration sources. Success in producing PicoBeacon paves the path to develop a number of sensor nodes that are powered purely by energy scavenging techniques. Toward this goal the PicoCube [52] sensor node has been developed by the same research group in 2008. The PicoCube employs the System on Chip (SoC) technology to fabricate a $1cm^3$ sensor node operating at an average 6µW. The PicoCube sensor node uses the TI MSP430F1222 microcontroller and two sensor boards that can measure pressure, temperature, acceleration, and supply voltage. This node uses a 0.8dbm transmitter based on Film Bulk Acoustic Resonator (FBAR) technology and powered by 0.7V battery. Figure 2.5 depicts some examples for the Pico sensor node family, figure 2.5(a) illustrates the Pico Beacon, figure 2.5(b) illustrates the PicoCube and figure 2.5(c) shows an early prototype for PicoSystem node.



PicoBeacon PicoCube Early prototype for the PicoSystem

Figure 2.5 Some examples for the PicoRaio family adopted from [53]

2-4 Overview of WSN applications

The concept of WSNs has attracted a huge number of applications in diverse areas including but not limited to environmental monitoring, logistic management, meter reading automation, medical telemetry, disaster alarming, energy management, industrial and building automation, ambiance controlling, animal populations monitoring and water supply protection. This section reviews some of these applications in selected areas while a comprehensive survey for these applications and many others can be found in [13-15,21-22].

2-4-1 Military applications

The concept of WSNs like many other technologies was developed in the context of military applications. The first known sensor network was the Sound Surveillance System (SOSUS) which developed in the late 1950s [12]. SOSUS deployed a set of acoustic sensor nodes in the Atlantic and Pacific oceans for submarine surveillance. The Cooperative Engagement Capability (CEC) [54] is another example for the applications of sensor networks in the military field. CEC allowed ships, aircraft and radars to share their sensory information together to create a detailed and consistent picture for battlefields. The Air Delivered Seismic Intrusion Detector (ADSID) system [55] was developed by the United States Air Force during the Vietnam War to detect vibrations of moving personal and vehicles. Each node of the ADSID system was equipped with a seismometer and a radio transmitter. The data gathered from each node was transmitted to aeroplanes via the unique frequency. The weight of each node was about 38 pounds and its dimension was about 48 inches in length by 9 inches in diameter. The ADSID system employed the ALOHA protocol [56] as a medium access control protocol, ALOHA had already been developed by the University of Hawaii at that time.

In additional to the applications that utilise WSNs in the monitoring and tracking of enemies, some recent applications utilise these networks to enhance the combat efficiency of soldiers. Towards this goal, the Defense Threat Reduction Agency has funded the intelligent clothing for rapid response to aid wounded soldiers project [57]. The main objective of this project is to develop an intelligent uniform such that a set of sensor nodes are woven into its fibres. These nodes assess the overall health of soldiers and alter them based on any maladies or exposure hazards (e.g., chemical, biological, radiological, nuclear and explosive). Moreover, these intelligent uniforms are able to collect detailed information about the wounded soldiers (e.g., location of bullet or shrapnel, depth of penetration and the effected organs) and communicate this information to emergency responders. Figure 2.6 shows some examples of intelligent uniforms.

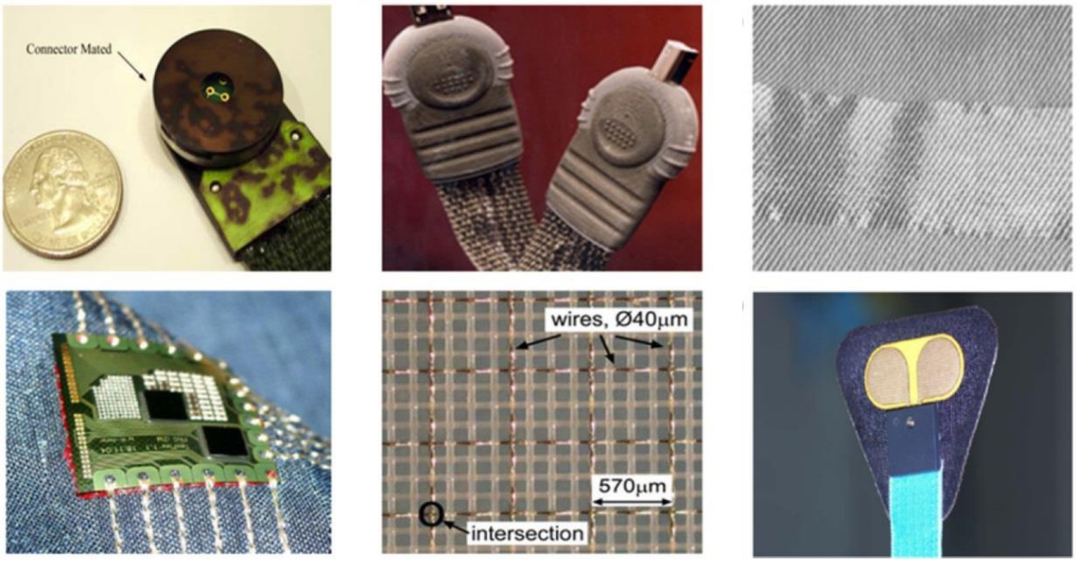


Figure 2.6 Examples of intelligent uniforms adopted from [57]

2-4-2 Smart home

The smart home [58] applications aim to exploit sensor nodes that can be fitted into artefacts and items in homes (such as an oven, washing machine, taps, beds and air conditioning units) in order to make these devices informative and convenient for the needs of the occupants. Additional benefits of smart home applications are improvement of safety measurements by preventing potential hazards from household devices, reduction of the maintenance costs of devices and minimisation of the energy consumption of these devices by turning appliances off whenever they detect that there is no further need for them.

One of the leading research projects that investigate the smart home approach is the Gator Tech Smart House (GTSH) [59] that has been developed by University of Florida's Mobile and Pervasive Computing Laboratory. The main aim of this project is to create assistive homes that can sense themselves as well as their residents in order to minimise human intervention. GTSH specifies the functions of the each component of a smart home, such as a smart mailbox that notifies residents whenever a new mail arrives, a smart bed that adapts itself according to the patterns of a sleeper and a smart floor that tracks the location of occupants, detects an injured occupant and reports data to the emergency services. A Conceptual diagram for the GTSH project is shown in figure 2.7.

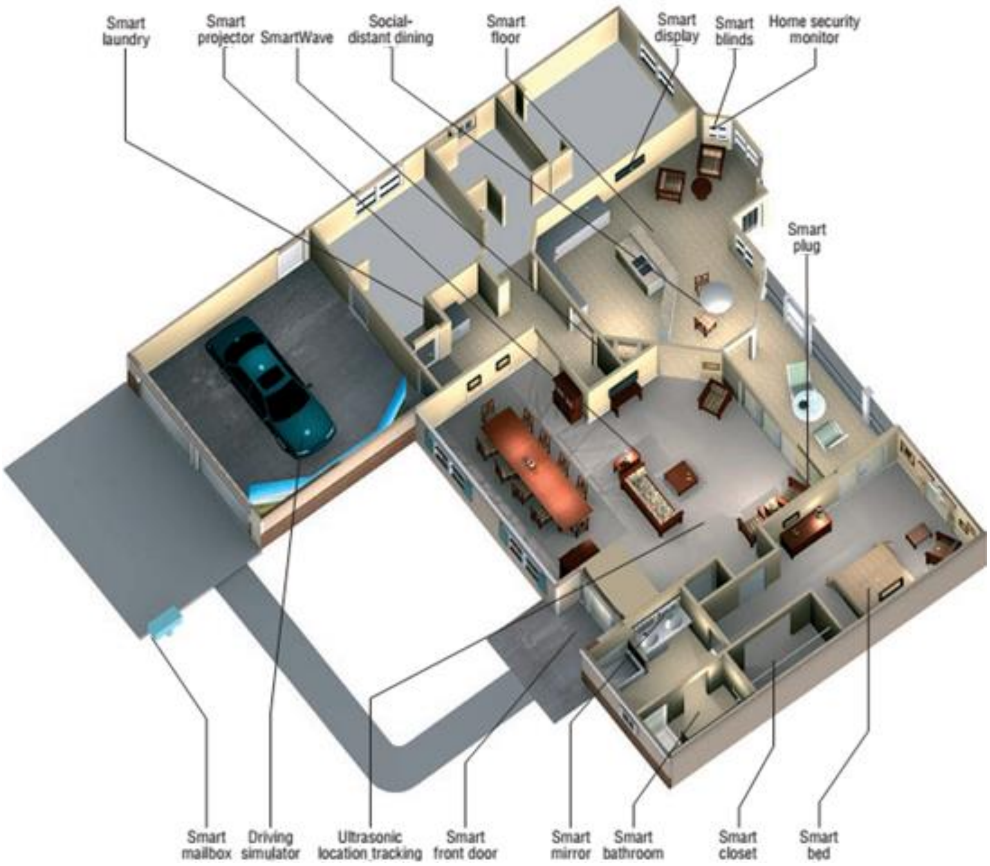


Figure 2.7 Conceptual diagram for the GTSH smart home adopted from [59]

Another example of the research efforts in the domain of smart homes is dubbed Managing An Intelligent Versatile Home (MavHome) project [60]. MavHome is funded by the Washington State University and the University of Texas at Arlington. The underlying approach of the project is to exploit the data collected from sensor nodes to infer the lifestyle patterns of residents and then adapt the behaviour of these devices according to the potential actions of the residents. The results of the MavHome project demonstrated its ability to reduce the daily interactions between humans and devices by more than 70% on average.

In additional to the aforementioned academic research projects, a number of industrial companies have introduced their smart home solutions. The Synco living system of Siemens [61] provides a complete solution for smart home including entertainment, energy saving and lighting control.

2-4-3 Smart energy

The main aim of smart energy (also known as smart grid) is to take advantage of sensing and actuating devices to improve the performance of electrical networks starting from the generation and transportation stages and downstream to the distribution stage [62,63]. Some major benefits of smart energy are adaptation of the generation capacities to market demands, automation of the grid configurations to alleviate power outages and facilitating seamless integration of different power sources such as distributed generators, electric vehicles and large scale renewable sources. A conceptual diagram for smart energy and its main components is depicted in figure 2.8.

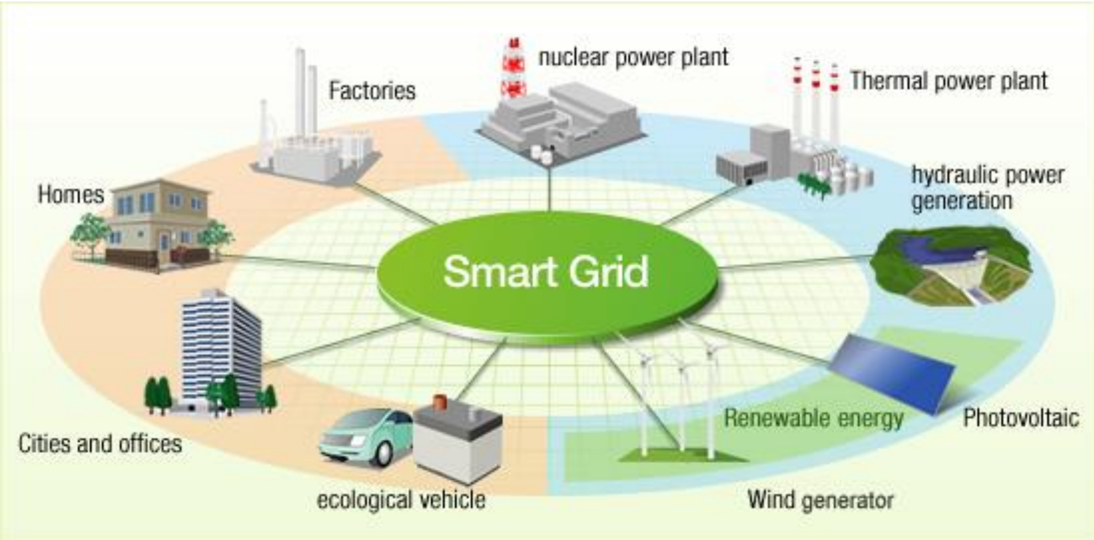


Figure 2.8 Conceptual diagram for smart energy adopted from [63]

Smart meter represents one of the most active research fields with more than 300 projects around the globe [66]. By the end of 2011, around 45 million smart reading devices were installed already in the European Union while it is planned to install about 240 million nodes by 2020. In the USA, 8 million smart meters were already installed by the end of 2011 and it is planned to install 60 million nodes by 2020. The main aim of the smart meters is to provide a real time feedback system that enables both consumers and utility suppliers to exchange their information and worked together to save energy.

2-4-4 Smart health

The main objectives of smart health are enhancement of the wellbeing and comfort of people, reduction of the healthcare costs and efforts and improvement of the early diagnosis of diseases. These goals are achieved by providing implantable or wearable WSNs that can collect the medical information and take the appropriate actions such as calling the emergency service. Smart health is stimulated by the low cost of sensor nodes and the amount of saving that can be achieved by utilising these networks. Some studies from Germany pointed out that employing the remote health monitoring applications to observe conditions of early discharged patients can save around €1.5 billion annually [66].

One of the pioneer research communities that considers these requirements and worked toward realising the concept of smart health is Intel. The proactive health group of Intel developed a complete solution that can enhance the quality of life of elderly peoples [67]. The underlying approach of such solution is to deploy WSNs that can collect medical information about people and share this information with the caregivers. Another example for the smart health applications is the PathFinder project that was proposed by the Ulster Community Hospitals Trust of Northern Ireland [68]. The PathFinder project deployed WSNs in more than 3000 homes in the community. Hospital staff can use these networks to track the status of the occupiers of these homes. The @Home project [69] is another example for the smart home, the @Home project use the ambulatory sensors that can monitor the electrocardiogram parameters, blood pressure and oxygen saturation level of patient. These sensors operate independently without patient intervention. When the system detects an abnormal case, the collected data is transmitted to the clinic via GSM networks. Another example for the application of WSNs in smart health is the Complete Ambient Assisted Living Experiment (CAALYX) project [70]. This project is funded by the European Union in the context of the sixth framework programme for Research and Technological Development (FP6). The main aim of this project is to increase the independence and self-confidence of elderly patients. CAALYX developed a wearable device that can measure specific vital signs, detect an injury and communicate this information with patient’s care providers. MyHeart is another FP6 research project [71] that aims to develop smart system that can monitor and prevent and cardiovascular diseases. The system employs wearable sensor nodes to monitor vital body signals and to provide the recommendations for patients after comparing the collected readings with normal values.

The recent applications of WSNs in the field of smart health are driven by the ability to fabricate tiny sensor nodes that can be implanted inside patient’s body. An example for the implantable WSNs is the Intra-Ocular Pressure (IOP) sensor that is used to diagnose glaucoma [72]. The IOP sensor node consists of golden ring antenna, microprocessor system based on the ASIC technology and strain-gage pressure sensor. An illustration for the IOP sensor is shown in figure 2.9 and a comprehensive survey for other applications of WSNs in different clinic areas (e.g., neurology, audiology, cardiology and ophthalmology) can be found in [66].

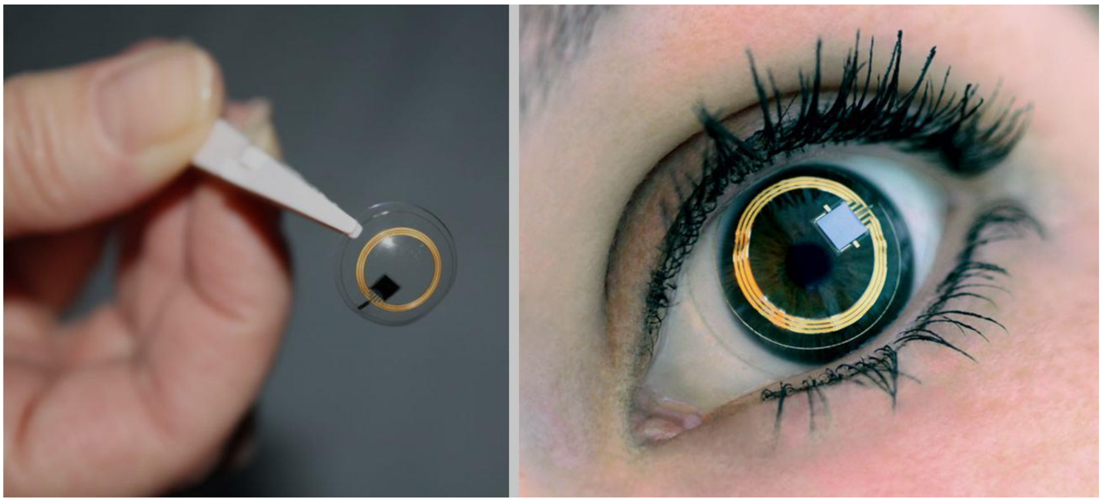


Figure 2.9 IOP implantable sensor node adopted from [72]

2-4-5 Smart environment

Smart environment is a collection of research efforts that employs the interconnections between WSNs and physical spaces to enhance their productivity, sustainability and safety. Towards these goals, diverse approaches and techniques have been devised; some of them focus on preservation and conservation of natural resources e.g., water and arable lands. Other efforts exploit the ability of WSNs to measure variation of physical quantities in order to build disaster management systems. This section explores some of these applications.

Precision Agriculture (PA) is a multidisciplinary application that employs WSNs in conjunction with other technologies (e.g., enhanced machinery, weather forecasting, variable-rate irrigation and herbicide-resistant) to improve the safety and quality of food products. Basically, PA exploits sensor nodes to harvest information about the crops or arable lands e.g., soil, drainage, isolation and topography. PA then uses this information to manage agricultural practices such as irrigation, crop rotation, and fertilisation [73-74]. Lofar Agro is one of the earliest PA projects that utilised the WSN to combat fungal diseases in potato fields [73]. This project deployed over 150 Mica2-like sensor nodes that are capable of measuring temperature and relative humidity of the crops. Furthermore the Lofar Agro project used a weather station to gather luminosity, air pressure, precipitation and wind strength of the field. This project used the TinyOS built-in routing protocol to facilitate multihop communication between nodes and gateways while global connectivity is provides by connecting these gateways to the Internet. Farm Level Optimal Water management Assistant for Irrigation under Deficit (FLOWAID) is a FP6 research project that aimed to resolve the irrigation deficit of limited water supply farms [74]. FLOWAID consists of a number of sensor nodes that collect relevant soil information such as volumetric water content and soil temperature. Each sensor node compares its readings with the typical values defined in the expert system provided by FLOWAID. Based on the results of comparison, sensor nodes take the appropriate decisions through controlling water supply valves. The results reported by the developers of the FLOWAID project demonstrated the ability of this project to produce the same quantities of crops with more than 10% save in water use.

Disaster management is another important application domain for WSNs. Such applications make use of the ability of WSNs to capture the transient events that precede disasters to anticipate their occurrences. Instigated by this ability, the University of Harvard has led a research programme to monitor active volcanoes [75]. This project used a number of sensor nodes each of which was equipped by a seismometer, a microphone, and a custom hardware interface board to capture volcanic events such as eruptions, tremor or earthquakes. This project employed the TinyOS as an operating system and the IEEE 802.15.4 standard to facilitate communications between nodes. The outcomes of this project highlighted the benefit of using WSNs in monitoring active volcanoes either in terms of deployment cost or resolution of collected data.

Another important application for the WSNs in the context of disaster management is to aid search and rescue of victims. One of the pioneer research projects that considered this requirement was developed by the University of Colorado and dubbed Connection-less Sensor-Based Tracking System Using Witnesses (CenWits) [76]. The CenWits consists of a number of in-situ sensor nodes that were distributed over the Yosemite National Park. Furthermore the CenWits provides wearable sensor nodes for the visitor of this park. All sensor nodes are connected through one of more access points that can determine the locations of nodes. Each sensor node emits beacon signals periodically and when two or more nodes are within communication range of each other, they record their presences at each other. The CenWits system exploits this information to track the locations of the lost or injured people and rescue them fast.

Intelligent Transportation Systems (ITS) is another example of the research efforts that utilise WSNs to realise the concept of smart environment. The main goals of ITS are to reduce traffic accidents; to improve the efficiency of vehicles and traffic networks and to minimise the disastrous effects of transportation systems on the environment. The underlying approach of ITS is to lessen deriver’s burden of making decisions by allowing vehicles, roads as well as pedestrians to interact with each other to predict the potential accidents and prevent them. The Vehicle Traffic Monitoring (VTM) [77] project is a good example for ITS, and this project has been developed as a part of the Libelium smart cities solution. VTM uses sensor nodes that can estimate the flow of pedestrians through measuring the flows of signals emitting from their phones. The gathered information is combined with other information about the traffic intensities of the roads and then used to advise the drivers and pedestrians about the possible congestion or accidents. An illustration for this project is presented in figure 2.10.

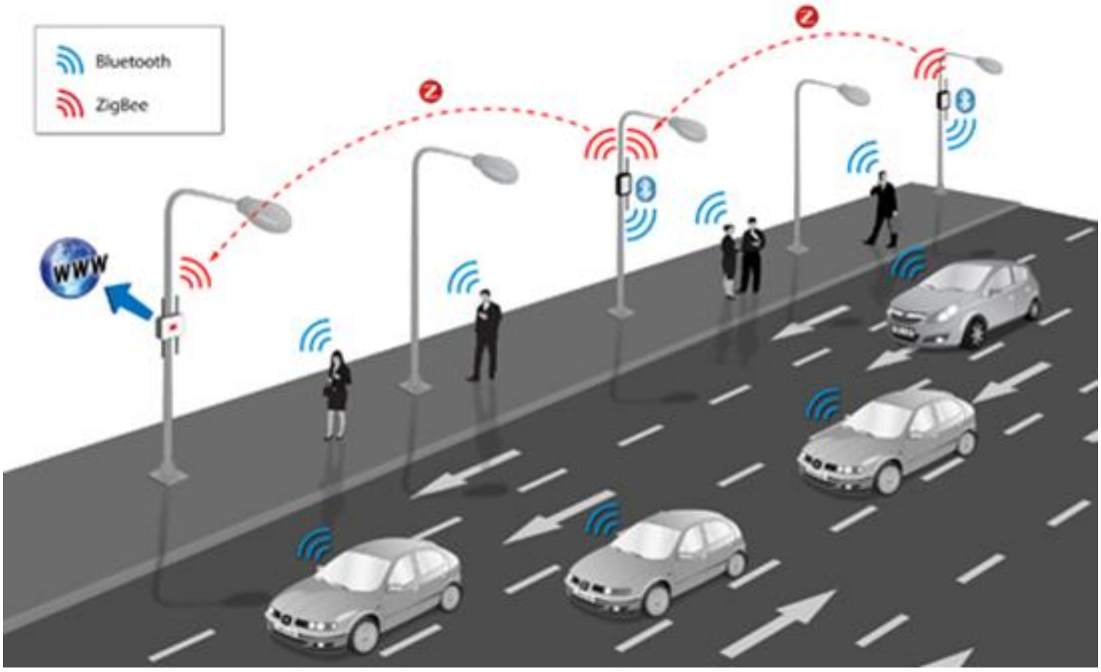


Figure 2.10 Conceptual diagram for vehicle traffic monitoring adopted from [77]

2-5 Design space of WSN communication protocols

The previous two sections introduced some of the state-of-the-art sensor platforms as well as some examples for real-life WSNs applications. It has been shown that there are strong demands to miniaturise the size of sensor nodes and to deploy them over large-scale networks. These demands pose great challenge in the design space of WSN communication protocols. This section summarises these key challenges.

1. Stringent resource constraints of the sensor nodes. A typical sensor node only has a low-speed processor, a few kilobytes of storage space, a short-range communication transceiver and is powered by a small battery. These constraints require development of lightweight protocols with a small memory footprint and low communication overheads.
2. Ad hoc topology. In most cases WSNs are deployed by scattering them over the area of interest to form an infrastructure-less network, moreover, this topology is changed over the time due to the depletion of node’s batteries. Hence the communication protocol should to be designed to suit different topology configurations.
3. Multihop communication capability. The severe energy constraint of sensor nodes dictates short transmission ranges of these nodes. Considering that the majority of WSNs applications require large-scale deployments with diverse topology scenarios demonstrates the need to account for disparate routing

policies (e.g., forwarding all packets over a single path or use of a load balancing policy) when devising communication protocols.

4. Heterogeneity of the traffic load of nodes. In WSNs each node has its own distinct traffic characteristics that differ from other nodes. The traffic of a typical node comprises of the internal traffic that is generated by the node itself as well as the external traffic that is received from other nodes for relaying. In WSNs, a node generates its internal traffic based on the status of the physical phenomena being monitored which varies between nodes. Similarly, the external traffic is received based on the location of a node with respect to the path of communications which also differs between nodes. Considering this requirement demonstrates the need to design a communication protocol in such a way that different traffic characteristics of different nodes can be accommodated without needing to reconstruct the protocol or causing performance degradation.

In summary, WSNs can be considered as a stack of stochastic processes, starting from the wireless communications that are controlled by interference and characteristics of the channel and up to the application layer when packets are generated according to the monitoring of random phenomena. Hence the communication protocol has to be designed to be powerful enough to cope with these stochastic characteristics of WSNs.

2-6 Conclusion

This chapter introduced WSNs and related technologies and explored some modern applications. The main purpose of this chapter is to provide the fundamental background required for the remainder of this thesis. It has been shown that the advances in fabrication technologies have quickened the development pace of sensor nodes and has miniaturised their size. It has also shown that the recent applications of WSNs demand large-scale development. The effects of miniaturisations of nodes and large-scale deployments of WSNs on the design space of WSN communication protocols have been discussed in this chapter.

Chapter 3

Medium Access Control, Duty Cycle Management and Routing Protocols

3-1 Introduction

The previous chapter introduced wireless sensor nodes and explored some of the major applications of Wireless Sensor Networks (WSNs). It has been shown that most of the current and envisaged applications of these networks demand deployment of scarce resources nodes over large-scale networks. Achieving such requirements falls on devising effective Medium Access Control (MAC), routing and duty cycle management protocols.

Fundamentally, MAC protocols attempt to regulate access to the shared channel amongst multiple nodes. The importance of the MAC protocols in WSNs is that these protocols can increase the probability of successful transmission which in turn can leverage the performance of networks. The routing protocols attempt to enhance the connectivity of the network by allowing a node to find the optimal route to the intended destination(s). The importance of the routing protocol for WSNs emerges from the fact that a typical WSN is a large-scale network which in turn requires an effective routing protocol that can handle the traffic flow of these networks. Finally the duty cycle management scheme aims to save the energy of a node by letting it sleep for a certain period if it has no role to play during this future period which in turn can prolong the lifetime of nodes. The main aim of this chapter is to introduce the underlying approaches of the aforementioned protocols and to overview their design space in the context of WSNs. This introduction includes the fundamental design principles of the protocols as well as the cross layer approach. Based on this introduction, some example MAC, routing, duty cycle and cross layer protocols are reviewed. The main aim of such reviewing is to highlight the main contributions and limitations of the work that has been proposed in the literature. This chapter then utilises both the introduction and review to present the motivation for this thesis.

The remainder of this chapter is organised as follows, sections 3-2 introduces the duty cycle management scheme. Overview for the key roles of the MAC protocol is given in section 3-3 and section 3-4 provides some example for this protocol. Section 3-5 introduces the routing protocol and section 3-6 considers some example for the routing protocols. In section 3-6 cross layer approach is discussed and the section 3-8 reviews some of the cross layer examples. Section 3-9 exploits review of duty cycle management, MAC and routing protocols to highlights the main motivation for using the inferential statistics to design these protocols and finally section 3-10 concludes this chapter.

3-2 Introduction to duty cycle management schemes

WSNs consist of a number of tiny nodes that are powered by small batteries and it is infeasible in some cases to replace these batteries, hence the lifetime of these networks is entirely dominated by the ability of communication protocols to provide an effective scheme that reduces consumed energy [24-25, 85]. In general, energy consumption by a node can be classified into two main categories: firstly, energy that is required to cater for traffic (e.g., transmission and reception of packets that need to be serviced by a node). Secondly, energy that is wasted while a node does nothing useful (e.g., overhearing: when a node receives other transmissions that are not destined to it or idle listening when a node waits for potential packets). One of the key strategies that is used to mitigate this needless consumed energy is to manage the duty cycle of nodes, i.e., to let a node sleep (turns its transceiver off) for a certain period if it has no role to play during this future period.

Broadly, most of the duty cycle schemes reported in the open literature are specified as a part of MAC or routing protocols hence the design space of these schemes is shaped by the main theme of the corresponding MAC or routing protocols. However, duty cycle schemes can be classified roughly into three categories: synchronous; asynchronous; and geographic.

Synchronous duty cycle management schemes assume that nodes within the network are perfectly synchronised and hence these nodes can wake and sleep together. Different techniques have been proposed to provide such synchronisation [24-25]. Some of these techniques utilise the cluster approach which divides the network into different zones and allows each zone to synchronise themselves. Other techniques suppose that a global synchronisation of all nodes across the network can be achieved by a central node. The main advantage of the synchronous duty cycle scheme is that it reduces the end-to-end delay of packets, as most of the nodes along specific paths are potentially awake. However, the main disadvantage of synchronous schemes is that it imposes a great overhead on the WSNs due to the requirement to disseminate synchronisation packets over the network. This overhead can lessen the lifetime of sensor nodes and reduce the throughput of the network significantly.

Asynchronous duty cycle schemes allow a node to manage its duty cycle independently of others. The main challenge of this type of schemes is how to maintain connectivity between different nodes [24-25]. Some of asynchronous schemes require neighbouring nodes to adjust their wake intervals in such a way that yields a temporal overlap between their wake intervals. Most of these schemes use hash functions to enable each node to figure out the overlapping parameters. Other asynchronous duty cycle schemes require a transmitter node to wake up the intended receiver prior to a transmission. Most of these schemes utilise the principles of the busy tone approach. These schemes assume that a sensor node can be equipped with two radio transceivers. One of them is used to exchange data packets and the other is used to wake the intended receiver. The main advantage of the asynchronous duty cycle approach is that it reduces the overhead of the synchronous schemes, however, the main disadvantage of the asynchronous scheme is that it increases the end-to-end delay of packets. The third category of duty cycle scheme is built on the assumption that a node is fitted with a GPS that can determine the location of a node [24-25]; nodes then exploit this location information to set their duty cycle. The main limitation of this type of duty cycle scheme is that most of the GPS systems demand a considerable amount of energy which can be inappropriate for scarce resources of sensor nodes.

3-3 Introduction to MAC protocols

A typical WSN is a large-scale wireless network in which packets traverse several hops. At each hop, the receiver does not only receive the intended signal but a superposition of all nearby on-going transmissions as well as the ambient noise [13, 22]. A receiver node can decode the intended signal correctly only when the ratio between the power of the intended signal to the total power of all interfering and noise signals is higher than a certain threshold. Thereby, the probability of receiving a successful signal can be increased either by increasing the power of the intended signal beyond the power of the expected interference and noise signals or by reducing the powers of the interfering and noise signals to a level that does not affect the quality of the intended signal. Diverse applications of WSNs have different ambient noise characteristics which in turn hinder their management; moreover, it is injudicious to increase the transmission power of sensor nodes due to the scarce resources of these nodes. Hence the best practice to increase the probability of receiving a successful signal is to reduce the number of interfering signals; however such reduction can result in a significant underutilisation of the channel. Therefore, the requirement for a mechanism to regulate access to the shared channel amongst multiple nodes is tremendously important for the proper operation of WSNs, such a mechanism is one of the main functions of the Medium Access Control (MAC) protocol.

MAC is the lowest half of the second layer of the Open Systems Interconnection (OSI) reference model. This sublayer operates on a top of the physical layer which makes the MAC protocol the main controller of the most significant energy consuming component of a sensor node (i.e., the radio transceiver). Hence the main function of MAC protocols in WSNs is to provide a lightweight and energy efficient mechanism to control the radio transceiver in such a way that allows a node to deliver its packets timely and over the highest possible reliable medium. Moreover, MAC is responsible to resolve or avoid the potential collisions between different nodes and to correct communication errors that may happen at the physical layer, and to perform other functions such as framing, and flow control. In order to achieve these goals, MAC for WSNs has to be designed to avoid all possible scenarios of performance degradation. Most of these scenarios fall broadly under four cases: overhearing, idle listening, collision and overhead [23, 32].

The term overhearing describes the case when a node receives a packet that is addressed to other nodes, hence a node consumes energy in reception of unwanted packets which in turn lessens the lifetime of sensor nodes. Moreover, the overhearing can reduce the throughput of a network and magnifies the end-to-end delay of packets as a node refrains from its transmissions whenever it overhears packets. The term idle listening refers the case when a node stays awake waiting for a potential transmission. The main limitation of idle listening is that it requires a transceiver to operate in idle mode which consumes a considerable amount of energy. Collision describes the case that the transmissions of two or more nodes are overlapped at a common receiver, hence the receiver cannot capture any transmission and thereby the energy expended in the transmission and reception is wasted. Another negative effect of the collisions is that they reduce the throughput of the network as they occupy the channel with useless transmissions. Moreover collisions can increase the end-to-end delay of packets, as colliding nodes have to retransmit their packets later. Overheads represent the control packets that are introduced by a protocol as a part of its operations; the negative effects of such overhead are multifaceted. Firstly, the overhead incurs a higher energy budget not only for the transmitter and receiver nodes but also for all the nodes that overhear the control packets. Secondly, the overheads may increase the probability of collisions between the control packets and data packets or even between control packets themselves, increasing the collision probability leads to reduce the throughput and magnify packet delay.

Different approaches and techniques have been proposed to design an effective MAC protocol that can mitigate aforementioned problems and that is suitable for scarce resource of WSNs. In general, these approaches are based on four main design schemes, namely: Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA) and contention based protocols. The following subsections discuss the fundamental characteristics of these schemes [22,13].

3-3-1 Frequency division multiple access

Frequency Division Multiple Access (FDMA) is one of the earliest techniques developed to regulate access to a shared channel [22-23]. The FDMA scheme divides the frequency band allocated to the network into sub-bands each of which is assigned to a node; hence different nodes can commence their transmissions simultaneously without a need for synchronisation between nodes. Interferences resulting from transmissions over adjacent channels are mitigated by separating these channels with guard bands. Figure 3.1 depicts a schematic diagram for the FDMA scheme.

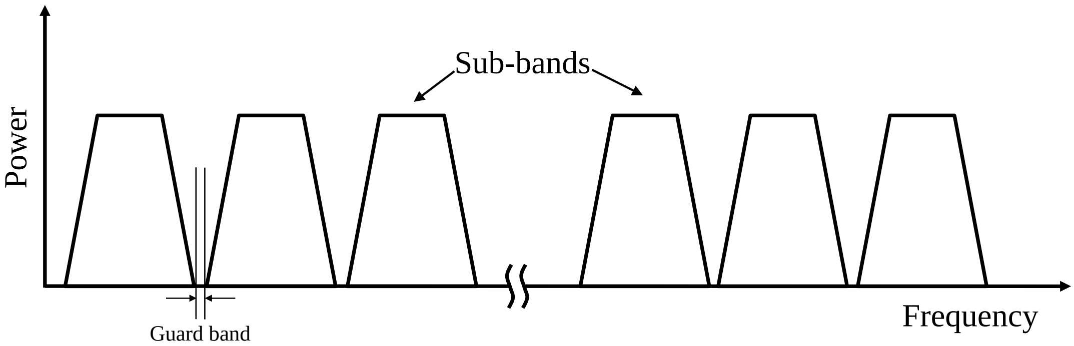


Figure 3.1 Schematic diagram for frequency division multiple access

The main advantage of FDMA is that it is easy to implement as each node use its own channel without need for coordination or synchronisation. However, the main disadvantage of FDMA is its requirement to equip each node with either multiple transceivers or frequency agility circuitry. This requirement increases the production cost of sensor nodes and more importantly consumes a considerable amount of energy which hinders the wide applicability of FDMA over WSNs.

3-3-2 Time division multiple access

Time Division Multiple Access (TDMA) divides the time axis of the shared channel into frames where each frame is further divided into slots [22-23]. These slots are assigned to nodes according to certain policy and each node transmits its packets only in its allocated slots. As a result, a transmission is carried out without interference which in turn reduces the collision probability and increases the channel utilisation significantly. Figure 3.2 shows a conceptual diagram for the TDMA scheme.

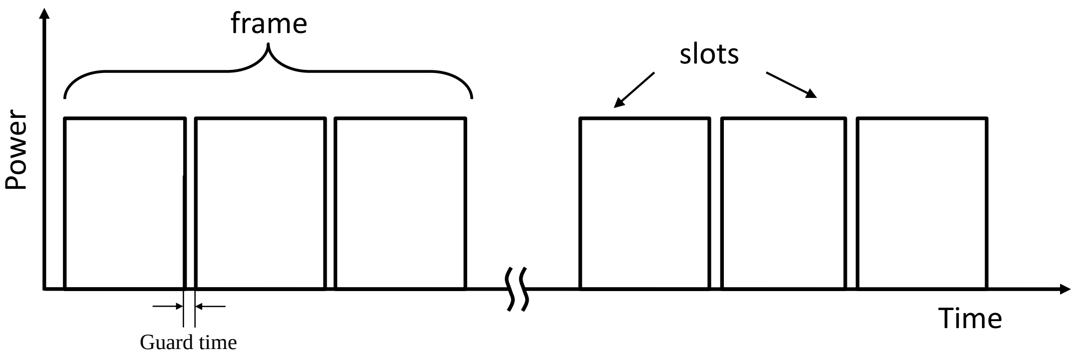


Figure 3.2 Schematic diagram for time division multiple access

The significant advantage of TDMA is that it provides a flexible scheme which can be adapted easily to accommodate traffic variations just by reconfiguring the allocation scheme without a need to readjust the hardware as the case with FDMA.

Another main advantage of the TDMA scheme is that it does not require special hardware equipment as the case with frequency agility in the FDMA scheme. However, the major downside of both TDMA and FDMA schemes is that their implementations waste a significant amount of resources (e.g., time or frequency) as both schemes require separation of the consequent time slots or sub-bands with guard times or guard bands respectively. Another major downside of TDMA is that it requires a perfect synchronisation across the entire network; this requirement consumes a considerable amount of energy and poses additional overheads on the network.

3-3-3 Code division multiple access

Code Division Multiple Access (CDMA) is a multiple access scheme that assigns each node a unique code that makes the transmission from a node distinguishable from transmissions from other nodes [78]. Hence different transmissions can be accommodated simultaneously over the same frequency band. CDMA utilises the approach of spread spectrum modulation to spread the original signal over a wide bandwidth. The direct sequence scheme is one of the traditional techniques used in CDMA. In this technique, each node multiplies its message by a unique wide band signal (dubbed spreading code) to generate a signal whose bandwidth is much greater than the bandwidth of the original message. Different signals from many nodes are sent synchronously over a common frequency band and the receiver reproduces the original signal by multiplying the received signal by the same spreading code of the transmitter. Typical direct sequence techniques require different nodes to use orthogonal spreading codes in order to reduce the interference between nodes. Other types of the CDMA techniques are based on frequency hopping, or hybrid methods of direct sequence and frequency hopping techniques.

Comparing the CDMA scheme with TDMA and FDMA demonstrates that the CDMA provides a soft capacity in which a new user can be added to the network without restriction on the number of frequency bands (as is the case with FDMA) or the number of time slots (as is the case with TDMA). However, increasing the number of nodes increases the interference which in turn degrades the channel utilisation. Moreover, the CDMA scheme suffers from the key limitation that it is difficult to devise a great number of orthogonal spreading codes that suite the large scale WSNs. Finally, the CDMA scheme requires a perfect synchronisation amongst all nodes within the network, since shifting the received signal can produce decoding errors.

3-3-4 Contention based multiple access

The previous sections reviewed the FDMA, TDMA and CDMA schemes. It has been shown that all of these protocols regulate access to the shared channel by scheduling the radio resources between nodes [22-23]. In general, these schemes require prior knowledge for the network topology and its traffic characteristics which make theses scheme unsuitable for the dynamic characteristics of most WSN applications. On the contrary, contention based multiple access schemes provide a distributed means that enable each node to decide the appropriate time for transmission without a requirement for a scheduling policy. Contention based protocols fall under two main classifications: ALOHA and Carrier-Sensing Multiple Access (CSMA) of which the CSMA with Collision Avoidance (CSMA-CA) is well established. Subsection 3-2-4-1 and 3-2-4-2 overview these two schemes respectively.

3-3-4-1 ALOHA

ALOHA was the first contention based MAC protocol developed at the University of Hawaii with the aim to build wireless networks that connected remote devices to mainframe computers [13,22]. The basic version of the ALOHA protocol is called pure ALOHA which allows a node to transmit its packets as soon as it ready to do so. A node upon receiving a successful packet sends an acknowledgment frame to inform the transmitter about the successful reception. When the acknowledgment frame is missed, a transmitter waits for some random time and then retransmits the packet again. The pure ALOHA protocol is simple as it does not require synchronisation or coordination between nodes; hence nodes can be added or removed easily without the need for rescheduling as is the case with the other MAC protocols (e.g., FDMA, TDMA and CDMA). Under light traffic intensity, the pure ALOHA protocol can provide good performance readings, however, increasing the traffic intensity increases the collision probability significantly which in turn decreases the throughput of the network and magnifies the delay of packets.

To overcome the limitations of pure ALOHA, the slotted ALOHA protocol was proposed. In slotted ALOHA all nodes are perfectly synchronised and all packets are of the same length. Furthermore, time is divided into slots of equal length where each slot equals the transmission time of a packet and a node can only commence a transmission at the boundary of a slot. The main advantage of the slotted ALOHA protocol is that it improves the throughput of network by avoiding the partially overlapping of packets found in the pure ALOHA protocol. However, this improvement requires perfect synchronisation amongst all nodes which hinders the wide applicability of this protocol over large scale WSNs. More importantly, the throughput readings of both pure and slotted ALOHA are still inferior under heavy or moderate traffic loads. This serious drawback is due to that in these schemes each node transmits its packets irrespective of the conditions of the shared channel. This drawback constitutes a motivation to develop Carrier Sensing Multiple Access (CSMA) schemes in which transmissions from nodes sharing the same channel can be respected. The underlying principle of CSMA schemes is that there is a wide class of wireless networks in which the propagation delay of a packet is negligible compared to its transmission time (as is the case in WSNs). Hence a node can detect ongoing transmissions from its neighbouring nodes and then uses this information to adjust its behaviour.

3-3-4-2 Carrier sensing multiple access

The Carrier Sensing Multiple Access (CSMA) protocol requires a node to listen to the shared channel before commencing a transmission (called the carrier sensing) to check if the channel is idle or occupied by a transmission [13,22]. If the channel is found idle then a node starts transmission otherwise a node defers its transmission (i.e., backs-off) for a random amount of time. Based on the policy by which a node assesses the channel and the back-off scheme, a number of variations of CSMA were proposed e.g., non-persistent, 1-persistent and p- persistent protocols.

In the non-persistent CSMA scheme, whenever a node assesses the channel as busy, it backs-off for a random amount of time before assessing the channel again. A node does not care about the state of the channel during the back-off intervals. The major drawback of the non-persistent protocol is that it may lead to underutilisation of the channel capacity as the channel may become idle while a node is backing-off.

In order to resolve this drawback, the 1-persistent CSMA protocol has been proposed. In the 1-persistent protocol, a node whenever assessing the channel as idle commences transmission immediately; otherwise a node assesses the channel persistently until the channel becomes idle. Hence the 1-persistent CSMA protocol can increase the channel utilisation as a node commences its transmission as soon as the channel becomes idle. However, the main drawback of the 1-persistent is that it consumes a considerable amount of energy in ceaseless sensing.

The p-persistent CSMA scheme was introduced to take advantage of both non-persistent and 1- persistent schemes. In the p-persistent CSMA scheme, a node that finds an idle channel commences transmission with probability p and with the complement of this probability a node waits for a predefined period and then attempts to transmit a packet again. This waiting period is set either to the propagation delay between the most distant nodes in the unslotted CSMA scheme or to a slot when the slotted CSMA scheme is used. When the waiting period has elapsed, a node senses the channel again; if it is found busy, a node senses the channel persistently until it becomes idle and then repeats the aforementioned process again. The main limitation of the p-persistent CSMA scheme is that it requires an accurate adjustment of the value of probability p with respect to the traffic intensity. Adjustment of p to a large value under high traffic intensity can increase the collision probability of packets and lessen the throughput of network. The main reason for this characteristic is that a large value of p allows nodes to commence their transmissions almost immediately, which in turn increases the contention to access the channel and magnifies the number of colliding nodes. On the other hand, setting the value of the probability p to a small value under light traffic rates can lead to considerable underutilisation of the channel, as a node potentially waits for longer periods while the channel is idle. The main challenge of employing the p-persistent CSMA scheme to multihop WSNs is that the traffic characteristics of these networks are highly dynamic and differ substantially amongst nodes. This in turn hinders prediction the appropriate value of p that suits these traffic conditions.

Comparing the performance of the ALOHA protocols with CSMA protocols demonstrates that the CSMA protocols are able to improve the throughput of network. CSMA enables a node to listen to the shared channel before commencing its transmissions which in turn eliminates the possibility of transmissions while the channel is occupied. However, the CSMA scheme suffers from high packet delay, as a node has to wait until the reception of the acknowledgment packet from the receiver to decide whether the transmitted packets have been received successfully or not.

With the aim of minimising the delay of the CSMA protocols, the Carrier Sensing Multiple Access with Collision Detection (CSMA-CD) protocol was proposed. CSMA-CD like the aforementioned CSMA protocols requires a node to assess the channel before transmission. If a node finds an idle channel then it transmits its packet immediately. Thereafter a node continues to monitor the transmitted signal and stops its transmission as soon as a collision is detected. Colliding nodes back off for a random duration generated according to a truncated binary exponential back-off algorithm based on the length of slot and the number of retransmissions. The main advantage of CSMA-CD is that it mitigates the delay of packets and increases the throughput of network by minimising the loss of channel capacity. However, this scheme cannot be used over most wireless networks, as detection of the occurrence or absence of collisions by the transmitter over the wireless channels is difficult due to the fluctuation of wireless channels and the decreasing of power of the signal over distance and time. Hence, a new contention based multiple access scheme called Carrier Sensing Multiple Access-Collision Avoidance (CSMA-CA) was proposed [13-32].

CSMA-CA leverages the performance of the CSMA protocols by reducing the possibility of collisions instead of detecting them as is the case in the CSMA-CD. The CSMA-CA protocol endeavours to solve two common problems found in wireless communications; namely, the hidden node and exposed node problems [13, 32]. A hidden node is a node that is within the range of the receiver but out of the range of the transmitter, conversely an exposed node is a node that is within the range of the transmitter but is out of the range of the receiver. Illustrations for the definition of the hidden and expensed nodes are given in figure 3.3 (a) and (b) respectively.

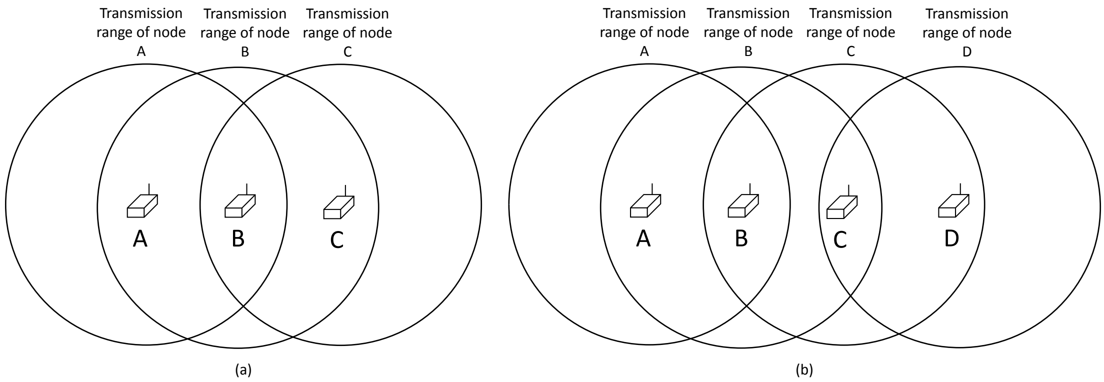


Figure 3.3 Hidden and exposed node scenarios

Figure 3.3(a) depicts a hidden node scenario; it shows that node B is within the transmission ranges of both nodes A and C but node A and C are outside their mutual transmission ranges. Hence a transmission from either A or C can be heard by B but a transmission from A cannot be received by node C and vice versa. In order to illustrate the scenario for hidden node collisions, let us assume that node A intends to transmit a packet to node B. Node A assesses the channel, finds it idle and commences its transmission. Assuming that during transmission of node A towards B, node C wishes to transmit a packet to node B. Here node C assesses the channel and finds it idle (as the transmission from node A cannot be heard by node C). As a result C commences its transmission to node B. Therefore, both transmissions from A and C will collide at node B. This collision is known as a hidden node collision.

Figure 3.3 (b) depicts the exposed node problem, it shows that node B is within the transmission range of both nodes A and C. Nodes A and C are outside their mutual transmission ranges and node D is within the transmission range of node C. Assuming that node B needs to send a packet to node A, node B assesses the channel, finds it idle and commences the transmission. During this transmission, node C wants to send a packet to node D. When node C assesses the channel, it finds a busy channel, as there is an ongoing transmission from node B to node A. So node C defers its transmission, however, it can be seen from figure 3.3 (b) that the transmission from node C to D and transmission from B to A can be carried out simultaneously and delivered successfully, as node D is outside the transmission range of node B.

In order to alleviate the hidden and exposed problems, several techniques have been proposed. The first technique is called the busy tone [13]. This technique is inspired by the fact that the collision takes place at the receiver while CSMA is performed at the transmitter nodes. Busy tone uses two distinct radio transceivers, one for data and the other for the control. The data channel is only used to transmit data packets while the control channel is used by the receiver to notify its neighbours that it is busy receiving a packet. Hence when a node wishes to transmit a packet, it first senses the carrier of the control channel. If there is busy tone signal then the node defers its transmission. The busy tone can solve both of the hidden and exposed node problems, however the need to equip each node with a two transceivers or a full duplex transceiver increases the complexity of tiny nodes which in turn pushes their production cost up and more importantly leads to fast depletion of their batteries.

The second technique that has been proposed to mitigate the effects of hidden node collision is dubbed Ready To Send (RTS) and Clear To Send (RTS) mechanism [79]. In such an approach, a transmitter sends the RTS packet to the intended receiver after assessment the channel as idle. The RTS packet is addressed to the intended receiver and contains information about the length of data packet. When the intended receiver receives the RTS packet, it assesses the channel and sends the CTS packet. If the transmitter receives a CTS packet it commences transmission of its data packet, otherwise the transmitter waits for a random time and then repeats the process again. Although the RTS/CTS mechanism can reduce the probability of hidden node collisions in some scenarios, the overheads associated with exchanging RTS/CTS packets as well as their energy costs make this mechanism unsuitable for the limited resource WSNs.

3-4 Examples of MAC protocol

This section reviews some of the MAC protocols that are reported in the open literature. The main aim of this section is to explore the design principles of these protocols and to highlight their main contributions and limitations.

3-4-1 IEEE 802.11 standard

The IEEE 802.11 standard was the first official standard that has been ratified especially for wireless networks. This standard is also known as Wi-Fi which stands for “Wireless Fidelity” [79]. The first version of IEEE 802.11 was published in 1997 and defined the physical and MAC layers specifications for Wireless Local Area Networks (WLANs). The maximum data rate of this standard was 2 Mbps obtained under three physical specifications: Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS) and Infrared (IR). Both DSSS and FHSS use the worldwide license-free 2.4 GHz Industrial, Scientific and Medical (ISM) band. A number of physical specifications were added to the standard to cope with the rapid development of wireless communication technology, for example, in 1999, the IEEE released two amendments IEEE 802.11b and IEEE 802.11a [89] which increased the maximum data rate to 11 and 54 Mbps respectively.

The IEEE defines two operational modes: Basic Service Set (BSS) and Independent Basic Service Set (IBSS). In the BSS, all communications between nodes traverse Access Points (APs) also known as Base Stations (BS). Conversely, the IBSS allows nodes to communicate with each other directly without need for an AP, which makes

IBSS more appropriate for the ad-hoc networks. Two MAC access mechanisms are defined by this standard: The Distributed Coordination Function (DCF) and Point Coordination Function (PCF). The DCF defines distributed channel access based on CSMA-CA while the PCF provides centrally controlled channel access based on a polling mechanism. In the PCF, the AP sends a specific packet called a poll packet to each node soliciting its transmission. A node upon receiving a poll packet sends it packet immediately to the AP without the need for a channel access mechanism.

DCF is based on the slotted CSMA-CA protocols. In this protocol, a node assesses the channel before commencing its transmission. If the channel is idle for at least a DIFS (DCF Inter Frame Space) time period, a node starts its transmission. Otherwise, if the channel is assessed as busy then a node generates a random number and stores it in a back-off counter. A node then assesses the channel in each slot. If the channel is found idle then a node decrements the value of the back-off counter otherwise a node freezes the value of this counter and resumes the decrement only if the channel is assessed as idle for a duration of DIFS. When the back-off counter reaches zero a node transmits its packet. Following a successful transmission, a receiver node sends an acknowledgment frame after waiting for a period called Short Inter Frame Space (SIFS). The value of SIFS is smaller than the value of DIFS which avoids the possibility that other nodes access the channel while the receiver transmits the acknowledgment frame. The CSMA-CA scheme of the IEEE 802.11 specification uses the Binary Exponential Back-off (BEB) policy in which the duration of the back-off is generated according to a uniform distribution whose initial width is set to the minimum Contention Window size (CW_{min}). This width is doubled after each collision and stops whenever it reaches the maximum Contention Window size (CW_{max}). The values of CW_{min} and CW_{max} are defined based on the physical specifications, e.g., the default values of these parameters are 31 and 1023 slots respectively for the DSSS modulation scheme. The CSMA-CA implementation defines the maximum retransmission limit which is the maximum number of times a node attempts to transmit its packet. When a node reaches this limit and a packet has not been received successfully, a node drops the packet. The IEEE 802.11 CSMA-CA protocol allows a network to use the RTS/CTS handshake mechanism to mitigate the effects of hidden node collisions. The RTS packet is sent prior to transmission of data packets and a receiver node sends the CTS packet after waiting for the SIFS time. Once the CTS packet is received successfully, the channel is reserved for the specific communication. Additional to this basic mechanism, IEEE 802.11 defines the virtual carrier sensing in which all nodes within the receiving ranges of the transmitter of RTS and CTS extract the value of the length of data packet from these control packets and store them in a data structure called Network Allocation Vector (NAV). A node then uses the NAV to predict the status of the channel without the need to sense the channel in each slot.

IEEE 802.11 specifies the Power Save Mechanism (PSM) as a means to manage the duty cycles schemes of nodes. The PSM assumes that the time is divided into beacon intervals and all nodes are perfectly synchronised. Each node wakes up at the beginning of the beacon frame and remains awake for a short period called the Ad-hoc Traffic Indication Map (ATIM). If a node wishes to send a packet to other node, the intended transmitter employs the CSMA-CA mechanism to send a control packet called the ATIM request message to the desired receiver. This receiver node upon reception of the request message replies with acknowledgment message called ATIM ACK and them remains awake for the remaining beacon period. Other nodes that do not send or receive ATIM request or acknowledgment packets during the ATIM interval go to sleep and awake at the beginning of the next beacon frame.

One of the major drawbacks of the IEEE 802.11 CSMA-CA protocol is that the performance of the network depends on the values of the contention windows. Small values reduce the delay and increase the probability of collisions (as the probability that two or more nodes reaching a zero back-off counter at the same time becomes high). Conversely setting the value of the contention window to a large value improves the throughput however it can increase the delay of packets. Employing the IEEE 802.11 CSMA-CA scheme in WSNs is impeded by two main limitations. Firstly this protocol is energy inefficient, as it requires a node to continuously monitor the channel before commencing transmission. This ceaseless assessment consumes a large amount of energy and makes this scheme inappropriate for low power networks. Furthermore, both of the RTS/CTS and NAV mechanisms are also energy inefficient, since the energy required to the exchange the RTS/CTS packets can exhaust the batteries of sensor nodes while the benefits of the NAV mechanism can be obtained only if a node is awake permanently which is inapplicable in WSNs as sensor nodes have to sleep for long durations in order to prolong their lifetime. The second main limitation of the employing IEEE 802.11 CSMA-CA over WSNs is related to the channel utilisation. The default values of the contention windows of this protocol are defined to suit the requirements of high data rate network, hence using this protocol over low-data rate such as WSNs can reduce the channel utilisation significant. Moreover, using the RTS/CTS mechanism imposes additional overheads and increases the probability of collisions between the control packets and data packets, which can reduce the throughput of WSNs significantly.

3-4-2 Sensor medium access control protocol

Sensor Medium Access Control (SMAC) [90] is one of the earliest MAC protocols that was designed for the particular requirements of WSNs; hence it serves as a base for a number of other MAC protocols. SMAC is a contention-based protocol that optimises the IEEE 802.11 CSMA-CA standard to save energy by introducing mechanisms to mitigate the energy wasted in overhearing, idle listening and collisions.

SMAC alleviates the effects of the idle listening by managing the duty cycle of nodes. SMAC divides the time axis into fixed intervals called frames where each frame is divided further into two intervals: a listen interval and sleep interval. In the sleep interval, a node turns its transceiver off to save energy whereas the listen interval is split into two periods: the synchronisation period which is used to exchange the synchronisation packets (also known as SYNC) and the communications period which is used to send data packets. During the synchronisation periods, a node listens to the channel for a certain period. If a node receives a SYNC packet from a neighbour then this node follows the scheme conveyed in the received SYNC and rebroadcasts it after waiting for a random delay. This node is called a follower. On the other hand if a node does not receive a SYNC packet during the listening period, then a node selects a scheduled scheme and broadcasts it in a SYNC packet. In this case a node is called a synchroniser, as other nodes can synchronise themselves using this packet. SMAC uses the term virtual cluster to refer to those nodes that have the same synchronisation scheme. A virtual cluster can contain a synchroniser and a number of followers. The main difference between the virtual cluster and the real cluster defined in other protocols (e.g., Low Energy Adaptive Clustering Hierarchy [91]) is that in virtual cluster nodes can communicate freely with nodes outside their cluster. Moreover, SMAC allows a node to move from a virtual cluster to others in order to find the duty cycle scheme that is more suitable for it; hence a node can reduce the energy consumed in idle listening further by moving to a low duty cycle virtual cluster.

The SMAC protocol adopts the overhearing reduction mechanism specified in the IEEE 802.11 standard. This mechanism allows those nodes that hear RTS and CTS packets and are not part of this communication to sleep during transmission of the data packet transmission. Moreover, SMAC reduces the probability of collisions of long data packets by enabling a node to fragment packets into small frames and send them in burst using a single RTS/CTS packet; hence when a frame is lost, a node retransmits a small frame instead of retransmitting a long frame which in turn saves energy of transmitter and receiver nodes.

Performance assessments reported by the developers of the SMAC protocol demonstrate the benefits of SMAC compared to IEEE 802.11. In terms of energy consumption, it has been shown that SMAC can save a considerable amount of energy compared to the IEEE 802.11 standard. However, it has been shown that the end-to-end delay of packets in SMAC is much higher compared to the IEEE 802.11 standard. This characteristic is attributed to the fact that in SMAC, packets traverse different virtual clusters each of which potentially has different duty cycle timings. Due to these shortcomings a number of enhancements for SMAC have been proposed. The work presented in [92] introduced a new mechanism called adaptive listening which allows neighbours of a receiver to wake up before the end of the transmission time. This proposal aims to reduce delay resulting when one of those nodes is the next hop of the recent received packet. Another enhancement for SMAC is presented in [93]. The underlying approach of this proposal is that a further saving in energy can be achieved if a node can commence its transmission immediately after reception of the CTS instead of waiting until the end of SIFS periods. The advantages of this proposal appear as a 5% reduction in energy consumption compared to the original SMAC.

3-4-3 Time out medium access control protocol

Time out MAC (TMAC) [94] has been proposed to overcome the fixed listening periods of SMAC. TMAC follows the underlying scheme of SMAC, e.g., notably the IEEE 802.11 CSMA-CA protocol with RTS/CTS handshake and frame structure.

The underlying approach of TMAC is that a node can go to sleep if it sends its packets in bursts at the beginning of the frame. Hence, TMAC allows nodes to wake up at the beginning of each frame to check whether there are transmissions intended to it. If the channel is assessed as idle for a pre-defined time called (TA) then a node goes to sleep immediately until the next frame. The TMAC protocol requires a node to send the RTS packet before commencing a data packet transmission and waits for a period of length TA. If the CTS packet is not received within this time, then a transmitter node resends the RTS packet again and waits for CTS packet. If the CTS packet is not received for the second time then a transmitter node goes to sleep and defers its transmission until the next frame.

Comparing the performance of TMAC and SMAC shows that the former protocol can reduce the energy by a factor of 5 with respect to the latter. However, this reduction depends on the length of the TA period with respect to the traffic variation rate [95]. Moreover, this energy saving leads to a significant decrease in the throughput of network, e.g., the throughput of TMAC is about 70% of the throughput of SMAC [94]. The other main limitation of TMAC is the early sleeping problem which occurs when an intermediate relaying node goes to sleep mode without consideration of the potential communications.

3-4-4 Berkeley medium access control protocol

The Berkeley MAC (BMAC) [96] was introduced to provide a flexible MAC protocol based on the CSMA mechanism. BMAC defined two core functions of the MAC protocol: Clear Channel Assessment (CCA) and Low Power Link (LPL). The CCA mechanism of the BMAC protocol allows a node to collect the signal strength readings of the channel when the channel is supposed to be free. A node then uses these readings to estimate the noise floor. Based on this estimation a node can distinguish between the noise signals and radio signals. Hence a node operating BMAC can provide a better prediction for the status of the channel compared to channel assessment mechanisms in traditional CSMA protocols. The LPL mechanism uses the preamble to manage the duty cycle of nodes. By default a node is asleep and wakes up only at a periodic interval called the receive check interval to check if there are activities on the channel. The LPL requires a node to send a preamble of length equal to or greater than the receive check interval before commencing a transmission, hence the receiver can awake and communication can be established. The obvious benefit of the LPL mechanism is that it reduces the overhead associated with transmission and reception of the RTC and CTS packets.

Moreover, LPL provides an asynchronous mean of managing the duty cycle of nodes.

The BMAC protocol provides a versatile framework that can be integrated with other MAC protocols or extended easily to cope with future requirements. However, this protocol suffers from the key limitation that is related to the preamble signals. As each transmission is preceded by a longer preamble signal to awake the receiver, a transmission is delayed by the length of these signals. Moreover these signals can consume significant amount of energy in dense networks as all the neighbours of a transmitter receive these signals, wake up and overhear the packet [13,23,95].

3-4-5 Traffic adaptive medium Access protocol

TRaffic Adaptive Medium Access (TRAMA) [97] is a scheduled based MAC protocol that uses the traffic characteristics of one-hop neighbours to manage the scheduling time of nodes. The TRAMA protocol assumes that all nodes within the network are synchronised and they share a single channel whose time axis is divided into fixed length intervals where each interval contains two periods: contention based and contention free periods. These two periods are subdivided into a number of slots. The contention-based period is used to setup and maintain the schedule of data transmission while the contention free period is used for data transmission. TRAMA consists of three algorithms: the Neighbour Protocol (NP), Schedule Exchange Protocol (SEP) and Adaptive Election Algorithm (AEA).

A node under the NP protocol selects a random slot and broadcasts a list of its one-hop neighbouring nodes; therefore each node can obtain information about its two-hop neighbours. The SEP protocol follows the same procedures of the NP except that a node broadcasts its future traffic characteristics (e.g., number of queued packets and the list of the intended receiver(s) of these packets). An additional difference between the NP and SEP protocols is that the broadcast domain of SEP is limited to one-hop neighbouring nodes. AEA uses the information gathered during the execution of the NP and SEP to schedule the transmission of the next interval. TRAMA determines the transmitter-receiver pair of each slot using a hash function that depends on the one-hop and two-hop neighbouring identifiers, the slot number as well as the traffic information.

The main benefit of the TRAMA protocol is that it enables nodes to access the channel based on their traffic patterns, which in turn can reduce the delay of packets and increase the throughput of network. However, the communication overheads required to maintain transmission-reception schedules and its computational complexity constitutes the main shortcoming of TRAMA.

3-4-6 Self-organising medium access control for sensor networks

Self-Organising Medium Access Control for Sensor networks (SMACS) [98] is a scheduled based MAC protocol that incorporates FDMA, TDMA and CDMA. SMACS mitigates the cost of the global synchronisation by allowing nodes within the network to negotiate with each other to arrange the appropriate scheduling scheme for their communications.

In SMACS, each node begins its operation by picking up a random frequency band and listening to the channel for a random period. If a node has not received a TYPE1 message at the end of this period a node broadcasts this message. The TYPE1 message is an invitation for other nodes to establish a communication link with the sender of such messages; a node upon reception of TYPE1 message waits for a random period and then responds to the inviter with TYPE2 message. When an inviter node receives TYPE2 messages, it selects the neighbouring node according to specific criteria (e.g., using signal strength or based on some information that is conveyed in TYPE2 messages) and sends a TYPE3 message to the selected neighbour. The TYPE3 message contains information about link schedule of the transmitter node. The intended receiver of the TYPE3 message uses this information to calculate the appropriate time and frequency of their communications and sends this information to the inviter node via a TYPE4 message. When the inviter node receives TYPE4 message, it modifies its schedule table accordingly. At the end of TYPE4 message communication links between the inviter and the sender of TYPE4 are formed.

The performance of the SMACS protocol relies mainly on the number of sub-frequency bands allocated to the network; a large number of bands alleviates collisions and can increase the channel utilisation of network, however it requires a powerful transceiver that can be tuned to different bands. On the other hand, using a small number of frequency bands leads to a high possibility of collisions. In order to overcome these shortages, the developers of SMACS defined an alternative means to regulate the channel access by replacing the FDMA with CDMA; however, the complexity associated with implementation of different encoding schemes hinders its applicability in WSNs.

3-4-7 Receiver initiated medium access control protocol

Receiver Initiated Medium Access Control (RIMAC) [101] is a MAC protocol that lets a receiver node control the reception time. The main aim of this protocol is to manage the duty cycle of nodes by letting transmitter and receiver nodes find a rendezvous time for exchanging data.

In RIMAC, each node wakes up periodically and broadcasts a beacon frame informing its neighbouring nodes that it is awake and ready to receive their data packets. Those neighbours that have pending packets commence their transmission immediately. Upon reception of a data packet, the receiver broadcasts another beacon frame acknowledging packet reception and soliciting other transmissions. If a node has not received a data packet within a pre-defined amount of time after transmission of a beacon frame, a node goes to sleep. The RIMAC protocol provides a collision resolution mechanism using the Back-off Window (BW) field in the beacon frame. The value of this field is set initially to zero, which dictates transmitter nodes to commence transmission without any back-off or channel assessment mechanisms. The receiver node increases the value of the BW field whenever it detects a collision, hence transmitter nodes commence transmission only after elapsing the period specified in the BW field. Transmitter nodes use clear channel assessment prior to a transmission which eliminates the possibility of commencing transmission while the channel is occupied with other ongoing transmissions. The value of BW is increased gradually if further collisions occur and when the value of BW reaches a threshold value and a transmission still suffers from collision, a receiver node goes back to sleep and wakes up on the next frame. The RIMAC protocol enables a transmitter node to announce its presence on the network by broadcasting a beacon-on-request frame following clear channel assessment. Thereafter, a receiver node waits for the BW duration specified in the beacon-on-request frame and then transmits a beacon frame indicating that it is awake and can receive data packets. Hence, the transmitter node starts transmission its data packet immediately rather than waiting for the next beacon.

The main advantage of the RIMAC protocol is that it provides a low overhead MAC protocol which can save energy consumed in exchanging control packets. However, this protocol suffers from a two main limitations. Firstly, a high level of delay results from the fact that a transmitter node has to wait for the waking intervals of a receiver. Secondly, the high collision probability that is a direct consequence of the fact that RIMAC does not provide a complete collision resolution mechanism but rather it requires a transmitter node to retry until BW of the receiver node reaches its threshold value. This mechanism magnifies the collision probability and increases packet delay.

3-4-8 IEEE 802.15.4 standard

The IEEE 802.15.4 standard [16] is the first official standard that specifies the Physical (PHY) and Medium Access Control (MAC) layers for a Low-Rate low-power Wireless Personal Area Network (LR-WPAN) which covers the requirements of WSNs. This standard provides simple and reliable wireless communications for low-duty-cycle LR-WPAN nodes and supports over 90 channels distributed across different frequency bands and features several topologies ranging from a simple star organisation to peer-to-peer networks. Furthermore, IEEE 802.15.4 provides contention based as well as scheduled based MAC protocols and allows nodes to be identified either by a 16 or 64bits address.

The IEEE 802.15.4 standard considers the regulations in different countries and supports a wide variety of frequency bands range from 780MHz to 10.6GHz. Among the most widely used bands are the world-wide unlicensed frequency bands, particularly the band at 2.4GHz. This frequency band provides 16 channels operating at 250kbits/second data rate and 62.5ksymbols/second symbols rate. These specifications imply 4bits/symbol are transmitted as 16 nearly orthogonal 32bit pseudo random (PN) noise sequences. These bits are modulated using a Direct Sequence Spread Spectrum (DSSS) scheme [16].

The IEEE 802.15.4 standard specifies transmission, sensing and detecting power threshold levels. The nominal transmission power is -3dBm and the maximum transmission power is left for local regulation. The receiver sensitivity is defined as the minimum average power level that causes a node to switch from an idle state to a receive state. The IEEE 802.15.4 standard specifies this value as less than or equal to -85dBm which enables a receiver to receive a signal with 1%frame error rate if interference is not present and the packet size is kept at 26bytes. The carrier-sense power threshold is used to determine if the channel is idle and ready for transmission or not. A node, before commencing a transmission, listens to the channel and compares the strength of the received signal with this threshold. If the received signal is below the carrier sense threshold, a node considers the channel as idle and commences transmission, otherwise the transmission is postponed. According to the standard [16], the value of this threshold is -75dBm (10dB above the receiver sensitivity) and the time required for this estimation is 8symbols which is 128µs for the 250kbits data rate.

This standard [16] defines a Wireless Personal Area Network (WPAN) as a group of nodes that share physical channels and have a main coordinator that initialises and synchronises nodes within the WPAN. According to the standard, nodes are classified into two types: Full Function Devices (FFDs) and Reduced Function Devices (RFDs). FFDs can communicate with all nodes over the WPAN and can play the role of coordinator. On the other hand, RFDs can communicate only with FFDs and work just as end nodes; i.e. FFDs can be used as routers to rely packets to other nodes while the RFDs cannot act as a router for other nodes' packets. According to the IEEE 802.15.4 standard, nodes in a WPAN can be organised in two different categories of topologies: star and peer-to-peer. In a star topology, there is only one WPAN and all nodes (either FFDs or RFDs) have to communicate with the WPAN's coordinator directly. This topology is suitable for traditional data-collection WSNs. Conversely, in a peer-to-peer topology, many WPANs might coexist which makes this topology suitable for multihop networks.

IEEE 802.15.4 WPANs can operate either in a beacon-enabled WPAN or a beaconless WPAN. In the former, coordinators emit regular beacons, while in the latter, beacons are sent upon request. A beacon is a frame that contains information about the coordinator and its corresponding WPAN and is used to synchronise other nodes within the same WPAN. In beacon-enabled WPANs, a coordinator uses a ‘superframe’ which contains a Contention Access Period (CAP) and a Contention Free Period (CFP, also called Guaranteed Time slots GTS). In the CAP, packets can be transmitted by different means: directly without any channel assessment (direct transmission mode), by using a slotted Carrier Sense Multiple Access with Collision Avoidance (slotted CSMA-CA) mechanism, or through an indirect transmission method. During the direct transmission mode, frames are sent without prior channel assessment, this occurs when a node can commence transmission between 320μs and 512μs and there is room in the current CAP period. Conversely, in indirect transmission, a sender node stores frames in its buffer and notifies the receiver. When a receiver node becomes ready to receive its pending frames, it sends a data request; thereafter, the originator node sends its pending frames. Finally, slotted CSMA-CA is employed when the direct or indirect modes are infeasible. In the CFP, on the other hand, a node commences communication directly during slots which were reserved previously without any channel assessment mechanism, i.e., a transmitter node just turns its transceiver on and starts transmission. The CFP is optional and limited to those nodes which have low latency data. In contrast, beaconless WPANs allow nodes to communicate each other directly without need for a coordinator makes which this channel access mode is more suitable for multihop WSNs.

In the CSMA-CA protocols, a node starts its contention to access the channel by backing-off for a random period distributed uniformly between 0 and $(B_0 - 1)$ slots, thereafter a node assesses the channel for one slot. If the channel is found idle, a node commences its transmission immediately, otherwise if the channel is assessed as busy, a node backs-off for a random period distributed uniformly between 0 and $2(B_0 - 1)$. This operation is repeated until either a node finds an idle channel and commences transmission or the maximum number of back-off stages is reached. In such a case a node reports a channel access failure to higher layers and drops packets. A transmitter node upon completing its transmission waits for the Inter Frame Space (IFS) period which is specified to give the receiver some time to process the data received over the physical channel. If a transmitter node has not requested an acknowledgment frame then it waits for the appropriate IFS and then proceeds immediately to service the next frame if it exists. However, if a sender requests an acknowledgment then the transmitter should wait for 12 slots (the time required by the receiver to switch its transceiver from the receive state to the transmit state). If the acknowledgement packet is not received within this time then the transmitter resends its packet.

