AN INTELLIGENT FUZZY CLUSTERING APPROACH FOR ENERGY-EFFICIENT DATA AGGREGATION IN WIRELESS SENSOR NETWORKS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SEYYİT ALPER SERT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

MAY 2018

Approval of the thesis:

AN INTELLIGENT FUZZY CLUSTERING APPROACH FOR ENERGY-EFFICIENT DATA AGGREGATION IN WIRELESS SENSOR NETWORKS

submitted by **SEYYİT ALPER SERT** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** __________

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** __________

Prof. Dr. Adnan Yazıcı
Supervisor, **Computer Engineering Department, METU** __________

Examining Committee Members:

Prof. Dr. Tolga Can
Computer Engineering Dept., METU __________

Prof. Dr. Adnan Yazıcı
Computer Engineering Dept., METU __________

Prof. Dr. İbrahim Körpeoğlu
Computer Engineering Dept., BİLKENT University __________

Prof. Dr. Ahmet Coşar
Computer Engineering Dept., University of TAA __________

Assoc. Prof. Dr. Ertan Onur
Computer Engineering Dept., METU __________

Date: __________

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

Name, Last Name: SEYYİT ALPER SERT

Signature :

ABSTRACT

AN INTELLIGENT FUZZY CLUSTERING APPROACH FOR ENERGY-EFFICIENT DATA AGGREGATION IN WIRELESS SENSOR NETWORKS

Sert, Seyyit Alper
Ph.D., Department of Computer Engineering
Supervisor : Prof. Dr. Adnan Yazıcı

May 2018, 127 pages

Technological developments have made the generation and usage of wireless sensor nodes possible. Although an individual node is capable of gathering data alone, these nodes generally cooperate to extract high-level semantic information from the sensed region. Networks consisting such nodes are referred to as Wireless Sensor Networks (WSNs). There is generally a balance between energy-efficiency and accuracy, which are two desirable but incompatible features of these networks, because of the resource-restricted nature of the utilized devices. The balance, which can also be called the trade-off, is tried to be optimized by efficient algorithms that mostly utilize manually-value-assigned parameters through trial-and-error processes. However, this assignment process nearly always fails in finding the optimum blend of parameters, renders the implementation vague and inapplicable for most cases, and generally biases the obtained result.

In this dissertation, an Intelligent Fuzzy Clustering Approach for Energy-Efficient Data Aggregation in Wireless Sensor Networks is proposed. The proposed approach

is a distribution-agnostic approach that runs and scales efficiently for sensor network applications. Additionally, along with the proposal, an optimization framework is utilized to tune the parameters used in the fuzzy clustering process in order to optimize the performance of a given WSN. This dissertation also includes performance comparisons and experimental evaluations of the proposal with the selected state-of-the-art algorithms. The experimental results reveal that the proposal performs better than any of the compared protocols under the same network setup considering metrics used for comparing energy-efficiency and network lifespan of the protocols. Besides, along with the proposed optimized fuzzy network clustering protocol, an empirical study on multi-modal object classification problem in wireless sensor networks is conducted in detail and obtained results are presented as well in order to corroborate the object classification accuracy aspect of the proposed protocol.

Keywords: wireless sensor networks, fuzzy clustering, optimization, multi-modal object classification

ÖZ

KABLOSUZ DUYARGA ŞEBEKELERDE ENERJİ-VERİMLİ VERİ TOPLAMA İÇİN AKILLI BULANIK KÜMELEME YAKLAŞIMI

Sert, Seyyit Alper
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Adnan Yazıcı

Mayıs 2018 , 127 sayfa

Teknolojik gelişmeler, kablosuz duyarga düğümlerinin üretilmesini ve kullanılmasını mümkün kılmıştır. Her bir düğüm tek başına veri toplayabilmesine rağmen, algılanan bölgeden yüksek seviyeli anlamsal bilgiyi çıkarmak için bu düğümler genellikle işbirliği yaparlar. Bu düğümlerden oluşan şebekeler, Kablosuz Duyarga Şebeke (KDŞ)'ler olarak adlandırılır. Genel olarak, kullanılan cihazların kaynak kısıtlı yapıları nedeniyle, bu şebekelerin arzu edilen ancak uyumsuz iki özelliği olan enerji verimliliği ve doğruluğu arasında bir denge vardır. Pazarlık olarak da adlandırılabilen denge, deneme-yanılma süreçleriyle çoğunlukla manuel değer atanmış parametreler kullanan verimli algoritmalar ile optimize edilmeye çalışılmaktadır. Ancak, bu atama süreci neredeyse her zaman, parametrelerin optimum bileşimini bulmakta başarısız olmakta, uygulamayı belirsiz ve çoğu durum için uygulanamaz kılmakta ve genellikle elde edilen sonucu tartışmalı hale getirmektedir.

Bu tez çalışmasında, Kablosuz Duyarga Şebekelerde Enerji-Verimli Veri Toplama için Akıllı Bulanık Kümeleme Yaklaşımı önerilmiştir. Önerilen yaklaşım, duyarga şebeke uygulamaları için verimli çalışan ve ölçeklenen dağılım-bağımsız bir yakla-

şımdır. Ek olarak, öneri ile birlikte belirli bir KDŞ'nin performansını optimize etmek için, bulanık kümeleme işleminde kullanılan parametreleri ayarlamada bir optimizasyon çerçevesi kullanılmaktadır. Bu tez aynı zamanda, seçilen son teknoloji algoritmalarla önerinin performans karşılaştırmalarını ve deneysel değerlendirmelerini de içermektedir. Deneysel sonuçlar, önerinin, protokollerin enerji verimliliğini ve ağ ömrünü karşılaştırmak için kullanılan metrikler göz önünde bulundurulduğunda, aynı şebeke kurulumu altında karşılaştırılan protokollerin her birinden daha iyi performans gösterdiğini ortaya koymaktadır. Ayrıca, önerilen optimize edilmiş bulanık ağ kümeleme protokolü ile birlikte, önerilen protokolün nesne sınıflandırma doğruluğu yönünü doğrulamak için kablosuz duyarga şebekelerde çok kipli nesne sınıflandırma problemi üzerinde ampirik bir çalışma ayrıntılı olarak gerçekleştirilmiş ve elde edilen sonuçlar sunulmuştur.

Anahtar Kelimeler: kablosuz duyarga şebekeler, bulanık kümeleme, optimizasyon, çok-kipli nesne sınıflandırma

***To My Family**...*

ACKNOWLEDGMENTS

The inception of inditing the acknowledgments made me feel as coming closer to the end of a long run. It was truly a long journey, with ups and downs, however quite enjoyable! That's why the accomplishment of this dissertation has been materialized with the invaluable aid of many people worth mentioning to whom I feel indebted.

First and foremost, I'd like to put my inmost gratitude into words for my advisor Prof. Dr. Adnan Yazıcı due to his perception and encouragement along my studies, guidance and support, and above all, trust on me. Sharing his wisdom is always a great honor for me. I am also sincerely grateful to him for our precious talks about nonacademic issues. Being his apprentice has presented me an invaluable experience, a widened perspective, and shed the light on my long-standing path.

Besides my advisor, I feel sincerely beholden to my Thesis Monitoring Committee members, Prof. Dr. İbrahim Körpeoğlu and Assoc. Prof. Dr. Ertan Onur, for their elaborate advices, constitutive criticisms, and excellent feedback which had crucial effects on this dissertation to be formed into its current structure. It was their questions and suggestions that incented me to broaden my studies from several perspectives.

I am very thankful to my colleague Metin Durmaz for his support, toleration, and suggestions throughout my studies. Especially, our talks on fuzzy optimization have proved to be very fruitful. It is an honor and rewarding experience for me to work with him during probably one of the toughest periods of our country.

As a member of Multimedia Research Lab., I'd also like to thank the whole research group for our precious talks throughout my study. Although we are not in the same study and working environment, we easily managed to meet frequently and exchange ideas throughout this dissertation. My special thanks go to Abdullah Alchihabi for the optimization framework and to Cihan Küçükkeçeci for the generic data simulator utilized in this study.

