RESEARCH REPORT

Hierarchical Clustering for Energy Efficient Routing in Wireless Sensor Networks

submitted to the
Department of Information Technology and Systems Management
at the
Salzburg University of Applied Sciences

by:
Kerstin Nageler, BSc

Head of Department:	FH-Prof. DI Dr. Gerhard Jöchtl
Supervisor:	Prof. Dr. (habil) Ulrich Hofmann

Salzburg, October 2012

Affirmation in Lieu of an Oath

Hereby I, Kerstin NAGELER, BSc, born on 06.05.1988 in St. Veit an der Glan, declare in lieu of oath that I composed the following thesis independently and have named all sources and additives which I have used completely.

Salzburg, on October 8, 2012

Kerstin NAGELER, BSc

Matriculation Number

General Information

First- and Lastname:	Kerstin NAGELER, BSc
Institution:	Fachhochschule Salzburg GmbH
Department:	Information Technology and Systems Management
Title of the Master Thesis:	Hierarchical Clustering for Energy Efficient Routing in Wireless Sensor Networks
Keywords:	wireless sensor networks, energy-efficient routing, hierarchical clustering, longer network lifetime, wide network expansion
Supervisor FH:	Prof. Dr. (habil) Ulrich Hofmann

Abstract

Wireless sensor networks highly depend on the efficient usage of energy as it represents a limited but vital resource of sensor nodes. The routing protocol impacts the energy consumption through the routes that are used to forward data to the destination. Sensor nodes consume the majority of their energy for data transmission because the actual energy demand depends exponentially on the distance between the communicating nodes. Therefore this thesis implements a new routing protocol which is based on LEACH and extends it with hierarchical clustering allowing multi-hop transmission. Through the hierarchical multi-hop transmission long distance communication can be realized more efficiently. Simulations showed that through this extension sensor nodes dissipate less energy and the lifetime of the individual nodes as well as the entire network can be increased. The new protocol revealed significant improvements in wide expanded networks while in small area networks it remains on a negligible level.

Contents

List of Tables

List of Figures

List of Abbreviations

BS	Base-Station
CH	Cluster-Head
CSMA	Carrier Sense Multiple Access
DBH-LEACH	Distance-based Hierarchy LEACH
DS-SS	Direct-Sequence Spread Spectrum
LEACH	Low-Energy Adaptive Clustering Hierarchy
LMS	Least Mean Square
MANETs	Mobile Ad-Hoc Networks
MSE	Mean Squared Error
ns-2	Network Simulator Version 2
RSSI	Received Signal Strength Indicator
SNR	Signal-to-Noise Ratio
TDMA	Time Division Multiple Access
uAMPS	Micro-Adaptive Multi-domain Power-aware Sensors
WMN	Wireless Mesh Network
WSN	Wireless Sensor Network

1

Introduction

WSN (Wireless Sensor Network) consist of up to thousands of constrained sensor nodes which are able to organize themselves autonomously. Therefore WSNs face additional challenges that distinguish them from common wireless networks, though many properties are similar to the wireless medium. Sensor nodes are typically designed for particular tasks and equipped with a limited power supply, restricted processing power and memory space. For most sensor nodes batteries are the only power supply, and in many applications it is not possible to replace or recharge them. Therefore efficient use of energy is critical for defining the operational time of the sensor nodes, which impacts the operability of the entire network. For that reason the improvement of energy awareness is topic of research for all components of the network. One of these aspects that has been investigated in recent years is the enhancement of the routing algorithm regarding energy demand. The characteristics of the protocols strongly depend on the requirements of the designated application. Thus many proposed routing algorithms differ greatly in their properties.

1.1. Problem Statement

Sensor nodes consume the majority of energy for data transmission: the critical resource regarding the network lifetime. Furthermore the amount of power required varies exponentially as the distance between the communicating nodes. Predictably, employing WSNs over large areas drains batteries quickly and decreases the operational time of the entire network. The same problem affects networks with low node densities. As a result the possible applications for WSNs are constrained by either network expansion or reduced battery life. Routing algorithms directly impact this behavior as they are responsible for organizing node communications and structuring. Moreover algorithms influence the energy demand for data transmission by defining how data is forwarded and how paths are defined and maintained. There exists great potential to improve the energy efficiency of WSNs and to reduce the energy that nodes dissipate extending the operational time of the entire network.

1.2. Motivation

The motivation for this research is the need to define a new energy efficient routing protocol for WSN that enables wide network expansion and is able to handle low node density. The new protocol is based on a combination of existing ideas and new concepts for hierarchy formation. The initial point of the research is the LEACH (Low-Energy Adaptive Clustering Hierarchy) protocol which is an accepted hierarchical routing protocol for wireless sensor networks. Due to the fact that LEACH has some drawbacks, several extensions of this protocol have already been proposed in order to improve its performance. Examples include TL-LEACH, which introduces a two-level hierarchy to LEACH and MR-LEACH adding multi-hop routing to the protocol.

The major enhancement of the resulting protocol is a hierarchical structuring of LEACH clusters that allows multi-hop transmissions based upon the remaining energy level and the distances between nodes. Through this structure energy-hungry long distance transmissions are replaced by multiple more economical short distance transmissions. Another important enhancement is the decentralized and autonomous organization of the nodes that provides the required WSN flexibility. Simulations are used to compare the effects of network expansion, energy consumption and network operation time of the new protocol with the original one. The new protocol is expected to decrease energy consumption for data transmission in wide expanded networks, enabling a longer network lifetime.

1.3. Outline

This report is structured as follows: Chapter 2 discusses the requirements and challenges of wireless sensor networks. Chapter 3 describes the WSN routing protocol LEACH in detail which serves as a fundament for the new protocol. In chapter 4, the implemented enhancements of the new protocol are introduced. Chapter 5 describes the executed simulations and discusses the results and observations. Chapter 6 concludes the work.

2

Wireless Sensor Networks

A wireless sensor network is defined as a network of numerous sensor nodes performing specified sensing tasks and exchanging sensed data through wireless communication. As there is no network infrastructure data transmission is performed by one or many nodes forwarding data to a destination node, called sink [3]. The sensor nodes are usually small and cheap devices that are responsible for particular sensing tasks only. In order to perform application-specific sensing, a high amount of sensor nodes is distributed in a target area. Sensor nodes are equipped with a communication unit for wireless data transmission, a sensing unit, which can include different physical sensors, a microprocessor and limited memory to perform preprocessing and calculations. The only power supply of a sensor node is a battery, which in many application scenarios is not replaceable. Therefore the extension of network lifetime is one major aspect of research in the area of sensor networks [8][15]. In general, WSNs share several issues with MANETs (Mobile Ad-Hoc Networks) but sensor networks are more than just a special subtype of ad-hoc networks. Thus most protocols developed for ad-hoc networks are not applicable for WSNs. The reason is that these protocols use concepts and techniques like flooding, which do not consider the constrained resources of sensor nodes and the communication requirements of sensor applications [11][3].

2.1. Characteristics: Wireless Sensor Networks vs. Wireless Ad-Hoc Networks

WSNs have several properties and limitations distinguishing them significantly from other types of wireless networks such as MANETs or WMN (Wireless Mesh Network). The aim of a WSN is to deliver application specific information from a target area. Typically the focus is not on a single value measure but on the general condition and development in a target area. Hence multiple sensor nodes can collaborate to achieve better and more reliable results. Furthermore there is always a many-to-one traffic pattern observable in WSNs as all sensed data is transferred to a defined sink. This is another important difference to MANETs where nodes communicate mainly according to the one-to-one pattern. The number of nodes in a WSN typically ranges from several hundred to thousands exceeding significantly the typical size of MANETs. As the area of network deployment is limited, a high density of sensor nodes exists in WSNs. This implies also that occurring events can be detected by multiple sensor nodes. For that reason redundancy and correlation is observable among data sensed by nodes located close to each other. Sensor nodes are highly constrained by available power, as they typically are battery-powered. In many applications, it is not possible to replace or recharge the batteries due to the environment nodes are deployed in. Therefore operability of the entire network depends on the energy dissipation of the sensor nodes. Nodes also need to be able to configure the network autonomously and handle topology changes caused by either mobile sensor nodes or nodes running out of energy [15].

Protocols for WSNs need to adapt to these characteristics to ensure an efficient use of the constrained resources. Furthermore, reducing overhead, adjusting to the data-centric approach of data forwarding and the many-to-one approach are important issues for protocols in WSNs. Unlike in MANETs, nodes do not compete with each other for bandwidth; nodes in WSNs cooperate in order to provide related data to the application. Transmission delay and data throughput are usually not the major metric for WSNs. As all network nodes depend on batteries the efficient usage of energy becomes one of the most critical tasks in WSNs. The lifetime and operability of the entire network depends on the way of energy consumption of the sensor nodes [9].

2.2. Applications of Wireless Sensor Networks

As a result of the availability of low-cost sensors with wireless networking capabilities and the support of fast and easy deployment in various environments, a widespread spectrum of applications for WSNs has emerged. The following list presents an overview of the fields of applications for WSNs [15].

- **Monitoring the Environment:** WSNs can be used to monitor various environmental parameters. This applies to monitoring the development or behavior of wild animals or plants as well as to the surveillance of the quality of water or air with respect to pollution, chemical hazards or temperature. Another application is disaster detection; for example detecting fires in a remote forest.
- **Military Applications:** Due to easy deployment, the number of applications for

WSNs in the military context is increasing strongly. Possible applications are monitoring of battlefield activities or remote detection of biological, nuclear or chemical attacks. Further WSNs can be used for surveillance of buildings or facilities and for navigation and coordination of unmanned vehicles.