The main advantage of the IEEE 802.15.4 standard is that it provides a MAC protocol with different mechanisms designed especially for WSNs. These features have facilitated the development and proliferation of WSNs by replacing numerous proprietary protocols [22-23]. Moreover, IEEE 802.15.4 has become an essential component of higher-level protocol stacks such as ZigBee [102] and 6LoWPAN

[17]. Moreover the CSMA-CA protocol of the IEEE 802.15.4 standard offers a lightweight and energy efficient MAC protocol that mitigates the ceaseless channel assessment found in the IEEE 802.11 standard. Another key advantage of the IEEE 802.15.4 standard is that it employs a truncated binary exponential back-off scheme that is more suitable for the low data rate of WSNs. However, the main limitations of the IEEE 802.15.4 CSMA-CA protocol are that it does not provide a mechanism to reduce the effects of the hidden node collisions or to manage the duty cycle of nodes.

3-4-9 Zebra medium access control protocol

Zebra Medium Access Control (ZMAC) [103] is a hybrid MAC protocol combining both contention based and contention free mechanisms with the aim to take advantage of these two mechanisms. The ZMAC protocol employs the CSMA mechanism under a Low Contention Level (LCL) condition, which is attributed to the fact that under this condition the probability of collision is low. Hence nodes can save the energy consumed in exchanging the scheduling packets without affecting network throughput. On the other hand, under a High Contention Level (HCL) condition, ZMAC employs the TDMA mechanism in order to alleviate the throughput degradation that results from increasing the collision probability.

ZMAC starts up in the setup phase in which the following four operations take place: neighbour discovery, slot assignment, local frame exchange and global time synchronisation. During the neighbour discovery, each node broadcasts a control packet to its neighbour soliciting their one hop neighbour information, thereby, at the end of the neighbour discover operation each node has information about its two hop neighbouring nodes. This information is used as an input set to the time slot assignment algorithm. The main aim of this algorithm is to allocate an owner for each time slot providing that no two nodes within a two-hop neighbourhood can commence transmissions in the same slot. During the local framing operation, each node announces its periodicity of assigned slots and finally the synchronisation operation is used to synchronise the clock cycles of all nodes within the network. The data transmission phase takes place after the completion of the setup phase. In ZMAC, a node can be in either HCL or LCL mode. The default operating mode is the LCL which is based on the CSMA mechanism. CSMA of the ZMAC protocol exploits the value obtained from the slot assignment algorithm to adjust the back-off window of a node, i.e., the owner of time slot uses a small back-off window compared to other contenders. A node switches to the HCL mode upon reception of Explicit Contention Notification (ECN) packets from a two-hop neighbour. The ECN packet is generated by a node whenever it observes that the current packet loss rate is higher than a threshold value. The ECN is propagated to the two-hop nodes and dictates them to use the TDMA scheme.

The main advantage of ZMAC is that it provides a flexible means to regulate the access to the shared channel which can prolong the lifetime of nodes and maintain an acceptable throughput. However, ZMAC suffers from several key limitations: firstly, the requirement to execute different functions in the setup phase. Secondly, ZMAC is not a scalable protocol as its operation depends on analysing the traffic intensity of two-hop neighbours. Finally, the ZMAC protocol requires nodes to switch between different MAC protocols which in turn increases the computation complexity and hinders its wide applicability over limited resource WSNs.

3-4-10 Crankshaft medium access control protocol

Crankshaft is a hybrid MAC protocol that is inspired by the operation of the internal combustion engine [104]. This protocol combines a contention-based mechanism with a contention free mechanism in order to take advantage of both mechanisms. Crankshaft is built on two main pillars: firstly to allow each node to wake periodically but asynchronously from other nodes and secondly to reduce the contention during these waking intervals by allowing transmitter nodes to exchange their readiness for transmission before awake intervals.

Crankshaft divides the time axis of the shared channel into a fixed length frame where each frame is subdivided into a group of slots. These slots fall under two categories: broadcast and unicast. The broadcast slots are used to reduce the contention amongst transmitter nodes prior to transmission which takes place in the unicast slots. The unicast slots are based on the principles of the TDMA mechanism in which each slot is allocated to a receiver node. Crankshaft provides a static allocation scheme in which a node determines its own unicast slot by computing the modulo of its MAC address with respect to the width of unicast slots. The main advantage of this allocation scheme is that it provides a lightweight algorithm which does not impose high communication overheads. However, as such a simple scheme can result in a disastrous consequence (as two or more neighbouring nodes can be assigned to the same slot), Crankshaft allows the owner of a unicast slot to propagate a signal during its allocated slots. A node then reverts to the receive mode if contention is resolved. A transmitter node when wishing to send a message to a receiver node selects a particular slot during the contention period and listens to the channel for a short amount of time to detect ongoing transmissions. If the channel is found idle, the transmitter node emits a preamble signal to notify other nodes regarding its readiness to send. Afterwards, a transmitter node waits for the unicast slot of the intended receiver and then transmits a symbol message followed by the data packet. An acknowledgement frame is sent upon successful reception. As the performance of this transmission scheme depends mainly on the moment in which a transmitter node emits its preamble signal, the Crankshaft protocol allows each node to select this moment based on a truncated geometric distribution.

The main advantage of the Crankshaft protocol is that it provides a simple means to regulate access to the shared channel. However, this simplicity endangers the stability of the protocol over large scale WSNs [23].

3-5 Introduction to routing protocols

Routing is a mechanism that enables a node to select the optimal path for its packets according to the application requirements; this mechanism operates at the third layer of the OSI reference model [32,80]. In general, routing begins by a neighbour discovery process in which each node attempts to gather information about its neighbours (e.g., addresses of neighbours, the signal strengths of reception for these neighbours). A node stores this information in a data structure called a routing table; depending on the design philosophy of a routing protocol, nodes exchange their routing information and compute the cost to a final destination based on specific criteria called routing metrics. Hence a node routes packets over the path that has the minimal cost (i.e., optimal path), moreover, a typical routing protocol provides techniques to maintain and correct the routing information.

In WSNs, routing protocols have developed as a direct consequence of the small transmission ranges of sensor nodes and the requirement to deploy these networks at large scale [13, 26]. The design space of routing protocols for WSNs is influenced by most of the challenges considered in section 3-2, e.g., the requirements for a lightweight and energy efficient scheme and the need to prolong the lifetime of nodes, increase the throughput of the network and reduce the end-to-end delay of packets.

One of the most important aspects that need to be considered when designing routing protocol is the routing metrics, as these metrics represent the criteria upon which the optimal route is selected. Hop count is the traditional and widely used routing metric. This metric computes the distance between a pair of nodes as the number of links that connect these nodes irrespective of the quality of the links. The main advantage of using the hop count metric is that it is simple to implement. However, the key disadvantage of this metric is that it does not consider the characteristics of the links which in turn can lead to serious performance degradation of the network [81]. This shortcoming motivated researchers to propose a number of routing metrics that measure different quality aspects of links or nodes (e.g., throughput, delay, reliability, memory usage or residual energy). The fundamental approach of these metrics is to send a probe packet periodically soliciting the quantity of interest. Some examples of this type of metric are [81]: Expected Transmission Count (EXT), Expected Transmission Time (ETT), Weighted Cumulative Expected Transmission Time (WCETT) and the Metric of Interference and Channel switching (MIC). The obvious advantage of these routing metrics is that they can improve the performance of routing protocol by providing dynamic characteristics of links or nodes. However, the disadvantage of this sort of routing metrics is that they poses additional overheads on the scarce resources of sensor nodes which in turn drains their batteries, increases the end-to-end delay of packets and reduces the throughput of networks.

The remainder of this section reviews the fundamental approaches that have been used to design routing protocols for WSNs.

3-5-1 Flooding routing approach

The flooding scheme is the simplest routing technique that disseminates data packets over networks [13, 22, 32]. The flooding protocol allows each node to broadcast its data packets to all of its neighbouring nodes. Each one of these neighbours that hears the packet for the first time rebroadcasts it. This process is performed by all nodes within the network and terminated only when either the destination node is found or every node within the network hears the packet and broadcasts it once. The obvious advantage of the flooding protocol is that it is simple and easy to implement. However, this scheme suffers from a number of limitations that make it unsuitable for most WSN applications. Firstly the number of transmissions required to deliver a packet is of the order of the number of nodes within the network, which makes it inefficient over large scale networks. Secondly, as each node rebroadcasts the packet once, a node receives several copies of the same packets from all of its neighbours which consumes energy of nodes significantly and occupies the channel with needless transmissions.

These limitations motivated researchers to propose some enhancement for this scheme. Efficient flooding [82] is an enhanced version of the pure flooding protocol in which each node except the originator of a packet rebroadcasts a packet with a certain probability p. A small value of this parameter reduces the rebroadcasting rate, however, it can cause the so-called flooding dying out problem which prevents flooding from approaching every node across the network. Conversely setting the value of p to a large value increases the overhead of broadcasting. Therefore the requirement to determine the best value of the p is the key factor for the success of this proposal.

Another mechanism proposed to enhance the performance of pure flooding protocol was introduced in [83]. This proposal allows each node to delay its rebroadcasting for a random period. If a node hears a packet at least k times then a node refrains broadcasting otherwise a node broadcasts the packet. Although this proposal can improve the performance of the pure flooding scheme, this improvement relies on the proper selection for the random delay periods and the value of k.

In general, the flooding routing scheme and most of its variants are not suitable for WSNs as these schemes pose high communication overheads that consume a large amount of energy, reduce the throughput of network and increase the end-to-end delay of packets significantly.

3-5-2 Proactive routing approach

The proactive routing scheme was developed to satisfy the requirements of ad hoc networks. The underlying assumption of the proactive scheme is that each node has to always maintain a route to all other nodes of the network [80-81]. Proactive routing schemes are based on the distance vector and link state protocols that were designed for use over the Internet such as the Routing Information Protocol (RIP) and Open Shortest Path First (OSPF) protocols.

In distance vector routing protocols [80,84] a node broadcasts a vector comprising a list of all known destinations along with their distances. Vectors from different nodes are disseminated across the network transitively and then paths are computed by each node. In link state protocols, a node propagates the status of its neighbouring links through the network and each node then computes routes to other nodes.

The proactive routing schemes provide different means to exchange routing information: periodic update, event trigger or combination of them [80,84]. The periodic exchanges occur regularly while the event trigger updates take place whenever a topology change happens. The primary advantage of proactive routing protocols is that they can provide a route to a potential destination without waiting for route discovery. However, the overhead associated with exchange the routing tables and their updates consume substantial amount of energy and occupy the channel with control packets. Moreover, proactive routing protocols require high computation complexity and memory usage as each node has to maintain information about all other nodes and compute their cost locally.

3-5-3 Reactive routing approach

Reactive routing schemes (also known as on-demand routing schemes) provide an alternative approach to designing routing protocols [80-82]. They aim to eliminate the overhead imposed by the requirement that each node should maintain fresh routes to all other nodes across the network. Reactive schemes relax this requirement and enable a node to discover the routes only when there is a need to use it, hence they are called on-demand schemes.

The fundamental mechanism of reactive schemes is route discovery which enables a node whenever it has no route to a destination to request this information from other nodes. A typical route discovery mechanism floods a network with request messages soliciting all known routes to the required destination. After a node obtains these routes, it selects the optimal path and forwards its packet consequently. Depending on the specification of the route discovery mechanism, a node maintains the routes for specific periods.

The main advantage of reactive routing schemes is that they reduce the resource overuse found in the proactive routing schemes. However, the disadvantage of the proactive routing schemes is that they require a node to delay the routing process until all routing information has been gathered. This amount of delay increase proportionally with the size of the network and the frequency with which topology of network is changed.

3-5-4 Geographical routing approach

Geographic routing schemes incorporate geographic information into proactive or reactive routing protocols in order to enhance their performance [80]. Geographic routing requires each node to determine its location either using the Global Positioning System (GPS) or with respect to a reference point. Therefore the requirement to exchange the routing tables or execute route discovery protocols can be relaxed which in turn reduces the overhead of the associated control packets. Although geographic routing schemes can meet the need of some applications of wireless networks, the scarce resources of the WSNs hinder its wide applicability.

3-5-5 Cluster routing approach

The aforementioned routing schemes can be classified as flat routing since each node within the network has to acquire routing information by itself, hence the overheads associated with exchanging routing packets increase proportionally with the number of nodes [80, 84]. Conversely, cluster routing schemes alleviate this overhead by dividing the network into groups often called clusters where each cluster contains a number of nodes. The cluster leader (also known as cluster head) is responsible for collecting the traffic from all the cluster members and forwarding it further to other clusters heads. Hence only a certain number of nodes (cluster heads) within the entire network maintain network-wide connectively whereas other nodes have to maintain a route to the cluster leader only.

The obvious advantage of cluster routing schemes is that they reduce the overheads of control packets required by the proactive and reactive schemes. Reducing these overheads can prolong the lifetime of nodes and increase the throughput of network. However, dependency on cluster heads to provide network-wide connectivity can lead to single point of failure when cluster heads die out or can lead to traffic bottlenecks around the cluster heads under high traffic intensity. Therefore, devising an effective mechanism to select the diameter of clusters with respect to the traffic intensity is the paramount design issue of cluster routing schemes.

3-6 Examples for routing protocol

This section reviews some of the routing protocols that are reported in the open literature. The main aim of this section is to explore the design principles of these protocols and to highlight their main contributions and limitations.

3-6-1 Low energy adaptive clustering hierarchy protocol

Low Energy Adaptive Clustering Hierarchy (LEACH) [91] is a cluster-based protocol that organises nodes within a network into clusters where each clusters comprises a number of nodes and a single cluster head. All nodes within a cluster forward their packets to the cluster head which aggregates these packets and forwards them further to the base station. As the aggregation and relaying of packets consumes a considerable amount of energy from cluster heads, the LEACH protocol rotates the role of the cluster head between all nodes in order to distribute the energy loads between all nodes.

The operation of the LEACH protocol consists of two phases: a setup phase and steady-state phase. In the setup phase, each node generates a random number between 0 and 1 and sends this value to the base station which uses an equation to select the number of clusters and elect their heads. When a node is elected as a cluster head, it uses a non-persistent CSMA scheme to broadcast advertisement packets to inform other nodes. Afterwards, non-cluster head nodes use the signal strength of the advertisement packets to select the nearest cluster head. After formation of the clusters, each cluster head creates the TDMA schedule and sends this schedule to other nodes within the cluster. In the steady-state phase, each cluster head collects and aggregates packets from all nodes within its cluster and then uses CDMA to send these packets to the based stations. CDMA is used to mitigate inferences between transmissions from different cluster heads. The LEACH protocol allows nodes after a pre-defined time to go back into the setup phase to select new cluster heads. This operation is specified to prolong the lifetime of sensor nodes that carry out the role of cluster heads.

The LEACH protocol suffers from key limitations: firstly, all nodes within the network should have enough transmission power to reach the base station which is unsuitable for large scale WSNs. Secondly, all nodes should have high computational power to support different MAC schemes (e.g., TDMA for intra-cluster transmissions, CSMA for cluster advertisements and CDMA for inter-cluster transmissions) which is inappropriate for tiny sensor nodes. Thirdly, the rotation of the cluster head imposes additional overheads that diminish the benefit of the rotation approach. Finally, there is no clear justification for the methodology by which the base station selects the cluster heads. Due to these limitations, LEACH-C [99] has been proposed, in which each node at the beginning of each setup phase, informs the base station about its position and residual energy. Based on this information, the base station works out the cluster structure for the entire network and broadcasts it. Although LEACH-C can enhance the cluster configuration, it still suffers from the other limitations of the LEACH protocol.

3-6-2 Power efficient gathering in sensor information systems

Power Efficient GAthering in Sensor Information System (PEGASIS) [100] is a routing protocol that aims to regulate traffic flow of the network by creating a chain from each node to the base station. The fundamental approach of this protocol is that a node can save its energy if it communicates only with its closest neighbour.

In the PEGASIS protocol, each node uses the signal strength of received packets to estimate the distance to its neighbouring nodes. Based on these estimations a node tunes its transmission power to the level such that only the nearest neighbour can receive the signal. Hence each node transmits its data packets to a single neighbour which in turn aggregates the packets and forwards them towards the base station. This mechanism forms a single chain connecting all nodes to the base station. The nearest node to the base station is dubbed a leader node. PEGASIS allows different nodes to act as a leader in different turns where the turn is the duration since the base station queries readings from sensor nodes until these readings are gathered and passed to the base station. The leader node of each turn is selected according to specific criteria defined by PEGASIS e.g., threshold distance to the base station. After a node is elected as a leader, it passes a token to its neighbours. This token is propagated downstream to collect the readings from nodes of the network.

The main advantage of the PEGASIS protocol is that it reduces the energy consumption of nodes and increases the channel utilisation by eliminating the overhead required to setup and maintain clusters, as is the case with the LEACH protocol. A comparison between the PEGASIS and LEACH protocols demonstrates that the former protocol can prolong the lifetime of network by a factor of two with respect to the latter protocol. However, the PEGASIS protocol suffers from a number of limitations that hinders its wide applicability in WSNs. Firstly; PEGASIS imposes an excessive delay for the data packets collected from nodes far away from the leader node. Secondly, the single leader node of PEGASIS represents a bottleneck of the traffic as the collision probability near the leader node increases significantly due to high traffic intensity and packet aggregation policy. Finally, the PEGASIS protocol is built on the assumption that most of sensor nodes are able to communicate with the based station directly which is impractical in many WSN scenarios [32].

3-6-3 Destination sequence distance vector protocol

Destination-Sequenced Distance-Vector (DSDV) [105] is one of the earliest proactive routing protocols that developed especially for ad-hoc networks. The design objective of the DSDV protocol is to enhance the operations of the Routing Information Protocol (RIP) [80] over wireless networks.

RIP like many distance vector protocols suffers from the counting-to-infinity problem which occurs when a node dies out or is isolated from the network and the neighbours of the isolated node update its routing table (i.e., by setting the hop count of the isolated node to infinity) without notifying their neighbours (i.e., two hop neighbours of the isolated node) about the isolated node. In such a case, when the routing updates are received by the neighbours of the isolated node, they think that a better path to the isolated node exists and consequently update their routing tables. This process continues until the hop count of the isolated node reaches infinity due to exchanging the routing tables between different nodes, hence it is called the counting-to-infinity problem.

The DSDV protocol overcomes the counting-to-infinity problem by tagging each routing update with a sequence number. A node operating DSDV maintains a monotonically increasing sequence number and conveys it in all routing update messages. Moreover, a node stores the highest known number sequence number for each destination in the routing table; hence a node can determine whether the received routing information is fresh or not just by comparing the sequence number of the received message with its peers in the routing table. DSDV uses both periodic and event trigger routing updates. In the periodic update, nodes exchange their complete routing tables whereas in the event trigger updates, nodes transmit portions of the routing table that have been changed since the last full update. If the portions of the event trigger updates are large to fit into a single packet, then full updates are used in the next update period.

The main advantage of the DSDV protocol is that it is simple to implement, however this protocol is inappropriate for most of the WSN applications, as DSDV imposes great overheads which can consume energy of sensor nodes, reduce the throughput of the network and increase the end-to-end delay of packets.

3-6-4 Optimised link state protocol

Optimised Link State Routing (OLSR) is a proactive routing protocol that enhances the traditional link state routing scheme [106]; OLSR mitigates the overhead associated with propagation of the link state updates over the network by using the Multi-Point Relay (MPRs). The MPR of a node is the smallest subset of its neighbours that can reach the two hop neighbours of this node. OLSR allows only the MPRs nodes to broadcast the link state updates and these updates are limited to the changes that occurred between the MPRs and their selectors, this approach can reduce both the number of updates and their broadcasting domains significantly.

The obvious advantage of OLSR is that it provides a low overhead routing protocol that can save the resources of tiny sensor nodes. However, the main disadvantage of this protocol is that it requires a precise adjustment for the update intervals with respect to the frequency with which the topology is changed. Using a small update interval can increase the overhead and consume a high level of energy while using a large update interval can lead to use of outdated routing information. Another major disadvantage of the OLSR protocol is that finding the MPRs sets is a Non Polynomial (NP) hard problem which depends mainly on the topology of the network. As a result, the benefits of OLSR differ substantially from one network to other.

3-6-5 Dynamic source routing protocol

Dynamic Source Routing (DSR) [107-108] is a reactive routing protocol that was specified by the Internet Engineering Task Force (IETF) to provide a flexible protocol suitable for ad-hoc networks.

DSR relies on two main mechanisms: routing discovery and routing maintenance. Routing discovery is initialised whenever an originator has to communicate with a destination whose route is not yet known. In such a case the originator broadcasts a control packet called a Route Request (RREQ). The essential field of the RREQ is the address of the originator node and the address of the destination. Each node that receives the RREQ and does not know a route to the destination appends its address to the header of the RREQ packet and rebroadcasts it. When the RREQ request is heard by a node that knows a route to the destination or the destination itself, then such a node constructs a new control packet called Route Reply (RREP) containing the complete path from the originator to the destination. The RREP packet is then unicasted to the originator of RREQ through the path that is specified in the header. Once this packet is received by the originator node, it forwards the packet through the specified path and then stores it in its routing cache. The DSR protocol uses the routing cache instead of the traditional routing table that is used in the proactive routing protocol. The routing cache enables a node to store multipath routes to the same destination, hence when a path is broken the alternative path can be used immediately instead of executing a routing discovery protocol. The second main mechanism of the DSR protocol is route maintenance. This mechanism is executed whenever a node recognises a broken link. In such a case a node generates a Route Error (RERR) control packet. This packet is sent to all nodes that use this broken link. These nodes then update their routing cache and perform a routing discovery process if there are no alternative routes.

The main advantage of DSR is that it avoids the routing loop problem found in some of the proactive routing protocols, as each node knows the complete routing paths. The second main advantage is that DSR provides a lightweight means to update the routing cache of nodes, as each node whenever it hears the RREQ or RREP updates its cache without the need to exchange the entire routing table as is the case in the proactive routing protocol. DSR provides another mechanism to gather routing information using the promiscuous mode. This mode enables a node to overhear the ongoing communication between its neighbours and acquires the routing information without the need to overwhelm the network with control packets. However, the main disadvantage of the DSR is that it requires a node to accumulate the complete path in each data packet which reduces the data payload of packets, increases their transmission time and energy and magnifies the probability of collision of packets. The second main disadvantage is that DSR does not provide a means to purge outdated routes, as each node keeps the all routes in its cache permanently unless it has been informed through the route error control packets. Keeping all routes all the time does not only lead to routing of packets over broken paths but also increases the occupancy of buffers which are severely scarce in WSNs.

3-6-6 Ad hoc on-demand distance vector protocol

Ad Hoc On-Demand Distance Vector (AODV) [109] is a reactive routing protocol that was proposed to overcome the main limitations of the DSR protocol. AODV replaces the routing cache with the routing table and removes the source routing from data packets.

The AODV protocol allows each node to maintain its routing information in a routing table as is the case with the proactive routing protocols; however, this table is local to the node and contains information about its neighbours only. In the routing table each entity is associated with a lifetime. If a route is not used during its lifetime then this route is deleted from the routing table. Like the DSR protocol, the AODV protocol is based on two mechanisms: routing discovery and routing maintenance. Both of these mechanisms are similar to their peers in the DSR protocol except that AODV replaces the source route with the destination sequence number that was proposed in the DSDV protocol [80]. Another main difference is that the AODV protocol allows a node to control the diameter of flooding of route request packets by modifying the Time To Live (TTL) field.

The main advantage of the AODV protocol is that it saves the energy of the sensor nodes and bandwidth of the networks by eliminating the source routing of the DSR protocol. However, the significant disadvantage of the AODV protocol is that nodes can store only a single path to each destination instead of multiple paths as is the case with DSR. Dependency on a single path can increase the overheads associated with routing discovery and path maintenance, since a node has to carry out these mechanisms each time this single path is broken.

3-6-7 Zone routing protocol

Zone Routing Protocol (ZRP) [110] is a hybrid routing approach that integrates both reactive and proactive routing schemes in order to reduce the communication overheads of proactive schemes and to minimise route acquisition latency of reactive schemes. ZPR divides networks into different zones and allows each node to use a proactive scheme for intra-zone routing and to employ a reaction routing protocol for inter-zone routing.

The ZRP protocol uses a parameter named “zone radius” to distinguish the area where the proactive routing scheme is used from the area where the reactive routing scheme is employed. All nodes whose distances in hops from the given node are less than the value of zone radius represent the zone of the given node, while all nodes that lie on the boundary of the zone radius are called peripheral nodes. The ZRP protocol uses the IntrA-zone Routing Protocol (IARP) to route packets between the nodes within its zones. The IARP is a link state routing protocol that requires each node to maintain fresh information about all nodes within its zone area. When a node is called upon to route a packet to a destination, it consults its routing table to find a route. If such a destination is within the zone radius, then the up-to-date route for that destination must already exist in the routing table, hence a node can use that route immediately. Otherwise, if the destination node is not within the zone’s area of a relaying node, the IntEr-zone Routing Protocol (IERP) is used to find the route. In the IERP protocol, a node sends a query message to its peripheral nodes. The query message is uniquely identified by a combination of the requestor’s address and query sequence number. If any one of peripheral nodes knows a route to the destination, it responds to a query message with a known route. On the other hand, if none of the peripheral nodes knows a route to the destination, the query message is propagated to all other peripheral nodes over the network. The ZRP protocol allows each peripheral node to wait for a random period between query reception and forwarding, which mitigates the possibility of collisions that may happen when two or more peripheral nodes forward the query message at the same time. The ZRP protocol enables a peripheral node to use the identifier of a query message to eliminate blind propagation of the query message.

The main advantage of the ZRP protocol is that it provides a flexible scheme that limits the scope of the proactive scheme to small areas and restricts the search space of the reactive scheme to certain nodes. The obvious result of this approach is that it reduces the energy consumption of nodes and the communication overheads of the channel which in turn provides better performance readings compared to pure reactive and proactive schemes. However, the disadvantage of ZRP is that its performance relies mainly on precise adjustment for the zone radius parameter with respect to the density of the network and traffic intensities of inter/intra zones.

3-6-8 ZigBee routing protocol

The ZigBee alliance [102,111] was one of the leading organisations to propose meshing protocols for IEEE 802.15.4 networks. Their proposal combines two kinds of routing algorithms: Tree-based Hierarchical routing and a lightweight version of the Ad-hoc On-demand Distance Vector routing protocol.

Tree-based hierarchical routing maps the network into a tree where the root of the tree is dubbed the coordinator of the network. Other nodes are classified as either: routers or end-devices. A router device can serve as a parent for one or more end-devices and is responsible to assign their addresses and route their packets while the end-devices cannot accept a child and do not perform a routing function. One of the important functions of the tree-based algorithm is to assign addresses to nodes within the network. The address assignment is a top-down scheme that is initiated by the coordinator and stops as soon as all nodes within the network are assigned an address. The address assignment process is controlled by three parameters: maximum number of routers, maximum number of children per a router and the depth of network. After setting these parameters, the coordinator divides the address space into a number of blocks each of which is assigned to a router child of the coordinator and the last block is reserved for end-devices attached directly to the coordinator. Each router thereafter uses the value of the aforementioned three parameters to assign each end-device child a unique address and each router a pool of addresses. This operation continuous until each leaf device (i.e., a device that has no child) is assigned an address.

The tree-based routing algorithm exploits the hierarchical address scheme to enable a router device to discover whether a destination node is within its children just by comparing the destination address with the addressing block of the router device. If the destination node is within the block of addresses of the router device, it forwards the packet towards its destination immediately; otherwise a router device forwards the packet towards its parent. A parent performs the same checking procedure and forwards the packet towards its parent if the destination is not within its children. This process is carried out until the destination is found or the packet is forwarded to the coordinator which can determine the appropriate branch of the tree.

The ZigBee routing protocol employs a simple version of the AODV protocol which is used to enhance the performance of the tree-based hierarchical routing protocol by allowing a node to establish mesh links between nodes that are not connected by tree links. In this protocol, route request messages are sent on demand whenever the route to the destination nodes cannot be allocated using the tree-based hierarchical routing, e.g., due to failure of tree links. In such a case, all router devices that know a route to the destination respond to the requestor router with a route reply message. If all nodes within the networks have no active path to the destination, the route error messages are sent to the requestor node. The route repair mechanism is used to repair the network topology by propagating a route discovery message.

The main advantage of the ZigBee routing protocol is that it provides a complete solution including addressing and routing schemes, however this advantage comes with many penalties. Firstly, the address allocation scheme of this protocol is not flexible enough to cope with the dynamic characteristics of the topology of WSNs. Secondly, incorporating the tree with reactive routing schemes poses a high communication overhead which in turn consumes a considerable amount of energy. Finally, the ZigBee routing protocol does not provide a scalable mechanism that can tackle the requirement of the envisaged applications of WSNs in which each node can serve as originator, router and final destination simultaneously.

3-6-9 IEEE 802.15.5 standard

IEEE 802.15.5 [112] is a recommended practice that was ratified by the IEEE 802.15 working group who specified the IEEE 802.15.4 standard; the main aim of the IEEE 802.15.5 standard is to extend a device's connectivity without needing to increase the transmission power or receiving sensitivity of tiny sensor nodes.

IEEE 802.15.5 introduced a new sub-layer called MESH above the IEEE 802.15.4 MAC layer. This MESH sub-layer maps the physical topology to a tree with a single root called a Mesh Coordinator (MC). The MC is responsible to create the network, allocate addresses to nodes and synchronise them. Other nodes join the tree by selecting the appropriate parent "a closer node to the MC" based on specific criteria (e.g., signal strength indicator or hop count). In the IEEE 802.15.5 standard, a node is restricted to one parent; however, a parent may have more than one child and many siblings which form the mesh links. The location of a node within a tree is identified by an 8bits level tree where the tree level of the MC is 0 and each child has a tree level one greater than its parent. IEEE 802.15.5 assigns addresses centrally by the MC using the adaptive block addressing scheme where a group of nodes under a common parent have consecutive addresses. These addresses are used to make decisions when forwarding packets. In practice, when a node receives a packet, it checks its routing table to determine the next hop by comparing the final destination address with the block addresses of its neighbours. If no entity matches the final destination address, a packet is forwarded until it is received by the MC which can decide to forward it to another branch or drop it.

The IEEE 802.15.5 standard uses a link state routing scheme to allow nodes to route their packets over mesh links (i.e., non-tree links). The Link State Tables (LSTs) of nodes is a data structure consisting of two parts. The first part is the neighbour list which stores information about address blocks and multicast groups of *k*-hops neighbours (where *k* is a parameter setting by the MC) and their one hop devices. The second part of LSTs maintains the connectivity matrix of a node. This matrix maintains the routing information of a node which is updated using periodic HELLO frames and over 20 control messages. Additionally, IEEE 802.15.5 uses the data packets to detect the mismatch between routing tables in different nodes. Each node sets the Up-Down flag of the header of data packet according to the location of the next hop with respect to the tree level. As the packet traverses from one node to other, this flag is changed accordingly. If any device observes any inconsistency, it sends the link state mismatch control frame and other devices replay to rectify information. When a device wishes to leave the network, it sends a leave control message requesting other nodes to update their link state tables.

In terms of energy saving mode, IEEE 802.15.5 specifies two energy saving modes: Synchronous Energy Saving (SES) and Asynchronous Energy Saving (ASES). In the SES mode, all nodes have the same active/inactive periods, whereas in the ASES mode, a node can enter an inactive period independently of others. Nodes under the

ASES mode negotiate with each other to schedule their future transmission times before going to sleep.

The main advantage of the IEEE 802.15.5 standard is that it provides a simple meshing mechanism that is compatible with the physical and MAC layers of the IEEE 802.15.4 standard. Moreover, the addressing scheme of IEEE 802.15.5 is more flexible than its peer in the ZigBee routing protocol as it does not pose restrictions on the depth of network or number of children that a node may have. However, the key limitations of the IEEE 802.15.5 standard are that it does not provide a means to stop immortal packets from becoming trapped in routing loops [2] and it imposes great overheads on the network resources by requiring a node to maintain both tree and mesh routing information.

3-6-10 Routing protocol for low power and lossy networks

The Routing Protocol for Low-power and lossy networks (RPL) [113] was released by the Internet Engineering Task Force (IETF) Routing Over Low power and Lossy (ROLL) working group to provide an Internet Protocol version 6 (IPv6) based routing protocol for low power and lossy networks e.g. IEEE 802.15.4, Low Power Wi-Fi, Bluetooth, and even wired links such as power line communication. The ROLL is a highly parametric protocol that defines the core routing functions and leaves the detailed specifications (e.g., addressing assignment mechanism, metric computation and energy saving mode) to be optimised according to specific applications.

The RPL protocol maps the physical topology to a Destination Oriented Directed Acyclic Graph (DODAG) logical topology using the Internet Control Message Protocol version 6 (ICMPv6). The main reason of using the DODAG topology instead of the tree or cluster topologies is that DODAG provides a flexible structure in which several roots can coexist, moreover each node can have multiple parents, children and siblings. After construction of the DODAG topology, root nodes initiate a RPL instance which specifies the Objective Function (OF) of the routing protocol e.g., reducing end-to-end delay of packets or avoiding high loss rate links. The RPL protocol allows different OFs to be defined over the same network and further each OF can contain more than one objective. A root upon selecting the OF disseminates control messages to inform other nodes about the new instance, thereafter all nodes except the root of a specific OF rank themselves with respect to the root. RPL introduces a strict ranking mechanism in which each node can be assigned a unique rank based on routing metrics appropriate for specific OF. In addition, RPL defines a number of mechanisms to rectify a rank inconsistency between nodes.

The RPL protocol employs a reactive distance vector routing protocol which allows nodes to update their routing tables on demand using a number of control packets. A node uses the Destination Advertisement Object (DAO) which is sent by a node to announce its address to its parent. When a DAO message is received by a parent, it adds a child's address to the routing table and forwards DAO to its parent (i.e., the first grandparent of the originator of the received packet). This operation continues until the DAO packet reaches the root of the RPL instance. When a device receives a packet it checks its routing table. If no entity matches the destination of the packet, then a received packet is forwarded up until a common ancestor of the source and destination is found which routes packets down to the destination. Furthermore, RPL uses the Measurement Object (MO) and Route Discovery Option (RDO) to measure the quality and existence of routes respectively. RPL defines a local and global routing repair mechanism to update and maintain consistent routing tables amongst all nodes.

The main advantage of RPL is that it facilitates employing IPv6 over WSNs which can widen the applicability of these networks and increase their proliferation. Moreover, RPL defines a complete routing protocol that works at the network layer which makes this protocol more robust compared to other standards such as the ZigBee and IEEE 802.15.5 protocols. However, the key disadvantage of the RPL protocol is that it provides a general framework which makes the performance of RPL depends to a great extent on its implementations.

3-6-11 Geographic adaptive fidelity protocol

Geographic Adaptive Fidelity (GAF) [114] is a geographical routing protocol that aims to manage the duty cycling of sensor nodes and maintains a reasonable level of routing fidelity. In the GAF protocol, nodes exploit their Geographical Position System (GPS) to determine their locations and form a so-called virtual grid which comprises a number of cells. All nodes within the same cell are treated equally for the purpose of routing, thereby only a node within a cell is awake and other nodes can go to sleep to save their energy. The GAF protocol defines two main operational stages: discovery and negotiation. In the discovery stage, a node determines its location and broadcasts this information periodically to its neighbours. In the negotiation stage, a node uses a rank-based algorithm to allow a node with the highest residual energy to wake for longer periods compared to other nodes within the same cell.

The main advantage of the GAF protocol is that it provides a modular duty cycle scheme that can be used with any routing protocol. However, the amount of energy that is saved by GAF depends heavily on the topology of the network and the diameter of the cells of the virtual grid. Increasing the cell diameter allows more nodes to save their energy; however, it requires each node within a cell to reach to all other nodes within the same cell which in turn requires some nodes to increase their transmission power to satisfy this requirement. Conversely, reducing the diameter of a cell can lead to a reduction in the benefits of the GAF protocol. Another main shortcoming of the GAF is that it requires a node to consume a significant amount of energy in determining its location and in negotiation with other nodes to manage their duty cycle scheme.

3-6-12 Span protocol

Span [115] is a cluster based routing protocol that forms a backbone of coordinators across the network. Each coordinator covers a certain area of a network and all of these coordinators are overlapped spatially. The underlying approach of Span is to keep coordinators awake perennially to carry out routing over the network and allow other nodes to stay asleep to save their energy.

The Span protocol allows each node to elect itself as a coordinator through the so-called coordinator eligibility rule which states that a non-coordinator node should elect itself as a coordinator if at least two of its neighbours cannot communicate either directly or through a coordinator. The Span protocol requires each node to delay its electoral announcement for a random period in order to mitigate the possibility that several nodes discover the lack of coordinators and elect themselves simultaneously. The function that is used to generate the random periods depends on two parameters: the residual energy of a node and the number of neighbours that can be connected via it. A node stops acting as a coordinator as soon as it discovers that all of its neighbours can communicate with each other either directly or through coordinators. In such a case, a coordinator node nominates itself as a tentative coordinator and still performs functions of coordinators until a new node is elected as a coordinator instead. The tentative policy is used to alleviate the loss of connectivity that may happen due to the transition between coordinators.

The main advantage of the Span protocol is that it provides a simple means to manage the duty cycle scheme of nodes and a reasonable routing scheme. However, this protocol increases the end-to-end delay of packets and reduces the throughput of the network due to the overheads associated with the election process.

3-6-13 Adaptive self-configuration sensor network topology

Adaptive Self Configuration sEnsor Network Topologies (ASCENT) [116] provides a general framework to manage the duty cycle scheme of nodes without specifying a particular routing mechanism. ASCENT defines two types of nodes: passive and active. Passive nodes turn their transceiver on to collect information about the status of the network without participation in a packet forwarding mechanism or routing information exchange. On the contrary, active nodes relay packet and exchange routing information until their batteries run out. The ASCENT protocol allows the sink or active nodes to send a help packet soliciting passive nodes to switch to the active state to overcome the high packet lost rate. When a passive node decides to be active, it broadcasts a packet announcing its presence on the network. The help messages stop as soon as the transmitter of such messages recognises that the packet lost rate is maintained below a pre-defined limit.

The main limitation of the ASCENT protocol is that it is built on the assumption that there are a very large number of sensor nodes that are allocated to monitor the same area, hence when a node dies out, other nodes can take its position immediately without losing connectivity. However, this assumption is not valid in some applications of WSNs in which each sensor node has a certain role that may differ from roles of other nodes e.g., smart home or smart health applications. The other major limitation of ASCENT is that it does not provide an effective means to save energy, as the passive nodes are still awake and listen to the channel, which consumes their energy.

3-7 Introduction to cross layer

A cross layer design approach aims to improve the performance of communication protocols by violating the hierarchical interactions of layered reference models such as Open Systems Interconnection (OSI) model. This violation can take different forms such as devising new interfaces between adjacent layers, merging two or more layers, sharing the information between different layers or developing new communication architecture [86-87].

The fundamental logic of a cross layer approach is that the layered model is more appropriate for wired networks since their communication channel is robust and hence each layer can be designed independently from the other layers [86-87]. Wireless networks, on the other hand, exhibit less independent characteristics. For instance, the conditions of the physical channel or residual energy of a node dominates the functionalities of other protocols. Hence incorporating such information in the design spaces of other layers can enhance their performance. However, different aspects related to the longevity, stability and proliferation have to be considered when designing cross layer protocols. This section discusses some of these aspects in order to highlight the strong and weak features of the cross layer approach with respect to the traditional layered model.

The first crucial aspect that needs to be considered when employing the cross layer approach to designing communication protocols is the modularity of the developed protocol [86-88]. In the traditional layered model, each layer abstracts the remaining layers in such a way that functionalities of a layer do not contradict with functionalities of other layers. Hence, developers of different layers can work in parallel without concern about the interoperation of their developed protocols. On the country, most cross layer approaches modify the layered model in such a way that reduces or even disengages its modularity. The significant disadvantages of the lack of modular architecture in cross layer approach are the increasing cost of development and maintenance of developed protocols and reduction in the reusability of existing protocols. More importantly, the cross layer approach minimises the longevity of the developed protocol as each modification or upgrading requires a redesign of the whole cross layer protocol.

The second important aspect related to the design of a cross layer protocol is the requirement to analyse the new interactions between different layers crossed by the developed protocol. In the classical layered model, each layer interacts with its adjacent layers via limited and restrained interfaces that are well-defined and thoroughly understood. Hence the behaviour of the layered model at runtime is predictable. Conversely, the cross layer approach optimises the performance of protocols by introducing additional interactions between different layers. Some consequences of these interactions are foreseen and can be tackled at the design phase whereas other consequences arise only at runtime and may result in undesirable characteristics. Accounting for such characteristics requires analysis of the developed protocols using complicated analytical techniques [86-88] which in turn increases the design-to-production cycles and limits the applicability of using a cross layer approach to design sophisticated protocols suiting modern applications of WSNs. Moreover, the unintended consequences that may result from the cross layer approach can endanger the stability of the networks [81, 86-87] and stifle their proliferation since most of the cross layer protocols cannot be standardised.

The aforementioned discussion demonstrates that fact that the improvements offered by the cross layer approach come with several serious limitations e.g., lack of stability and longevity and high development cost. In WSNs, these limitations can outweigh the potential benefits of a cross layer approach since most of WSN applications require a robust design approaches that can be extended easily and adopted to suit vast and diverse applications of these networks. Moreover, as most of the current and envisaged applications of WSNs rely mainly on integrating these networks with global communication systems, hence using the cross layer protocol can hinder such integrations.

3-8 Examples of cross layer protocols

This section reviews some of the cross layer protocols that are reported in the open literature. The main aim of this section is to explore the design principles of these protocols and to highlight their main contributions and limitations.

3-8-1 Sensor protocol

Sensor Protocol (SP) [117] is a cross layer approach that provides unified abstraction for the MAC and routing layers. The main aim of this abstraction is to enable different MAC and routing protocols to coexist over the same architecture. The fundamental logic of SP is that different WSN applications need different MAC and/or routing characteristics, thereby abstracting the functionalities of these two layers can provide a rich design space for WSNs.

The SP protocol is built on three main operations: data transmission, data reception and neighbour management. The data transmission and reception operations as their names imply are used to manage the transmission and reception of the data packet respectively. The SP protocol does not provide a special behaviour for these operations except that it supports quality of service by specifying two flags: urgent and reliable. The urgent flag demonstrates the importance of the packet with respect to other packets. When this flag is set, the MAC protocol starts transmission of this packet regardless of other queued packets or the energy expenditure of transmission. Likewise, setting the reliable flag dictates the MAC protocol to request an acknowledgment frame. The neighbour management operation of SP is used by both of the MAC and routing protocols to manage the neighbouring information. The neighbour management uses two data structures: neighbour table and message pool. The neighbour table maintains some information from both MAC and routing layers, e.g., link states, route costs and duty cycle of neighbours. The main function of the neighbour table of SP is that it provides a common repository for both MAC and routing protocols, which in turn saves the storage space of nodes, time and energy required to access the data. The Message pool is a data-structure that serves as a message-passing interface between the MAC and routing protocols. The message pool is a bi-directional interface. The control channel is used to pass the data from the routing to MAC and the feedback channel used to pass the data from the MAC to routing. The control interfaces are used to set the urgent and reliable flags while the feedback interfaces are used by a routing protocol to report whether the desired results are obtained or not, e.g., if the reliable transmission is performed or not. Additionally, the feedback interfaces are used to notify the routing layer about the status of the medium, e.g., existence of congestion or packet loss rate.

The main advantage of SP is that it provides a modular architecture that can work with different MAC and/or routing protocols or even a particular function of them. The developers of SP demonstrated this approach by simulating a MAC protocol that incorporates the clear channel assessment of BMAC into the CSMA-CA mechanism of IEEE 802.15.4 standard [16]. The second main advantage of SP is that it provides a lightweight structure of information sharing between the MAC and routing layers. However, the main drawback of the SP protocol is that it leaves many issues open, e.g., the addressing scheme, time stamping and security policy which hinders its implementations. More importantly, as the SP protocol is limited to the MAC and routing layers, addition of a layer may require a reconstruction of the whole protocol or modifying it in such a way that negates its advantages.

3-8-2 Routing enhanced medium access control protocol

Routing enhanced Medium Access Control (RMAC) [118] is a cross layer protocol that utilises the routing information to regulate the access of the shared media. The main aim of RMAC is to mitigate the end-to-end delay of packets resulted from different duty cycles of different nodes along the path of packet. The RMAC protocol aligns the sleep/wake periods of all nodes within the path from the originator to the final destination by sending control frame prior to data packets. The RMAC protocol employs the IEEE 802.11 CSMA-CA scheme and replaces the RTS/CTS packets by the Pioneer frame (PION). The PION packets contain information from both of the MAC and routing layers, e.g., the length of data packet and the addresses of packet originator, final destination and next hop. The Time To Live field of the PION is set to zero by the originator of packets and is increased by each node within packet, each next hop consults its routing table, selects the appropriate neighbour and forwards the PION packet.

RMAC divides the time axis of the shared channel into three intervals: synchronisation, data and sleep periods. In the synchronisation periods all nodes within the network synchronise their clock cycle. The data period is used to announce the readiness of data packet transmissions while the sleep period is used to transmit the actual data packets. During the data period a node that wishes to transmit a packet waits for a random time in additional to the DIFS period during which a node sense the channel. If the channel is found idle then a node transmits a PION packet which traverses along the path from the originator to the final destination. In the sleep periods, all nodes that have responded the PION packet remain awake to forward data packets while all other nodes can go to sleep to save their energy. The RMAC protocol provides another means to save energy of originator and intermediate relaying nodes, as each node along the path of data packet can turn its transceiver off and go to sleep as soon as it transmits the data packet and receives the acknowledgment frame from the next hop.

The main advantage of the RMAC protocol is its ability to reduce the end-to-end delay by transmitting a data packet within a single duty cycle. Another advantage of this protocol is that it can prolong the lifetime of the network as it provide flexible duty cycle schemes that suit different traffic patterns. However, the main limitation of the RMAC protocol is the high collision probability which is a direct consequence of the fact that RMAC does not have a tidy schedule for the transmission time in the sleep periods. The Demand Wakeup Medium Access Control (DWMAC) [119] protocol has been developed to overcome the collision limitation of RMAC. DWMAC alleviates this situation by providing a one-to-one mapping between data and sleep periods, this mapping is computed through a function that uses the starting time of both of these periods as references with respect an identifier that conveyed in a new control packet dubbed thresholding (SCH) frame. Other functions of DWMAC are similar to their peers in the RMAC protocols, which in turn enhance the performance of the DWMAC protocol compared to the RMAC. However, the downside of DWMAC is that it requires a perfect synchronisation amongst all nodes within network which is difficult to achieve over large scale WSNs.

3-8-3 Receiver based auto rate protocol

Receiver Based Auto Rate (RBAR) [120] is a cross layer protocol that exploits the conditions of physical layer to adjust the data rate of transmission. This protocol employs the IEEE 802.11 CSMA-CA protocol with the RTS/CTS mechanism. In the RBAR protocol a node measures the signal strength of the received packet and then uses it to find the maximum possible data rate of the intended communication and encodes this value in the CTS packet. A transmitter node, when receiving the CTS packet, selects the modulation scheme that is appropriate for requested data rate. The authors of the RBAR demonstrate the advantage of using the cross layer approach by comparing RBAR with IEEE 802.11 RTS/CTS mechanism, and the results show that the RBAR can improve the throughput of network significantly.

A recent study [86] uses the RBAR over a multihop networks. This study employs the Destination Sequenced Distance Vector routing protocol (DSDV) which is based on the minimum hop count metric. The results reported in the latter study show that using the DSDV routing protocol over the RBAR cross layer protocol halves the throughput of networks compared to the case when the DSDV is used over the original IEEE 802.11 protocol. This reduction in the throughput of the DSDV/RBAR is attributed to the fact that the DSDV protocol routes packets over the minimal number of hops or equivalently DSDV selects the routing path which has the longest distance between intermediate relaying nodes. As the long distance reduces the signal strength of RTS packets, RBAR always uses the modulation scheme that provide a low data rate which in turn reduces the throughput of DSDV/RBAR compared to DSDV/IEEE 802.11 protocols. This characteristic highlights the unintended consequences of the cross layer approach.

3-9 Motivation for using inferential statistics to design MAC, routing and duty cycle protocols

This section is devoted to discuss the main approaches used to design the MAC and routing and their cross layer protocols in order to point out their advantages and disadvantages in the domain of WSNs and to provide a motivation for this thesis.

In terms of the MAC protocols, it can be seen that CSMA-CA as a distributed mechanism with low overhead is appropriate for most of the WSNs applications, i.e., CSMA-CA does not require a node to be equipped with a special transceiver (as the case in FDMA), does not impose great overheads on the network (as is the case in TDMA) or does not assign each node a unique encoding scheme (as is the case in CDMA). However, CSMA-CA suffers from high collision probabilities and considerable packet delays which are primarily due to the limited design space of most of the existing CSMA-CA protocols. In particular, most of these protocols require all nodes to generate their back-off intervals using a uniform distribution with the same width and to double this width after each collision or channel assessment failure. Hence, the time required to service a packet increases multiplicatively based on the history of unsuccessful transmissions or failed channel assessments. Operating such a scheme over multihop WSNs makes the situation even worse, as in these networks, each node has distinct traffic characterises that differ substantially from others which in turn requires each node to contend to access the channel differently. Furthermore, in multihop networks the collision probability of each transmitter–receiver pair differs significantly than others as each transmitter and receiver nodes have different number of neighbours and these neighbours contend to access the channel according to different traffic loads. As a result using the same contention procedures by all nodes as the case with existing CSMA-CA scheme can be a source of performance degradation. Motivated by these limitations, this thesis proposes a novel MAC protocol that re-engineers the CSMA-CA scheme by replacing the uniform distribution with the Gamma distribution. The main aim of using the Gamma distribution is to provide a flexible means to generate the back-off intervals that can reflect the distinct contention characteristic of each node. The proposed CSMA-CA protocol also specifies an intelligent collision resolution mechanism that can identify the cause of the collision and remediate them instead of just doubling the back-off width blindly as the case in traditional CSMA-CA protocols. Interestingly, the proposed protocol exploits the principles of inferential statistics to allow a node to adjust the parameters of the Gamma distribution according to prediction for the status of the channel. The main advantage of the proposed protocol is that it reduces the end-to-end delay of packets as it allows a node to access an idle channel with high probability which in turn reduces the service time of packets and the associated contention energy. Moreover the proposed protocol increases the throughput of the network due to the intelligent collision resolution mechanism.

Reviewing the literature related to duty cycle management schemes demonstrates that most of these schemes either require nodes to be fitted with special equipment (as is the case with asynchronous or geographical cycle schemes) or impose overhead (as is the case with synchronous duty cycle schemes). As both of these approaches demand high energy consumption, the amount of energy saved by them is reduced significantly by the additional overhead or equipment required to maintain their operations correctly. Another main limitation of most duty cycle schemes is that they are devised as a part of a MAC or routing protocol which in turn makes their adaption to other configurations infeasible in some cases. Motivated by these limitations, this thesis proposes a novel duty cycle management scheme that can operate independently of routing or MAC protocols and that is applicable to a general multihop network without restrictions on the topology configurations, traffic patterns or routing policies. Basically, the proposed scheme enables a node to determine the time and durations of both sleeping and waking intervals individually using a nonparametric Bayesian inference technique in which each node infers the potential communications and adapts its duty cycle without need to use a control packet or special equipment. This approach is derived based on the metaphors between the wireless communication networks and chemical reaction networks. The main benefit of the proposed scheme is that it imposes no overheads on the network which in turn saves energy consumed in exchanging control packets and eliminates collisions between these control packets or between them and data packets. This in turn leads to further energy saving, reduction in the end-to-end delay of packets and magnified in throughput of network.

Finally, it can be seen from reviewing literature of routing protocols that the main challenge of developing an effective routing protocol for WSNs is that determination of the optimal routes requires fresh and accurate information to be gathered about all possible routes to the intended destination. However, obtaining such information imposes a great overhead which in turn can lessen the lifetime of nodes, reduce the throughput of network and increase the end-to-end delay of packets. Most of the routing approaches mitigate the effect of overheads either by reducing the size of routing information (as is the case with link state routing protocols) or by letting a node collect the routing information only when a route is required (as is the case with reactive routing protocols). However, such reduction increases the route acquisition latency problem. Motivated by these limitations, this thesis proposes a novel routing protocol that formulates routing as a solution for multi-objective optimisation problem in which each node attempts to route its packets over those paths that maintained the global optimisation for the entire network. The proposed protocol exploits one of the state-of-the-art optimisation algorithms, called the Harmony Search (HS) algorithm, to develop the routing protocol. The proposed protocol mimics the thinking strategy and logic reasoning of a musician ensemble during improvisation of the most pleasing harmony. Moreover, the proposed protocol is lightweight, scalable and highly adaptive as it allows each node to infer the routing metrics from the characteristics of its neighbour and utilises the principle of the spatial reasoning to guide a node to discover those areas of networks that presumably yield the optimal or near optimal paths. In additional, the routing protocol provides an error-correction mechanism that reduces the possibility of routing packets over suboptimal paths. The main advantage of the proposed protocol is that prolong the lifetime of nodes, reduce the end-to-end delay of packet and increase the throughput of network

3-10 Conclusion

This chapter has introduced MAC, routing, duty cycle schemes as well as their cross layer protocols with the aim to explore the fundamental logic of these protocols. Based on this exploration, some example protocols are reviewed in order to highlight their main contributions and limitations. This chapter has also highlighted the main motivation for the work presented in this thesis.

Chapter 4

Simulation Techniques and Validation Methods

4-1 Introduction

The main objective of this thesis is to develop medium access, routing and duty cycle management schemes for WSNs that can leverage the performance of these networks without consideration for the specific topology configurations, traffic patterns or routing policies. Towards this goal, there is a need to address the methods and techniques that can be used to assess the performance of the proposed protocols and to collect and analyse the data.

In general, there are three performance prediction methods [121]: real testbeds, analytical models and simulation tools. Although the real testbed is the method by which comprehensive evaluation can be carried out, it is difficult to use it in the research domain. The main reason is that WSN testbeds consume a lot of time and money, need to be updated continuously in order to reflect changes that are proposed during the design phase; and they require long periods of time to collect the data. On the other hand, analytical models mitigate the main shortcomings of using testbeds by quantifying the interaction parameters of WSNs mathematically. However, the usability of the analytical models depends to great extent on deriving rigorous mathematical systems that can capture the real behaviour of WSNs without affecting the solvability of the systems. The last performance prediction method is simulation which virtualises the WSNs over computer systems in such a way that eliminates the adversity of real testbeds and eradicates the complexity of analytical models.

Notwithstanding the performance evaluation methods being used, verification and validity of a model is an important factor to ensure accuracy of results [121-122]. Furthermore, the methodology of collecting and analysing data is the final important step toward qualification of the results. This thesis attempts to present a high quality research study, hence it follows strict methodologies to ensure that any proposal or result is assessed carefully and well attributed. Therefore, the analytical models for predicting the operation of medium access control protocols and energy consumption performance are developed and then used in conjunction with the simulator. This not only enhances the integrity of the results but also facilitates investigation of the protocols from different perspectives.

This chapter focuses on the simulation techniques while the analytical models are left for the subsequent chapters. The chapter starts by reviewing some of the platforms that are developed for WSNs. Thereafter, the NS3 simulator [123] that is selected to assess the performance of the proposed protocols is considered in more detail. The main reason for selecting NS3 is that it provides a reliable simulation tool that is widely used.

The remainder of this chapter is organised as follows. Section 4-2 introduces the simulation tools and considers some of the well-known WSN simulation platforms. Section 4-3 outlines the main feature of the NS3 simulator as the selected simulation platform and section 4-4 provides an overview for the simulation scenarios and result validation methodology. Finally section 4-5 concludes this chapter.

4-2 Overview of the simulation techniques in WSNs

The key benefit of using a simulator to design WSN protocols is that it provides a flexible and affordable tool that can assess the performance of a network under different conditions, from different aspects and without the need for real deployments. Broadly, most current WSN simulation platforms are based on the discrete events approach. This approach represents a simulated network as two time-dependent vectors: event and state [121-122]. The event vector maintains a chronological list of the actions that have to be executed by the simulation while the state vector keeps track of the variables that describe performance figures. Management of the event and state vectors and their interactions is the main responsibility of the simulation engine. Different aspects related to credibility, usability, extensibility and interoperability have to be considered when selecting the appropriate simulation platforms. Owing to diverse applications of WSNs and the complexity of building a simulator that is suitable for all applications, different platforms have been developed.

The following subsection explores some of the WSN simulators in order to highlight their underlying approaches before considering the selected simulator in more detailed in the next section.

4-2-1 OPNET

OPtimised Network Engineering Tool (OPNET) [124] was the first commercial network simulator launched by the Massachusetts Institute of Technology (MIT) in 1987 to help companies to diagnose or reorganise their networks. OPNET is a discrete event simulator that represents a network using a three level hierarchy consisting of the network, node, and process levels. The network layer is the area where the topology is defined, the node level defines the physical characteristics of nodes and finally the process level is the deepest level which defines the protocol operations using Finite State Machines (FSMs) and the Proto-C language [125]. In addition, the OPNET simulator provides a number of standard protocol libraries such as Transmission Control Protocol/Internet Protocol (TCP/IP), Asynchronous Transfer Mode (ATM), IEEE 802.11, Fibre Distributed Data Interface (FDDI) and many others. OPNET simulates wireless channels using a fourteen-stage radio link model called the radio transceiver pipeline and supports a number of indoor and outdoor propagation models. Moreover, OPNET features several tools to animate and analyse the outcomes of simulation that are capable to illustrate the live operations of the simulated network and to plot graphs of the different performance metrics such as packet delay, buffer occupancy, packet loss ratio and throughput.

The main advantage of the OPNET simulator is that it provides Graphical User Interface (GUI) tools to simulate and analyse performance of communication networks with comprehensive documentation. However, the main drawback of OPNET is related to its cost.

4-2-2 TOSSIM

TOSSIM is an acronym for TinyOS SIMulator [126]; it is a free and open source software developed by University of Californian at Berkeley to test, debug and analyse the TinyOS codes of MICA family Motes. The TOSSIM simulator is a bit level emulator that runs on UNIX machines or Microsoft Windows via Cygwin and uses the nesC language to code the simulation modules. In TOSSIM, all sensor nodes have identical code images whereas the physical components of motes are replaced with their counterparts software coding; the simulator engine uses a multivariate vector to manage the interactions amongst nodes. In terms of wireless propagation model, TOSSIM represents a network as a graph where a pair of nodes is connected via unidirectional links. Each link is associated with a probabilistic bit error rate which quantifies the possibility with which a node receives packets successfully from the linked node.

The main advantage of the TOSSIM emulator is that its results match closely their peers in the real testbeds [127]. However, this advantage is weighed against that these results are restricted to specific node platforms.

4-2-3 Castalia

Castalia is open source and free software that was developed by the National Information and Communication Technology of Australia (NICTA) [128] with the aim of providing a powerful simulator platform for low-power embedded networks e.g. WSNs and Body Area Networks (BANs). One of the main design objectives of Castalia is to provide realistic propagation models and towards this goal Castalia defines a map for temporal variations of path losses between nodes within the networks. The readings of path losses can be obtained either from empirically measured data or using theoretical models. Castalia provides most of the well-known modulation schemes (e.g. phase shift keying and frequency shift keying) and enables developers to define a custom modulation scheme by importing the signal noise ratio-bit error rate curves into the simulation engine. The Castalia simulator comes with built-in real radio modules e.g. CC2420 and also enables a user to specify custom radio parameters by specifying different states, their corresponding power levels and the time required to switch between them [129]. Moreover, the Castalia developer team provides implementations for a number of MAC and routing protocols for WSNs including the IEEE 802.15.4, TMAC, Baseline BAN and tree based routing protocols. Other protocols can be written in C++ language. In order to enhance the credibility of the outcomes of Castalia, a number of modules have been packaged into the simulation kernels such as sensor manger, mobility manager and resource managers. The sensor manager defines the number of sensor devices per node, the power consumed during sensing and the distribution of the number of sensing events per unit time. The mobility manager defines the movement pattern of nodes and finally, the resource manager keeps tracks of the energy consumption and the clock drift of sensor nodes. In terms of the result analysis, Castalia provides rudimentary built-in tools for manipulating and displaying results either in tabular or graphical forms.

The key advantage of Castalia is that it provides a highly parametric simulation engine developed especially for WSNs and that enables developers to modify different parameters at runtime without need to recompiling the modules. However, the main drawback of the Castalia simulator is that it is not a standalone simulator but rather it operates on the OMNeT++ simulation engine platforms [128].

4-2-4 NS2

Network Simulator 2 (NS2) [130] is open source and free software that is distributed under the GNU General Public License. NS2 was developed under the Virtual InterNetwork Testbed (VINT) project. This project was funded in 1997 by the Defense Advanced Research Projects Agency and conducted by several research communities including the University of California at Berkeley, Xerox PARC and the University of Southern California. The main aim of VINT was to develop a robust and evolvable software tool that can investigate the interactions of network protocols and engineer them without needing for real deployments. Owing to the long evolution period and the large online support society of NS2, it has become one of the most widely used and highly acceptable simulators.

The underlying approach of NS2 is to utilise the Object Oriented Programming (OOP) in building an effective simulation engine [131]. Basically NS2 specifies the main components of the simulator as a set of classes, e.g., simulator class, packet class and node class. Each one of these classes is associated by several attributes and a number of member functions. NS2 exploits powerful aspects of OOP such as inherences, encapsulation and polymorphism to enhance the code reusability and maintainability of both the simulation codebase and the developed modules. Besides these advantages, employing the OOP approach leverages the extendibility and longevity of the NS2 simulator as new modules can be developed and integrated with others seamlessly and without need for significant modification. Another key feature of the NS2 simulator is the use of dual languages; simulation routines and modules are coded in C++ whereas the script of simulation scenarios are coded in Object Tool command language (OTcl). The main advantage of using dual languages is to reduce the overhead associated with recompiling the simulated modules under different scenarios. However, the downsides of using dual languages are that it increases computational complexity and reduces the usability of NS2. In addition, NS2 provides an emulation capability that enables a simulation engine to execute developed protocols over real networks attached to the host where the NS is running. NS2 provides a flexible means to analyse and visualise the simulation outcomes. The basic idea is that to generate standard file formats that stores simulation session. These files are interpreted by software packets to display the simulated results. Hence a number of analysis tools for NS2 have been introduced; some examples are visual trace analyser, network Animator and Xgraph [130-131].

Despite the many advantages offered by NS2, it has a number of drawbacks. Firstly, the high computation complexity required to organise different classes and to manage their interactions. Secondly, the long learning curve which results from the fact that developers are required to use two different languages. Thirdly, NS2 suffers from inadequate documentations that describe its source code or internal structures.

4-3 NS3

NS3 is free and open source software that was launched in 2006 as a joint research project by a number of institutes including Inventors for the digital world, Georgia Institute of Technology and the University of Washington [132]. The NS3 simulator is not an extension for NS2 but rather a new platform that aims to provide a powerful simulation tool suitable for modern communication networks. Hence substantial enhancements as well as a plenty of new features have been introduced by NS3.

NS3 improves the scalability aspect of NS2 by reorganising the underlying structure of the class hierarchy [133]. NS3 is built as a set of C++ libraries that are combined together as well as with other external libraries at runtime to form the simulation engine. These libraries can be classified into four main categories: core, simulation, common and node libraries. The core library specifies the functions of the simulator kernel such as debugging functions, random number generator, smart pointers and callbacks. The simulator library is used to control the time, schedulers and event objects and common library contains the public objects that can be called by other libraries such as tracing and the monitor objects. Finally the node library defines the classes that are used to represent the simulated networks such as node, net devices and communication channel. Moreover, NS3 utilises modern concepts of OOP to enhance robustness of interactions between classes. Dynamic-casting is one of these concepts that is used to manage the access between subclasses and their base classes dynamically without need for explicit operations. The obvious advantage of dynamic-casting is that it can ease the development process and avoid runtime errors.

Smart pointer is another mechanism that is employed by NS3 to improve the garbage collection process. The smart pointer keeps track of memory locations of objects and de-allocates them automatically when all of the caller functions do not further need them. Using the smart pointer mechanism mitigates the memory lack problems which in turn can leverage the performance of the simulator. Aggregation is the one of the highly sophisticated mechanism that is adopted by NS3 to reduce the tightly coupling between different classes. The aggregation mechanism enables a subclass of certain type to be associated with other type subclasses dynamically at runtime. The main benefit of aggregation is that it reduces the overhead imposed by nesting different type classes. The NS3 simulator overcomes the limitation of NS2 related to use of dual language by employing C++ exclusively to code both protocols modules and simulation scripts. Another key enhancement of the NS3 simulator is the automatic documentation that is provided by the Doxygen tool. This tool parses the source code to identify the comments and definitions of variables and parameters, generates the corresponding documents and links them to other documents. The key result of this approach is that the documents of NS3 are consistent and up to date.

NS3 uses the node library to abstract real networks as a set of classes each of which represents a specific component of simulated networks [134]. The four main base classes of NS3 are: node, application, channel and net device. The node class represents the communication points of the simulated networks, e.g. wireless sensor nodes or base stations. The application class represents the user programmes that generate the activities of the simulated scenarios, such as the generation of packets. The third main class of NS3 is the channel with mimics the medium that carries the communications between nodes. NS3 supports varies channels of both wired and wireless communications e.g., fibre optic, Ethernet as well as wireless media. Finally, the net device class defines the communication interfaces that are responsible to connect nodes to corresponding channels. Some examples of the net device classes that are supported by NS3 is the Ethernet network interface controller and radio transceivers. The NS3 simulator employs the aggregation approach to establish the relations between the aforementioned classes in a similar way that a real network is constructed. For example, the net device classes are aggregated to the node classes to imitate connecting of transceivers to sensor nodes. This approach provides a richly modular abstraction in which a class can be aggregated to more other classes, for example, multiple applications classes can be aggregated to the same node in order to model different traffic patterns.

Figure 4-1 illustrates the NS3 high level abstractions for a network consisting of i nodes communicate over n channels using m applications. It is worth noting that all the interactions amongst applications, net devices and protocol stack (where the protocol modules are coded) as carried out through the interfaces manager as shown in this figure. The interfaces manager provides Application Interfaces (APIs) of standard formats supported by most operating systems. The main benefit of the interfaces manager is twofold: firstly to enhance the emulation capability of the simulator by facilitating the integration between the simulation kernel and real network. Secondly to increase the interoperability between different modules implemented by different developers.

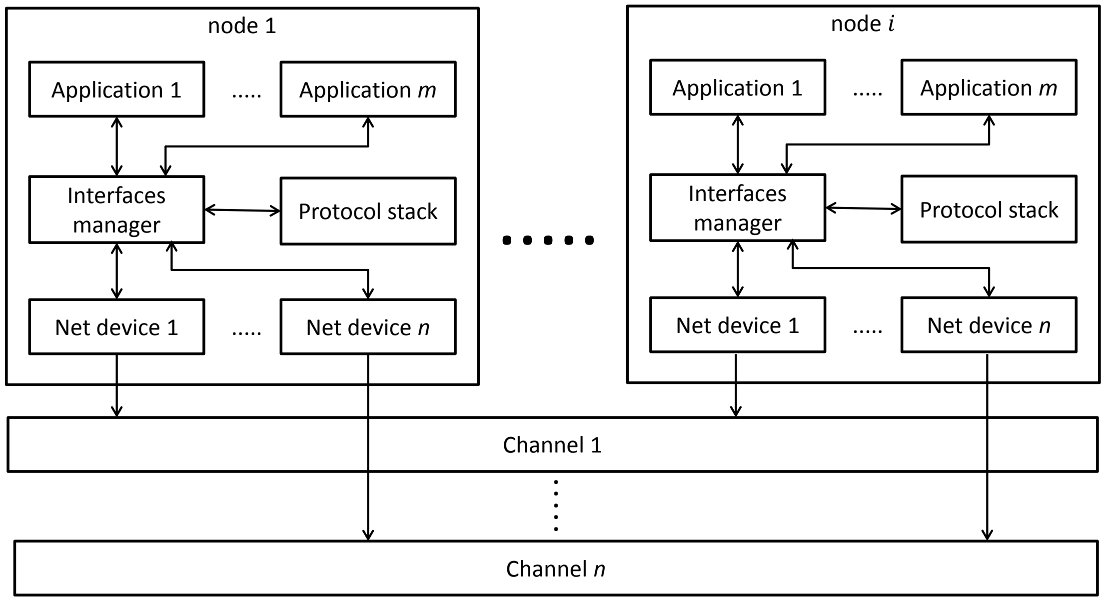


Figure 4.1 High level abstraction for the base classes of the NS3 simulator

The NS3 simulator provides a flexible means [130] to build protocol stack; basically, developers create the required classes and then use the low-level APIs provided by the interfaces manager to define interactions between classes. However, as this approach can consume considerable efforts and time, NS3 features a helper classes to aid developers in building their modules faster. Helper classes comprise a set of subclasses and their associated APIs that is defined to perform a specific function, helper classes facilitates coding a protocol stack by passing the simulation parameters as arguments to these classes without need to use low-level APIs. NS3 provides extensive collections of helper classes that cover almost all core operation of the simulations as well as most of the well-known protocols. Some examples for helper classes are topology, CSMA, physical status helpers, IEEE 802.15.4 , IEEE 802.11, 6LowPAN, ADOV, OLSR helpers.

The NS3 simulator supports several propagation models [135] that are designed to meet different deployment environments. Roughly these models can be classified into two main categories. Firstly, deterministic propagation models that are based on electromagnetic wave theory such as free space, two-ray ground, Nakagami-m fast fading and log-distance path loss propagation models. Secondly, propagation models that are based on the empirical measurements gathered from real fields, some examples for this type are the COST 231-Hata, International Telecommunication Union ITU-1411 and Okumura propagation models. Additionally, the NS3 simulator allows developers to define their own propagation model or create hybrid propagation models by combining the deterministic with empirical models.

The NS3 simulator offers a robust energy module [136] that can account for the energy consumption of different components of sensor nodes including radio transceiver, processing and sensor units. Like many other components of the NS3 simulator, the energy modules are defined as a class. This class has just two members: the energy source and energy device. The energy source class defines the type of energy sources (e.g., batteries or energy scavenging sources) and their associated attributes (e.g., initial amount of energy and discharge characteristics). The energy device class defines the amount of energy demanding by each component of sensor node and their different states, (e.g., power consumed in each state of radio transceivers). The interactions between the energy source and energy device members are carried out through a number of APIs that can be called from within the same energy module or from other classes in protocol stack. The main benefit of separating the energy providing members from the energy consumption members is to maintain a high level of modularity as each member can be modified without disrupting others. Moreover this approach allows a node to aggregate more than energy sources and energy devices readily.

NS3 supports the NAM animation tool provided by NS2 and also introduces a new set of tools e.g. the PyViz and iNSpect [121,123]. The main difference between these tools and NAM is that these new tools are fed from the trace hooks defined in the common library, hence these tools provide live visualisation that enable developers to debug their modules at runtime. Other advanced feature offered by the NS3 simulator is the distributed simulation approach which enhances the scalability of the simulation by allowing execution of the simulation codes on multiple computing hosts. One of the design objectives of NS3 is to enhance the credibility of the simulation outcomes by minimising the discontinuities between their outcomes and results of testbeds. Toward this goal a number of emulation techniques have been introduced by NS3, e.g., Network Simulation Cradle (NSC) [121] and emulation net device. The NSC allows the actual protocol stack implemented on the host machine to be incorporated into simulated networks. The emulation net device technique works in the opposite direction by allowing the NS3 engine to use real networks in sending and reception of simulated packets.

4-4 Simulation scenarios and results validation

This section reviews the main characteristics of simulation scenarios and the methods of collecting and analysing the readings of these scenarios. In general, this thesis considers a number of sensor nodes that have identical equipment (radio transceivers, battery capacity, etc.) and are distributed uniformly over a given area to form a multihop network. The border effect is avoided by allowing nodes that are near the edge of the simulated area to communicate with the other nodes on the opposite sides. All the transmissions and receptions from nodes are performed with the same and fixed data rate and using the same power level that is obtained from a real data sheet [137].

This research employs an ideal propagation model with free space parameters. The main reason for this selection is that all the proposed protocols operate at the MAC and routing layers and hence incorporating a non-ideal model can hinder the discrimination between the effects of the proposed model and the effects of the randomness of the propagation model. Another reason for using an ideal propagation model is that WSNs are deployed over wide range of environments including indoor, outdoor and underwater. Since each one of these environments has its own characteristic that differ substantial than others, considering a particular propagation model can limit the applicability of the proposed protocols.

The physical parameters of the network (e.g., data rate, channel assessment mechanism and slot time) are adjusted according to the IEEE 802.15.4 physical specifications [16]. IEEE 802.15.4 is the official standard for low power wireless personal area networks which covers a wide communication systems and this standard has become an essential component of higher-level protocol stacks such as ZigBee [111] and 6LowPAN [17]. Hence adopting the physical layer parameters of the IEEE 802.15.4 standard in this thesis can widen the applicability of the proposed protocols and make them ready to be integrated with future standards.

Each simulation scenario is typified by four parameters $\langle k, N, X, Y, Z \rangle$ where the first two parameters (i.e., k, N) are used to characterise the topology of network, the second two parameters (i.e., X, Y) represent the traffic patterns and finally the Z parameter defines the routing policy.