I am also indebted to TÜBİTAK (The Scientific and Technological Research Council of Turkey), which has financially supported my doctoral studies in part by a research grant under Grant No. 114R082.

In the world, there is nothing like the love, patience, toleration and self-sacrifice of a family. For this reason, I feel owe to all my family members and would like to express my heartfelt thanks. They generally had no idea what I was doing, what this thesis is about, or why I was struggling to stay awake and sharp late in the midnights. They were quite interested in the completion of this thesis rather than how I manage to do it. Nevertheless, the interminable support and encouragement they maintained have always been a great help throughout my research and are very very much appreciated.

Finally, my wife Deniz, who has absolutely provided the foremost contribution by putting herself in my shoes in daily life and aiding me to focus extensively on my research, deserves my heartfelt gratitude and inmost thanks. Not only the dissertation part, but my whole Ph.D. study period progressed with her excellent toleration, understanding and love. As a Ph.D. candidate's wife, she was compulsorily a Ph.D. widow for quite a long time. Although we could not spend most our spare times together or as we wish throughout my research, I feel deep in my heart that she never stops preying for my success. Feeling her everlasting love, understanding, and above all respect on me, is like a miracle. So, this dissertation is devoted to her and our beloved son Alp Yaman, with great passion and gratitude.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ALGORITHMS

ALGORITHMS

LIST OF ABBREVIATIONS

ALRE	Average Link Residual Energy
ANN	Artificial Neural Network
CH	Cluster-Head
CHEF	Cluster Head Election using Fuzzy
CIS	Communication and Information Systems
CLONALG	Clonal Selection Algorithm
CLONALG-M	A Modified Clonal Selection Algorithm
COG	Center of Gravity
Coor.	Coordinate
CS	Clonal Selection
DBCP	Density-Based Clustering Protocol
DECA	Density-based Energy-efficient Clustering Algorithm
DESS	Discrete Event System Simulator
DPCC	Dynamic Power Control Clustering
EAUCF	Energy Aware Unequal Clustering with Fuzzy
EECS	Energy Efficient Clustering Scheme
EEUC	Energy-Efficient Unequal Clustering
FC	Fusion Center
FLC	Fuzzy Logic Controller
FND	First Node Dies
fps	Frame Per Second
GA	Genetic Algorithm
GPSR	Greedy Perimeter Stateless Routing
HNA/HND	Half of the Nodes Alive/Half of the Nodes Die
HEED	Hybrid Energy-Efficient Distributed
IP	Internet Protocol
IR	Infra-Red
j	Joule

LEACH	Low Energy Adaptive Clustering Hierarchy
LND	Last Node Dies
MATLAB	Matrix Laboratory
MB	Megabyte
METU	Middle East Technical University
MMC	Multimedia Card
MOFCA	Multi-Objective Fuzzy Clustering Algorithm
MOPSO	Multi-Objective Particle Swarm Optimization
NAND	Not AND
NOR	Not OR
N/A	Not-Applicable
PIR	Passive Infra-Red
PSO	Particle Swarm Optimization
QoS	Quality-of-Service
RAM	Random Access Memory
RD	Relative Distance
RF	Radio Frequency
RR	Routing Route
RREP	Route Reply
RREQ	Route Request
SA	Simulated Annealing
SOFM	Self-Organizing Feature Map
SOS	Self Organizing Sensor
SPI	Serial Peripheral Interface
SSD	Solid State Drive
SWSN	Surveillance Wireless Sensor Network
TAA	Turkish Aeronautical Association
TPGF	Two-Phase geographical Greedy Forwarding
TRE	Total Remaining Energy
TS	Tabu Search
TTDFP	Two-Tier Distributed Fuzzy Logic Based Protocol
WMSN	Wireless Multimedia Sensor Network
WSN	Wireless Sensor Network

CHAPTER 1

INTRODUCTION

Technological developments have made the generation and usage of wireless sensor nodes possible. Although an individual wireless node is capable of gathering data alone, these nodes nearly always cooperate to extract high-level semantic information from the sensed region. Networks of such nodes are named as Wireless Sensor Networks (WSNs). These wireless nodes possess various components such as a computation board, a communication interface, a circuitry for interconnecting the available components, and one or more power supplies. Required sensors demanded by the application domain are integrated into these nodes prior to the deployment phase. A reference component schema of a wireless sensor unit is depicted in Figure 1.1 [1].

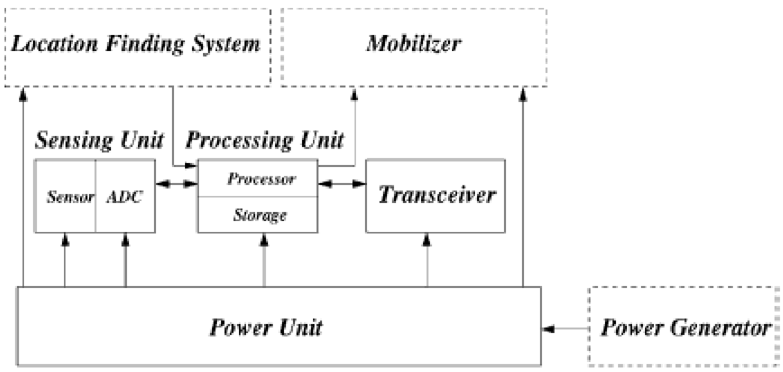

Figure 1.1: A reference component schema of a wireless sensor unit [1].

In the sections of this chapter, the increasing utilization of WSNs together with sample application domains are initially presented. Then, capabilities of these networks and faced challenges are discussed. Finally, the problem addressed in this dissertation is declared. At the end of the chapter, the organization of this dissertation is provided.

1.1 Wireless Sensor Networks

In fact, generation of very small wireless nodes and the collaborative working principle among them are the leveraging ideas behind WSNs [1]. It is possible to encounter sensor nodes at varying sizes from a few inches to some meters. A deployed WSN may contain a great number of these units according to the application domain and devices may either be deployed densely or sparsely depending on the requirements. When it comes to the deployment scheme of these wireless nodes, it can either be randomly or manually. This decision is made considering the working environment and according to the resources in hand.

Initial utilization aim of these early wireless nodes is for passive indoor applications. The first generation of the wireless nodes has the capability to sense and gather low-cost scalar data such as the location of detections, pressure, humidity, or temperature. This is due to the completely restricted resources such as storage or computation capabilities. For this reason, the sole purpose of these first generation nodes is to deliver obtained data to the demanded locations such as next-hop nodes or the base stations (sinks). However, next-generation wireless nodes possess larger storage spaces, more computation capabilities, and better or redundant power solutions when compared to their predecessors. Geared with these augmented capabilities, their primary utilization effort has also evolved from passive to active environments. In this respect, field monitoring, healthcare delivery, military applications, disaster recovery, or even industrial process controls become in the scope of usage areas.

Advances in technology and proliferation of user requirements have given birth to the derivation of WSNs. Wireless Multimedia Sensor Networks (WMSNs) are among the prominent derivatives of WSNs. WMSNs may include various types of wireless units that are geared with abundant forms of sensing units. From this perspective, in addition to scalar data, multimedia and/or multi-modal data become the focus. In these type of networks, a phenomenon is not only measured through its scalar features but also through its multi-modal aspects such as images, audio, and video streams. A Multimedia or multi-modal content requires higher-bandwidth channels when compared to scalar data that can be transmitted in a delay-tolerant manner. A reference schema of a WMSN is described in [2] and can be depicted as in Figure 1.2.