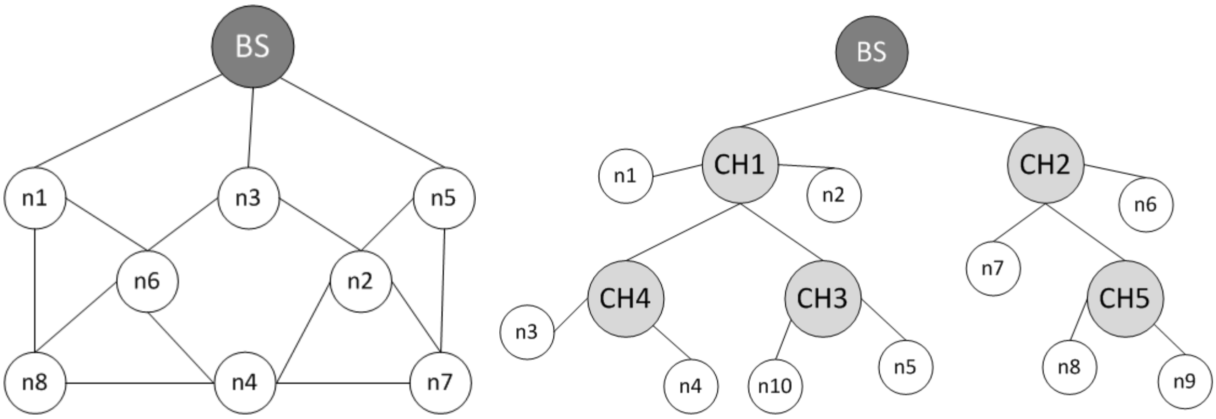
- **Health Care Applications:** WSNs can also improve the quality of health care and reduce the load on staff for monitoring patients. For example, it can be applied to monitor the patients' behavior and alert a doctor in case patients need help. Currently nurses need to enter the room of patients periodically or even stay in the room to monitor patients. Thus the usage of WSNs makes the work of nurses and doctors easier. One step further is to connect the sensing of vital signs to environmental monitoring in order to be able to send an appropriate treatment to the place of an emergency.

More areas of application for WSN range from industrial process control, surveillance to a smart home intelligence [15].

2.3. Network Architectures

The application of a WSN defines the requirements for protocols and also impacts the network architecture. Independent from these factors, energy efficiency remains the most important factor influencing the network lifetime. In general, energy can be wasted or conserved through measures at all levels of the network from the physical layer through the MAC and network layer reaching also to the application layer [3]. Due to the different requirements and constraints of WSNs compared to other wireless networks like MANETs or WLANs their protocols cannot be applied to WSNs. The reasons are the limited computation power, memory and energy of sensor nodes. Therefore, it is necessary to provide a separate protocol stack for WSNs. As sensor networks consist of a high number of randomly spread nodes, network control and management becomes complex and requires numerous protocols handling tasks like medium access, synchronization, self-configuration and routing [14].

Approaches: There are different approaches which can be applied to achieve optimum network performance. The evaluation of energy consumption of sensor nodes showed that communication requires the predominant part of the nodes energy. For example, transferring one bit to destination 100 meter away requires the same amount of energy as performing 3000 instructions on a microcontroller [10]. Furthermore, it turns out that data transmission is much more energy demanding than receiving. Moreover, the energy demand for sending data depends on the distance between sender and receiver; it increases exponentially with distance. Based on this knowledge, the reduction of the amount of data and the distance for data transmission to the sink represents high energy conserving potential. Therefore network architectures distinguish between single-hop and multi-hop networks. In single-hop architectures every node transmits sensed data directly to the sink. This causes high energy dissipation for distant nodes in order to reach the sink, meaning that these nodes run out of energy early. Contrary to this, multi-hop architectures use multiple short-distance transmissions via intermediate nodes between the distant source and the sink. This approach is expected to save energy. Two different types of multi-hop network architecture are distinguished: flat and hierarchical which are showin in figure 2.1 [14].



(a) Flat Architecture (b) Hierarchical Architecture

Figure 2.1.: Comparing a flat and a multi-hop clustering network architecture [14]

- **Flat Architecture:** The flat network architecture is characterized by the fact that all nodes provide the capability to forward data of other nodes. There is no specific structure how data is forwarded, but each node is able relay data to a neighboring node instead of directly transmitting it to the sink. The data gathering is performed through data-centric routing where the sink sends queries through flooding [14].
- **Hierarchical Architecture:** Different from the flat architecture, in hierarchical networks nodes are arranged in a structure. Commonly clustering is used to setup a hierarchical structure among the sensor nodes. Inside a cluster one node is selected to become CH (Cluster-Head). This node is responsible for forwarding data from the entire cluster to the sink. Furthermore, the CH may process or aggregate the cluster data, reducing the quantity of data it has to transmit. Due to these additional tasks cluster-heads dissipate more energy. But it also allows other cluster nodes save on transmission energy because these nodes only need to transmit over the short distance to the CH. Therefore it is necessary to rebuild clusters and choose different cluster-heads periodically to balance load and energy consumption among all nodes. Otherwise CHs would run out of energy early. The dynamic cluster setup and effective cluster-head selection represents a main challenge in decentralized

networks. Thus various strategies to solve this issue are introduced in different routing algorithms. Another differentiation among hierarchical strategies can be made according to the way the CH transmits data to the sink. One way is single-hop clustering, where every cluster-head transmits directly to the sink independent from its distance to the sink. Another is multi-hop clustering, where cluster-heads are able to relay their data through neighboring cluster-heads [14].

2.4. Design Issues in Wireless Sensor Networks

In wireless sensor networks the design and architecture is always influenced by the application. On the other hand the architecture has a strong impact on the performance of protocols and on the network lifetime. The key issues are the following:

- **Network Dynamics:** The majority of applications get by with stationary sensor nodes that remain at the same place for the whole network operation. But for some applications mobility of sensor nodes or the sink is required. Moreover, monitored events might be static or dynamic. Dynamic events like target tracking require periodic transmission to the sink. To the contrary, monitoring of static events like fire detection in a forest only requires a reactive transmission whenever events are detected. In general, mobility increases the complexity of routing and network management significantly. Algorithms have to consider mobility and handle changing external factors like interference and noise on the wireless channel [1][14].
- **Network Deployment:** The topology of a WSN depends on the deployment of nodes but it strongly influences the performance and efficiency of protocols. Typically sensor nodes are deployed in a random or ad-hoc manner. Therefore it is necessary that nodes are self-organizing and self-configuring to dynamically set up a working network. As nodes are not distributed uniformly the network needs to be organized or structured through methods such as clustering to ensure connectivity and energy awareness in the network. The second possibility is a deterministic deployment, where every node is assigned to a fixed position and paths between nodes are defined in advance. The deterministic approach produces less control overhead but it is not adaptable to changes and can only be used for few applications. Hence handling random topologies is more complex and produces more overhead but it is more flexible and simplifies the deployment [1][11].
- **Energy Considerations:** The energy dissipation for data transmission is related to distance between nodes. The demand for energy may increase also with obstacles in the transmission range of the sensor nodes. Data transmission can be done in a multi-hop or single-hop manner. Using single-hop transmission every node directly

sends the data to the target or sink node. This is a simple method which is applicable only for networks with a small expansion or a low node density. On the other hand multi-hop architecture utilizes several intermediate nodes to forward the data through short distance transmissions. Through the usage of multi-hop transmission wider ranges can be covered more effectively. The disadvantage of multi-hop routing is the overhead produced by the route management and maintenance [1][14].

- **Data Delivery Model:** There are four primary data delivery models for wireless sensor networks: continuous, event-driven, query-driven and hybrid. Continuous data delivery implies periodic data transmission from the sensor nodes to the sink. Event-driven and query-driven delivery is triggered by an occurring event or by a query of the sink. In some networks a hybrid of continuous and event-driven or query-driven delivery model is implemented. The selection of the data delivery model is related to the application but it influences also the stability and energy consumption of a routing protocol [1].
- **Node Capabilities:** Sensor networks either consist of a homogenous or a heterogeneous set of nodes. In a heterogeneous network, nodes may be equipped with different resources or sensing capabilities. Therefore it is reasonable to assign more powerful nodes with more complex and energy demanding tasks like routing or aggregating. In contrast in homogenous networks all nodes possess the same resources. Therefore homogenous nodes distribute tasks evenly among each other in order to achieve balanced energy consumption. The heterogeneity of nodes increases the complexity of routing as not all nodes are able to perform the same tasks [2][1].
- **Data Aggregation:** Sensor nodes located near to each other frequently generate redundant data when sensing the same events. In order to reduce the number of required transmissions nodes perform data aggregation using functions like suppression, min, max or average. This is based on the fact that data transmission represents the most energy consuming task of sensor nodes. Therefore data aggregation is a method to optimize the traffic, save energy and extend the operation time of the network [1].
- **Sink Configuration:** A sink is a node that collects all sensed data from a WSN. Particular applications require either one or multiple sinks in a network, which may be stationary or mobile. A single stationary sink is easy to manage but causes hotspot effects. These occur when the nodes nearest to the sink are used frequently to forward data and therefore consume more energy which shortens the lifetime of these nodes. Mobile sinks balance the load for relaying on many nodes as they move but also complicate network control. The advantage of multiple sinks is load balancing network traffic and avoiding hotspot effects [14].

3

The LEACH Protocol

LEACH is a cluster-based, application-specific WSN protocol designed for monitoring the environment. Its aim is to make nodes self-organizing by using a randomized cluster formation. Further it is able to preserve energy on applying low-power transmissions and data-processing for reducing data to transmit. The data-processing employed by LEACH is an aggregation of sensor node data. Sensor nodes deployed in the same area sense highly correlated data because these nodes are exposed to the same environmental events. Therefore data sensed in a cluster can be aggregated without losing important information and the amount of data transmitted can be cut down to a minimum. In combination with the assumption, that data-processing is less power demanding than communication significant energy savings can be achieved [5].