The parameter k defines the connectivity degree of the network, i.e., the average number of mutually independent paths between non-neighbouring nodes. The parameter N defines the average number of nodes per a channel which is the product of the average number of nodes per a square metre by the average area of the transmission range. The main reason for specifying this parameter is to provide a meaningful quantity that describes both the topology and transceiver parameters. The values of N and k are set by adjusting the transmission range and the number of nodes per unit area in order to generate a variety of network topologies [138].

The parameter X defines the probability distribution that is used by nodes to generate their internal inter-arrival times between packets and the parameter Y specifies the rate of this traffic. This thesis uses three inter-arrival distributions, namely: Constant Bit Rate (CBR), Exponential (EXP) and Weibull (WBL). CBR represents the case when the intervals between consecutive packets are of a fixed length equals to the reciprocal of the traffic rate specified by the parameter Y. In the EXP case, the intervals between consecutive packets are random variables following exponential distribution whose rate is defined by the value of parameter Y. Finally, WBL represents the case when the intervals between consecutive packets are generated according to the Weibull distribution. The Weibull distribution is characterised by two parameters: shape and rate. The shape parameter is adjusted to low value (e.g. less than one) in order to obtain a heavy-tailed distribution, while the rate parameter is set to the value of parameter Y. The main reason for using these three inter-arrival distributions is to assess the performance of the proposed protocol under different traffic patterns. CBR is suitable for the periodic reporting applications while EXP accounts for the event driven applications and finally WBL represents the heavy tailed traffic patterns found in some networks such as the Internet of Things [9]. The length of packets that is generated by all of these distributions are set at the Maximum Transmission Unit (MTU) of the IEEE 802.15.4 standard which is 133 bytes unless other values are specified in a particular assessments.

Finally, the parameter Z defines the routing policy. This study employs two policies. The first policy distributes traffic evenly amongst all other nodes within the network. This policy is denoted by U and used to ensure that each node acts as a source, router and final destination simultaneously. The second traffic distribution policy is denoted by A and refers the case when the traffic that is generated by all nodes is destined to a single receiver selected randomly. This traffic distribution mode has been selected as it is suitable for cluster based WSNs. In chapters 6-7, the DSDV [105] routing protocol is used while chapter 8 proposes a novel routing protocol.

For each simulated scenario (tuple) fifty random topologies are generated and for each topology, the simulation sessions are repeated one hundred times each elapsing for 10^9 seconds. The readings of each scenario are averaged over all nodes within the networks and then the 95% confidence level with respect to the mean is considered in the results.

Other parameters that are closely related to the proposed protocols such as contention parameters of the CSMA-CA protocol or the routing metrics are stated in the corresponding chapters.

4-5 Conclusion

This chapter has reviewed the simulation techniques and result validation methods that will be used in this thesis. A review of some of the widely used simulation platforms demonstrates that the NS3 simulator offers appealing features which make it an excellent candidate to assess the performance of the proposed protocols in this thesis. This chapter has also provided an overview for the main characteristics of simulation scenarios and highlighted the methods of collecting and analysing the simulation outcomes. It can be seen from this overview that these configurations account for different cases which facilitates predicting the performance of the proposed protocols under wide variety of configurations.

Chapter 5

Analytical Framework for Statistical Analysis of CSMA-CA Protocols over Multihop WSNs

5-1 Introduction

The first main objective of this thesis is to design an effective MAC protocol for multihop WSNs that can achieve three concurrent goals: increasing the throughput of a network, minimising the end-to-end delay of packets and prolonging the lifetime of nodes. The discussion presented in chapter 3 demonstrates that the Carrier Sensing Multiple Access with Collision Avoidance (CSMA-CA) protocols as decentralised schemes with low overheads are suitable for WSNs. Moreover, it has been shown that the design spaces of most of the current CSMA-CA protocol are inadequate to accommodate the unique characteristics of multihop networks, e.g., heterogeneous traffic patterns and distinct channel conditions for each transmitter-receiver pairs. In order to reveal the main limitations of the current CSMA-CA protocols, and thereby to overcome them during the development of the proposed protocol, there is a need to acquire a deep and accurate understanding for the behaviour of CSMA-CA under different operational conditions.

An exact and complete understanding of the CSMA-CA protocol can be acquired through statistical analysis as such analysis can abstract the stochastic behaviour of the protocol as a probability distribution. Therefore all the possible states of the system can be predicted readily and moreover the design of CSMA-CA protocols can be accomplished without missing information. Another great advantage of the statistical analysis is its ability to quantify the relationships between hidden variables that affect the performance of the CSMA-CA protocol and that cannot be derived from other analytical or simulated models: e.g., the effect of variation of the back-off on the channel assessment [6].

In general, analysing the CSMA-CA protocol statistically is a challenge, and this challenge increases tremendously when the analysis targets multihop WSNs as these networks constitute a highly random system. Explanations for the randomness of a typical multihop WSN can be drawn by considering the following facts. Firstly, the lack of a de-facto topology that can represent wide deployments of multihop WSNs, and thus a typical representation of a multihop WSN is a random planar process. Secondly, the heterogeneity of the traffic load of nodes, as in a multihop WSN the traffic intensity of a node comprises the traffic generated by the node itself as well as the traffic forwarded by other nodes for relaying. Thus, the traffic intensity of a node is a time-spatial random process. Finally, as a typical multihop WSN is a large-scale network in which each node has a different number of neighbours and each node has its own traffic intensity, the view of the channel conditions of each transmitter–receiver pair differs substantially from others. Moreover due to the low-data rate of WSNs, a packet encounters a high amount of both queuing and medium access delay, thus the contention between nodes becomes subject to the randomness of the topology, traffic and service time. Accounting for these factors in analysing the statistical characteristics of CSMA-CA protocols over a multihop WSN requires a new paradigm that is able to reflect these different levels of randomness.

This chapter proposes a novel approach to analyse the statistical characteristics of CSMA-CA protocols over multihop WSNs [6]. The proposed approach represents the multihop WSNs as a three layered model: a topology model, a routing model and a queuing model. The topology model abstracts the underlying structure of the network using set theory to describe the physical relations between nodes, e.g., transmission range and carrier-sensing range. The routing model quantifies the traffic load of each node based on the originator-final destination traffic rates and the multipath routing policy. The queuing model represents a node as a GI/G/1 queue [139] in which the inter-arrival distribution is constructed from the routing model. The service time distribution of a queue is derived from the characteristics of the CSMA-CA protocol considering the physical constraints from the topology model; this model then uses queuing theory to obtain the other quantities, including the distributions of inter-departure and waiting times. The benefit of the proposed modelling approach is threefold: firstly, it facilitates modelling a multihop network as a network of queues considering the effects of both the physical and routing layers; secondly it enables assessment of the performance of a MAC under a wide variety of situations just by modifying the corresponding models without requiring reconstruction of the entire analytical framework; and finally, it presents a general framework that can be easily adapted for use with any other medium access control protocols (either contention based or scheduled based) by varying the service time distribution. Using the proposed framework, the Moment Generating Functions (MGFs) of the probability distributions of the MAC service time, inter-departure, and waiting times are obtained. From these distributions, comprehensive assessment of the behaviour of the CSMA-CA protocol is derived including different performance metrics (average end-to-end delay, probability of dropping a packet and throughput). In order to demonstrate the accuracy of the proposed analytical framework, we use it to analyse the IEEE 802.15.4 CSMA-CA protocol and then compare its results with simulation outcomes. The main aim of this comparison is not only to validate the accuracy of the proposed model but also to highlight the main limitations of the existing CSMA-CA protocol and enable us to overcome them when developing the proposed MAC protocol. Moreover, this framework constitutes the building block for the other analytical model presented in the subsequent chapters.

The remainder of the chapter is organised as follows: in section 5-2 the key existing analytical models and their limitations are summarised, section 5-3 overviews the statistical theorem used in deriving the analytical model and section 5-4 presents the proposed analytical framework. Section 5-5 uses the proposed analytical model to evaluate the performance of the IEEE 802.15.4 CSMA-CA protocol and Section 5-6 presents and discusses the results. Finally, section 5-7 concludes the chapter.

5-2 Overview of the existing analytical models

Evaluating the performance metrics of Medium Access Control (MAC) protocols including CSMA-CA has been previously proposed in the literature using various assumptions and techniques. However, most of these models are inadequate to analyse the behaviour of the CSMA-CA protocols over multihop WSNs. This section reviews the key techniques that are used in modelling these protocols; this reviewing demonstrates limitations of these models and highlights motivations for the proposed model.

5-2-1 Throughput-offered traffic ($S - G$) analysis

The throughput-offered traffic ($S - G$) is one of the earliest modelling techniques used to analyse the throughput of contention based MAC protocols, e.g., ALOHA and CSMA-CA [140]. This analysis is based on assumptions that the number of nodes sharing a single channel is infinite each of which is equipped with a buffer for just a single packet. In this approach, it is assumed that the sum of new and retransmitted packets offered to the channel by all nodes “offered traffic” follows a Poisson distribution with rate G packets per unit of time. This approach estimates the throughput of the network under steady state conditions as the fraction of offered traffic load that is transmitted successfully over the channel, i.e., $S = P_{suc}G$, where P_{suc} is the probability of successful transmission. The value of P_{suc} is determined by computing the probability that no packets are generated during the vulnerable time of transmission of another packet.

The obvious limitation of this approach is its oversimplifying assumptions. In the practical cases, the number of nodes is finite and each one can generate its traffic based on the application requirements which may differ from a Poisson distribution, for example, the case with a constant bit rate that is used in periodic monitoring WSNs. Moreover, the assumption that a node has a buffer enough for a single packet means that a node cannot generate a new packet while it contends to access the channel which makes $S - G$ inappropriate for computing the queuing delay. The other key limitation of this technique is the assumption that the offered traffic always follows a Poisson distribution regardless of effects of the MAC protocol. All of these limitations hinder applying the $S - G$ approach to analysis of multihop WSNs.

5-2-2 Markov chain analysis

This technique represents the operational states of a MAC protocol as a state space in a Markov chain. The main challenge of using Markov analysis is how to select the state space that represents accurate characteristics of the protocol while maintaining the tractability and solvability of the chain. The pioneer work of using Markov chains to model contention based MAC protocols was introduced in [141] to investigate the performance of a slotted ALOHA with a single buffer packet where a node has only two states: backlogged and thinking. In the backlogged state a node contends to access the channel, while in the thinking state a node waits for a random period before generating a new packet. In this model, the number of nodes is used as the state space which yields a single dimension Markov chain.

The increased complexity of modern MAC protocols increases the state space of their corresponding Markov chain model, which in turn requires developers to use simplifying assumptions in order to maintain tractability. For instance, the model presented in [142] analysed the CSMA-CA protocol specified in the IEEE 802.11 DFC standard [79]. This study assumed a finite number of nodes sharing a single channel where all nodes are within transmission ranges of each other and all nodes have visibility of all other nodes, so hidden node collisions are neglected. Furthermore, this model assumed that a collided packet is retransmitted until it is successfully received. Moreover, the study considers just the saturated region (where a node generates a new packet as soon as it finishes servicing the current one). Based on these assumptions, a two dimensional Markov chain is constructed to represent the back-off stage and back-off counters of a single node while the interactions between nodes (i.e., channel modelling) is determined by another Markov chain. The main outcome of this model was to compute the maximum saturation throughput without consideration for other performance metrics, e.g., delay, jitter and probability of dropping a packet. However, the unempirical assumptions used in this model (e.g., the saturated traffic condition and ignoring the hidden node collisions) impede its applicability over multihop WSNs. Extension for this model to account for real cases such as unsaturated traffic condition, limited retransmission mode or hidden node collision have been introduced in some work, e.g., [143-147]. However, such extensions require using three- or four- dimensional Markov chains which in turn increases the modelling complexity.

Comparing Markov analysis with $S-G$ analysis demonstrates that the former technique can resolve the oversimplification assumptions of $S-G$ e.g., infinite number of nodes. However, the main drawback of Markov chain analysis is that its complexity is increased exponentially with the sophistication of the modelled protocol.

5-2-3 Network of queues

This technique has been introduced by Jackson [148] to analyse the performance of multi-programming computer systems. In this analysis, processors are modelled as an open queuing network (i.e., where jobs can enter or leave the queuing networks anywhere). This approach provides a product form solution for the number of jobs in each queue and then uses it to determine the performance metrics of networks.

However, this computation is feasible only when the inter-arrival and service time of nodes follow exponential distributions.

Due to this limitation, a number of approximations for the Jackson networks have been proposed, and these approximations can be classified into five categories: diffusion approximation [149], mean value analysis [150], operational analysis [151], exponentialisation approximations [152] and decomposition method [153]. The diffusion approximation analyses queuing networks at heavy traffic situation by approximating the inter-arrival and service distribution as a normal distribution. The mean value analysis approximates the length of queues for just a closed queuing network, where no external traffic enters the network. The operational analysis derives a mathematical model based on premises that are assessed practically, e.g., using testbeds or simulation outcomes. The exponentialisation approach approximates the general service time distribution as the exponential distribution and then applies the Jackson equations. The decomposition approximation decomposes the network into a set of individual queues, analyses each node individually and then investigates the relations between these queues.

Applying the network of queue's approach to predict the performance of CSMA-CA protocols requires substantial modifications. Firstly, in network queuing theory there is no consideration of losing a task while it moves from a queue to other. Conversely, in wireless networks, a packet (task) can be lost during transmission due to collision or channel conditions. Secondly, queuing network theory assumes that the service time distribution of a queue is independent of the service time of other nodes. In contrast, the service time of node (queue) operating CSMA-CA protocols is affected by the behaviour of other nodes due to the contention between them. A third reason that impedes using the network queuing theory in modelling CSMA-CA protocols directly is the lack of closed-form solutions to the medium access delay (service function) of the CSMA-CA protocols.

However, some works provide approximations towards this goal. In [154], the authors use the exponentialisation approach to approximate the distribution of medium access delay as an exponential distribution. This model shows some inaccuracy compared with simulation outcomes. In [155], the authors use the simulation results to compute the variance of the medium access delay and then approximate the distribution of the service time using the Edgeworth approximation method to estimate the service distribution. In [156] the authors used the moment matching approach to approximate the distribution of medium access delay based on just the first two moments which are obtained from simulation. The key limitation of these approaches is that the accuracy of the distribution depends on the length of the simulation runs. The model proposed in [157] uses the diffusion approximation to analyse the performance of the IEEE 802.11 protocol, however, this study assumes that the traffic as well as the service distributions follow a normal distributions which is impractical.

The aforementioned discussion highlighted the fact that most of the modelling techniques used to predict the performance of the CSMA-CA protocols cannot reflect the random characteristics of multihop networks e.g., ad-hoc topology configurations, heterogeneous traffic patterns, multipath load balancing and queuing delay. Motivated by this lack and the importance of developing a rigorous analytical model for designing and development of WSNs, this chapter is devoted to provide such modelling framework. The proposed framework is built on accurate assumptions that represent real and general cases and with novel methodology and results [6]. In terms of realistic assumptions, it is assumed that each node has its own unique traffic load and that each transmitter–receiver pair has its own distinct view of the channel conditions. Moreover, the proposed framework relaxes the widely-made assumption that the probability of channel assessment is independent of the length of the back-off windows and is distributed uniformly over the whole service time by the more accurate assumption that the probability of assessment of the channel varies from one node to another as well as between different back-off stages. In order to facilitate this computation, a number of statistical theorems are employed. These theorems are reviewed in the next section.

5-3 Overview of statistical theorems used in the analytical framework

This section summarise the key statistical theorems that will be used in deriving the proposed analytical framework.

5-3-1 Moment generating function

A moment Generating Function (MGF) is defined as a reflected bilateral Laplace transform of a Probability Mass Function (PMF) [158]. If X is a discrete random variable whose PMF is given by $F_X(x) = P[X = x]$; $\forall x \in \mathbb{R}$, then its MGF is denoted here by $x(t)$ and given as:

$$x(t) = E\left[e^{tx}\right] = \sum_{x=-\infty}^{\infty} e^{tx} F_X(x) \tag{5.1}$$

Where $E[\,]$ is the expectation operator and t is a dummy variable whose exponents are placeholders for the coefficients of the PMF.

The MGF has been selected due to many reasons, firstly, the MGF possesses a uniqueness property which means that if the MGF of a distribution exists then its probability mass function also exists, is unique, and can be determined by an inverse Laplace transform or by extracting the ordered coefficient of the MGF. Secondly, other transforms that characterise probability distributions can be derived from the MGF just be replacing the dummy variable in equation (5.1), e.g., the probability generating function is derived by replacing the e^{tx} by t and the characteristic function of a distribution [158-160] is determined by substituting e^{tx} by e^{itx} where i denotes the imaginary unit. This property enables us to express the results in multiple forms which in turn facilitate investigation of performance from different view of points. More importantly, by using the MGF, all moments of a distribution can be found easily just by differentiating the MGF with respect to t and evaluates the result at $t = 0$. Let μ_x^k represent the k^{th} moment of a random variable X whose MGF is $x(t)$ then μ_x^k can be expressed as:

$$\mu_x^k = E\left[X^k\right] = \frac{d^k}{dt^k} x(t)\bigg|_{t=0} \tag{5.2}$$

Finally, as the MGF is a based on the Laplace transform, it inherits its linear property which facilities adding two independent random variables just by multiplying them. Let $X_1, X_2, X_3, \ldots, X_n$ be n independent random variables whose MGFs are given respectively as $x_1(t), x_2(t), x_3(t), \ldots, x_n(t)$, and let further the sum of these random variables be denoted by $Y = \sum_{i=0}^{n} X_i$. Then the MGF of Y which is denoted by $y(t)$ is given as:

$$y(t) = \prod_{i=1}^{n} x_i(t) \tag{5.3}$$

5-3-2 Compound probability distribution

While most random variables are generated according to well-defined probability distributions, e.g., uniform distribution, normal distribution, etc. In some cases there is a need to repeat generation of a random variable that follows a specific distribution a number of times, this number of times is itself another random variable that follows another distribution. The resultant distribution is called a compound probability distribution [161].

In mathematical notations, let N be a random variable following a certain distribution and $X_1, X_2, X_3, \ldots, X_N$ be mutual independent and identically distributed (i.i.d.) random variables that follow another distribution. Further let N be independent of all X_i's, then the sum of X_i denoted by S can be given as:

$$S = X_1 + X_2 + X_3 + \ldots. + X_N \tag{5.4}$$

In statistical parlance, the PMF of S is called the compound distribution, the PMF of N is called the primary distribution and the PMF of $X_i's$ is known as the secondary distribution. In the domain of moment-generating functions, let $n(t)$ and $x(t)$ be the

MGFs of the primary and secondary distributions respectively and $s(t)$ be the MGF of the compound distribution then equation (5.4) becomes [161]:

$$s(t) = n(x(t)) = n(\ln(x(t))) \tag{5.5}$$

5-3-3 Mixture probability distribution

The mixture probability distribution theorem [161] is used to derive the probability distribution of a collection of random variables that are generated according to different distributions and mixed together according to certain weights. In mathematical notation, suppose that there are N random variables, $X_1, X_2, X_3, \ldots, X_N$ each of which is generated according to different probability distributions. These N random variables are mixed according to specific weights where a random variable X_i is mixed with a weight w_i. Then the sum of these N random variables is called a mixture random variable, denoted by S, and given as:

$$S = \sum_{i=1}^{N} w_i X_i \tag{5.6}$$

The condition that all weights equal one, i.e., $\sum_{i=1}^{N} W_i = 1$, is required to maintain a proper distribution for the random variable S. However, if this condition is not satisfied then the legitimation techniques can be used to derive a proper distribution. Using the notation of an MGF, let $s(t)$ be the MGF of the resultant random variable S, and $X_i(t)$ be the MGF of underlying distributions, then $s(t)$ can be expressed as:

$$s(t) = \sum_{i=1}^{N} w_i x_i(t) \tag{5.7}$$

5-4 Proposed analytical framework

The main contribution of this chapter is to propose a novel approach to analyse CSMA-CA protocols over multihop WSNs that can account for general cases without restrictions on the topology configurations, traffic patterns or routing policies. The proposed approach derives accurate results based on studying these networks over a three layered framework comprising a topology model, a routing model and a queuing model, each of which captures the specific characteristics of multihop network. A conceptual diagram for this approach is illustrated in figure 5.1.

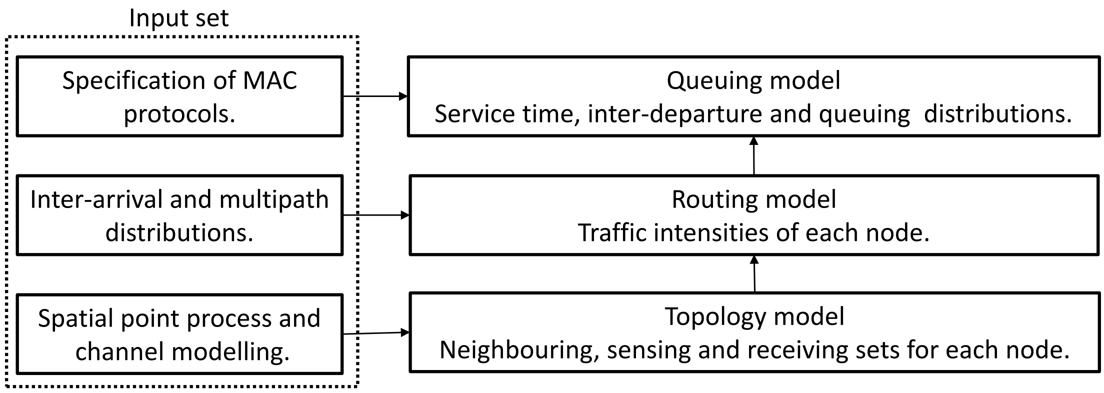


Figure 5.1 Conceptual diagram for the proposed analytical framework

The benefit of the proposed framework is threefold: firstly, it facilitates modelling of a multihop WSN as a network of queues considering the effects of both the physical and routing layers. Secondly, the proposed framework enables assessment of the performance of the protocol under a wide variety of situations just by modifying the corresponding models without requiring reconstruction of the entire analytical framework, Finally, this model presents a general framework that can be easily adapted for use with any other MAC protocol (either contention based or scheduled based) by varying the service time distribution. Subsection 5-4-1,5-4-2 and 5-4-3 discuss the topology, routing and queuing models respectively while subsection 5-4-4 describes the system perspective for the framework.

5-4-1 Topology model

The topology model is the first building block of the proposed framework; the main objective of this model is to describe the underlying structure of the multihop network mathematically using set theory. Thereby, the routing and queuing models can recognise the relationships between different nodes readily.

This chapter considers a multihop network consisting of a set of nodes Υ scattered uniformly over a given area. All of these nodes are equipped with identical transceivers whose transmission range is fixed at (TR m) and whose carrier sensing range is (CS m) where $TR \leq CS$. The capture effect is not considered so if two packets are overlapped for any period then both are considered lost. It is assumed further that the network is connected and the path between any pair of nodes is known. A packet is routed based on the shortest path scheme (hop count scheme); if there is more than one path between a pair of nodes that have the same length and these paths do not have a common intermediate node (mutually independent paths) then the load is distributed between them according to a given multipath weighting factor. We define $\|n - x\|$ as the Euclidean distance between nodes n and x and for each node the following sets:

1. $\Omega(n)$ is a set of nodes that can correctly decode the signal received from node n in the absence of any distortion, i.e. $\Omega(n) = \{x \in \Upsilon : \|n - x\| < TR\}$.
2. $\Psi(n)$ is a set of nodes that share the channel with node n, thus node n before starting a transmission has to ensure that there is no ongoing transmission by any member of this set, i.e. $\Psi(n) = \{x \in \Upsilon : \|n - x\| < CS\}$.
3. $\Lambda(n)$ is a set of nodes such that the transmission of any member can be received successfully by node n in the absence of collisions, i.e. $\Lambda(n) = \{x \in \Upsilon : \Omega(x) \cap \{n\} = \{n\}\}$. It is worth noting that both $\Omega(n)$ and $\Lambda(n)$ can be treated equally in the case that the ideal physical channel is considered (i.e., links between nodes are bidirectional). However, these two sets are defined distinctly in order to maintain the proposed framework general to represent different channel models.
4. $P(n,u) = \{P_1(n,u), P_2(n,u), .., P_x(n,u)\}$ is a set containing all the mutually independent shortest paths from node n to node u; each member of $P(n,u)$ is given as $P_x(n,u) = \{n, n_1, n_2, .., u\}$ in which n_i is the next hop of node n_{i-1}.The mutually independent condition implies that $P_x(n,u) \cap P_y(n,u) = \{n,u\}, \forall \{P_x(n,u), P_y(n,u)\} \in P(n,u)$.

5-4-2 Routing model

The routing model is the second layer in the proposed analytical model; the main aim of this model is to quantify the traffic intensity of each node. The routing layer utilises the collections of sets that are provided by the topology model with the inter-arrival distributions of packets and the multipath load balancing policy to determine the traffic load of each node.

It is assumed that the MGF of the inter-arrival times of packets that are generated internally by node n at its application layer is $a_n^i(t)$ and the rate of this traffic is λ_n^i packets per slot. It is assumed further that $\lambda_{nu}^{\mathrm{i}}$ is the fraction of the traffic generated by node n and destined to node u as a final destination over all possible paths between these nodes where $\sum_{\forall u \in \Upsilon} \lambda_{nu}^i = \lambda_n^i$. Further let $\omega_{P_x(n,u)}^i$ be a multipath weighing factor defined as the fraction of λ_{nu}^i that is sent over the specific path $P_x(n,u)$, where $\sum_{\forall P_x(n,u) \in P(n,u)} \omega_{P_x(n,u)}^i = 1$; this ensures that the sum of the weighting factors over all paths between any two nodes is unity. Let $\xi(P_x(n,u), v)$ be the probability that a packet sent over the path $P_x(n,u)$ is received successfully by node v where v is an intermediate relaying node in the path. If $v = u$ then $\xi(P_x(n,u), u)$ represents the probability that a packet sent over the path $P_x(n,u)$ is received successfully at its final destination (node u in this case). As these probabilities depend on the characteristics of the MAC protocol, their values that correspond to the IEEE 802.15.4 CSMA-CA protocol are given in equation (5.32) and (5.33) respectively. Employing the aforementioned parameters facilitates quantifying the traffic flow using the following probabilities:

1. The probability that the external traffic received successfully by node n has node

n itself as final destination is denoted by r_n and is the ratio between the aggregated traffic received successfully to node n from all nodes (numerator) and the aggregated traffic that is received successfully at node n either as a final destination or to be forwarded to other nodes (denominator).

$$r_n = \frac{\sum\limits_{\substack{\forall y \in \Upsilon \\ \forall P_x(y,n) \in P(y,n)}} \left(\lambda_{yn}^i \omega_{P_x(y,n)}^i \xi\left(P_x(y,n), n\right) \right)}{\left(\begin{array}{l} \sum\limits_{\substack{\forall y \in \Upsilon \\ \forall P_x(y,n) \in P(y,n)}} \left(\lambda_{yn}^i \omega_{P_x(y,n)}^i \xi\left(P_x(y,n), n\right) \right) \\ + \sum\limits_{\substack{\forall y,z \in \Upsilon \\ \forall P_x(y,z) \in P(y,z)}} \left(\sum\limits_{\substack{\forall n \in P_x(y,z) \\ n \neq y, n \neq z}} \lambda_{yz}^i \omega_{P_x(y,z)}^i \xi\left(P_x(y,z), n\right) \right) \end{array} \right)} \tag{5.8}$$

2. The probability that the traffic departing from node n is destined to node u as the next hop $u \in \Omega(n)$ is denoted by r_{nu} and given as the ratio between the total traffic sent over the link between node n and u (numerator) to the sum of the traffic generated internally by node n and the external traffic routed by n (denominator).

$$\begin{aligned} r_{nu} &= \frac{\sum\limits_{\substack{\forall y,z \in \Upsilon \\ \forall P_x(y,z) \in P(y,z)}} \left(\sum\limits_{\forall (n,u) \in P_x(y,z)} \lambda_{yz}^i \omega_{P_x(y,z)}^i \xi\left(P_x(\mathrm{y},z), n\right) \right)}{\lambda_n^i + \sum\limits_{\substack{\forall y,z \in \Upsilon \\ \forall P_x(y,z) \in P(y,z)}} \left(\sum\limits_{\substack{\forall n \in P_x(y,z) \\ n \neq y, n \neq z}} \lambda_{yz}^i \omega_{P_x(y,z)}^i \xi\left(P_x(\mathrm{y},z), n\right) \right)} \\ &= \left(E\left[A_n\right]\right)^{-1} \left(\sum\limits_{\substack{\forall y,z \in \Upsilon \\ \forall P_x(y,z) \in P(y,z)}} \left(\sum\limits_{\forall (n,u) \in P_x(y,z)} \lambda_{yz}^i \omega_{P_x(y,z)}^i \xi\left(P_x(\mathrm{y},z), n\right) \right) \right) \end{aligned} \tag{5.9}$$

5-4-3 Queuing model

The queuing model represents the top layer of the proposed analytical framework. In this model each node is represented as a GI/G/1 queue [139] in which both inter-arrival and service distributions follow general probability distributions; the queues are first-come, first-served.

For an arbitrary node n, the MGF of the inter-arrival distribution of packets is denoted by $a_n(t)$. This function is composed of two parts: firstly, the inter-arrival distribution of packets that are generated internally by node n with MGF of $a_n^i(t)$; and secondly, the inter-arrival distribution of packets received successfully from all neighbours of node n and which need to be forwarded to other nodes, here denoted by $a_n^e(t)$. As these distributions are statistically independent [163], the MGF of the resultant inter-arrival distribution is given as:

$$a_n(t) = a_n^i(t) a_n^e(t) \tag{5.10}$$

The service function of this queuing model is also general as it represents the distribution of the time that is spent in servicing a successful packet. The MGF of this distribution is $s_n(t)$ and its derivation is described in section 5-5-7. Based on the values of $a_n(t)$ and $s_n(t)$, the MGF of the inter-departure distribution of node n denoted by $d_n(t)$ can be approximated as [163-167]:

$$d_n(t) = \rho_n s_n(t) + (1 - \rho_n) s_n(t) a_n(t) \tag{5.11}$$

where ρ_n is the utilisation factor of node n, (i.e., $\rho_n = E[S_n]/E[A_n]$) and $E[S_n]$ is the average time required to service a successful packet, as this parameter depends on the specification of MAC protocol, its value that is corresponding to the IEEE 802.15.4 CSMA-CA protocol is given in equation (5.36).

Packets departing from node n are split into multiple traffic flows and distributed amongst the node's neighbours according to probabilities r_{nu} , $\forall u \in \Omega(n)$. The MGF of the inter-departure time of packets that depart from node n and are destined to node u, denoted by $d_{nu}(t)$ can be given as [167-168]:

$$d_{nu}(t) = \frac{r_{nu} d_n(t)}{1 - (1 - r_{nu}) d_n(t)} \tag{5.12}$$

Following the same notations, the MGF of the inter-arrival times of the external distribution (i.e., packets that are received successfully by a node n and need to be serviced (forwarded further)) which is denoted by $a_n^e(t)$ is the aggregated traffic received by node n from all other nodes except that destined to node n itself. Recalling that r_n is the probability that the received traffic is destined to node n itself enables us to express the value of $a_n^e(t)$ as:

$$a_n^e(t) = \frac{(1 - r_n)\left(\prod_{\forall x \in \Lambda(n)} d_{xn}(t)\right)}{1 - r_n\left(\prod_{\forall x \in \Lambda(n)} d_{xn}(t)\right)} \tag{5.13}$$

Substituting equation (5.13) into equation (5.10) yields the MGF of the aggregated inter-arrival distribution of each node. In this chapter, we assume that the internal inter-arrival distribution $a_n^i(t)$, probabilities of routing traffic r_{nu} and r_n and multipath weighting factors $\omega^i_{P_x(n,u)}$, $\forall P_x(n,u) \in P(n,u)$, $\forall n, u \in \Upsilon$, are given, while the value of $s_n(t)$, $E[S_n]$ and $\xi(\mathrm{P_x}(n,u), v)$ are derived in section 5-5-7.

5-4-4 System perspective

From a system perspective, a multihop network is considered as a network of queues, in which the behaviour of each node is affected by the physical constraints described in the topology model and the traffic distribution obtained from the routing model. It is assumed that the network is operating in an equilibrium region which is defined by the following two conditions:

1. The utilisation factor of each node is less than or equal to one, i.e., $\rho_n \leq 1\ ; \forall n \in \Upsilon$. This is required to prevent buffer overflow at a node.
2. The aggregated traffic sent over a channel is less than or equal to the data rate of the network; this ensures that the network does not reach an oversaturated condition.

These conditions facilitate employing the main assumption of the tagged user analysis [169-171] which concludes that a network of interacting queues can be decoupled to a collection of independent queues when each queue operates at its equilibrium probability, where equilibrium probabilities quantifies mutual dependence between nodes due to the shared resources. Considering that in a network operating CSMA-CA protocols the interactions between nodes are carried out through the conditions of channels implies that the equilibrium probability can be described in terms of the probability of assessing the channel. Hence this chapter assumes that the probability of assessing the channel as idle by node n written here as α_n is constant. It is worth noting that this assumption constitutes the key assumption for the majority of models presented in the literature. Moreover, proving the existence of the equilibrium region for CSMA-CA networks has been the subject of a number of previous works, e.g. [172-175], while investigations for the equilibrium region in general GI/G/1 systems has been introduced in [176]. Table 5.1 summarises the key symbols used in the proposed analytical framework.

Table 5.1 Key symbols of the proposed analytical framework

Symbol	Definition
X	Random variable X.
$x(t)$	Moment Generating Function (MGF) of the probability distribution of random variable X.
$E[X]$	Expectation of the random variable X.
$V[X]$	Variance of the random variable X.
D	Data rate of the network in bits per slot.
Υ	Set containing all nodes in a MPAN.
N	Average number of nodes per a single channel.
L	Random variable represents the length of packet in slot.
$a_n^i(t)$	MGF of the inter-arrival of packets generated by node n itself.
$a_n^e(t)$	MGF of the inter-arrival of packets received successfully by neighbours of node n and required forward further.

$a_n(t)$	MGF of the aggregated inter-arrival of packets at node n.
λ_n^i	Rate of packets that are generated internally by node n, given in packets per slot.
λ_{nu}^i	Rate of packets that generated internally by node n and destined to node u as a final destination over all possible paths, given in packets per slot.
$P(n,u)$	All possible mutually independent shortest paths from node n to node u.
r_{nu}	Probability that the traffic departing from node n is intended for node u as the next hop.
r_n	Probability that traffic reaching node n is destined for node n itself as a final destination.
$s_n(t)$	MGF of time required to service a successful packet by node n.
$d_n(t)$	MGF of the inter-departure time of packets of node n.
$q_n(t)$	MGF of total delay (both for queuing and servicing) time of packets at node n.
ρ_n	Utilisation factor of node n.
ξ_n	Average probability that a packet that is sent by node n to a neighbour is received successfully.
α_n	Probability that node n assesses the channel as idle.
(n,u)	Ordered pair indicates that node u is the next hop of node n, where $u \in \Omega(n)$.

5-5 Performance evaluation of the IEEE 802.15.4 CSMA-CA protocol

This section employs the proposed analytical model to assess the performance of the IEEE 802.15.4 beaconless CSMA-CA protocol [16] over multihop WSNs. This protocol has been selected as it is the first official standard that has been ratified especially for WSNs and has been used widely in both of the academic and industrial disciplines [13, 21-23]. The key objective of this assessment is not only to exemplify the proposed analytical framework and to evaluate its integrity but also to reveal the deep characteristics of IEEE 802.15.4 CSMA-CA protocol and to quantify the relationships between different parameters that affect its performance over multihop WSNs. More importantly, the outcomes of this assessment facilitate overcoming the limitations of this protocol when designing an effective CSMA-CA mechanism for multihop WSNs in the next chapter.

The remaining of this section is organised as: subsection 5-5-1 introduces the IEEE 802.15.4 CSMA-CA protocol. Subsection 5-5-2 derives the MGF of the probability distribution of service time, and the probability of assessing the channel and probability of an idle channel is computed in subsections 5-5-3 and 5-5-4 respectively. The probability of transmission, collision and MGF of the time required to service a successful packet is given in subsections 5-5-5, 5-5-6 and 5-5-7 respectively and finally subsection 5-5-8 computes the performance metrics.

5-5-1 Overview of the IEEE 802.15.4 CSMA-CA protocol

The IEEE 802.15.4 beaconless CSMA-CA protocol [16], like other CSMA-CA protocols, requires a node to assess the channel before commencing transmission. However, it has been optimised to save energy by allowing a node to stay idle during the back-off periods and assesses the channel only at the end of these periods. This protocol can be modelled by the state transition diagram shown in figure 5.2 in which there is a group of states: the back-off states (denoted by B) where a node stays idle for a random period before assessing the channel; the channel assessment states (denoted by C) where a node assesses the channel to determine if it is free or occupied with another transmission; and the transmission state (denoted by T) where a node transmits its packets.

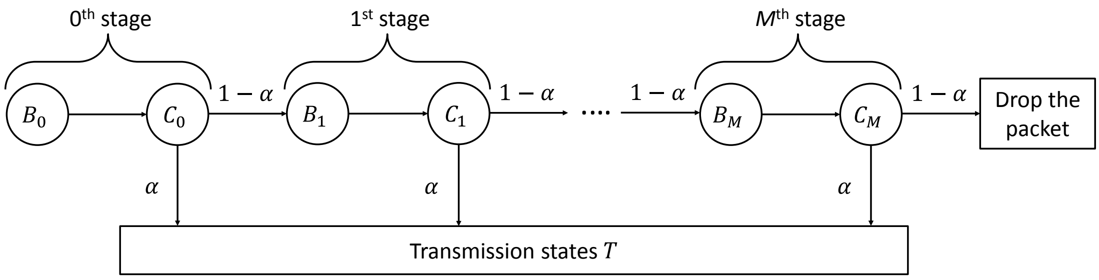


Figure 5.2 State transition diagram of IEEE 802.15.4

All durations of these states are measured in a time unit referred to as a *slot* where the length of a slot is 320 μs for the 250 kbps data rate used in this study. A node that has a packet to send starts the CSMA-CA process by backing-off for an integer random number of slots (denoted by B_0 in figure 5.2) uniformly distributed between 0 and $W_0 - 1$ slots (where W_0 is eight) and then assesses the channel for one slot (denoted by C_0 in figure 5.2). We refer to the total period comprising both the 0$^{\text{th}}$ back-off and channel assessment period as the 0$^{\text{th}}$ stage. If a node finds a free channel after this random back-off period B_0 then it commences transmission in the next slot. Otherwise (if the channel is found busy) a node backs-off for another random number of slots (denoted by B_1 in figure 5.2) uniformly distributed between 0 and $W_1 - 1$ slots (where W_1 is sixteen) before assessing the channel again (denoted by C_1 in figure 5.2). This back-off period and channel assessment period comprise the 1$^{\text{st}}$ stage. In order to save energy, the back-off and channel assessment cycle is repeated until either a node finds an idle channel and then starts transmission, or exceeds the number of back-off stages specified by the parameter M (which has a default value of four) [16]. Based on this method, the CSMA-CA protocol can be modelled as $M + 2$ stochastic stages where the m^{th} stage consists of a single slot for the channel assessment and a random period representing the back-off durations uniformly distributed between 0 and $W_m - 1$ where $W_0 = 8$ slots, $W_1 = 16$ slots and $W_2 = W_3 = W_4 = 32$ slots. The transition between the m^{th} stage and the transmission state occurs when a node assesses the channel as idle in that stage; this occurs with probability α as illustrated in figure 5.2, while transition to the next stage $m+1^{th}$ occurs with the complement of this probability.

5-5-2 MGF of the probability distribution of service time

Here, $z_n(t)$ represents the MGF of the probability distribution for the time (measured in slots) required to service a packet from the instance that a node starts the 0th back-off stage until it either finishes the transmission or drops the packet due to exceeding the maximum number of attempts to transmit the packet. This distribution is affected by two sources of randomness: the time that a node spends in each stage (back-off, channel assessment and transmission if possible) and the status of the channel when a node assesses it. As a result, $z_n(t)$ can be found using the theorem of mixture probability distributions [161] where the distributions to be mixed are those distributions of the time that a node spends in each stage, and the mixture weighting is the probability of assessing the channel as idle. Let $b_m(t)$ be the MGF of the back-off period that a node spends in the m^{th} stage, $c(t)$ be the MGF of the time spent in the channel assessment and $l(t)$ be the MGF of the time spent in transmission (the MGF of the length of packets). By examining the state transition diagram of the CSMA-CA process as shown in figure 5.2, we find that there are $M+2$ possible cases for servicing a packet: $M+1$ components refer to paths that end with packet transmission after a node finds an idle channel and the final case occurs when a node always assesses the channel as busy and drops the packet. Table 5.2 summarises these cases, their probabilities and MGFs.

Table 5.2 MGF of time spent to service a packet and their corresponding probabilities

Case ID	Event	Probability	MGF
1	Transmit a packet in the 0th stage.	α_n	$b_0(t)c(t)l(t)$
2	Transmit a packet in the 1th stage.	$\alpha_n(1-\alpha_n)$	$b_0(t)b_1(t)\big(c(t)\big)^2 l(t)$
3	Transmit a packet in the 2th stage.	$\alpha_n(1-\alpha_n)^2$	$b_0(t)b_1(t)b_2(t)\big(c(t)\big)^3 l(t)$
…	…	…	…
$M+1$	Transmit a packet in the Mth stage.	$\alpha_n(1-\alpha_n)^M$	$\left(\prod_{m=0}^{M} b_m\right)\big(c(t)\big)^{M+1} l(t)$
$M+2$	Drops a packet due to exceeding M busy channel assessments.	$(1-\alpha_n)^{M+1}$	$\left(\prod_{m=0}^{M} b_m\right)\big(c(t)\big)^{M+1}$

The first case represents the event of transmitting a packet at the 0th stage which occurs with probability α_n (i.e. when a node finds the channel idle at the first assessment); in this case the time required to service a packet is the random number of slots uniformly chosen between 0 and W_0-1 before the channel assessment, plus one slot for the channel assessment and L slots for the transmission of the packet. As these variables are pairwise independent, the MGF of the total time in this case is the product of their individual MGFs which is given as $b_0(t)c(t)l(t)$. The second case represents the situation of servicing a packet at the 1st stage which occurs with probability $\alpha_n(1-\alpha_n)$; this happens when a node assesses the channel as busy at the 0th stage and then assesses the channel as idle at the 1st stage. In this case, the total time spent is the sum of two random back-off periods in the 0th and 1st stages, plus two slots for channel assessments and L slots for packet transmission. The MGF of the total time required in this case is the product of their individual MGFs: $b_0(t)b_1(t)(c(t))^2l(t)$. The remaining $M+1$ cases can be found using the same methodology. It is worth noting that the last case accounts for the situation in which a packet is dropped; this occurs when the number of busy channel assessments exceeds M; the probability of this happening is $(1-\alpha_n)^{M+1}$ and the length of packet is not considered in this term. Summing all of these functions according to their probabilities gives the following general equation:

$$\begin{aligned} z_n(t) &= \alpha_n b_0(t)c(t)l(t) + \alpha_n(1-\alpha_n)b_0(t)b_1(t)\left(c(t)\right)^2 l(t) + .. \\ &+ \alpha_n(1-\alpha_n)^M b_0(t)b_1(t)...b_M(t)\left(c(t)\right)^{M+1} l(t) \\ &+ (1-\alpha_n)^{M+1} b_0(t)b_1(t)...b_M(t)\left(c(t)\right)^{M+1} \end{aligned} \tag{5.14}$$

In the IEEE 802.15.4 protocol, the values of $b_m(t)$ and $c(t)$ can be given as:

$$b_m(t) = \frac{e^{tW_m}-1}{W_m\left(e^t-1\right)} \tag{5.15}$$

$$c(t) = e^t \tag{5.16}$$

Equation (5.15) can be obtained by applying the definition of the MGF to the uniform distribution while equation (5.16) reflects the fact that the channel assessment period is constant and equal to one slot, therefore their MGFs are just to shift the service function. Substituting equations (5.15) and (5.16) together with default values of the IEEE 802.15.4 protocol ($M=4, W_1=2W_0$ and $W_2=W_3=W_4=4W_0$) into equation (5.14) gives the MGF of the service function in the IEEE 802.15.4 as:

$$z_n(t) = l(t)\left(\sum_{m=0}^{M} \frac{\alpha_n(1-\alpha_n)^m e^{t(m+1)}}{W_0^{m+1}\left(e^t-1\right)^{m+1}} \prod_{j=0}^{m} \frac{\left(e^{tW_0 2^{\min(2,j)}}-1\right)}{2^{\min(2,j)}}\right. \\ \left. + \frac{\left(e^t(1-\alpha_n)\right)^{(M+1)}}{W_0^{M+1}\left(e^t-1\right)^{M+1}} \prod_{m=0}^{M} \frac{\left(e^{tW_0 2^{\min(2,m)}}-1\right)}{2^{\min(2,m)}}\right) \tag{5.17}$$

Using the MGF enables us to find the moment of a distribution just by differentiating it with respect to t and equating it to 0. Therefore, the average and variance of the service time are given respectively as:

$$E[Z_n] = z_n'(t)\big|_{t=0} = \left(1-(1-\alpha_n)^{M+1}\right)E[L] + \frac{1}{2}\sum_{m=0}^{M}(1-\alpha_n)^m\left(1+W_0 2^{\min(m,2)}\right) \tag{5.18}$$

$$V[Z_n] = \left(z_n''(t)\big|_{t=0}\right)^2 - z_n'(t)\big|_{t=0} \\ = \frac{V[L]}{12}\left(\sum_{m=0}^{M}\left(\alpha_n(1-\alpha_n)^m\right)^2 - (m+1) + \sum_{j=0}^{m} W_0^2\left(2^{\min(2,j)}\right)^2\right) \\ - \frac{(1-\alpha_n)^{2M}\left((M+1)\right)}{12}\sum_{m=0}^{M} W_0^2\left(2^{\min(2,m)}\right)^2 \tag{5.19}$$

Equation (5.18) and (5.19) express the average and variance of the service time using the default values of the IEEE 802.15.4 protocol in terms of the initial back-off window and the incremental policy of the back-off.

The MGF of the service time given in equation (5.17) is a function of the probability of assessing the channel as idle α_n hence there is a need to find the value of this probability. The probability that node n assesses the channel as idle α_n is the joint probability of two events: the first event is that node n assesses the channel in an arbitrary slot and the second event is that the channel is idle during this slot. The probability of node n assessing the channel depends on the behaviour of the node itself, while the probability of the channel being idle depends on the activities of the other nodes that share the channel with node n. This interaction relation implies that evaluating the value of α_n requires an analysis of both component probabilities, and this can be achieved through the solution of simultaneous equations derived in subsections 5-5-3 and 5-5-4 below.

5-5-3 Probability of assessing the channel

In most of the previous work, it has been commonly assumed that the probability of assessing the channel is same for all nodes within a network, independent of the length of the back-off windows and is distributed uniformly over the whole service time e.g., [143-147]. Although this assumption reduces the complexity of modelling, it does not fully reflect the stochastic behaviour of the protocol. This chapter proposes a more realistic approach by assuming that the probability of assessing the channel varies between stages as well as differing from one node to another. Justification for the first part of this assumption is drawn from the fact that a node assesses the channel only after its choice of back-off duration has elapsed, and while this duration is selected randomly it is still dependent on the probability distribution of the back-off, which differs between stages. The second part of the assumption is justified by considering that a node assesses the channel only when it has a packet ready to transmit, and this depends on the traffic intensity that typically differs between nodes in multihop WSNs.

Let $\phi_n^{(m)}(k)$ be the probability that a node n assesses the channel in the k^{th} slots during the m^{th} stage (i.e., probability that node n is in the m^{th} stage and k slot and it assesses the channel), where k represents the number of slots that are elapsed since a node n commences servicing a packet, i.e., $1 \leq k \leq \sum_{i=0}^{M} W_i$ and m is an arbitrary stage, i.e., $0 \leq m \leq M$. Let us assume further that ϕ_n is the average probability that node n assesses the channel per the complete service time, hence the value of ϕ_n can be derived as a weighted sum of the moments of $\phi_n^{(m)}(k)$ as:

$$\phi_n = \sum_{m=0}^{M} (1-\alpha_n)^m \left(\frac{\sum_{k=1}^{W^{(m)}} \phi_n^{(m)}(k)}{W^{(m)}} + \sum_{f=2}^{F} \left(\frac{\sum_{k=1}^{W^{(m)}} \left(\phi_n^{(m)}(k) - \frac{\sum_{k=1}^{W^{(m)}} \phi_n^{(m)}(k)}{W^{(m)}} \right)^f}{W^{(m)}} \right)^{\frac{1}{f}} \right) \tag{5.20}$$

$$W^{(m)} = \left(\sum_{i=0}^{m} W_i \right); 0 \leq m \leq M$$

where $(1-\alpha_n)^m$ is a weighting factor that qualifies the stages of service time (as a node goes to the next stage only when it has assessed the channel as busy in the previous stage). The term $\frac{\sum_{k=1}^{W^{(m)}} \phi_n^{(m)}(k)}{W^{(m)}}$ is the average of all probabilities by which node n assesses the channel during the m^{th} stage. The term $\left(\left(W^{(m)}\right)^{-1} \sum_{k=1}^{W^{(m)}} \left(\phi_n^{(m)}(k) - \left(W^{(m)}\right)^{-1} \sum_{k=1}^{W^{(m)}} \phi_n^{(m)}(k) \right)^f \right)$ represents the f^{th} moment of the probability of assessment the channel during the m^{th} stage (i.e., the f^{th}

moment of $\phi_n^{(m)}(k)$) while the sum from 2 to F accounts for the first F^{th} higher moments of probabilities of assessment of the channel in a stage. Equation (5.20) expresses the average probability of assessing the channel based on the behaviour of the node itself which is the first equation required to obtain the value of α_n.

It is of interest to note that equation (5.20) is able to account for stochastic characteristics of the probabilities of channel assessment using the higher moments rather than the average value as is the case in most of the analytical models reported in the literature. Moreover, this equation is simple as it only requires computing the values of $\phi_n^{(m)}(k)$ in each stage which is determined in the rest of this section for the IEEE 802.15.4 protocol.

For the IEEE 802.15.4 protocol, in the 0^{th} stage, $\phi_n^{(0)}(k)$ refers to the probability that a node assesses the channel in the 0^{th} stage and slot k, where $1 \leq k \leq W_0$. As a node in this stage selects a slot to assess the channel randomly based on a uniform distribution between 1 and W_0, thereby $\phi_n^{(0)}(k)$ is:

$$\phi_n^{(0)}(k) = \begin{cases} \frac{1}{W_0}; & 1 \leq k \leq W_0 \\ 0; & \text{otherwise} \end{cases} \tag{5.21}$$

In the 1^{st} stage, the probability of assessing the channel in the k^{th} slot, denoted by $\phi_n^{(1)}(k)$, is the probability that a node assesses the channel in slot j where $0 \leq j \leq k-1$ in the 0^{th} stage and then backs-off for $k-j-1$ slots in the 1^{st} stage. For example, the probability that a node assesses the channel in the 3^{rd} slot in the 1^{st} stage is the sum of the probabilities that a node assesses the channel in slot 0, 1, or 2 in the 0^{th} stage, finds the channel busy, and then backs-off for 2, 1 or 0 slots respectively in the 1^{st} stage. An illustration of this concept is given in figure 5.3.

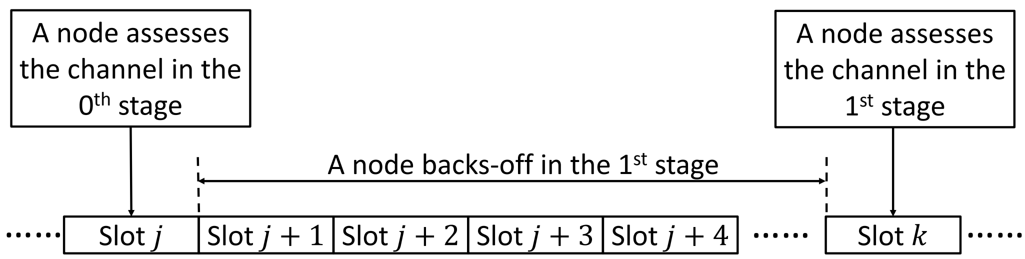


Figure 5.3 Relation between sensing and backing-off in the 0^{th} and 1^{st} stages

Accordingly, $\phi_n^{(1)}(k)$ can be given as:

$$\phi_n^{(1)}(k) = \sum_{j=1}^{j=k-1} \phi_n^{(0)}(j)\Phi(1, k-j-1); 1 \leq k \leq W_0 + W_1 \tag{5.22}$$

where $\Phi(m, j)$ is defined as the probability that a node selects to back-off for j slots at the beginning of stage m which is given as:

$$\Phi(m, j) = \frac{1}{W_m}; 1 \leq j \leq W_m; 0 \leq m \leq M \tag{5.23}$$

The methodology used to derive $\phi_n^{(1)}(k)$ can be generalised for the other stages, as the relationship between the channel assessment and back-off of any two consecutive stages follows the same procedure. This gives:

$$\phi_n^{(m)}(k) = \sum_{j=1}^{j=k-1} \phi_n^{(m-1)}(j)\Phi(m, k-j-1) \\ ; 2 \leq m \leq M;\ 1 \leq k \leq \sum_{i=0}^{i=m} W_i \tag{5.24}$$

Figure 5.4 depicts the probability of assessing the channel in the IEEE 802.15.4 protocol. It is shown that in the 0th stage, a node assesses the channel with a constant probability equal to $1/W_0$. This results from the fact that the probability of assessing the channel in this stage depends on the selection of a back-off period with a uniform distribution from one to W_0. In the remaining stages, the probability of assessing the channel differs significantly between slots, and this variation increases with an increasing number of stages. For example, $\phi_n^{(1)}(k)$ is an isosceles trapezoid function that results from the convolution between two independent uniform distributions: $\phi_n^{(0)}(k)$ and $\Phi(1, j)$ each having different ranges, while $\phi_n^{(4)}(k)$ is closer to a normal distribution which results from the central limit theorem.

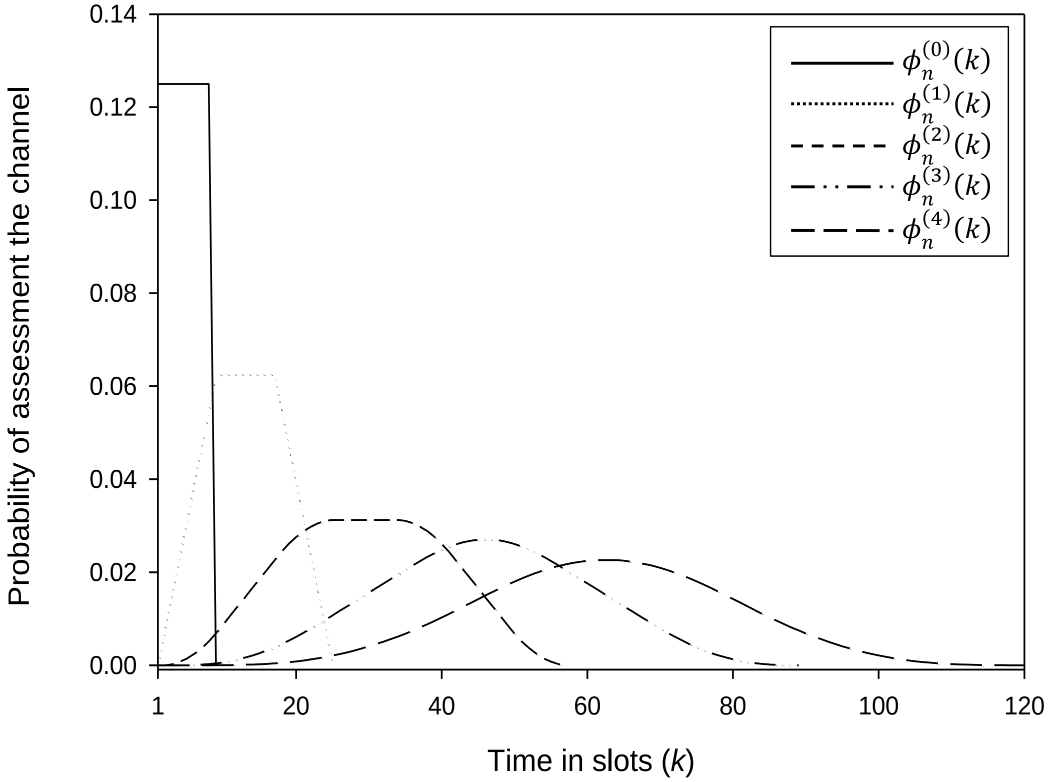


Figure 5.4 Probability of channel assessment in IEEE 802.15.4

Using the values of $\phi_n^{(m)}(k)$ shown in figure 5.4, the average probability of assessment the channel per service time ϕ_n can be derived using equation (5.20) up to predefined moments (by assigning the value for F in equation 5.20).

Figure 5.4 demonstrates one of the key limitations of the IEEE 802.15.4 CSMA-CA protocol, that is the probability of channel assessment in a back-off stage decreases considerably compared to the previous stages. This characteristic makes a node that has failed to access the channel in a specific back-off stage is less sensitive for the channel conditions in the subsequent stages. Hence a node can defer its transmission for longer periods even if the channel is idle which in turn leads to magnify the end-to-end delay of packets and to reduce the throughput of the network. Evaluations for the effect of this characteristic over multihop networks are presented in the results of this chapter.

5-5-4 Probability of an idle channel in an arbitrary slot

The probability that the channel of node n in an arbitrary slot k is idle is the probability that nodes sharing the channel with node n including n itself have not assessed the channel in the previous L slots, i.e.,

$$\Pr\left[\text{the channel of node } n \text{ is idle}\right] = \left(\prod_{\forall x \in \Psi(n) \cup \{n\}} \left(1 - \rho_x \phi_x\right) \right)^L \quad (5.25)$$

Equation (5.25) considers the necessary and sufficient conditions of the channel being idle, as if the channel was occupied with a transmission, it will finish after L slots, but if the channel was idle then if no node assesses the channel during these slots, the channel will remain idle. Based on this equation, the probability of assessing the channel as idle from the viewpoint of node n can be given as

$$\begin{aligned} \Pr\left[\text{node } n \text{ asseses idle channel}\right] = \Pr\left[\text{node } n \text{ assesses the channel}\right] \\ \times \Pr\left[\text{the channel of node } n \text{ is idle}\right] \end{aligned} \quad (5.26)$$

Using equation (5.25) and considering that node n senses the channel only when it has a packet to transmit, which happens with probability ρ_n, yields the following equation:

$$\alpha_n = \rho_n \phi_n \left(\prod_{x \in \Psi(n) \cup \{n\}} (1 - \rho_x \phi_x) \right)^L \quad (5.27)$$

Equation (5.27) is the second of the equations that is required to find the values of α_n ; this equation expresses the probability of assessing the channel based on the behaviour of neighbouring nodes. By solving equation (5.20) and (5.27) the values of α_n and ϕ_n can be obtained.

5-5-5 Probability of transmission

Let $\mathcal{T}_n$ be the event that node n does not transmit in a slot, $\mathcal{H}_n$ be the event that node n has a packet in a slot and let $\overline{\mathcal{T}_n}$ and $\overline{\mathcal{H}_n}$ are the complement of these events respectively. Then the probability that the node n does not transmit in a slot, written as γ_n , can be given using the law of total probability [177] as:

$$\gamma_n = \Pr[\mathcal{T}_n | \overline{\mathcal{H}_n}] \Pr[\overline{\mathcal{H}_n}] + \Pr[\mathcal{T}_n | \mathcal{H}_n] \Pr[\mathcal{H}_n] \quad (5.28)$$

The conditional probability $[\mathcal{T}_n | \overline{\mathcal{H}_n}]$ is one as a node never transmits when its queue is empty. The conditional probability that a node has a packet and does not transmit $\Pr[\mathcal{T}_n | \mathcal{H}_n]$ is the probability that a node does not access the channel, i.e., the complement of all probabilities by which a node can access the channel in different stages, hence the value of $\Pr[\mathcal{T}_n | \mathcal{H}_n]$ can be given as:

$$\begin{aligned}\Pr[\mathcal{T}_n|\mathcal{H}_n] &= 1-(\alpha_n+\alpha_n(1-\alpha_n)+\alpha_n(1-\alpha_n)^2+\alpha_n(1-\alpha_n)^3 \\ &\qquad +\cdots+\alpha_n(1-\alpha_n)^M) \\ &= (1-\alpha_n)^{M+1}\end{aligned} \tag{5.29}$$

Substituting theses values into equation (5.28) and considering that the probability that a node has a packet $\Pr[\mathcal{H}_n]$ is ρ_n , i.e., $\Pr[\mathcal{H}_n]=\rho_n$ and $\Pr[\overline{\mathcal{H}_n}]=1-\rho_n$ yields:

$$\gamma_n=(1-\rho_n)+\rho_n(1-\alpha_n)^{M+1}=1-\rho_n(1-(1-\alpha_n)^{M+1}) \tag{5.30}$$

5-5-6 Probability of collision

A packet that is sent from node n to node u where $u\in\Omega(n)$ is lost due to collision if at least one node of those that are within the carrier-sensing range of the receiver (node u) other than the intended transmitter (node n), i.e., $\Psi(u)\backslash\{n\}$ commences transmission during the reception of packet from node n. From the viewpoint of the transmitter, members of the set $\Psi(u)$ are categorised into two subsets: the subset of nodes that lies within the carrier sensing range of the intended transmitter (node n), i.e., $\Theta_1(n,u)=\{x\in\Psi(u):\Psi(u)\cap\Psi(n)=\{x\},x\notin\{n,u\}\}$ and those nodes hidden from the intended transmitter, $\Theta_2(n,u)=\{x\in\Psi(u):\Psi(u)/\Psi(n)=\{x\},x\notin\{n,u\}\}$. Considering that the receiver cannot receive a packet while it is transmitting, and that there is a single opportunity for a collision from a member of $\Theta_1(n,u)$ (due to the carrier sensing) and there are L different opportunities for a collision from a member of $\Theta_2(n,u)$ gives the probability that the transmission sent from node n to node u is collided, denoted by χ_{nu}, as:

$$1-\chi_{nu}=\gamma_u(1-\gamma_n)\left(\prod_{\forall x\in\Theta_1(n,u)}\gamma_x\right)\left(\prod_{\forall y\in\Theta_2(n,u)}\gamma_y\right)^L \tag{5.31}$$

5-5-7 MGF of the time required to service a successful packet

The probability that a packet sent from node n to node u over a specific path, e.g., $P_x(n,u)$ is received successfully by node v where v is an intermediate relaying node, i.e., $\xi(P_x(n,u),v)$ and the probability that a packet sent over the path $P_x(n,u)$ is received successfully to the final destination, i.e., $\xi(P_x(n,u),u)$ are given in the following two equations respectively.

$$\xi(P_x(n,u),v)=\prod_{\substack{\forall n_i\in P_x(n,u)\\ n_i\notin\{v,\dots,u\}}}1-\left(\beta_{n_i}+\left(1-\beta_{n_i}\right)\chi_{n_in_{i+1}}\right) \tag{5.32}$$

$$\xi(P_x(n,u),u) = \prod_{\substack{\forall n_i \in P_x(n,u) \\ n_i \notin \{u\}}} 1 - \left(\beta_{n_i} + \left(1 - \beta_{n_i}\right)\chi_{n_i n_{i+1}}\right) \tag{5.33}$$

where β_n is the probability that a node drops a packet after assessing the channel as busy $M+1$ times which is given in equation (5.29) i.e., $\beta_n = (1-\alpha_n)^{M+1}$, as a node, when it assesses the channel and finds it busy $M+1$ times, drops a packet. The set $\{v, \dots, u\}$ that appears in equation (5.32) is a subset of $P_x(n,u)$ that contains node v and all of the next hops towards node u as well as node u itself, equations (5.32) and (5.33) are the probabilities that are required to compute the routing probability in section 5-4-2.

Let ξ_n be average the probability that a packet sent by node n to a member of its neighbours is received successfully, the value of this probability is:

$$\xi_n = 1 - \left(\beta_n + (1-\beta_n) \sum_{\forall x \in \Omega(n)} r_{nx}\chi_{nx}\right) \tag{5.34}$$

Then the MGF of the time required to service a successful packet by node n, here written as $s_n(t)$, is considered as a compound probability distribution [161] with $z_n(t)$ as a secondary distribution and with a geometric distribution with success probability ξ_n as a primary distribution. Therefore $s_n(t)$ can be given as:

$$s_n(t) = \xi_n\left(1 - (1-\xi_n)z_n(t)\right)^{-1} \tag{5.35}$$

A justification for this definition can be drawn from the fact that the time required to service a successful packet is a compound of one or more random variables each one distributed according to the service time distribution $z_n(t)$. Using the properties of the compound probability distribution, the average and variance of the time required to service a successful packet can be given as:

$$E[S_n] = {\xi_n}^{-1}E[Z_n] \tag{5.36}$$

$$V[S_n] = {\xi_n}^{-2}((E[Z_n])^2 + \xi_n V[Z_n]) \tag{5.37}$$

where the values of $E[Z_n]$ and $V[Z_n]$ are given in equations (5.18) and (5.19) respectively. It is worth noting that equation (5.37) is that required to find the utilisation factor in equation (5.11).

5-5-8 Performance metrics of multihop WSNs

Using the expressions of the MGFs for inter-arrival time of packets and the successful service time, the MGF of the total delay (both for waiting and servicing) that a packet experiences in node n can be approximated as [178]:

$$q_n(t) = \frac{(1 - E[A_n]E[S_n])(t-1)s_n(t)\left(1 - a_n\big(s_n(t)\big)\right)}{E[A_n]\left(1 - s_n(t)\right)\left(t - a_n\big(s_n(t)\big)\right)} \tag{5.38}$$

where $a_n(s_n(t))$ refers to a compound probability distribution whose primary distribution is the inter-arrival distribution and whose secondary distribution is the successful service distribution [161]. Using the properties of the MGF, the mean and variance of the total delay at node n can be approximated as:

$$E[q_n] = E[S_n] + \frac{E[S_n]V[A_n] + V[S_n]E[A_n]}{2\left(1 - E[S_n]E[A_n]\right)} \tag{5.39}$$

$$V[q_n] = \frac{\left(E[S_n]\right)^2 V[A_n] + \left(E[A_n]\right)^2 V[S_n]}{4\left(E[S_n]E[A_n]\right)} + \frac{\left(V[S_n]\right)^2 V[A_n] + \left(V[A_n]\right)^2 V[S_n]}{\left(V[S_n] + V[A_n]\right)^2} \tag{5.40}$$

The average delay and jitter of packets at node n are the average and standard deviation of $q_n(t)$ respectively. Likewise the average end-to-end delay and jitter over a path $P_x(n,u)$ can be determined as the aggregate of the averages and standard deviations of each node within this path. Hence, the average end-to-end delay, jitter and throughput over the path $P_x(n,u)$ can be given respectively as:

$$E\left[q_{P_x(n,u)}\right] = \sum_{\substack{\forall y \in P_x(n,u) \\ y \notin \{u\}}} \xi_y^{-1} E\left[q_y\right] \tag{5.41}$$

$$J\left[q_{P_x(n,u)}\right] = \sum_{\substack{\forall y \in P_x(n,u) \\ y \notin \{u\}}} \xi_y^{-1} \sqrt{V\left[q_y\right]} \tag{5.42}$$

$$\eta_{P_{x(n,u)}} = L \sum_{\substack{\forall y \in P_{x(n,u)} \\ y \notin \{u\}}} \xi_y (E[Z_n])^{-1} \tag{5.43}$$

where the throughput of a node is determined as the fraction of the MAC service rate (the reciprocal of the average MAC service time) that has been used to send packets on the condition that they have been received successfully [5].

5-6 Validation of the analytical results

This section assesses the integrity of the analytical model by comparing its results with the outcomes of simulations. It starts by summarising the simulation parameters and how the analytical results are obtained in subsection 5-6-1, then in subsection 5-6-2 the breakdown analysis for the average probability of lost packets are provided. Subsection 5-6-3 presents the statistical characteristics of the service and inter-departure distributions. Finally, subsection 5-6-4 considers the characteristics of the average end-to-end delay and probability of lost packets over multihop networks.

5-6-1 Simulation scenarios

This chapter utilises the simulation scenario and labelling scheme presented in section 4-4 in which each case is represented by the five tuple scheme $\langle k, N, X, Y, Z\rangle$. The parameter k denotes the average number of mutually independent paths between non-neighbouring nodes (i.e., connectivity degree of the network) and parameter N denotes the average number of node per a channel. The parameter X specifies the probability distribution used to generate the inter-arrival times between packets, this parameter can take three values: CRB (for Constant Bit Rate), EXP (for exponential distribution) and WBL (for Weibull distribution). The parameter Y defines the rate with which a node generates its internal packets given in packet per slot. Finally, the parameter Z defines the traffic distribution scheme which can take two possible values: U when all traffic that is generated internally by a node is distributed evenly to other nodes within the network and A to indicate that all traffic generated by all nodes within the network is destined to a single final destination.

From the perspective of the analytical framework, the value of the parameter X is set at $a_n^i(t)$ and the value of the parameter Y represents $\lambda_n^i;\ \forall n \in \Upsilon$. Finally the values of the U and A cases of the parameter Z are obtained from the analytical model by setting the value of λ_{nu}^i to $(|\Upsilon - 1|)^{-1}$ and $\lambda_{nx}^i = \lambda_n^i$ and $\lambda_{nu}^i = 0, \forall u \neq x, \lambda_x^i = 0$ respectively. Furthermore, the value of the F parameter of equation (5.20) is set to 20, which means that the average probability of assessment the channel during the service time is found up to the first 20^{th} moments. The results provided in this chapter assumes that all nodes are permanently awake (i.e., unity duty cycle scheme) and employs the source routing policy with hop count as routing metrics. Table 5.3 summaries the simulation parameters.