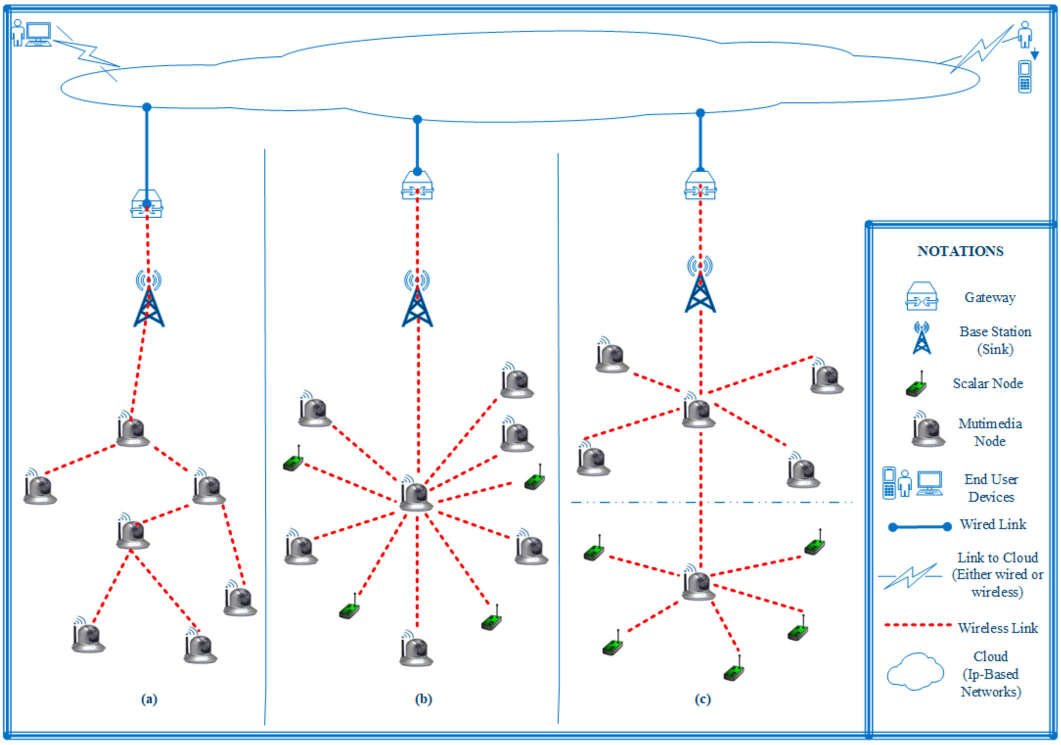

Figure 1.2: An exemplary schema of a WMSN: a) A homogenous single-tier network, b) A heterogeneous single-tier network, c) A heterogeneous multi-tier network.

Similarly to WMSNs, Surveillance Wireless Sensor Networks (SWSNs) are among derivatives of WSNs that generally possess multimedia capabilities targeting surveillance application domain. SWSNs begin to proliferate as a result of incorporating low-energy communication substructure with low-cost multi-modal content providers such as seismic or acoustic sensors together with imaging cameras and microphones. With these multi-modal augmented capabilities, WMSNs and SWSNs are able to obtain, process, store, and correlate multimedia content which is highly consulted in decision making.

1.2 Multi-Modal Content and Challenges

As generally is, there is a trade-off between the gain and the cost which enforces the beneficiary to review the cons and pros of the subject matter. This is also the case for multi-modal content in WSNs. Although multi-modal information is generally crucial for high-level semantic information extraction, it has an unprecedented acquisition, storage, and computation cost in this resource-constrained environment. For this reason, preliminary design models implemented for WSNs, considering the network stack and the expected requirements at each layer, need revisions for the aforementioned derivatives of these networks in consequence of the characteristics of multi-modal content.

The component structure of a multimedia node pretty much falls apart from that of a scalar node and hardware restriction is the core constraint over the multimedia nodes. Computation power, storage size, and short-endurance battery lifetime are among the challenges and main disadvantages faced in these derivatives of WSNs. Moreover, data transmission requirements of multimedia or multi-modal content such as expected Quality of Service (QoS) levels and bursty traffic are among the challenges that need addressing in addition to hardware constraints.

Efficient storage and management of data in WSNs are other problems in the field. Ordinary data storage design is not suitable for the derivatives of WSNs because of the constrained physical structure of a sensor node in wireless multimedia networks. Bearing in mind the restricted structure, implemented strategies in order to overcome this difficulty divide into three approaches:

- In one of the approaches, wireless nodes possess little storage space so as to efficiently utilize restricted energy resources and deliver obtained data immediately to the next-hop and let the receiving node have the burden. If the next-hop node is not the destination, the process is repeated until data reaches the destination where long-term storage and data processing take place. The reason behind the first argument is that the final destination in these networks is generally the sink, and the sink mostly resides in a constraint-relaxed locations.

- Another strategy is to equip each wireless device with high-capacity and energy efficient flash storage. Available approaches following this strategy struggle to achieve a storage-centered network as depicted in [3]. The reason behind the second argument is that consumed energy due to storage of data is insignificant when it is cross-checked with the transmission of data. For this reason, energy consumption of flash storage devices targeting sensor networks is extensively evaluated to realize storage-centered networks and presented in [4].

Tested device characteristics, flash consumption, sleep current and power-up consumption are introduced in Tables 1.1-1.3, respectively.

Table 1.1: Tested Device Characteristics [4]

Manufacturer and Type	Interface and Capacity	Page Size	Erase Block
Atmel (NOR Serial)	SPI 512 KB	256	1
ST (Serial NOR)	SPI 512 KB	256	256
Hitachi (MMC)	SPI 32 MB	512	16
Toshiba (NAND Flash)	8-bit bus 16 MB	512	32
Micron (NAND Flash)	8-bit bus 512 MB	2048	64

Table 1.2: Consumed Energy by Flash Device (μJ) [4]

	Read	Write	Erase	Bulk Erase	Total
NOR (Atmel)	0.26	4.3	2.36	N/A	6.92
NOR (Telos)	0.056	0.127	N/A	0.185	0.368
MMC (Hitachi)	0.06	0.575	0.47	0.0033	1.108
NAND (Toshiba)	0.004	0.009	N/A	0.004	0.017
NAND (Micron)	0.027	0.034	N/A	0.001	0.062

Table 1.3: Current in Sleep & Consumption in Power-Up[4]

	Current in Sleep (μA)	Consumption in Power-On (μJ)
Serial (NOR)	2	0
Hitachi (MMC)	84	1130
Toshiba (NAND)	5	0

- The final approach to be mentioned is hybrid implementations including the pros of each aforementioned approach. The motivation behind the hybrid implementation lies in the fact that each application domain has its own requirements and it should be possible to modify the utilized storage structure.

The aforementioned challenges lie in the internal structure of the wireless nodes. Apart from these challenges, there are external challenges. Type of external challenges stems from operation environment in which the nodes operate. The environmental effects such as water, temperature, or pressure also have important effects on the nodes together with typically cruel or hostile field conditions. One final classification of the challenges can be made in the application layer. Application level challenges are the result of domain-specific usages. These summarized challenge headlines are detailed in [5] as industrial application challenges in WSNs.

When multimedia streaming becomes the core focus, as in WMSN and SWSN applications, it necessitates additional attention due to low-bandwidth channels and timeliness requirements. In this respect, additional challenges that need addressing lie at the transport layer and in stream coding level and can be stated as follows:

- At the transport layer, transmission of an obtained stream from the generating node to the destination is not an easy task. This is because the network may contain broken paths, which are also referred to as holes. There are two types of holes: static and dynamic. These holes may be caused by the improper deployment of the available wireless devices, the dying of the nodes, or the overlapping. A region that cannot be covered by sensing instruments because of improper node deployment is referred to as a *static hole*, whereas a *dynamic hole* is due to the overlapping of stream data in a densely deployed network.

- For the stream coding, restricted computational capabilities and low power levels of sensor nodes lie at the heart of challenges that are faced.

In order to address transport layer challenges, it is studied thoroughly in the literature since reliable and efficient data transmission is crucial in these networks, especially when the transmission occurs in streams which are more energy consuming operations when confronted to scalar data. Among available studies, Two-Phase geographical Greedy Forwarding (TPGF) routing algorithm and its differences from Greedy Perimeter Stateless Routing (GPSR) are evaluated in [6] and elected as to be mentioned here. However, in order not to degrade the readability of this introductory chapter, elaborations are left for the big picture and literature survey chapter.