3.1. General Protocol Flow of LEACH

In general the operation of LEACH is split into rounds consisting of two phases each. Every round starts with a setup-phase, which defines clusters for the entire round. Therefore all nodes calculate the probability to become cluster-head and some nodes declare themselves CH for the round. All nodes that remain non-cluster-heads determine the nearest cluster-head and join the cluster of that node. The second phase is called steady-state phase and represents the actual communication. The steady-state phase itself is split into frames. During each frame all cluster-nodes send data once to their CH. The cluster-head is responsible to collect data from all nodes in its cluster, perform data aggregation and transmit the resulting data to the base-station at the end of every frame. Figure 3.1 illustrates the timeline of LEACH consisting of several rounds. Each round starts with the setup-phase represented by the white fields. The steady-state operation consists of multiple frames. The cluster-head is also responsible for managing the communication inside a cluster. As the data exchange within clusters is organized though TDMA (Time Division Multiple Access), the CH creates and distributes a schedule defining a time slot

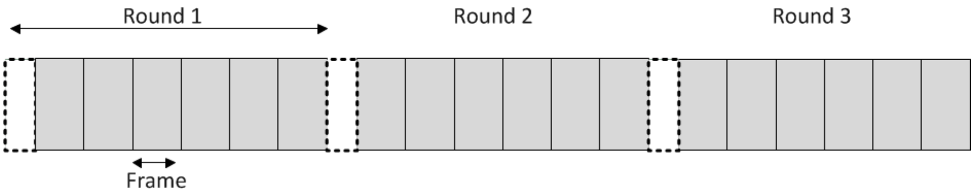


Figure 3.1.: Timeline of LEACH [5]

for every node. Based on the schedule nodes are able to determine the exact time when to send data to the cluster-head. In the meanwhile nodes remain in sleep state and thereby save energy. Due to this distribution of tasks the cluster-heads consume more energy than ordinary nodes. In order to prevent CHs from draining the battery early, the role of the cluster-head is changed every round. The aim of this approach is to distribute the power dissipation evenly among all nodes of the network and thereby extend the complete network lifetime [5].

3.2. Cluster Formation Phase

The cluster formation in LEACH is a decentralized process and defines clusters for a round. At the beginning of each round every node decides independently whether or not it becomes cluster-head. No additional communication between nodes and the base-station or other nodes is necessary to define the cluster-heads. Therefore the nodes calculate the probability of becoming CH for round $k+1$ which starts at time t as shown in equation 3.1. This process intends to distribute the time being a cluster-head equally among all nodes in order to spread energy consumption evenly and keep all nodes alive as long as possible [5].

$$P_i(t) = \begin{cases} \frac{k}{N - k * \left(r \mod \frac{N}{k}\right)} & : C_i(t) = 1 \text{ was not CH} \\ 0 & : C_i(t) = 0 \text{ was CH within the last } \left(r \mod \frac{N}{k}\right) \text{ rounds} \end{cases} \tag{3.1}$$

The expected number of cluster-heads per round is defined as k. Further N represents the total number of nodes in the network. Consequentially every node is intended to become cluster-head once within $\frac{N}{k}$ rounds. Hence nodes may select themselves only as CH that have not been cluster-head within a period of $r \mod \frac{N}{k}$ rounds. Applying this probability avoids unbalanced power draining due to nodes becoming cluster-head with different frequency. k is defined based on the assumption that every node transmits data within every frame and nodes are equipped with equal starting energy. Equation 3.2 outlines the derivation of the expected number of CH in order to achieve an roughly equal level of energy at all nodes after $\frac{N}{k}$ rounds [5].

$$\begin{aligned} E\left[\#CH\right] &= \sum_{i=1}^{N} P_i\left(t\right) \\ &= \left(N - k * \left(r\frac{N}{k}\right)\right) * \frac{k}{N - k * \left(r \mod \frac{N}{k}\right)} \\ &= k \end{aligned} \tag{3.2}$$

Moreover the equations suppose that all sensor nodes provide sensor data to transmit in every frame, which means that every node sends the same amount of data. Only in this case the even distribution of the probability for becoming CH is reasonable [5].

The expected number of CHs in a WSN can be optimized in order to reduce the energy consumption. The analytical estimation of an ideal k depends on the used energy models and the network topology. For LEACH the network area is defined as a $M * M$ square containing N uniform distributed sensor nodes. Another important factor is the location of the BS (Base-Station) which is defined to be outside of sensor area. The fundamental energy models are the Free-Space model and the Two-Ray Ground Reflection model presented in 4.2. Based on these models and the assumptions mentioned here, [5] proved that for an example with $N = 100$ nodes and $M = 100$ meter 5 isthe ideal number of CH [5].

The process of cluster formation starts with nodes defining themselves as CHs according to the probabilities described in equation 3.1. In order to set up clusters every designated CH broadcasts an advertisement. The new CHs send this message with a transmission power high enough that every node can receive it, using CSMA (Carrier Sense Multiple Access). This method eliminates the hidden terminal problem and at the same time ensures that all nodes are able to find a CH. The high transmission power required for the advertisement is not critical regarding the nodes total energy dissipation because the advertisement is a short message transmitted only once in a round. All non-cluster-head nodes collect all advertisements of the CH and store the ID of the potential CH locally. Additionally nodes determine the distance to a CH, thereto the nodes evaluate the signal strength of the received advertisements. This method of distance estimation is only possible when propagation channels are symmetric. The nodes select the best CH which is the one located the nearest to the node, as the lowest amount of energy for data transmission is required. Because of the used method to estimate the distance from the signal strength, it is also possible that a CH physically further away from a node is a better choice than another CH nearer to the node where the transmission is interfered by obstacles or environmental factors. After deciding for a CH a node must inform the chosen CH that it joins the cluster. For that reason all ordinary nodes send a join-request (join-REQ) packet using also the CSMA protocol to the selected CH. Even though nodes are aware of the required transmission power to reach the CH they use the high transmission power for sending the join-REQ message. It turned out that is more energy-efficient sending also this short message with high power instead of using other access methods like RTS-CTS to avoid collisions and hidden terminals. When the CHs receive join-REQ packets clusters are fixed for the entire round. As the CH is responsible for coordinating the communication within its cluster it defines a TDMA schedule and sends it to all cluster nodes. The schedule defines sending slots for each node in every frame. Nodes can go to sleep mode and turn off the radio unless they are in the sending slot and as a result save energy when they are not intended to communicate. This step finalizes the set-up phase and initiates the steady-state phase [5].

An schematic overview of the steps in the cluster-formation phase of LEACH is presented in figure 3.2.

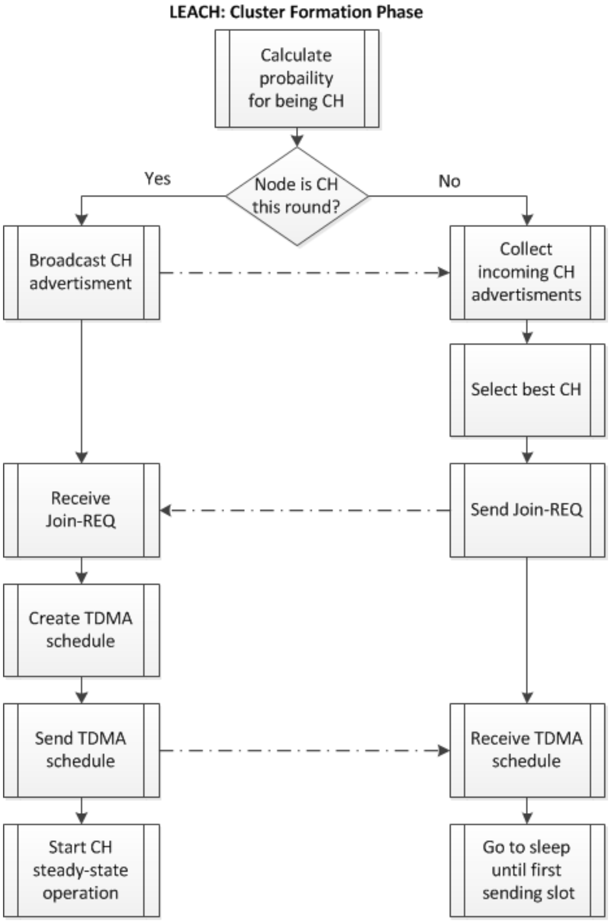


Figure 3.2.: Steps of the cluster-formation phase [5]

3.3. Steady-State Phase

In the steady-state phase nodes perform sensing and transmit data to the BS. The phase is split into frames, each frame contains a sending slot for every cluster node. The length of a frame depends on the number of nodes in a cluster. The slot time reserved is constant for all nodes. The decentralized cluster setup algorithm does not guarantee that the actual number of CHs each round corresponds with the expected number of k. The algorithm neither ensures CHs being spread evenly in the sensing area. This implies that the number of nodes in a cluster may vary strongly and with it the length of a frame and the number of frames in one round. Basically CHs remain in receiving mode for the duration of a complete frame and collect data of all nodes in their cluster. At the end of each frame CHs perform data aggregation and afterwards transmit aggregated data directly to the BS applying CSMA(see chapter 3.4). CHs may be located anywhere in the network, also on the far other side of the network area than the BS. Therefore it is possible that data transmission to the BS requires a high amount of energy and drains batteries of CHs fast [5].

The communication within a cluster operates with TDMA, which is collision free, accomplishes low latency and provides efficient bandwidth and energy usage. In general CHs create a schedule that defines the order of the cluster nodes transmitting data to their CH. Every frame passes through the schedule once, and after the last node CHs perform data aggregation and transmit data to the BS [5]. Figure 3.3 illustrates the structure of two frames in a round.