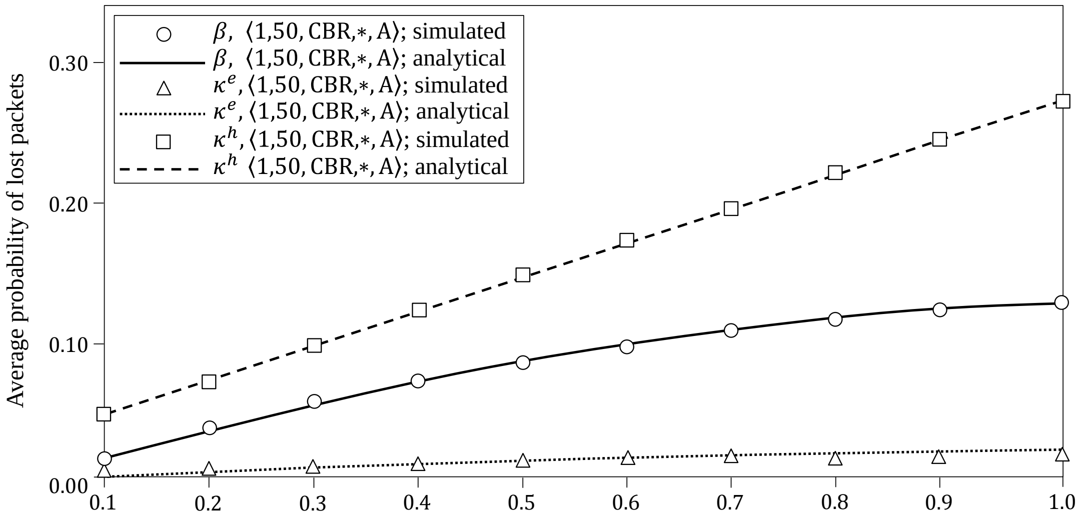
Table 5.3 Simulation parameters

Parameter	Value
Data rate (D)	250kbps
Channel model	Free space
Area	$10^4 \times 10^4 \text{m}^2$
Number of nodes	800-1600

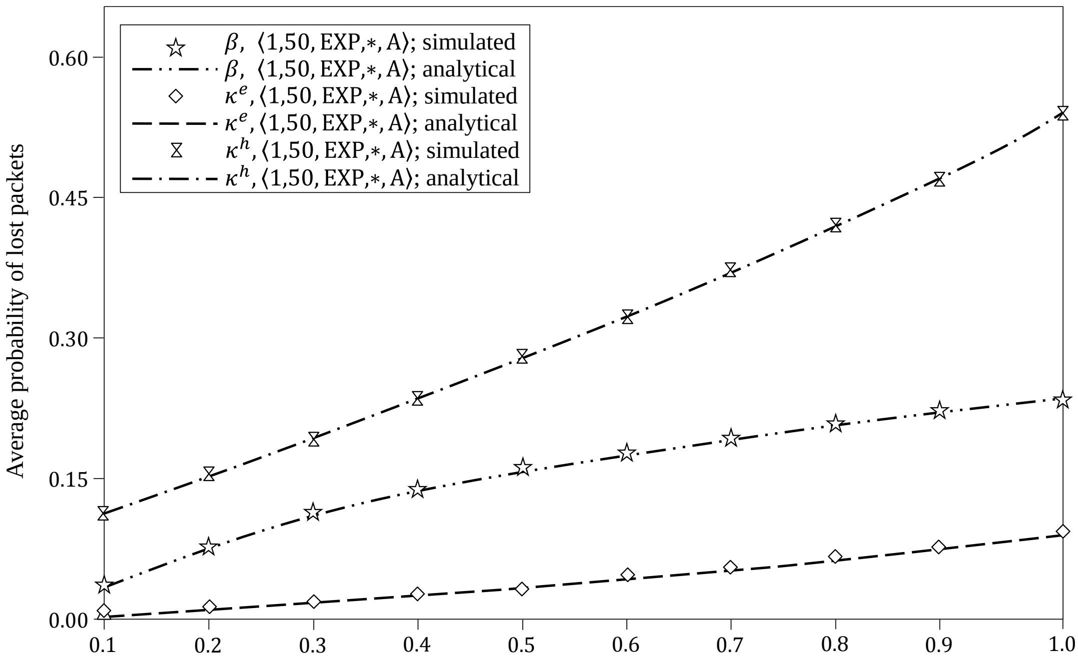
Packet length (L)	30-130 bytes
Power readings	According to [137]
Transmission range (TR)	22-52 m
Carrier-sensing range (CS)	$2 \times TR$
$X = a_n^i(t)$	CBR for Constant Bit Rate
	EXP for Exponential distribution.
	WBL for Weibull distribution.
$Z = \lambda_{nu}^i$	$\frac{1}{\lvert\mathcal{N}\rvert}$ for evenly traffic distribution.
	$\begin{cases}\lambda_n^i; & \text{for a randomly selected node } x\\ 0; & \text{otherwise}\end{cases}$
$\omega_{P_x(n,u)}^i$	$\frac{1}{\lvert P(n,u)\rvert}$
CSMA-CA parameters	$M = 4, W_1 = 2W_0$ and $W_2 = W_3 = W_4 = 4W_0$

5-6-2 Breakdown of average probability of lost packets

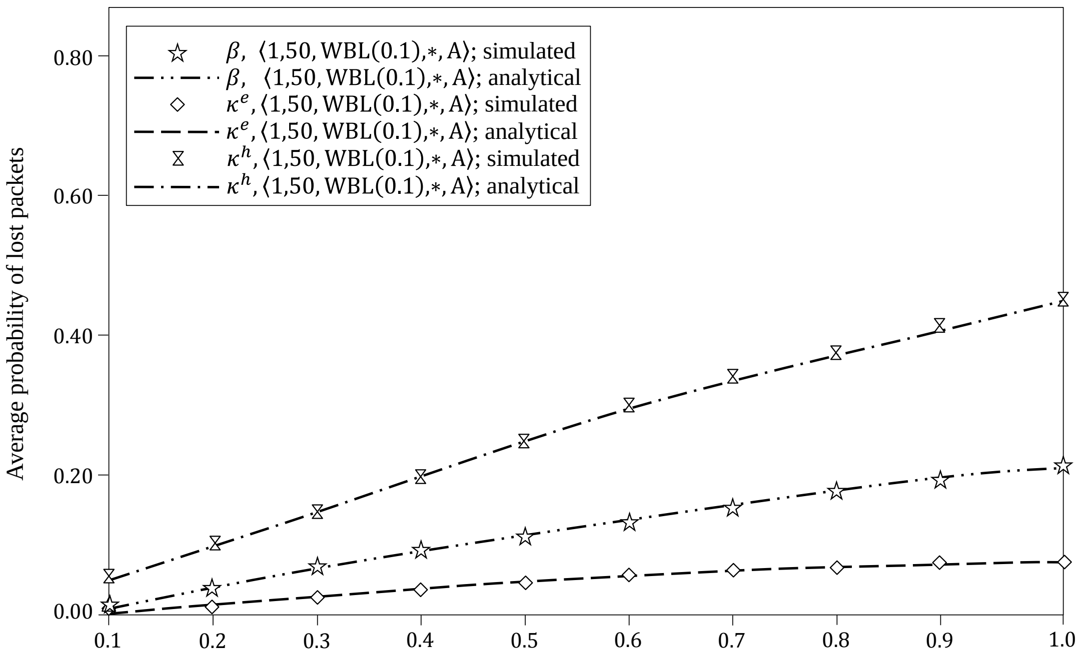
A breakdown for the average probabilities of lost packet over 16 hops versus the normalised traffic intensity for different network configurations is shown in figure 5.5.



(a)



(b)



(c)

Figure 5.5 Breakdown of the average probability of lost packets of IEEE 802.15.4

Figure 5.5 depicts the breakdown of the average probabilities of lost packets when there is on average a single path between non-neighbouring nodes, the average number of nodes per a channel is 50, inter-arrival times between packets are generated according to different distributions: Constant Bit Rate (CBR), Exponential (EXP) and Weibull (WBL), traffic rates vary from 0.1 to 1 with respect to the data rate and all packets are destined to a single final destination. It can be seen that in all cases, the probabilities of lost packets increase proportionally with traffic intensity; the probability of losing a packet due to a hidden node collision (denoted by κ^h) is the most significant contributor to this effect, followed by the probability of dropping the packet due to a failure to access the channel (β) and finally the probability of lost a packet due to simultaneous transmission by a node that is a neighbour of the receiver and within the carrier sensing of the transmitter (denoted by κ^e). This characteristic is attributed to the fact that there are L opportunities for a hidden node collision while there is only a single opportunity for a collision caused by the other case as shown in equation (5.31). The increasing of the failure to access the channel (β) is due mostly to the increase in channel occupancy due to the increase in traffic intensity.

Figure 5.5-b shows that generating inter-arrival times between packets using EXP increases the probabilities of lost packet compared to the cases when the inter-arrival times between packets are generated according to CBR as the case in figure 5.5-a. this characteristic is attributed to the fact that in the EXP distribution packets are being generated at irregular intervals, thereby, there is an increased possibility of two or more packets are generated in the same interval (i.e. the time interval between generation can be smaller than the time required to sense the channel and the length of packet) and in this case a node immediately starts to service a packet after finishing the transmission of the previous packet; this increases the contention on the channel which leads in turn to magnify the probabilities of lost packets. However, in the case of a CBR distribution, packets are generated periodically so the probability that two or more packets are generated in the same interval is small. Figure 5.5-c shows that using the Weibull distribution with shape parameter 0.1, i.e., WBL(0.1) increases the probabilities of lost packets notably. This is attributed to that the WBL(0.1) is a heavy-tailed distribution (i.e., the masses of such distribution are spread over large area with small amount of probabilities between inter-arrival times). Hence the probability that the inter-arrival times between packets are small is high; this leads to an increase in contention and consequently the probabilities of lost packets.

5-6-3 Statistical characteristics of service time and inter-departure distributions

This section reveals some important characteristics of the CSMA-CA protocol by investigation of the Cumulative Distribution Function (CDFs) and Probability Mass Functions (PMFs) of service time and inter-departure distributions. From an analytical viewpoint, the PMFs of medium access delay and inter-departure are derived from their MGFs by extracting the ordered coefficients of equations (5.17) and (5.11) respectively, and then from these PMFs the CDFs are computed, where the value of α is taken as the average over the whole network and number of hops is fixed at 16 hops.

Figure 5.6 shows that the probability of completing the service of a packet within a small number of slots decreases when the average number of nodes contending to access the channel N and the packet arrival rate increase. It can also be seen that the minimum number of slots required to service a packet is 5 (the number of slots required to drop a packet when a node finds a busy channel in the first slot in each stage) while the maximum number of slots is 128 (the sum of the maximum back-off periods in each stage plus the length of a packet, $L = 8$ slots).

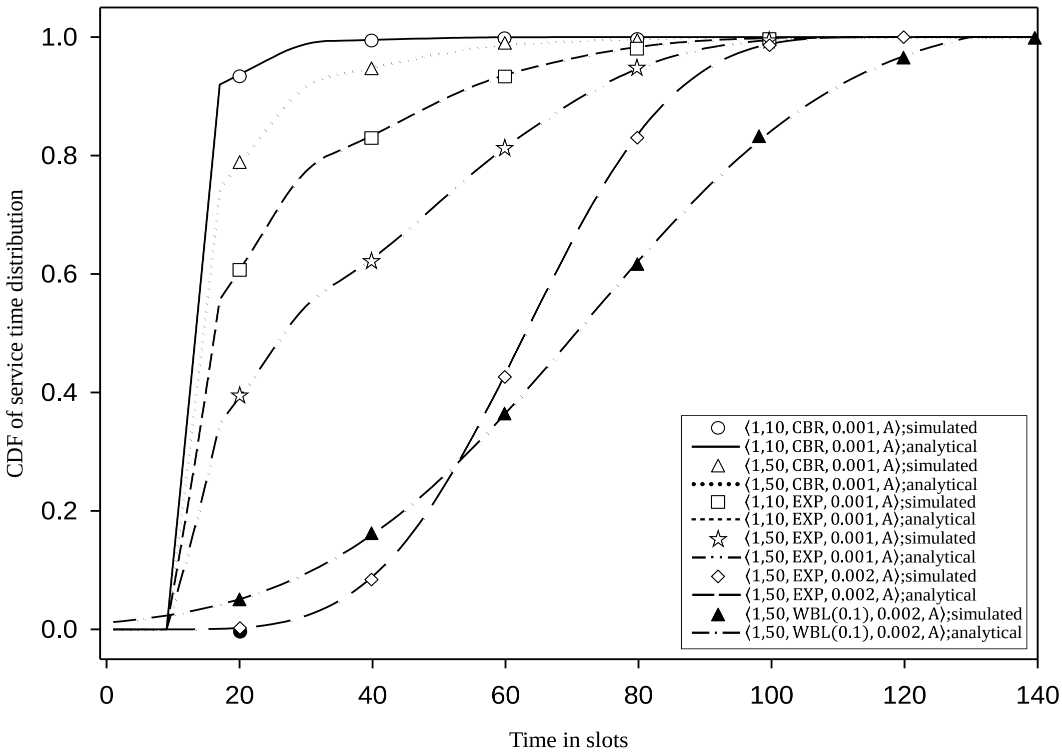


Figure 5.6 CDFs of the service time distribution of IEEE 802.15.4

Figure 5.6 demonstrates that when the average number of nodes per channel $N = 10$ and packets are generated according to a CBR with a low rate, the probability of a packet completing its service in less than 20 slots is around 0.95 which means that under these conditions a packet is typically backed-off for no more than 2 stages. When the average number of nodes per channel is increased to $N = 50$ and other factors are kept at the same values (the same constant distribution, the same packet generation rate and the same packet length are used), the probability of completing the service of a packet by 20 slots decreases by more than 20% to reach 0.75 while with probability 0.95 a packet completes its service within 50 slots. Comparing these results shows an inverse relation between an increasing number of nodes and the probability of assessing the channel as idle. When packets are generated according to an exponential distribution (EXP) and for a small average number of nodes per channel $N = 10$, the probability of completing the servicing of a packet in around 20 slots is about 0.6 and when the number of nodes increases to 50 the probability of completing the servicing of a packet in the same number of slots 20 slots decreases to about 0.4. This is attributed to an increase in the probability that the channel becomes busier with an increasing number of nodes and packets are thereby involved in a greater number of back-off periods. When the number of nodes is kept at the same value $N = 50$ but the packet arrival rate is increased from 0.001 to 0.002 packets per slot, the probability of servicing a packet in the same number of slots (20) decreases dramatically to less than 0.2. This gives an indication that the inter-arrival rate of packets has a greater effect than the number of nodes on the service time. Under light traffic conditions, a node stays idle for a long period after finishing serving a packet which gives time for other nodes to service their packets. In contrast, under heavy traffic conditions, a node often starts to immediately service another packet after completing servicing the current packet, so the number of nodes contending to access the channel remains roughly constant. The characteristic of service time distribution when the inter-arrival times between packets are generated according to a heavy tailed distribution is illustrated in the last cases of figure 5.6. It can be seen that the CDF of the scenario $\langle 1{,}50, \mathrm{WBL}(0.1){,}0.002, \mathrm{A}\rangle$ converges slowly to unity compared to other CDFs of all other scenarios. This is attributed mainly to that using a heavy tailed distribution to generate the inter-arrival times between packets increases the variation of traffic over the network which in turn makes the contention to access the channel vary greatly between nodes and slows ergodicity of the service time.

Figure 5.7 illustrates the PMF of the case $\langle 1{,}10, \mathrm{EXP}, 0.001, \mathrm{A}\rangle$, that is there is on average a single path between non-neighbouring nodes, the average number of nodes per a channel is 10, inter-arrival times between packets are generated according to exponential distribution with low rate, 0.001 packet/slot and all packets are destined to a single final destination. It can be seen from the results of this figure that even under these light operational conditions, the tail of PMF is heavy. This characteristic is a direct result of the Binary Exponential Back-off (BEB) mechanism that is used to generate the back-off intervals. Recall that the IEEE 802.15.4 CSMA-CA standard requires a node to generate its back-off durations based on a uniform distribution whose initial window is set to predefined value and is doubled whenever the channel is assessed as a busy. Thereby, the time required to service a packet increases multiplicatively based on the number of busy channel assessments. From a statistical perspective, the BEB mechanism represents a power law process which yields a distribution whose tail decays slower than exponential functions; hence these distributions are typified as heavy tailed distributions [197]. Using a heavy tailed distribution to generate the service time of packets over multihop WSNs results in a poor performance as it increases the average end-to-end delay due to the high variation of the service time. Moreover, the heavy tailed service time distributions reduce the throughput and lifetime of the network, as it requires a node to contend for the channel for a long time. Chapter 6 considers these issues and demonstrates that replacing the BEB with low variable process can improve the performance of the network significantly.

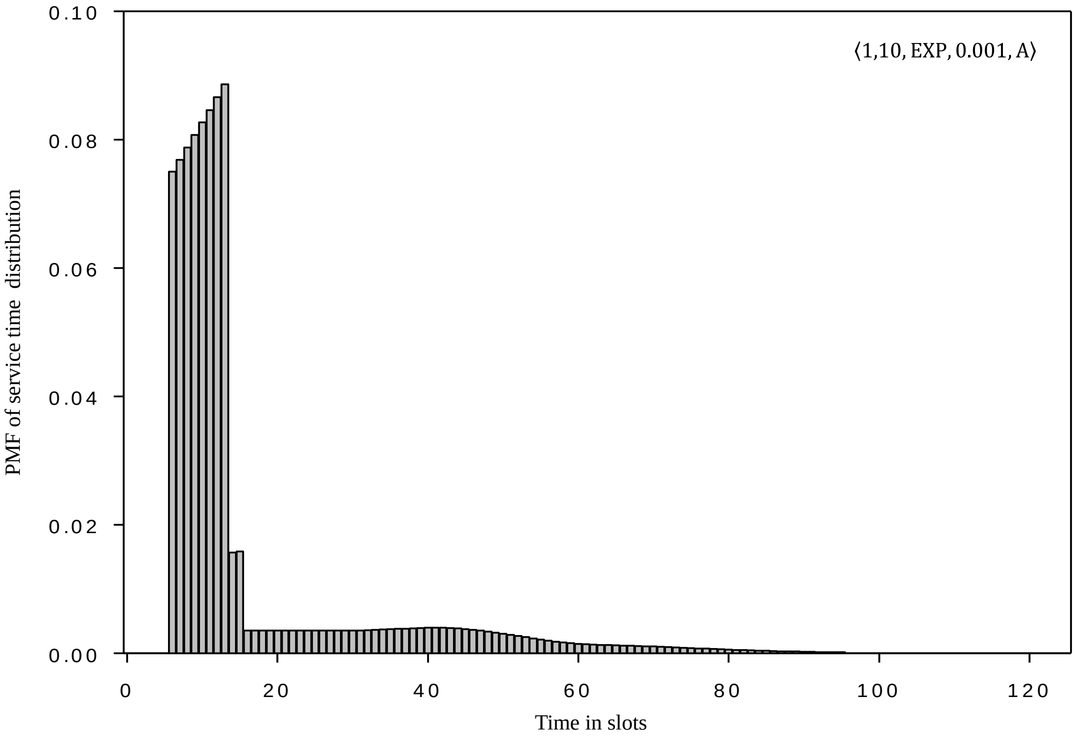


Figure 5.7 PMF of the service time distribution of IEEE 802.15.4 for a selected case

Figure 5.8 shows the CDFs of the inter-departure time of the scenarios considered in figure 5.6. For a small average number of nodes per channel $N = 10$ with low packet arrival rate of 0.001 packet/slot, we have shown in figure 5.6 that the time required to access the medium is low (a packet completes its service in no more than 20 slots with probability 0.95) and in this case the inter-departure time is dominated by the time between the departure of a packet and the arrival of the next packet; most of the inter-departure time is spent while a device is idle. In contrast, when the average number of nodes per channel N increases to 50, the time required to access the medium increases (recall that this time increases by 20% due to the greater number of nodes) and therefore packets encounter a queuing delay consistent with the sum of the medium access delays of other packets already in the queue. This makes the inter-departure times longer due to the higher contribution of the service function. For exponential distributions, the aforementioned observations remain true, however, when the rate increases from 0.001 to 0.002 packets/slot, it can be seen that the inter-departure time increases significantly. This increasing is attributed to the additional contributions from both the medium access delay and the increase in the traffic intensity. Comparing the inter-departure distribution of the case $\langle 1{,}50, \mathrm{WBL}(0.1), 0.002, \mathrm{A}\rangle$ with the remaining cases shows that employing the Weibull distribution to generate the inter-arrival times between packets starched out the inter-departure distribution over larger interval. This characteristic exhibits the impact of high level of variation of WBL(0.1) on reducing the probabilities of inter-departure timesbetween packets. A mathematical quantification for the impact of the variation of the inter-arrival and service time distributions on the inter-departure distribution has been introduced by Marshall's equation [180] which can be approximated as:

$$CV[D_n] = \sqrt{\left(\rho_n CV[S_n]\right)^2 + \left(1-\rho_n^2\right)\left(CV[A_n]\right)^2} \tag{5.49}$$

where $CV[A_n]$, $CV[S_n]$ and $CV[D_n]$ are the coefficients of variation of inter-arrival, service and inter-departure distributions respectively. Considering that, in this assessment, the value of $CV[A_n]$ of WBL(0.1) is the highest followed by EXP and CBR justifies the characteristics of inter-departure distributions shown in figure 5.8.

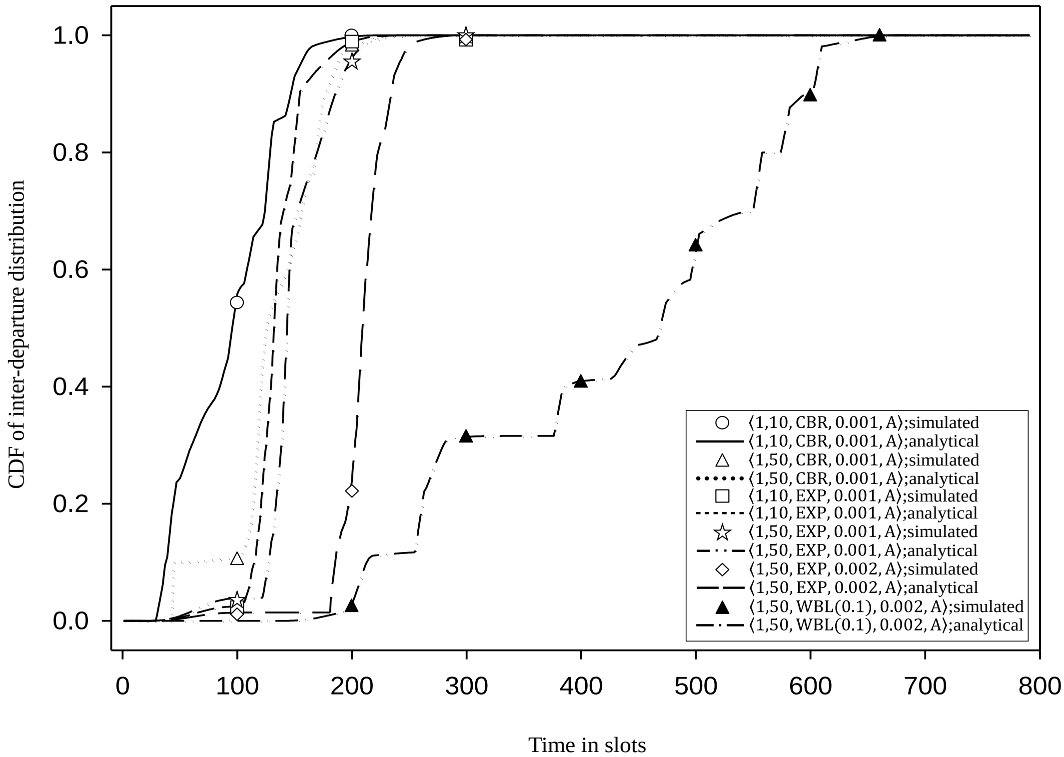


Figure 5.8 CDFs of the inter-departure distribution of IEEE 802.15.4

Figure 5.9 illustrates the inter-departure PMF of the cases $\langle 1,10, \mathrm{EXP}, 0.001, \mathrm{A}\rangle$ whose service time distribution depicted in figure 5.7. Comparing the PMFs of figures 5.9 and 5.7 shows that the tail of the inter-departure distribution is heavier than of the medium access delay; this is a result of queuing delay.

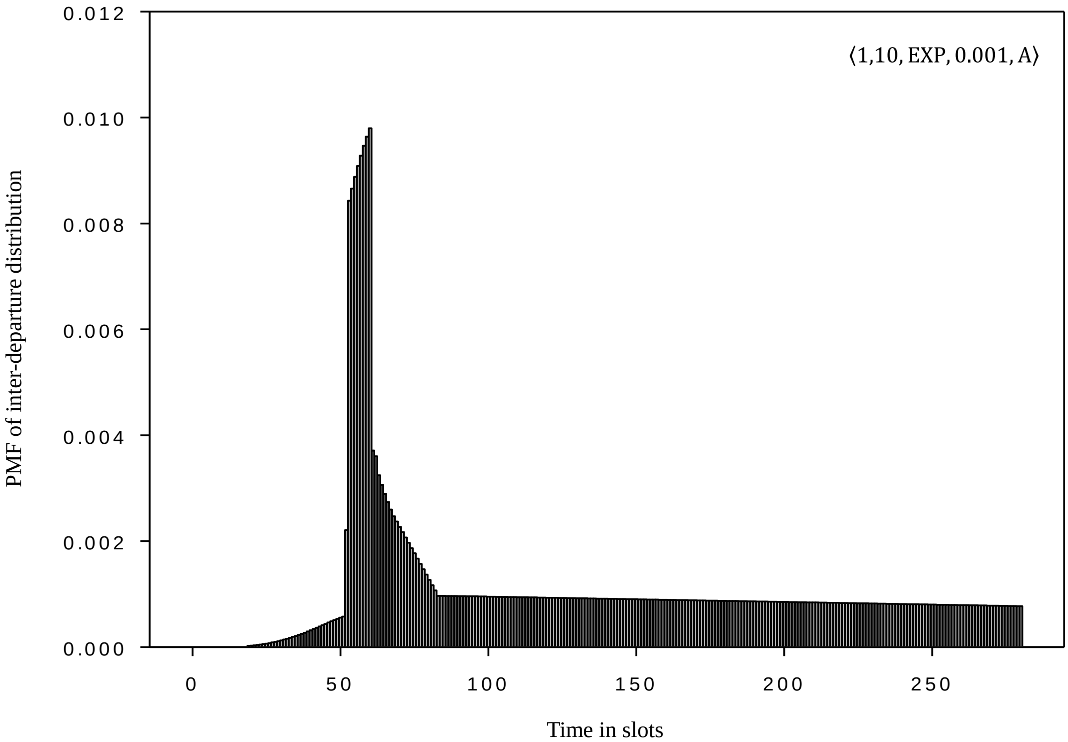


Figure 5.9 PMF of the inter-departure distribution of IEEE 802.15.4 for a selected case

Assessment of the accuracy of the CDFs of the analytical framework is carried out by comparing them with the simulated readings using the Kolmogorov-Smirnov index (K-S index) [181] which quantifies the maximum difference between the simulated samples and analytical CDFs in the range [0,1] where 1 means there is no match while 0 means the simulated samples matches the analytical CDF very well. This assessment uses the number of packets of the simulation as an indicator for the K-S index, the results of the K-S assessments of the service time and inter-departure distributions are given in figure 5.10.

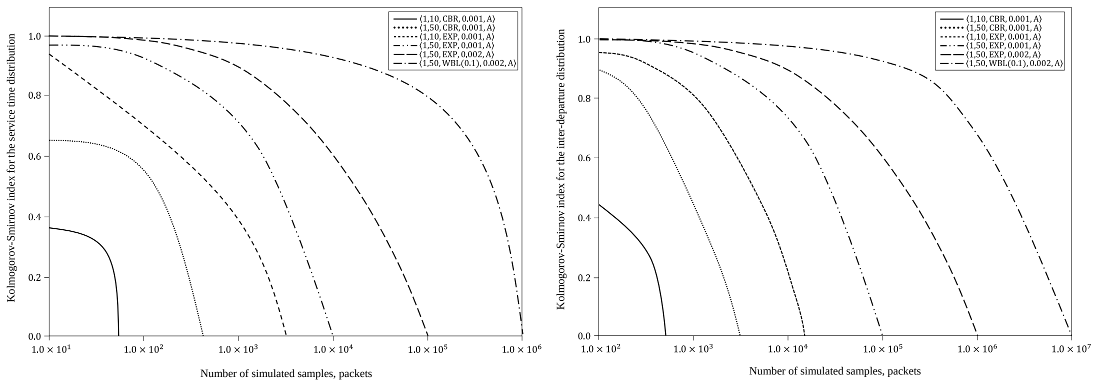


Figure 5.10 K-S index of service time and inter-departure distributions of IEEE 802.15.4

Results of figure 5.10 show that the number of packets required to obtain close match between the simulated and analytical CDFs increases with the traffic intensity and the average number of nodes per channel. Moreover, the results of these figures demonstrate that the inter-arrival distributions with higher variations require more simulated samples. It is also shown that there is a need for a large number of packets from the simulation to obtain a closer match of the K-S index in inter-departure times compared to service time distributions.

5-6-4 Performance metrics over multihop networks

Evaluation for the average end-to-end delay and probability of lost packets over a multihop link in the different network configurations are shown in figures 5.11-5.14. The good agreement between theory and simulation demonstrates the accuracy of the analytical model in accounting for realistic assumptions: heterogeneous traffic patterns, different channel viewpoints and queuing delay.

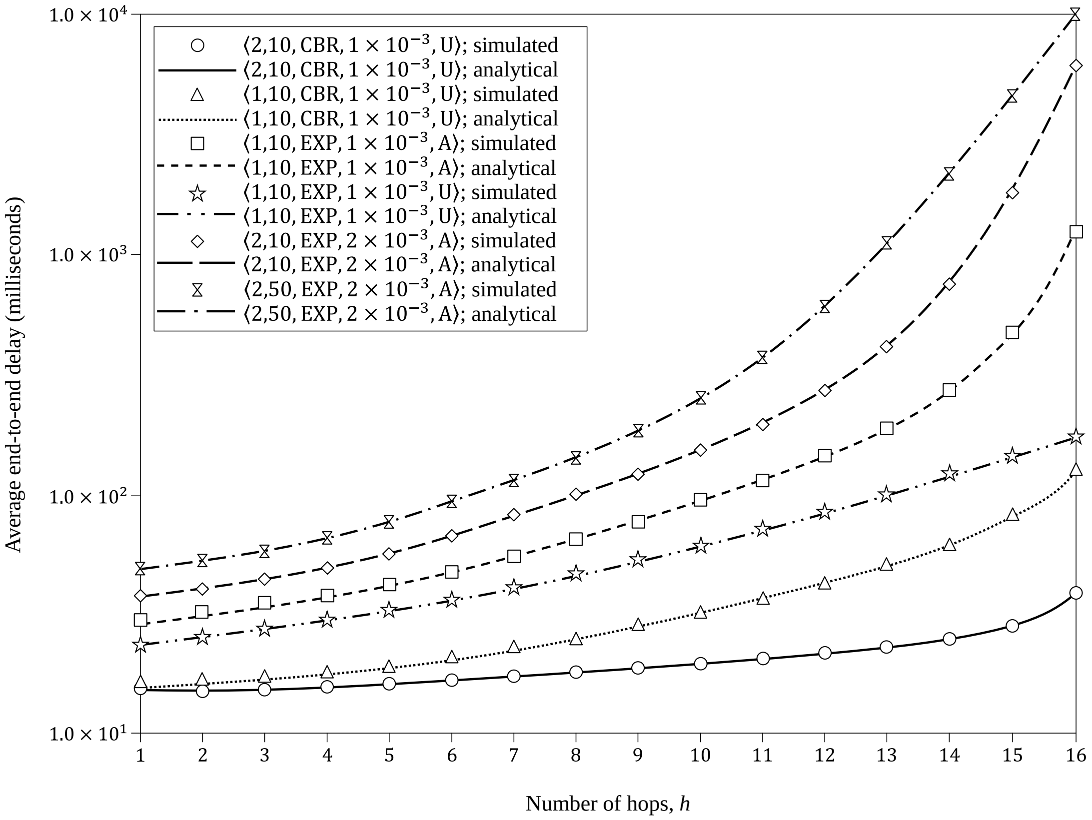


Figure 5.11 Average end-to-end delay of IEEE 802.15.4

The results of figure 5.11 illustrate the effects of the connectivity degree on the average end-to-end delay of packets. Comparing the first two scenarios, i.e., $\langle 2,10, \mathrm{CBR}, 1 \times 10^{-3}, \mathrm{U}\rangle$ and $\langle 1,10, \mathrm{CBR}, 1 \times 10^{-3}, \mathrm{U}\rangle$ shows that when the number of hops is high ($h > 5$) the readings of the former cases is much lower than their peers in the latter cases, however when $h \leq 5$ then the readings of both cases are about the same. This characteristic is attributed to the fact that increasing the number of mutually independent paths over a small number of hops does not reduce the contention to access the channel which keeps the readings of average end-to-end delay similar, while for a larger number of hops the contention is nearly eliminated and hence the average end-to-end delay is reduced. It can also be seen from figure 5.11 that when all traffic is forwarded to a single receiver (indicated as A, in the last element of the tuple) the average end-to-end delay exceeds the equivalent result when all nodes act as originator, router and final destination (U in the last element of tuple); this is attributed to the increase in the collision probabilities with number of hops as all nodes attempt to send their traffic to the same node. It is clear that when the traffic intensity or the average number of nodes per channel increase, the average end-to-end delay increases as a result of the increase in the contention to access the channel as well as the probability of collision as shown in section 5-6-2.

Assessments for the characteristics of the average end-to-end delay under heavy tailed inter-arrival distributions (WBL) are presented in figure 5.12.

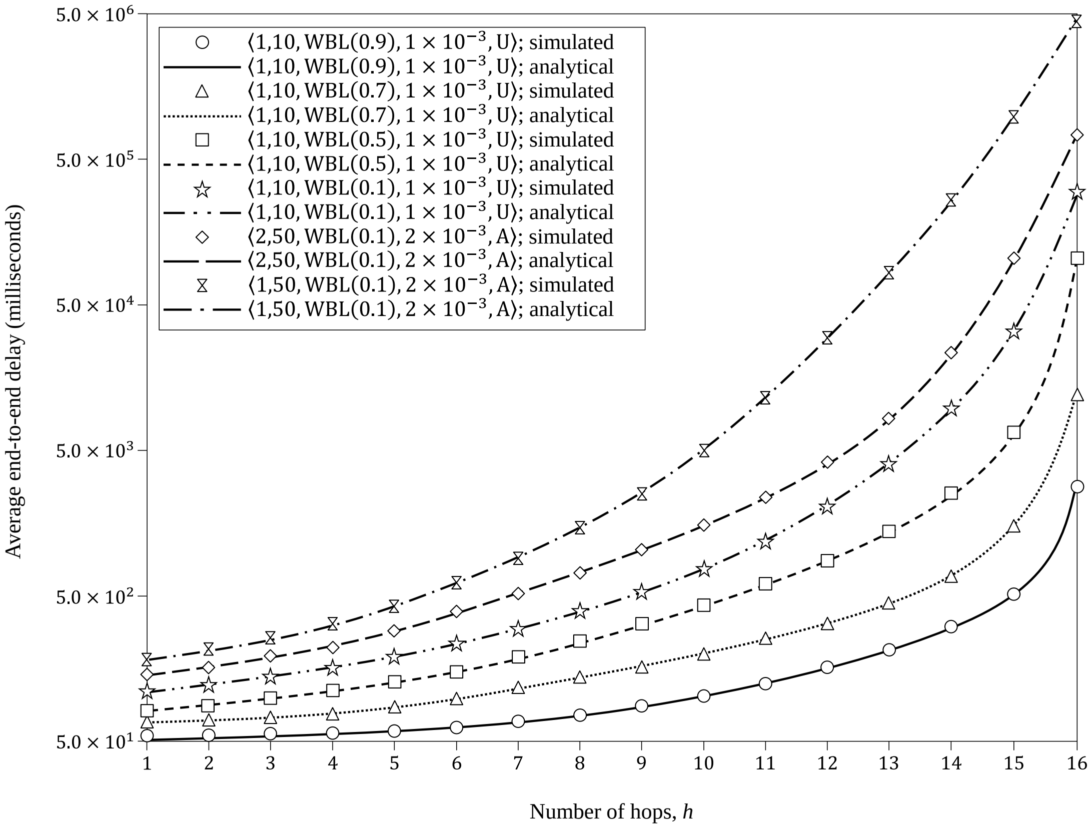


Figure 5.12 Average end-to-end delay of IEEE 802.15.4 under heavy tailed distributions

Results of figure 5.12 show that reducing the shape parameter of WBL increases the average end-to-end delay; this is attributed mainly to the inverse relationship between the value of the shape parameter of the WBL distribution and the heaviness of its tail. Considering that increasing the heaviness of the tail of a WBL distribution increases the variations of inter-arrival timings and recalling the inverse relation between the variation and the fast convergence of service and inter-departure distributions (which is shown in figures 5.6 and 5.8 respectively) explains the significant increases in average end-to-end with the shape parameter of the WBL distribution.

The characteristics of the average probability of lost packets over multihop networks for different network scenarios are illustrated in figures 5.13 and 5.14.

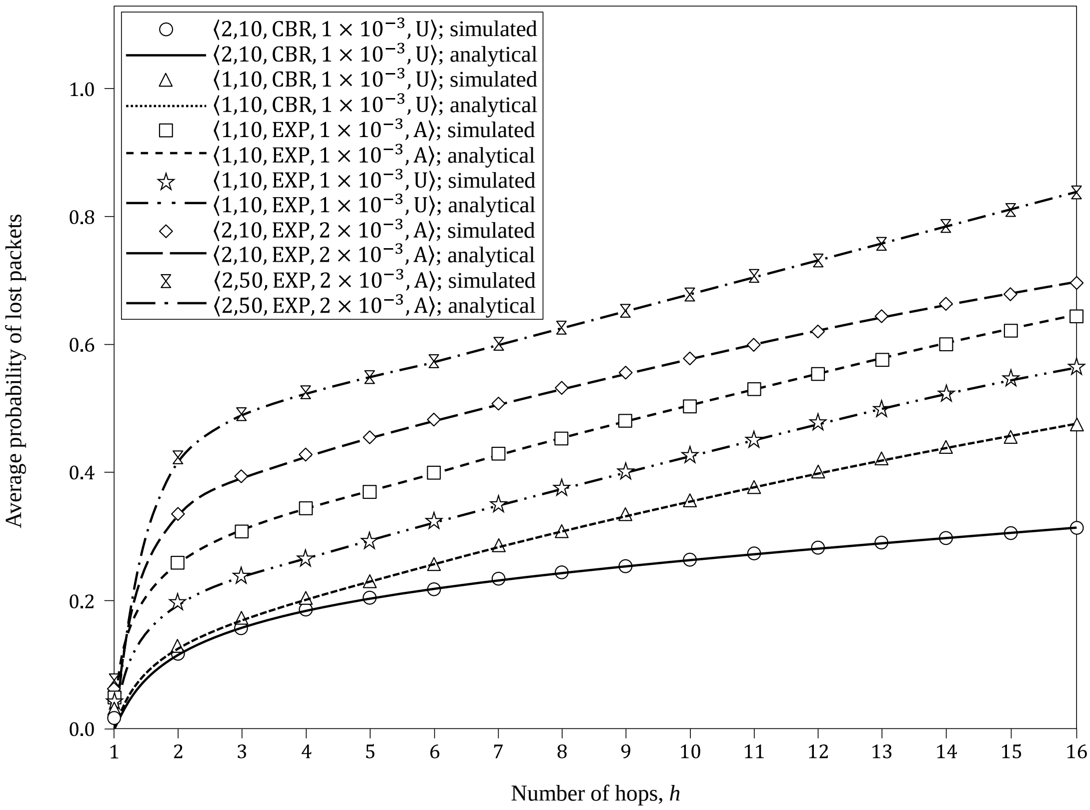


Figure 5.13 Average probability of lost packet of IEEE 802.15.4

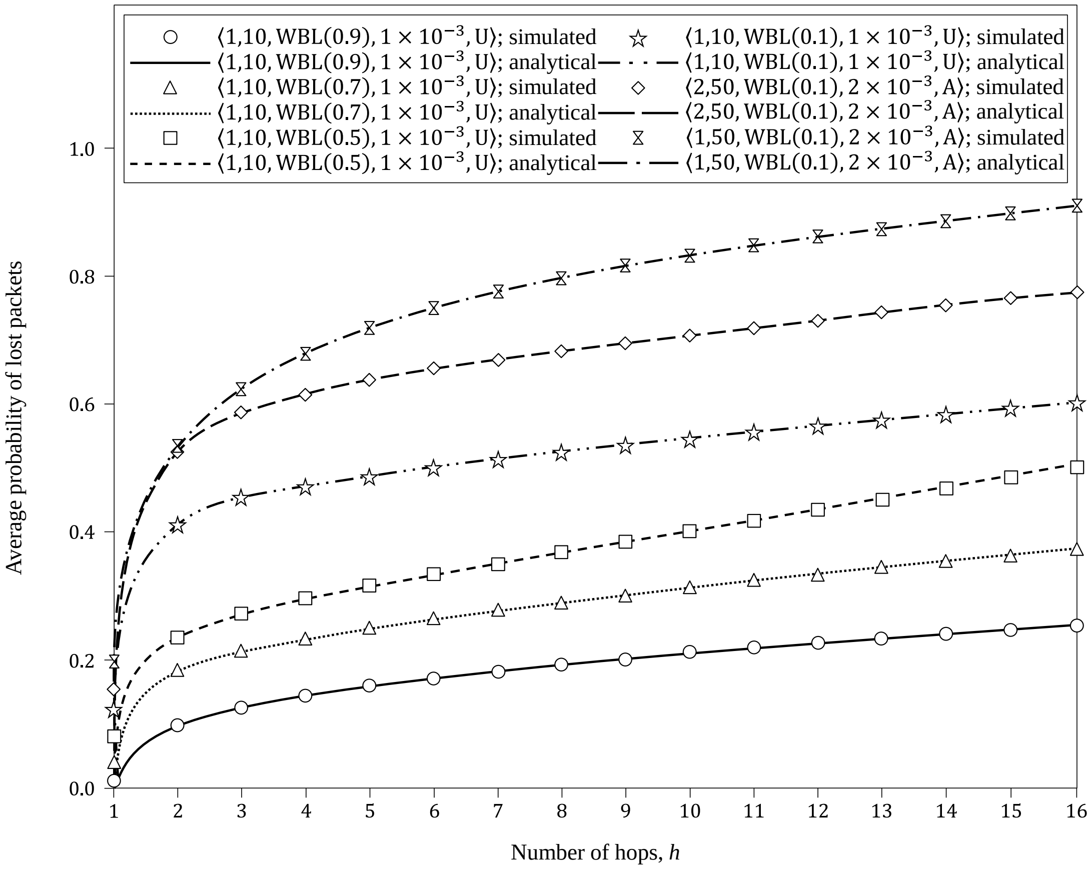


Figure 5.14 Average probability of lost packet of IEEE 802.15.4 under heavy tailed distributions

It can be seen that the probability of lost packets increases substantially with number of hops. This characteristic is attributed mainly to the fact that the IEEE 802.15.4 CSMA-CA protocol does not provide an effective means to resolve or even to identify the underlying reason for lost packets. This protocol only doubles the back-off window after each channel assessment failure and resets this window after each unsuccessful transmission. Although this doubling can lessen the possibility that two or more nodes access the shared channel at the same instance, it cannot mitigate hidden node collisions that are considered as the most significant contributor for the packet lost in has been seen in section 5-6-2. Another main reason for increasing the average probability of lost packets over multihop is that doubling the back-off windows reduces the channel utilisation as the IEEE 802.15.4 CSMA-CA protocol does not enable a node to sense the channel during the back-off intervals in order to save the energy of nodes.

5-7 Conclusion

This chapter has presented the first step toward the main goal of this thesis which is to design an effective MAC, routing and duty cycle management protocols for multihop WSNs. This chapter has proposed a versatile and accurate analytical framework by which the statistical analysis of the CSMA-CA protocols can be obtained. The proposed model is versatile as it can predict the performance of CSMA-CA without restriction on topology configurations, traffic patterns or multipath routing policy. In terms of accuracy, the proposed framework accounts for the effect of variation of the back-off intervals on channel assessments, queuing delay and all possible cases of collisions. The proposed model is exemplified by analysing the statistical characteristics of the IEEE 802.15.4 CSMA-CA and its integrity is demonstrated through comparison with simulations. The results of this chapter highlight that this protocol suffers from significantly poor performance over multihop, moreover, it has been shown that the main reason for this limitation is its back-off mechanism that yields a heavy tailed service time distribution. The next chapter exploits these findings to design an effective MAC protocol.

Chapter 6

Gamma Distribution Based CSMA-CA Back-off Scheme

6-1 Introduction

The previous chapter assessed the statistical characteristics of the Carrier Sensing Multiple Access with Collision Avoidance (CSMA-CA) protocols over multihop networks using the IEEE 802.15.4 CSMA-CA beaconless protocol as an example [16]. Although this protocol has been specified especially for WSNs, the results of chapter 5 highlighted the performance limitations of this protocol over multihop networks, i.e., it has been shown that the average end-to-end delay and probability of lost packets increase significantly with increasing number of hops, average number of nodes contending to access the channel, or the traffic intensity. The deep investigation conducted in chapter 5 showed that this degradation is attributed mainly to the heavy tail of the probability distribution of the medium access delay.

The IEEE 802.15.4 CSMA-CA protocol, like other existing CSMA-CA protocols, consists of two fundamental mechanisms: channel assessment and back-off. The channel assessment process is used to prevent a node from commencing transmission while the channel is occupied with transmissions from other nodes. The back-off algorithm is used to regulate channel access by making a node wait for a random period of time after a collision or busy channel assessment. Therefore, the efficiency of the CSMA-CA protocol depends mainly on the probability distribution used to generate the back-off periods.

Most existing CSMA-CA protocols employ Binary Exponential Back-off (BEB). In these schemes, back-off durations are generated according to a uniform distribution. The window of this distribution is set to an initial value at the start of contention to access the channel and is reset to the same initial value after each successful transmission or when a packet is dropped. The window is doubled whenever a collision or busy channel assessment occurs. Thus, the windows of back-off stages in BEB increase multiplicatively based on the status of the channel (either busy channel or collision) which means that the process that is yielded from BEB is a power law process. From a statistical perspective, the probability distribution of a power law process is a heavy tailed distribution and hence we can conclude that the BEB scheme of CSMA-CA protocol is the main cause of the performance degradation of these protocols over multihop WSNs.

Two other important limitations of the BEB scheme are: firstly, its inability to adapt the back-off periods of a node according to its distinct traffic demands. In a typical multihop network, the traffic pattern is heterogeneous since each node has its own traffic characteristics that differ substantially from other nodes and hence each node needs to contend to access the channel differently. Comparing this requirement with the BEB scheme demonstrates that in BEB, all nodes execute the same procedures irrespective of their traffic loads. The second main limitation of the BEB scheme is that this scheme does not provide a mechanism to resolve the cause of unsuccessful transmission. BEB doubles the back-off windows blindly without providing remedial action for the underlying causes of failed transmission. An illustration for this shortcoming can be acquired by considering the two BEB mechanisms of two standards, the IEEE 802.11 [97] and IEEE 802.15.4 [16]. In the IEEE 802.11 the back-off window is doubled after each collision regardless of whether this collision was due to a non-hidden or hidden node. Conversely, in the IEEE 802.15.4 standard, the back-off window is doubled after a busy channel assessment.

Although a number of ideas have been proposed to improve the performance of BEB, most of them are not specifically designed for multihop networks which make them inadequate in achieving acceptable performance readings for these networks in terms of network throughput, node lifetime and packet delay. Hence there is a high demand to develop a CSMA-CA mechanism for multihop networks. The importance of this chapter for the remaining of the thesis is that it provides the back-off scheme required to make the MAC protocol effective for multihop networks. Upon this proposed scheme other enhancements, e.g., the duty cycle management scheme, routing protocols that are developed in the remaining chapters are incorporated.

This chapter proposes a novel CSMA-CA scheme for multihop networks. The proposed algorithm allows each node to adjust the parameters of the probability distribution used to generate the back-off intervals. This aim of this tuning is to achieve the following objectives: reducing the end-to-end delay, increasing throughput and minimising the energy consumption. End-to-end delay is reduced by letting a node to assess the channel with a high probability when it is presumably idle, and back-off with a high probability when the channel is potentially busy, hence a node can avoid deferring access to the channel for the additional random period. The collision probability is reduced by employing a novel collision resolution algorithm that accounts for both non-hidden and hidden node collisions. Reduction of the time required to service a packet and the probability of collision prolong the lifetime of nodes by reducing the energy consumed in contention to access the channel. The proposed protocol employs the Gamma distribution [182] to generate the back-off intervals. The Gamma distribution has been selected because it is a highly parametric distribution which provides an excellent basis for adapting to the wide variety of potential traffic distributions. For example, the exponential, Poisson, Erlang, Chi-square and normal distributions can be derived from a Gamma distribution simply by adjusting its parameters [182]. Another great benefit of employing the Gamma distribution in the CSMA-CA protocol is to mitigate the high level of variation found in the BEB protocols. Employing the Gamma distribution generates a Gamma process that has a low level of variation compared to that yielded from the power law process in the BEB scheme. Importantly, the protocol is lightweight and can be applied to any wireless network, as it does not require modification of the physical layer standards. The simulated outcomes demonstrate the bountiful benefits of the proposed protocols.

The rest of this chapter is organised as follows. Section 6-2 provides an overview of the binary exponential back-off process. Section 6-3 explores the related work in order to highlight their contributions and limitations and to motivate the proposed work. Section 6-4 discusses the statistical characteristics of the Gamma distributions and section 6-5 exploits this discussion to highlight the underlying approach of the proposed protocol. Section 6-6 provides the pseudo-code of the proposed protocol, section 6-7 assesses the benefit of the proposed protocol and finally section 6-8 concludes this chapter.

6-2 Overview of binary exponential back-off

This section overviews the Binary Exponential Back-off (BEB) mechanism from different viewpoints. Section 6-2-1 provides a brief history of this mechanism, section 6-2-2 explains the relationship between the BEB and the power-law process while section 6-2-3 reveals some reported limitations of BEB.

6-2-1 Brief history of binary exponential back-off

The exponential back-off algorithm is a mechanism that is used in a multiple agent interacting system to control the rate with which these agents access shared resources [183-185]. The key attribute of exponential back-off as its name implies is that the process execution's rate is changed multiplicatively rather than in a linear manner. i.e., the rate is changed exponentially which gives enough room to accommodate more activities by other parties in the system. The exponential back-off mechanism has been used to solve problems in different disciplines including the Resource Constrained Project Scheduling Problem (RCPSP) [183] to provide an effective utilisation for those resources that are shared by different tasks; Software Transactional Memory (STM) [184] to mitigate conflicts that may happen when two or more transactions access the same memory locations and Thwart Dictionary Attacks (DA) [185] to control the lockout period between the number of authentication attempts.

Employing the exponential back-off mechanism in Medium Access Control (MAC) protocols dates back to the work presented in 1973 [186] that assessed the performance of satellite system operating the slotted ALOHA protocol. The author reported that delaying the retransmission of a collided packet for a random period distributed uniformly with length $2^{W_{min}}$ slots (where W_{min} is the minimum back-off exponent of the BEB mechanism) increases the throughput of the network and when

W_{min} approaches infinity, the throughput comes close to its optimal value, i.e., e^{-1}. Moreover, it was shown that if the value of W_{min} is kept constant irrespective of the number of collisions, then the throughput collapses to zero while if the W_{min} is doubled after each collision then more stable throughput is maintained. Although this work considers only a single hop satellite network, its conclusion constitutes the framework of the Binary Exponential Back-off (BEB) mechanism in MAC protocols.

Another milestone of employing BEB to manage the traffic flow in communication networks was credited to Jacobson in 1986 when the throughput of the network connecting the Lawrence Berkeley Laboratory and the University of California at Berkeley dropped from 32kbps to just 40bps [187]. The investigation carried out by Jacobson concludes that the main reason of this degradation is the overwhelming of the traffic which caused congestion at the network links or overflow at devices. Based on this conclusion, the exponential retransmit back-off timer was proposed; the fundamental principle of this policy is to let a node double its retransmission timeout based on the congestion level of the network.

6-2-2 Binary exponential back-off and its relation to power-law process and heavy tailed distributions

The BEB algorithm in MAC protocols can be specified using four parameters:

1. The minimum and maximum back-off exponents, denoted here by W_{min} and W_{max} respectively. These two values define the initial value and upper limit of windows of uniform distributions that a node selects its back-off durations from.
2. The maximum number of back-off stages, denoted by M.
3. The incremental policy of BEB which specifies the procedure by which the back-off exponent is incremented. Two main policies are widely used: Truncated Binary Exponential Back-off (TBEB) and BEB. In the TBEB the maximum back-off exponent is fixed at a predefined value while in BEB, the maximum back-off exponent is set according to the maximum number of back-off stages.
4. The action upon which the back-off exponent is incremented; this action is defined based on the design philosophy of the BEB algorithm. For instance, in the IEEE 802.11 DCF standard, the back-off exponent is increased after a collision while in IEEE 802.15.4 the back-off exponent is incremented upon sensing the a busy channel.

Using the aforementioned parameters, the maximum length of the back-off windows in the m^{th} stage for the BEB and TBEB schemes can be given as $2^{W_{min}+m-1}$ and $2^{\min(W_{min}+m-1, W_{max})}$ respectively. These expressions represent canonical forms of the power law process which is the underlying cause of being the probability distribution of BEB based MAC protocols are heavy tailed.

The relationship between the power law process and the heavy-tailed distribution has been explained in [188-189]. This work shows that the heavy-tailed distribution is produced from a power-law process as a result of mixing large numbers of small observations with a small numbers of large observations. Thus, the majority of the observations are small and most of the contributions of the mean of these observations are due to the rare large observations. This characteristic of heavy-tailed distributions has been named by other researchers as the expectation paradox [190] which states that for a random variable sampled from a heavy-tailed distribution, the longer we have waited, the longer we should expect to wait. This paradox demonstrates one of the fundamental statistical characteristics of heavy-tailed distributions that is slow convergence of its expectations and in more general most of its moments.

Assessment of the impact of the heavy tailed distribution on performance indicators has been presented in the work presented in [191-192] which concludes that employing the heavy-tailed distribution as a service function in a given queuing system degrades its performance considerably. The results reported in [192] quantifies this conclusion by comparing the performance indicators of a queuing system under both exponential and Pareto distributions (as an example for the heavy-tailed distribution). The results show that using the Pareto distribution imposes up to a threefold increase in the average length of customers in the queue and doubles the average waiting time compared to the exponential distribution.

6-2-3 Limitations of employing the binary exponential back-off mechanism in CSMA-CA protocols

Although the BEB has been widely used in the design and development of the CSMA-CA protocols, a number of studies reported in the literature highlight limitations of using this scheme in design the CSMA-CA protocols. This subsection reviews some of these results.

The high level of delay of BEB is the key limitation that has been reported in many works. An explanation for this high level of delay has been carried out using different means. The work presented in [5,193-197] attributes this high level of delay to the unfairness of BEB, since a node that finishes transmission resets its back-off window to the minimum value it can access the channel faster than those nodes that are still waiting to access the channel. This behaviour makes the some nodes dominate channel utilisation which in turn increases the delay over the network and reduces the throughput. Approaches to resolving the unfairness have been proposed in a number of works [5,193-197]. Other studies relate the high level of delay of BEB to an increase in the number of nodes contending to access the channel. These studies consider a single hop network with homogenous traffic patterns. It has been shown in [198] that the optimal performance of network can be obtained if the back-off windows are adjusted to the number of active nodes; however, this study provides no details about the initial back-off window and its incremental policy. Our work presented in [5] considers the same situation and proves that the optimal performance can be obtained if the channel assessment rate of a node is adjusted to the reciprocal of the number of nodes contending to access the channel.

The second major limitation of the BEB scheme is its poor design space, i.e., BEB depends on doubling the windows of a uniform distribution to resolve the collision or contention to accesses. This doubling ignores the relationship between the main causes of the failure and the suitable action required to resolve it. For instance, it can be seen that a collision due to hidden nodes cannot be resolved by doubling the back-off windows as occurrence of this type of collision depends on the length of packets. One of the earliest schemes proposed to resolve hidden node collisions is the Request to Send/Clear to Send (RTS/CTS) algorithm proposed by the IEEE 802.11 standard [79]. Although this algorithm can provide some improvements, these improvements are weighed against the overhead imposed by exchanging the RTS and CTS packets.

These limitations motivate researchers to propose modifications or enhancements for BEB or even replacement it with a more powerful back-off strategy. The next section explores some of these works.

6-3 Related Work

This section is devoted to explore the main back-off approaches that are proposed in the literature, as there are plenty of these approaches; they are divided into three groups. Section 6-3-1 considers BEB as specified in the standards. Section 6-3-2 covers the X increase-Y decrease algorithms and section 6-3-3 explores those algorithms that use non-uniform distributions.

6-3-1 Binary Exponential Back-off standards

The Distributed Coordination Function (DCF) of the IEEE 802.11 standard [79] is one of the earliest CSMA-CA protocols that is specified especially for wireless networks. In this protocol, a node contends to access the channel by assessing it continuously. When the channel is found busy, the back-off duration is decreased, otherwise the duration is frozen until the channel becomes idle. The back-off duration is chosen based on a uniform distribution and the window of this distribution is doubled following a collision, and is reset to the minimal back-off exponent after successful transmission or upon dropping a packet. While this protocol can provide good throughput, the energy consumed ceaselessly in sensing the channel limits its applicability in low power networks such as WSNs. More discussion has been introduced in section 3-4-1.

Based on the requirement for saving the energy consumed in sensing the channel, the IEEE standards community introduced the IEEE 802.15.4 CSMA-CA protocol [16].

This protocol has been overviewed and investigated deeply in chapter 5; the results of this chapter illustrate the poor performance of this protocol over multihop WSN.

6-3-2 X increase Y decrease Back-off scheme

The BEB algorithms outlined in the previous sections are considered as exponential increase-reset decrease algorithms, as a node increases its back-off window exponentially after each collision or busy channel assessment and resets the back-off window to its minimum value after successful transmission or dropping a packet. As this behaviour can provide unfairness problems in homogenous single hop networks, a number of proposals have been devised to address this problem. The main theme of these proposals is to replace the exponential increase-reset decrease mechanism with a mechanism that can alleviate the unfairness. These works can be typified as X increase-Y decrease back-off scheme, some of them are summarised in this section.

The Multiplicative Increase and Linear Decrease (MILD) algorithm [193] was developed under the Medium Access with Collision Avoidance (MACAW) scheme. The MILD algorithm allows nodes to share their back-off parameters by including these parameters in all transmitted packets. In the MILD algorithm, a node upon a collision increases its back-off window by multiplying it by 1.5 instead of doubling it as in the traditional BEB scheme. Conversely, when a node delivers a successful packet then it decreases its back-off window linearly; while all other nodes that overheard the successful packet use the back-off parameters specified in this packet. The MILD mechanism is based on the assumption that a successful node has approached the back-off parameters that match the contention level of its neighbouring area. Although the MILD algorithm alleviates the unfairness problem found in the traditional BEB algorithm, it suffers from a number of key shortcomings. Firstly, MILD reduces the throughput of the network as the MILD algorithm adds a new header including the back-off parameters; this new header increases the overhead and the probability of collisions which in turn reduces the throughput. Secondly, the MILD algorithm introduces the back-off interval migration problem which occurs when two or more areas with different traffic loads are adjacent; as nodes in these areas propagate their back-off parameters, nodes contend to access the channel using wrong parameters which in turn degrades the performance of the networks.

The Linear MILD (LMILD) algorithm [194] has been proposed to overcome the shortages found in the MILD algorithm. The LMILD algorithm maintains the linear decrease policy of MILD but introduces a different increment policy in which the collided nodes increase their back-off windows multiplicatively, whilst other nodes that overheard the collision increase their back-off windows linearly. However, the adjustment for the back-off parameters of nodes that overhear the packets requires determining whether the packet is collided or not which is obtained from physical layer. Assessment of the performance of LMILD demonstrates that it outperforms MILD in terms of the fairness and the back-off interval migration problem. However, these advantages are dominated by the accuracy with which a physical layer can detect the on-going collisions.

The sensing Back-off Algorithm (SBA) [195] is another approach that proposed to overcome the limitation of BEB. In SBA a node that experiences a collision, multiplies its back-off window by a constant $\omega > 1$. While the transmitter and receiver of successful packets multiply their back-off windows by a constant $\beta < 1$ and all nodes that have overheard the successful transmission decrease their back-off windows linearly by the time required to transmit a packet. It has been shown that SBA provides better fairness readings compared to the MILD and LMILD, however, this improvement relies heavily on initial setting of these parameters.

6-3-3 Binary exponential back-off with non-uniform probability distributions

The Sift algorithm [196] was designed especially to cope with the spatially correlated contention of the event driven WSNs. The Sift algorithm resolves the correlated contention that occurs when more than two sensors report identical events for the same coordinator. The fundamental logic of the Sift algorithm is to replace the uniform distribution of the traditional BEB scheme with a skewed distribution in order to minimise the collision probability of the first few slots. In the Sift algorithm each node maintains a parameter called the probability of transmission whose initial value is set depending on the number of nodes. A node after setting these parameters monitors the activities of its neighbours and decreases the probability of transmission upon overhearing a transmission from a neighbour. Hence the contention to access to the shared channel follows a geometrically decreasing mechanism. Although the Sift algorithm is able to reduce the delay for the event-driven applications, it consumes a large amount of energy as it requires a node to listen to the channel continuously. Furthermore, Sift ignores the effect the hidden nodes and exhibits poor performance over dense networks. Following the same approach, the work presented in [197] replaces the uniform distribution of BEB with a geometric distribution whose mean is same as the mean of the uniform distribution of BEB. The results reported in [197] demonstrate slight improvements in terms of throughput compared to the traditional BEB scheme.

In collusion, it can be seen that most of the work reported in the open literature attempt to improve specific aspects of the BEB scheme without considerations for typical characteristics of multihop networks, e.g. heterogeneous traffic patterns, ad-hoc topology configurations and multi path load balancing. Moreover, most of these works are either require a node to listen to the channel continuously or impose additional overheads. Since both of these approaches are not suitable for WSNs, there is a need for a novel design methodology that considers the limited resources of these networks and caters for challenging characteristics of multihop network. This chapter exploits the principles of the inferential statistics in conjunction with the statistical characteristics of Gamma distribution to design an effective MAC protocol. The next subsection explores the fundamental properties of Gamma distribution while the underlying approach of the proposed protocol is discussed in section 6-5.

6-4 Statistical characteristics of the Gamma distribution

Introduction of Gamma distributions is dates back to the seminal work proposed by Pearson in 1895 [199] in which all probability distributions are classified into five types according to their statistical moments. The Gamma distributions were defined under type III, as they provide limit range skewness curves in one direction [182]. Since that time, the Gamma distributions and their processes have found their way into wide range of disciplines including statistical physical problems, reliability analysis, Bayesian inference applications and economical models. Comprehensive surveys can be found in [200-203] and their reference lists.

The Probability Density Function (PDF) of the Gamma distribution is given as [182]:

$$f_X(x|r,s) = \frac{r^s}{\Gamma(s)} x^{s-1} e^{-rx}; x, r, s \in (0, \infty) \tag{6.1}$$

where X is the random variable, $\Gamma(s)$ is the Gamma function [182], r is the rate parameter and s is the shape parameter of the Gamma distribution. One of the most important properties of the Gamma distribution is its versatility in representing a wide variety of other distributions just by adjusting the values of the shape and rate parameters. Figure 6.1 illustrates the PDFs of Gamma distribution for selected rate and shape values.

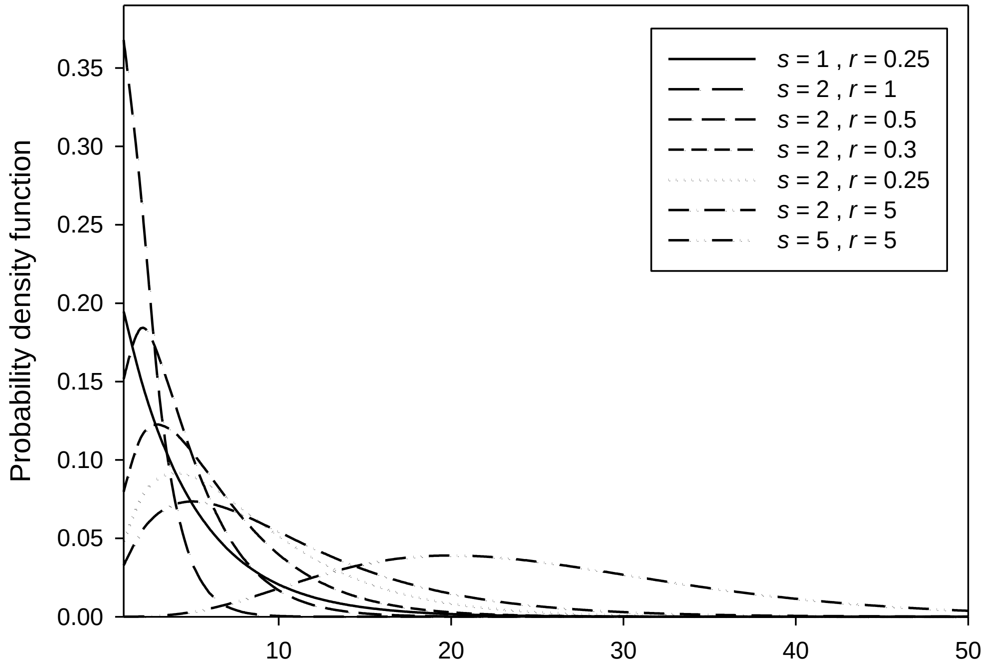


Figure 6.1 PDFs of Gamma distribution under different shape and rate parameters

Figure 6.1 shows that that the shape parameter (i.e., s) has a great effect on the skewness of the PDF. It can be seen that setting the value of the shape parameter to 1 reduces the Gamma distribution to an exponential distribution whose rate is adjusted to the same value specified by the parameter r (i.e., the rate) of the Gamma distribution. It can also be seen that setting the shape parameter of a Gamma to a value larger than 1 reduces its skewness, while increasing the shape parameter speeds up approaching the Gamma distribution to normal distribution. In terms of the rate parameter, figure 6.1 shows that the rate parameter controls the dispersion of the Gamma distribution through stretching or squeezing the PDFs to the left or the right according to overall support of the random variable. Figure 6.1 shows that setting the value of r to a large value squeezes the PDF to the left and risen its height while setting the value of r to a small value stretches the PDF of the Gamma distribution to the right and reduces its height.

Besides this versatile feature, the Gamma distribution has some appealing statistical characteristics that are highly required to design an effective MAC protocol for WSNs. Some of these characteristics are:

1. The Gamma distribution is the conjugate prior for most of the probability distributions that are used to model the traffic over communication networks. e.g., Exponential, Poisson, Normal and even the heavy tailed distribution such as Pareto, Weibull and lognormal distributions. In has been shown from figure 6.1 that when $s = 1$, then Gamma distribution is reduced to the exponential distribution. Similarly, the Erlang distribution can be obtained by adjusting the

shape parameter of the Gamma distribution to positive integer values, i.e., $r \in \mathbb{Z}^{+}$. Considering that the Erlang distribution is used to model the waiting time until the occurrence of the k^{th} event of a Poisson process whose rate is set at r, reveals the relationship between the Gamma distribution and the Poisson process. Setting the value of the shape parameter of the Gamma distribution to half of the freedom parameter of the chi-squared distribution and keeping the rate parameter to 2 facilitates representing chi-squared distributions as a Gamma distribution. The Pareto distribution can be expressed as a mixture of the exponential distributions with Gamma mixture weights [182]. Furthermore, the gamma distribution can be used to model the traffic characteristics of the real-deployment networks. In [205-207], the authors conclude that the Gamma distribution is an excellent candidate to represent Internet traffic. The work presented in [206] demonstrates the ability of the Gamma distribution to model the self-similar traffic patterns. Moreover, the work presented in [208] shows that using the Gamma distribution as an inter-arrival and service function of a given queuing system provides the isomorphism which implies that majorly of other queuing outcomes can be approximated through it.

2. The Gamma distribution has the infinitely divisible property [204] which means that a random variable generated according to Gamma distribution can be expressed in terms of n independent random variables each of which is distributed according to a Gamma distribution.
3. The Gamma distribution is closed under convolution [182], i.e., the sum of n independent random variables, each one of them distributed according to a Gamma distribution is a random variable that follows a Gamma distribution.
4. The tail of a Gamma distribution decays exponentially which minimises the probability of rare events and provides a low level of variation around the mean. This feature results from the fact that all moments of the Gamma distribution vanish and converge around its mean e.g., the coefficient of variation of the Gamma distribution is given as s^{-1} while the skewness coefficient is $2s^{-0.5}$.
5. The Gamma distribution is unimodal which means that when a random number is generated according to a Gamma distribution, these variables are concentrated in a bounded range which makes the inference process easy.

6-5 Underlying approach of the proposed MAC protocol

The main contribution of this chapter is to design an effective MAC protocol for multihop WSNs that can achieve three concurrent goals: increasing the through of network, reducing the end-to-end delay packets and minimising the energy consumed during service a packet independently of topology configurations, traffic patterns or routing policies. The fundamental logic of the proposed protocol is to enable each node to generate its back-off intervals according to a Gamma distribution whose parameters are inferred for the status of the channel. In particular, a node infers the rate with which its channel is busy and the type of collision (either due to hidden or non-hidden nodes). A node then uses the inferred values as the rate and shape parameters for the Gamma distribution.

Justification for using the Gamma distribution to generate the back-off interval can be acquired by considering the versatility aspects of the Gamma distribution and the fact that the Gamma distribution is the conjugate prior for most of the probability distributions that are used to model the traffic over communication networks. Hence using the Gamma distribution whose rate parameter is set according to the rate with which the channel is found busy enables a node to back-off with high probability when the channel is potentially busy and to assess the channel when it is presumably idle. This in turn saves the energy consumed in assessing busy channels and more importantly allows a node to deliver its packets without deferring them for random periods. Another key advantage of enabling a node to generate its back-off intervals according to the rate with which the channel is found busy is that such adjustment can reduce the end-to-end delay of packets and thereby increases the throughput of the network. Furthermore, a justification for using the shape parameter to mitigate the collision probability is that the shape parameter controls the skewness of the Gamma distribution which in turn enables colliding nodes to manage their channel assessment activities without spending long back-off intervals. Justification for using the inference process is that this process enables a node to cater for its distinct traffic demands and to adapt its contention parameters in accordance with the traffic varying of network adequately. More importantly, the inference process prolongs the lifetime of nodes by enabling them to gather the required information without a need to listen for the channel for long periods or overwhelming the network with control packets.

Besides the aforementioned advantages, using the Gamma distribution to generate the back-off intervals yields a Gamma process. The Gamma process is a pure-jump stochastic process with stationary and independent increment [200-203]. A Gamma process is a mathematically tractable process whose sample paths increase monotonically, hence construction of its likelihood is straightforward. Moreover, the Gamma process features a much lower variation compared to the power law process used in BEB. Reducing the variation of the back-off process has the advantage that it improves the traffic flow over the network. Due to the infinite divisibility property of the Gamma distribution, the value of the Gamma process at any instance follows a Gamma process. This property is particularly important in designing the proposed back-off scheme as it ensures that the back-off intervals generated from the proposed model will always have the appealing statistical characterises of Gamma distributions. An illustration for the underlying approach of the proposed protocol is shown in Figure 6.2.

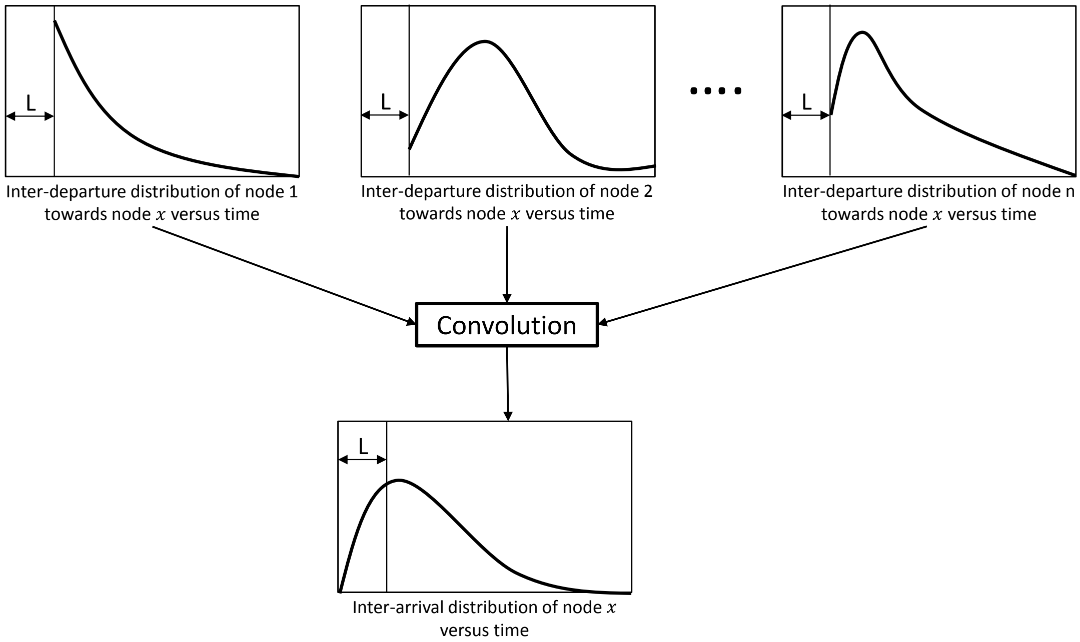


Figure 6.2 Underlying approach of the proposed protocol

Figure 6.2 shows the inter-departure distributions of a number of nodes that contend to send their packet to a common receiver where L refers to the length of packets in time unit. It is worth noting that all of these inter-departure distributes are truncated by L since the departure time between two consequence packets from the same nodes has to be separated by at least L, hence the probability of the event that inter-departure is less than equal L is zero. Using the independent assumptions [163], the inter-arrival distribution at the common receiver is the convolution of all inter-departure distributions of all contenders. As the convolution is a smooth operator, the inter-arrival distribution at the common receiver will be extended over larger intervals including $[0, L]$. These probabilities in the interval $[0, L]$ represent the probability of overlapping at the common receiver, i.e., the collisions. Hence the obvious solution is control the inter-departure distributions of the contenders in such a way that reduces the probabilities of overlapping which is the fundamental logic of the proposed protocol. Recall that equation (5.11) shows that the inter-departure distribution is a function of inter-arrival distribution and service distribution; as the inter- arrival distribution is application-specific then the proposed protocol uses the Gamma distribution to control the contending behaviour of nodes by allows each node to adjust the shape and rate parameter based on the states of the channel. A detailed description for such adjustment is given in the pseudo-code in the next section.

6-6 Pseudo-code of the proposed protocol

This chapter follows the same assumptions and notations presented in chapter 5 that are recalled here, a multihop network consisting of a set of nodes Υ distributed randomly over a given area and operating at a fixed data rate denoted by D bits per slot. For an arbitrary node n where $n \in \Upsilon$, $\Psi(n)$ is a set of nodes that share the channel with node n, and λ_n is the inter-arrival rate of packets to node n. This rate includes the aggregated packets forwarded to it from other nodes as well those packets generated by the node itself.

This chapter proposes to use the rate parameter of a Gamma distribution to adjust the rate with which the back-off durations are generated and use the shape parameter of a Gamma distribution to tune the back-off duration after collisions. The pseudo-code of the proposed CSMA-CA protocol as shown in figure 6.3 consists of five functions: Startup, Initialization, Channel Access, Update Rate and Update Shape. The proposed protocol is specified to be a global CSMA-CA protocol that can be used with any network configurations hence we use ε to denote the duration required to assess the channel in slots [5].