1.3 The Problem

Although there are various attractive problems in the WSN field as of today as stated in the previous section, this dissertation focuses on energy efficiency in data aggregation problem. Expected from a WSN is to fulfill its defined task, whether it is object detection, localization, tracking, or classification, in order to be useful for the end user. In this respect, although data aggregation and the lifetime of the individual nodes initially look relatively unimportant, in fact, they are crucial since they serve as minimum requirements for the mentioned operations.

If a V-node wireless network is modeled by an V-vertex graph described by G, where N and E are its nodes and edges between nodes as given in Eq. 1.1 and Eq. 1.2, then a two-dimensional network topology matrix can be constructed using the node coordinates as depicted in Table 1.4.

$$G = (N, E) \tag{1.1}$$

$$N = \{1...V\} \tag{1.2}$$

Table 1.4: Topology Matrix

V_Num	**1 , 2 ... V**
Coor. X (X_i)	$X_1, X_2...X_v$
Coor. Y (Y_i)	$Y_1, Y_2...Y_v$

Problem *Given G as an input WSN, intelligently clustering the given WSN which is efficient considering the energy aspect of data aggregation operations.*

The main problem in this context arises from the resource-constrained operation architecture and clustering approaches targeting WSNs are either sophisticated but not lightweight or lightweight but not efficient enough to meet expectations. Additionally, most available solutions targeting the efficient aggregation problem do not consider the clustering and routing phases jointly which diverts the implementation far from being applicable in real-world scenarios. And finally, it is also valid to state that most available solutions are fixed such that the evolving environment is not taken into consideration in these processes. In the light of the clarifications made, sub-problems of this dissertation considering the problem statement are as follows:

Sub-Problem 1 *Designing an energy-efficient clustering methodology.*

The designed clustering methodology should not only improve the energy-efficiency in the process when it is compared against the available solutions but also should not incur an additional cost which would enforce the beneficiary to prefer available solutions to the designed methodology.

Sub-Problem 2 *Designing an energy-efficient routing methodology.*

The designed routing methodology should contribute to the energy-efficiency of the aggregation operation carried out by the utilization of the designed clustering methodology and also should be computationally feasible under a resource-constrained sensor network environment.

Sub-Problem 3 *Flourishing the designed aggregation methodology with intelligence.*

The intelligence aspect should mimic the behavior of humans in the decision processes so that the data aggregation substructure suits to the requirements of the evolving environment.

1.4 Scope & Contributions

The main scope of the thesis concentrates on the design of an intelligent clustering approach by performing a literature survey over the aforementioned sub-problems in the context of WSNs. For the energy-efficient clustering problem, firstly, a two-tier distributed fuzzy logic-based clustering protocol is proposed. For the energy-efficient routing problem, it is tried to overcome in the designed clustering methodology as an additional tier which extends a crisp routing protocol using fuzzy variables in path determination process. Then, the designed approach is extended using the proposed Clonal Selection principle-based optimization algorithm. In addition, the designed aggregation methodology is flourished with intelligence by considering the changes in the deployment environment. Finally, proposed clustering methodology is empirically tested over multi-modal object classification domain in WSNs.

Considering the problem definition, related sub-problems, and scope of the thesis, this dissertation has three primary aims and its contributions are as follows:

- An energy-efficient data aggregation approach is designed and proposed. In this dissertation, a Two-Tier Distributed Fuzzy Logic Based Protocol (TTDFP) for the purpose of addressing data aggregation problems is proposed. TTDFP is a competitive and a fully-distributed protocol considering the lifetime requirements of the WSNs. TTDFP does not require the inclusion of a central decision-making point during any of its phases. This distributed operation architecture protects the protocol from the single point-of-failure situations. In the first tier, TTDFP decides upon the final CHs through an energy-based competition of provisional leaders, which are primarily chosen by means of a probabilistic model. Fuzzy clustering phase handles uncertainty occurring in the clustering phenomena more efficiently than its crisp and fuzzy counterparts. It should be noted that the focus of many studies in the literature is on energy-efficient clustering and, to the best of knowledge, none takes the efficiency of clustering and routing phases jointly into account. The fuzziness in the second tier (Tier II) is a novelty which also enhances routing performance when compared to its crisp counterpart.

- Developed data aggregation methodology is flourished with the intelligence sugar in order to mimic the human behavior. An optimization framework is utilized in order to tune the input parameters of the first tier in TTDFP rather than using a trial-and-error approach to find the right blend of these parameters. The optimization framework employs the Simulated Annealing algorithm to tune the pair of parameters with the aim of optimizing the performance metrics of WSNs. Additionally, the fuzzy output function of TTDFP is optimized by designing and employing a Clonal Selection principle-based approximation algorithm. And finally, Simulated Annealing procedure is triggered with respect to the change in the number of alive nodes so that the algorithm modifies its input parameters according to the changing environment which, in turn, breaks the fixed operation structure of the algorithm.

- The impact of the designed methodologies is assessed in a multi-modal object classification problem of the WSN domain. A verification procedure consisting multiple methodologies is implemented in order to test the real-world usability and highlight the effect of the proposal. Additionally, the same optimization framework that is used for parameter tuning is employed with the aim of obtaining the approximate weights of sensor modalities in fusion rather than utilizing an expert knowledge in the weighted averaging procedure.

1.5 Organization

This dissertation starts with the essential properties of WSNs. This initial part normally addresses an audience of novices to the WSN field. Thereafter, related information about WSN application challenges and multi-modal content together with streaming issues are given. Then, the problem, scope, and contributions of this thesis study are provided.

In Chapter 2, background, the framework of the research, and related work which make up the big picture in the scope of the dissertation are presented. The motivation behind the study is stated in Section (2.1). In Section (2.2), data aggregation concept including clustering and routing aspects are introduced. Fuzzy logic utilization in WSN applications is elaborated in Section (2.3). Thereafter in Section (2.4), optimization in WSNs including the selected sample algorithms (approaches) together with their usage areas are presented. Object classification in WSNs is introduced in Section (2.5); this section also covers related information about multi-modality (2.5.1) and data-fusion (2.5.2) which concludes the chapter.

This guides the reader to Chapter 3 where a novel clustering methodology, entitled Two-Tier Distributed Fuzzy Logic Based Protocol (TTDFP), is introduced and described in detail. In this chapter, an overview of the methodology and system model are covered in Sections (3.1) and (3.2), respectively. Section (3.3) presents extensive information regarding the operation architecture of TTDFP. Since TTDFP is the applied methodology in the conducted empirical study presented in Chapter 5, comprehensive tests and analysis together with experimental evaluation of the approach are covered in Section (3.4) of the chapter. Concluding remarks are also given in the final Section (3.5) of this chapter prior to detailing into the type of optimization that can be utilized for improving the performance of fuzzy clustering algorithms.

Chapter 4 encompasses information and related descriptions of the proposed fuzzy optimization methodology for rule-based clustering algorithms which this dissertation partly based upon. Presented are an overview in Section (4.1), the methodology in Section (4.2) which is inspired from the Clonal Selection Principle (4.2.1) together with its modified and proposed version CLONALG-M (4.2.2), and experimental evaluation (4.3). At the end of the chapter, in Section (4.4), brief remarks are provided.

The discussion of topics until Chapter 5 puts the background into words and provides required knowledge for comprehending the empirical study in this chapter. Although some parts of this chapter are still targeting the readers with restricted background on clustering, optimization and multi-modality topics in WSNs, many crucial aspects by the way they collaborate are blended into a more coherent representation in Sections (5.3) and (5.4) of this chapter and also tested in a problem from the WSN domain throughout the rest of the chapter.

The final chapter of this thesis, Chapter 6, consists of conclusions, discussion, and foreseen future work which conclude the dissertation.