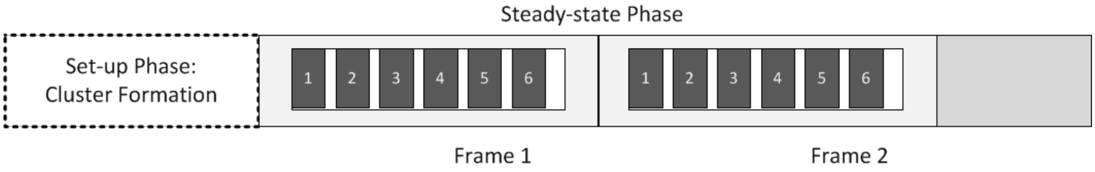


Figure 3.3.: Structure of frames in LEACH [5]

However, this method does not consider surrounding clusters communicating at the same time. Due to a low physical distance it is possible that the intra-cluster communication (node to CH) interferes with communication in neighboring clusters. Collisions may occur and corrupt the data transmitted making it unusable for the receiver. A solution for this problem is DS-SS (Direct-Sequence Spread Spectrum). In fact, the approach implemented in LEACH is called transmitter-based code assignment which refers a unique spreading code to each cluster. All nodes in a cluster use this spreading code for transmitting data to the CH. In order to receive data properly the CH correlates the incoming signal with the spreading code. Nodes receiving signals from nodes of other clusters that used another spreading code perceive these signals as noise. Therefore nearby data transmission of other clusters does not interfere or corrupt the communication of surrounding clusters. Additionally cluster nodes adopt the sending power to their distance to the CH helping to save energy but also reducing the overlapping transmissions. The advantage of this approach is the simplified filtering of received packets based on one spreading code for the CH despite reducing the number inter-cluster collisions to a minimum [5].

The communication between CHs and the BS utilizes CSMA incorporating a fixed spreading code. Therefore CHs must sense the medium to ensure the wireless channel is free before starting to transmit. In case a node detects the channel to be occupied it must wait until the channel is free again and an additional random delay to avoid multiple nodes starting to send simultaneously. The spreading code is also used to ensure the CH-to-BS transmission not interferes with any other intra-cluster communication happening at the same time [5].

3.4. Data Correlation and Aggregation

There are essential differences between data exchanged in typical wireless networks such as MANETs or WLANs and data in WSNs. In typical wireless networks it is important to transmit individual packets as these data is usually part of one-to-one traffic. In contrast in WSNs all network nodes perform a common application. Thus not single sensor values are crucial but the evaluation of the development of sensed values. As LEACH is designed for monitoring applications with periodic sensor updates the user needs to be informed about events that occur concluded from the evaluation of sensor values. For that reason methods of data aggregation are applied to convert a large amount of individual sensor data into less but meaningful data. Data aggregation and data fusion improves also the quality of the data as it eliminates outliers and produces more reliable results emphasizing communalities out of multiple unreliable sources. The aggregation of data is either performed at the BS as it is a central point that collects all data or locally at the CH for each cluster. Calculations to determine the suitable type of data aggregation for a WSN depends on the fundamental energy-dissipation models for processing and communication of the sensor nodes. When considering the assumption, that data processing requires essentially less energy than data transmission the use of local data aggregation at the CH becomes reasonable. The aggregation of data at the CH reduces the amount of data transmitted to the BS and therefore saves energy at the CH [5].

Beamforming is a common algorithm for data aggregation. It combines the signals of multiple sensors using a weighting filter. The function of weighting filters is to optimize characteristics of the data such as the MSE (Mean Squared Error) or the signal-to-noise ratio. Several algorithms to determine suitable weighting filters exist, for example the LMS

(Least Mean Square) error approach or the maximum power beamforming algorithm. In principle these algorithms differ in the provided tradeoffs between data quality and energy saving. In general the choice of the weighting filter depends on the application specific requirements for the WSN as no ideal solution exists [5].

In order to perform data aggregation without losing significant information data needs to be correlated. In WSNs sensor nodes are frequently randomly distributed in the area of interest. Typically the node density in a network is high causing multiple overlapping sensing areas. Hence occurring events are detected by more than one sensor nodes. The beamforming aggregation in LEACH assumes that events are detectable within a $2p$ distance. This implies that in a cluster with a diameter less than $2p$ sensed data correlates allowing to perform aggregation with the ratio of L:1 (L represents the number of nodes in the cluster) [5].

4

Extensions of LEACH

The analysis of the original LEACH protocol and existing extensions initiated the idea of creating a new extension of LEACH. The aim of this new protocol is to overcome some drawbacks and weaknesses of the original protocol. The major change is to enable the CHs to set up a distance-based multi-hop hierarchy, therefore the new protocol is called DBH-LEACH (Distance-based Hierarchy LEACH).

The modifications introduced in this thesis are based on the original LEACH protocol as defined in [5]. The implementation of the extensions employs code of LEACH for ns-2 (Network Simulator Version 2) provided by the MIT uAMPS (Micro-Adaptive Multi-domain Power-aware Sensors) project [6]. The objective of the extensions is to improve the network lifetime and the supported network expansion of LEACH by adding the capability to create multi-hop hierarchies for data transmission. The effects of the changes are evaluated through simulations.

4.1. Hierarchical Cluster Formation and Multi-Hop Transmission

This section presents enhancements in the cluster-formation phase made in DBH-LEACH in order to enabled CHs to build a hierarchical structure among each other. The aim is to save energy on the required transmission power for direct data transmission between CHs and the BS. Due to the CH selection method of LEACH, nodes in the boundary area of the network are as probable to become CH than nodes in the center of the network. Consequently, these nodes run out of energy earlier as they require a higher amount of energy to transmit data to the BS than other nodes nearer to the BS. The relation between node distance and required transmission power depends on the applied radio energy and radio propagation model that are described in detail in chapter 4.2.

In order to enalbe CHs to structure themselves hierarchically new steps and packets are added to the cluster-formation process of LEACH. The following paragraphs describe these steps in detail.

4.1.1. Distance-To-BS Advertisement

A fundamental requirement for setting up a distance-based hierarchy is that every single node is aware of its distance to the BS. Based on this knowledge nodes are able to find the best choice for the next-hop CH and compare it with the distance for direct transmission. Nodes are able to determine distances only by measuring the signal strength of incoming packets when they know the transmission power used by the sender. As LEACH does not use confirmation packets, nodes never receive packets from the BS and are therefore not able to calculate the distance to it. The solution of this problem is an additional packet called Distance-To-BS advertisement. This packet contains only header information and a flag declaring it as Distance-To-BS advertisement. The BS broadcasts this packet using the maximum transmission power at beginning of the network operation. Every node is intended to receive this packet, determine its clearance from the BS based on it and store this information. If a node is not able to receive the advertisement it uses the maximum distance which is defined by default. In order to avoid unnecessary overhead the Distance-To-BS advertisement is transmitted only once in the entire network operation. This implementation does not consider nodes that join the network later. The distance information is used for the hierarchy formation as specified in the next sections.

4.1.2. Extended Cluster-Head Advertisement

The CH-advertisement (CH-ADV) is the first packet exchanged in the cluster-formation phase of LEACH. Originally, nodes that have decided to become CH in a round broadcast this packet containing their node-ID. DBH-LEACH adds information about the distance of the node to the BS to the advertisement. Moreover nodes append information about their remaining energy to the CH-ADV packet. This information is intended only for other nodes that are CH in the same round. For that reason also all CHs have to listen for CH-advertisements broadcasted by other CHs. When receiving these packets, every CH maintain lists of all other CHs at this round, their direct distance to the BS, the distance to this node and the remaining energy of a particular node. Based on this information nodes can calculate the optimal next-hop node for transmitting data to the BS.

4.1.3. Determining the Best Next-Hop CH

The next step in setting up a hierarchical multi-hop transmission structure among cluster-heads is determining the ideal next-hop CH. In a WSN with the conditions stated in **??** the best choice is the node that requires the least amount of energy for transmitting over the complete path to the BS.

The required power for transmitting data to a destination depends on the distance between transmitter and receiver. The fundamental radio energy models describing the relation between distance and transmission power are the Free Space model and the Two-Ray Ground Reflection model. These models define, that within a certain boundary the sending power is proportion to d^2. When the distance to transmit exceeds this boundary, the sending power is proportional to d^4 [5]. Both models and relevant calculations are explained in chapter 4.2.

Based on these models CHs can determine whether it is reasonable to transmit via another CH to conserve energy. In case there are multiple nodes representing a more energy efficient link to the BS than a direct transmission does, the one with the lowest energy demand for the entire transmission is selected. The decision is based on the energy required for transmitting to the next-hop node and the energy the next-hop CH needs for directly transmitting to the BS. This approach does not consider that a next-hop CH may use another next-hop CH itself. In case the next-hop CH uses another CH for transmitting data to the BS, the actual energy for the transmission is less than calculated by the other CHs. One disadvantage of this approach might be, that nodes select a next-hop CH which performs a direct transmission to the BS that is more expensive regarding energy than another CH that uses multi-hop transmission itself. Since the cluster formation works decentralized, every node decides autonomously about the ideal next-hop CH. As DBH-LEACH is time-synchronized, all nodes perform this task simultaneously which makes it impossible for CHs to know about next-hop choices made by other CHs and to consider it in the next-hop selection process.

Energy considerations when making a cluster-head decision: During the development of DBH-LEACH the awareness emerged, that multi-hop transmission may have the side-effect of an increased data loss. The reason was that nodes were choosing next-hop cluster-head nodes with too little remaining energy. Due to the higher energy dissipation as CH and also as receiver of next-hop messages the node ran out of energy while being a next-hop CH. As a result, additional to the data of the nodes cluster also all data of those clusters that employed the node as next-hop CH was lost. Furthermore other CHs are not able to detect that a next-hop node is unavailable and continue transmitting data to this node. This happened frequently when the network reached a late phase in its operation and many nodes run at a low level of energy. In order to avoid losing a high amount of information this way, additional criteria for the next-hop CH selection regarding the remaining energy of potential next-hop nodes have been introduced. As information about the energy level of CHs is exchanged through the extended CH-ADV message (4.1.2), nodes are able to compare the energy levels. A threshold was introduced allowing nodes only to choose CHs with at least 25% of the initial nodes energy left. Simulations showed, that below this level it is very probable that these nodes run out of energy during the next round as a CH.