Startup ( )
  $r \leftarrow 1, T \leftarrow 0, B \longleftarrow 0, L \longleftarrow 0, \Xi \leftarrow \emptyset, \omega \longleftarrow 0$
end Startup ( )

Initialisation ( )
  $c \leftarrow 0, m \leftarrow 0, s \leftarrow 1$
end Initialization ( )

Channel Access ( )
  If ($m \leq M$ and $c \leq C$)
    Assess the channel for duration ε
    If (channel is idle)
      Commence transmission
      Update shape ( )
    else
      Back-off for a duration $\Gamma(r, s)$
      $m \leftarrow m++$
      Update rate ( )
    end if
  else
    Drop a packet
  end if
end Channel Access ( )

Update rate( )
  If (packet is received)
    $r \longleftarrow r - \dfrac{(L + Length\ of\ packet\ in\ bits)}{\mathrm{T} \times \mathrm{D}}$
  else

$$r \longleftarrow r - \frac{(B + Backoff\ duration\ in\ seconds)}{T}$$

end if
end Update rate( )

Update shape ( )
If (acknowledgment is not received)
$c++$
Ξ [receiver address]++
For each $y \in \mathbf{\Omega}$

$$\omega = \frac{\omega + \Xi[y]}{|\Xi|}$$

End for
if (Ξ [receiver address]) $\leq \omega$
$s = s + \varepsilon$
else
$s = s + L$
end if
end if
end Update shape ( )

Figure 6.3 Pseudo-code of the proposed CSMA-CA protocol

6-6-1 Startup function

The first function, Startup, is executed once during the lifetime of a node. This function sets the following parameters to their initial values: the rate of the Gamma distribution (r) is set to 1, Time counter (T) to 0, total time intervals spent in back-off by a node (B) to 0 and the aggregated length of received packets (L) to 0. The r parameter is used to adjust the rate parameter of the Gamma distribution with which back-off intervals are generated. The T parameter is a counter that measures the time duration since a node started up in slots. The B and L parameters are used to sum the total time that a node has spent backing-off and the total length of received packets in bits respectively. In addition to these parameters, an arbitrary node n maintains a set of a dimension $|\Psi(n)|$ called the collision number set Ξ to an empty set. This set is indexed by the receiver address and used to record the number of collisions occurring at each receiver. From the collision number set, the average number of collisions encountered by a node (transmitter) is updated after a collision and kept in the variable (ω).

6-6-2 Initialisation function

The Initialization function is executed at the beginning of each CSMA-CA process to set the contention parameters to their default values. The shape of the Gamma distribution (s) is set to 1, the number of times the channel has been assessed as busy

(m) and the number of collisions (c) are set at 0. The shape parameter is used to tune the shape of the Gamma distribution that is employed to generate the back-off durations. The m and c parameters are used to control the number of stages of the CSMA-CA process. The maximum values of these parameters are network wide and indicated by M and C respectively.

6-6-3 Channel access function

The Channel Access function starts by comparing the number of collisions (c) and number of busy channel assessments (m) with their maximum values (C and M respectively). If the c or m parameters exceed their maximum values then a node drops a packet. Otherwise, a node assesses the channel for duration ε. This value is set according to the physical specification of the network which widens the applicability of the proposed protocol. If a node finds an idle channel, it commences transmission and then executes the Update Shape function. Otherwise, if a node finds a busy channel, it backs-off for a random duration generated according to the Gamma distribution. After this back-off period has elapsed, a node increments the number of busy channel assessments m by one and goes to the Update Rate function.

6-6-4 Update rate function

The Update Rate function is used to adjust the rate of the Gamma distribution that is used to generate the back-off durations. In the proposed protocol, the optimal value of this rate is defined as the rate which keeps a node backing-off while the channel is busy. The advantage of this adjustment is twofold: saving the energy consumed in sensing a busy channel and reducing the end-to-end delay by avoiding deferring the packet for another back-off duration following busy channel assessment. In order to determine the value of the optimal rate, we consider the channel utilization from the viewpoint of node n which is denoted here by ϑ_n and given by:

$$\vartheta_n = \frac{\sum_{\forall x \in \Psi(n)} \lambda_n}{D} \tag{6.1}$$

ϑ_n is the rate with which node n finds a busy channel and hence the optimal rate for generating the back-off interval is ϑ_n , i.e., $r_n = \vartheta_n$.

Determining the value of r_n requires a node to know the traffic rate of all nodes sharing the channel with it, which is expensive in terms of energy and computation as a node has to monitor and analyse the channel at all times. However, in this chapter, a simple and accurate algorithm is proposed to approximate the optimal value of the rate parameter.

From the viewpoint of node n, the channel is considered busy if it is occupied with transmissions from its neighbours, either if this transmission is intended to node n itself or another node. Although a node can determine the rate with which it received traffic exactly, it can infer the traffic that is sent from its neighbours through the back-off and channel assessment processes. Hence a node executes the Update Rate function in two situations: upon receiving a packet and upon assessing the channel as busy. Recall that the initial value of the rate parameter is one, as in the startup phase a node has no accurate information about the traffic characteristics of its neighbours and hence it assumes that it is the sole node operating on the channel.

When a node receives a packet, it updates the aggregate length of received packets (L) and divides the resultant by the maximum channel utilization over duration (T) expressed in units of bits, i.e., the maximum number of bits that could be sent over the channel if the network operated in the saturated region which is computed as a product of data rate and time counter. This computation methodology has been proposed as it quickens the convergence of the rate parameter to its ergodic value even when the traffic is highly fluctuating. Following the same methodology, a node updates the value of the rate parameter upon assessment of the channel as busy, except that in this case a node aggregates the total time spent in back-off and divides it by the time counter.

6-6-5 Update shape function

The Update Shape function is used to adjust the shape of the Gamma distribution that is used to generate the back-off intervals. The main objective of this adjustment is to eliminate or at least reduce the collision probability. From the viewpoint of the transmitter, a collision occurs if there is a simultaneously transmissions from other nodes towards the same receiver. These potential transmitters can be categorised into two main groups. Firstly, the hidden nodes group which contains all nodes that are not within the carrier sensing of the transmitter and secondly, those nodes that are within the carrier sensing of the intended transmitter (called here non-hidden nodes). Further details are given in section (5-5-6).

Non-hidden node collisions happen when the transmitter and at least a member of those nodes that are sharing the channel with it, i.e., $\forall x \in \Psi(n)$ accesses the channel at the same instance. I.e., when differences between the back-off timings of those nodes that are involved in a collision are less than the time required to consider the on-going transmission of others. Hence the obvious solution is to shift the shape parameter by the time required to assess the channel ε. Conversely, a hidden node collision occurs when at least a neighbour of a receiver commences transmission while the receiver is receiving a packet, i.e., transmissions from two or more nodes overlap at any time during the reception time. Hence shifting the shape parameter by the length of a packet can avoid this type of collision.

Determining whether the collision occurred due to hidden or non-hidden nodes can be estimated by considering that a non-hidden node collision affects all transmissions of a node (as it depends on the transmitter), while a hidden node collision affects some traffic (as it depends on the receivers). The Update Shape function allows a node to infer this characteristic by comparing the number of collisions that occur with a particular receiver address, $\Xi[\text{receiver address}]$ to the average number of collisions encountered by a node ω. The Update Shape function uses acknowledgment packets as an indicator of the occurrence of a collision. If the acknowledgment packet is received, then a node supposes that the packet has successfully arrived and exists in the CSMA-CA process. Otherwise, if the acknowledgment packet has not been received, a node will increment the number of collisions c, the value of the member of the collision set Ξ that corresponds to the receiver address and the average number of collisions ω. Following this update, a node compares the number of collisions that have occurred at the receiver with the average number of collisions ω.

If the number of collisions of that receiver is more than the average number of collisions then a node increments the shape parameter by the length of packet (supposing that a hidden node collision has occurred). On the other hand if the number of collisions per receiver is less than or equal to the average number of collisions per node then a node assumes a non-hidden collision has happened and hence it increments the shape parameter with the time required to assess the channel.

6-7 Results and discussion

This section evaluates the benefits of the proposed Gamma based CSMA-CA protocol versus the IEEE 802.15.4 CSMA-CA standards, the CSMA-CA parameters of both protocols are given in table 6.1. The outcome of the simulation scenarios are labelled using the five tuple $\langle k, N, X, Y, Z \rangle$ where the k parameter refers to the connectivity degree (the average number of mutually independent paths between non-neighbouring nodes). The parameter N denotes the average number of nodes per channel, and the parameter X denotes the inter-arrival distribution of times between packets that are generated internally by each node. This assessment uses three inter-arrival distributions: CBR for Constant Bit Rate, EXP for exponential distribution and WBL for the Weibull distribution. The parameter Y specifies the rate with which a node generates its internal packets given in packet per slot. Finally parameter Z defines the traffic distribution scheme which can take two possible values U, when all traffic that is generated internally by a node is distributed evenly to other nodes with the network and A to indicate that all traffic generated by all nodes within the network is destined to a single final destination. Further discussions about these parameters or others that are related to duration of simulated sessions, physical channel and network topology are provided in section 4-4.

Table 6.1 CSMA-CA parameters for Gamma and IEEE 802.15.4 simulation scenarios

Parameter	802.15.4	Gamma
Data rate	250kbps	250kbps
Channel assessment time	$320\mu s$	$320\mu s$
Initial back-off exponent	3	N/A
Maximum back-off exponent (M)	5	N/A
Maximum number of busy channel assessment	5	5
Maximum number of retransmission (C)	3	3
Initial shape parameter	N/A	1
Initial rate parameter	N/A	1

6-7-1 End-to-end delay

Figures 6.4 and 6.5 illustrate the average end-to-end delay of the proposed protocol (indicated by Gamma in these figures) compared to the IEEE 802.15.4 standard under different network configurations.

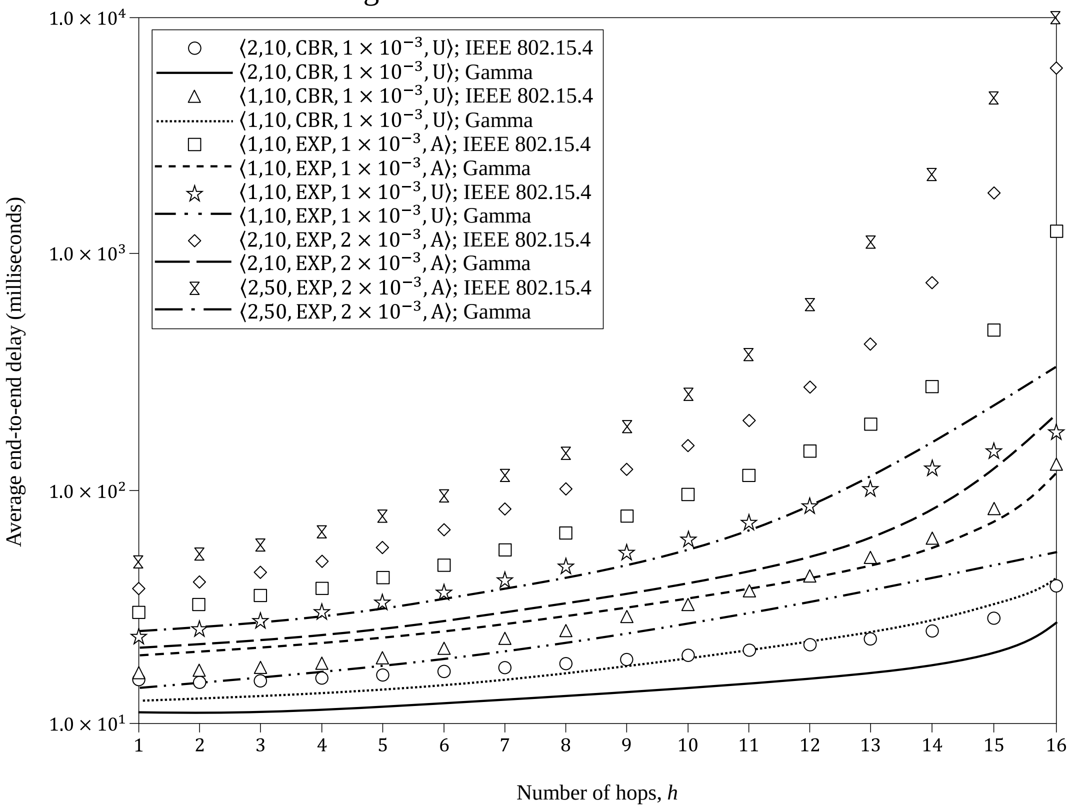


Figure 6.4 Average end-to-end delay of Gamma CSMA-CA and IEEE 802.15.4

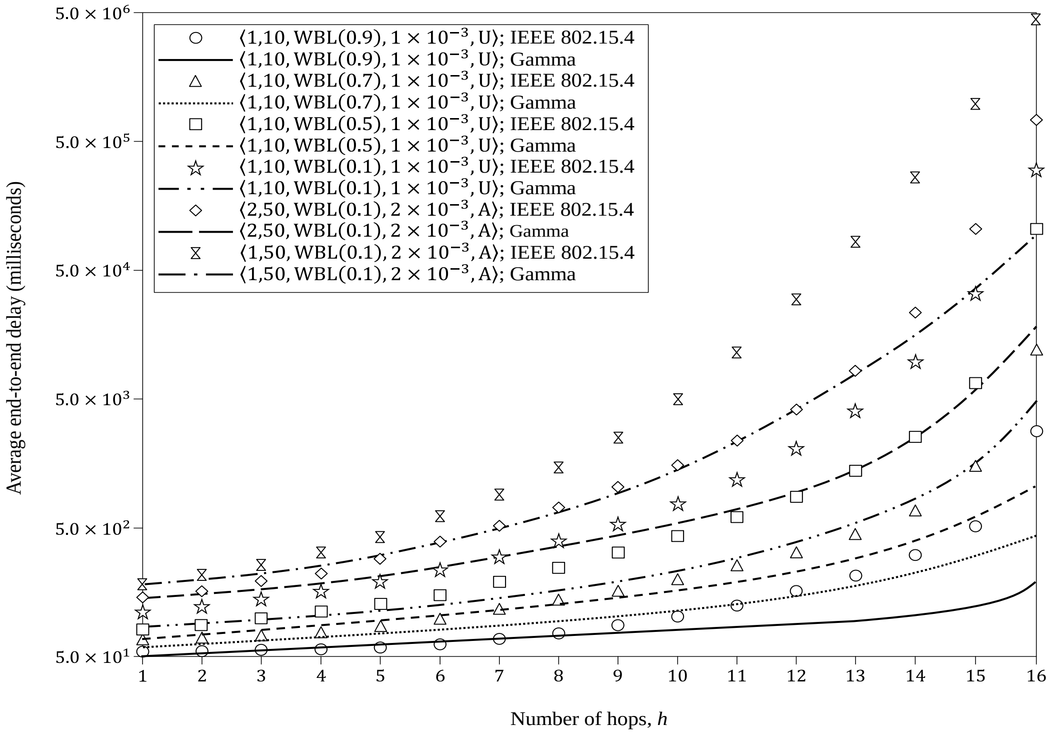


Figure 6.5 Average end-to-end delay of Gamma CSMA-CA and IEEE 802.15.4 under heavy tailed inter-arrival distributions

The results of figures 6.4 and 6.5 demonstrate that, under all cases, the proposed protocol provides the minimal average end-to-end delay readings compared to IEEE 802.15.4. These characteristics are attributed to the ability of the proposed protocol to manage the contention parameters of each node according to its traffic demands which in turn facilitates delivering timely packets. Another main reason for the significant reduction in the average end-to-end delay of the proposed protocol is its collision resolution mechanism that enables a node to resolve the underlying causes of collisions and thereby this mechanism improves the successful traffic flow across the network. Conversely in the Binary Exponential Back-off based protocols (i.e., the IEEE 802.15.4 standards), each node generates its back-off intervals according to a uniform distribution that does not reflect the traffic patterns of the multihop networks. Moreover, this protocol doubles the back-off windows whenever a node fails to access the channel which leads to defer transmission for long periods. Another important reason for the substantial increase in the average end-to-end delay of the IEEE 802.15.4 CSMA-CA protocol over number of hops is its collision resolution mechanism. This protocol requires colliding nodes to reset their back-off windows to the same initial value (which is typically small) and then precede their contention to access the channel. This mechanism magnifies the end-to-end delay as reducing the back-off window of colliding nodes increases the probability that two or more members of these nodes commence their transmission simultaneously which leads to increase the collision probabilities and the end-to-end delay.

Deep insights into the characteristics of the end-to-end delay over multihop networks can be obtained by measuring its Squared Coefficient of Variation (SCV), i.e., the ratio between the variance and the squared average of the end-to-end delay. The SCV metric is used to quantify the variability of the end-to-end delay, the higher SCV the greater dispersion of the end-to-end delay around its average value and vice versa. Figure 6.6 shows the characteristics of this metric for both the proposed and the IEEE 802.15.4 CSMA-CA protocols under different network configurations.

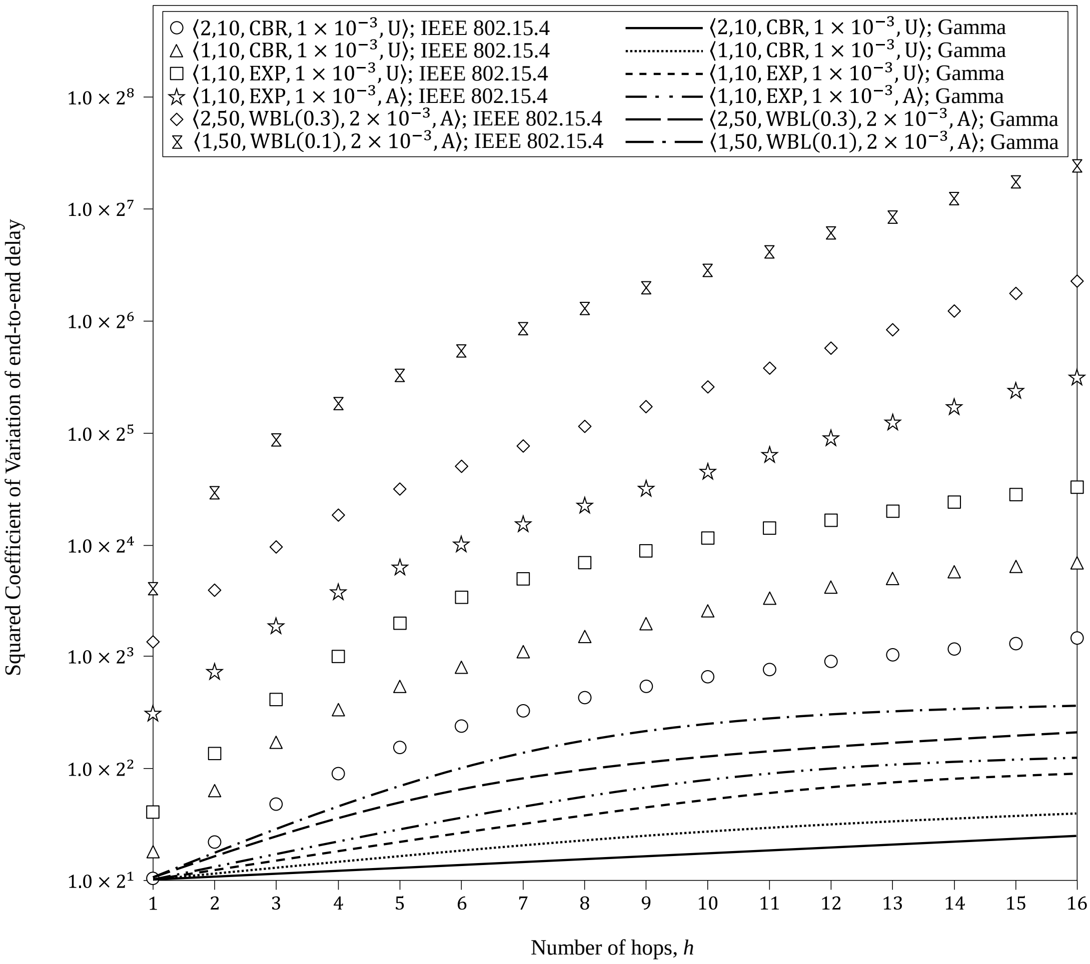


Figure 6.6 Squared coefficient of variation of end-to end delay of Gamma CSMA-CA and IEEE 802.15.4

Results of figure 6.6 demonstrate that the IEEE 802.15.4 CSMA-CA protocol provides the highest level of variations under all scenarios. For instance, when there is on average two mutually independent paths between non-neighbouring nodes, the average number of nodes per channel is 10 each of which generates its inter-arrival times between packets according to Constant Bit Rate (CBR) with low rate (0.001packet/slot) and these packets are distributed evenly to other nodes, i.e., the case $\langle 2{,}10, \mathrm{CBR},\ 1 \times 10^{-3}, \mathrm{U} \rangle$. Under these conditions, the SCV of the end-to-end delay provided by the IEEE 802.15.4 CSMA-CA protocol increases dramatically from about 2 over single hop to more than 7 over 16 hops. Conversely, the squared coefficient of variation of the proposed protocol of this scenario does not exceed 3 over 16 hops. An explanation for this characteristic can be acquired by considering the fact that the BEB mechanism specified in IEEE 802.15.4 CSMA-CA is a power law process which suffers from high variation due to its weak ergodicity [209]. Moreover, this behaviour is the underlying causes of being the medium access distribution of IEEE 802.15.4 CSMA-CA is heavy tailed. On the other hand, the proposed protocol is based on a Gamma process which has a very low level of variation [200]. For the same reason it can be seen that using the EXP or WBL distributions to generate the inter-arrival times between packets leads to great increases in SCV of the IEEE 802.15.4 CSMA-CA protocol compared to slight increases in SCV of the proposed protocol. This characteristic is attributed to the ability of the proposed protocol to mitigate the variation of different inter-arrival distributions, as the Gamma distribution is the conjugate prior for these inter-arrival distributions.

Figure 6.6 also shows that reducing the connectivity degree or sending all packets to the same final destination (indicated by “A” in the last element of tuple) magnifies SCV of IEEE 802.15.4 greatly whereas it keeps the SCV of the proposed protocol about the same. This is attributed mainly to that reducing the connectivity degree or using the "A" routing policy lessens the routing redundancy which in turn increases the collision probability of the IEEE 802.15.4 protocol and variability of the end-to-end delay. By contrast, the proposed protocol is able to mitigate the collision probability over multihop links and hence it can reduce the readings of the SCV metric. Considering the relationships between SCV and probability of collision and noting that increasing the traffic intensities or average number of nodes per channel increases the probability collision justify the high variations of the IEEE 802.15.4 compared to the proposed protocol.

6-7-2 Throughput

Figure 6.7 illustrates the throughput of the proposed and IEEE 802.15.4 CSMA-CA protocols versus number of hops for different scenarios. In general, it can be seen that the characteristics of throughput follow the same trends as the average end-to-end delay in terms of the effects of the connectivity degrees, average number of nodes per channel, traffic intensity and routing policies. It can be seen also from the results of this figure that the throughput of the IEEE 802.15.4 CSMA-CA protocol decreases sharply with increasing number of hops. The key reasons for this characteristic are the increasing of the collision probability and of the variation of the IEEE 802.15.4 protocol over multihop; such increasing minimises the channel utilisation and yields low throughput.

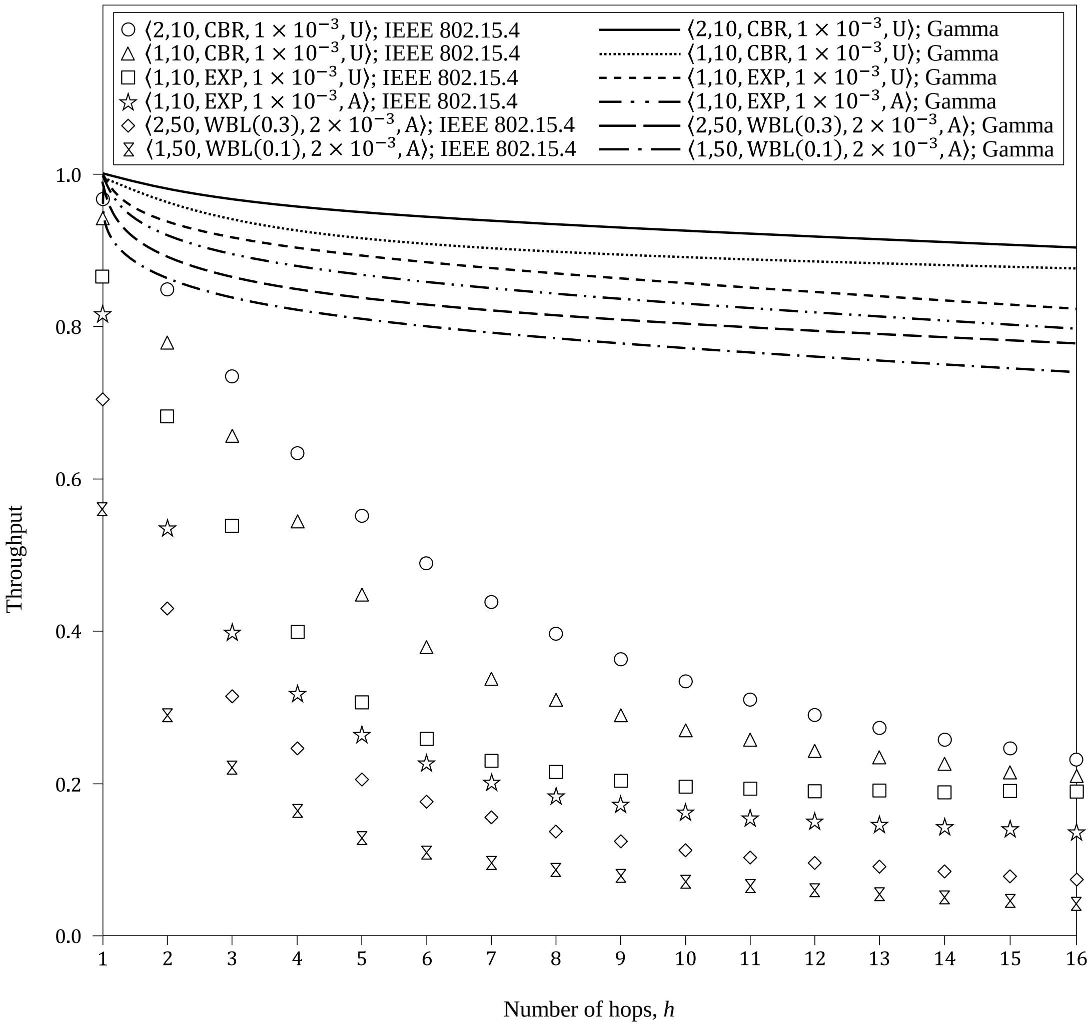


Figure 6.7 Throughput of Gamma CSMA-CA and IEEE 802.15.4

6-7-3 Energy

An assessment for the average energy consumed per packet of both IEEE 802.15.4 and the proposed protocols is presented in figure 6.8. This figure shows that the IEEE 802.15.4 protocol consumes the highest level of energy compared to the proposed protocol. This is attributed to that the service times of IEEE 802.15.4 are much higher than their peers in the proposed protocol which in turn requires a node to contend to access for the channel for longer periods. A further support for the high amount of the energy consumed in the contention procedures of the IEEE 802.15.4 protocol has been introduced in [251]. This study shows that nodes operating in a saturated single hop network consume around 25 % of their energy just in contention to access the channel. The increasing in energy per packet over multihop links is due to the increasing in the collision probability over these links, hence considering the ability of the proposed protocol to alleviate such probability explains the low energy expenditure of the proposed protocol compared to IEEE 802.15.4.

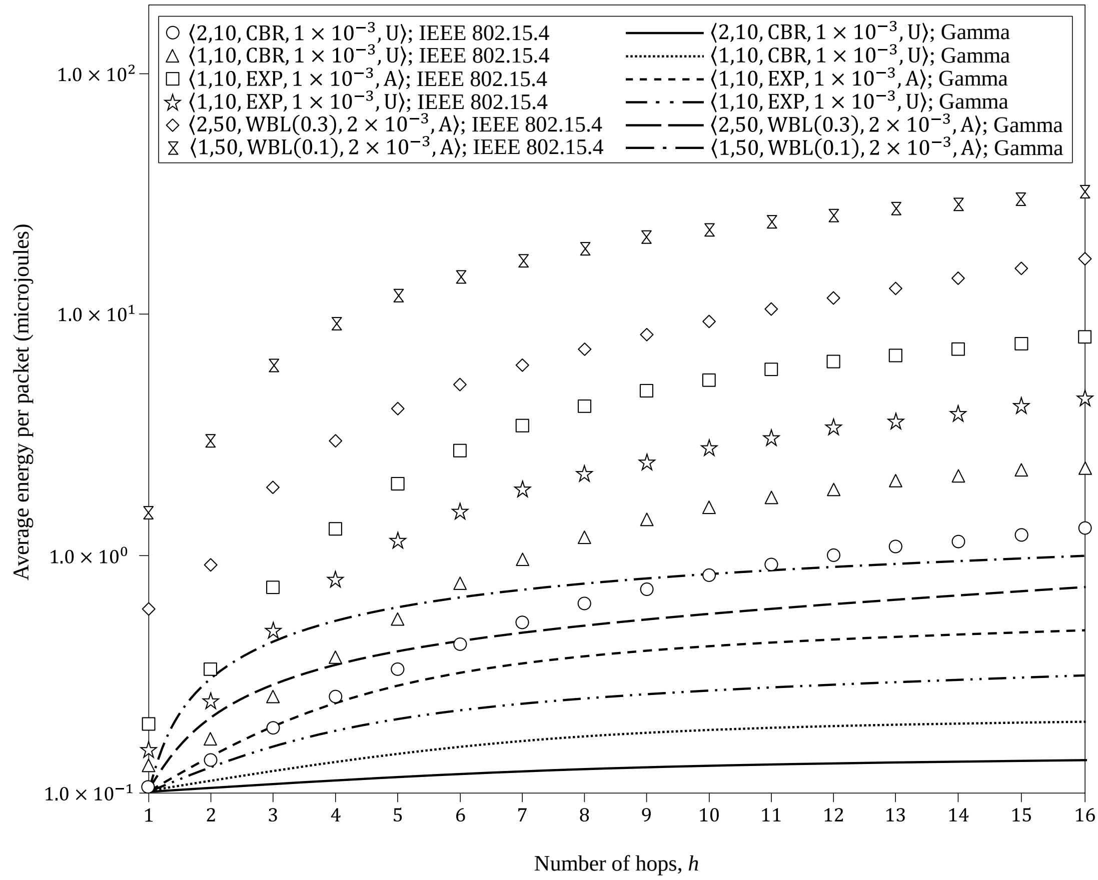


Figure 6.8 Average energy per packet of Gamma CSMA-CA and IEEE 802.15.4

6-8 Conclusion

This chapter has introduced the first main objective of this thesis which is to design an effective MAC protocol for multihop WSNs that can increase throughput, reduce end-to-end delay and save energy consumed during contention. The design methodology used in developing the proposed protocol is to provide a flexible contention scheme which reflects the dynamic characteristics of multihop networks, so each node can adjust its contention parameters to achieve the aforementioned goals. The new protocol replaces the uniform distribution that is used in the existing binary exponential back-off based protocol with a Gamma distribution which is more flexible and replaces the blind doubling of back-off windows by a more intelligent collision resolution algorithm. The simulation outcomes demonstrate the substantial benefits of the proposed protocol with respect to the IEEE 802.15.4 protocol. It is worth noting that although the proposed MAC protocol can improve the three aforementioned performance metrics compared to one of the state-of-the-art MAC protocol, this does not mean that the proposed protocol is always able to optimise all three metrics simultaneously. The protocol exploits the fact that it enables nodes to reduce the time required to deliver successful packets, can save contention energy (which prolongs lifetimes) ,and can improve the channel occupancy (which enhances the throughput).

Chapter 7

Duty Cycle Management Scheme for Multihop Networks Inspired by Artificial Chemistry

7-1 Introduction

The previous chapter introduced the first main step towards achieving the main objectives of this thesis by specifying a CSMA-CA (Carrier Sensing Multiple Access-Collision Avoidance) protocol that satisfies the sophisticated requirements of multihop networks. The advantages of the proposed CSMA-CA protocol have been assessed by comparing its performance readings with the IEEE 802.15.4 protocol. These comparisons demonstrate great benefits of the proposed protocol in terms of increasing the throughput of the network, reduction of the end-to-end delay of packets and minimisation of energy required to service a packet. Minimisation of energy required to service a packet can save a considerable amount of energy and prolong the lifetime of network; however, the main causes of energy waste are due to other sources, e.g., overhearing and idle listening. Hence, the need to develop a scheme that is able to mitigate the waste of energy from the aforementioned sources is important for attaining the main objectives of this thesis.

Multihop Wireless Sensor Networks (WSNs) like other low-power personal area networks consists of a number of tiny nodes that are powered by small batteries and it is infeasible in some cases to replace these batteries. The lifetime of these networks is entirely dominated by the ability of communication protocols to provide an effective scheme to reduce energy consumption. In general, energy consumption by a node can be classified into two main categories: firstly, energy that is required to cater for traffic, (e.g., transmissions and reception of those packets that need to be serviced by a node) and secondly, energy that is wasted while a node does nothing useful (e.g., overhearing when a node receives other transmissions that is not destined to it or idle listening when a node waits for potential packets). One of the key strategies that is used to mitigate this needless consumed energy is to manage the duty cycle of nodes, i.e., to let a node sleep (turns its transceiver off) for a certain period if it has no role to play during this future period.

Although a duty cycle scheme can provide a significant saving in wasted energy, it can cause serious performance degradation, e.g., reducing the throughput or magnifying the end-to-end delay if it is mismanaged. Discussion presented in chapter 3 showed that most of the duty cycle management schemes reported in the open literature are tailored to specific network characteristics. Hence there is a need to develop an effective and general multihop duty cycle management scheme that can operate independently of the topology, traffic pattern or routing policy and that can achieve three main goals: prolonging the lifetime of nodes, increasing the throughput of networks and reducing the end-to-end delay of packets.

The main reason for the shortage in the design of an effective duty cycle management scheme is the absence of a framework that is capable of accounting for dynamic interaction between nodes in the presence of varying traffic demands while being able to provide a means of measuring the level of optimisation of the aforementioned three objectives. An explanation for this challenge can be drawn by considering the characteristics of a typical multihop network. A typical wireless multihop network is a large-scale infrastructure-less network in which each node has its own traffic load and each transmitter receiver-pair has its own distinct view of the channel. A justification for this argument can be drawn by considering that a multihop network is ad-hoc in nature where the load of a typical node comprises both traffic generated by the node itself as well as traffic forwarded by other nodes for relaying. Moreover, as each communication is subject to interference from neighbours of the transmitter or receiver, and as each of these neighbours contends for a channel based on different loads, the conditions of the channel as seen by each transmitter-receiver pair differ.

This chapter overcomes the challenge in the design of an effective duty cycle management scheme for multihop networks through Artificial Chemistry (AC). AC is a mature multidisciplinary research field that investigates the fundamental principles of chemical processes and then exploits these principles to engineer effective computational paradigms [210-211]. AC connects different research fields including laboratory practice, mathematical modelling and computational algorithms. This chapter utilises the modelling aspect of AC, in particular, molecular statistical mechanics [212] to derive the time evolution of the consumed energy in multihop networks. It then employs the computational algorithm aspect of AC to design the effective duty cycle management scheme. The underlying approach that is used in this design is called artificial decision making [213-214]. According to this approach, chemical species are represented as a set of autonomous interacting systems in which each system can decide its future action by inferring and classifying the communications into useful and useless communications. The importance of this chapter for the remaining of this thesis is that it provides an effective duty cycle scheme that can prolong the lifetime of nodes, and hence other protocols can work for longer periods.

The underlying approach of the proposed duty cycle scheme is to enable nodes to construct the probability distributions from which they select their duty cycle parameters. The main aim of this selection is to achieve the following objectives: minimising the energy consumed, reducing the end-to-end delay and increasing throughput. The energy is saved by allowing a node to sleep to avoid the overhearing an idle listening; this saving prolongs the lifetime of nodes and allows nodes to route their packets over shortest path for longer period. Besides prolonging the lifetime of network, maintaining the shortest path for longer periods enhances route redundancy which reduces the end-to-end delay and increases the throughput by reducing the probability of collisions over longer paths.

The rest of this chapter is organised as follows. Section 7-2 provides an overview of artificial chemistry and its related research disciplines in the context of computer and communication engineering. Section 7-3 introduces the modelling and optimisation techniques that will be used to evaluate the energy consumption of the multihop networks and to design the duty cycle management scheme. In section 7-4 the analogies between artificial chemistry and multihop networks are introduced and the application of the molecular statistical mechanics to model the energy consumption in wireless network is given in section 7-5. Section 7-6 and 7-7 present the underlying approach of the proposed duty cycle management scheme and its pseudo-code respectively and in section 7-8 the results and discussion are presented. Finally section 7-9 concludes this chapter.

7-2 Overview of artificial chemistry

The ultimate goal of artificial chemistry is to build an artificial system that can benefit from the fundamental principles of chemical processes; due to the widespread nature of chemistry and advances in computing and communication technologies the artificial chemistry represents an active research field. Subsection 7-2-1 explores the main approaches under which chemistry is employed in the design computing systems and subsection 7-2-2 provides brief history of the artificial chemistry and its related research disciplines.

7-2-1 Introduction to chemistry

Chemistry is a branch of science that investigates the nature, properties and composition of substances and the changes that they undergo [215]; chemistry is described as the ‘central science’ or ‘servant science’ as it links the applied sciences with physical and life sciences. This integration between chemistry and other research fields yields a number of interdisciplinary research activities, and some examples [216] are agrochemistry, environmental-chemistry, electrochemistry, nanochemistry, photochemistry, nuclear-chemistry, astrochemistry, biochemisty and chemical-computing.

Chemistry can be studied from two perspectives: substance perspective and process perspective [215-216]. The substance perspective studies chemical properties of substances and the conditions under which these properties are changed; conversely, the process perspective is concerned with analysing chemical reactions. A chemical reaction can be seen as a transformation of some substances into other substance; the properties of yielded products is not necessary a resultant of the properties of the reactants, i.e., the properties of the yielded products can differ substantially from the properties of individual reactants or of all of which taken together. Hence, a chemical reaction can be characterised as a process that evolves its entities qualitatively and quantitatively. This exemplary behaviour is one of the most important reasons of employing the fundamental principles of chemical processes to model and design a wide variety of dynamical processes in vast area including ecology, proto-biology, computer science, linguistic, mechanical and social systems [215-222]. It is worth noting that this chapter considers the process perspective of chemistry, as it is more appropriate for designing the duty cycle scheme for multihop networks.

Employing chemistry in computer and communication engineering has been introduced under two main approaches: real chemical computing and artificial chemical computing. Real chemical computing [219] uses real chemical substances to perform computations. A real chemical computing system is a spatially extended chemical system which uses the concentration profile of chemical species to represent the input and output of the computations and uses the diffusion and phase waves to accomplish data and information communications. The advantage of this type of computing compared to conventional computers (silicon-based) is threefold: firstly, the massive parallelism as all information processing as well as the input and output of data are performed using the chemical reactions that are accomplished simultaneously. Secondly, fault tolerance and auto-configurations result from the ability to replace any faulty substrate almost immediately, and finally, the reusability and disposability since the chemical species that are used in these computers can be recycled. Based on these advantages real chemical computing has been used to solve various problems relating to image processing, routing planning, robot navigation and logic gates [219-222].

Artificial chemistry computing (AC) [210-211], on the other hand, is a field of research that studies the fundamental principles of chemical processes and then exploits these principles to engineer effective computational paradigms on conventional silicon-based processors. A typical chemical reaction is a highly sophisticated process that is affected by many factors including: the chemical properties of a species (e.g., atomic structure of each species); the interaction mechanism between the species (bond forming and breaking between different species); chemical conditions of the reaction (presence of a catalyst or solvent) and conditions of the physical environment of the reaction (temperature, pressure or intensity of electromagnetic waves). The ability for AC to account for these factors facilitates its ability to be used to model other dynamic systems. Thus AC has been employed extensity to solve a wide variety of problems in different disciplines including life science, mechanical systems and economics [218-219].

Interestingly, AC has been employed in many aspects of electrical and computer engineering. The work presented in [223] reveals the analogies between AC and circuit theory and demonstrates that artificial chemistry can be used to model a class of multi-port electrical networks. Other work presented in [224] investigated the relationship between chemical reaction networks and Boolean gates that are used in digital design. The work presented in [225] employs the principles of the artificial chemistry to develop a computer simulator called Tierra. In this simulator, the machine instructions compete to access the central processing unit and memory using the mutation, replication and recombination processes that are defined in chemical processes.

7-2-2 Brief history of AC and its related research disciplines

Chemical reactions are dynamic processes that have some interesting behaviour such as self-maintained self-maintenance, self-replicating, emergence and evolution. Employing these characteristics in the design of artificial systems is the prime motivation behind the appearance and rapid development of artificial chemistry. Widespread use of chemistry in different disciplines and advances in computing and communication techniques has led to formulation a mature multidisciplinary research field. This highly multidisciplinary field does not only enrich AC but also leads to develop novel research lines, e.g., autonomic computers and artificial-life. This section provides a brief overview with emphasis on the chemical processes perspective and computational applications.

One of the pioneering works devoted to the design of a self-replicating system was proposed by John von Neumann in the late 1940s [226]. The main aim of such a system was to build a system able to construct other systems or even itself from raw materials given the description of the required system. The system proposed by John von Neumann consisted of three main components: universal machine, constructor arm and memory tape. The information encoded on the memory tape is used as blueprint instructions of the required machine; the universal machine reads the information that is encoded on the memory tape and instructs the constructor arm to assemble the required machine part by part. Neumann demonstrated the ability of his system to replicate itself by feeding the description of the universal machine in the memory tape. The concept proposed by Neumann has been developed further to form the cellular automata system [226-227]; cellular automata system replaced the mechanical machinery of Neumann’s system by intelligent automata that are distributed on a lattice. Each automaton maintains a set of finite states and updates its states in coherent with its present state as well as the states of its neighbours according to specific algorithms. The relation between self-replicating principles and AC was discovered by Watson and Crick, who demonstrated that biotic organisms can be abstracted as a self-replicating system in which DeoxyriboNucleic Acid (DNA) acts similarly as the memory tape of Neumann’s machine [228-229]. Revealing the chemical processes of DNA and its mechanism has important contributions in deriving a number of computational algorithms, e.g., genetic algorithms [230], artificial neural networks [231].

Success in building self-replicating machines motivated scientists to investigate a more sophisticated behaviour; that is the mechanism by which an interacting system can evolve its pattern and structure to cope with unpredictable changes. A detailed description for this mechanism has been proposed by Turing in 1950s, his seminal work introduced the ‘chemical morphogenesis’ [232] as a means to understand the spatial pattern formation in interacting system. According to Turing, an interacting system can be represented as a set of chemical morphogens “chemical substance” that react and diffuse through the system. Turing considered two cases for interacting processes: firstly, the case when these morphogens just react without diffusion, then the system maintains its spatial uniform development. On the other hand, if morphogen react and diffuse then the spatial heterogeneous patterns are generated; these heterogeneous patterns are dominated by concentrations of the chemical morphogens. Turing represented his hypothesis using an ingenious mathematical theory. This theory was considered one of the first broad frameworks that quantified pattern formation using chemical processes [233]. Based on this theory, a number of computational algorithms have been introduced, some of which mimic the pattern formation of biological systems to form the short-range-activation long-range inhibition which used was been to solve a number of problems related to computers and communication networks [234-235]. Other work inspired by Turing’s mechanism and introduced the concept of autonomic computing [236].

Instigated by Turing’s theoretical work of and Neumann’s machine, the approach of self-creating system has been proposed by Maturana and Varela in 1972 dubbed autopoietic. The autopoietic system is defined as “the system that is capable to produce and maintain itself from within itself” [233,237]; this system mimicked the chemical reactions inside a cell by abstracting a cell as a machine consisting of a set of reactions and a collection of components, each with its own distinct properties that govern reactions with other components. A reaction is able to produce new components that have different reaction characteristics, which in turn leads to new reaction mechanisms. This work introduced the emergence principle of chemical processes, i.e., the ability of a chemical process to produce a sophisticated organism from collections of primitive constituents [238]. The emergence principle becomes one of the most important approaches that has been used to understand and design creative processes, e.g., flocks of birds, school of fish, giant termite, Wikipedia and software filtering advertisement system [238-239].

Another milestone of AC has been introduced in 1994 by Fontana [240]. This work investigated the evolutionary principle of chemical process (i.e., the ability of chemical reactions to modify the characteristic of its components internally over the time). This work paves the way toward developing of chemical organisation theory [241] that has been used to analyse and design dynamic interacting systems. The chemical organisation theory together with the remaining works contributed to formation a novel research discipline called Artificial Life (AL) [211] which aims to build artificial systems that exhibit behaviours of natural organisms. I.e., the artificial system that can act like living in systems in many aspects such as, survival, intelligence, self-reproduction, homeostasis, civilization, creativity and forecasting [211, 241].

7-3 Modelling and optimisation techniques of AC

This section provides detailed descriptions for the techniques that will be used to model the energy consumption and to design the proposed duty cycle management scheme for multihop networks. These techniques are explained in the chemical domain in order to get deep insight about them; then based on this section the analogies between the chemical reaction networks and wireless communication networks is presented in the next section. Subsection 7-3-1 presents the molecular statistical mechanics which will be employed to model the energy consumption of wireless networks while section 7-3-2 introduces the artificial decision making of AC that will be used to design the proposed duty cycle scheme.

7-3-1 Molecular statistical mechanics

The molecular statistical mechanics [212] represents a set of chemical species that accommodate a given volume as a chemical reaction network, in which each species is characterised by its properties (e.g., chemical or physical) and is quantified by its concentration. A reaction between a pair of species over a specific trajectory is parameterised by the so-called propensity function which is the time dependent function that quantifies a reaction between these species over a given trajectory. The main aim of the molecular model is to determine the timing evolution of the concentrations of chemical species in a given reaction network.

From a mathematical perspective, a chemical reaction network is modelled using two sets (Υ, Θ) where Υ is a set of all chemical species within a reaction network and Θ is a set of the reaction trajectories. The propensity function of a reaction between a pair of species (e.g.,$i, j \in \Upsilon$) that occurs over a trajectory $k \in \Theta$ is given by $p_{i,j}^{k}(\tau)$. Concentrations of all chemical species is denoted by $\boldsymbol{\varepsilon}(\tau)$ and defined as time dependent vector over Υ, where $\boldsymbol{\varepsilon}(\tau_0) = \boldsymbol{K}$ is the value of the concentration of all species at the beginning of the reaction, i.e., the initial conditions. As each reaction changes the concentrations of some or all chemical species of the network, $g\left(\boldsymbol{\varepsilon}(\tau); p_{i,j}^{k}(\tau)\right)$ is defined as a function that quantifies the changes of the concentration of chemical species that participate in a reaction whose propensity function is $p_{i,j}^{k}(\tau)$, i.e., cost function. Utilising these definitions facilitates expression for the time evolution of the concentrations of the chemical species in a network as:

$$\frac{d\boldsymbol{\varepsilon}}{d\tau} = \sum_{i,j \in \Upsilon} \sum_{k \in \Theta} \boldsymbol{g}\left(\boldsymbol{\varepsilon}(\tau); p_{i,j}^{k}(\tau)\right); \ \boldsymbol{\varepsilon}(\tau_0) = \boldsymbol{K} \tag{7.1}$$

It is worth noting that the system of delayed differential equations given in equation (7.1) represents one of the canonical forms of dynamic systems, in which the metric space is the concentrations of the chemical species and time evolution operators are the propensity functions. Hence, the existence and uniqueness of the solution of such a system has been proven in the literature, e.g., [244-245].

As an illustrative example let us consider the following chemical reaction network:

$$2\mathrm{NaN_3} \xrightarrow{k_1(\tau)} 2\mathrm{Na} + 3\mathrm{N_2} \tag{7.2-a}$$

$$10\mathrm{Na} + 2\mathrm{KNO_3} \xrightarrow{k_2(\tau)} \mathrm{K_2O} + 5\mathrm{Na_2O} + \mathrm{N_2} \tag{7.2-b}$$

This chemical reaction network is a simple representation of the reaction that takes place in the airbag to protect the car's driver from a head-on collision [246]. The main aim of this reaction network is to fill the airbag with Nitrogen (N_2) gas. However, as this gas cannot be stored in a car, so the Sodium azide (NaN_3) which has a solid state form (salt) is stored in the airbag and the reaction depicted in equation (7.2) is used to produce Nitrogen. Reaction (7.2-a) is triggered based on the signal received from the deceleration sensor of the car, this electrical pulse decomposes the Sodium azide (NaN_3) into Nitrogen (N_2) and Sodium metal (2Na). The amount of the Nitrogen produced from this first reaction is used to fill the airbag. However, as the second product (2Na) is highly reactive and a potentially explosive material, a number of reactions are carried out to convert it into harmless materials; one of those reactions is shown in equation (7.2-b). The reaction (7.2-b) converts Sodium metal into harmless chemical species (potassium oxide (K_2O) and sodium oxide ($5Na_2O$) in the presence of Potassium Nitrate (KNO_3).

The propensity function of the first reaction which is given as $k_1(\tau)$ is a function of the deceleration of the car. This function decomposes the concentration of the (NaN_3) and increases the concentration of (N_2) and (Na). Similarly the propensity function of the second reaction which is denoted by $k_2(\tau)$ reduces the concentration of (Na) and (KNO_3) and increases the concentration of (K_2O), (Na_2O) and (N_2).

7-3-2 Artificial decision making of AC

From the perspective of molecular statistical mechanics, a chemical reaction network is a distributed interacting system that is governed by a set of rules specified by propensity functions. This system can be expressed in the parlance of artificial decision making as a set of interacting particles whose uncertain spaces are propensity functions. This analogy between molecular statistical mechanics and artificial decision-making is the key principle behind utilising the artificial decision making techniques in optimising the reactions processes of AC [212,214].

Using the artificial decision making techniques in AC requires the definition of two functions. Firstly the objective function which describes the main target that is required to be achieved by the system, e.g., increase the concentration of the desired yield, improve its quality or even reduce the wasted species. Secondly, the renormalized-group [214,247] which are the environmental conditions that can be altered by a species (agent) in order to achieve the predefined objective function. Based on these two functions, a species infers a potential reaction and categorises it based on the objective function as either constructive or deconstructive reactions. Based on this classification a species alters its renormalized-group in accordance to maximise the occurrence of the constructive reactions and to minimise the occurrence of the deconstructive reactions considering any defined constraints.

The two main advantages of this approach are that it is lightweight and versatile. The lightweight nature is due to the fact that a species needs only to know about the reactions that affect its objective functions and to cluster them into two mutually exclusive groups. Versatility is demonstrated by the ability of a species to modify its situation with respect to the potential reaction without requirement to alter the reaction itself and without affecting the reaction sequence. The following hierarchical Bayesian model [248] is used to infer the probability distribution of the reaction timings:

$$\begin{gathered} R_i(\tau|\phi) \\ \phi|G \sim G_x(\tau) \\ G_x|G_0, \alpha \sim DP(\tau|G_0, \alpha) \end{gathered} \tag{7.3}$$

where $R_i(\tau|\phi)$ is the probability distribution that is used by species i to predict the time for the next reaction conditioning on parameter ϕ , DP denotes the Dirichlet Process which is characterised by two parameters: the dispersion α and the base distribution G_0, $G_x(\tau)$ is the higher order process that is used to generate the parameter ϕ.

A justification for selecting the Dirichlet process [214,248-249] to represent the timings of reactions that are observed by a species is that in real reaction networks, there are infinite sources of reactions, hence the aggregated time of reactions, as seen by a species is heterogeneous which in turn requires an effective process to describe it. Since the Dirichlet process has complete random measures and its support is a distribution, (i.e., any sample from a Dirichlet process is distribution by itself) makes it the best candidate to represent the reactions. Moreover, employing the Dirichlet process as a prior over a given distribution provides a posterior distribution that follows the Dirichlet process also.

7-4 Analogies between chemical reaction networks and multihop wireless communications networks

This section describes the analogy between chemical reaction networks and multihop wireless communication networks. Let us reconsider the reaction network shown in equation (7.2) under the following assumptions:

1. The chemical species represent wireless nodes.
2. The reactions between the chemical species represent the communications between the wireless nodes.
3. The concentrations of the chemical species represent the energy contained in the wireless nodes.
4. The propensity functions of the reactions represent the inter-arrival distributions of packets.
5. The trajectories of the reaction represent the communication routes.

Applying these assumptions to equation (7.2) facilitates seeing the chemical reaction networks given in equation (7.2) as a communication network with six nodes A, B, C, D, E and F, each of which represents a chemical species, (i.e., node A represents NaN_3, node B represents Na, node C represents N_2, node D represents KNO_3, node E represents K_2O and node F represents NaO).

An illustration for these analogies is shown in figure 7.1 where each node is represented with circle and arrows are used to depict the transmission-reception directions, where an arrow is based at a transmitter node and headed towards a receiver node.

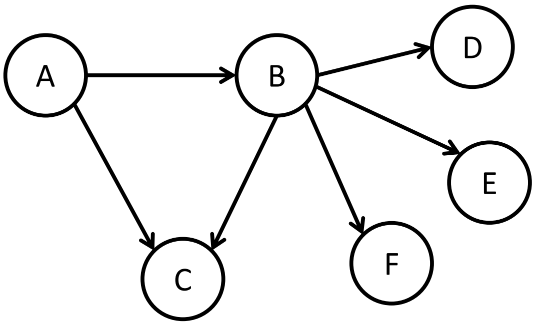


Figure 7.1 Wireless network representation for chemical reaction of equation (7.2)

With accordance with equation (7.2-a), node A generates its packets with function $k_1(t)$. As node A transmits these packets it consumes transmission energy and both nodes B and C (neighbours of node A) consume receiving energy. In the reaction (7.2-b) the traffic that is received at node B from node A is distributed to other nodes. Similar to the aforementioned description, when node B commences transmission it consumes transmission energy while its neighbours (nodes C, D, E and F) consume receiving energy. This example shows that node A does not receive the traffic sent by node B which reflects the situation of an asymmetric link between node A and B where node B can receive data from node A but not vice-versa, hence node A does not consume receiving energy when node B is transmitting.

It is worth noting that a wireless communication network is a simple form of chemical reaction networks. Wireless communication networks are mono-stable, i.e., they have a single equilibrium state as the energy of all nodes approaches zero as time approaches infinity. This is due to the fact that all communications consume the energy of nodes while in the chemical reaction networks a reaction can increase or decrease the concentration of species. More importantly, while the renormalized-group in AC contains a number of freedom parameters that a chemical species can control, the renormalized-group in wireless communication networks is composed of a single binary variable (awake or asleep).

7-5 Molecular statistical mechanics for modelling energy in multihop networks

This section presents the first main contribution of this chapter that is utilising the molecular statistical mechanics of AC introduced in section 7-3-1 to derive the time evolution of energy consumption in multihop network operating CSMA-CA. Towards this goal, subsection 7-5-1 presents symbols and definition of the proposed energy model, subsection 7-5-2 derives the energy consumed during service of packets and finally section 7-5-3 uses these findings to develop the energy model.

7-5-1 Symbols and definitions of the proposed energy model

This energy model follows the methodology, assumptions and notations of the analytical framework that was proposed in chapter 5. This integration allows computing the time evolution of consumed energy with a high level of accuracy and independently of the topology configurations, traffic patterns or routing policies. The main assumptions that made in chapter 5 are recalled here. There are a number of wireless nodes that have identical equipment (radio transceivers, battery capacity, etc.) and they are distributed randomly over a given area to form a multihop network. All the transmissions and receptions of all nodes are performed with the same and fixed data rate and using the same power level. It is assumed that there are four states for a transceiver: transmit, receive, idle and shutdown. The transmit state represents the case when a node is transmitting its packets; similarly the receive state refers the case when a node is receiving signal. In the idle state, some part of the transceiver is functioning, however, it is not in transmit or receive states. Therefore, a node can switchover to receive state as soon as it hears a signal whose power is greater than or equal to its receiving sensitivity threshold. Finally, in the shutdown state, the transceiver is turned off to save the energy. Additional to these states a node consumes some energy to switch the transceiver from a state to others. Table 7.1 summarised the key symbols used to derive the energy model.

Table 7.1 Key symbols of the energy model

Symbol	Definition
R	The length of slot in second, this value is constant and defined in the specification of a CSMA-CA protocol, e.g., in the IEEE 802.15.4 based networks that are operated at 250kbps a slot is $320\mu s$. This value is used here to convert the duration of slot (that is used in deriving the service time) into seconds which is required to maintain the amount of energy in SI unit, Joule.
$\mathcal{E}^{init}$	Initial energy of nodes given in joules.
$Pow(idle)$	The amount of power that is consumed by a node in idle state given in watt.
$Pow(rx)$	The difference between the receiving power and the idle power given in watt.
$Pow(tx)$	The difference between the transmission power and the idle power given in watt.
$\mathcal{E}_B^A$	The amount of energy that is consumed by a node to switchover between state A to state B, where A and B can be any one of these states: idle, receive or transmit. The value of $\mathcal{E}_B^A$ is given in joule.
$E[TX_n]$	The average energy that is consumed by a node in servicing a packet, the value of this energy is derived in equation (7.4).
RX	The amount of energy that is consumed by a node during receiving a packet given in joule, i.e., $RX = \mathrm{R} \times L \times \left(Pow(rx) + Pow(idle)\right)$ where L is the length of packets in slot.

7-5-2 Energy consumption during service time of a packet

This subsection derives the amount of energy that a node spends during servicing a packet using a CSMA-CA protocol. During the CSMA-CA process a node utilises different power levels: idle level in back-off periods; reception power when it is performing channel assessment (CCA); and transmission power when it commences transmission. Since a node spends most of its time in the back-off states compared to the other states (as the back-off period is the longest period) a simple computation can be performed if the power level during the idle state is considered as the reference and the transmission and reception power levels are added to it. The proposed model refers to the power level that a node spends in the idle state by $Pow(idle)$, $Pow(rx)$ is the difference between the receiving power and the idle power and $Pow(tx)$ is the difference between the transmission power and the idle power. The energy consumed to switchover from mode A to mode B is denoted by $\mathcal{E}_B^A$ (modes A and B can be: idle, receiving or transmission). Figure 7.2 shows the schematic diagram for the proposed energy model and its notations. In this figure, it is assumed that a packet backs-off twice to show all possible switchover energy. It is also shown that the power consumed in the receiving state is higher than in the transmission state which is common in WSNs, due to the short transmission range in these networks [250-251].

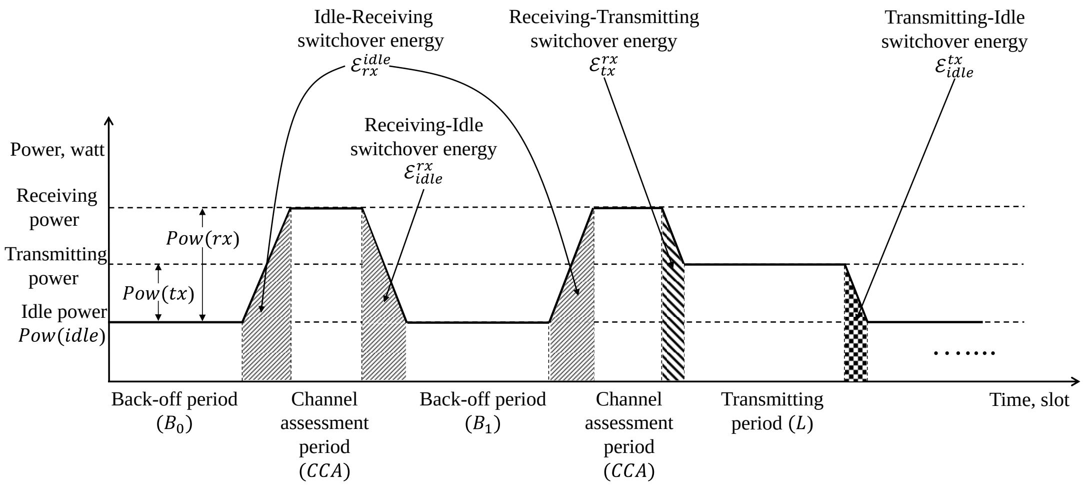


Figure 7.2 Schematic diagram for energy model and its notation

Based on the aforementioned definitions, the average energy consumed during the service time of a packet by an arbitrary node n is denoted by $E[TX_n]$ and given as:

$$\begin{aligned} E[TX_n] = \mathrm{R}\big(&E[S_n] \times Pow(idle) + E[Q_n] \times CCA \times Pow(rx) \\ &+ (1-(1-\alpha_n)^{M+1}) \times L \times Pow(tx)\big) \\ &+ E[Q_n] \times \mathcal{E}_{rx}^{idle} + (1-(1-\alpha_n)^{M+1}) \times \mathcal{E}_{tx}^{rx} \\ &+ (1-\alpha_n)^{M+1} \times \mathcal{E}_{idle}^{rx} \end{aligned} \quad \text{Joules} \quad (7.4)$$

A node consumes at least the idle power $Pow(idle)$ during the whole average service time $E[S_n]$. Throughout this time, a node performs on average $E[Q_n]$ channel assessments, each one takes CCA slots with receiving power $Pow(rx)$. After that period, a node consumes $Pow(tx)$ power during the transmission of a packet which varies according to packet length L and occurs only when a device accesses the channel, which it does with probability$(1 - (1-\alpha_n)^{M+1})$. Since all of the timings are specified in slots, they are multiplied by R, (i.e., length of slot in second) to express the results in joules.

Additionally, a node expends some energy to switchover between different modes: $\mathcal{E}_{rx}^{idle}$ joules to switch from the idle mode (which is used when a node is backing-off) to the receiving mode (which is used when a node is assessing the channel) each time a node assesses the channel, which occurs on average $E[Q_n]$ times during servicing a packet. A node after performing a channel assessment switches either to the transmission mode or to the idle mode. Switching to the transmission mode consumes $\mathcal{E}_{tx}^{rx}$ joules and happens when a node accesses the channel which occurs with probability $(1-(1-\alpha_n)^{M+1})$ whilst a node expends $\mathcal{E}_{idle}^{rx}$ joules to switch to idle mode with complement of the aforementioned probability, i.e., $(1-\alpha_n)^{M+1}$.

The values of the energy and power are obtained from the transceiver datasheet [137]. While the value of $E[S_n]$ and α_n are given in equations (5.36) and (5.27) respectively and the value of $E[Q_n]$ can be computed by multiplying the number of channel assessments by their corresponding probabilities as given in table 5.2.

7-5-3 Energy model

This section derives the propensity function of traffic flow and then uses it to compute the energy model. The propensity function is denoted here by $F(P_x(u,v);n)$ and defined as a time dependent function that characterises the traffic that is generated at node u, destined to node v as a final destination and sent over the path $P_x(u,v)$ as seen by intermediate node relaying node n. The value of $F(P_x(u,v);n)$ is given as:

$$F(P_x(u,v);n) = \left(\lambda_{uv}^{i}\omega_{P_x(u,v)}^{i}\xi(P_x(u,v);n)\right)\left(\tau - \mathrm{R}\sum_{\substack{\forall y\in P_x(u,v)\\ y\notin\{n_{+1},\dots,v\}}} E\left[S_y\right]\right) \qquad (7.5)$$

Where $\lambda_{uv}^{i}\omega_{P_x(u,v)}^{i}$ is the rate of traffic that is generated by node u, destined to node v as a final destination and sent over the path $P_x(u,v)$ and $\xi(P_x(u,v);n)$ is the probability that a packet sent over the path $P_x(u,v)$ is received successfully by node n where n is an intermediate relaying node in the path. If $n=v$ then the $\xi(P_x(u,v),v)$ represents the probability that a packet sent over the path $P_x(u,v)$ is received successfully to its final destination (node v in this case).The values of these two parameters are derived section 5-5-7, moreover n_{-1} and n_{+1} are defined as the previous and next hop of node n in the path $P_x(u,v)$ respectively. Hence the first term of this equation (i.e., $\lambda_{uv}^{i}\omega_{P_x(u,v)}^{i}\xi(P_x(u,v);n)$) represents the rate with which node n sees the traffic that sends from u to v over path $P_x(u,v)$ .The second term of equation (7.3) represents the average delay that a packet encounters since it is transmitted by node u until received by n. Based on the aforementioned definitions, the energy consumed during the lifetime of node n is given as:

$$
\begin{aligned}
&\frac{d\mathcal{E}_n(\tau)}{d\tau} \\
&= \lambda_n^i \big(\mathcal{E}_n(\tau)\big)^{E[TX_n]} \prod_{\forall y \in \Omega(n)} \Big(\mathcal{E}_y(\tau)\Big)^{RX} \\
&+ \big(\mathcal{E}_n(\tau)\big)^{RX} \left(\sum_{\substack{\forall u,v \in \Upsilon \\ \forall P_x(u,v) \in P(u,v)}} \sum_{\substack{n \in P_x(u,v) \\ n \notin \{u,v\}}} \left(\begin{matrix} F(P_x(u,v); n_{-1}) \\ \Big(\mathcal{E}_{P_x(u,v),n_{-1}}(\tau)\Big)^{E\left[TX_{P_x(u,v),n_{-1}}\right]} \\ \prod_{\forall y \in \Omega(P_x(u,v),n_{-1}) \setminus \{n\}} \Big(\mathcal{E}_y(\tau)\Big)^{RX} \end{matrix} \right) \right) \\
&+ \big(\mathcal{E}_n(\tau)\big)^{E[TX_n]} \left(\sum_{\substack{\forall u,v \in \Upsilon \\ \forall P_x(u,v) \in P(u,v)}} \sum_{\substack{n \in P_x(u,v) \\ n \notin \{u,v\}}} \left(\begin{matrix} F(P_x(u,v); n) \\ \prod_{\forall y \in \Omega(n)} \Big(\mathcal{E}_y(\tau)\Big)^{RX} \end{matrix} \right) \right) \\
&+ \big(\mathcal{E}_n(\tau)\big)^{RX} \left(\sum_{\substack{\forall u \in \Upsilon \\ \forall P_x(u,n) \in P(u,n)}} \left(\begin{matrix} F(P_x(u,n); n_{-1}) \\ \Big(\mathcal{E}_{P_x(u,n),n_{-1}}(\tau)\Big)^{E\left[TX_{P_x(u,v),n_{-1}}\right]} \\ \prod_{\forall y \in \Omega(P_x(u,n),n_{-1}) \setminus \{n\}} \Big(\mathcal{E}_y(\tau)\Big)^{RX} \end{matrix} \right) \right) \\
&+ \big(\mathcal{E}_n(\tau)\big)^{RX} \left(\sum_{\substack{\forall u,v \in \Upsilon \\ \forall P_x(u,v) \in P(u,v)}} \sum_{\substack{\forall y \in P_x(u,v) \\ n \notin P_x(u,v) \\ y \in \Lambda(n)}} \left(\begin{matrix} F(P_x(u,v); y) \\ \Big(\mathcal{E}_y(\tau)\Big)^{E[TX_y]} \\ \prod_{\forall z \in \Omega(y) \setminus \{n\}} \Big(\mathcal{E}_z(\tau)\Big)^{RX} \end{matrix} \right) \right)
\end{aligned} \tag{7.6}
$$

The first term of equation (7.6) expresses the amount of energy consumed in transmission the traffic that is generated internally by node n whose rate is given as λ_n^i. Hence, node n consumes transmission energy while all of its neighbours consume the receiving energy. The second and third terms account for the energy consumed in routing the external traffic of node n, the first two summations of these terms represent all paths that are passed through node n. The second term uses the propensity function of the traffic as seen by the previous hop of node n in the path $P_x(u,v)$, hence this node consumes transmission energy while its neighbours as well as node n expend receiving energy. Following the same approach, the third term uses the propensity function as seen by node n and indicates that node n spends transmission energy while all of its neighbours consume receiving energy. The fourth term of this equation quantifies the energy consumption in receiving the traffic that is destined to node n as a final destination and finally the fifth term accounts for the energy consumed by node n and its neighbours when it overhears traffic that is not destined to it. It is worth noting that this equation considers a general case, hence accounting for the energy consumed while a node is idle can be added to the equation based on its duty cycle.

7-6 Underlying approach of the proposed duty cycle scheme

This section utilises the decision making principle of AC presented in section 7-3-2 to design an effective duty cycle management scheme for multihop WSNs that can prolong the lifetime of nodes, decrease the end-to-end delay of packets and increase the throughput of network. A schematic diagram for the duty cycle of an arbitrary node u is illustrated in figure 7.3. It shows that the time of this node is divided into random intervals, $T_1, \dots, T_n, \dots$ .An arbitrary interval T_n is subdivided into two random durations, W_n when a node is awake and V_n where a node is asleep.

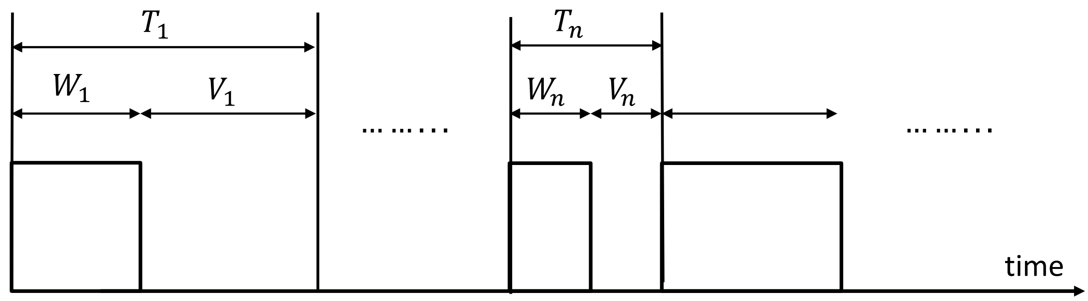


Figure 7.3 Schematic diagram for the duty cycles

Node u needs to wake up when it has a packet to send, when it receives a packet either to route it to the next hop or to absorb it (as node u is the final destination of the received packet). Conversely a node needs to sleep when one of its neighbours commences a transmission that is not destined to it (to avoid energy wasted during the overhearing). As these times are random variables then the effective method to find them is by specifying their probability distributions, which are dubbed constructive and deconstructive distributions respectively for the wake up and sleep durations.. This section discusses the high level abstraction of the proposed duty cycle management scheme and its mathematical equations. It is worth noting that this discussion considers the constructive distribution since the deconstructive distribution follows the same approach. Figure 7.4 illustrates the high level abstract for the key assumptions of the proposed scheme.

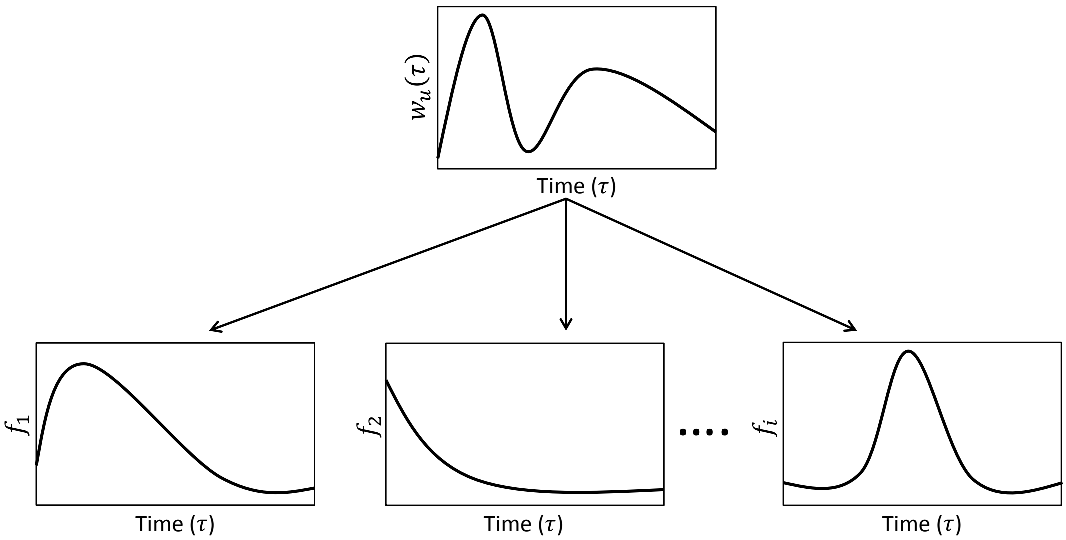


Figure 7.4 High level abstract for the key assumptions of the proposed scheme

This figure shows that the constructive distribution of node u, dented by $w_u(t)$, comprises a number of sub-distributions, (i.e., f's) each of which represents the constructive traffic destined to node u from another node. Considering the appealing statistical properties of the Gamma distribution that was discussed in section 6-4 (e.g., the Gamma distribution has the infinitely divisible property, is closed under convolution and is the conjugate prior for most of the probability distributions that are used to model traffic in communication network) enables us to safely assume that all these distributions following a Gamma distribution. Therefore, the $f_i(r_i, s_i)$; $\forall\ i \in \Psi/\{n\}$ refers to the constructive traffic received by node u from node i, and r_i, s_i are the rate and shape parameters of this distribution respectively. Although it is easy to determine the rate of the constructive distributions (i.e., $r_i; \forall\ i \in \Psi/\{u\}$) as long-run frequency of the number of constructive packets, the main challenge is how to derive the shape parameters of each constructive distributions, i.e., the values of s_i; $\forall\ i \in \Psi/\{u\}$.