CHAPTER 2

THE FRAMEWORK OF THE STUDY

In the course of this chapter, the conducted survey of the literature is presented in a self-contained manner in order to clarify the big picture of the study. The survey starts with the motivation behind and proceeds with depicting the design principles of a generic data aggregation system in WSNs. This chapter is structured according to several well-defined aspects of multi-modal object classification domain implemented in WSNs. Additionally, a flow of general data fusion procedure which helps to exemplify the structure utilized in the empirical study on multi-modal object classification system in Chapter 5 is also presented. The chapter elaborates into the descriptions of the interacting variables of the WSN systems, which target multi-modal object classification domain, that serve as the building blocks of this thesis study and includes citations to the state-of-the-art researches.

As shortly depicted in the previous chapter, technological developments enabled the generation and utilization of wireless sensor nodes. Although a wireless unit is independently capable of gathering data, the deployed nodes generally collaborate in order to extract high-level semantic information from the viewed region. Networks of such nodes are referred to as WSNs. These nodes are powered by one or more energy sources which cannot be recharged easily in most cases, there are emerging energy harvesting methodologies though. As an apparent consequence, gearing these wireless devices with lightweight energy-harvesting techniques is a hot study field among researchers [7]. However, a great number of nodes in the market does not readily possess this technology because of its current cost, which prevents the implementation of such available methods. In this sense, decreasing consumed energy by the utilization of efficient approaches is still among primary aims to be achieved [8].

2.1 The Motivation Behind The Study

Prior to the dissertation study, a number of crucial aspects together with related gaps in the literature which must be investigated and addressed are identified if the interest is in real-world implementation. In this respect, the identified gaps are the insufficiency in data aggregation considering the energy aspect, optimized and distributed run-time behavior, and adaptability to the changing environment. It is valid to state that when the identified gaps are bridged, it is envisioned that the real-world usability of the studies in the literature becomes more concrete. For this reason, this study struggles to fulfill these gaps. To put it briefly, efficiency in data aggregation operations along with optimized and distributed computation necessities are the main motivations behind this dissertation.

From the dissertation title point of view, the study includes issues on Data Aggregation, Fuzzy Logic, Optimization, and Object Classification in WSNs. Corresponding sections regarding the survey of the literature are presented in the rest of this chapter.

2.2 Data Aggregation in WSNs

Data aggregation in WSNs tackles the compilation and delivery of data from generating wireless devices to the intended locations such as the base station (sink) where it can be utilized for required purposes. In general, there can be different sources of information and these sources are sensors geared onto wireless devices in WSNs. There are several sensor types including thermal, vibration, seismic, Passive Infra-Red (PIR), fluid, acoustic, and imaging sensor classes, and so on. From this perspective, different classifications of sensors exist in the literature and the major classification method relies on the internal principles of the sensor in hand. Because of the fact that the energy consumption of a single wireless node is utterly important for the whole network lifespan, it is a smart decision to utilize those types of sensors which demand or require the least energy. Seeing that the so-called data aggregation operation is based upon both the delivery and the compilation of data, routing approaches on top of clustering methodologies in this perspective provision energy-efficient infrastructure for the demanded task, respectively.

2.2.1 Network Clustering

Clustering is an advantageous procedure that can categorize sets of alike subjects (objects in hand) into piles or bunches, which are hereby referred to as clusters, in the given space. The subjects that are the members of a specific cluster are more alike to each other than those subjects in distinct clusters. A representative snapshot of a clustering procedure is denoted in Figure 2.1, where (a) captures a scene from the input subjects to be grouped and (b) presents the obtained clusters from the same scene. The subjects that are left ungrouped are not more alike to any subject in one cluster than in another cluster. Here, it is noteworthy to state that the subjects in the snapshot are clustered with respect to their locations on the two-dimensional space.

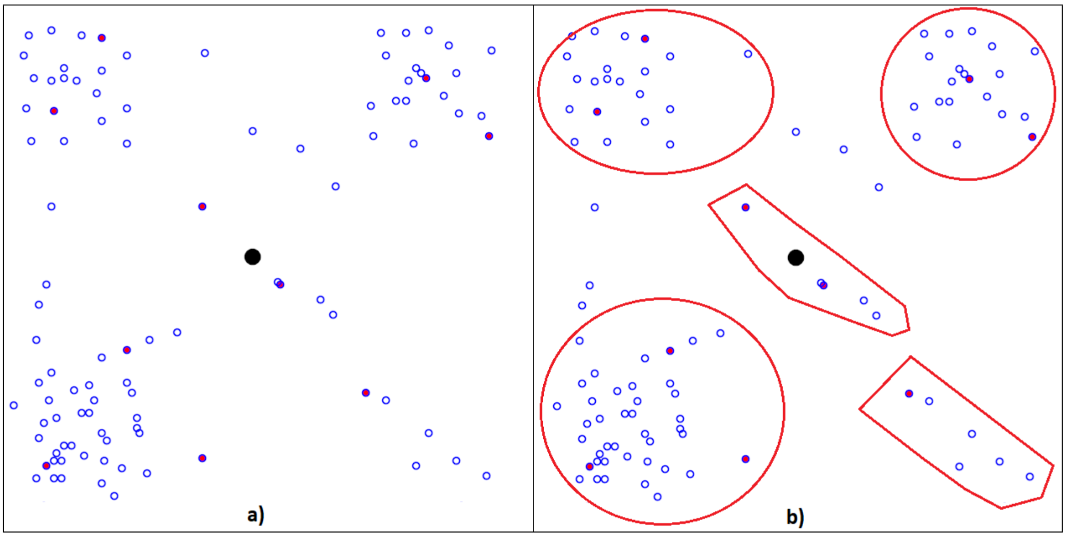

Figure 2.1: A representative snapshot of clustering.

In accordance with the clarifications made for clustering in general, deployed wireless devices in a WSN can be split into clusters. Although there are variations in specific application domains, in general, there exists a single cluster-head (CH) and sometimes one but rather more member nodes in each cluster. In the literature, CHs are also sometimes referred to as leaders since they lead data compilation process from member nodes then deliver obtained data to the intended localities. With the use of clustering techniques, the demanded performance can be provisioned [9][10] and the network scalability can also be augmented [11]. Additionally, there are further pros of clustering such as conservation of the restricted bandwidth, stabilization of topology, and preservation of energy [12].

The necessity of clustering in WSNs arises from various demands like reducing the amount and length of relayed packets and provisioning effective handover techniques to these packets. In the literature, there exist several studies clustering methodologies for sensor networks. In the upcoming parts, the main distinctive properties of selected prevalent clustering approaches are briefly discussed. However, an exhaustive analysis concerning clustering categories can be found in [13].

Local decisions are promoted to elect CHs in Low-Energy Adaptive Clustering Hierarchy (LEACH) [14], which depicts the protocol as a distributed algorithm. A probability model is utilized in its CH election phase. Data compression is performed by CHs in LEACH prior to sending packets to the base station. However, the operation of LEACH is not efficient due to its negligence of the remaining energy-levels of nodes.

The Hybrid Energy-Efficient Distributed (HEED) protocol is based upon the equal node assumption [15]. Parameters like energy, node degree, and distance are considered in two phases to elect CHs. However, HEED bears the hotspots problem inherently and results in an unbalanced energy consumption since it has an inclination towards generating more clusters than necessary [13].

Due to the inexactness in WSNs, like most application domains, various clustering algorithms such as [16] and [17] utilize fuzzy sets to cope with uncertainty. In the study of [16], the selection of CHs is done in the base station, which promotes this approach to be a centralized one. Cluster-Head Election mechanism using Fuzzy logic (CHEF) [17] is an alike proposal. However, the inclusion of the sink in the CH selection is not required.

Methodologies like the aforementioned ones experience the hotspots problem since they incline to compose equal-sized clusters in WSNs. Unequal clustering techniques are proposed so as to serve as a cure to the pitfalls of equal clustering studies. The basic motivation behind unequal clustering is the distribution possibility of energy consumption through efficient cluster-size adjustment. This situation differentiates the impact of inter and intra-cluster energy consumption pursuant to the distances of CHs with respect to the base station.