4.1.4. Joining the Next-Hop CH

When a node succeeded in finding an appropriate next-hop CH, the node has to inform this CH about its choice. This is done through the Join-Info packet introduced for that reason. For the next-hop CH it is necessary to get this information in order to reserve a spot in the schedule for each of these additional nodes. The placement of the other CHs in the schedule is a major challenge to make the multi-hop transmission work and is described in section 4.1.5.

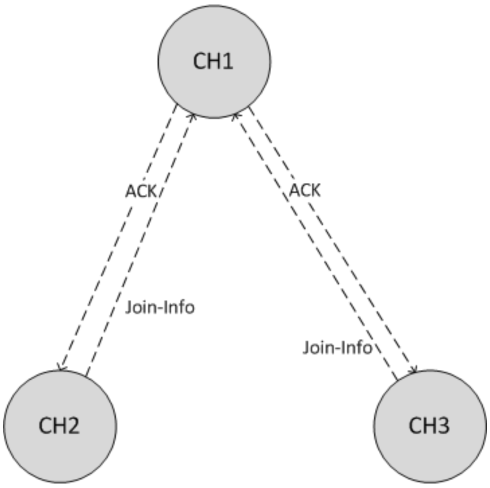


Figure 4.1.: CH2 and CH3 join CH1 as next-hop CH

After successfully integrating all CHs that sent a Join-Info packet into the schedule , the next-hop CH responds to these nods with an ACK packet. Figure 4.1 visualizes the packets exchanged by CH2 and CH3 for joining CH1 as next-hop CH. Nodes that receive the ACK packet are accepted by the selected next-hop CH. These nodes reserve a corresponding slot in their own schedules for sending data to the next-hop node. The ACK packet also contains the number of nodes in the schedule of the next-hop CH. This number is important for the nodes to ensure the length of their schedules fits the one of the next-hop CH. Furthermore it is necessary to make sure nodes are really sending in the slot the next-hop CH expects to receive data from this node. Otherwise collisions with nodes in the cluster of then next-hop CH (CH1) that send according to the local schedule occur.

4.1.5. Scheduling for Multi-hop Communication

The entire communication within a cluster is defined by a TDMA schedule created at the beginning of each round. This schedule defines an order of all nodes in a cluster. Based on this order every node is able to calculate the time when it is expected to transmit data to the CH. Nodes can go into sleep mode to save energy until this time. This is one of the significant advantages of TDMA. At the same time the schedule represents a disadvantage when trying to introduce additional communication [13]. Usually CHs expect to receive data from all of their cluster nodes following the order in the schedule, providing a fixed time slot for every transmission. After running through the schedule once and receiving one set of data of all cluster nodes, aggregated data is transmitted directly to the BS. In order to enable CHs to transmit data to a next-hop CH this next-hop node must expect to receive data from these nodes and reserve a slot in its schedule for all previous-hop CHs. Even though next-hop transmissions replace sending to the BS, it is not possible to perform them within the same time slot. There are several reasons for this. First, the communication with the BS uses CSMA because there are only few nodes that need to transmit data to it and these nodes are sending after different time periods. Second, CHs receive data from tightly sending nodes, hence any additional transmission causes collisions. Third, the number of nodes within a cluster may vary excessively. As a result the CHs reach the slot for transmitting to the BS at different times and frequencies. In order to establish a working communication between CHs, the next-hop CH must expect receiving data from another CH in the same slot as the previous-hop CH sends data to the next-hop CH. Therefore both have to provide an additional slot in their schedule for the next-hop communication, which includes following steps. As soon as the node sends a Join-Info packet indicating that it wants to use a certain CH as next-hop, both nodes must reserve an additional time slot at the same position of their schedule for the next-hop transmission. Preventing additional communication for defining this slot the position is selected according to something unique and commonly known, in this case the unique spreading-code of each CH was chosen. The spreading-code can be determined by every CH due the complete list of all other CH it keeps. Figure 4.2 illustrates the allocation of the slot in the schedule of both sides of a multi-hop transmission. In this example CH3 with the spreading code 3 selects CH1 as its next-hop CH. Both reserve the slot with the index 3 in their schedule for CH3 transmitting to CH1. The node that originally possess this slot and all following nodes are pushed back to the next index.

Moreover it is crucial to consider the different number of nodes in both schedules. Otherwise nodes become desynchronized after the first round. The CH with the lower amount of nodes in its cluster starts the second round earlier than its counterpart with more cluster nodes. To avoid this it is necessary to ensure both schedules remain synchronized. Two cases can be distinguished: Either the lower level CH (CH3) or the next-hop CH (CH1) has more nodes in its cluster. In case the lower-level CH is the one with less nodes in the cluster the schedule is simply filled up with dummy-nodes to meet the amount of nodes of the next-hop CH. When the situation is the other way around the lower-level CH adds dummy-nodes to fit a multiple of the size of the next-hop CH schedule. In this case the lower level CH sends only every n round to the next-hop CH in the other rounds the slot of the next-hop CH schedule remains unused.

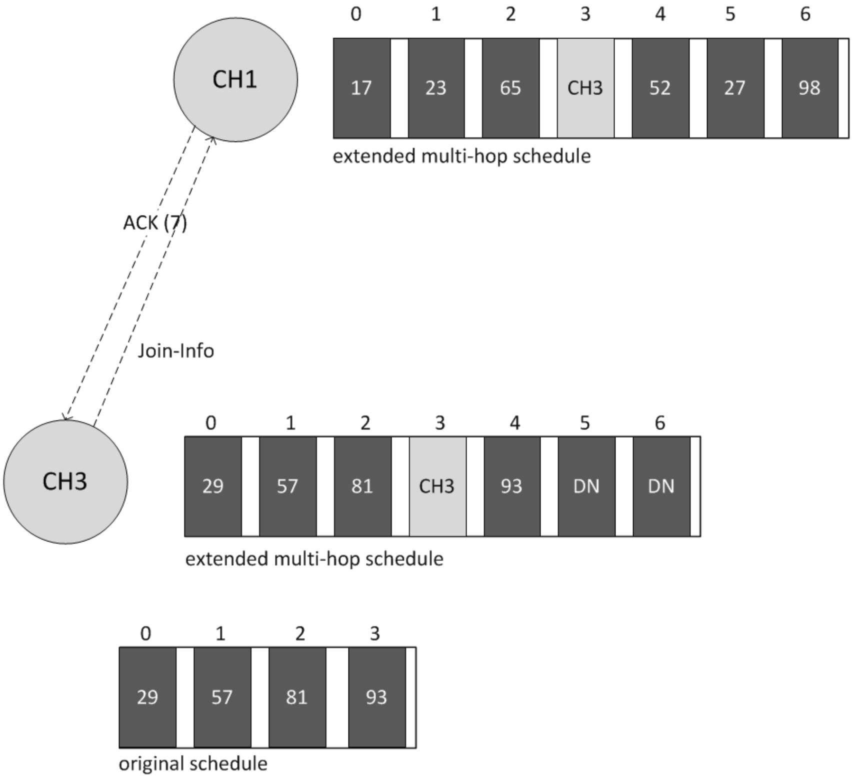


Figure 4.2.: Schedule extended by DBH-LEACH

In order to be able to adopt the schedule length the cluster-sizes are exchanged with the Join-ACK packet. Ensuring the correct size is used the previous-hop CH which sends the Join-Info packet to its next hop waits until it receives the Join-ACK before it starts building the schedule for its cluster.

In this example in figure 4.2 the number of nodes of CH1 is 7 as transmitted in the ACK message. The lower-level CH (CH3) has originally 4 nodes in its schedule. For transmitting to the next-hop node one entry with its own node-ID is added. As the size of the schedule remains smaller than the size of the schedule of CH1 two dummy-nodes

(DN) are added to reach the same length of the schedule as in the next-hop node.

4.1.6. Steady-state Phase in the Next-hop Communication

The steady-state phase remains basically as defined in LEACH. The only difference concerns the transmission to the BS which is not done by CHs that found an eligible next-hop CH. As explained in 4.1.5 for the next-hop communication an additional slot in the schedule of the sending and the receiving CH is reserved.

4.2. Models and Calculations

Models play a crucial role in network simulations as they define the properties of the simulated environment. The employed models may also have a major impact on the simulation results and on the comparability with real world [4].

4.2.1. Radio Propagation Models

Radio propagation models define the characteristics of the radio transmission between wireless nodes. These models represent the electromagnetic wave propagation on a wireless channel in the simulation. Simulations use radio propagation models to determine the signal strength of incoming packets. Typically the decision whether or not data is received strong enough to interpret it correctly depends on it. The wireless channel is influenced by the environment and obstacles within the transmission range. In general there are three effects that may occur and influence the signal propagation in the wireless medium. These effects are refection, diffraction and scattering, which appear when a signal hits objects or edges. Despite the awareness of these effects, most simulations apply simplified models [4].

Basically two types of propagation models can be distinguished: large-scale and small-scale propagation models. Small-scale models define the relationship between transmission quality and node movement. Large-scale propagation models are used for calculating the transmission power based on the distance between communicating nodes over a long time. Fundamental ratios in these models are the RSSI (Received Signal Strength Indicator) and the path loss. The RSSI is simply defined as a threshhold of the signal strength that arriving packets have to exceed to be distinguishable from noise. The path loss is defined as the difference between transmitted and received power. The exact calculation method depends on the particular propagation model used. ns-2 applies the RSSI on incoming packets. Therefore it compares the RSSI value to a threshold to determine whether the node was able to receive a packet [4][5][7].