Here the key assumption is that these shape parameters are random variables following stochastic process; this assumption is made in order to account for the fact that traffic variations in a typical multi-hop WSNs follows a stochastic process that is affected by many sources such as collisions, rerouting of packets due to battery depletion and ad-hoc topology. Based on this assumption, the shape parameter can be represented as $s \sim (G(\tau)|\ \tau = x)$ where $(G(\tau)|\ \tau = x)$ is a sample of Dirichlet Process (DP) at time instance $\tau = x$ (also called Dirichlet distribution). This sample is defined as a weighted collection of point masses (denoted by z's) where the location and weight of an arbitrary point z are denoted by ℓ_z and r_z respectively, i.e., $G(\tau)$ can be given as $G(\tau) = \sum_{z=1}^{K} r_z \delta(\ell_z)$ [214]. Here $\delta(\ell_z)$ refers to the Dirac delta function which is defined as $\delta(\ell_z) = 1$ at location ℓ_z and zero otherwise and K is the total number of mass points while the condition $\sum_{z=1}^{K} r_z = 1$ is defined in order to maintain $G(\tau)$ as a proper probability distribution. It is of interest to note that the definition of $G(\tau)$, (i.e., the sample of DP) can be used to express any probability distribution which justifies the argument that Dirichlet process can be used to model any random process which is presented in section 7-3-2 under equation (7.3). The next step in determining the shape parameters of the constructive distributions is to define the mechanism by which the values for the weights and locations are assigned to the new received packet based on empirical data; the flowchart of this mechanism is depicted in figure 7.5.

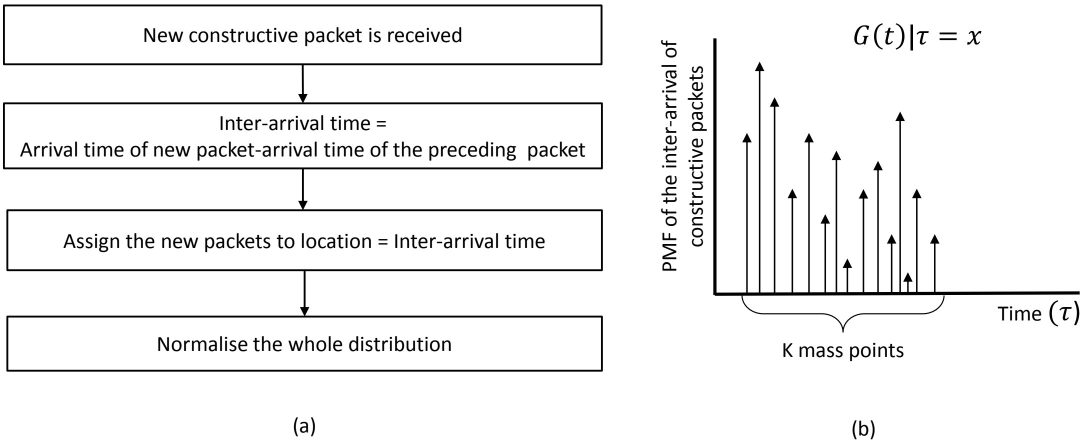


Figure 7.5 Flowchart of constructing the shape parameter of constructive distribution

Figure 7.5(a) shows that node u whenever receives a new constructive packet, it subtracts its arrival time from the arrival time of the preceding received packet in order to determine the inter-arrival times between constructive packets. Based on the inter-arrival times, node u increases the weight of mass point corresponding to the inter-arrival time and normalise the whole distribution in order to maintain a proper distribution. Figure 5.7(b) shows a sample for the Probability Mass Function (PMF) of inter-arrival distribution of constructive packets at an arbitrary time instance $\tau = x$. The final step in the proposed duty cycle management scheme is to enable node u to infer the location and weight parameter of the inter-arrival time of the next constructive packet. This operation is performed by passing the empirical distribution derived in the previous section, i.e., $G(\tau)$ to Dirichlet Process (DP). DP uses the dispersion parameter, denoted by α, to perform this inference; the dispersion parameter is used to measure the variation of the estimated data with respect to the empirical data. Based on the values α and points masses of $G(\tau), 1 \leq k \leq K$, DP computes the location and weight of the next constructive packet, i.e., the location and weight of z_{K+1} according to the following equation:

$$DP(G(\tau), \alpha) = \begin{cases} r_k(\alpha + 1)^{-1}; \text{the new packet is located at location } k \\ \alpha(\alpha + 1)^{-1}; \text{create new mass point at } K + 1 \end{cases} \tag{7.7}$$

In the first case of equation (7.7) node u assigns the new packet to an old location k while in the second case node u assigns the new packet to new location. Thereafter, node u normalises the distribution of $DP(G(\tau), \alpha)$, finds its mean and assigns it the shape parameter of the distribution $w_u(t)$. The following section shows the pseudo-code that is used to a node to initialise and maintain these parameters.

7-7 Pseudo-code of the proposed duty cycle management scheme

The pseudo-code for the proposed scheme is given in figure 7.6 and explained in the following subsections.

Initial ( )
 $\boldsymbol{H} \leftarrow \emptyset, \alpha \leftarrow 0$
end Initial ( )

Receive packet ( )
 If (final destination address OR next hop) != my address
  Insert new record $\boldsymbol{H}(Arrival\ Time, 0)$
 else
  Insert new record $\boldsymbol{H}(Arrival\ Time, 1)$
end Receiving a new packet ( )

Generate sleep intervals ( )
 $X_v = E\left[DP\left(\boldsymbol{H}(Arrival\ Time, 0)\right)\right]$
 $Y_v = count\ (\boldsymbol{H}(Arrival\ Time, 0))/timecouter$
 $V = Gamma(Y_v, X_v)$
end Generate sleep intervals ( )

Generate wake intervals ( )
 $X_w = E\left[DP\left(\boldsymbol{H}(Arrival\ Time, 1)\right)\right]$
 $Y_w = count(\boldsymbol{H}(Arrival\ Time, 1))/timecouner$
 $W = Gamma(Y_w, X_w) + (Number\ of\ queued\ packets * \varpi)$
 $\alpha \leftarrow Cov(\boldsymbol{H}(Arrival\ Time) -$ expected receiving times)
end Generate wake intervals ( )

Figure 7.6 Pseudo-code of the proposed duty cycle management scheme

7-7-1 Initial function

This function is executed once during the lifetime of a node with the aim to set the following parameters to their initial values: bivariate vector $\boldsymbol{H}$ to $\emptyset$, the dispersion parameter α of the Dirichlet process to 0. The $\boldsymbol{H}$ is a bivariate vector, the first variable of $\boldsymbol{H}$ stores the arrival time of the received packets and the second variable of $\boldsymbol{H}$ is a flag that indicates the category of this packet. The value of this flag is set to one if this packet is a member of the constructive category (i.e., if this packet is destined to the node either as final destination or next hop) or to zero if packet is member of the deconstructive group (i.e., if a node overhears this packet). The parameters of the Dirichlet process are set to the shown values based on the assumption that in the initial phase, a node is in the minimal information state hence it is uses a symmetrical Dirichlet distribution with zero dispersion.

7-7-2 Receive packet function

This function is executed each time a node receives a packet with the aim to classify that packet into either constructive or deconstructive clusters. The function exploits the final destination and next-hop fields to determine the required information. The key concept of the function is that if the value of the final destination or the next-hop fields match the address of the receiver node then this packet is designed for the receiver. Hence a node creates a new record in the $\boldsymbol{H}$ vector, and the value of the first variable of the new created record of $\boldsymbol{H}$ is filled with the arrival time of the received packet. The second variable of the new created field is set to 1 to reflect that the received packet is amongst the constructive traffic. On the other hand, if the values of the final destination and the next-hop fields of the received packet are not equal to the address of the receiver node then a new record in $\boldsymbol{H}$ is also created but in this case its first variable stores the arrival time of the received packet while the second variable of the new created record stores 0.

7-7-3 Generate sleep intervals function

This function is executed whenever a node sits idle (i.e., a node is not in a reception or transmission state) for a period longer than the expected duration until the next reception or transmission states. The function extracts the arrival times that are stored in set $\boldsymbol{H}$ and that are categorised as deconstructive cluster (those records of $\boldsymbol{H}$ that have 0 in their second variable). The arrival times are then passed as arguments to a Dirichlet process and the average of this process is stored in a local variable X_v. The local variable Y_v is used to store the rate of the packets that are received and classified as a deconstructive cluster. This rate is computed by dividing the number of packets of this type by the time since a node starts until it executes this function. The function then uses X_v and Y_v as the shape and rate parameters respectively to generate the sleep interval according to the Gamma distribution.

A justification for using the arrival times of packet of deconstructive packets to derive the sleeping durations is that these times represent the intervals within which a node can sleep without missing a packet destined to it. However, as the accuracy of this determination depends mainly on the amount of data (packets that are received), the proposed function uses this data as empirical data. It then applies the Dirichlet process to compute the complete probability distribution. This step improves the accuracy of inference process over a small sample space [248].

7-7-4 Generate wake intervals function

This function is executed immediately after execution of the Generate sleep intervals function. The main reason for this is to keep the decision of a node about the sleeping and waking durations coherent and avoid any ill-defined update. This function follows the same procedure and arguments given in the Generate sleep intervals function except that it uses those records of $\boldsymbol{H}$ that have 1 in their second variable which account for the constructive cluster. Another major difference between the Generate wake intervals and Generate sleep intervals functions is that the length of packets that are queued in the buffer of a node are multiplied by the average time required to service a packet (ϖ) and then added to the length of waking interval that is generated from the Gamma distribution. This addition quantifies the time required to service the current queued packet and prevents a node from sleeping while it has a packet in its queue. It is shown in the results of this chapter that this summation has the advantage of reducing the end-to-end delay of packets. The last step of this function updates the value of the dispersion parameter of the Dirichlet process according to the coefficient of variation of the differences between the actual arriving times of packets and the expected receiving times. This is proposed to accelerate the convergence between the inferring process (Dirichlet) and the empirical traffic.

It is worth noting that the proposed protocol is lightweight and can be implemented over tiny resources of sensor nodes. A justification for this argument can be acquired by discussing the requirements of the proposed scheme from three main aspects: storage space, energy requirement and computational complexity. In terms of the storage space, the proposed scheme uses a single data structure of a bivariate vector, the first variable stores the value of the arrival time of the received packets and the second variable is a flag storing whether the received packet is intended the node or not. In terms of the energy consumption the proposed scheme does not impose a communication overhead, hence it saves the energy required to exchange control packets and avoids collisions that may happen between these packets and data packets or between control packets themselves. In terms of computational complexity, the proposed protocol does not require a node to perform a complicated computation, as a node uses the values stored in its bivariate vector to construct a Dirichlet process whose parameters are used to generate the waking and sleeping durations based on Gamma distribution. Considering that the Dirichlet distribution can be generated from Gamma distribution [252] and there are many simple algorithms that can be used to generate Gamma distribution, e.g., [253-254] demonstrates the lightweight nature of the proposed scheme.

7-8 Results and Discussion

This section is devoted to assess the integrity of the modelling technique presented in section 7-5 and to demonstrate the benefit of the proposed duty cycle management schemes proposed in section 7-7. These evaluations are carried out upon two different CSMA-CA protocols, the IEEE 802.15.4 protocol that employs the Binary Exponential Back-off (BEB) and Gamma based back-off scheme proposed in chapter 6. Analytical results as well as simulation outcomes are labelled using the same policy that was introduced in section 4-4, i.e., using a five parameter tuple $\langle k, N, X, Y, Z\rangle$. The k parameter refers to the connectivity degree (the average number of mutually independent paths between non-neighbouring nodes) and the parameter N denotes the average number of nodes per channel. The parameter X specifies the probability distribution used to generate the inter-arrival times between packets at each node. This chapter uses three inter-arrival distributions: CBR, EXP and WBL to refer to the cases where inter-arrival times are generated according to Constant Bit Rate, exponential and Weibull distributions respectively. The rate with which a node generates its internal packet is denoted by Y and given in packet per slot. Finally, the parameter Z defines the routing policy which can take two possible values: U when all traffic that is generated internally by a node is distributed evenly to other nodes with the network and A to indicate that the traffic generated by all nodes within the network is destined to a single final destination. Other simulation parameters are set similarly to section 4-4. The results of this section are organised into three subsections: subsections 7-8-1 and 7-8-2 evaluate the energy consumption of IEEE 802.15.4 and Gamma based CSMA-CA protocols respectively, and subsection 7-7-3 assesses the benefit of the proposed duty cycle management scheme.

7-8-1 Energy consumption in IEEE 802.15.4 protocol

This subsection assesses the integrity of the molecular statistical mechanics proposed in section 7-5 to predict the energy consumed in multihop networks operating IEEE 802.15.4 protocol. Figures 7.7 and 7.8 compare the simulated and analytical results of the average energy consumed in the networks versus the number of hops for different scenarios under unity duty cycle using equation (7.6). The results of these figures show that there is a close match between the simulated outcomes and the results obtained from the proposed energy model. This match reflects the accuracy of the assumptions that are made in deriving the proposed model.

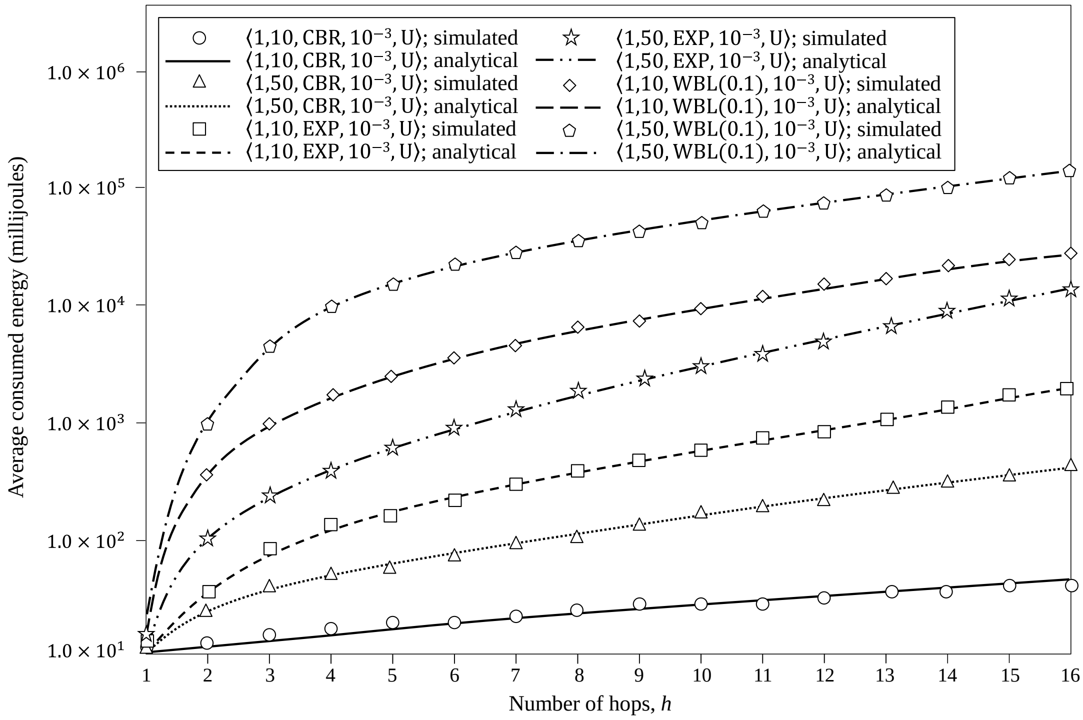


Figure 7.7 Average energy consumption of IEEE 802.15.4

The effect of increasing the average number of nodes per channel on the average energy consumed per hop can be obtained by comparing the first two cases of figure 7.7. In both of these cases, there is on average a single path between non-neighbouring nodes, the inter-arrival times between packets are generated according to the same distribution (CBR) with an identical rate of 10^{-3} packet/slots and each node acts as a source, router and final destination simultaneously. However, in the first case the average number of nodes per channel is small, N=10, (i.e., the scenario $\langle 1{,}10, CBR, 10^{-3}, U\rangle$) while in the second case N is increased to 50 (i.e., the scenario $\langle 1{,}50, CBR, 10^{-3}, U\rangle$). It can be seen from the results of this figure that the average energy consumed per hop is at least doubled when the number of hops is seven and when number of hops increases to sixteen the average energy consumed per hop increases by more than ten times. This characteristic is attributed to two main reasons: firstly, increasing the amount of energy consumed in overhearing with increasing average number of nodes per channel. The second reason is that increasing the average number of nodes per channel increases the contention to access the channel, which in turn increases the service time and its energy expenditure. The effects of inter-arrival probability distributions of packets on the average energy consumed per hop are also depicted in figure 7.7. It can be seen that under the same average number of nodes per channel, traffic intensity, connectivity degree and routing policy employing the CBR distribution to generate the inter-arrival times between packets saves much energy compared to the remaining cases (i.e., EXP and WBL). An explanation for this characteristic can be drawn by considering that CBR has the lowest coefficient of variation (the coefficient of variation of CBR is zero as the variance of this distributions is zero) which in turn regulates the traffic over multihop networks. Based on this explanation, it can be seen that when the exponential distribution is used to generate the inter-arrival times between packets, the readings of the average energy consumed per hops are lower than their peers when the WBL distribution is used. These results supports the use of the coefficient of variation in the Dirichlet process as proposed in section 7-6.

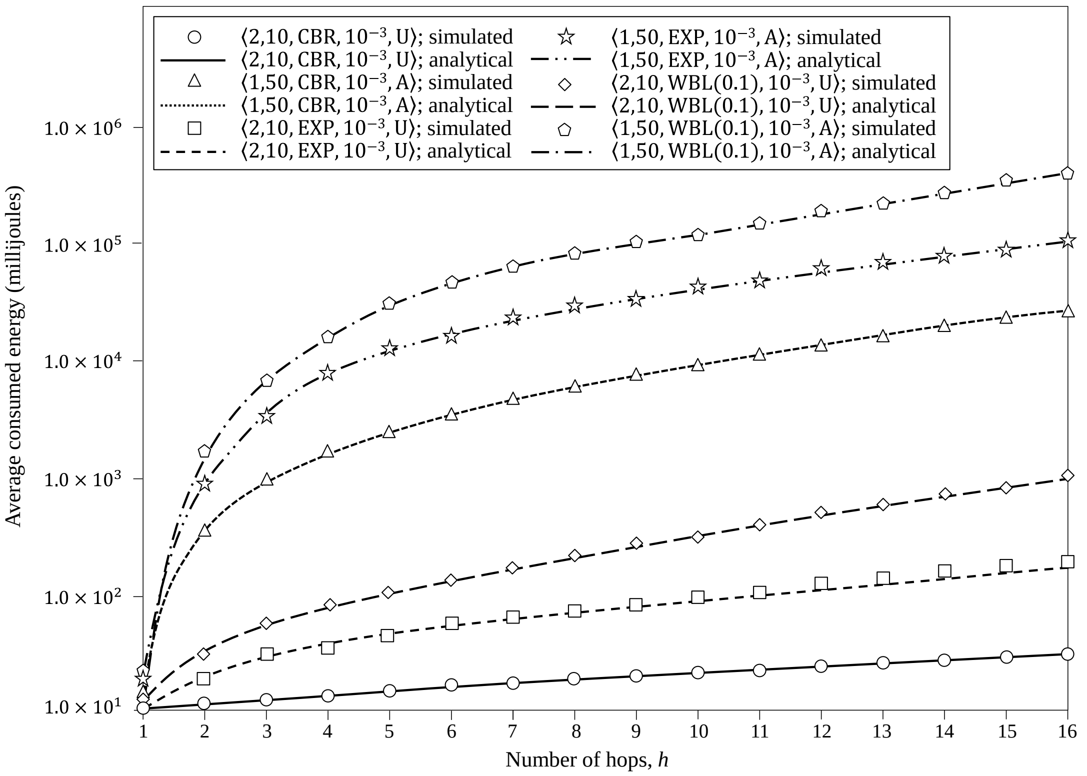


Figure 7.8 Effects of connectivity degree and routing policy on energy consumption of IEEE 802.15.4

Figure 7.8 evaluates the effects of the connectivity degree and the traffic policy on the average energy consumed per hop for different network configurations. The first case of figure 7.8 represents the scenario that is similar to the first scenario of figure 7.7 except that the connectivity degree is doubled. It can be seen that doubling the connectivity degree reduces the consumed energy which demonstrates advantages of using route redundancy on saving the wasted energy, i.e., sending the traffic over more than a path reduces the collision probability and mitigates overhearing by reducing the traffic intensity per channel. Compared the second case of figure 7.8 with its peer in figure 7.7 shows that sending all packets to a single final destination (as is the case in figure 7.8) increases the consumed energy significantly compared to the case when traffic is distributed evenly to all nodes (as is the second case in figure 7.7). This behaviour is attributed to the increase in the probability of collision near the sole receiver which in turn increases energy consumption. Although these two cases consider the CBR, comparing the remaining cases of figure 7.8 with its peers of figure 7.7 shows that both EXP and WBL have the same trend but with at different scales.

7-8-2 Energy consumption in Gamma based CSMA-CA protocol

Chapter 6 provided the first main contribution of the thesis, which is specifying an effective MAC protocol for multihop network. The CSMA-CA mechanism that was proposed in chapter 6 replaced the uniform distribution of the IEEE 802.15.4 CSMA-CA protocol with the Gamma distribution and replaced the blind doubling of back-off windows by a more intelligent collision resolution mechanism. This section is devoted to assess the impact of these modifications on the consumed energy. The same scenarios that have been investigated in the previous section are considered here and the results are normalised with respect to the results in figures 7.7 and 7.8 based on equation (7.5).

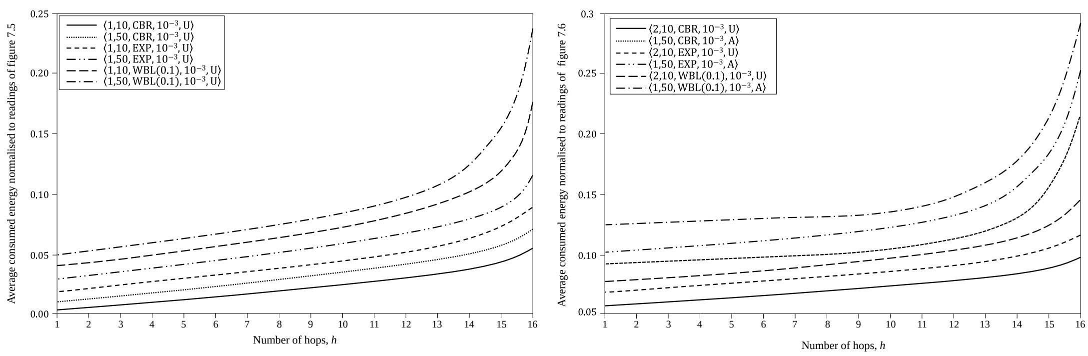


Figure 7.9 Average energy consumption of the Gamma CSMA-CA normalised to IEEE 802.15.4

It can be seen from results in figure 7.8 that the Gamma based CSMA-CA protocol can always save energy compared to the IEEE 802.15.4 protocol. This saving is attributed to the ability of the Gamma based CSMA-CA protocol to adjust the contention parameters of nodes according to the dynamic characteristics of the network. By this adaptation, a node can access the channel with the high probability when it is idle and delivers its packet with low collision probability. As a result, a node saves a significant amount of energy that is wasted in long back-off durations and repeated channel assessments imposed by the IEEE 802.15.4 protocol. Another key reason for this saving in energy is that the Gamma based CSMA-CA protocol can mitigate the amount of energy that is spent in switching-over the transceiver between different modes. A study for the significance of switching-over energy has been introduced in [251]; that work concludes that reducing the switchover time by a factor of two reduces the average consumed energy by more than 10%.

7-8-3 Benefits of the proposed duty cycle scheme

This subsection evaluates benefits of the proposed duty cycle management scheme (that provided in section 7-7) over the two MAC protocols: the IEEE 802.15.4 and the Gamma based CSMA-CA protocol. These evaluations consider three performance metrics: average lifetime of nodes, average end-to-end delay of packets and throughput of network. The results of these assessments as illustrated in figures 7.10-7-12 are normalised to the unity duty cycle scheme of the IEEE 802.15.4 protocol. In general, it can be seen from results of these figures that the proposed duty cycle scheme is able to prolong the lifetime, increase the throughput and reduce the end-to-end delay over different network configurations for both IEEE 802.15.4 and Gamma based CSMA-CA protocols. These improvements demonstrate the advantages and wide applicability of the proposed duty-cycle scheme which is attributed mainly to the fact that the proposed duty-cycle scheme employs a non-parametric inferential procedure. In particular, the proposed scheme is not based on pre-assumptions about the dynamic behaviours of networks but it infers the potential communications based on the interacting mechanism of nodes, which in turn makes this approach effective under different configurations.

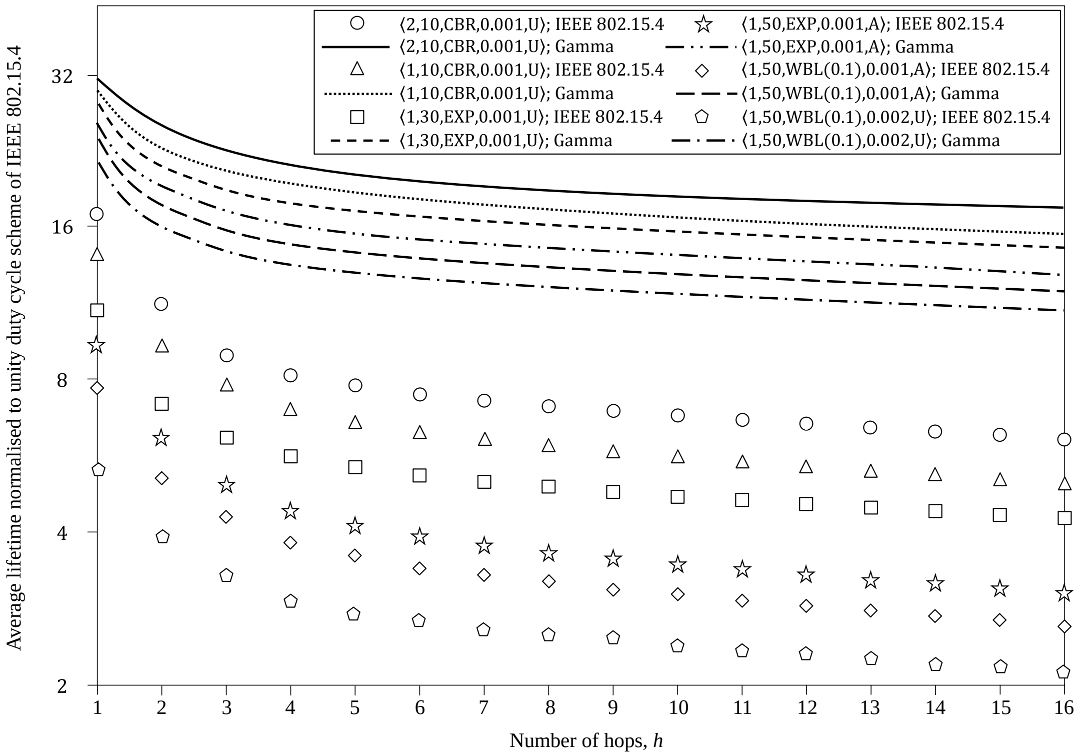


Figure 7.10 Average lifetime of AC/IEEE 802.15.4 and AC/Gamma normalised to unity duty cycle/IEEE 802.15.4

The results of figure 7.10 show that employing the proposed duty cycle management scheme over the Gamma based CSMA-CA protocol provides better performance readings compared to the cases when the proposed duty-cycle scheme is used over the IEEE 802.15.4 protocol. The main reason for this characteristic is the ability of the Gamma based CSMA-CA protocol to minimise the service times of packets as well as the probability of collision. As a result, a node spends a large portion of its life in idle state, which in turn saves the consumed energy and expedites the inferring of the waking and sleeping intervals. Another main reason for the significant increasing in the lifetime of nodes that are yielded from incorporating the proposed duty cycle management scheme with the Gamma CSMA-CA protocol is that the Gamma distribution is the conjugate prior for the probability distribution that are used to generate the inter-arrival times between packets. Hence using the Gamma distribution to generate the back-off intervals as well as the waking and sleeping durations enhances the consistency of interference process and reduces the uncertain space of interference process. In comparison, the IEEE 802.15.4 protocol suffers from high level of end-to-end delay and significant variation, which in turn requires a node to be awake for longer durations in order not to miss the routed traffic. This characteristic is attributed to the fact that the IEEE 802.15.4 protocol is based on a power law process, which has a heavy tailed distribution with higher probability of rare events [6].

The average end-to-end delay of the proposed duty cycle scheme for different network scenarios under both the IEEE 802.15.4 and Gamma based CSMA-CA protocols are illustrated in figure 7.11.

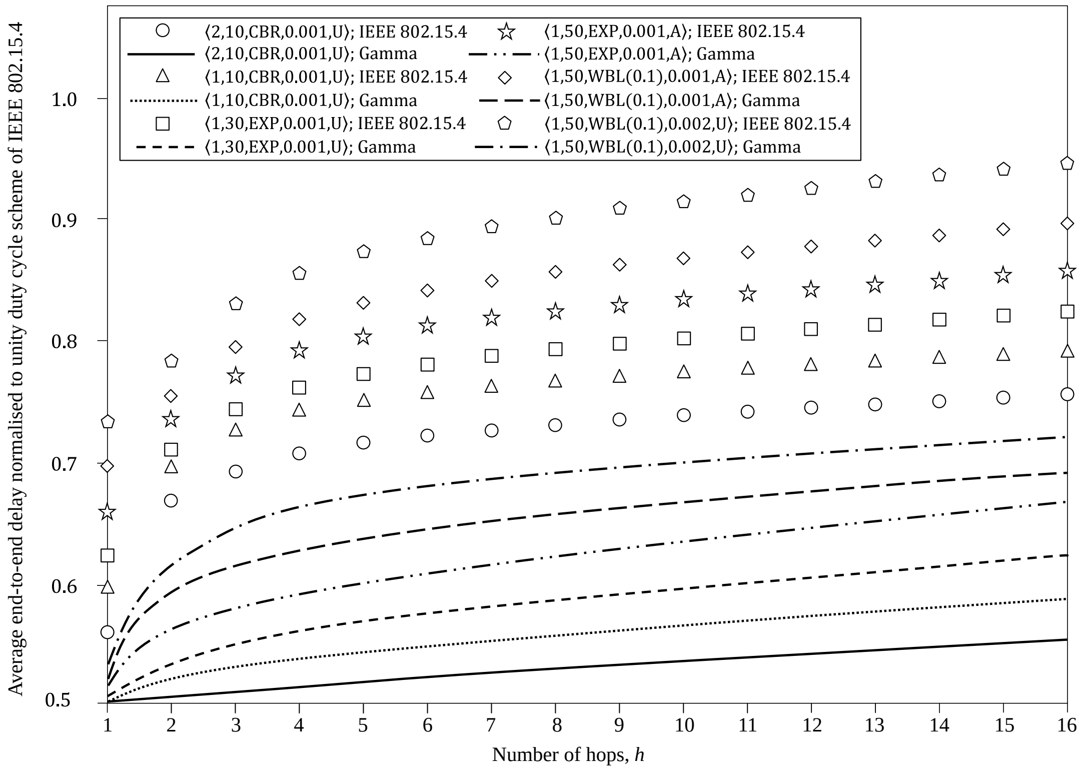


Figure 7.11 Average end-to-end delay of AC/IEEE 802.15.4 and AC/Gamma normalised to unity duty cycle/IEEE 802.15.4

It can be seen from results of figure 7.11 that using the proposed duty cycle management scheme over either the Gamma based CSMA-CA or IEEE 802.15.4 protocol can reduce the average end-to-end delay considerably compared to the cases when the unity duty cycle scheme is used over the IEEE 802.15.4. The main reason for this behaviour is that the proposed duty cycle scheme prolongs the lifetime of network hence it can maintain the shortest paths between nodes for longer time. Forwarding packets over these shortest paths leads to reduce the end-to-end delay. Conversely, in the unity duty cycle scheme, the topology of the network is changed frequently due to exhaustion of batteries of some nodes. These changes in the topology require a node to route its packets over longer paths which in turn increase the end-to-end delay of packets. Another key reason for the reduction in the end-to-end delay of packets is attributed to the mechanism of the proposed duty cycle scheme which requires a node to not to sleep unless it has forwarded its queued packets.

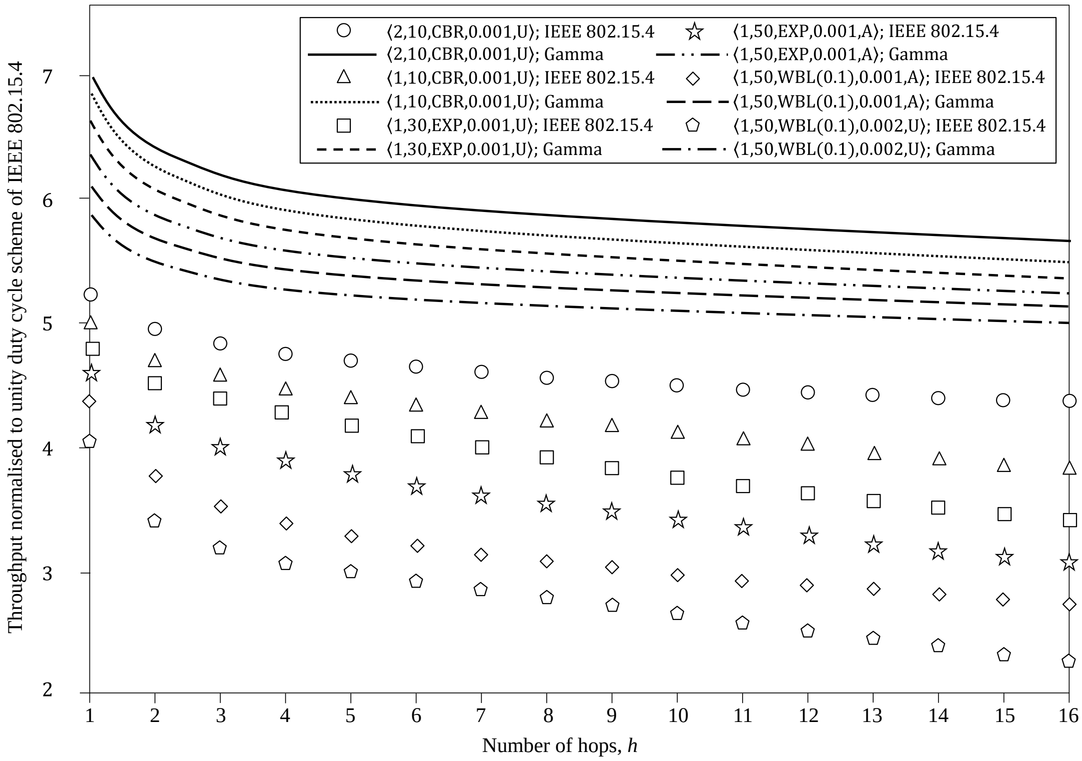


Figure 7.12 Throughput of AC/IEEE 802.15.4 and AC/Gamma normalised to unity duty cycle/IEEE 802.15.4

The results of figure 7.12 assess the advantage of the proposed duty cycle management scheme on the throughput of the networks, these results show that the proposed scheme can at least double the throughput of the network operating a highly variable MAC protocol (i.e., IEEE 802.15.4) under extremely operational conditions, i.e., the case $\langle 1,50,\mathrm{WBL}(0.1),0.002,\mathrm{U}\rangle$. This characteristic is attributed to the fact that the proposed scheme does not impose overheads on the channel as each node infers the parameters of its duty cycle without a need for control packets. Hence prolonging the lifetime of nodes enables them to deliver more packets over channels that are free from control packets, which in turn increases the throughput of the network. Moreover, prolonging the lifetime of node maintains the shortest path for longer period which in turn avoids the higher collision probability that would be resulted from sending packets over longer paths as the case in the unity duty cycle scheme.

7-9 Conclusion

This chapter has proposed a novel duty cycle management scheme for multihop wireless networks. The proposed scheme is able to achieve three concurrent goals: prolonging the lifetime of nodes, minimising the end-to-end delay of packets and leveraging the throughput of networks. The scheme is applicable to a general multihop network without restrictions on the topology configurations, traffic patterns or routing policies. Moreover, this scheme is lightweight and does not require any modification to the MAC or routing protocols. The approach that is used in deriving the proposed scheme is based on drawing analogies between the reactions of chemical species and the communication of wireless nodes. Based on this metaphor, the powerful modelling and optimisation techniques of artificial chemistry (AC) are exploited to design the proposed scheme. The molecular statistical mechanics of AC are used to quantify the energy consumed and the artificial decision making techniques of AC are employed to develop an effective duty cycle management scheme for multihop networks. The discussion and results reported in this chapter demonstrate the accuracy of the proposed model as well as the benefits of the proposed duty cycle management scheme compared to a unity duty cycle scheme over two MAC protocols: the IEEE 802.15.4 and Gamma based CSMA-CA protocols. The contributions of this chapter together with the contribution of chapter 6 represent the first two main objectives of this thesis. The next chapter will present the third objective by designing an effective routing protocol.

Chapter 8

Routing Protocol for Multihop Networks Inspired by the Harmony Search Algorithm

8-1 Introduction

The main objective of this thesis is to design effective MAC, duty cycle and routing protocols for multihop Wireless Sensor Networks (WSNs). Towards this objective, chapter 6 proposed an effective CSMA-CA protocol based on Gamma distribution back-off scheme and chapter 7 proposed an effective duty cycle management scheme inspired by principle of artificial chemistry. Comprehensive assessments of the performance of these two proposals demonstrate that they can achieve the stated design goals of this thesis, i.e., reduce the end-to-end delay of packets, prolong the life time of nodes and increase the network throughput. These results are based on a simple routing scheme that distributes traffic evenly over the shortest hop count paths. However, the hop count scheme uses a sole routing criterion depending on the number of hops between the originator and final destination despite the fact that there are more important criteria that have to be considered such as the end-to-end delay and probability of lost packets. Thereby, the need to develop a more powerful routing protocol is crucial to enhance the performance of multihop WSNs and constitutes the last step in accomplishing the main objective of this thesis.

Routing in wireless sensor networks has emerged as a direct consequence of the small transmission range of sensor nodes and the requirement to deploy these networks at large scale. Although the design of a routing protocol for communication networks is a longstanding practice that has been carried out using different approaches and techniques, most of these approaches and techniques cannot be directly employed to develop a routing protocol for WSNs. The main rationale behind this is that WSNs possess unique characteristics that need special consideration, e.g., low computation and communication capabilities of sensor nodes, ad-hoc deployment of networks, dynamic changes of topology due to depletion of batteries and heterogeneous traffic patterns. Accounting for these stochastic characteristics imposes a new set of design goals for WSN routing protocols. Some of these design goals characterise the requirements of nodes, e.g., the need to minimise the computational complexity and the communication overheads of the protocol (which is required to prolong the lifetime of nodes and save the needless channel occupancy). Other design goals typify the requirements of the operational conditions e.g., maximising the scalability, and adaptively (which is required to ensure a correct operations of the protocol when the size of network is enlarged or to cope with dynamic changes of these networks). These new design goals lead to paradigm shifts in the design of WSN routing protocols, such as using the principles of reinforcement learning [255], bio-inspired [256] or evolutionary computing [257]. Most of these approaches employ the principles of narrow artificial intelligence [258] in which a routing protocol is designed to mimic the behaviour of mindless species. Although these approaches can provide effective routing protocols, they suffer from the key limitation that their further developments are dominated by the underlying metaphors made to mimic the behaviour of mindless species.

This chapter proposes a broader perspective for designing WSN routing protocols. It conceptualises a routing protocol as a solution of a multi-objective optimisation problem [259-260] in which each node attempts to route its packets over those paths that maintain the global optimisation for the entire network. The key advantage of the proposed perspective is that it facilitates employing a wide variety of methodologies and techniques that are proposed to solve global optimisation problem in designing a routing protocol for WSNs considering their limited resources. Furthermore, it provides a unified means to obtain deep insights into the fundamental mechanisms of the well-known existing routing protocols, which in turn enable us to highlight their main contributions and limitations and draw key conclusions about the future trends in protocol design. More importantly, this chapter exploits one of the state-of-the-art optimisation algorithms, called the Harmony Search (HS) algorithm [261,262], to develop the routing protocol. The proposed protocol mimics the thinking strategy and logic reasoning of a musician ensemble during improvisation of the most pleasing harmony. Hence our proposal protocol utilises the artificial general intelligence principle which eliminates any restrictions towards future development. Moreover, the protocol is lightweight, scalable and highly adaptive as it allows each node to infer the routing metrics from the characteristics of its neighbours and utilises the principle of spatial reasoning to guide a node to discover those areas of networks that presumably yield the optimal or near optimal paths. In additional, the routing protocol provides an error-correction mechanism that reduces the possibility of routing packets over suboptimal paths. It is of interest to note that the kernel mechanisms of the proposed protocols are to enable nodes to harvest the routing metrics from the traffic, to analyse them statistically and to infer the most promising route based on these operations without a need to overwhelm the network with control packets; while the HS algorithm has been used as an inspirational source to devise these mechanisms. The benefits of the proposed protocol are assessed from different aspects, by comparing its outcomes with two existing routing protocols, IEEE 802.15.5[112] and ANT [263] under different scenarios. These assessments are also carried out over other algorithms that were proposed in this thesis, i.e., the Gamma based back-off scheme that was proposed in chapter 6 and the duty cycle management scheme that was proposed in chapter 7. The results demonstrate that the proposed protocol is able to achieve the three stated design goals of this thesis, (i.e., increase in throughput of network, prolong the lifetime of nodes and minimise the end-to-end delay of packets) without restrictions to specific topology, traffic patterns or MAC protocols.

The remainder of this chapter is organised as follows. Section 8-2 reveals the relationships between the routing protocols and optimisation problem, subsection 8-2-1 presents the mathematical formulation of the routing protocol as a solution for the multi-objective optimisation problem and subsection 8-2-2 summarises the key criteria that are used to evaluate the performance of optimisation algorithms. Section 8-3 employs these evaluation criteria to explore the fundamental approaches of well-known existing routing protocols and section 8-4 highlights the main limitations of the exiting approaches and draws some conclusions about the future design trends. Section 8-5 introduces the harmony search algorithm, section 8-6 provides the underlying approaches of the proposed protocol and section 8-7 presents the pseudo-code of the protocol. Section 8-8 assesses the performance of the proposed protocol under different cases and finally section 8-9 concludes this chapter.

8-2 Routing protocol as optimisation problem

The main contribution of this chapter is to propose an effective routing protocol for WSNs that is able to achieve three concurrent goals: increasing the throughput of networks, reducing the end-to-end delay of packets and prolonging the lifetime of nodes. The proposed protocol can achieve these goals without particular considerations for the specific topology, traffic patterns or MAC protocols. The fundamental approach of our proposal is that a routing protocol is a distributed optimisation algorithm that is able to route packets over those paths that maintain global optimisations for the entire network. Considering that WSNs are a stack of stochastic processes and there are multiple routing metrics that have to be taken in account (e.g., energy, throughput and delay) leads to the conclusion that a routing protocol can be conceptualised as a solution for a multi-objective optimisation problem. Motivated by the importance of the abstraction model for the routing protocol and connections between routing protocols and optimisation problems, this section presents a mathematical formulation for routing protocols in the context of optimisation algorithms. This formulation has two main advantages: firstly, it provides thorough understanding for the underlying routing mechanism upon which a routing protocol is built and secondly, it facilitates using the evaluation criteria of optimisation algorithms to assess the performance of routing protocols. Subsection 8-2-1 provides the mathematical formulations and subsection 8-2-2 presents some of the key evaluation criteria. The remainder of this section overviews some works that are devoted for the same purpose.

Formal analysis [264] can be considered as one of the earliest approaches that has been employed to validate the integrity of the routing protocol. This approach models a routing protocol as collections of states and rules and then uses mathematical logic to discover the correctness of the protocols under different situations. The main advantage of this approach is that it provides deep insights into the underlying characteristics of the routing protocol; however, it suffers from several limitations including overblown description and verbose manipulations. Routing algebra [265] is another approach that was proposed to abstract routing protocols. Routing algebra abstracts a network as a tuple of seven symbols, some of these symbols are used to typify characteristics of paths and nodes (e.g., weights, label, signature and attributes) and other symbols are used to define operations over these characteristics. Routing algebra utilises these parameters to express the general conditions for routing protocols (e.g., conditions under which loop may happen). Although the routing algebra provides a rigorous model for describing routing protocols, the complexity associated with its definition limits its ability to account for key characteristics of routing protocols demanded by the current and envisaged applications of WSNs, e.g., multiple routing metrics and load balance policy strategies.

8-2-1 Formulation of a routing protocol as an optimisation algorithm

This section utilises the symbols and notations that were presented in chapter 5 to formulate a routing protocol as a solution for the optimisation problem [260]. Recall that Υ was defined as a set of all nodes within a given multihop network, and $\Omega(n)$ was defined a set of neighbours of an arbitrary node n, i.e., those nodes that can correctly decode the signal received from an arbitrary node n in the absence of any distortion. Let $\mathcal{L}$ be defined as a set of ordered pairs containing directional links between neighbouring nodes, i.e., $\mathcal{L} = \{(n, x): n, x \in \Upsilon;\ x \in \Omega(n)\}$. The first element of an arbitrary member of $\mathcal{L}$ is the transmitter of that link while the second element is the receiver. All possible paths between a pair of nodes (e.g., n and u) is denoted by $\mathcal{P}(n, u)$ and defined as a set of all mutual independent paths between these two nodes. Each member of $\mathcal{P}(n, u)$ e.g., $\mathcal{P}_j(n, u)$ is a set of a sequence of nodes that starts at node n and ends at node u conditioning that any two consecutive nodes are joined by a link, (i.e., member of $\mathcal{L}$). Employing these definitions, the optimal path between nodes n and u can be defined as a solution for the optimisation problem with the following parameters:

1. Solution space which is all possible paths between node n and node u ,i.e. $\mathcal{P}(n, u)$.
2. Evaluation function which is a multivariate function that assesses the quality of the solution space from different perspectives, the criteria of this assessment are a set of decision variables (i.e., routing metrics) numbered from 1 to k in equation (8.1).

$$\mathcal{M}\left(\mathcal{P}_j(n,u)\right) = \left(m_1\left(\mathcal{P}_j(n,u)\right), \dots, m_k\left(\mathcal{P}_j(n,u)\right)\right) \tag{8.1}$$

It is worth noting that one of the most important aspects of specifying a multi-objective optimisation problem is the method by which a node articulates its preferences about the routing metrics. The main reason for this importance is that the output of an articulation method biases a given path towards the optimal solution regions. A number of articulation methods have been proposed in the literature, e.g., weighted-sum, multiplicative-sum, fuzzy logic, utility function and lexicographic approach [266].

3. Objective function which specifies goals of a routing algorithm. This function selects those paths that yield the minimal values of evaluation functions considering zero or more constraints. The output of the objective function defines as the optimal solution and denoted by $\mathcal{P}^*(n,u)$ as shown in equation (8.2) in which $\mathcal{C}_1\big(\mathcal{P}(x,y)\big),\dots,\mathcal{C}_i\big(\mathcal{P}(x,y)\big)$ represent a set of i constraints.

$$\mathcal{P}^*(n,u) = \begin{array}{c} \min\limits_{\forall \mathcal{P}_j(n,u)\in\mathcal{P}(n,u)} \left(\mathcal{M}\left(\mathcal{P}_j(n,u)\right)\right) \\ \text{subject to } \left(\mathcal{C}_1\big(\mathcal{P}(x,y)\big),\dots,\mathcal{C}_i\big(\mathcal{P}(x,y)\big)\right) \end{array} \tag{8.2}$$

4. Allocation function which is denoted by $\mathcal{H}\big(\mathcal{P}^*(n,u)\big)$ and defined as the policy by which the traffic load between node n and node u is allocated to all members of the optimal paths, i.e., $\mathcal{P}^*(n,u)$.

As an illustrative example, equations (8.3) specify a routing algorithm that distributes load between the originators and final destinations, (e.g., n and u) evenly over all shortest paths. It can be seen that the evaluation function of such a system is a single variable function that maps a given path into its cardinality, (i.e., number of nodes within a given path). The second equation selects those paths with the minimal cardinality and populates them in the set $\mathcal{P}^*(n,u)$. Finally, the last equation distributes load between nodes n and u evenly between all optimal paths (cardinality of optimal path set). Interestingly, this system of equations describes the routing scheme that has been used in previous chapters.

$$\begin{gathered} \mathcal{M}\left(\mathcal{P}_j(n,u)\right) = \left|\mathcal{P}_j(n,u)\right| \\ \mathcal{P}^*(n,u) = \min_{\forall \mathcal{P}_j(n,u)\in\mathcal{P}(n,u)} \left(\mathcal{M}\left(\mathcal{P}_j(n,u)\right)\right) \\ \mathcal{H}\big(\mathcal{P}^*(n,u)\big) = \left(\left|\mathcal{P}^*(n,u)\right|\right)^{-1} \end{gathered} \tag{8.3}$$

8-2-2 Evaluation criteria for routing protocol based on the optimisation problem

The aforementioned subsection demonstrated that a routing protocol can be formulated in the context of an optimisation problem. This section provides a general overview for the techniques that have been proposed to design optimisation algorithms and then discusses key criteria that are used to evaluate them. In section 8-3, these criteria are employed to get deep insights into well-known routing protocols that are reported in the open literature.

Solving optimisation problems is one of the most active and mature research disciplines, hence enormous numbers of algorithms have been developed, e.g., greedy algorithm, branch and bound and out-of-killer algorithm [260,266]. These algorithms can be broadly classified into three categories: complete, approximate and meta-heuristic. Complete algorithms visit the entire solution space in order to solve the optimisation problem, albeit that these techniques can always derive the optimal solution, their computational complexity, resource usage and time required to obtain these solutions increase extraordinarily with the size of the solution space. The approximation algorithms mitigate the shortcomings of the complete algorithms by eliminating some regions of solution space based on some criteria. However, these algorithms cannot always guarantee to find the optimal solution. Finally, meta-heuristic algorithms combine the benefits of complete and approximation algorithms by defining a higher-level framework to control one or more underlying optimisation procedures [267-268], hence these algorithms can approach the optimal or near optimal solutions within responsible time and resource utilisations. Instigated by these advantages, the meta-heuristic algorithms have become the de-facto approach to solving most of the optimisation problems.

Basically, meta-heuristic algorithms can be considered as a continuous interaction between two processes: diversification and intensification. [261-262, 266-268] The former process guides algorithms to discover the entire solution space while the latter process directs algorithm to focus on those regions of solution space that are presumably more promising than others. Owing to the proliferation of meta-heuristic approaches, a number of criteria have been developed to evaluate them [268-269]. Some of the key criteria that are appropriate to evaluate the performance of routing protocols are summarised as follows:

1. Balancing between diversification and intensification processes, the main reason for the importance of using this criterion to evaluate routing protocols is that emphasis on intensification can lead to discovering a part of all possible paths which may yield a suboptimal routing. On the other hand, emphasis on the diversification process leads to consumption of large amount of needless energy and may slow convergence to the optimal solution.
2. Speed of convergence to the optimal or near optimal paths, which characterises the rate with which an algorithm can approach the optimal solution in each execution cycle. The importance of this criterion stems from the fact that fast approaches to the optimal paths mitigate jumping around feasible solution regions, which can lead to inconsistency in routing protocols.
3. Resource utilisation which qualifies the amount of resource that is used by a routing protocol in order to find the optimal path.

8-3 Overview of existing routing protocols

The design of an effective routing protocol for WSNs can be considered as one of most active research areas that has received assiduous attention by a cadre of researchers in recent decades. Consequently, vast and diverse approaches and techniques have been employed to design WSN routing protocols. Comprehensive surveys for these protocols have been introduced in number of works e.g., [270-273]. The main aim of this section is not to provide a survey for all routing protocols but rather to discuss the fundamental approaches that have been used to design routing protocols for WSNs. This discussion focuses on the relationships between the routing protocols and optimisation problem provided in the previous section. The main advantage of this discussion is twofold: firstly it provides applications-independent classifications where the core approaches of the routing protocol are evaluated without consideration for particular implementation. Secondly, it demonstrates the paradigm shifts associated with development of routing protocols for WSNs.

8-3-1 Classical routing protocols

One of the earliest approaches used to design routing protocols for WSNs was re-engineering those protocols specified for mobile ad-hoc networks to suit the scarce resources of WSNs. Some well-known examples for this approach are the ZigBee routing protocol that re-engineered the ad hoc on-demand distance vector routing protocol and IEEE 802.15.5 protocol that reengineered the optimized link state routing protocol.

In this sort of routing protocol, nodes, before routing their packets, have to obtain complete information about all possible paths using proactive or reactive means. Thereafter, a node employs a shortest path algorithm such as Dijkstra, Breadth-first or Bellman-Ford algorithms to compute the optimal path based on a predefined metric. The diversification process in these protocols is based on the Lagrangian relaxation [274], i.e., a node assumes that the cost for the intended destinations is binary (either zero for the neighbouring nodes or infinite for unknown destination) a node then gradually replaces the routing cost with accurate values. Conversely, these protocols have no concise specification for the intensification process.

The obvious advantages of classical protocols is their simple implementation, however, these protocols suffer from a number of key limitations. Firstly, they depend on the greedy algorithm, which requires a node to route its packets based on the current available information without consideration for historical information or future consequences. Secondly, high overheads associated with exchanges routing information between nodes. This overhead increases channel occupancy and collision probability of data packets and exhausts the battery of nodes. Thirdly, as each node has to collect and analyse all possible paths to the final destination nodes, the computational complexity and memory usage of nodes are significant. Moreover, these protocols require a considerable amount of time to consider changes for network which limit their scalability and adaptivity. Finally, the objective function of these routing protocols is defined to find the shortest path to the final destination nodes rather than to find the optimal path that can achieve a global optimisation for the whole network.

8-3-2 Admissible heuristic based routing protocols

The key reason of the serious performance limitations of the classical-based routing protocols is that they employ a complete approach, i.e., they require a node to visit the entire solution space in order to find the shortest paths. Such approach takes a long time before obtaining complete information and requires a high computational complexity and memory utilisation. In order to mitigate this drawback, some researchers have proposed to employ the admissible heuristic approach to design routing algorithms. An admissible heuristic algorithm enhances the performance of classical routing protocols through definition of an intensification process that enables a node to focus on the promising areas of solution space.

The pioneer admissible shortest path search algorithm was called A* (also known as first-best search algorithm) [275]. This algorithm defines a new function called a heuristic function in addition to the conventional evaluation function that is used in classical routing algorithms. The main aim of the heuristic function is to underestimate the cost to the intended final destination nodes by computing the cost under optimal operational conditions. The A* algorithm distinguishes regions of the optimal or near optimal solution spaces from other region by maintaining two lists: an open list and a closed list. The open list contains solution space (nodes) that need to be visited while the closed list contains those nodes that have been examined previously. A node upon receiving routing information from its neighbour updates both the routing table and the open and closed lists accordingly.

Comparing the outcomes of the A* algorithm with the classical based routing algorithms demonstrates that the former algorithm outperforms the latter in terms of the resource utilisation and speed of convergence in finding the shortest path. This characteristics is attributed mainly to the fact that the heuristic function of A* constitutes the intensification processes which quickens finding the shortest path and saves resources. Hence it can be concluded that the classical routing algorithms can be considered a special case of the A* algorithm in which the values of the heuristic function are always equal the outcomes of the evaluation function [276]. However, these advantages of the A* lie mainly on the quality of the heuristic function. Hence a number of variations for the A* routing protocol have been proposed e.g., Iterative Deeping A*(IDA*) [277], Simplified Memory Bounded A* (SMA*) [278]. Based on these algorithms, a number of routing protocols have been proposed e.g., Ad-Hoc On Demand Distance Vector-Best Frist routing protocol (AODV-BeFS) [279], A* Algorithm for Energy Efficient Routing in Wireless Sensor Network [280] and A-Star algorithm based Energy Efficient Routing (ASSER) [281].

One of the key advantages of the admissible heuristic based routing algorithms besides their outperformance compared to the classical based routing algorithms is that it facilitates employing the principles of artificial intelligence, in particular machine learning techniques [282]. These techniques have been employed in order to enhance the quality and accuracy of the heuristic function defined in the A* by letting a node to adapt the value of the heuristic function based on the routing information received from its neighbours. This adaptation is carried out iteratively, where each iteration is defined according to specific time constraints. Geographical and Energy Aware Routing (GEAR) [283], Statistically Assisted Routing Algorithms (SARA) [248], Constraint-Based Learning Real Time A*(CB-LRTA*) [285] are some examples of the routing protocols that incorporated the admissible heuristic algorithm and learning techniques. Although admissible heuristic routing protocols can improve performance readings of networks compared to classical protocols, this improvement comes at the expense of additional overhead associated with specifying and maintaining the heuristic function.

8-3-3 Reinforcement learning based routing protocols

Success in incorporating machine learning techniques in the admissible heuristic search algorithm encouraged researchers to focus strongly on these techniques to develop routing protocols. The reinforcement learning strategy lent itself to the design of routing protocols. Reinforcement Learning (RL) [255] is a branch of machine learning concerned with building an artificial system that can learn from its experience without the need for external supervision. In RL, a learning agent (node) interacts with its environment using a trial-and-error approach. Based on this interaction, a node receives feedback qualifying the appropriateness of its action. If a node conducts a good action then a node receives a reward, otherwise a node receives a punishment. Considering these rewards and punishments, a node reinforces those actions that magnify its rewards.

Most of the RL techniques can be represented as a Markov decision process with four main components: set of all possible states, set of all potential actions, state transition probability matrix and finally a matrix of reward/punishments that an agent can receive when it transits from one state to another through a particular action. Based on these definitions, the policy function is defined as a time dependent function that characterises the behaviour of a node at any time. The main target of an agent is to derive the optimal policy that maximises its cumulative rewards. Determining reward values is carried out through a predefined function such as Q-function or least square [255-286] approaches. These functions are controlled mainly by two parameters: learning rate and discount factor. The learning rate determines the frequency with which a node replaces its current reward value with a new learned value. Precise adjustment of this learning factor is critical, as overestimation can replace high-quality current values by low-quality training data; conversely underestimation can lead to use of outdated information. The discount factor determines the weighting factor of long term rewards; mainly there are two types of rewards: short term reward (which is obtained immediately after execution of the action) and long term reward (which is obtained after a predetermined time). In RL, a node selects the next action based on the historical rewards which imposes the limitation that a node tends to use the same old solutions rather than attempting to discover alternative solutions. In order to resolve this shortcoming, the RL algorithms specify a mechanism by which trade-off between diversification and intensification processes can be adjusted; these trade-off mechanisms can be defined as a deterministic function (e.g., ϵ-greedy) or probability distribution (e.g., Boltzmann distribution) [255-286].

Q-routing [287] is a routing protocol that is based on the Q-learning approach; this protocol is model free as it assumes that a node has no prior knowledge about the environment states or the transition probabilities. Q-routing employs the bootstrapping approach in which a node samples the environment gradually and then straps new information with old. The Q-function is defined using the end-to-end delay which includes queuing, transmission as well as propagation delay of packets. Q-routing utilises real data packets to construct the routing table, i.e., an intermediate relaying node upon receiving a packet sends a backward packet containing its Q-values to the final destination. Thereafter, the previous hop updates its routing table according to the feedback from its next hop nodes. The obvious advantage of this scheme is that it alleviates the control overhead found in other routing protocols; however, the Q-routing protocol suffers from a serious disadvantage called the hysteresis problem which limits the capability of Q-routing to find the shortest paths under traffic variation [288]. Based on this limitation a number of enhancements for the Q-routing have been proposed, e.g., Dual Reinforcement Learning (Dual RL) [289] which defines the Q-values using the next state and previous states, hence a node can trace the evolution of the traffic characteristics. However, the implementation of Dual RL is more complicated compared to the Q-routing. Another enhancement for the Q-routing was proposed in the Adaptive Routing Protocol (AdaR) [290] which replaced the Q-function with least square policy iteration with the aim to reduce the state space and batch processing. Moreover, AdaR replaced the end-to-end delay of Q-routing protocol with three metrics: hop count, residual energy and aggregated ration of link reliability. Although AdaR provides better performance readings compared to Q-routing, its implementation requires high computation complexity and memory usage which limits its scalability.

The key difference between the Q-learning routing protocols and the admissible heuristic based routing protocols is that the admissible heuristic routing protocols require global knowledge for the network and environment while in Q-learning, a node constructs this knowledge over the course of the time. Thus the Q-learning algorithms are lightweight compared to the admissible heuristic algorithms; however, the disadvantage of Q-learning is that nodes require longer time to obtain enough information about the environment. Another major difference between the Q-learning and the admissible heuristic algorithms is related to the diversification process, (the admissible heuristic algorithms use the relaxation assumption). Conversely, Q-learning employs more intelligent mechanisms such as ϵ-greedy or the Boltzmann distribution. On the other hand, Q-learning imposes some major drawbacks. Firstly, an agent operating Q-learning can approach its local optimal policy only when it has a sufficient experience which is typically requires long periods over dynamic environment as the case in multihop WSNs. The second major drawback of the Q-learning is the accuracy with which a node can determine its current states, as most of wireless networks are not Markovian environments a node within these networks cannot always determine its states accuracy. Although this drawback can be overcome by using hidden Markov chains or partial observable Markov process [291], the difficulties required to implement these processes hinder their applications in low-power networks.

8-3-4 Game theory based routing protocols

The ability of reinforcement learning to achieve better performance readings compared to the heuristic and admissible heuristic algorithms encouraged a cadre of researchers to generalise single agent reinforcement learning algorithms to multi-agent algorithms. Multiple agents can share their experiences which speed up the computations and enhance the scalability and robustness of networks. This generalisation represents a category of game theory called a stochastic game [292-293]. Stimulated by this fact, game theory has been utilised to design routing algorithms either as a standalone or by incorporating it with reinforcement learning techniques.

Game Theory is a mathematical framework that investigates the competitive interactions amongst self-interested and rational agents. Game theory models an interacting system as a game of a number of players and each player has a set of actions, each player defines its payoff function in terms of the cost and rewards of a possible action. The ultimate goal of a player is to find the optimal strategy that increases its rewards and reduces the incurred cost, when all players apply their optimal strategies, it is said that the system is operating in the equilibrium point. One of the crucial contributions of the game theory is to determine mathematically whether or not a system can reach the equilibrium point under specific strategies, accordingly a number of analytical models has been developed, e.g., Nash, correlated, Wardrop equilibriums [293]. Game theory provides comprehensive classification for games based on different perspectives, e.g., according to the ability of players to cooperate (cooperative vs. non-cooperative games); according to the information that is available for players (complete vs. incomplete information games); according to the computation of reward and cost (zero-sum vs. non-zero sum games) and according to the dynamic characteristics of the environment shared by players (stationary vs. stochastic games) [292-293].

The concrete pillar upon which the game theory is constructed is the assumption that all players are rational, the definition of rational players has been introduced by some authors using three axioms: completeness, transitivity and continuity [294]. However, design of an artificial system that is able to act rationally is associated always with great challenges, hence a number of dilemmas and paradoxes has been emerged when game theory is applied to the design of routing protocols. For example, the forwarder's dilemma [295] in which a node drops packets received from other nodes instead of routing them in order to reduce its cost (amount of energy required to route packets). Another example is known as a selfish routing phenomenon [295-296], which results from being all nodes (players) always selecting paths that optimise their payoffs regardless of others. This leads to misutilisation amongst links of networks. In order to overcome this phenomenon, sophisticated economic theories such as the price of anarchy and price stability have to be used [296]. A third example for the difficulties of straightforward application of game theory to design routing protocols is Braess's paradox [296] which states that when a network reaches an equilibrium regime using specific strategies then adding more resources can lead, in some cases, to another equilibrium regime in which the rewards of all players are reduced. However, some studies e.g., [297-298] demonstrated that the main reason for this behaviour is that the game theory requires each agent to use the same strategy to determine the best actions which causes crowds at additional resources.

Incorporating the principles of game theory with reinforcement learning to design routing protocols has been proposed under Multi-agent reinforcement learning (MARL) protocols, e.g., [299]. In these protocols, all nodes within a network share their experiences and can transfer their knowledge to new nodes. The main advantages of MARL are obvious, i.e., parallel computation, scalability and robustness of the network. However, this approach has some limitations, such as exponential growth of the state-action space which results from the requirement that each node has to keep trace of the shared environment as well as the status of other nodes in order to adapt its behaviour accordingly. Another limitation is the requirement to conduct comprehensive and continuous diversification and intensification processes in order to keep a node up to date with the stochastic characteristics of the shared environment. Another major drawback is the need for coordination between nodes in order to obtain mutually consistent actions between all nodes [255-299].

8-3-5 Swarm intelligent based routing protocols

Admissible heuristic, RL and game theory based routing algorithms improve performance readings compared to the classical routing algorithms. The main reason for this improvement is that these protocols are able to discriminate between the optimal or near optimal solution regions from other regions. This discrimination is carried-out by allowing a node to interact with its environment and evaluate the benefits of these interactions using the cost-to-go function (i.e., the heuristic function in admissible heuristic; Q-function of Q-learning algorithms and payoff function in game theory). However, the downside of these approaches is that increasing the solution space leads to an increase in the overhead required to maintain the quality of the cost-to-go function. Motivated by this shortcoming, the swarm intelligence approach has been proposed to design routing protocols. Swarm intelligence [300-302] differs from the aforementioned algorithms substantially as it provides the discrimination between the optimal or near optimal solution regions from other regions through modification of the environment using indirect communications between nodes.

Swarm Intelligence (SI) is a branch of the artificial intelligence discipline that aims to design computational algorithms or problem solving techniques inspired by the collective behaviour of social insets and other animal societies. The SI algorithms are built mainly on two design aspects: firstly, the emergence principle that exhibits the ability of mindless creatures, (e.g., ants or bees) to accomplish collectively a sophisticated global objective (e.g., nest building or larva sorting). Hence the term intelligence in SI is the property of the whole system rather than the result of the intelligence of individual constituents. Secondly, the Stigmergy mechanism [300-302] that enables agents to coordinate their actions without a need for external coordination. One of the most well-known and inspirational sources for designing routing protocols is the behaviour of ants when they are searching for food. Foraging ants deposit chemical substance (phenomena) on the path from the nest to the food source. The concentration of the phenomena on a path is enforced by a number of ants walking on the path and is evaporated with time [303]. Ants can sense the pheromone and can follow the path that is discovered by other ants; hence ant colonies can find the shortest path from the nest to the food source collectively through indirect communication.

Ant colony routing algorithms (ANT) [263] have been proposed to imitate the behaviour of foraging ants. In these algorithms, the artificial ants (control packets) mimic real ants while the pheromone trails is defined based on the routing metrics (arrival time at each destination, end-to-end delay, number of hops). A typical ant colony routing algorithm uses two types of artificial ants: forward and backward. The forward artificial ants are constructed by each node and sent to other nodes within the network periodically. Forward packets traverse network by selecting the next-hop randomly and exchange routing metrics on their path, each node stores the concentration of pheromones in the pheromone table. This table maintains a real-value entity for each neighbour/final destination pair qualified by the percentage of traffic that is routed to the final destination via that neighbour. Similar to the real ant colony, increasing the value of the pheromone concentration of the neighbour/final destination pair reflects the goodness of using that neighbour to reach the given final destination. When a forwarded packet arrives to its final destination, it is replaced with the backward packets that traverse the network from the final destination to the originator following exactly the same path. Additional to the artificial ants and pheromone table, ant colony based routing algorithms define the decay function which mimics the evaporation of pheromone in a real ant colony. This function is used to update the routing information in pheromone tables.

The ant colony routing algorithms demonstrate two key advantages compared to other routing algorithms that are adaptiveness and robustness. These advantages result from the fact that ant colony routing algorithms employ a huge number of agents (artificial ants) that work in parallel, hence these agents can learn about changes in a network quickly and persistently even if some of these agents are lost due to collisions or communication errors. However, there are a number of disadvantages of ant colony routing algorithms. First of all, the communication overhead associated with sending and receiving the forward and backward artificial ants. This overhead has a detrimental effect when the size of these packet increase either due to the increase in the number of routing metrics collected by these packets or with the diameter of network. Secondly, the convergence of the ant-colony algorithm to finding the optimal solution is not always guaranteed [303-304]. This characteristic is attributed mainly to the fact that the quality of the diversification process depends mainly on the intensity and coverage area of artificial ants while the quality of the intensification process lies mainly on the parameters of the decay functions.

8-3-6 Genetic algorithm based routing protocols

It has been shown in the previous subsection how the swarm intelligence routing algorithms allow nodes to cooperatively explore the optimal routes, which leads to better performance readings compared to the aforementioned routing algorithms. However, the weaknesses of this approach are that it requires considerable communication overhead in order to maintain this cooperation effectively and can diverge from the optimal or near optimal solution due to the probabilistic forward diversification mechanism. These shortcomings motived researchers to employ evolutionary algorithms [257] which provide powerful manipulation tools for the solution space.

Evolutionary algorithms represent a subfield of evolutionary computing that aim to solve continuous and combinational optimisation problems inspired by the Darwinian theory of evolution [305]. The Darwinian theory of evolution depicts a biological system as a population of individuals that encounters environmental pressure, due to this pressure some of these individuals are extinct while other individuals survive and reproduce next generations (offspring). Offspring inherit some characteristics of their parents while other characteristics are modified as a result of natural variation, hence the biological system can be considered as a product of continuous process of natural selection. Inspired by this approach, the evolutionary algorithms abstract the solution space of an optimisation problem as population individuals and define a set of operators to mimic the effects of the environmental pressure. The main aim of the evolutionary algorithm is to eliminate suboptimal solutions which in turn lead to approach the optimal or near optimal solution rapidly. Evolutionary algorithms encompass three main branches: evolutionary programming, evolution strategies, and genetic algorithms, however as the genetic algorithm is the most widely used algorithm in designing routing protocols; the remainder of this subsection considers it in more detail.

Genetic Algorithms (GAs) [230] are inspired by a mechanism of genetic inheritance that takes place in biotic organisms. GAs can be considered as a microevolution system in which the population individuals are set of chromosomes and the environmental pressure is simulated using three operators: selection, mutation and crossover. The main aim of GA is to mimic the competing behaviour of the chromosomes which leads to dominate traits of the strongest encoded genes. The first step in GA based routing protocols is representation, which encodes a path between a pair of nodes as a vector of symbols (e.g., binary, number, character). The main aim of the representation is to facilitate applying the genetic operators rigorously and to parameterise the characteristics of each path thoroughly (e.g., in terms of routing metrics). After encoding routing paths, a node starts manipulating these routes using the genetic operators, the first operator is selection. The main aim of the selection operator is to choose individuals based on some criteria in order to pass them to other operators. GAs employ a number of selection strategies including Elitist, Roulette-wheel, Fitness-proportionate, scaling and tournament [306]. Those individuals that have been selected are passed to the crossover operator. Crossover is a binary operator that merges two existing solution (parents) in order to generate a new solution (offspring) similar to the matting process used in chromosomes. The key principle behind the crossover is to generate a new solution, whose attributes are closer for the optimal or near optimal region than the original solutions. The crossover operator is based on a stochastic process i.e., the selection of parents as well as the merging techniques are random processes. There are a plenty of techniques that can be used to perform the crossover, e.g., the single point crossover which splits each parents randomly into two parts and then combines these parts randomly to form the offspring; or uniform crossover which selects the value of a location in the offspring from its parents with equal probabilities. However, as these schemes can produce a loop in routing protocols, more sophisticated crossover techniques have been employed, e.g., multiple crossover, partially matched crossover, cycle crossover and order crossover [307]. After performing the crossover operations on the selected chromosomes, a node passes them to the mutation operator. The mutation operator mimics the effect of natural variation in real chromosomes by modifying one or more attributes of the candidate solution in order to fasten its approach to the optimal or near optimal solution region. The mutation operator is unary as it takes one routing path and modifies its attributes stochastically using some functions (e.g., bit flipping, interchanging, inversion, insertion [306-307]). The output of the mutation operator is validated and then assessed using the objective function.

Comparing the GA with other routing algorithms shows some key advantages: firstly, GAs are particularly suitable for large scale networks with a huge solution space and multi-objective criteria [307-308]. The main reason for this characteristic is that genetic operators are able to diversify and intensify the solution space exhaustively. For instance, the crossover facilitates transferring information between candidate solutions while the mutation operator allows each candidate solution to discover its immediate neighbours. Secondly, the GA mitigates the communication overhead found in some routing algorithms, (for instance swarm intelligence or classical based routing protocols) as all of the genetic operators are applied to the solution space locally. Thirdly, the solutions generated by GAs converge to the optimal or near optimal solution with high speed. This characteristic is attributed mainly to the fact that better offspring solutions replace their low-fitted parents immediately which in turn reduces the number of unfitted solutions over the time. However, the key limitation of the GA based routing protocol is its requirement to define the robust encoded scheme that is able to reflect the dynamic characteristics of the network. Without this effective scheme the result of GA is meaningless.

8-3-7 Fuzzy logic based routing protocols

Previous subsections have reviewed some approaches that have been employed to design routing protocols for WSNs. Each one of these approaches is built on specific assumptions that devise techniques and control the behaviour of nodes. However, all of the aforementioned approaches can be treated equally from the perspective that all of them, more or less, represent non-knowledge based approaches, i.e., the decision making of these approaches depends solely on information acquired from the shared environment. Hence the efficiency of these approaches lies mainly on the status of the shared environment which could be affected by some factors like: uncertainty and ambiguity. One of the powerful approaches that has been proposed to overcome this shortcoming is fuzzy logic.