The properties of Energy-Efficient Unequal Clustering (EEUC) are its distributed and competitive operation mechanisms. Each node has an assigned competition radius in the protocol and local competition determines leaders [18]. As nodes get closer to the base station, their corresponding competition radii are reduced. Global data-gathering in EEUC results in increased energy cost and degrades the network efficiency, but it improves the network lifespan when compared to both HEED and LEACH [13].

The Multi-Objective Particle Swarm Optimization (MOPSO) protocol attempts to optimize the number of clusters generated together with the consumed energy [19]. The specific design of MOPSO is based on mobile networks and hence ignores the location of the nodes with respect to the base station, which causes MOPSO to be liable to the hotspots problem.

Multi-Objective Fuzzy Clustering Algorithm (MOFCA) is introduced to address the energy hole and the hotspots problems and increases the lifespan of WSNs [20]. It is a semi-distributed algorithm and struggles to solve the energy hole and hotspots problems with its fuzzy and distribution-agnostic approach. However, it does neither pursue an optimization framework nor an efficient operation considering the routing of the relayed packets.

Density-based Energy-efficient Clustering Algorithm (DECA) utilizes two parameters in clustering, namely the density of each node and energy levels [21]. DECA, like some aforementioned algorithms, does not take the node locations into account and cannot address depicted problems.

Dynamic Power Control Clustering (DPCC) methodology is based on multi-packet receipt technique [22], in which an interference cancellation algorithm in the sink is implemented. Although the study proliferates the clustering phenomenon with the interference cancellation approach, it is a centralized algorithm and does not fit into the requirements of WSNs since the cancellation algorithm runs on the base station.

Another related study is a fuzzy logic-based clustering algorithm using super CHs, which extends the lifetime of WSNs [23]. Here, a super CH is selected among CHs, and only this super CH can send packets to the sink. However, it is designed for direct communication and comparisons are only drawn against LEACH.

2.2.2 Packet Routing

Most depicted clustering algorithms of the previous subsection put emphasis only on the CH selection and do not consider how aggregated data is relayed to the sink, despite the fact that the efficiency of aggregation also depends on routing efficiency. Routing of the packets in a sensor network operates dissimilarly to traditional routing: Firstly, links are unreliable and there is generally no infrastructure due to some ad-hoc nature. Secondly, node failure-ratio is quite high and the next-hop node may not operate properly. Finally, energy conservation is a must and considered among primary goals.

Overall, wireless routing algorithms can be categorized into two as single-hop or multi-hop routing protocols with respect to the communication scheme. In single-hop communication, generated packet is directly sent to the receiving node without any intermediate node. Due to this feature, they are also known as direct communication protocols. These protocols are not suitable to be utilized for most sensor networks due to energy conservation requirements together with transmission range barriers. For this reason, multi-hop communication is often preferred over the direct transmission. However, there exist exceptions to this case which are beyond the scope of this dissertation.

In general, routing protocols in sensor networks can be categorized into three as *flat*, *hierarchical*, and *location-based* routing with considering the network structure. In the case of flat routing, nodes are equally important or have more-or-less equal functionality. In the case of hierarchical routing, which is the case for clustering, nodes possess different roles such as the generating node, the intermediate node (as in the case for CHs), or the final destination node (as in the case for the sink). In location-based routing, node locations over multi-dimensional space are utilized in order to efficiently route data in the network. Although there are various proposed methodologies considering each different category in the literature, elaborations are not provided in this subsection since routing itself is a broad topic. Additionally, it is valid to state that this dissertation focuses on hierarchical routing which is a type of multi-hop routing protocol. However, a detailed survey about the main routing protocols resides in [24] for further investigation.

In a WSN, it is possible to control the topology of the network, which is referred to as "topology control", by transmission range adjustment and/or specific message-forwarding node selection. This behavior can truly control the neighbor set of a node in a WSN [25]. Topology control directly relates to packet routing since the routed packets follow the established topology. There exist two major topology control categorizations which are homogeneous and non-homogeneous ones [26]. In homogeneous systems, transmission ranges of all the nodes are fixed and the same. However, in heterogeneous counterparts, nodes with differentiating transmission ranges exist.

Following the clarifications made in the previous paragraph, it is valid to state that clustering is also regarded as a way of controlling the topology since the primary aim is the organization of a network for the purpose of load balancing among nodes and extending the network lifespan [27]. In clustering, since member nodes are only allowed to send aggregated data to their CHs, there is no parameter like *k-connectivity* for leaf nodes; however, CHs may route data to the sink/s in a multi-hop manner where this parameter become crucial if the task is a mission-critical one.

In the literature, there are studies dealing with topology control problems. Specifically, the studies like [28], [29], and [30] are dealing with the power adjustment in the transmission which relies on the capability of nodes to adjust their transmission power. Similarly, the studies like [31], [32], and [33] utilize sleep schedule with the aim of decreasing the consumed energy when nodes are idle. There also exist other studies that employ geometrical structures, direction and location information for controlling the topology. Moreover, combinations of the mentioned methodologies are also possible [34] [35]. There exists prior work trying to achieve *k-connectivity* of any two nodes or between sensor nodes and super-nodes in heterogeneous sensor networks [36]. However, to the best of knowledge, none focus on or seek a more efficient transmission path than the available ones in a clustered environment.

2.3 Fuzzy Logic

A *Fuzzy* set is a set which consists of elements with differentiating degrees of membership. In the classical, also known as the crisp set theory, element membership degrees are evaluated in binary terms, an element is either a member of the set or not. In contrast, fuzzy set theory enables the fractional evaluation of the element memberships. This fractional evaluation is performed with the utilization of a *membership function* valued in the real unit range [0, 1].

A fuzzy set is the generalization of a crisp set because the membership function of a fuzzy set is the generalized version of an indicator function in a classical set [37]. There is an ever-increasing utilization of the fuzzy set theory in the literature seeing that it can be used in various fields such as robotics, dynamics, or monitoring where information is not complete, precise, or determined. The WSN field is among these domains due to inexactness of the operational nature of the wireless nodes.

The definition of a fuzzy set $\bar{A}$ is presented in mathematical terms in Eq. 2.1. In the equation, an element of the fuzzy set is depicted as x and the membership degree of the element x to the set $\bar{A}$ is depicted as $\mu(x)$.

$$\bar{A} = \{x, \mu(x) \mid x \in X\} \tag{2.1}$$

Major objectives of fuzzy sets are depicted in [38] and put forward as follows: inexact modeling, generalization (also known as relaxation), complexity reduction (compactification), and meaning preserving reasoning with the use of linguistic approximation. In the WSN research, fuzzy sets are mostly utilized in clustering, object detection, classification, and data fusion oriented proposals generally through applying linguistic approximation.

Although there are various studies on network clustering and the operation can be done either following a crisp approach or a fuzzy methodology, a significant number of studies propose fuzzy-based solutions to the clustering problem due to the power of fuzzy logic for effectively handling inexact values. The fuzzy-based methodologies are found to be superior to their crisp counterparts, especially when the number of clusters or the boundaries among them are uncertain. Additionally, fuzzy logic is utilized to produce solutions to a wide range of problems in different domains since it provides flexible solutions. However, the success of the solution greatly depends on the internals of the logic controller.

Fuzzy rule-based systems utilize fuzzy logic as a means of knowledge representation about the problem and model the correlations together with interactions existing among its variables. Rule bases are generally represented using matrices or tables and a generic form of a rule-table is given in Table 2.1. These rules depict the interaction between variables and generally lie in the fashion of IF (a) and (b) and (c) $THEN$ (R_1), which means that for example IF weather is COLD (X_1) and light is LOW (X_2) and weight is LIGHT (X_3), $THEN$ the $OUTPUT$ value of the function is (Y_1). In these systems, rules are evaluated by using a controller as an inference technique and a known method such as the Center of Gravity (COG) technique for the defuzzification of linguistic functions.