All simulations of LEACH presented here utilize deterministic models which are presented in this section [7].

4.2.1.1. Free Space Model

The Free Space model is based on the assumption of having a clear line-of-sight path between sender and receiver without any obstacles in the transmission range [7]. The calculation of the RSSI is done with the Friss free space equation shown in equation 4.1:

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi d)^2 L} \tag{4.1}$$

where the variables are defined as following:
$P_r(d)$ [J] the receive power for a given distance d
P_t [J] the transmit power
G_t [m] the transmitting antenna gain
G_r [m] the receiving antenna gain
λ [m] the wavelength of the carrier signal
d [m] distance between sender and receiver
$L \geq 1$ system loss factor

4.2.1.2. Two-Ray Ground Reflection Model

The second model applied in this work is the Two-Ray Ground Reflection model, which considers both the direct path and ground-reflection path. In this model the electrical field received (E_{TOT}) takes both components into account and sums them up [7]. This model considers multipath fading emerging in case of no direct line-of-sight. In this case the signal is reflected by the obstacle or the ground extending the transmission distance and delaying the arrival time of the signal [5]. Additionally, this model considers the height of the transmitter and the receiver. Further the received signal strenght is regarded independent from the wavelength of the carrier signal. For simulations in ns-2 with this model the height of sender and receiver must be equal. Equation 4.2 shows the relation between received power and distance defined by the Two-Ray Ground Reflection model [4],

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4} \tag{4.2}$$

where h_r is the height above ground of the receiving antenna and h_t for the height above ground of the receiving antenna.

The channel propagation model of ns-2 defines the usage of the Free Space model when communicating nodes are within the crossover-distance $d_{crossover}$. If the distance between sender and receiver exceeds the crossover-distance the Two-Ray Ground Reflection model is applied. The crossover-distance is defined by equation 4.3 [5]:

$$d_{crossover} = \frac{4\pi\sqrt{L}h_r h_t}{\lambda} \tag{4.3}$$

4.2.2. Radio Energy Model

The radio energy model defines the energy consumption characteristics for sending and receiving data of a radio. The selection of a radio energy model has a significant impact on the performance of the protocol applied on it. In this case a basic model is used, which takes up energy for supplying the radio electronics at both participants of a communication and additionally requires energy for the signal amplification at the sender. The required energy for amplification depends on the distance between sender and receiver. As mentioned in 4.2.1, the crossover-distance defines which propagation model applies for estimating the required transmission power. Based on the Free Space model, the power for transmitting within the crossover-distance is proportional to d^2 while at distances exceeding $d_{crossover}$ the Two-Ray Ground Reflection model applies requiring transmission power related to d^4. In general the transmission power must be set high enough to ensure a certain signal strength at the receiver, exceeding the threshold $P_{r-thresh}$, to be able to distinguish data packets from noise. The simulations of LEACH and DBH-LEACH are based on the assumption of an environment with a receiver noise figure of $17dB^2$ and a thermal noise floor of $99dBm$. In order to receive data properly the SNR (Signal-to-Noise Ratio) of incoming signals may not fall below $30dB$. Therefore the threshold for receiving is defined as in equation 4.4 [5].

$$P_{r-thresh} \geq 30 + (-82) = -52dBm = 6.3nW \tag{4.4}$$

Moreover the required transmission power is defined by the following equations which are derived from the equations for the Free Space model (4.1) and the Two-Ray Ground Reflection model (4.2)[5]:

$$P_t = \begin{cases} \alpha_1 P_{r-thresh} d^2 & : d < d_{crossover} \text{ where } \alpha_1 = \frac{(4\pi)^2}{G_t G_r \lambda^2} \\ \alpha_2 P_{r-thresh} d^4 & : d \geq d_{crossover} \text{ where } \alpha_2 = \frac{1}{G_t G_r h_t^2 h_r^2} \end{cases} \tag{4.5}$$

4.2.3. Data Aggregation Energy Model

The aim of data aggregation in LEACH is to reduce of the amount of data transmitted through exploiting correlation and redundancy in sensor data. A lower amount of data to transmit implies also less energy consumption for data transmission. In LEACH and DBH-LEACH the CHs are responsible for data aggregation after receiving data from all cluster nodes and before transmitting it to the BS. In order to perform data aggregation, calculations need to done which consume energy as well. The actual energy demand depends on the implemented aggregation algorithm and on the used hardware of the sensor nodes. Based on experiments with a StrongARM processor the LMS beamforming algorithm was selected for LEACH [5]. This algorithm turned out to dissipate less energy than other beamforming algorithms (e.g. Maximum Power) and also to scale linear with the number of nodes in the network, which is important in sensor networks. Based on the experiment results, the energy dissipation for performing data aggregation is defined as $5nJ/bit/signal$ for all simulations [5].

5

Simulations

In order to evaluate the actual effects of the enhancements introduced by DBH-LEACH, as described in chapter 4, a scenario consisting of three different configurations was simulated; using the original LEACH protocol and the new protocol. The aim of the simulations is to determine whether DBH-LEACH achieves its design objectives.

5.1. Simulation Environment

The discrete event simulator ns-2 is used for all simulations performed in this thesis. ns-2 is intended for network research and provides capabilities to simulate various transport, routing and multicast protocols for wired and wireless networks. The development started in 1989 as a version of the REAL network simulator and was expedited through different projects and institutes over the last years. This simulator is in continuous development and consists also of numerous contributions from different research projects and companies [12].

5.2. Conditions and Assumptions for the Simulations

The simulations are based on several assumptions and general conditions that define the environment and the application of the network. For all simulations performed in the scope of this thesis the following assumptions apply:

1. The deployment area of the sensor nodes is a square with a fixed side length. The nodes are distributed randomly within this area.
2. There is only one type of sensor node used in all simulations. All nodes are equipped with the same components meaning all nodes have the same energy resources and processing power.

3. All nodes are stationary, there is no mobility or movement considered in the simulations.
4. Nodes dissipate energy for performing data transmission and data aggregation. But also in sleep and idle mode nodes consume a low amount of energy.
5. The random positions of the nodes are defined and stored in a topology file that all simulation runs apply for both protocols.

5.3. Definition of Metrics

The analysis of simulations is based on log-files that document the core characteristics of every single simulation run. In order to interpret the simulation results this information must be analyzed and evaluated statistically regarding the relevant factors. Therefore metrics strongly depend on the goals and effects that are intended to analyze through the simulations. The aims of DBH-LEACH were the improvement of the network lifetime and the enhancement of the performance of the protocol in wide expanded networks. Furthermore it is expected that energy consumption is spread more evenly among the nodes causing nodes running out of energy slower. Following metrics based on the data collected from the simulations are used:

- **Quantity of Active Nodes:** Every simulation logs the number of active nodes every 10 seconds during the run. Based on this information the average number of remaining active nodes at each time step of the network operation is calculated. The results are shown in a graph of nodes at the average number of active nodes at any time during the network operation. This allows concluding about the energy consumption behavior of a protocol.
- **Network Lifetime:** The network lifetime is measured by evaluation of the reached maximum operational time of the network. Usually a percentage of the number of nodes defines when the network is considered not functioning any more. Basically this number depends on the application until what amount the results are meaningful. For all simulations in this thesis this boarder was set to 5, which means, that the network stops its operation when the number of active nodes drops below this limit.
- **Data Transmission Ratio and Data Loss:** The amount of data transmitted to the BS and received by it are accumulative values, summing up the bytes transmitted by all CHs towards the BS or received by the BS. As these values are logged together with the quantity of active nodes every 10 seconds, it is possible to determine average over all simulation runs for each step. Thus it is possible to analyze and visualize the mean development of transmitted and received data. Furthermore another log keeps

track of all collision and desynchronized sending and receiving activities causing packet loss. These values are extracted at the end of each simulation and provide a fundament to calculate the mean data loss.

5.4. Simulation Scenario

The aim of this scenario is to analyze the effects on network lifetime and energy consumption of the DBH-LEACH protocol compared to the original LEACH protocol in a wide expanded area. Therefore three simulations with a different quantity of nodes in the same sensing area are executed and the results are compared in the following. The list below describes the parameters of the three simulations performed for this scenario.

- Network expansion area: 200 x 200 meter (for all three simulations)
- Position of BS: x=275, y=100 (outside the sensor node area)
- Number of nodes in simulation 1: 100
- Number of nodes in simulation 2: 150
- Number of nodes in simulation 3: 200

Each of the simulations performs 100 runs using the same topology in order to gain representative results. The nodes are distributed randomly in the sensing area. Furthermore all simulation are executed with the same settings for LEACH and DBH-LEACH. The results presented in this chapter compare the characteristics of the two protocols but also relate to the effects of the varying number of nodes in the sensing area.