Fuzzy logic can be seen as an extension for Boolean logic: in fuzzy logic everything is associated with a degree that determines its certainty with respect to fuzzy system [309-310]. For example, a number can be partially member of a given fuzzy set according to a specific degree that is assigned to it; a statement can be simultaneously partly true and partly false depending on its certainty degree, and the fuzzy function can be defined using hedge expressions (e.g., few, several and usually). A fuzzy system is defined in term of three components: fuzzifier, fuzzy inference controller and defuzzifier. The fuzzifier converts the information received from environment (called crisp in fuzzy logic) into their corresponding fuzzy representation. The fuzzifier performs this mapping with the help of a membership function, which is a function that allocates the certainty degree for each possible input. The membership function can be considered as the knowledge base in the fuzzy system. This function is fed into the system from experts and could take different forms, e.g., triangle waveform, trapezoidal, Gaussian or Sigmoidal functions. The second component of a fuzzy system is a fuzzy inference controller which can be considered as the heart of system. This controller uses a set of IF-THEN statements in order to derive the fuzzy output (hence the fuzzy logic is characterised as rule based reasoning). The last component is the defuzzifier which converts the fuzzy output to their crisp values. A number of methods have been proposed in order to accomplish this defuzzification, and some of these are: mean of maximum method, centre of gravity method and height method.

Employing fuzzy logic in designing routing protocols has been concentrated heavily on the ability of the fuzzy system to solve a multi-objective optimisation problem by assigning a certain degree to different objectives. For instance, work [311] uses buffer occupancy and the residual energy of nodes to construct the membership function and then defines the stability of network as a function of the output of the membership function. That work uses the optimized link state routing protocol as a packet forwarding mechanism, and the performance assessment of [311] demonstrates the outperformance of incorporating the fuzzy logic. Comparing fuzzy logic with the remaining approaches demonstrates its ability to mitigate the uncertainly associated with the received data by biasing this input against a membership function (knowledge based). However, as these functions are fed from an external expert, they cannot always account for the runtime dynamic characteristics of the network. Alleviation for this limitation has been proposed by incorporating the learning techniques with fuzzy logic. The second main limitation of the fuzzy logic is the computational complexity required to perform fuzzification and defuzzification [312].

8-4 Summary of the current situations and future trends

The previous subsection provides a solid walkthrough of the well-known approaches that have been used to design routing protocols for WSNs. Although each one of them has its own design philosophy, deep consideration of these approaches facilitates obtaining some insights about the current situations and future trends.

In general, it can be seen that most of the routing protocols suffer from one of more key limitations that hinder their wide applicability for current and envisaged WSNs applications. For instance the classical and admissible heuristic routing protocols suffer from large communication overheads, and the reinforcement learning and game theory approaches mitigate these shortcoming at the expense of computational complexity. Swarm intelligence provides a good balance between computational complexity and communication overhead but it cannot always provide the optimal routes. The Genetic algorithm enhances the optimal convergence of the swarm intelligence; however, it suffers from resource overuse. Finally, fuzzy logic is proposed to achieve better performance over uncertain environments, however, it requires detailed prior-knowledge for the characteristics of networks. Considering these limitations highlights the need for a new paradigm for designing WSN routing protocols. The remaining of this section exploits the aforementioned discussion to draw the future trends of design routing protocols for WSNs.

1. There is a strong trend in employing the principle of the meta-heuristic approaches (i.e., by defining a routing protocol as a general framework and leaving the detailed specification for the implementers) in designing routing protocols. Following this approach exhibits the inapplicability of a single specification protocol to cope with the complexity of the current and envisaged applications of WSNs. A justification for the aforementioned conclusion can be acquired by considering the development timeline of the aforementioned approaches. Comparing reinforcement learning to classical based routing protocols demonstrates that reinforcement learning provides a general framework within which a node can learn from its environment, whereas the detailed specifications of routing protocols (e.g., reward computation, packet forwarding mechanism) are left for the implementers. Following the same argument, SI can be considered as a generalisation for RL in which a node not only learns through explicit interaction but also through the modification of environment, (i.e., using indirect communication between nodes). Genetic algorithms extend the process of the direct and indirect communication by adding the environmental pressure (genetic operators).
2. There is, to some extent, a growing tendency towards utilising strong artificial intelligence instead of narrow artificial intelligence. Support for this conclusion can be drawn by assessment of artificial intelligence based routing protocols from the perspective related to self-awareness and the independence in solving a given problem. For instance, in RL, it is assumed that agents are not aware about the global goal but rather that each node endeavours to increase its rewards earned from the environment. Moreover, as the agents of RL are reliant on the environment, they cannot solve a given problem independently. In SI based protocols, agents are aware about their global goal and hence they cooperatively learn each other. However, agents in the SI protocols cannot solve a problem independently, as the underlying assumption of these protocols is that a wireless node represents the alive parts (e.g., either ant's nests or food sources) while the biotic parts are the artificial ants (i.e., control packets). Finally, in the GA algorithms, nodes are not self-aware but agents of a GA can work independently to solve a given problem (i.e., by applying the genetic operators). Further support for moving towards strong artificial intelligence has appeared in recent researches which recommend moving from artificial intelligence to artificial general intelligence [285].

Motivated by the discussion of this section and the limitations of the existing routing protocols reported in this section, this chapter proposes to employ the Harmony Search (HS) algorithm to design an effective routing protocol for WSNs.

8-5 Introduction to the harmony search algorithm

A Harmony Search algorithm (HS) is a meta-heuristic algorithm proposed by Geem in 2000 [261-262]. Although this algorithm is relatively new, it has been used successfully to solve a number of optimisation problems in a vast area including economy, logistics management, game design, web page clustering, visual tracking, robotics, energy system dispatch, photo-electronic detection, water network design, structure design, hearing aids, genetic structure prediction and computer vision. A comprehensive survey of the applications of HS can be found in [261-262,313] and their reference lists.

HS is inspired by the behaviour of a musical ensemble when they are endeavouring to improvise the most pleasing harmony. In such an ensemble, a number of skilled musicians, each one of them playing a different instrument, collaborate spontaneously in order to produce a highly aesthetic harmony without the aid of a maestro, manuscript or even preparation. A musician engaged in an improvisation utilises almost all of her/his aptitudes including physical competence (to play the instrument); cognitive and memory capabilities (to devise and retrieve musical notes) and emotion intensity (to capture attentions, unravel events and predict the potential actions of other musicians) to respond on the spur of the moment to unexpected actions (either from the remaining musicians or from audience). The improvisation is a creative process that has been investigated from different perspectives, e.g., psychology, thinking strategy, error-correction multi-culture interactions and logic reasoning [314-320]. Moreover, it constitutes rich inspirational sources for computing research including: emotion systems [321] and live algorithms [322].

From an engineering perspective, a HS algorithm can be considered as a set of activities that are conducted by a collection of agents in real time. Each one of these agents has its own distinct capabilities and its own perspective of the interaction and these agents share an error-correction system that refines their actions towards achieving their global goal. The underlying approach of the HS is the improvisation process that allows different strategies to be incorporated effectively by the same agent under a common management scheme that guides these strategies according to their efficiency in achieving the desired objective and in accordance to the available resource. The ability of HS to provide such general framework that is not restricted to a specific mechanism is the main reason behind the wide applicability of HS in solving optimisation problems in different disciplines. The improvisation process of HS provides rich design aspects that can satisfy sophisticated challenges imposed on the design space of WSN routing protocols. Some of the key aspects are summarised in the following:

1. The HS imitates the behaviour of skilled human beings (musicians) and hence HS enables us to incorporate the higher cognitive capability of humans (e.g., inference and reasoning) which can increase the performance of the developed protocols. More importantly it provides an open-ended approach that can be extended to mimic operations of real WSNs without being restricted to the specific behaviour of mindless species as is the case in the swarm intelligence or genetic algorithms. Another key advantage of using HS to design routing protocols for WSNs is that it is realises using one of the future research disciplines which is general artificial intelligence.

2. Improvisation is a heterogeneous collaborative approach in which each agent participates in achieving the global goal using its own capabilities and strategies that may differ substantially from their peers. Improvisation can be considered as a process that manages different activities from different agents and refines them towards achieving the global goal. Conversely, other techniques such as swarm intelligence, reinforcement learning or evolutionary algorithms are based on the fundamental assumption that all agents have identical capabilities and all of them follow the same strategies. Hence an attempt to incorporate agents with different capabilities can be complicated or may lead to inconsistent outcomes. The importance of such an approach in the design of routing protocols for WSNs is that it facilitates routing of packets over different characteristics networks using different schemes as is the case with the Internet of Things where packets traverse from WSNs to the Internet.
3. Improvisation has a number of feedback mechanisms at different levels including the feedback received from outside the improvisation system (appreciation received from audiences of a musical ensemble); feedback received from peer agents (by improving the aesthetic quality) and self-reflection mechanism that enables an agent to use another strategy when it realises that the current one is not sufficient enough. These different feedback systems are integrated to form an effective error-correction policy that enables improvisers to approach their global goal fast. This is important in designing an effective routing protocol as it provides a means by which nodes can reduce the possibility of routing packets over suboptimal paths.
4. HS provides a flexible framework that can interweave other meta-heuristic algorithms which is important when designing a routing protocol that is able to deal with different stochastic processes in different communication layers.
5. HS can solve optimisation problems that involve continuous or discrete parameters or combinations of them, which allows a routing protocol to find the optimal path using both discrete (e.g., inter-arrival of packets) or continuous (e.g., mobility) routing metrics.
6. HS is a lightweight algorithm and does not require pre-setting of initial conditions. The impact of this property in the design of an effective routing protocol emerged from the fact that a WSNs is a set of stochastic processes. Hence it is always difficult to define the initial conditions that ensure finding the optimal route.
7. Solutions that are generated by the HS algorithm are absolutely convergent to the optimal value. This convergence has the advantage that it ensures that the path that is estimated by a node approaches the optimal or near optimal route over the time.

Table 8.1 compares the main characteristic of HS with other approaches that are proposed in the literature and overviewed in the previous sections.

Table 8-1 Comparison between the well-known approached used to develop WSN routing protocols

Criteria	Classical	Admissible heuristic	Reinforcement learning
Agents	Not defined	Not defined	Not defined
Inspirational source	Not defined	Not defined	Animal learning
Diversification	Lagrange relaxation	Lagrange relaxation	ϵ-greed or Boltzmann distribution
Intensification	Not defined	Heuristic function	
Convergence to optimal or near optimal solution	Absolutely	Probabilistically	Probabilistically
Computational complexity	High	High	High
Communication overhead	High	High	Low

Criteria	Ant colony	Genetic	Harmony search
Agents	Artificial ants	Chromosomes	Musicians
Inspirational source	Emergence	Natural selection	Musical Improvisation
Diversification	Randomness	Selection operator	Divergent thinking strategy, spatial reasoning and error correction
Intensification	Pheromone concentration and the evaporation rate	Mutation and crossover operators	
Convergence to optimal or near optimal solution	Probabilistically	Absolutely	Absolutely
Computational complexity	Low	High	Low
Communication overhead	High	Low	Low

8-6 Underlying approach of the proposed protocol

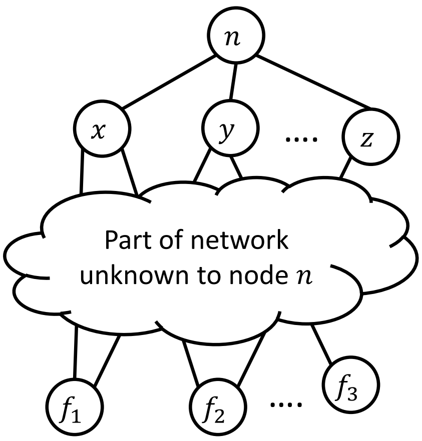
This chapter utilises the principle of the HS algorithm to design an effective routing protocol for WSNs. The design objective of this protocol is to enable each node to route its packets over those paths that maintain the global optimisation for the entire network without particular considerations to the specific topology configurations, MAC protocols or traffic patterns. The proposed protocol represents a wireless network as a musical ensemble in which each node plays the role of a musical player while the interactions between these nodes are carried out by mimicking the improvisation process artificially. The proposed protocol abstracts the behaviour of a musical player from two aspects: thinking strategy and logic reasoning. The thinking strategy enables a node to conceptualise the network in the perspective that is similar to the vision of a musical player to the ensemble. The logic reasoning enhances this thinking strategy by allowing a node to direct its future searching activities based on the conclusions that are drawn from the status of the network. Moreover, the proposed protocol provides a powerful error-correction mechanism that can reduce the possibility of routing packets over suboptimal paths. This section discusses the underlying approaches of the proposed protocol from these three aspects. It is of interest to recall that the proposed routing is based on the assumptions that each node can harvest the routing metrics form the traffic, analyse them statistically and then infer the most promising route based on the outcomes of these operations and without a need to overwhelm the network with control packet; whilst HS has been used as an inspirational source to devise these assumptions.

8-6-1 Thinking strategy

The foremost fundamental design objective of the proposed protocol is to imitate the thinking strategy of a musician during the improvisation process. Imitation of the thinking strategy of a musician (or a human being in general) is a highly complicated task that is under continuously investigation [315]. However it has been shown that, in a broad sense, a human utilises two different thinking strategies: convergent and divergent. A human uses the convergent thinking when he/she attempts to solve a well-defined problem with predictable results (such as when performing routine tasks). On the contrary, the divergent thinking strategy is employed to solve problems that convey uncertainty or when attaining to produce a creative solution as in the case of improvisation. One of the well-established theories that has been used to explain the divergent thinking strategy is the schema theory [323]. This theory states that the knowledge inside a human mind is organised into units (called schema) within which information is stored. These schemata represent knowledge about situations, events, actions as well as the concepts in terms of the objects and a set of relationships between these objects. Based on this theory, divergent thinking can be abstracted as recalling the previous schemata and regrouping their information in a new perspective corresponding to the problem under consideration. Providing a complete implementation of such an approach in an artificial system is extremely difficult as it requires the development of a higher-level process that can recognise information and their relations and to devise a scheme that is able to provide a meaningful regrouping of the stored information in the context of the problem under consideration. However, some work has attempted to emulate the object-relationship aspects of the divergent thinking, e.g., object-oriented analysis [324] and meta-programming [325].

This section employs a simple mean to mimic the fundamental approach of the divergent thinking strategy. Basically, an arbitrary node n classifies any other node as either a neighbour (a node that can receive packets of node n directly) or a final destination (a node that needs one or more intermediate relaying nodes to receive packets of node n). Based on this classification, a node abstracts the network as a collection of objects (i.e., neighbours) that are characterised by a set of attributes. Some of these attributes typify the stochastic behaviours of a neighbour (e.g., rate of lost packets, utilisation factor, and expected delay). Other attributes describe the spatial characteristics of network (e.g., dense and reachability). The proposed protocol allows a node to infer most of these attributes from the received packets without a need for probe packets. These attributes are then used as routing metrics as well as guiding parameters for the intensification and diversification processes. The main advantage of this approach is twofold: firstly, it provides a modular framework in which new attributes can be appended to the existing attributes readily. Furthermore, it enables nodes with different capabilities to use different attributes without requiring redesign of the routing protocol. Secondly, it fastens approaching to the global optimal paths by eliminating the inconsistency of using different parameters for routing metrics and searching processes.

Figure 8.1 shows the conceptual diagram for a simple network from the perspective of an arbitrary node n. It can be seen that node n recognises other nodes of networks as either neighbour (nodes x, y and z) or final destination nodes (f_1, f_2 and f_3). The table on the right shows the values of attributes for each neighbour, these values are inferred from the received packets without a need for probe packets. For example the final destination attribute (FISD) is defined as a bivariate attribute, the first element of FISD stores the address of the final destination while the second elements stores the distance to that final destination which is obtained from the Time To Live (TTL) field of the received packet.



Neighbour	LOST	DELY	UTIS	DIST	FISD
x	0.25	0.4	0.3	4	$(f_1,4)$,.., $(f_2,3)$
y	0.35	0.6	0.25	10	$(f_2,6)$,.., $(f_3,5)$
z	0.78	0.36	0.54	2	$(f_1,6)$,.., $(f_3,5)$

Figure 8.1 Conceptual diagram for the network from perspective of node n

8-6-2 Spatial reasoning

The second main design objective of the proposed protocol is to mimic the reasoning capability of a musician during the improvisation process. This reasoning capability enables a musician to make his/her future actions based on the conclusions that are drawn from his/her interpretation of the explicit or implicit actions of other musicians or spectators (e.g., body language and domain specific cues). The importance of the reasoning in the musical improvisation is that it facilitates devising a future plan based on the incomplete information that is acquired instantaneously and without need to keep track of historical data. Moreover, considering that each musician may interpret the same action differently and thereby he/she may devise different strategy to improve the aesthetic quality of the played harmony enables us to recognise the advantage of using reasoning in unifying the effort of heterogeneous agents each of which has his/own distinct vision and capabilities to achieve a global goal.

Inspired by the importance of reasoning in the improvisation process, the proposed protocol allows a node to use the principles of spatial reasoning [326] to enhance its search process effectively. Instead of conducting a blind search in each direction of the network (as the case in the classical or ant colony based routing protocols), a node in our proposal focuses on those areas of networks that presumably have the optimal or near optimal paths which in turn saves resources and fasten approaching to the global optimal paths for the whole networks. The spatial reasoning is a branch of inferential statistics that qualifies the spatial representation of a collection of objects over a geometrical space. Spatial reasoning uses the first order predicate calculus which enables it not just to describe the relationships between objects spatially but also to draw some conclusions about their hidden relationships [11,326]. Spatial reasoning is a mature research discipline that has many different approaches and techniques ranging from geospatial [327] that is used in geographical information systems to mereotopology [328] that is used to describe the connection and part-hood over a topology.

This chapter considers the scarce resources of the WSNs and uses a lightweight version of the spatial reasoning known as Region Connection Calculus 5 (RCC5) [326,329]. RCC5 regards a geometrical space as a collection of regions whose relationships can be described by five jointly exhaustive and pairwise disjoint relationships. In order to demonstrate the benefit of such approaches in designing an effective routing protocol, let us starts by casting definitions of the RCC5 in the context of wireless networks. From a perspective of node n each final destination with all neighbours that know this final destination can be considered as a region. Although node n has no complete information about all intermediate nodes between the final destination and its neighbour, node n recognises the distance between the final destination and each neighbour through the value of TTL of the received packet as shown in figure 8.1. Following the notations of RCC5, the relationships between a pair of regions of final destinations (e.g., f_1 and f_2) can be typified by one of the following Boolean relationships:

1. $DR(f_1, f_2)$: the regions of final destinations f_1 and f_2 are disconnected, i.e., both final destinations f_1 and f_2 cannot be reached by a common neighbour. Figure 8.2.a shows that the regions of the final destinations f_1 and f_2 are disconnected.

2. $EQ(f_1, f_2)$: the region of the final destination f_1 equals the region of the final destination f_2, i.e., both of these destinations can be reached by the same neighbours and through the same values of TTL. Figure 8.2.b shows that both regions of f_1 and f_2 are equal as both of them can be reached by neighbours x and y and through the same TTL values. It is worth noting that the equal relationship does not mean that these regions are dimensionally identical but rather are coincide.
3. $PP(f_1, f_2)$: the region of the final destination f_1 is a proper part of the region of the final destination f_2, i.e., all neighbours that can reach final destination f_2 can also reach the final destination f_1 but with smaller values of TTL. Figure 8.2.c shows that the region of f_1 is a proper part of the region of f_2.
4. $PPI(f_1, f_2)$: the region of the final destination f_1 is not a proper part of the region of final destination f_1. This relation is the complement of the relation $PP(f_1, f_2)$. Figure 8.2.c shows that the region of f_2 is not a proper part of the region of f_1.
5. $PO(f_1, f_2)$: the regions of the final destinations f_1 and f_2 are overlapped, i.e., the final destination f_1 and f_2 share a common neighbour at least and both f_1 and f_2 are not related by proper part of equal relations. Figure 8.2.d shows that the regions of final destinations f_1 and f_2 are overlapped.

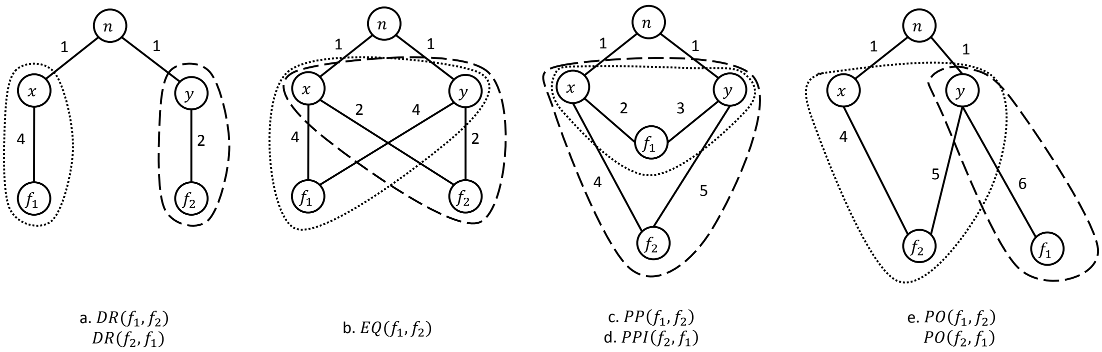


Figure 8.2 RCC5 relationships in the context of wireless networks

The proposed protocol utilises RCC5 relationships to enhance the effective of the routing protocol by introducing the so-called goal oriented search. The goal oriented search enables a node to guide its diversification or intensification processes to achieve a certain goal such as finding alternative paths for disconnected regions of final destinations or discovering the minimal connected dominating set. Finding the alternative paths for final destinations that are disconnected from other regions can be carried out simply by letting a node to focus its search on those regions of final destinations that have the largest number of disconnected relationships. The obvious benefit of this search is that it enables a node to reach those final destinations through more than a single neighbour which in turn enhances reliability of routing protocol by providing redundant paths and saves resources of networks by distributing the load over multiple paths. Likewise discovering the minimal connected dominating set of network can be attained by working out the not proper part relationships. The main benefit of obtaining information about the connected dominating set is that this set represents the virtual backbone of the network. Hence a node can use members of this set to find routes to unknown final destinations. Following the same approach enable us to realise that many different search goals can be defined and obtained from the RCC5 relationships readily.

8-6-3 Error correction mechanism

The last main design objective of the proposed protocol is to mimic the error-correction machine of the improvisation process which enables a musician to correct a mistake that are commenced either by themselves or by other musicians. The proposed protocol provides two error correction mechanisms. The first mechanism enables a node when receiving a packet over a suboptimal path to notify the originator of the received packet regarding other paths. The second error correction mechanism enables a node to refine its search mechanism based on the fluctuation of the inferred attributes. The key advantage of this mechanism is that it can enhance the routing stability by predicating the potential traffic congestion of topology changes.

8-7 Pseudo-code of the proposed protocol

The proposed protocol is a generic routing protocol that does not require any pre-setting or assumption regarding the traffic patterns, topology configurations or MAC protocols. The routing protocol uses the four essential fields that exist in any packet format: originator address (ORG) which specifies the address of the node that initiated the packet; final destination address (FIN) which specifies the address of the last receiver; source address (SRC) which specifies the address of the neighbour that forwarded the packet; and finally Time-To-Live (TTL) which is a counter of the number of hops that are traversed by a packet during its journey from the originator to the current receiver. The proposed protocol employs the hop-by-hop routing strategy which enables the originator as well as intermediate relaying nodes to select the next hop upon receiving a packet. The benefits of this strategy is twofold, firstly, it increases the robustness of the routing protocol by allowing an intermediate relaying node to select the next hop based on the current conditions of the network and secondly, it reduces the overhead associated with specifying the complete path in the header of packets. The pseudo-code of the routing protocol is given in figure 8.3 and explained in the following subsections.

Startup ( )
  Create MAP
End Startup ( )

Receive packet ( )
If packet is acknowledgment

```
  ACK(SRC)++
  LOST(SRC) ← 1 − ACK(SRC)/TPKT(SRC)
else
  if (ORG , SRC, TTL) is unknown
     Add MAP (ORG, SRC, TTL)
  RPKT (SRC) ++
  UTIZ (SRC) ← RPKT(SRC)/Time counter
  DELY(ORG , SRC, TTL) ← IRTP(SRC)/RPKT(SRC)
  VART(ORG , SRC, TTL) ← COV[IRTP(SRC)]
if (FINS = = my address) and (there is a better path or SBOPT is set)
  send NTFC packet to the originator
else
  Forward packet ( )
end Receive Packet ( )

Forward packet ( )
Π = all neighbours that  can reach FIN
  nexthop ← min( (LOST(x)+UTIZ(x)+DELY(x)+VART(x)+DSTC(x)+DENS(x)) / (Σ_{x∈Π} LOST(x)+UTIZ(x)+DELY(x)+VART(x)+DSTC(x)+DENS(x)) )
If Π is empty then and the minimal connected dominating set is empty
  Nexthop = max(PO(*,*))
else
  Nexthop = neighbour within minimal connected dominating set
end if
set SBOPT
end Forward packet ( )

Goal Oriented search ( )
If there is a disconnected region
  Finaldestination= max(DR(*,*))
else if minimum connected dominating set is not found
  Finaldestination= max(PPI(*,*))
else if fluctuation of attributes <=30%
  Finaldestination= final destination of the region with the maximum fluctuation
end if
Send SCHR(Finaldestination)
end Goal Oriented search ( )
```

Figure 8.3 Pseudo-code of the proposed routing protocol

8-7-1 Startup function

This function is executed once during the lifetime of the node with the aim to create the MAP data structure which is used to maintain information about the neighbouring nodes and their attributes. The proposed protocol uses the following key attributes: final destination list (FISD) which stores the final destinations known by a neighbour and their TTL values; rate of lost packets (LOST) which estimates the probability with which a neighbour loses its packets; utilisation factor (UTIZ) which approximates the rate with which a neighbour uses its resources; expected delay (DELY) and (VART) which estimates the mean and coefficient of variation of the inter-arrival times of packets. In additional, the proposed protocol uses the following variables to store information about neighbours: number of acknowledgment frames received from a neighbour (ACK), aggregated inter-arrival times between packets that are received from a neighbour (IRTP) and finally number of the packets that are received and transmitted to a neighbour (TPKT) and (RPKT) respectively. The initial values for all of these attributes are set to zero and updated whenever a node receives a new packet or conducting a search.

8-7-2 Receive packet function

This function is executed whenever a node receives a packet with the aim to update the attributes of the neighbour that has sent the packet and to notify the originator if it has routed its packet over suboptimal path. A node upon receiving a packet checks its type, if the received packet is acknowledgment for a data packet that has been sent by the receiver then the number of acknowledgment frames of this neighbour ACK(SRC) is incremented by one and then divided by the total number of packets that transmitted to this neighbour TPKT(SRC). The resultant is then subtracted from one to obtain the LOST attribute of the neighbour SRC, hence this attribute estimates the throughput of the link between the receiver and SRC. Otherwise, if the received packet is of type data and the values of the ORG, SRC and TTL fields specified in the received packet are not existed in the MAP of the receiver node, then this node appends them to its MAP. Thereafter, a node updates the UTIZ attribute of SRC by incrementing the number of packets received from this neighbour (i.e., RPKT(SRC)) by one and divides the results by the time counter. Hence the UTIZ attribute approximates the rate with which SRC uses its resources, e.g., battery as well as the resources shared between SRC and the receiver, e.g., channel. The expected delay attributes, i.e., DELY is updated by dividing the aggregated inter-arrival times of packets received from the neighbour SRC i.e., IRTP(SRC) by the total number of packets received from this neighbour RPKT(SRC). Finally, the variation attribute VART is updated by computing the coefficient of variation of the inter-arrival times of packets received from the SRC. It is of interest to note that the DELY and VART attributes measure the end-to-end traffic characteristics from the originator to the receiver node including the traffic load, the transmission time as well as the queuing delay, hence these attributes are identified by the complete path, i.e., ORG, SRC,TTL. The main advantage of this approach is that it improves the routing decision by allowing a node to estimate the minimal delay that can encounter a packet over this path.

It is worth noting that the main advantage of employing the inference to collect the attributes from the received packet is twofold. Firstly it allows a node to route its packets based on the current status of its neighbours and secondly it saves resources that would by consumed in sending probing packets periodically to collect the routing metrics. Besides these advantages, the receive packet function provides an error correction mechanism that enables the final destination of the received packet to enquire its MAP for the optimal path towards the originator. If such path is found, then a node sends a notification packet (NTFC) towards the final destination. The notification packet contains the FISD attribute of all neighbouring nodes of the receiver and is sent to the neighbour that has forwarded the data packet (i.e., SRC). Each node receives NTFC updates its MAP and the contents of the NTFC packet and then relays it towards the final destination of the data packet. Moreover, the proposed protocol allows the final destination to send the NTFC packets on demand, this occurs when the originator or an intermediate relaying node sets the suboptimal (SBOPT) flag. Otherwise if the receiver node is an intermediate relaying node then it calls the forward packet function.

8-7-3 Forward packet function

This function is executed whenever a node receives a packet and has to find the optimal route towards the final destination (FIN). The receiver node consults its MAP to obtain all neighbours that can reach FIN; this is performed by checking the attribute (FISD) of all neighbours existing in the MAP data structure. If more than a neighbour is found, then these neighbours and their attributes are stored in a temporary data structure called Π.

Afterwards, a node selects the next hop that has the minimally aggregated attributes. The key benefit of using the minimal operator instead of using a fixed load balancing factor (as the case with other routing protocols) is that using the minimal operator enables a node to cope with dynamic characteristics of the network adequately. The forward packet function uses the four stochastic attributes discussed in the previous section (i.e., LOST, UTIZ, DELY and VART) in additional to two spatial attributes: DSTC(x) and DENS(x). The DSTC(x) attribute counts the number of disconnected regions reached by neighbour x. Selecting the next hop with minimal disconnected relationships has the advantage that it enables a node to maintain a global connectivity by avoiding overburdening the neighbours that can reach distinct final destinations. The DENS(x) attribute sums the number of final destinations that can be reached by the neighbour x within TTL values less than or equal to the TTL values with which the neighbour x reaches the intended final destination (i.e., FIN). Hence this attribute estimates the dense/spare characteristics of the regions towards the final destination. The importance of incorporating the DENS attribute in selecting the next hop can be acquired by considering the fact that routing packets over dense regions increases the amount of energy and resources waste in overhearing packets by the neighbours of the intermediate relaying nodes.

On the other hand if the final destination of the received packet is not established in the MAP, (i.e., Π is empty) then the receiver node routes the packet to the neighbour that has the maximum overlapped regions unless a node finds the minimal connected dominating set. In both cases, a node sets the “SBOPT” flag to indicate that this packet has been forwarded over a suboptimal path, hence when the final destination receives a packet sends a NTFC packet.

8-7-3 Goal oriented search function

The main aim of the goal oriented search function is to direct the diversification and intensification processes of a node in order to satisfy certain goals, e.g., finding alternative paths to disconnected regions or finding the minimal connected dominating set. The underlying approach of this function is to utilise the RCC5 calculus and the defined attributes to enable a node to recognise the spatial relationships between nodes as well as their stochastic characteristics.

Finding alternative paths to those final destinations reached by the minimal number of neighbours is one of the key goals that the goal oriented search function works towards. A node achieves this goal be identifying the final destination that has the maximum disconnected relationships. Thereafter, a node sends a SCHR (i.e., search request packet) containing the FISD lists of its MAP to the intended final destination. The procedure used to route this packet is identical to the procedure used to route the NTFC packet, i.e., each node upon reception of SCHR updates its MAP as well as the content of this packet and relays it. When the SCHR packet is received by the final destination, it compares its MAP with the contents of the SCHR packet in order to find common region(s) between them. If such region(s) is found, then a notification packet (i.e., NTFC) is sent to the originator of the SCHR packet. Otherwise if common region(s) not found, then the final destination of the SCHR packet realises that the region specified in the SCHR packet is disconnected from its known regions and thereby it conducts searching from its perspective. This final destination selects the largest region that is disconnected from the region specified in the received SCHR packet and sends to the final destination of this region a SCHR packet. This process continuous until a common region(s) is found and the originator of SCHR is received a NTFC packet. Although it is well known that searching over a graph can degrade performance of the network due to its high computational complexity and communication overheads, the proposed algorithm does not suffer from these limitations. A justification for this argument is that the proposed algorithm is based on a decentralised approach that unifies searching endeavours carried out by different nodes from different perspective which in turn leads to significant reduction in the computational complexity of finding a common region(s). A further support for our argument can be acquired by considering that the proposed searching procedures are carried out by the most distant nodes of the largest disconnected regions. This in turn fastens approaching to the common region(s) and more importantly reduces the communication overheads [239].

The second main goal defined by the search oriented goal is to find the minimal connected dominating set of the network, i.e., the smallest cardinality subset of connected nodes such that a node within a network is either a member or a neighbour of a member of this subset. Finding the minimal connected dominating set is carried out using the procedure used to find common region(s) of disconnected final destinations described in the previous paragraph except that a node uses the not proper part relationship (i.e., PPI) instead of disconnected relationship. A justification for using not proper part relationship is that the region of final destination with the maximum PPI relationships has a broad view for the network and hence it is preassembly within the virtual backbone of the network. The last main goal of the goal oriented search function is to anticipate the traffic congestion or topology changes of the network and updates its MAP data structure accordingly. This goal is achieved by monitoring the fluctuation in the values of the attributes stored in the MAP data structure. If a node notices significant fluctuation of the values of attributes, then this node sends a SCHR packet to the final destination of the region with maximum fluctuation. Thereafter, when the NFTC is received a node can determine whether there is a need to conduct a search or not.

8-8 Results and discussion

This section assesses the integrity of the proposed routing protocol from different perspectives. It starts by comparing its performance with the IEEE 802.15.5 [112] and ANT routing protocols [236] in section 8-8-1. These two protocols have been selected as they represent two different routing approaches. The IEEE 802.1.5.5 protocol can be considered as an example of classical based protocol while the ANT protocol is the one of the most cited routing protocols that is inspired by the swarm intelligent approach. This assessment is carried out under the standard IEEE 802.15.4 CSMA-CA protocol. In section 8-8-2 the benefits of the proposed routing protocol are assessed under the Gamma based CSMA-CA protocol that was proposed in chapter 6 and the duty cycle management scheme that was proposed in chapter 7. Like the previous chapters, the results of this section are labelled using the five tuples introduced in section 4-4, i.e., $\langle k, N, X, Y, Z \rangle$ where k denotes the connectivity degree (the average number of mutually independent paths between non-neighbouring nodes). The parameter N denotes the average number of nodes per channel. The parameter X refers to the probability distributions that are used to generate the inter-arrival times between packets by nodes, this chapter uses three different distributions: Constant Bit Rate (CBR) , Exponential (EXP) and Weibull (WBL). The rate with which a node generates its packets is defined in the parameter Y and given in packet per slot. Finally the parameter Z specifies the traffic routing policy which can take two possible values: U when traffic that is generated internally by a node is distributed evenly to other nodes with the network and A to indicate that traffic generated by all nodes within the network is destined to a single final destination. Other simulation parameters are set similarly to section 5-4.

8-8-1 Performance of the proposed protocol versus IEEE 802.15.5 and ANT

This subsection assesses the performance of the proposed protocol by comparing its results with the outcomes of other two routing protocols: namely IEEE 802.15.5 [112] and ANT [263] under the IEEE 802.15.4 CSMA-CA standard [16]. The results of this comparison are illustrated in figures 8.4-8.6.

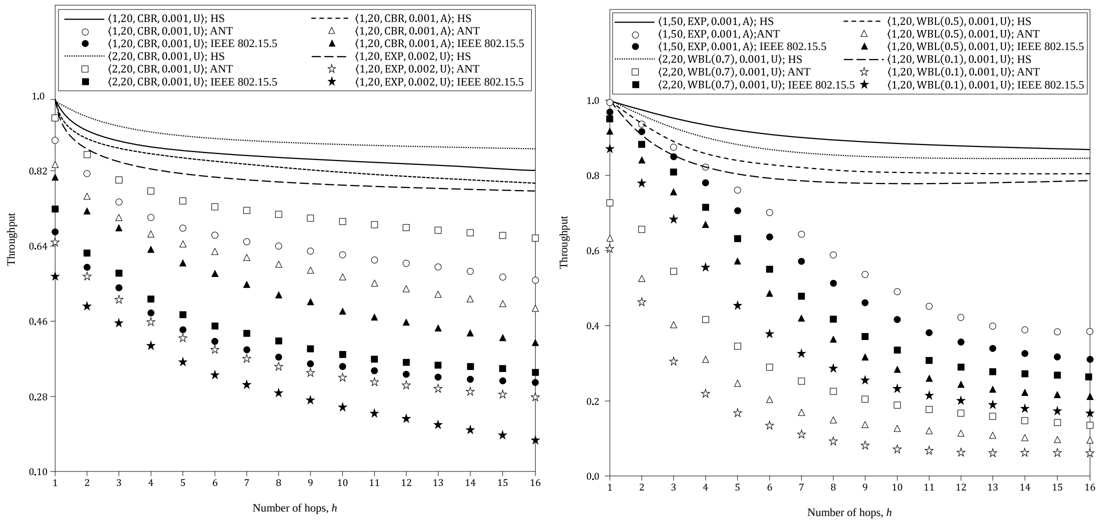


Figure 8.4 Throughput of HS, ANT and IEEE 802.15.5

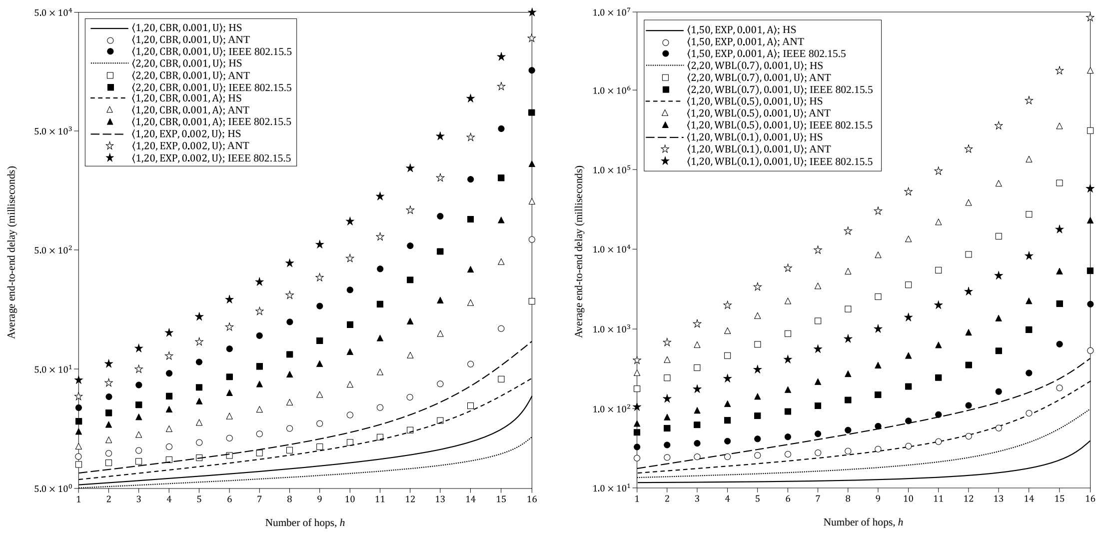


Figure 8.5 Average end-to-end delay of HS, ANT and IEEE 802.15.5

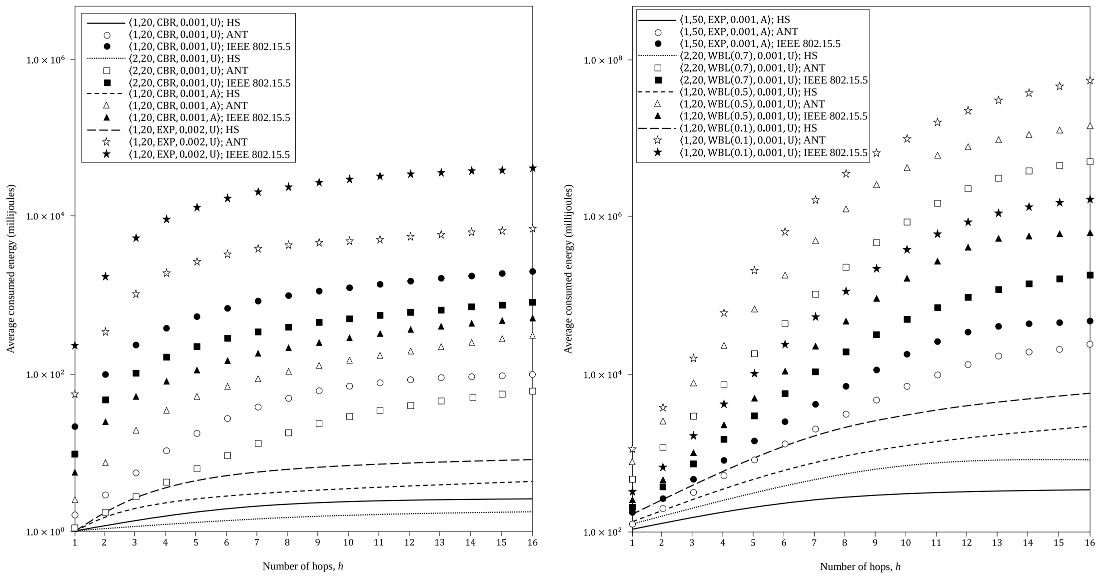


Figure 8.6 Average energy consumption of HS, ANT and IEEE 802.15.5

Figure 8.4 shows that when there is on average a single path between a pair of non-neighbouring nodes, the average number of nodes per a channel is small, $N = 20$, traffic is generated according to a Constant Bit Rate (CBR) with low traffic rate (i.e. 0.001 packet per a slot) and all packets are destined to a single final destination, i.e., the case $\langle 1{,}20, \mathrm{CBR}, 10^{-3}, \mathrm{A}\rangle$. It can be seen that under these conditions, the throughput of both IEEE 802.15.5 and ANT over a small number of hops are about the same. This characteristic is attributed to the fact that the overhead imposed by the IEEE 802.15.5 protocol in exchanging routing tables roughly equals the overhead of the artificial ants required by the ANT protocol. Hence the collision probabilities resulted from these overheads are about the same which in turn yields approximate throughputs. However, increasing the number of hops leads to significant reduction in the throughput of the IEEE 802.15.5 protocol compared to the ANT protocol. This characteristic is attributed to that increasing the number of hops increases the size of routing table and thereby magnifies the sizes of control packets exchanged between nodes, this in turn reduces the channel utilisation and increases the collision probability between the control packets and data packets. The high throughput of the proposed protocol is attributed to the fact that the proposed protocol enables nodes to infer the routing metrics from the network without imposing communication overheads on the network.

Comparing the readings of the average end-to-end delay of the IEEE 802.15.5, HS and ANT protocols of the scenario $\langle 1{,}20, \mathrm{CBR}, 0.001, \mathrm{A}\rangle$ as illustrated in figure 8.5 shows that these readings increase proportionally with number of hops and these increments are always higher in the IEEE 802.15.5 protocol. This characteristic is attributed to the fact that IEEE 802.15.5 uses the mesh links to route packets that are within its second hop and uses the tree links to route packets that are far than that, hence increasing the number of hops increases the traffic over the tree links which leads to an increase in the average end-to-end delay of packets. On the other hand, the reduction of the average end-to-end delay offered by ANT compared to IEEE 802.15.5 is due mainly to that ANT employs a distributed mechanism that enables a node to route its packets over the optimal paths without distinguishing between mesh and tree links. This avoids traffic congestion over certain links and thereby reduces the end-to-end delay of packets. Comparing the characteristics of the proposed protocol with ANT and IEEE 802.15.5 shows that HS can reduce the average end-to-end delay substantially. This is attributed to the fact that HS enables nodes to select the optimal routes based on the attributes that represent the current status of the network; while in ANT, nodes have to wait for the backward artificial ants to make a sound routing decision.

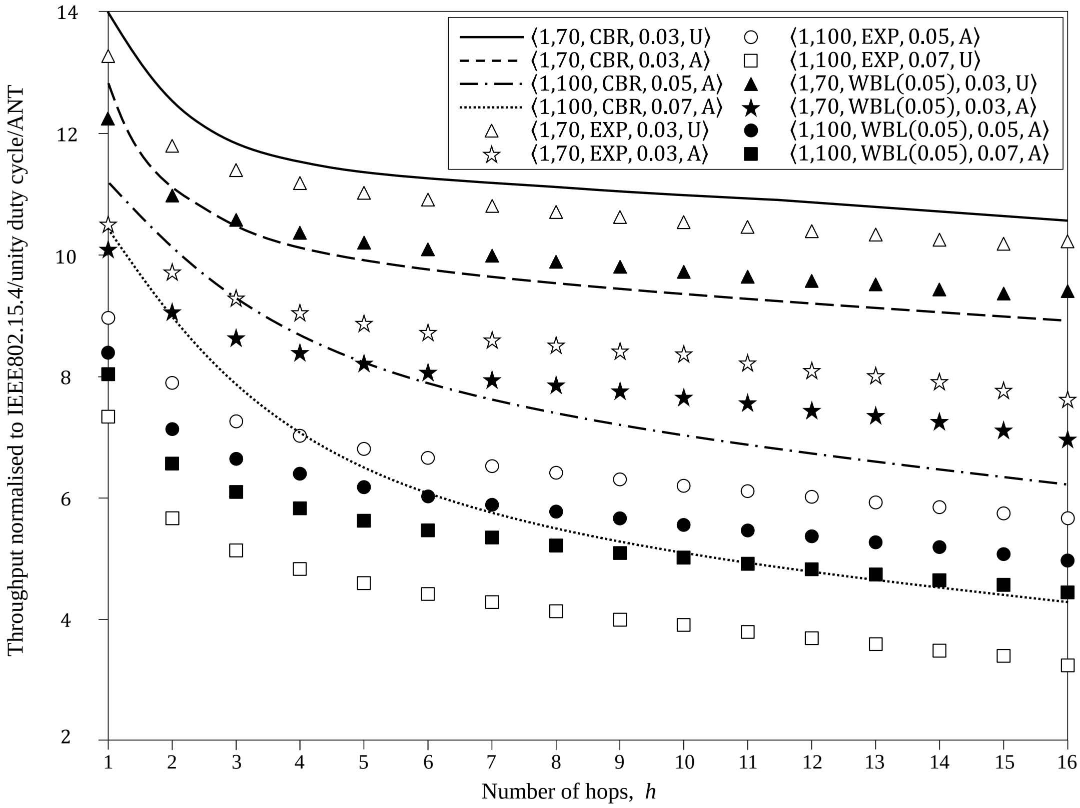
Results of figure 8.6 show that under the same network configuration $\langle 1{,}20, \mathrm{CBR}, 0.001, \mathrm{A}\rangle$, the proposed protocol (HS) provides the minimal energy consumption compared to other protocols. This characteristic is due mainly to that the HS protocol uses the DENS attribute to select the optimal paths which enables a node to avoid routing its packets over highly dense regions. This in turn mitigates the overhearing energy along the route from originators to final destinations and hence reduces the energy consumption of nodes. Another main reason for the reduction in the average energy provided by the HS protocol is that this protocol prevents overburdening of the intermediate relaying nodes by selecting the next hop that has the minimal UTIZ attribute. Comparing this characteristic with the behaviour of the IEEE 802.15.5 protocol shows that the high energy consumption of IEEE 802.15.5 is due to the need to reform the tree links several times. In particular, nodes within tree topology are exhausted due to the extreme traffic that they are received, hence the batteries of these nodes are depleted faster than other nodes which in turn requires exchanging control packets to form new tree links. It can also be seen from the results of this figure that the ANT protocol provides moderate energy consumption which exhibits the advantage of using decentralised means to gather routing information.

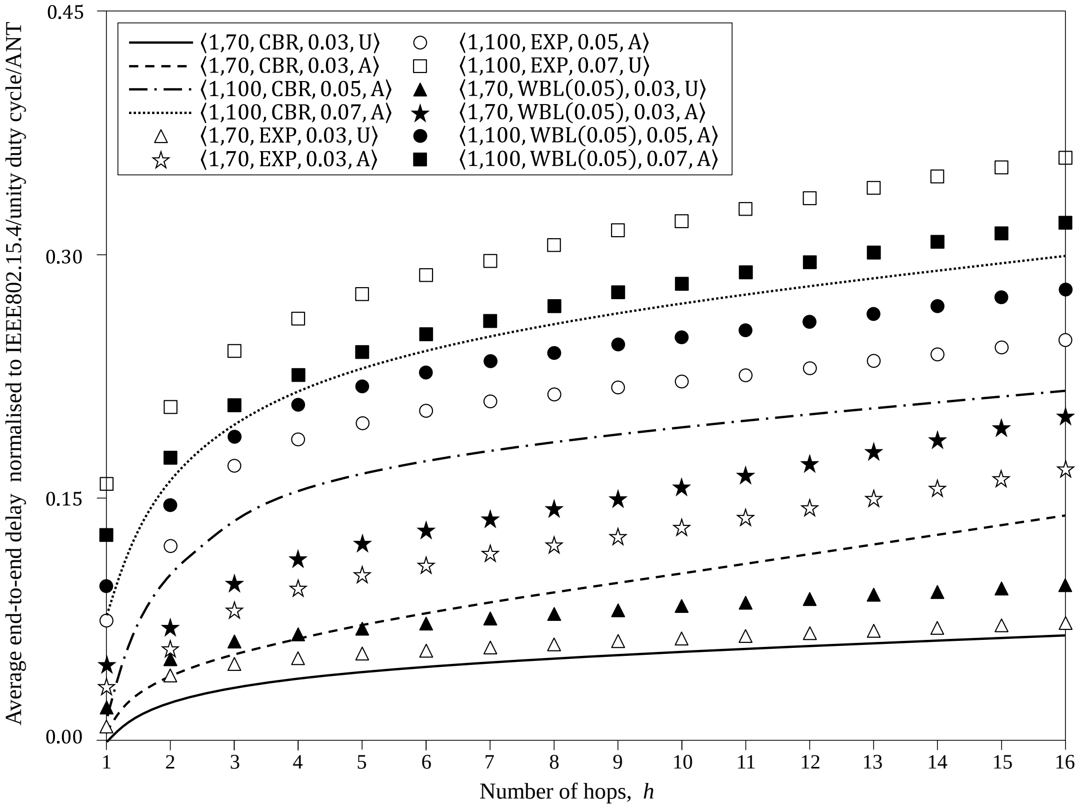
The results of figures 8.4-8.6 show that doubling the connectivity degree of network improves the performance of the network operating the HS protocol notably compared to other protocols. This characteristic is attributed to that HS is designed to route packets over multipath, hence providing a network with high connectivity degree enables nodes to distribute their loads between different paths effectively which in turn reduces average end-to-end delay of packets, minimises consumed energy and leverage throughput. Although the ANT protocol can also cater for multipath routing, the outperformance of the HS protocol is due to that the differences between the diversification processes of these protocols. Recall that in HS nodes are required to conduct searching to recognise the underlying structure of the network, while the routing metrics are inferred from the network. Hence employing the HS over connected networks reduces the volume of control packets which leads to leverage the performance readings. Conversely, in ANT the control packets (i.e., forward and backward artificial ants) are used to harvest the routing metrics as well to discover the network topology which in turn increases the overheads associated with control packets. The results of these figures shows that doubling of the connectivity degree of network has no considerable effects on the performance readings of the IEEE 802.15.5 protocol, which is attributed to that the packet forwarding mechanism of the IEEE 802.15.5 protocol does not recognise the multi-paths links, – i.e., all packets are still routed over either mesh or tree links. Comparing the outcomes of cases $\langle 1{,}20, \mathrm{CBR}, 0.001, \mathrm{A}\rangle$ and $\langle 1{,}20, \mathrm{CBR}, 0.001, \mathrm{U}\rangle$ shows that changing the traffic distribution policy from A (when all packets are destined to a sole final destination) to U (when packets that are generated by a node are distributed evenly to all other node within the network) degrades the performance of the network operating the IEEE 802.15.5 protocol significantly while improves the performance of network operating the HS and ANT protocols. This characteristics is attributed to that distribute traffic over all nodes requires a node to maintain paths to each node within the network which increases the sizes and intensity of control packets used by the IEEE 802.15.5 protocol. This increasing in the overheads leads to increase the collision probability, channel occupancy, end-to-end delay and energy consumption. The improvement of both ANT and HS is attributed to that distributing traffic over multiple nodes enhances the routing redundancy which in turn leverages the performance of the network. Moreover, the outperformance of HS compared to ANT is emerged from the fact that HS can approach the optimal path faster than ANT. The explanation for this is that the HS conceptualises the network as a collection of regions that are spatially connected and then enables a node to search over regions rather than to visit each node as the case in ANT.

The effects of using different distributions to generate inter-arrival times between packets on the performance metrics of the IEEE 802.15.5, ANT and HS protocols are illustrated in figure 8.4-8.6. It can be seen that employing heavy tailed distribution (i.e., WBL) to generate inter-arrival times between packets degrades the performance of ANT drastically compared to other protocols. It can also be seen that this degradation increases inversely with the value of the shape parameter of the WBL distribution. An explanation for these characteristics can be acquired by considering the fact that the ANT protocol uses the pheromone concentrations to select the optimal paths towards the final destination. The values of these concentrations are increased proportionally with number of artificial ants and data packets that are passed over paths. Conversely, the values of pheromone concentrations are decreased (evaporated) over the course of the time according to an exponential function [263, 303] and when the value of the pheromone concentration corresponding to specific path reaches zero, a node purges its routing table from this path. In low variation inter-arrival distributions (as the case in CBR and EXP) the evaporation mechanism of ANT performs well as it enables a node to maintain the optimal or near optimal paths to other destinations almost permanently. However, using WBL increases the variation between packets which in turn increases the possibility of being a node has no or suboptimal paths to specific destinations. This in turn degrades the performance of ANT under heavy tailed distributions. Moreover, considering the fact that the variation of WBL distributions increases inversely with its shape parameters justifies the reduction in the performance metrics of ANT when the shape parameter of WBL distribution is reduced. The characteristics of the IEEE 802.15.5 protocol under WBL distribution demonstrate the fact that this protocol requires nodes to exchange their routing table periodically which in turn makes the routing information less sensitive to the variation of inter-arrival of data packets. Finally, the outperformance of the proposed protocol under different inter-arrival distributions is attributed to that this protocol use the variation attribute to select the optimal route, hence a node can even account for the temporal variations of the inter-arrival distribution which in turn leverages the performance readings of the network.

8-8-2 Benefits of the proposed protocol under the Gamma based CSMA-CA protocol and artificial chemistry cycle management scheme

This section assesses the performance of the proposed routing protocol (HS) under Gamma based back-off scheme that was proposed in chapter 6 and the duty cycle management scheme that was proposed in chapter 7. The CSMA-CA mechanism that was proposed in chapter 6 replaced the uniform distribution of the IEEE 802.15.4 CSMA-CA protocol with Gamma distribution and replaced the blind doubling of back-off windows by more intelligent collision resolution algorithm. Chapter 7 employs the principles of Artificial Chemistry (AC) to propose a novel duty cycle management scheme for networks based on the non-parametric inferential approaches to generate the waking and sleeping durations. This section considers extremely operational conditions and depicts the results normalised to the cases (ANT/unity duty cycle/IEEE 802.15.4), i.e., when the IEEE 802.15.4 CSMA-CA protocol is used as MAC protocol with the unity duty cycle management scheme and the ANT routing protocol. The results of this assessment are given in figure 8.7.


⟨1,70, CBR, 0.03, U⟩
⟨1,70, CBR, 0.03, A⟩
⟨1,100, CBR, 0.05, A⟩
⟨1,100, CBR, 0.07, A⟩
⟨1,70, EXP, 0.03, U⟩
⟨1,70, EXP, 0.03, A⟩
⟨1,100, EXP, 0.05, A⟩
⟨1,100, EXP, 0.07, U⟩
⟨1,70, WBL(0.05), 0.03, U⟩
⟨1,70, WBL(0.05), 0.03, A⟩
⟨1,100, WBL(0.05), 0.05, A⟩
⟨1,100, WBL(0.05), 0.07, A⟩
Throughput normalised to IEEE802.15.4/unity duty cycle/ANT
Number of hops, h
2
4
6
8
10
12
14
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16


⟨1,70, CBR, 0.03, U⟩
⟨1,70, CBR, 0.03, A⟩
⟨1,100, CBR, 0.05, A⟩
⟨1,100, CBR, 0.07, A⟩
⟨1,70, EXP, 0.03, U⟩
⟨1,70, EXP, 0.03, A⟩
⟨1,100, EXP, 0.05, A⟩
⟨1,100, EXP, 0.07, U⟩
⟨1,70, WBL(0.05), 0.03, U⟩
⟨1,70, WBL(0.05), 0.03, A⟩
⟨1,100, WBL(0.05), 0.05, A⟩
⟨1,100, WBL(0.05), 0.07, A⟩
Average end-to-end delay normalised to IEEE802.15.4/unity duty cycle/ANT
Number of hops, h
0.00
0.15
0.30
0.45
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

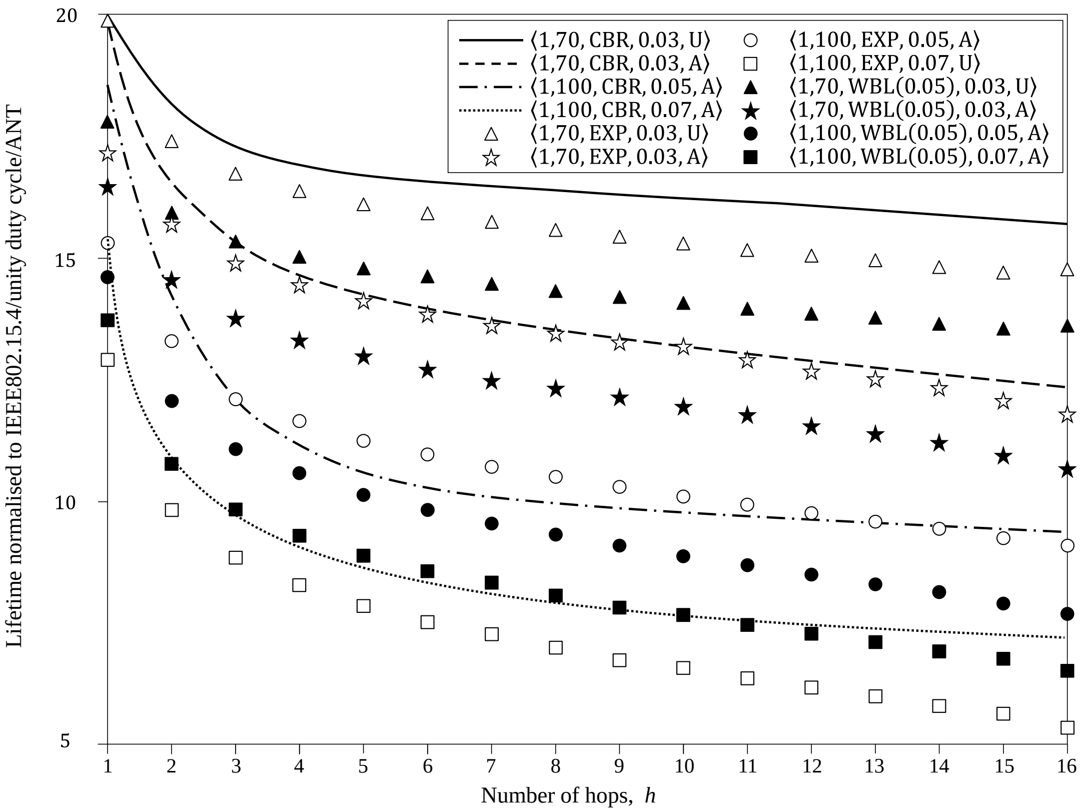


Figure 8.7 Throughput, average end-to-end delay and lifetime of HS/AC/Gamma normalised to ANT/unity duty cycle/IEEE 802.15.4

It can be seen from results of figure 8.7 that incorporation of the Gamma based CSMA-CA protocol and AC based duty cycle management scheme with the proposed routing protocol always improves the performance of networks. It can be seen that under the most extremely operational conditions, $\langle 1{,}100, \mathrm{WBL}(0.05), 0.07, \mathrm{A}\rangle$ that throughput is magnified by about four-fold, average end-to-end delay is reduced by about three times and lifetime is prolonged by about four times compared to case when the IEEE 802.15.4 CSMA-CA is used with unity duty cycle and ANT. These characteristics demonstrate the ability of all proposed protocols to be integrated seamlessly and provide significant enhancements for the performance of network. Indeed the Gamma based CSMA-CA protocol reduces the time required to service successful packets which in turn increases the channel utilisation and enables a node to save the energy consumed in contention to access the channel. The AC duty cycle management scheme mitigates the energy consumed in overhearing and idle listening which in turn prolongs the lifetime of nodes without imposing overheads on the network. This enables nodes to communicate for longer periods and over low-utilised channel which in turn leverages the throughput of network and reduces the end-to-end delay of packet. Finally, the HS routing protocol proposed in this chapter enhances the performance of network by allowing nodes to route their packets over the routes that yield global optimisation for the entire network. It can also be seen from the results of this figure that the improvements provided by the proposed protocols are not restricted to specific topology configurations, traffic patterns or routing policies. This exhibits the advantages of using the inferential statistics approach to designing a communication protocols for WSNs.

8-9 Conclusion

This chapter has proposed the final main contribution of this thesis which is to design an effective routing protocol for WSNs that can achieve three concurrent goals: prolonging the life time of nodes, minimising the end-to-end delay of packets, and increasing the throughput of networks. The main objective of the proposed protocol is to enable nodes to route their packets over those paths that maintain the global optimisation for the entire network without particular considerations to specific topologies, MAC protocols or traffic patterns. The proposed protocol is inspired by the harmony search optimisation algorithm that mimics the behaviour of a musical ensemble when they are endeavouring to improvise the most pleasing harmony. Moreover, the essential functions of the proposed protocol is to enable each node to harvest the routing metrics from the traffic, analyse them statistically and then infer the most promising route without a need to overwhelm the network with control packets. The results and discussion of this chapter demonstrate benefits of the proposed protocol by comparing its outcomes with two state-of-the-art routing protocols: IEEE 802.15.5 and ANT under different scenarios. Moreover, these assessments are carried out over other algorithms that were proposed in this thesis, i.e., the Gamma based back-off scheme that was proposed in chapter 6 and the duty cycle management scheme that was proposed in chapter 7.

Chapter 9

Future Work

9-1 Overview

The previous chapters have employed the principles of inferential statistics to design effective Medium Access Control (MAC), routing and duty cycle management strategies for Wireless Sensor Networks (WSNs). These strategies considered general WSN scenarios without restrictions on topology configurations, routing policies or traffic patterns. Moreover the previous chapters endeavoured to achieve three concurrent goals: increase the network throughput, reduce the end-to-end delay of packets and to prolong the lifetime of nodes. Comprehensive simulated assessments in conjunction with developed analytical models demonstrated the ability of the proposed protocols to achieve the stated goals.

During the development of the work presented in the previous chapters, a number of potential future research directions have emerged. This chapter discusses the most promising research areas and highlights the potential benefits that can be obtained from conducting further research in these areas. The future research directions are grouped into three main categories. Firstly, future work related to the analytical framework proposed in chapter 5. This work presents the modifications that are required to extend the proposed analytical framework to the cases not covered in chapter 5 such as non-ideal channel models or non-unacknowledged retransmission modes. The second category of the potential future work discusses the modern concepts and approaches related to the inferential statistics such as meta-reasoning and anticipatory computing. The underlying approaches of these modern fields and their possible applications in the context of designing WSNs communication protocols are provided. The final category of the future work discusses how the metaphors presented in this thesis can be utilised further to solve other problems related to WSNs.

The remainder of this chapter is organised as follows. Section 9-2 outlines the future work related to the proposed analytical modelling. Section 9-3 highlights the future work related to the inferential statistics and proposed protocols and finally section 9-4 considers the future work related to the multidisciplinary metaphors.

9-2 Future work related to the proposed analytical framework

The analytical framework presented in chapter 5 has been developed to analyse the statistical characteristics of the Carrier Sensing Multiple Access with Collision Avoidance (CSMA-CA) protocols. It has been shown that this framework is able to provide deep insight into the fundamental operations of this sort of protocol under different traffic patterns, routing policies and topology configurations. However, this framework made some simplifying assumptions in terms of the channel models and retransmission modes. Hence it would be useful to extend the proposed framework to cover these cases. This section considers the main benefits of such extensions and their corresponding modifications.

One of the important areas for future work is to extend the proposed framework to non-ideal channel conditions. The proposed framework use the assumption that the network is deployed over an ideal channel in which communications between neighbouring nodes are dominated exclusively by their Euclidean distances with respect to a predefined range. This assumption has been made in order to derive general results that are not restricted to specific channel models. However, in real deployment scenarios, additional factors (e.g., presence of obstructions and reflections of the signals) [330] have to be taken into consideration when determining the quality of received signals. Incorporating non-ideal channel models into the proposed analytical framework can widen its applicability and enhance the resemblance of its outcomes compared to real deployments. Although there are an enormous number of channel models reported in the literature, the main challenge is to find the appropriate model that can reflect the characteristics of real deployments. It is worth noting that the proposed framework is able to account for different channel models without requiring substantial modifications. Recalling that the proposed framework uses set theory to represent the channel model and its impact on the ability of a node to decode a correct signal. In the case of an ideal channel, these sets are populated according to the Euclidean distances between nodes as illustrated in section 5-4-1. Hence accounting for different channel models can be accomplished by populating the sets according to the criteria of the selected channel model and without the need to reconstruct the entire framework.

Another direction for future research involves extension of the proposed framework to other retransmission modes. The proposed framework considers the unacknowledged retransmission mode in which a receiver node does not transmit an acknowledgment frame to confirm successful reception. This policy has been widely used in WSNs in order to save energy consumed in transmission and reception of acknowledgment frames. However, some WSNs applications require confirmation for successful receptions of packets. Hence incorporating retransmission mode in the proposed analytical framework can enhance its capability of predicting the performance metrics under different operational conditions. The modular approach upon which the proposed analytical framework is built facilitates adapting different retransmission modes readily. From the mathematical perspective, the proposed analytical framework derives the probability distribution of the time spent in a single CSMA-CA cycle and then uses the theorem of compound probability distributions to construct the successful service time distribution. Hence the successful service time distribution is defined as a compound probability distribution whose primary component is the probability distribution of repetitive generation and whose secondary component is the distribution of the time spent in a single CSMA-CA cycle. In the case of the unacknowledged retransmission mode (considered in chapter 5) the geometric distribution is used as the primary component of the compound probability distributions. In the case of acknowledged retransmission mode this geometric distribution can be replaced with a truncated geometric distribution whose cut-off limit is the maximum number of retransmissions. A justification for this selection is that in the acknowledged retransmission mode a node drops a packet after transmitting it for a predefined number of times without receiving an acknowledgement frame; hence the repetitive generation of the CSMA-CA distribution is limited by this parameter.

9-3 Future work related to the inferential statistics and proposed protocols

This thesis has utilised the principle of inferential statistics to design effective MAC, routing and duty cycle management schemes for WSNs. It has been shown that the use of inferential statistics provides a rich design space with powerful tools and techniques that can leverage the performance of communication protocols without overburdening the scarce resources of WSNs. Besides these advantages, inferential statistics and its related fields are suitable for the envisaged applications of WSNs. In such applications sensor nodes act on the behaviour of human beings to adapt the environments in such a way that make them more informative and convenient. These envisaged applications need a powerful communication protocols that can interpret the high cognitive levels of humans and take the most appropriate actions. Towards this goal a number of new research disciplines have been introduced; this section reviews some of these disciplines, highlights their underlying approaches and explores their applicability in the context of designing WSN communication protocols.

One of the most promising research fields that can be used to design communication protocols for WSNs is meta-reasoning. The term meta-reasoning or “reasoning about reasoning” has been coined in 1979 by Flavell during his investigation for the mechanism that is used by human beings to develop a creative solution for the problem under consideration [331]. The study conducted by Flavell concluded that one of the most powerful activities that empowers this development is self-reflection, which is the ability of humans to monitor and control their reasoning activities. A decade after work of Flavell, Nelson introduced the meta-reasoning model [331]. The model not only enhances our understanding for meta-reasoning but also facilitates using this approach to design algorithms for artificial systems. Fundamentally, Nelson’s model represents creative solving mechanism of a human as a two layer model: a meta-reasoning layer and an object layer. The meta-reasoning layer regulates and controls reasoning activities while the object layer specifies the solution strategies. The meta-reasoning layer imitates the self-reflection mechanism of humans; this layer guides a human to select amongst different strategies in order to solve the problem under consideration. The meta-reasoning layer has a global view of the current situation of the problem and available resources as well as the efficiency of different strategies defined in the object layer. The interfaces between the meta-reasoning layer and object layer are carried out through two main functions: monitor and control. The monitor function is used to audit performance of the executed strategy while the control function is used to instruct the object layer, e.g., to adapt the current strategy more effectively or to switch to better strategy. Based on the outcomes of these two functions the meta-reasoning layers can direct the problem solver to obtain further information or to devise a new solution strategy. One of the main techniques used by the meta-reasoning layer to enhance its ability to take decision is inferential statistics. Basically, the meta-reasoning layer uses the outcomes of each strategy to infer the unseen information and then determine whether the current strategy is adequate to solve the problem at hand or if there is a need to find an alternative way.

The meta-reasoning approach features several powerful aspects that can enhance the design space for WSN communication protocols. The most powerful aspect of meta-reasoning is that it aims to provide a high-level process that is not restricted to specific operational conditions. Therefore, developing protocols according to the principles of meta-reasoning can widen their applicability and enhance their adaptiveness. The second main strong aspect of the meta-reasoning approach is that it facilitates transferring information between different solution strategies which is turn enables a sensor node to look at the problem from different perspectives. Hence using the meta-reasoning to design WSN communication protocols can leverage their ability to find the optimal solution within a reasonable time and resources. However, the main challenge of the using meta-reasoning to design WSN communication protocols is the requirement to consider the scarce resources of these networks.

Anticipatory computing [332] is another research discipline that can be utilised to design effective communication protocols for WSNs. Anticipatory computing is the research discipline that aims to build an artificial system that are able to control their present behaviours based on their predictions for the futures states. This research discipline is inspired by the cognitive capabilities of humans, and the relations between anticipation and human’s intelligence have been investigated in number of studies e.g., [213,229,258,332]. The main conclusion of these studies is that the capability of humans to devise intelligent solutions lies mainly on their ability to represent a problem mentally beforehand and to imagine the future consequences of potential solutions. This conclusion constituted a novel perspective upon which the anticipatory computing is developed. The models represented in [322] showed that an anticipatory system can be realised if each agent within such system is fed with a model that can pace more rapidly than the actual system. Hence these agents can examine the effects of their possible actions before practicing them in the actual system. Towards this goal, a number of anticipatory algorithms such Bayesian estimation and anticipatory classifier systems [332-333] have been introduced. These approaches employ the principles of the inferential statistics to provide lightweight and accurate anticipatory methods. The obvious benefit of employing an anticipatory computing approach to design WSN communication protocols is that this approach is the natural means for the envisaged applications of WSNs. As most of these applications aim to anticipate the requirements of humans and adapts the physical environments accordingly. However, the main challenge towards obtaining this benefit is the need to develop appropriate anticipatory algorithms that are suitable for scarce resources of WSNs.

Another direction for future research involves evaluating the performance of the proposed protocols over real-world scenarios either by incorporating the real-world data into simulation environments or by implementing these protocols over real testbeds. The key advantage of conducting research in this direction is that it can bridge the gap between the simulated results and the outcomes of real-world applications. One of the effective techniques that can be used to achieve this goal is feeding the simulation environment with real-world data such as real traffic distributions, topologies of real deployments and empirical channel characteristics. The benefit of such an approach is that it provides a flexible and affordable means to test the protocols from different perspectives and under different conditions. However, as some performance aspects depend on the data processing and storage spaces of sensor nodes, which are typically limited compared to simulation engines, thereby implementation of the proposed protocols over real testbeds can provide the most realistic performance outcomes. However, the main challenge of such implementation is that it requires a lot of effort, time and cost to setup and maintain the network; this is typically unmanageable due to the practical problems and generally needs long timescales to collect the data.

Identification of limitations of the proposed protocols is amongst the interesting areas for future research. In particular, this thesis proposed effective MAC, duty cycle management and routing protocols for multihop WSNs considering general scenarios, and validates the benefits of these protocols using both simulations and analytical models when applicable. These validations are carried out by comparing the performance of the proposed protocols with some widely used protocols under different network configurations. The majority of these comparisons demonstrate the performance results of our protocols in terms of throughput, average end-to-end delay and lifetime; however further assessment for each one of these protocols under different scenarios or using different performance metrics can highlight their key limitations. The key advantage of such further assessments is that they can provide meaningful insights into the operations of these protocols from new perspectives, which in turn facilities enhancing the performance of the proposed protocols. One of the main methods that can be used to undertake the further assessments is to analyse the dependency between the proposed protocols. Fundamentally, each one of the proposed protocols has been designed to work independently; this assumption is made in order to widen the applications of these protocols as different applications may require different protocols. Nevertheless, these protocols can be integrated to form cross layer protocols, and hence some functions and procedures would be redundant; thereby removing such redundancies can save storage space and reduce the computational complexity. However, the main challenge towards achieving this goal is that there is a need to analyse the dependency between the functions and procedures of different protocols. Dependency graphs [86-87] represent one of the widely-used techniques that have been used to figure out the dependency of cross layer protocols.

9-4 Future work related to the multidisciplinary metaphors

One of the key reasons behind selecting the inferential statistics to design WSN communication protocols in this thesis is that inferential statistics is a highly multidisciplinary research field. This multidisciplinary not only facilitates viewing WSNs from new perspectives but also enables us to exploit other disciplines as an inspirational sources. This thesis has developed two protocols based on drawing the metaphors between WSNs and artificial chemistry and improvisation process of harmony search algorithm. This section discusses the future work related to these metaphors.