Table 2.1: A Generic Rule-Table

X_1 **Input Ling. Var. #1**	X_2 **Input Ling. Var. #2**	X_3 **Input Ling. Var. #3**	Y **Output Ling. Var.**
Cold	Low	Light	Y_1
....	....	....	Y_j
Warm	Medium	Normal	Y_k
....	....	....	Y_m
Hot	High	Heavy	Y_x

2.4 Optimization in WSNs

In this section, some of the basics of optimization and how it is used in WSN context in the literature are presented in accordance with the general framework of the dissertation.

In mathematics, optimization is a decision problem on the best element from an available set of alternatives considering one or more criterion [39]. An optimization problem is a maximization or minimization of a real function through the consistent choice of a set of input values from a range of values for each function parameter. A formal definition of an optimization problem can be stated as follows:

Definition (Optimization) *Given $f : D \rightarrow R$, where D is a domain of vectors of the* *form* $< x_1, x_2, ..., x_n >$ *and each* x_i *has a range of feasible values, then optimization is the procedure of determining an element* $x_{opt} \in D$*, such that* $f(x_{opt}) \leq f(X), \forall X \in D$.

In the definition, the f is named as an objective function and a feasible solution that the objective function can reach is named as an *optimal solution*.

2.4.1 Intelligent Clustering: What to Optimize?

As a term, *Intelligent Clustering* refers to the usage of *Artificial Intelligence (AI)* techniques in the process of clustering. In fact, this is true in most cases. In the literature, there are abundant studies using AI techniques in the clustering process. However, in this dissertation, it possesses the meaning of less-intervention of humans in the clustering process *parameter value assignment* rather than the utilization of AI in the clustering process itself.

Decreasing the number of CHs in a WSN is among vital problems in order to reduce contention of the channel and to increase the algorithm efficiency when realized at CH level. There are *few* studies including intelligence in the clustering process in the same respect as we do, the others differ in their approach. In [40], authors propose a Self Organizing Sensor (SOS) network which does not necessitate many user-defined inputs because of being based upon an intelligent clustering methodology. However, although they refer to their study an intelligent approach, their study includes user-defined parameters and authors manually assign values to the defined parameters. In [41], authors propose a Genetic Algorithm (GA)-based method in order to offer a solution to a problem of network optimization. Their approach utilizes a GA for the clustering process itself, not for the manual parameter assignment problem. As stated previously, in our proposal, intelligence possesses the meaning of less-intervention of humans in the clustering process *parameter setting*, not the clustering process itself. In this respect, this intelligent point of view can be utilized in two distinct and generally not related parameter sets in the WSN context:

- Algorithm-Dependent Parameters: Most algorithms in the literature possess various input parameters for the clustering process itself. This parameter set may consist of parameters such as remaining energy level, distance to base station, or density of a given wireless node. In the calculation of the values of output parameters such as the chance, probability of a wireless node to become CH, the depicted parameters are utilized. Seeing that these parameters are algorithm-dependent, they may be employed in some algorithm, but not in another algorithm. There is no general way of determining the presence of a parameter in a given algorithm since the same parameter might be utilized under different names or aliases. In this regard, each different algorithm assigns various values to these input parameters in an effort to obtain the output parameter value. This process is executed almost always in a trial-and-error manner, and whenever a suitable combination of the input parameters is encountered in the process, this combination is employed for the evaluation of the algorithm. This situation generally biases the final result and contribution of the study.

- Algorithm-Agnostic Parameters: There are parameters, such as tolerance and connectivity, which may be valid for the whole network. All nodes must adhere to this property. These parameters can be independent from the utilized algorithm and may be referred to as algorithm-agnostic parameters. The determination of the values takes the form of a brute-force search, which is exponential in a computational point-of-view in the average case.

In the scope of this dissertation, the focus is on optimizing the algorithm-dependent parameters. Because, to the best of knowledge, once the optimization methodology is decided, it is quite straightforward to adapt the operation of architecture from one parameter set to another.

Designing a rule-based fuzzy logic system as depicted in the previous section covers the definition of fuzzy sets that are generally depicted by membership functions and rules. As soon as the fuzzy system is designed, one of the foremost problems to handle is to establish the optimal placement of these described membership functions. In general, most membership functions in fuzzy control systems are presumed to be triangular or trapezoidal in shape and linear in a computational manner. For this reason, the actual problem is to decide on the value of the variables that characterize the shapes of these functions. If a suitable representation can be chosen, then the membership function approximation is reduced to a discrete optimization problem that could be modeled by using parameters in general.

The parameters that define the shape of fuzzy output functions in rule-based clustering algorithms are usually obtained from a field expert or generated automatically. Therefore, it consumes a considerable amount of time to determine the shape of these functions together with the tuning operations, and it is often impractical to design the optimal fuzzy system in detail. Various search algorithms are proposed in the literature with the aim of improving the behavior of fuzzy systems.

Genetic Algorithms (GAs) commence with an initial population of randomly constituted solutions (chromosomes) and moves toward better chromosomes with the modeled use of genetic operators in nature. For any given problem, GAs maintain a population of chromosomes and this population endures evolution through natural selection. Relatively better chromosomes give offspring to reproduce in order to replace relatively worse chromosomes which die in every generation. Fitness function serves as the environment which discriminates between better and worse chromosomes.

GA is employed the first in [42] for the specification of membership functions. In that study, authors apply GA in order to model a Fuzzy Logic Controller (FLC) for a problem. Two examples are presented in the study: an adaptive and a non-adaptive GA-designed controller where membership functions are transcribed in real time. There are membership functions which follow Gaussian distribution and the fitness function is a minimization which minimizes the difference between the center of the track and the cart. Similarly, [43] also studies the cart problem. In that study, authors follow a holistic methodology by utilizing GA in the system. Meredith et al. apply GA in order to fine-tune the membership functions in an FLC of an air vehicle [44]. Cheng et al. [45] choose image thresholds through the minimization of the measures of fuzziness. In that study, authors employ peak locations that are obtained from the histogram using peak selection criteria in an image segmentation problem for the automatic determination of fuzzy function bandwidths.

Self-Organizing Feature Maps (SOFMs) are a kind of Artificial Neural Networks (ANNs) which can be trained by utilizing unsupervised learning methodologies in order to generate a low-dimensional discretized exemplification of the input training samples. The mapping is done from the input to output space and the representation is called a map. For this reason, they are utilized as a method to conduct dimensionality reduction. SOFMs employ competitive learning and utilize a neighborhood function in order to maintain the topological features of the input space, which distinguish them from remaining ANNs that employ error-correction learning by using methods such as back-propagation. As one of its examples, Yang et al. [46] delineate a methodology in order to produce fuzzy membership functions utilizing SOFM. Their methodology is applied to a pattern recognition problem.

Tabu Search (TS) is a meta-heuristic search methodology which uses local search approaches utilized for mathematical optimization. TS gets a possible solution to a problem and inspects its immediate neighbors with the aim of finding a better solution. Because of its local search features, TS has an inclination towards suboptimal regions or on plateaus where many solutions are equally fit, which cause it to become stuck in local minimums or maximums. In the study of [47], an approach for determining membership functions sing TS is proposed. Additionally in the study of Cerrada et al. [48], an approach that allows the integration of the attitude of the system parameters with fuzzy membership functions is proposed.

In the light of aforementioned methodologies, it is wise to state what Clonal Selection (CS) is, since it is the principle adopted in the proposed optimization methodology in Chapter 4. CS is utilized in order to elucidate the main properties of an immune response to an antigenic stimulus. It is built upon the concept that solely the cells which are susceptible to identifying a stimulus eventually multiply and diversify into effector cells, therefore are chosen when compared to those that are not. Hyper-mutation and affinity proportional reproduction are among the main principles of the theory. The greater the affinity is the greater number of offspring produced. Every cell undergoes a mutation throughout reproduction which is inversely proportional to its affinity with the antigen. When compared to GAs, a standard GA does not preserve these immune properties that the Clonal Selection does [49]. When it comes to CLONALG, it is an exemplary CS algorithm and essentially formed to tackle unconstrained optimization problems, especially for combinatorial and multi-modal optimization. Antigens are exemplified by the fitness function, and antibodies are exemplified by the candidate solutions. CLONALG is primarily cultivated to conduct pattern recognition and machine learning tasks and then applied to the optimization problems [50].