5.4.1. Quantity of Active Nodes

The first characteristic investigated is the number of active nodes remaining in relation to the elapsed network operation time. Based on the energy model used in the simulations for both protocols, it is suggested that DBH-LEACH extends the network operation time and delays nodes running out of energy by preserving it for long-distance (multi-hop) transmission. Figure 5.1 presents the average number of active nodes at any time over all 100 simulation runs for simulation 1 with 100 nodes and for simulation 3 with 200 nodes. The dashed line and the dot and dashed line show the devolution of DBH-LEACH. The continuous and dotted lines illustrate the result of the original LEACH. Evaluating the two curves for simulation 1 indicates that the DBH-LEACH nodes operate longer. Moreover the number of active nodes declines more slowly than with LEACH, thereby increaseing the advantage of DBH-LEACH in extending the network´s lifetime. It also shows that the overall pattern of energy consumption is not changed by DBH-LEACH. The reason is

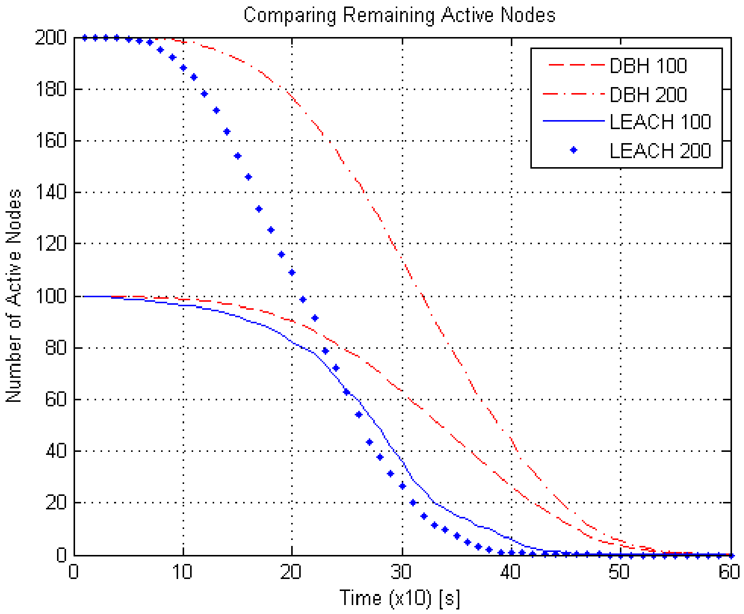


Figure 5.1.: Comparing the development of active nodes in the network for simulations with 100 and 200 nodes

that DBH-LEACH uses the same method for CH selection as does the original protocol which does not consider influencing factors such as node location or energy level.

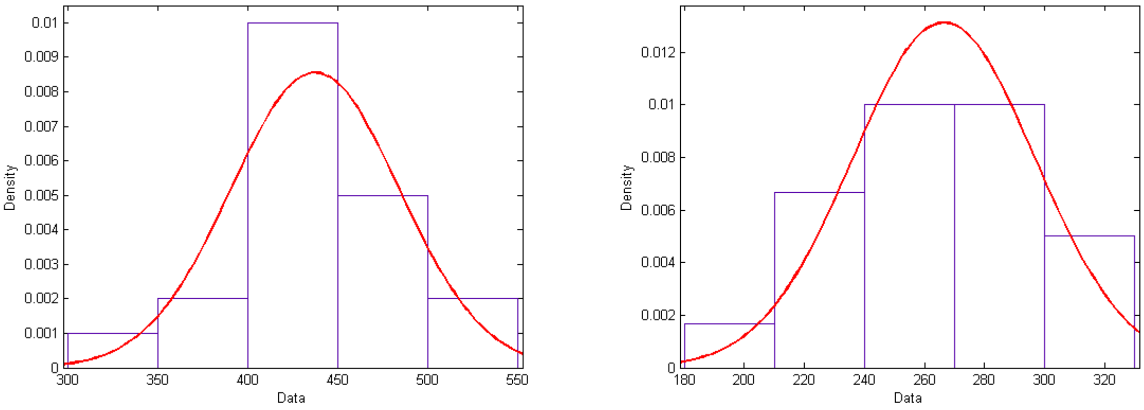
Also for simulation 2 with 150 nodes the identical characteristics are observable. Furthermore the advantage of DBH-LEACH compared to LEACH grows with an increased number of nodes in the network. This implies that the protocol can better handle a higher node density.

The results of the simulation with 200 nodes also confirm previously observed trends. The original version of LEACH deteriorates its performance regarding the development of active nodes when the number of nodes increases. Especially at the beginning, when all or the majority of nodes are active LEACH nodes deplete the energy resources fast. This effect slows down as the number of nodes in the network decreases. Contrary, the energy dissipation in DBH-LEACH remains similar for all three simulations. Based on these observations it is possible to conclude that hierarchical multi-hop transmission counters the drawback of the original version that dissipates energy fast in higher density and wide expanded networks. The reason for this development is that with LEACH also remotely located nodes become cluster-heads, which need to transmit data directly over long distances to the BS. Thus it is reasonable that the energy saving of multi-hop transmissions in DBH-LEACH gains relevance in networks with larger distances between nodes. This underlines the comparison of the devolution of nodes alive for the simulation with 100 and 200 nodes as shown in figure 5.1.

5.4.2. Network Lifetime

The network lifetime is one of the most important metrics for WSN protocols. In order to ensure an approximate normal distribution of the simulation results the batch means method combining the results of five simulation runs was applied. The resulting values roughly represent normal distribution but outliers impact the result above average as it is based only on 100 samples.

Figure 5.2 illustrates the distribution of the network lifetime for both protocols at the simulation with 200 nodes. In general, it shows that the network lifetime in DBH-LEACH has a wider variance but also a higher mean value than in LEACH.



(a) DBH-LEACH (b) LEACH

Figure 5.2.: Distribution of network lifetime in the simulation with 200 Nodes

In order to underline the difference in the network operation time based on the simulation results a 99% confidence interval was calculated. Table 5.1 compares the confidence intervals for LEACH and DBH-LEACH for all three simulations of this scenario. To verify that the simulation results are normal distributed within a defined level of significance the chi-square test was performed for each simulation of this scenario. The detailed results are documented in Appendix (A.1). The null hypothesis was not rejected at the 5% significance level for any of the simulations; therefore it is possible to determine the confidence interval.

Figure 5.3 visualizes these confidence intervals for an easier comparison. It is noticeable that the intervals for all three simulations with DBH-LEACH vary only slightly. This graphic also shows that DBH-LEACH performs best with 150 nodes in the sensing area. Even though it is only a minor difference to the other simulations, it appears that at this node density the protocol is able to utilize hierarchical next-hop transmissions best. However it must be considered that all simulations are performed with random topologies.

Simulation	μ	σ	Lower Limit	Upper Limit
100 Nodes with DBH-LEACH	444.91	24.43	429.28	460.54
100 Nodes with LEACH	365.49	24.38	349.89	381.09
150 Nodes with DBH-LEACH	450.12	23.83	434.88	465.36
150 Nodes with LEACH	298.09	44.70	269.50	326.69
200 Nodes with DBH-LEACH	437.33	46.65	407.48	467.17
200 Nodes with LEACH	266.43	30.41	246.98	285.88

Table 5.1.: Confidence intervals of the network lifetime

Therefore it is possible that the one used for the simulation with 150 nodes had a more suitable node distribution for multi-hop transmissions, as another reason for the increased lifetime. In general DBH-LEACH performs quite stable at all three simulations made in this scenario. The results of the original LEACH protocol are essentially different. The mean value of the network lifetime decreases significantly with the increasing number of nodes. In general it performs considerable worse than DBH-LEACH in the 200 x 200 meter area. The performance of LEACH deteriorates when the network conatins a higher number of nodes.

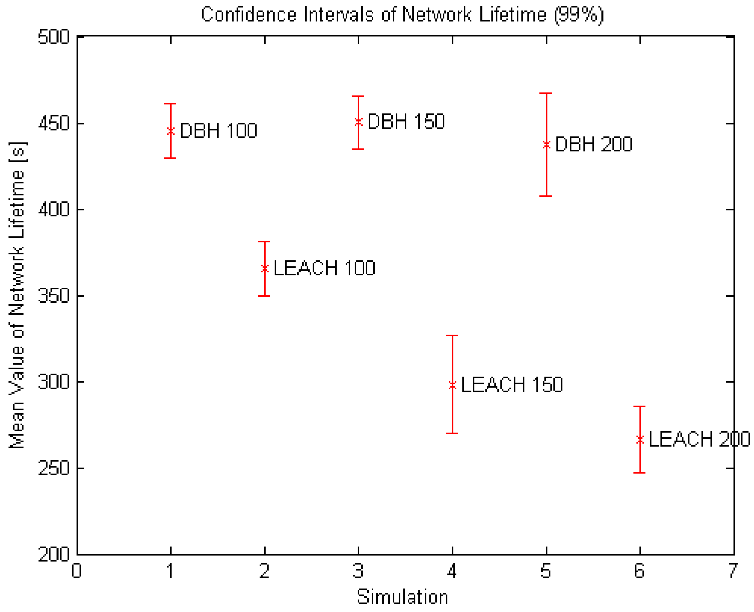


Figure 5.3.: Confidence intervals for network lifetime of all simulations

5.4.3. Data Transmission and Packet Loss

More characteristics to evaluate the performance of a network protocol are the packets loss and the successful data delivery rate. When comparing the amount of data transmitted to the BS, a significant decrease of approximately 40% is noticeable for DBH-LEACH as figure 5.4 exemplifies for simulation 1 with 100 nodes.