The proposed duty cycle management scheme presented in chapter 7 introduced the metaphors between WSNs and Artificial Chemistry (AC). It has been shown that a sensor node can be represented as a chemical species and the communication between nodes can be represented as a chemical reaction. Based on this metaphors chapter 7 utilised the modelling technique of AC to derive time evolution for the energy consumption in WSNs and the optimisation technique to devise the duty cycle management scheme. Considering that AC is a mature research discipline with rich tools and techniques facilitates utilising the proposed metaphor to design communication protocols for more sophisticated environments such as nanoscale WSNs. Nanoscale WSNs is a new paradigm that aims to utilise the rapid development of nanotechnology to construct wireless communication networks consist of sensor nodes in order of nanometres in size. The main challenge towards realising the nanoscale WSNs is to find the appropriate approaches to design their communication protocols, as most of the traditional approaches are not applicable at nanoscale. The molecular communication concept [334] has been proposed to overcome this challenge, the molecular communication exploits the chemical processes that take place between DNAs as an inspiration source to devise communication protocols. Basically the molecular communication approach follows the metaphors represented in chapter 7; however this approach aims to build a complete communication system starting from the physical layer and up to application layer.

Chapter 8 provided another metaphor approach for WSNs. It has been shown that these networks can imitate one of the creative processes conducting by the skilled human beings (musicians). This metaphor represents WSNs as a musical band in which each sensor node acts like a musical player and the interactions between nodes are carried out via an improvisation process. Chapter 8 developed an effective routing protocol by allowing each sensor node to mimic the divergent thinking strategy of musician to recognise the relationships between nodes and to infer the routing metrics. Moreover spatial reasoning is used to enable a node to guide its route-discovering mechanism without need to overburden the scarce resources of nodes. In general the rapid developments of the WSNs impose unprecedented challenges that require devising WSN communication protocols based on the strong artificial intelligent discipline that attempts to mimic the behaviour of human. Instigated by this paradigm a number of strong artificial intelligence algorithms have been introduced e.g., brain storm optimisation algorithm, human group formation algorithm and social emotional optimisation algorithm [335]. Moreover, it is expected that these trends will be the dominant means to design communication protocols for WSNs as they can be adapted to cope with the diverse applications of WSNs.

Chapter 10

Summary and Conclusions

10-1 Overview

This thesis has introduced the issues and techniques associated with designing effective Medium Access Control (MAC), routing and duty cycle management strategies for multihop Wireless Sensor Networks (WSNs) using the principles of inferential statistics. These three protocols have been developed in order to exemplify the central thrust of the thesis that is to provide design principles and methodologies that can be used to develop communication protocols using the inferential statistic. WSNs represent an important infrastructure for an enormous number of applications in diverse areas including medical telemetry, disaster alarming, environmental monitoring and home and industrial automation. Moreover, most of the envisaged applications of WSNs aim to integrate these networks into global communication networks to form a world-wide digital skin over the physical environments. The main challenges in the design and development of communication protocols for WSNs are the scarce resources of these networks, in particular the need to design a lightweight and scalable protocol that can provide timely, reliable and energy efficient communications.

This research has overcome the challenge in the design of WSN communication protocols by employing the principles of inferential statistics. This approach has been selected due to many reasons. Firstly, inferential statistics provides powerful tools and techniques that utilise incomplete information of the observed data to draw conclusions about the underlying behaviour of the system under study. Hence a sensor node can make sound decisions without overburdening its resources. Secondly, inferential statistics features rigorous mathematical theorems that are spread across large multidisciplinary research fields. Therefore, using inferential statistics in the analysis and design of WSN communication protocols enhances understanding for hidden relationships between the underlying processes and behavioural responses of nodes under different operational conditions. Moreover, inferential statistics facilitates exploiting the methods and techniques of other disciplines as inspirational sources for developing WSN communication protocols. This thesis utilised inferential statistics to devise WSN protocols that can achieve three concurrent goals: increase the network throughput, reduce the end-to-end delay of packets and prolong the lifetime of nodes independently of topology configurations, traffic patterns or routing policies.

This chapter provides a general summary and conclusion of the work reported in this thesis, highlights its main findings and identifies the original contributions to the field. More detailed and specific summaries were provided at the end of each chapter.

Technologies related to WSNs and some of their applications were introduced in chapter 2 with the aim of providing background material for the subsequent chapters. A review of some of the hardware platforms of sensor nodes demonstrated the limited resources of WSNs. It has been shown that a typical sensor node is a tiny device with a low-speed processor, a few kilobytes of storage space, a short range communication transceiver and is powered by a small battery. A review of software platforms for WSNs illustrated that most of their operating systems are rudimentary. The overview for WSN applications presented in chapter 2 showed that most of the current and envisaged applications of WSNs are sophisticated and that there is strong trend to integrate these networks into global networks. Based on these reviews, chapter 2 discussed the main characteristics of WSNs and their impact on the design space of communication protocols. It has been shown that WSNs can be considered as a stack of stochastic processes, starting from ad-hoc topology where sensor nodes are scattered randomly to monitor the field of interest and up to the application layer where packets are generated to report random phenomena. Considering these characteristics led to the conclusion that using the inferential statistics is the natural means to design effective WSN protocols.

A comprehensive literature review on the Medium Access Control (MAC), routing and duty cycle management schemes for WSNs was introduced in chapter 3. It has been shown that the contention based MAC protocols and in particular the Carrier Sensing Multiple Access with Collision Avoidance (CSMA-CA) protocols are more suitable for multihop WSNs, since these protocols provide decentralised channel access mechanism with low-overheads. This argument has been supported by the discussion of the key characteristics of the time division multiple access, frequency division multiple access and code division multiple access schemes. Chapter 3 then reviewed the duty cycle management scheme reported in the literature. It has been shown that most of these schemes are developed as a part of MAC or routing protocols. Hence, these schemes are tailored to specific network configurations without consideration for the general requirements demanded by the wide applications of WSNs. It has been shown that most of the duty cycle management schemes either require nodes to be equipped with special equipment (as is the case with asynchronous or geographical cycle schemes) or they impose overheads (as is the case with synchronous duty cycle schemes). As both of these approaches demand high energy consumption, their resultant energy saving is reduced significantly by the additional overhead or equipment required to maintain their correct operations. A review of the basic concepts of routing in WSNs was also introduced in chapter 3. This review demonstrated the different classifications of routing protocols and highlighted their strong and weak aspects. It has been shown that the proactive and reactive routing protocols provide a trade-off between the route acquisition latency problem and the cost associated with obtaining routing information in advance. The discussion presented at the end of chapter 3 provided the motivations for using the inferential statistics to design communication protocols by considering the underlying approaches of the proposed protocols.

Chapter 4 has introduced the simulation techniques and validation methodologies used in this research. Some of the simulation platforms used to assess the performance of WSNs were overviewed and then a detailed review for the chosen simulator, NS3, was presented. NS3 has been selected as it provides a reliable simulation tool whose underlying assumptions imitate the real characteristics of WSNs closely. The importance of using analytical models in conjunction with the simulation tools has been highlighted in chapter 4. It has been shown that such validation can increase the integrity of results and facilitate assessing the operations of WSNs from different perspectives. Finally, chapter 4 presented the general simulation parameters that were used to evaluate the performance of the proposed protocols.

Chapter 5 has proposed a novel analytical framework for statistical analysis of the CSMA-CA protocols over multihop networks. The proposed framework represents the network as a three layered model: a topology model, a routing model and a queuing model. The topology model uses set theory to describe the spatial relationships between nodes considering the channel model. The routing model quantifies the traffic characteristics of each node based on the inter-arrival distributions and the multipath load balancing strategies. The queuing model represents a sensor node as a GI/G/1 queue whose inter-arrival distribution is obtained from the routing model and whose service time distribution is derived for the characteristics of the CSMA-CA scheme under consideration. Chapter 5 outlined the multifaceted advantages of the proposed analytical framework. It has been shown that this framework provides a rigorous approach that can account for real characteristics of the network including heterogeneous traffic characteristics over nodes, distinct channel conditions of each transmitter-receiver pair and the impact of the variation back-off windows on the probability of channel assessment. A second main advantage of the analytical framework is that it offers a versatile means that can assess the performance of a MAC protocol under a wide variety of situations just by modifying the corresponding layer without the need to reconstruct the entire analytical framework. Moreover, the proposed framework is able to derive the complete probability distributions of the inter-arrival, inter-departure, service and queuing times of packets which in turn facilitates obtaining the first and higher order performance metrics readily. These features make the proposed analytical framework one of the most powerful tools in the analysis and design of MAC protocols for WSNs.

A demonstration for the value of the proposed analytical framework was introduced in chapter 5 through assessment of the statistical characteristics of the IEEE 802.15.4 CSMA-CA protocol. This protocol has been selected because it is the official standard that was ratified for low power networks and has been used widely in both industrial and academic disciplines. Assessment of the performance metrics of the IEEE 802.15.4 CSMA-CA protocol highlighted that this protocol provides poor performance over multihop networks. It has been shown that the end-to-end delay and the throughput degrade significantly with increasing number of hops. A deep investigation of these characteristics has been introduced in chapter 5 through assessment the probability distribution of the service time. This assessment reveals the fact that the IEEE 802.15.4 CSMA-CA protocol produces a heavy tailed distribution in which the majority of mass functions are spread out over a large area with small probabilities between large service time intervals. These characteristic has been attributed to the power law process that results from the Binary Exponential Back-off (BEB) mechanism.

Chapter 6 proposed an effective CSMA-CA protocol for multihop WSNs. This chapter demonstrated that most of the CSMA-CA protocols suffer from two main limitations. Firstly, the lack of a flexible mechanism via which nodes can adjust their contention parameters to satisfy the dynamic characteristics of multihop networks. Secondly, most of the CSMA-CA protocols provide no means to resolve or even identify the main causes of unsuccessful transmissions. Chapter 6 showed that these limitations can be overcome by replacing the uniform distribution with the Gamma distribution and specifying an intelligent collision resolution mechanism. The main reason for selecting the Gamma distribution to generate the back-off intervals is that the Gamma distribution is the conjugate prior for most of the probability distributions that are used to model traffic in communication networks, e.g., Exponential, Poisson, Normal and even the heavy tailed distribution Pareto, Weibull and lognormal distributions. Another key reason of using the Gamma distribution is that this distribution can mitigate the high level of variation found in the BEB mechanism. Since employing the Gamma distribution produces a Gamma process that has a low level of variation compared to that yielded from the power law process in the BEB scheme. Other statistical properties of the Gamma distribution and their importance in designing an effective back-off scheme have also been considered in chapter 6. The second main pillar of the proposed CSMA-CA protocol was to develop an intelligent collision resolution mechanism which is able to identify the underlying causes of unsuccessful transmissions and to take remedial actions by adjusting the contention parameters.

The pseudo-code presented in chapter 6 showed that the proposed protocol allows each node to utilise the principles of inferential statistics to adjust the parameters of the Gamma distribution according to its prediction for the contention behaviour of other nodes. The main objectives of this adjustment are to reduce the end-to-end delay of packets, to increase the throughput of the network and to prolong the lifetime of nodes. The end-to-end delay is reduced by allowing a node to access an idle channel with high probability and to back-off with high probability when the channel is busy. Therefore, a node can avoid deferring transmission of its packet for random periods which in turn reduces the service time of packets. Reducing the service time increases the network throughput by leveraging the channel utilisation and prolonging the lifetime of nodes by reducing the energy consumed in contenting to access the channel. A comprehensive comparison between the proposed protocol and the IEEE 802.15.4 CSMA-CA standard presented in chapter 6 shows the outperformance of the proposed CSMA-CA protocol.

The design of an effective duty cycle management scheme for WSNs was the topic of chapter 7. This chapter discussed the main reason behind the shortage in devising a general duty cycle management scheme. Notably the absence of a framework that is capable of accounting for the dynamic interaction between nodes while being able to provide a means to measure the level of optimisation. Motivated by this shortage, chapter 7 exploited the Artificial Chemistry (AC) discipline to design duty cycle management scheme. AC was selected as it provides multidisciplinary and powerful tools that can be used to design, analyse and optimise one of the most sophisticated dynamic systems (i.e., chemical reaction processes). A brief introduction to AC and its related research fields in the context of computer and communication engineering has been considered in chapter 7. It has been shown that AC has played major roles in developing large number of computational algorithms including evolutionary algorithms, autonomic computing and artificial life. Chapter 7 revealed the analogies between WSNs and chemical reaction networks and then exploited these analogies to model the energy consumption and design the duty cycle management scheme. The modular statistical mechanism was introduced in chapter 7 as the means to model the energy consumption of multihop WSNs. The modular statistical mechanism represents the energy consumption of WSNs as a set of the delayed differential equations in which the metric space is the residual energy of sensor nodes and time evolution operators are the inter-arrival times of packets. The integrity of this modelling approach has been assessed in chapter 7 by comparing its outcomes versus simulation results under different scenarios.

Chapter 7 reviewed the artificial decision making of AC, highlighted its underlying approach and demonstrated the advantages that can be obtained by using it to design an effective duty cycle management scheme for multihop WSNs. It has been shown that the underlying approach of the artificial decision making of AC is non-parametric Bayesian inference. In such an approach, a WSN is represented as a collection of smart agents each of which infers the potential communications and then takes the appropriate action autonomously considering the interacting relationships. The two key advantages of this approach as outlined in chapter 7 are its lightweight nature and versatility. The lightweight nature was attributed to the fact that the artificial decision making process of AC does not require a node to acquire complete information about the whole communications of the network. The versatile advantage of the artificial decision making was highlighted by considering the fact that in such approach a node has to modify its position with respect to the potential reaction without need to alter the reaction itself.

The pseudo-code provided in chapter 7 demonstrated the mechanism by which a node can infer the potential communication and the techniques used to generate the sleeping and waking intervals. Moreover, chapter 7 evaluated the benefits of the proposed duty cycle management scheme by comparing it performance with the unity duty cycle scheme under different MAC protocols including the IEEE 802.15.4

and Gamma based CSMA-CA protocols. It has been shown that that proposed duty cycle management scheme is able to achieve the three objectives stated in this thesis (i.e., prolonging the lifetime of the node, reducing the end-to-end delay of packets and leveraging the network throughput). Chapter 7 explained how the proposed scheme can achieve the aforementioned objectives. It has been shown that the lifetime of nodes is prolonged as a result of alleviation of the detrimental effects of overhearing and idle listening and without the need to overwhelm the network with control packets. The end-to-end delays of packets are reduced due to ability of the proposed duty cycle scheme to route packets over the optimal path for longer periods. This in turn avoids transmission packets over sub-optimal paths and leverages the throughput of networks.

Chapter 8 considered the routing protocol for WSNs, it has been shown that the greatest challenge in designing an effective routing protocol is to provide the optimal balance between the cost associated with obtaining fresh information about the status of the network and the scarce resource of WSNs. Towards this goal, chapter 8 formulate the routing protocol as an optimisation problem with two basic mechanisms: diversification and intensification. The diversification mechanism was defined as the process by which nodes can discover the paths of the network while the intensification mechanism is used to direct nodes to focus on those area of network that are presumably have an optimal paths. Chapter 8 exploited the formulation of the routing protocol as an optimisation problem to review the fundamental logic of the routing protocols that were appeared in the literature. It has been shown that the key limitation of these approaches is the imbalance between the diversification and intensification mechanisms. In particular, the classical, admissible and swarm intelligent based routing concentrates on the diversification mechanism and overwhelm the network with control packets. Conversely, other approaches such as the reinforcement learning, game theory and fuzzy logic focus on the intensification mechanism which in turn slow approaching the optimal routes. Chapter 8 discussed the future trends in designing effective routing protocols for WSNs. It has been shown that there is a strong trend in employing the principles of strong artificial intelligence in order to provide a more robust and adaptive approaches that can satisfy the rapid development of the WSNs.

Chapter 8 proposed a novel routing protocol for WSNs inspired by the Harmony Search (HS) algorithm. The algorithm has been selected due to many reasons. Firstly, HS is able to imitate the creative process that is conducted by musicians during their endeavours to improvise the most pleasing harmony on the spur of the moment. Hence higher cognitive capabilities of humans such as the inference and reasoning can be utilised in designing the routing protocols. Secondly, HS considers the distinct characteristics of different agents and provides number of feedback mechanisms that can enhance the reliability of routing protocols. Other appealing features of the HS algorithm and their importance roles in devising an effective routing protocol have been discussed in chapter 8. This chapter provided the metaphors between HS and WSNs. It has been shown that WSNs can be abstracted as a musical band in which each sensor node represents a musician while the interactions between these nodes are carried out by mimicking the improvisation process artificially.

Based on these metaphors, chapter 8 specified the proposed routing protocol in terms of three mechanisms: divergent thinking; goal oriented search and error correction. Divergent thinking enables a node to recognise the network as a set of objectives that are typified by a set of attributes. The proposed protocol defined the inference process by which a node can gather these attributes and illustrated the advantages of this approach in reducing the side effects of overhead of control packets. The routing protocol proposed in chapter 8 defined the goal oriented search to enable a node to guide its search process based on certain preferences compared to the node's perspective to the network. Chapter 8 exploits a simple yet powerful inferential statistics technique called spatial reasoning to aid a node in deriving its perspective of the network. Finally the error correction mechanism is defined in the proposed protocol to avoid routing packets over sub-optimal paths. The error correction mechanism, divergent thinking and spatial reasoning form the intensification and diversification processes of the proposed protocol. Comprehensive assessments of the proposed routing protocol have been introduced in chapter 8 using different network configurations and compared to two widely used protocols, IEEE 802.15.5 and an ant colony protocol. It has been concluded that the proposed routing protocol is able to reduce the end-to-end delay of packets, prolong the lifetime of nodes and leverage the network throughput.

10-2 Original Contributions

This section outlines the main novel contributions represented in this thesis, this contributions are categorised into three areas: developed protocols, analytical models and multidisciplinary metaphors.

10-2-1 Proposed protocols

This thesis introduced a number of WSN communication protocols; all of these protocols are able to achieve three concurrent goals: leverage the network throughput, prolong the lifetime of nodes and reduce the end-to-end delay of packets. Moreover, the proposed protocols consider general cases without restrictions on topology configurations, traffic patterns and routing policies. These protocols are:

- A novel CSMA-CA protocol based on Gamma distribution back-off mechanism has been developed in chapter 6 [6].
- A novel duty cycle management scheme inspired by artificial chemistry has been introduced in chapter 7 [7].
- A novel routing protocol inspired by the harmony search algorithm has been designed in chapter 8 [8].

10-2-2 Analytical Models

Two analytical models have been introduced in thesis with the aim to get deep insights into the behaviours of WSNs and to provide rigorous prediction means for the operations of these networks. These analytical models are:

- A novel analytical model to analyse the statistical characteristics of CSMA-CA protocols under general case has been developed in chapter 5.
- A novel analytical model to predict the energy consumption of the multihop WSNs operating CSMA-CA has been introduced in chapter 7.

10-2-3 Multidisciplinary metaphors

Besides the proposed protocols and the analytical models, this thesis provided broader perspectives of the WSNs by casting their operations in the domain of other research fields. This in turn facilitates using these fields as inspirational sources for design effective communication protocols. The following two metaphors are presented in this thesis:

- The artificial chemistry metaphor which represents WSNs as a chemical reaction network.
- The harmony search metaphor which represents WSNs as a musical band.

References

[1] M. Baz, and D. A. J. Pearce, "An introduction to DANSE," *in Proc.* 12th *Annual PostGraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting* (PGNet), 2011.

[2] M. Baz, and D. A. J. Pearce, "Header Compression Scheme for IEEE 802.15.5," *in Proc. International Symposium on Wireless Communication Systems* (ISWCS), 2011.

[3] M. Baz, P. D. Mitchell, and D. A. J. Pearce, "Improvements of CSMA-CA in IEEE 802.15.4," *in Proc. International Conference on High Performance Computing and Communication* (HPCC), 2012.

[4] M. Baz, P. D. Mitchell, D. Grace, and D. A. J. Pearce, "Gamma distribution based CSMA-CA back-off scheme for multihop wireless networks," *in Proc. International Conference on Electronics, Communications and Photonics Conference* (SIECPC), 2013.

[5] M. Baz, P. D. Mitchell, and D. A. J. Pearce, "Versatile Analytical Model for Delay and Energy Evaluation in WPANs: A Case Study for IEEE 802.15.4 CSMA-CA," *Wireless Personal Communications*, vol.75, no.1, pp. 415-445, 2014.

[6] M. Baz, P. D. Mitchell, and D. A. J. Pearce, "Analysis of Queuing Delay and Medium Access Distribution over Wireless Multihop PANs," *IEEE Transaction on Vehicular Technology*, Accepted, 2014.

[7] M. Baz, P. D. Mitchell and D. Grace, "Duty Cycle Management Scheme for Multihop Networks Inspired by Artificial Chemistry," *Submitted to IEEE Transaction on Wireless Communications*, 2014.

[8] M. Baz and P. D. Mitchell, "Effective Routing Protocols for IoT inspired by Improvisation Process," *in preparation*.

[9] J. Vasseur, and Adam Dunkels, *Interconnecting Smart Objects with IP-The Next Internet*, Morgan Kaufmann, 2010.

[10] J. Momoh, *Smart Grid: Fundamentals of Design and Analysis*, John Wiley & Sons, 2012.

[11] H. Nakashima, et al., *Handbook of Ambient Intelligence and Smart Environments*, Springer, 2009.

[12] C. Chong, and S. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges," *IEEE Journal and Magazines*, vol.91, no.8, pp. 1247-1256, 2003.

[13] K. Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks: Technology, Protocols, and Applications*, John Wiley & Sons, 2007.

[14] H. Eren, *Wireless Sensors and Instruments: Networks, Design, and Applications*, CRC Press, 2005.

[15] M. Ilyas, and I. Mahgoub, *Smart Dust: Sensor Network Applications, Architecture And Design*, CRC PressINC, 2006.

[16] *Wireless Medium Access Control (MAC) and physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Network (LR-WPANs)*,

IEEE 802.15.4 Standard, 2011.
[17] Z. Shelby and C. Bormann, *6LoWPAN: The Wireless Embedded Internet*, John Wiley & Sons, 2011.
[18] International Telecommunications Union, “ITU Internet Reports 2005: The Internet of Things. Executive Summary,” [Online] Available at: http://www.itu.int/osg/spu/publications/internetofthings/InternetofThings_summary.pdf.
[19] O. Vermesan, and P. Friess, *Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems*, River Publishers, 2013.
[20] E. Lee, and S. Seshia, *Introduction to Embedded Systems: A Cyber-physical Systems Approach*, Lee & Seshia, 2011.
[21] L. Gavrilovska, et al., *Application and Multidisciplinary Aspects of Wireless Sensor Networks: Concepts, Integration, and Case Studies*, Springer Science & Business Media, 2010.
[22] H. Karl, and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*, John Wiley & Sons, 2007.
[23] A. Bachir, et al., “MAC Essentials for Wireless Sensor Networks,” *IEEE Communications Surveys & Tutorials*, vol.12, no.2, pp.222-248, 2010.
[24] R. Carrano, et.al, "Survey and Taxonomy of Duty Cycling Mechanisms in Wireless Sensor Networks," *IEEE Communications Surveys & Tutorials*, vol.16, no.1, pp.181-194, 2014.
[25] G. Anastasi, et.al, "Energy conservation in wireless sensor networks", *Ad Hoc Networks*, vol. 7, no.3, pp 537-568, 2009.
[26] K. Akkaya, and M. Younis, “A Survey of Routing Protocols in Wireless Sensor Networks", *Ad Hoc Network*, vol.3, no.3, pp.325 -349 2005.
[27] D. Gabbay, et.al, *Philosophy of Statistics*, Elsevier, 2011.
[28] A. Hald, *A History of Probability and Statistics and Their Applications before 1750*, John Wiley & Sons, 2005.
[29] N. Johnson, Ed., and S. Kotz, Ed., *Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present*, John Wiley & Sons, 2011.
[30] L. Vaughan, *Statistical Methods for the Information Professional: A Practical, Painless Approach to Understanding, Using, and Interpreting Statistics*, Information Today, Inc., 2001.
[31] A. Welsh, *Aspects of Statistical Inference*, John Wiley & Sons, 2011.
[32] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks: Theory and Practice*, John Wiley & Sons, 2010.
[33] I. Stievano, Ed., *Flash Memories*, InTech, 2011.
[34] S. Misra, et.al, *Guide to Wireless Sensor Networks*, Springer Science & Business Media, 2009.
[35] P. Levis, and D. Gay, *TinyOS Programming*, Cambridge University Press, 2009.
[36] H. Cha, et.al, “The RETOS operating system: kernel, tools and applications,” In *Pro. The 6th International Conference on Information Processing in Sensor Networks* (IPSN), pp. 559-560, 2007.
[37] H. Abrach et al., ‘‘MANTIS: System Support for Multimodal Networks of

In-Situ Sensors,'' in *Pro. The 2nd Workshop on Sensor Networks and Applications* (WSNA'03), pp. 50–59, 2003.

[38] F. Zhao, and L. Guibas, *Wireless Sensor Networks: An Information Processing Approach*, Morgan Kaufmann, 2004.

[39] M. Kuorilehto, *Ultra-Low Energy Wireless Sensor Networks in Practice: Theory, Realization and Deployment*, John Wiley & Sons, 2008.

[40] J. Hill, and D. Culler, "Mica: a wireless platform for deeply embedded networks," *IEEE Micro Magazine*, vol.22, no.6, pp.12-24, 2002.

[41] V. Raghunathan et.al, "Energy-Aware Wireless Microsensor Networks," *IEEE Signal Processing Magazine*, vol.19, no.2, pp. 40–50, 2002.

[42] S. Hollar, "Cots dust," M. S. thesis, University of California, Berkeley, 2000.

[43] Atmel Corporation, "AT90S/LS8535 Microcontroller Datasheet," [Online], Available at: http://www.atmel.com./Images/doc1041.pdf.

[44] Murata Manufacturing Co., "TR1000 Transceiver Datasheet," [Online], Available at: http://www.rfm.com/products/data/tr1000.pdf.

[45] J. Polastre, et al., "The mote revolution: Low power wireless sensor networks," *in Proc. The 16th Symposium on High Performance Chips* (HotChips), 2004.

[46] Atmel Corporation, "ATmega128/128L Microcontroller Datasheet," [Online], Available at: http://www.atmel.com/images/doc2467.pdf.

[47] Texas Instruments, "CC100 Single Chip Very Low Power RF Transceivers," [Online], Available at: http://www.ti.com/lit/ds/symlink/cc1000.pdf.

[48] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," In *Pro. The 6th International Conference on Information Processing in Sensor Networks* (IPSN), pp.364-369, 2005.

[49] A. Savvides, M. Srivastava, "A distributed computation platform for wireless embedded sensing," *in Proc. IEEE Conference on Computer Design: VLSI in Computers and Processors* (ICCD), pp.220-225, 2002.

[50] J. Hill, "System Architecture for Wireless Sensor Networks," Ph.D thesis, University of California, Berkeley, 2003.

[51] J. Rabaey, et.al, "PicoRadio: Ad-hoc Wireless Networking of Ubiquitous Low-Energy Sensor/Monitor Nodes," *in Proc. The IEEE Very large Scale Integrated Systems* (VLSI), pp.9-12, 2000.

[52] Y. Chee, et al., "PicoCube: A 1cm^3 sensor node powered by harvested energy," *in Proc*. The 45th ACM/IEEE conference on Design Automation Conference (DAC), pp.114-119, 2008.

[53] D. Bend, et al., "PicoSystem, Self-Powered, Wireless Sensor Platform," [Online], Available at: https://ei.haas.berkeley.edu/c2m/PICO_Cube.pdf.

[54] Military Analysis Network, "Cooperative Engagement Capability," [Online], Available at http://fas.org/man/dod-101/sys/ship/weaps/cec.htm.

[55] A. Tambini, *Wiring Vietnam: The Electronic Wall*, Scarecrow Press, 2007.

[56] N. Abramson, "The Aloha System — Another Alternative for Computer Communications", *Proc. Fall Joint Computer Conference*, 1970.

[57] L. Castano, and A. Flatau, "Smart fabric sensors and e-textile technologies: a review", *Smart Materials and Structures*, vol.23, no.5, pp.1-27, 2014.

[58] J. Augusto, Ed., and C. Nugent, Ed., *Designing Smart Homes: The Role of Artificial Intelligence*, Springer Science & Business Media, 2006.

[59] S. Helal, et.al, "The Gator Tech Smart House: a programmable pervasive space," *IEEE Computer Magazine*, vol.38, no.3, pp.50-60, March 2005.

[60] D. Cook, et.al, "MavHome: an agent-based smart home," *in Proc. The 1st IEEE Conference on Pervasive Computing and Communications* (PerCom), pp.521 – 524, 2003.

[61] Siemens, "Synco living – save energy with attractive home automation," [Online], Available at http://www.buildingtechnologies.siemens.com/bt/global/en/buildingautomation-hvac/home-automation-system-synco-living/Pages/synco-living.aspx.

[62] M. Llyas, et al., *Sensor Networks for Sustainable Development*, CRC Press, 2014.

[63] Y. Strengers, *Smart Energy Technologies in Everyday Life: Smart Utopia*, Palgrave Macmillan, 2013.

[64] Hitachi, Ltd, "Smart Grid", [Online], Available at http://www.hitachi.com/environment/showcase/solution/energy/smartgrid.html.

[65] V. Giordano, et.al, "Smart Grid projects in Europe," [Online], Available at: http://ses.jrc.ec.europa.eu/sites/ses/files/documents/smart_grid_projects_in_europe_lessons_learned_and_current_developments.pdf.

[66] D. Cooka, J. Augustob, and V. Jakkulaa, "Ambient intelligence: Technologies, applications, and opportunities," *Pervasive and Mobile Computing*, vol.5, no.4, pp.277-298, 2009.

[67] S. Giroux, Ed., and H. Pigot, Ed., *From Smart Homes to Smart Care*, IOP, 2005.

[68] Ulster Community and Hospitals Trust, "Interim policy on research governance process," [Online], Available at: http://webarchive.nationalarchives.gov.uk/20130805112926/http://media.dh.gov.uk/network/18/files/2011/11/interim-governance-arrangements-7-nov-2011.pdf.

[69] J. Sachpazidis, "@HOME: A modular telemedicine system," *in Proc. The 2nd Conference on Mobile Computing in Medicine*, Pages: 87–95, 2002.

[70] B. Kamel, "CAALYX: a new generation of location-based services in healthcare," *International journal of health Geographics*, vol.9. no.6, pp.1-6 2007.

[71] J. Habetha, "The MyHeart Project: Fighting cardio-vascular diseases by prevention and early diagnosis," *in Proc. The 28th IEEE Engineering in Medicine and Biology Society* (EMBS), pp. 6746 – 6749, 2006.

[72] Contact lens Headline, "Smart Contact Lens Measures Intraocular Pressure; Helps Glaucoma Patients Manage Condition," [Online], Available at: http://www.contactlensheadlines.com/6447/triggerfish/

[73] A. Baggio, "Wireless Sensor Networks in Precision Agriculture," *in Proc. ACM Workshop Real-World Wireless Sensor Networks*, 2005.

[74] J. Balendonck, "Farm Level Optimal Water management Assistant for Irrigation under Deficit (FLOWAID)," European Community's Sixth Framework Programme through, Public report, [Online], Available at: http://cordis.europa.eu/documents/documentlibrary/124584981EN6.pdf,

2009.

[75] G. Werner-Allen, et al., “Deploying a Wireless Sensor Network on an Active Volcano,” *IEEE Internet Computing*, vol.10, no.2, pp.18-25, 2006.

[76] J. Huang, S. Amjad and S. Mishra, “CenWits: a Sensor-based Loosely Coupled Search and Rescue System using Witnesses,” *in Proc. The Third ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp. 180-191, 2005.

[77] Vehicle Traffic Monitoring Platform with Bluetooth Sensors over ZigBee, [Online], Available at http://news.ne2ne.com/articles/share/50249/.

[78] K. Zigangirov, *Theory of Code Division Multiple Access Communication*, John Wiley & Sons, 2004.

[79] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 1997.

[80] S. Basagni, Ed., et.al, *Mobile Ad Hoc Networking*, John Wiley & Sons, 2004.

[81] I. Akyildiz and X. Wang, *Wireless Mesh Networks*, John Wiley & Sons, 2009.

[82] P. Mohapatra, Ed., and S. Krishnamurthy, Ed., *AD HOC NETWORKS: Technologies and Protocols*, Springer Science & Business Media, 2005.

[83] S. Ni, et. al., “The Broadcast Storm Problem in a Mobile Ad Hoc Network,” *in Proc. The ACM Annual Conference on Mobile Computers and Networks* (MobiCom), pp. 151-162, 1999.

[84] J. Zheng and A. Jamalipour, *Wireless Sensor Networks: A Networking Perspective*, John Wiley & Sons, 2009.

[85] H. Jie, Z. Baoxian and H. Mouftah, "Routing protocols for duty cycled wireless sensor networks: A survey," *IEEE Communications Magazine*, vol.50, no.12, pp.116-123, 2012.

[86] V. Kawadia and R. Kumar, "A cautionary perspective on cross-layer design," *IEEE Wireless Communications*, vol.12, no.1, pp.3-11, 2005.

[87] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol.43, no.12, pp.112-119, 2005.

[88] Raja Jurdak, *Wireless Ad Hoc and Sensor Networks: A Cross-Layer Design Perspective*, Springer Science & Business Media, 2007.

[89] B. O’Hara and A. Petrick, *IEEE 802.11 Handbook: A Designer's Companion*, IEEE Standards Association, 2005.

[90] W. Ye, J. Heidemann and D. Estrin "Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks", *IEEE/ACM Transaction on Networking*, vol. 12, no. 3, pp.493 -506 2004.

[91] W. Heinzelman, W. Chandrakasan, and H. Balakrishnan, "Energy-efficient Communication Protocol for Wireless Microsensor Networks," *in Proc. The 33rd Annual Hawaii International Conference on System Sciences*, pp.4-7, 2000.

[92] W. Silva, and J. Heidemann, “Ultra-low duty cycle MAC with scheduled channel polling,” *in Proc. The Fourth ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp. 321–334, 2006.

[93] F. Hamady, et.al, “Enhancement of the S-MAC protocol for wireless sensor

networks, *in Proc. The 4th IEEE Conference on Energy Aware Computing* (ICEAC), pp.16-18, 2010.
[94] T. Vandam, and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," *in Proc. The 1st ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp. 171–180, 2003.
[95] Y. Zhao, et.al, "A survey and projection on medium access control protocols for wireless sensor networks," *ACM Computing Surveys*, vol.45, no.1, 2012.
[96] J. Polastre, J. Hill and D. Culler "Versatile Low Power Media Access for Wireless Sensor Networks," *in Proc. The 2nd ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp. 95-107, 2004.
[97] V. Rajendran, K. Obraczka and J. Garcia-Luna-Aceves, "Energy-Efficient, Collision-Free Medium Access Control for Wireless Sensor Networks," *in Proc. The 1st ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp.181-192, 2003.
[98] K. Sohrabi, J. Gao, V. Ailawadhi and G. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Personal Communication*, vol.7, no.5, pp. 16-27, 2000.
[99] W. Xinhua and W. Sheng, "Performance Comparison of LEACH and LEACH-C Protocols by NS2," *in Proc. The 9th IEEE Conference on Distributed Computing and Applications to Business Engineering and Science* (DCABES), pp. 254-258, 2010.
[100] S. Lindsey and C. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," *in Proc. The Conference for Aerospace Experts, Academics, Military Personnel, and Industry Leaders*, pp. 1125-1130, 2002.
[101] Y. Sun, O. Gurewitz, and D. Johnson, "RIMAC: A receiver initiated asynchronous duty cycle MAC protocol for dynamic traffic loads in wireless sensor networks," *in Proc. The 6th ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp.1-14, 2008.
[102] S. Farahani, *ZigBee Wireless Networks and Transceivers*, Newnes, 2011.
[103] I. Rhee, et.al, "Z-MAC: A hybrid MAC for wireless sensor networks," *in Proc. The 5th ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp.90-101, 2007.
[104] G. Halkes and K. Langendoen, "Crankshaft: An Energy-Efficient MAC-Protocol For Dense Wireless Sensor Networks" *in Proc. European Conference on Wireless Sensor Networks* (EWSN), pp.228-244, 2007.
[105] C. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers", *in Proc. The ACM conference on Communications Architectures, Protocols and Applications* (SIGCOMM), pp.234 -244 1994.
[106] P. Jacquet, et al., "Optimized link state routing protocol for ad hoc networks," *in Proc. IEEE Multi Topic Conference* (INMIC), pp.62-68, 2001.
[107] D. Johnson and D. Maltz, "Dynamic source routing in ad-hoc wireless networks," *Mobile Computing, Kluwer International Series in Engineering and Computer Science*, vol.353, pp.153-181,1966.
[108] J. Broch, D. Johnson, and D. Maltz, "The Dynamic Source Routing Protocol

for Mobile Ad Hoc Networks," IETF Internet draft, [Online] Available at: http://www.ietf.org/rfc/rfc4728.txt, 2007.

[109] C. Perkins, E. Royer, and S. R. Das, "Ad Hoc on Demand Distance Vector (AODV) Routing", IETF Internet Draft, [Online] Available at: http://www.ietf.org/internet-drafts/draft-ietfmanet-aodv-06.txt, 2000.

[110] Z. Haas, "A new routing protocol for the reconfigurable wireless networks," *in Proc. The 6th IEEE Conference on Universal Personal Communications Record*, pp.562-566, 1977.

[111] ZigBee Alliance, "ZigBee Network Layer Specifications", [Online] Available at: http://www.zigbee.org/Standards/Overview.aspx, 2004.

[112] *Recommended Practice for Information technology Telecommunications and information exchange between systems Local and metropolitan area networks-Specific requirements Part 15. 5: Mesh Topology Capability in Wireless Personal Area Networks (WPANs)*, IEEE 802.15.5 Standard, 2009.

[113] T. Winter, et al., "RPL: Routing Protocol for Low Power and Lossy Networks," [Online] Available at: http://tools.ietf.org/html/draft-ietf-roll-rpl-19.

[114] S. Roychowdhury and C. "Geographic adaptive fidelity and geographic energy aware routing in ad hoc routing," *International Journal of Computer & Communications Technology*, vol.1, no.2, 2010.

[115] B. Chen, "SPAN: an Energy-efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks", *Wireless Networks*, vol.8, no.5, pp.481-494, 2002.

[116] A. Cerpa and D. Estrin, "Ascent: Adaptive Self-Configuring Sensor Network Topologies," *IEEE Transactions on Mobile Computing*, vol.3, no.3, pp. 272-285, 2004.

[117] J. Polastre, et.al, "A Unifying Link Abstraction for Wireless Sensor Networks". *in Proc. The 3rd ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp.76-89, 2005.

[118] S. Du, K.Saha and D. Johnson, "RMAC: A routing-enhanced duty-cycle MAC protocol for wireless sensor networks," *in Proc. The 26th Annual IEEE Conference on Computer Communications*, pp. 1478–1486, 2007.

[119] Y. Sun, et.al., "DW-MAC: A low latency, energy efficient demand wakeup MAC protocol for wireless sensor networks," *in Proc. The 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (MobiHoc), pp.53–62, 2008.

[120] G. Holland, N. Vaidya, and P. Bahl, "A Rate-Adaptive MAC Protocol for Multihop Wireless Networks," *in Proc. The ACM 7th Annual Conference on Mobile Computers and Networks* (MobiCom), pp. 236-251, 2001.

[121] K. Wehrle, M. Günes and J. Gross, *Modeling and Tools for Network Simulation*, Springer Science & Business Media, 2010.

[122] M. Guizani, et.al, *Network Modeling and Simulation: A Practical Perspective*, John Wiley & Sons, 2010.

[123] NS3 Simulator, [Online] Available at: http://www.nsnam.org/.

[124] OPNET Simulator, [Online] Available at: http://www.riverbed.com /products/performance-management-control/opnet.html?redirect=opnet.

[125] A. Sethi and V. Hnatyshin, *The Practical OPNET User Guide for Computer Network Simulation*, CRC Press, 2012.

[126] TOSSIM Simulator, [Online] Available at: http://cs.ecs.baylor.edu/~donahoo/classes/5321/projects/SensorNet/tossim.html.

[127] P. Levis, et.al, "TOSSIM: Accurate and Scalable Simulation of Entire TinyOS Applications," *in Proc. The 1st ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp. 126-137, 2003.

[128] Castalia Simulator, [Online] Available at: https://castalia.forge.nicta.com.au /index.php/en/.

[129] A. Boulis, "Castalia: revealing pitfalls in designing distributed algorithms in WSNs," *in Proc. The 5th ACM Conference on Embedded Networked Sensor Systems* (SenSys), pp. 407–408, 2007.

[130] NS2 Simulator, [Online] Available at: http://www.isi.edu/nsnam/ns/.

[131] T. Issariyakul and E. Hossain, *Introduction to Network Simulator NS2*, Springer Science & Business Media, 2011.

[132] NS3 Simulator, [Online] Available at: http://www.nsnam.org/.

[133] J. Font, et.al, "Analysis of source code metrics from ns-2 and ns-3 network simulators," *Simulation Modelling Practice and Theory,* vol.19, no.5, pp.1330-1346,2011.

[134] G. Carneiro, P. Fortuna, and M. Ricardo, "Flowmonitor: A network monitoring framework for the network simulator 3 (ns-3)," *in Proc. The 4th International ICST Conference on Performance Evaluation Methodologies and Tools*, pp.1-10, 2009.

[135] M. Stoffers and G. Riley, "Comparing the ns-3 Propagation Models," *in Proc. The 20th IEEE Conference on Modeling, Analysis & Simulation of Computer and Telecommunication Systems* (MASCOTS), pp.61-67, 2012.

[136] H. Wu, S. Nabar and R. Poovendran, "An energy framework for the network simulator 3 (ns-3)," *in Proc. The 4th International ICST Conference on Simulation Tools and Techniques*, pp.222-230, 2011.

[137] CC2420 Transceiver Datasheet, [Online]. Available at: http://www.ti.com/lit/ds/symlink/cc2420.pdf.

[138] C. Bettstetter, "On the Connectivity of Wireless Multihop Networks with Homogeneous and Inhomogeneous Range Assignment," *in Proc. Vehicular Technology Conference* (VTC- Fall), pp.1706-1710, 2002.

[139] L. Kleinrock, *Queuing System: Theory*. John Wiley & Sons, 1976.

[140] S. Tasaka, *Performance Analysis of Multiple Access Protocols*, MIT Press, 1986.

[141] L. Kleinrock and S. Lam, "Packet Switching in A Multiaccess Broadcast Channel: Performance Evaluation," *IEEE Transactions on Communications*, vol.23, no.4, pp. 410–423, 1975.

[142] G. Bianchi, "Performance analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol.18, no.3 pp.535-547, 2000.

[143] J. Mišic and V. Mišic, "Access delay for nodes with finite buffers in IEEE 802.15.4 beacon enabled PAN with uplink transmissions," *Computer Communications*, vol. 28, no. 10, pp.1152–1166, Jun. 2005.

[144] X. Li, et al. “Four-dimensional Markov Chain Model of Single-hop Data Aggregation with IEEE 8 .15.4 in Wireless Sensor Networks,” *Wireless Network*, vol.18, no.5, pp. 469-479, 2012.

[145] A. Faridi, et al. “Comprehensive Evaluation of the IEEE 802.15.4 MAC Layer Performance With Retransmissions,” *IEEE Transactions on Vehicular Technology*, vol.59, no.8, pp.3917-3932, 2010.

[146] Y. Huang, et al. “A Comprehensive Analysis of Low-Power Operation for Beacon-enabled IEEE 802.15.4 Wireless Networks,” *IEEE Transactions on Wireless Communications*, vol.8, no. 11, pp.5601-5611, 2009.

[147] C. Jung, et.al, “Enhanced Markov Chain Model and Throughput Analysis of the Slotted CSMA/CA for IEEE 802.15.4 Under Unsaturated Traffic Conditions,” *IEEE Transactions on Vehicular Technology*, vol.58, no.1, pp.473-478, 2009.

[148] R. Boucherie and N. Dijk, *Queueing Networks: A Fundamental Approach*, Springer Science & Business Media, 2010.

[149] M. Reiman, “Asymptotically exact decomposition approximations for open queueing networks,” *Operations Research Letters*, vol.9, no.6, pp.363-370, 1990.

[150] A. Seidmann, P. Schweitzer, and S. Shalev-Oren, “Computerized Closed Queueing Network Models of Flexible Manufacturing Systems,” *Large Scale Systems*, vol. 12, pp. 91-107, 1987.

[151] P. Denning and J. Buzen, "The operational analysis of queueing network models", *ACM Computer Surveys*, vol. 10, no. 3, pp.225 -261, 1978.

[152] D. Yao and J. Buzacott, “The exponentialization approach to flexible manufacturing system models with general processing times,” *European Journal of Operational Research*, vol.24, no.3, pp.410-416, 1986.

[153] G. Bitran, and R. Morabito, “Open Queuing Networks: Optimisation and Performance Evaluation Models for Discrete Manufacturing Systems,” *Production and Operations Management*, vol.5, no.2, pp.163-193, 1996.

[154] M. Goyal, et al. “A Stochastic Model for Beaconless IEEE 8 .15.4 MAC Operation,” *Computer Communications*, vol. 34, no.12, pp.1460- 1474, 2011.

[155] Y. Huang, et al. “A Comprehensive Analysis of Low-Power Operation for Beacon-enabled IEEE 802.15.4 Wireless Networks,” *IEEE Transactions on Wireless Communications*, vol.8, no. 11, pp.5601-5611, 2009.

[156] P. Park, et al., “A Generalized Markov Chain Model for Effective Analysis of Slotted IEEE 802.15.4,” *in Proc. The 6th IEEE International Conference on Mobile Ad hoc and Sensor Systems*(MASS), pp.130-139, 2009.

[157] T. Czachórsi, et al., “A Diffusion Approximation Model for Wireless Networks Based on IEEE 802.11 Standard,” *Computer Communications*, vol. 33, Supplement.1, pp.S86-S92, 2010.

[158] G. Levitin, *The Universal Generating Function in Reliability Analysis and Optimization*, Springer Science & Business Media, 2006.

[159] H. Kobayashi, B. Mark and W. Turin, *Probability, Random Processes, and Statistical Analysis: Applications to Communications Signal Processing, Queueing Theory and Mathematical Finance*, Cambridge University Press,

2011.
[160] N. Ushakov, *Selected Topics in Characteristic Functions*, VSP, 1999.
[161] Ch. Charalambides, *Combinatorial Methods in Discrete Distributions*, Wiley-Interscience, 2004.
[162] R. Schinazi, *Classical and Spatial Stochastic Processes*, Birkhäuser, 1999.
[163] L. Kleinrock, *Communication Nets: Stochastic Message Flow and Delay*. McGraw- Hill, 1964.
[164] P. Burke, "The Output of a Queuing System", *Operation Research*, vol.4, no.6, pp.699–704, 1956.
[165] W. Whitt "Approximations for Departure Processes and Queues in Series," *Naval Research Logistics Quarterly*, vol. 31, pp.499 -521, 1984.
[166] E. Gelenve, and I.Mitrani, *Analysis and Synthesis of Computer Systems*, Academic, new York, 1980.
[167] T. Czachórsi, et al., *Network Performance Engineering, A Handbook on Convergent MultiService Networks and Next Generation Internet*, Springer, 2011.
[168] M. Girish, et al., "Higher Order Approximations for the Single Server Queue with Splitting, Merging and Feedback," *European Journal of Operation Research*, vol. 124, no.3, pp.447- 467, 2000.
[169] A. Sheikh, et al., "A Unified Approach to Analyze Multiple Access Protocols for Buffered Finite Users," *Network and Computer Applications*, vol.27, no.1, pp.49 –76, 2004.
[170] T. Wan, et al., "Performance and stability analysis of buffered slotted ALOHA protocols using tagged user approach," *IEEE Transactions on Vehicular Technology*., vol. 49, no. 2, pp. 582-593, 2000.
[171] S. Rasool, et al., "An Approximate Analysis of Buffered S-ALOHA in Fading Channels Using Tagged User Analysis," *IEEE Transactions on Wireless Communications*, vol.6, no.4, pp.1320-1326, 2007.
[172] C. Singh, et al. "Performance evaluation of an IEEE 802.15.4 sensor network with a star topology," *Wireless Network*., vol. 14, pp.543 -568, 2008.
[173] F. Shu, and T. Sakurai, "A New Analytical Model for the IEEE 802.15.4 CSMA-CA protocol," *Computer Networks*. vol. 55, no.11,pp.2576-2591,2011.
[174] V. Ramaiyan, et al. "Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness and Multistability," *IEEE/ACM Transactions on Networking*, vol.16, no.5, pp.1080-1093, 2008.
[175] A. Kumar, et al. "New insights from a fixed point analysis of single cell IEEE 802.11 WLANs", *in Proc. IEEE INFOCOM*,2006.
[176] D. Bertsimas, et al., "On the large deviations behavior of acyclic networks of G/G/1 queues," *Annuals of Applied probabilities*, vol. 8, no. 4, pp.1027-1069, 1998.
[177] A.Gut, *Probability: A Graduate course*. Springer, 2005.
[178] H. Takagi, *Queuing Analysis as a Foundation for Performance Evaluation*. North-Holland, 1993.
[179] A. Weiss, et al., *Large Deviations for Performance Analysis: Queues,*

Communications and Computing. Chapman & Hall, 1995.

[180] W. Whitt "The Queuing Network Analyzer", *Bell System. Technol*., vol. 62, no. 9, pp.2779 -2815, 1983.

[181] F. Massey, "The Kolmogorov-Smirnov Test for Goodnees of Fit," *American Statistical Association*, vol. 46, no. 256, pp.68-78, 1951.

[182] Bowman, *Properties of Estimators for the Gamma Distribution*, CRC Press, 1987.

[183] R. Kolisch and S. Hartmann, "Experimental Investigations of Heuristics for RCPSP: An Update," *European Journal Of Operational Research*, vol.174, no.1, pp.23–37, 2006.

[184] M. Herlihy and N. Shavit, *The Art of Multiprocessor Programming*, Elsevier, 2012.

[185] S. El Hajj, et.al, "Random backoff control to thwart malicious behavior in WLANs," *in Proc. The 19th IEEE Local & Metropolitan Area Networks*, (LANMAN), pp.1-6, 2013.

[186] L. Kleinrock and S. Lam "Packet-swtiching in a slotted satellite channel," *National Computer Conference* (AFIPS), pp.703 -710, 1973.

[187] V. Jacobson, "Congestion avoidance and control," *in Proc*. *The ACM conference on Communications architectures, protocols and applications* (SIGCOMM), pp.314 -329, 1988.

[188] S. Foss, et.al, *An Introduction to Heavy-Tailed and Subexponential Distributions*, Springer Science & Business Media, 2013.

[189] M. Crovella, "Performance Evaluation with Heavy Tailed Distributions," *In. Proc. The 7th International Workshop on Job Scheduling Strategies for Parallel Processing*, pp.1-10, 2001.

[190] B. Mandelbrot, *The Fractal Geometry of Nature*. Henry Holt and Company, 1983.

[191] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*, Wiley, 2000.

[192] R. Martinez, et.al, "The Consequences of Heavy Tailed Service Time Distribution on a Basic Queuing Model and its Performance Indicators," *Ingeniería e Investigación*, vol.30, no.2, pp.136-146, 2010.

[193] V. Bharghavan, et.al, "MACAW: a media access protocol for wireless LAN's," *in Proc. the ACM conference on Communications architectures, protocols and applications* (COMM), pp.212–225, 1994.

[194] L. Taijun, T. Tiebin and C. Cheng, "A New Backoff Algorithm for IEEE 802.11 Distributed Coordination Function," *in Proc. The 6th Conference on Fuzzy Systems and Knowledge Discovery* (FSKD), pp.455,459, 2009.

[195] J. Haas and D. Jing "On optimizing the backoff interval for random access schemes," *IEEE Transactions on Communications*, vol.51, no.12, pp.2081-2090, 2003.

[196] K. Jamieson, H. Balakrishnan and Y. C. Tay, "Sift: A MAC Protocol for Event-Driven Wireless Sensor Networks," *in Proc*. Third *European Conference on Wireless Sensor Networks* (EWSN), pp.260-275, 2003.

[197] L. Bononi, M. Conti, and E. Gregori, "Design and Performance Evaluation of an Asymptotically Optimal Backoff Algorithm for IEEE 802.11 Wireless

LANs," *in Proc. The 33rd IEEE Annual Hawaii Conference on System Sciences* (HICSS), pp.1-10, 2000.

[198] H. Anouar and C. Bonnet, "Optimal Constant-Window Backoff Scheme for IEEE 802.11 DCF in Single-Hop Wireless Networks Under Finite Load Conditions," *Wireless Personal Communications*, vol.43, no.4, pp.1583-1602, 2007.

[199] C. Kleiber and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, 2003.

[200] J. Noortwijk, "A survey of the application of gamma processes in maintenance," *Reliability Engineering & System Safety*, vol.94, no.1, pp.2-21, 2009.

[201] D. Dey and C. Rao, *Handbook of Statistics: Bayesian Thinking, Modeling and Computation*, Gulf Professional Publishing, 2005.

[202] M. Nikulin, *Advances in Degradation Modeling: Applications to Reliability, Survival Analysis, and Finance*, Springer Science & Business Media, 2010.

[203] L. Sakhnovich, *Levy Processes, Integral Equations, Statistical Physics: Connections and Interactions: Connections and Interactions*, Springer Science & Business Media, 2012.

[204] F. Steutel and K. Harn, *Infinite Divisibility of Probability Distributions on the Real Line*, CRC Press, 2003.

[205] K. Sunggon, J. Lee and D. Sung, "A Shifted Gamma Distribution Model for Long-Range Dependent Internet Traffic," *IEEE Communications Letters*, vol.7, no.3, pp.124-126, 2003.

[206] F. Chatelain, "Parameter Estimation for Sums of Correlated Gamma Random Variables. Application to Anomaly Detection in Internet Traffic," *in Proc. the IEEE Conference on Acoustics, Speech and Signal Processing* (ICASS), pp.3489-3492, 2008.

[207] S. Gebert, et.al, "Internet Access Traffic Measurement and Analysis," *in Proc. The 4th International Workshop on Traffic Monitoring and Analysis* (TMA), pp.29-42, 2012.

[208] A. Ghosal and S. Madan, "Isomorphs of a class of Queueing systems," *International Journal of Mathematics and Mathematical Sciences*, vol.2, no.1, pp.103-110,1979.

[209] T. Meinguet, *Heavy Tailed Functional Time Series*, Presses university de Louvain, 2010.

[210] P. Dittrich, J. Ziegler and W. Banzhaf, "Artificial Chemistries-a review," *Artificial Life*, vol.7, no.3, pp.225-275, 2001.

[211] C. Langton, *Artificial Life: An Overview*, MIT Press, 1997.

[212] H. Sawai, *Biological Functions for Information and Communication Technologies: Theory and Inspiration*, Springer, 2011.

[213] A. Madureira, C. Reis and V. Marques, *Computational Intelligence and Decision Making: Trends and Applications*, Springer, 2012.

[214] V. Mo, et al., "Bayesian analysis of systems with random chemical composition: renormalization-group approach to Dirichlet distributions and the statistical theory of dilution," *Physical Revision*, vol.65, no.011112,

pp.1-8, 2002.

[215] A. Woody, Ed., R. Hendry, Ed., and P. Needham, Ed., *Philosophy of Chemistry*, Elsevier, 2012.

[216] Henry Watts, *A Dictionary of Chemistry and the Allied Branches of Other Sciences*, Longmans, Green, and Company, 1875.

[217] K.Hosokawa, et al., "Dynamics of self-assembling systems: Analogy with chemical kinetics," *Artificial Life*, vol.1, pp. 413–427, 2004.

[218] B. Straatman, et al., "An Artificial Chemistry-based Model of Economies," *Artificial Life*, vol.11, pp.592-599,2008.

[219] C. Müller-Schloer, H. Schmeck and T. Ungerer, *Organic Computing — A Paradigm Shift for Complex Systems*, Springer, 2011.

[220] N. Rambidi, K. Shamayaev and G. "Peshkov, "Image processing using light-sensitive chemical waves," *Physics Letters*, vol. 298, no.5-6, pp.375-382, 2002.

[221] A. Adamatzky and B. De Lacy Costello, "Experimental logical gates in a reaction-diffusion medium: The XOR gate and beyond," *Physical Review*, vol.66, no.4, 2002.

[222] N. Rambidi and D. Yakovenchuk, "Chemical reaction-diffusion implementation of finding the shortest paths in a labyrinth," *Physical Review*, vol.63, no.2, 2001.

[223] G.Oster, et al., "Chemical reaction networks," *IEEE Transactions on Circuits and Systems*, vol.21, no.6, pp.709-721, 1974.

[224] S. Andrew et al., "Implementation of a multifunctional logic gate based on folding/unfolding transitions of a protein," *Future Generation Computer Systems*, vol.19, no.1, pp. 87–97, 2003.

[225] T. Ray, "Population dynamics of digital organisms," *In Proc. The Workshop on Artificial Life II Video*. 1991.

[226] G. O'Neill, *Technology Edge*, Simon and Schuster, 1985.

[227] A. Adamatzky, *Game of Life Cellular Automata*, Springer Science & Business Media, 2010.

[228] W. Poundstone, *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge*, Courier Dover Publications, 2013.

[229] Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press, 1988.

[230] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer Science & Business Media, 1996.

[231] N. Karayiannis and A. Venetsanopoulos, *Artificial Neural Networks: Learning Algorithms, Performance Evaluation, and Applications*, Springer Science & Business Media, 1992.

[232] A. Turing, "The Chemical Basis of Morphogenesis," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol.237, no. 641, pp.37-72, 1952.

[233] S. Cooper and J. Leeuwen, *Alan Turing: His Work and Impact*, Elsevier, 2013.

[234] B. Kerner, V. Osipov, Autosolitons: A New Approach to Problems of Self-Organization and Turbulence, Springer Science & Business Media, 1994.

[235] K. Mainzer, *Thinking in Complexity: The Computational Dynamics of Matter, Mind, and Mankind*, Springer Science & Business Media, 2007.
[236] P. Lalanda, J. McCann and A. Diaconescu, *Autonomic Computing: Principles, Design and Implementation*, Springer Science & Business Media, 2013.
[237] E. Thompson, *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, Harvard University Press, 2007.
[238] G. Minati, E. Pessa and M. Abram, *Systemics of Emergence: Research and Development, Springer Science & Business Media, 2006*.
[239] Greg Van Alstyne, Robert K. Logan, "Designing for Emergence and Innovation: Redesigning Design," *ARTIFACT*, vol.1, no.2, pp.120-129, 2007.
[240] W. Fontana, and L. Buss, "The arrival of the fittest': Toward a theory of biological organization," *Bulletin of Mathematical Biology*, vol.56, no.1, pp.1-64, 1994.
[241] P. Dittrich and P. di Fenizio, "Chemical Organisation Theory," *Bulletin of Mathematical Biology*, vol. 69, no.4, pp.1199-1231, 2007.
[242] M. Boden, *The Philosophy of Artificial Life*, Oxford Univesity Press, 1996.
[243] D. Parisi, "Artificial Life and Higher Level Cognition," *Brain and Cognition*, vol.34, no.1, pp.160-184, 1997.
[244] M. Feinberg, "Existence and uniqueness of steady states for a class of chemical reaction networks," *Archive for Rational Mechanics and Analysis*, vol.132, no.4, pp.311-370, 1995.
[245] P. Érdi, *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models*, Manchester University Press, 1989.
[246] M. Clugston, and R. Flemming, *Advanced Chemistry*, Oxford University Press, 2000.
[247] G. Chan and S. Sharma, The Density Matrix Renormalization Group in Quantum Chemistry, *Annual Review of Physical Chemistry*, vol. 62, no.2, pp.465-481, 2011.
[248] D. Heckerman. *Learning in Graphical Models*, MIT Press, 1998.
[249] D. Dey, and C. Rao, *Handbook of Statistics: Bayesian Thinking, Modeling and Computation*, Gulf Professional Publishing, 2005.
[250] S. Pollin, et al., "Performance analysis of slotted carrier sense IEEE 802.15.4 medium access layer," *IEEE Transactions on Wireless Communications*, vol. 7, no. 9, pp.3359–3371, 2008.
[251] B. Bougard, et.al, "Energy efficiency of the IEEE 802.15.4 standard in dense wireless microsensor networks: Modeling and improvement perspectives," *in Proc. Design Automation and Test in Europe Conference and Exhibition*, pp. 196–201, 2005.
[252] L. Devroye, *Non-uniform Random Variate Generation*, Springer-Verlag, 1986.
[253] D. Kundu and R. Gupta, "A convenient way of generating gamma random variables using generalized exponential distribution," *Computational Statistics & Data Analysis*, vol.51, no.6, pp.2796-2802, 2007.

[254] G. Marsagla and W. Tsang, “A simple method for generating gamma variables,” *ACM Transactions on Mathematical Software*, vol.26, no.3, pp.363-372, 2000.
[255] M. Wiering, and M. Otterlo, *Reinforcement Learning: State-of-the-Art*, Springer Science & Business Media, 2012.
[256] S. Olariu, Ed. and A. Zomaya, Ed., *Handbook of Bioinspired Algorithms and Applications*, CRC Press, 2005.
[257] A. Ghosh, Ed. and S. Tsutsui, Ed., *Advances in Evolutionary Computing: Theory and Applications*, Springer Science & Business Media, 2003.
[258] B. Goertzel and C. Pennachin, *Artificial General Intelligence*, Springer Science & Business Media, 2007.
[259] M. Cheng, Y. Li and D. Du, *Combinatorial Optimization in Communication Networks*, Springer Science & Business Media, 2006.
[260] Y. Collette and P. Siarry, *Multiobjective Optimization: Principles and Case Studies*, Springer Science & Business Media, 2003.
[261] Z. Geem, Ed., *Music-Inspired Harmony Search Algorithm: Theory and Applications*, Springer Science & Business Media, 2009.
[262] Z. Geem, *Recent Advances in Harmony Search Algorithm*, Springer Science & Business Media, 2010.
[263] M. Gunes, U. Sorges, and I. Bouazizi, “ARA-the ant-colony based routing algorithm for MANETs,” *in Proc. The IEEE Workshop on Workshop on Ad Hoc Networks* (IWAHN), pp.79-85, 2002.
[264] K. Bhargavan, D. Obradovic, and C. Gunter, “Formal verification of standards for distance vector routing protocols,” *Journal of the ACM*, vol. 49, no. 4, pp. 538–576, 2002.
[265] J. Sobrinho, "An algebraic theory of dynamic network routing," *IEEE/ACM Transactions on Networking*, vol.13, no.5, pp.1160-1173, 2005.
[266] J. Andersson, “A survey of multiobjective optimization in engineering design,” Department of Mechanical Engineering, Linköping University, Technical report LiTH-IKP-R-1097, 2000.
[267] T. Cormen, Ed., *Introduction To Algorithms*, MIT Press, 2001.
[268] T. Gonzalez, *Handbook of Approximation Algorithms and Metaheuristics*, CRC Press, 2007.
[269] X. Gandibleux, *Metaheuristics for Multiobjective Optimisation*, Springer Science & Business Media, 2004.
[270] K. Akkaya and M. Younis, “A survey on routing protocols for wireless sensor networks,” *Ad Hoc Networks*, vol.3, no.3, pp.325–349, 2005.
[271] J. Al-Karaki, and A. Kamal, “Routing techniques in wireless sensor networks: a survey,” *IEEE Wireless Communications Magazine*, vol.11, no.6, pp.6-28, 2004.
[272] A. Zungeru, Li. Ang and K. Seng, “Classical and swarm intelligence based routing protocols for wireless sensor networks: A survey and comparison,” *Journal of Network and Computer Applications*, vol.35, no.5, pp.1508–1536,2012.
[273] R. Kulkarni, and A. Forster A; Venayagamoorthy, G.K., “Computational Intelligence in Wireless Sensor Networks: A Survey,” IEEE

Communications Surveys & Tutorials, vol.13, no.1, pp.68-96, 2011.

[274] M. Resende, and P. Pardalos, *Handbook of Optimization in Telecommunications*, Springer Science & Business Media, 2008.

[275] P. Hart, N. Nilsson, and B. Raphael, “A Formal Basis for the Heuristic Determination of Minimum Cost Paths,” *IEEE Transactions On Systems Science and Cybernetics*, vol.4, no. 2, pp.100–107, 1968.

[276] L. Rios and L. Chaimowicz, “A Survey and Classification of A* Based Best-First Heuristic Search Algorithms,” *in Proc. The 20th Brazilian Symposium on Artificial Intelligence* (SBIA), pp.253-262, 2010.

[277] E. Korf, “Depth-first iterative-deepening: An optimal admissible tree search,” *ACM Journal on Artificial Intelligence*, vol.27, no.1, pp.97–109, 1985.

[278] Russell, S., “Efficient memory-bounded search methods,” *in Proc. The 10th European Artificial Intelligence Conference* (ECAI-92), pp.1-5, 1992.

[279] A. Sobeih, et, al. “Finding Bugs in Network Protocols Using Simulation Code and Protocol-Specific Heuristics,” *in Proc. 7th International Conference on Formal Engineering Methods*, (ICFEM) ,pp.235-250, 2005.

[280] K. Rana and M. Zaveri, “A-Star Algorithm for Energy Efficient Routing in Wireless Sensor Network,” *in Proc. The 4th Conference on Network Security & Applications* (CNSA), pp.232-241, 2011.

[281] K. M. Rana and M. A. Zaveri, “ASEER: A routing method to extend life of two-tiered wireless sensor network,” *International Journal Advanced Smart Sensor Network System*, vol. 11, no. 2, pp.1–16, 2011.

[282] R. Korf, “Real-time heuristic search,” *Artificial Intelligence*, vol.42, no.2–3, pp.189–211, 1990

[283] Y. Yu, R. Govindan and D. Estrin, “Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks,” Computer Science Department, University of California, Los Angeles, Technical Report UCLA/CSD-TR-01-0023, 2001.

[284] M. Rossi, M. Zorzi, and R. Rao, “Statistically assisted routing algorithms (SARA) for hop count based forwarding in wireless sensor networks,” *Wireless Networks*, vol.14, no.1, pp.55-70, 2008.

[285] Y. Shang, et.al, "Constraint-based routing for ad-hoc networks," *in Proc. International Conference on Information Technology: Research and Education* (ITRE), pp.306-310, 2003.

[286] L. Busoniu, *Reinforcement Learning and Dynamic Programming Using Function Approximators*, CRC Press, 2010.

[287] J. Boyan and M. Littman, “Packet routing in dynamically changing networks: A reinforcement learning approach,” *in Proc. Advances in Neural Information Processing Systems Conference* (NIPS), pp.671 -678, 1994.

[288] S. Choi and D. Yeung, “Predictive Q-routing: A memory-based reinforcement learning approach to adaptive traffic control,” *Advances in Neural Information Processing Systems Conference* (NIPS), pp.945 -951, 1996.

[289] S. Kumar and R. Miikkulainen, “Dual reinforcement q-routing: An online adaptive routing algorithm,” *in Proc. The Artificial Neural Networks in*

Engineering (ICCN), 1997.

[290] P. Wang and T. Wang, "Adaptive Routing for Sensor Networks using Reinforcement Learning," *in Proc. The 6th IEEE Conference on Computer and Information Technology* (CIT), pp.219-224, 2006.

[291] M. Hauskrecht, "Value-function approximations for partially observable Markov decision processes," *Journal of Artificial Intelligence Research*, vol.13 no.1, pp.33-94, 2000.

[292] Z. Han, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*, Cambridge University Press, 2012.

[293] S. Lasaulce, and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*, Academic Press, 2011.

[294] S. Hargreaves Heap, and Y. Varoufakis, *Game Theory: A Critical Text*, Psychology Press, 2004.

[295] M. Felegyhazi, and J. Hubaux, "Game Theory in Wireless Networks: A Tutorial," Swiss Federal Institute of Technology Lausanne, Technical report LCA-REPORT-2006-002, 2007.

[296] T. Roughgarden, *Selfish Routing and the Price of Anarchy*, MIT Press, 2005.

[297] J. Koronacki, Z. Ras, and S. Wierzchon, *Advances in Machine Learning II: Dedicated to the Memory of Professor Ryszard S. Michalski*, Springer Science & Business Media, 2009.

[298] D. Easley, and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

[299] L. Busoniu, R. Babuska, and B. de Schutter, "A Comprehensive Survey of Multiagent Reinforcement Learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol.38, no.2, pp.156-172, March 2008.

[300] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, 1999.

[301] C. Blum, and D. Merkle, *Swarm Intelligence: Introduction and Applications*, Springer Science & Business Media, 2008.

[302] J. Kennedy, J. Kennedy, and R. Eberhart, *Swarm Intelligence*, Morgan Kaufmann, 2001.

[303] M. Dorigo, and T. Stützle, *Ant Colony Optimization*, MIT Press, 2004.

[304] X. Yu, and T. Zhang, "Convergence and Runtime of an Ant Colony Optimization Model," *Information Technology Journal*, vol.8, no.3, pp.354-359, 2009.

[305] D. Fogel, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, John Wiley & Sons, 2006.

[306] M. Gen, R. Cheng, L. Lin, Network *Models and Optimization: Multiobjective Genetic Algorithm Approach*, Springer Science & Business Media, 2008.

[307] K. Dahal, K. Tan, and P. Cowling, *Evolutionary Scheduling*, Springer Science & Business Media, 2007.

[308] M. Črepinšek, S.Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms: A survey," *ACM Computing Surveys*, vol.45, no. 3, 2013.

[309] T. Ross, J. Booker, and W. Parkinson, *Fuzzy Logic and Probability Applications: Bridging the Gap*, SIAM, 2002.
[310] G. Feng, *Analysis and Synthesis of Fuzzy Control Systems: A Model-Based Approach*, CRC Press, 2010.
[311] A. Kots, and M. Kumar, The fuzzy based QMPR selection for OLSR routing protocol, *Wireless Networks*, vol. 20, no.1, pp.1-10, 2014.
[312] P. Khanpara, "A Review on Fuzzy Logic Based Routing In Ad Hoc Networks," *International Journal of Advanced Research in Engineering and Technology*, vol.5, no. 5, pp.75-81, 2014.
[313] D. Manjarres, et al., "A survey on applications of the harmony search algorithm," *Engineering Applications of Artificial Intelligence*, vol.26, no.8, pp.1818–1831, 2013.
[314] J. Sloboda, *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, OUP Oxford, 2000.
[315] C. Lewis, and P. Lovatt, "Breaking away from set patterns of thinking: Improvisation and divergent thinking," *Thinking Skills and Creativity*, vol. 9, no.1, pp.46– 58, 2013.
[316] C. Moorman and A. Miner, "Organizational Improvisation and Organizational Memory," *Academy of Management Review*, vol.32, no.4, pp. 698-723, 1998.
[317] A. King, and A. Ranft, "Capturing knowledge and knowing through improvisation: what managers can learn from the thoracic surgery board certification process," *Journal of Management*, vol.27, no.3, pp.255–277, 2001.
[318] E. Landgraf, *Improvisation as Art: Conceptual Challenges, Historical Perspectives*, Bloomsbury Publishing, 2011.
[319] S. Franklin, *Artificial Minds*, MIT Press, 1997.
[320] A. Berkowitz, *The Improvising Mind: Cognition and Creativity in the Musical Moment*, Oxford University Press, 2010.
[321] P.Petta, C. Pelachaud, and R. Cowie, *Emotion-Oriented Systems: The Humaine Handbook*, Springer Science & Business Media, 2011.
[322] J. McCormack, and M. d'Inverno, *Computers and Creativity*, Springer Science & Business Media, 2012.
[323] M. Leman, *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*, Springer Berlin Heidelberg, 2011.
[324] S. Ramnath, and B. Dathan, *Object-Oriented Analysis and Design*, Springer Science & Business Media, 2010.
[325] V. Štuikys, and R. Damaševičius, *Meta-Programming and Model-Driven Meta-Program Development: Principles, Processes and Techniques*, Springer Science & Business Media, 2012.
[326] M. Aiello, I. Pratt-Hartmann, and J. van Benthem, *Handbook of Spatial Logics*, Springer Science & Business Media, 2007.
[327] Y. Murayama, *Progress in Geospatial Analysis*, Springer Science & Business Media, 2012.
[328] L. Polkowski, *Approximate Reasoning by Parts: An Introduction to Rough Mereology*, Springer Science & Business Media, 2011.

[329] I. Düntsch, H. Wang, and S. McCloskey, “A relation – algebraic approach to the region connection calculus,” *Theoretical Computer Science*, vol. 255, no.1–2, pp.63–83, 2001.
[330] M. Simon, and M. Alouini, *Digital Communication over Fading Channels*, John Wiley & Sons, 2005.
[331] P. Mitchell, and K. Riggs, *Children's Reasoning and the Mind*, Psychology Press, 2014.
[332] R. Rosen, *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*, Springer Science & Business Media, 2012.
[333] S. Lynch, *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*, Springer Science & Business Media, 2007.
[334] B. Atakan, *Molecular Communications and Nanonetworks: From Nature To Practical Systems*, Springer Science & Business Media, 2014.
[335] B. Xing, and W. Gao, *Innovative Computational Intelligence: A Rough Guide to 134 Clever Algorithms*, Springer International Publishing, 2013.