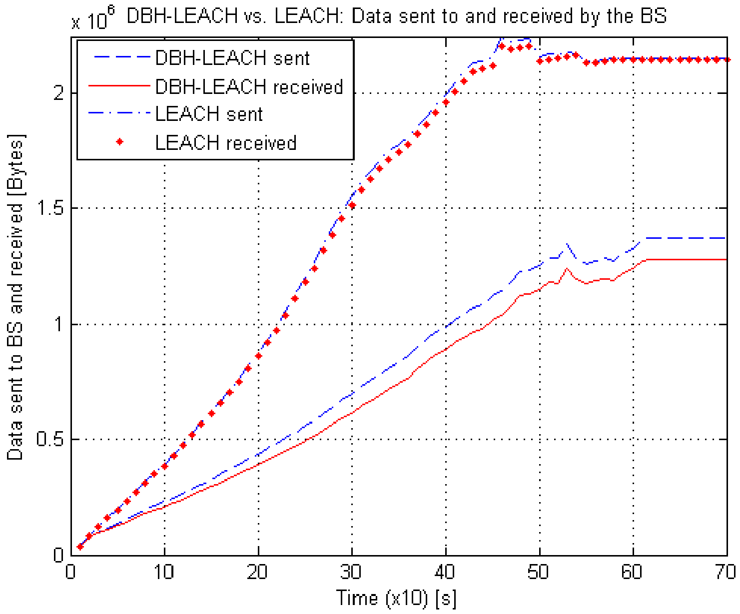
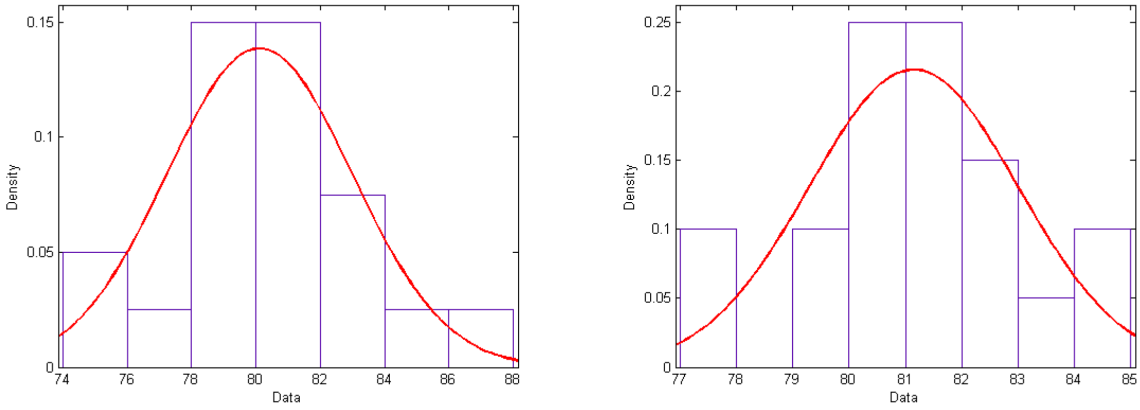


Figure 5.4.: Data transmitted to the BS and received by the BS

This effect is caused by adding general nodes to the schedules in order to stay synchronized with the schedule of the next-hop CH. As a result, cluster-heads that use next-hop transmission reach the sending slots less frequent. For example, the number of nodes in the schedule might be supplemented to fit a multiple of the number of nodes in the next-hop CHs schedule. Therefore in DBH-LEACH data is not transmitted as frequently as in the original LEACH. The effect occurs in all simulations of this scenario. Additionally a gap between the curve of data sent to the BS and of data received by the BS is apparent for DBH-LEACH. The reason for the difference is an increased number of collisions that happen when a node has a high amount of previous-hop CH. In this case the packet to transmit to the BS reaches a size that often takes more time to transmit than reserved. Consequently a large packet containing multiple packets of other CHs collides with a packet from an ordinary node in the cluster. Thus data of multiple CHs is lost through a single collision. Furthermore there is also a chance of data loss at the transmissions between CHs in DBH-LEACH. Figure 5.5 shows the distribution of successful data transmission between CHs for the simulations with 150 and 200 nodes. It indicates that the success rate of next-hop data transmission is approximately normal distributed around 80%.



(a) 150 Nodes

(b) 200 Nodes

Figure 5.5.: Distribution of successful data transmission between CHs

5.5. Analysis of the Simulation Results

The analysis of the simulations in a 200 x 200 meter area network shows, that the aim of improving the network lifetime in wide expanded WSNs was achieved. But it is important to notice that the effective amount of data transmitted to the BS decreases through the changes in the protocol. Furthermore it is observable that the original protocol performs poor with increasing node density. Hence DBH-LEACH achieves also an improvement regarding the supported node density in a network.

It is important to notice that it was not possible to implement the best possible solution within the scope of this research. As the implementation is based on the uAMPS code for LEACH existing shortcomings of the implementation had to be accepted. It was not possible to find an ideal solution for the next-hop CH selection, as it not considers possible next-hop choices of other CHs or the schedule synchronization. The reason is the complexity of the decentralized setup process and the tight timing through TDMA. Solving these issues requires an extensive redesign of the protocol or the processes involved.

6

Conclusion

This research analyzed energy efficiency in the context of WSN routing. A new routing protocol was developed, called DBH-LEACH which is derived from the established and well known LEACH protocol. The major enhancement that was introduced by the new protocol is the addition of the ability of clusters to create hierarchical structures among each other. Through these structures cluster-heads are able to use less energy-hungry multi-hop paths for data transmission, thereby extending the lifetime of nodes and the entire network. The effects of the extended lifetime were analyzed through simulations comparing the new protocol with the original LEACH protocol. These simulations showed that hierarchical multi-hop paths can reduce energy consumption for long distance data transmissions. This reduction is mainly observed in large, expanded networks where sensor nodes can operate with measurable reductions in energy use. As a result the operation time of individual nodes is extended, as is the network´s lifetime. It was also demonstrated that for small network areas multi-hop transmission does not achieve a significant benefit over single-hop transmission, as used in LEACH. Moreover the simulation results implied that the properties of the decentralized cluster-formation of LEACH adversely impact energy dissipation and cannot be compensated by the multi-hop transmission. In another important finding, the amount of data transmitted decreases despite a longer operation time. This is related to modifications in the scheduling that are necessary to conduct the multi-hop hierarchy in DBH-LEACH.

This work identifies relationships between network expansion, the number of nodes, and the levels of energy consumption. This can support future research on the energy dissipation of sensor nodes. It suggests ideas to refine and improve metrics for new routing protocols. DBH-LEACH provides the potential for greater energy efficiency and the delivery of more data by improving the scheduling and the handling of next-hop nodes. While beyond the scope of this work, it is probable that the DBH-LEACH protocol can assist in future improvements in node organization, cluster formation hierarchy management.

A

Additional Evaluation of Simulations

This appendix contains additional evaluations made on the simulation results in order to verify particular preconditions.

A.1. Chi-Square Test for Simulation Scenario 1

The chi-square test has been performed for every simulation of the scenario. The results have the following meaning:

- h = 0 - shows that the null hypothesis for a significance level of 5% cannot be rejected.
- p - The p value is the probability, when assuming the null hypothesis, to observe the given statistic or one more extreme.
- chi2stat - The chi-square statistic.
- df - Degrees of freedom.

Results of Chi-Square Test for Simulation Scenario 1 are shown in table A.1.

Simulation	h	p	chi2stat	df
100 Nodes with DBH-LEACH	0	0.928	0.008	1
100 Nodes with LEACH	0	0.729	0.119	1
150 Nodes with DBH-LEACH	0	0.474	0.513	1
150 Nodes with LEACH	0	0.569	0.323	1
200 Nodes with DBH-LEACH	0	0.734	0.115	1
200 Nodes with LEACH	0	0.754	0.098	1

Table A.1.: Results of chi-square test for all simulations in scenario 1

References

[1] Akkaya, K. and Younis, M.: *A survey on routing protocols for wireless sensor networks*. http://www.csee.umbc.edu/~kemal1/mypapers/Akkaya_Younis_JoAdHocRevised.pdf, visited on 17.12.2011.

[2] Al-Karaki, J. N. and Kamal, A. E.: *Routing techniques in wireless sensor networks: A survey*. IEEE Wireless Communications, 11:6–28, 2004.

[3] Buratti, C., Conti, A., Dardari, D., and Verdone, R.: *An overview on wireless sensor networks technology and evolution*. In *Sensors 2009*, volume 9. 2009. www.mdpi.com/1424-8220/9/9/6869/pdf.

[4] Günes, M. and Wenig, M.: *Models for realistic mobility and radio wave propagation for ad hoc network simulations*. In Misra, S., Woungang, I., and Misra, S. C. (editors): *Guide to Wireless Ad Hoc Networks*, pages 255–280. Springer-Verlag, London, 2009, ISBN 978-1-84800-327-9.

[5] Heinzelman, W. B.: *Application specific protocol architectures for wireless networks*, 2000. http://www-mtl.mit.edu/researchgroups/icsystems/pubs/theses/wendi_phd_2000.pdf, visited on 20.01.2012.

[6] Massachusetts Institute of Technology: *The mit uamps ns code extension*, 2000. http://mtlweb.mit.edu/researchgroups/icsystems/uamps.

[7] Matos, J. and Miranda, H.: *Ad hoc routing under randomised propagation models*. In *INForum 2010 - II Simposio de Informatica*, pages 379–390. http://inforum.org.pt/INForum2010/papers/computacao-movel-e-ubiqua/Paper034.pdf, visited on 13.04.2012.

[8] Padmanabhan, K. and Kamalakkannan, P.: *A study on energy efficient routing protocols in wireless sensor networks*, 2011. http://www.eurojournals.com/EJSR_60_4_04.pdf, visited on 20.01.2012.

[9] Perillo, M. and Heinzelman, W. B.: *Wireless sensor network protocols*, 2005. http://www.ece.rochester.edu/courses/ECE586/readings/perillo.pdf.

[10] Pottie, G. J. and Kaiser, W. J.: *Wireless integrated network sensors*. Commun. ACM, 43(5):51–58, 2000, ISSN 0001-0782.

[11] Singh, S. K., Singh, M. P., and Singh, D. K.: *Routing protocols in wireless sensor networks – a survey.* 2010.

[12] University of Southern California: *The network simulator - ns-2.* `http://www.isi.edu/nsnam/ns/`, visited on 13.06.2012.

[13] Zheng, J.: *Medium access control.* In Zheng, J. and Jamalipour, A. (editors): *Wireless Sensor Networks.* John Wiley & Sons, Inc., New Jersey, 2009, ISBN 978-0-470-16763-2.

[14] Zheng, J.: *Network architectures and protocol stack.* In Zheng, J. and Jamalipour, A. (editors): *Wireless Sensor Networks.* John Wiley & Sons, Inc., New Jersey, 2009, ISBN 978-0-470-16763-2.

[15] Zheng, J. and Jamalipour, A.: *Introduction to wireless sensor networks.* In Zheng, J. and Jamalipour, A. (editors): *Wireless Sensor Networks.* John Wiley & Sons, Inc., New Jersey, 2009, ISBN 978-0-470-16763-2.