2.5 Object Classification in WSNs

Detection and localization are two related procedures performed about a phenomenon when the major goal is object classification in a WSN. Firstly, detection, or sometimes referred to as discovery, of an object is done through the use of sensors. Then, the location of the object is acquired, and thereafter the recognition (identification) process commences. In this respect, classification is a convenient process employed for the recognition of an object. In classification, the detected object is categorized with respect to its features and the target types or classes are generally defined prior to the deployment phase after a learning cycle which is referred to as training of the network.

Although localization and classification of objects are based upon distinguishable features, the process of classification could occur between arbitrary boundaries with respect to the problem domain and generally independent of detection. In a technical point-of-view, detection either occurs or not, that's to say, ambiguity is not an option, perhaps a variable with *True* or *False* boolean values. Available features of an object are extracted by the utilization of extraction methodologies and these features are fed into classification process. In general, WSNs are trained offline and discriminative features of the target classes are introduced to the network for accurate classification of objects into predefined classes. The classification process could take place in a leaf node, where the real detection occurs, at an intermediate node, at the sink(s) where the restriction on resources is not as tight as the other nodes mostly, or with the collaborative effort of these distinct type of nodes. If the context is sensor networks, it is possible to extract depicted discriminative features by utilizing single or multiple-modalities for the classification process. Here, it is wise to elaborate on the details of multi-modality which would enlighten the object classification path.

2.5.1 Multi-Modality

In this dissertation, the vocable "modality" is utilized in a meaning-preserving sense from the *semiotics* domain. In this domain, this vocable is described as follows:

Definition (Modality) *A modality is a certain type of information encoding for rep-*

resentation purposes. A particular class or type of representation format in which information is stored is meant by the use of the word "modality". [51]

In the definition, modality is ambiguously put forward and needs elucidation. Various definition efforts are encountered in the literature and the *source of information* aspect is seen to be adopted in the domain of object classification. Objects are represented by their features. Each feature can be regarded as a distinct modality at one extreme and the combination of all features can be regarded as one modality at the other extreme [52]. A trade-off can be found by clustering the features with respect to defined criteria.

The choice of modalities in the WSN community is performed with respect to one of the mentioned possibilities. However, there exists no definite evidence about which method of feature combinations offer the expected result [52]. In sensor network context, each sensor type can be depicted as a modality due to its distinctive source of information property. From this perspective, multi-modal object representation in the WSN context can be done using a different type of sensor such as velocity (accelerometer), seismic, vibration and/or acoustic sensors. If the application type in the sensor network targets multi-media or surveillance applications, these depicted scalar sensors can be augmented with camera (imaging) sensors. Due to the fact that there exist various types of sensors employable in sensor networks other than the depicted ones here, the aforementioned are chosen with respect to their profoundness in today's WSNs.

In studies targeting object classification in the literature, the multi-modality subject frequently appears in biometrics or multimedia domains since these fields of research utterly utilize the multi-modality property to be able to represent and reveal the structure of the complex data. For instance, humans are recognized and differentiated according to physical and behavioral characteristics in the biometrics domain. Since recognition by the use of a single trait is rarely possible or enough, each trait is regarded as a different modality. More or less the same situation is valid for multimedia usages in WSN context targeting surveillance applications.

If multiple-modality is the case for various sensors in the domain, then the object classification process may achieve higher accuracy results when the combined effort among available modalities are utilized. However, the vital point here to be taken into account is the computational complexity and the energy-efficiency of the integration process.

2.5.2 Data Fusion

Fusion can be expressed as the composition of multiple source defining the same entity into a single but more coherent, accurate, and beneficial state. The utilized data can either be in the raw or processed state as in the case of knowledge/information. The fused data generally provides a more informative representation of the subject matter than that provisioned by any *single* source. Data fusion is a tempting process since it is applied for better understanding the complex structure of the entity. A similar expression "information fusion" is sometimes utilized interchangeably for data fusion, or also some other times utilized in order to distinguish the fusion of information (a form of processed data) from the fusion of raw data.

In the WSN context, data fusion is among hot topics recently since it has abundant usage areas. What makes data fusion alluring for WSNs is its impact over energy-efficiency. In other terms, the primary duty of most WSNs is data aggregation and the so-called collected data can be compacted largely by using fusion techniques, which in turn, decreases energy consumption in data transmission, where resource-constrained nature is an inherent pressure over energy-powered nodes. Although there are various and sometimes completely intriguing definitions of data fusion, the adapted definition in this dissertation can be found in [53]. A flow of common data fusion procedure implemented in most sensor network applications is presented in Figure 2.2.

The "source" component of the procedure depicted in Figure 2.2 can be in any of the modalities such as sensors or features. The primary reason for fusion strategies applied in WSN applications can be said as energy-efficiency and accuracy issues, which are also the direct results of working with combined sources. The reasons below are the main arguments in order not to rely on a single source:

- Fusion of supplementary data ensures a more descriptive exemplification of the

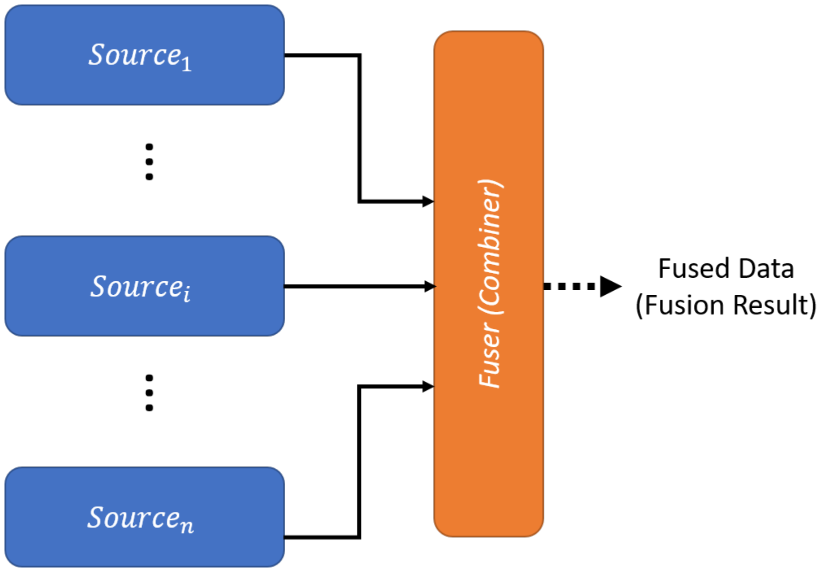

Figure 2.2: A flow of general data fusion procedure.

subject matter. Additionally, when different fusion sources consist of redundant data then inexactness or vagueness of a decision decreases. Thus, robustness and reliability of a possible decision increases [54],

- One of the functional benefits of data fusion resides in its unreliable source filtering property. In real life applications, it is often unknown to deduce the performance or added value that is brought by additional modality at the design time. Not only performance but also the reliability of a modality cannot be predicted correctly in real-world environments. For this reason, dependency on any of the available modality or their weights on the overall result can together be decreased [54],

- Noise is an undesirable property which may cause incorrect classification results. Multiple sources in this respect can operate as a filter and decrease the effect of noise on the result [51],

- While no source is impeccable, the collaborative effort among them generally produces better results hence improving usability [55],

- Each source has an implicit upper bound on the overall performance. Therefore, no system performance can repeatedly be tuned through modifying its components or operation steps [56].

2.5.2.1 Fusion Methodology

The algorithm utilized for the combination of input sources is referred to as "Fusion Methodology". There exist extensive proposed research in the literature and experimentally evaluated in the context of fusion. However, there is no consensus on a superior methodology among researchers regardless of the application domain [57]. The prediction of whether or not a combination to perform superior is difficult, and from this perspective, no clear preference can be made over a combination methodology [58].

Available algorithms in the literature proposed in the context of fusion can be divided into two main categories: *trained* and *non-trained* approaches. Trained approaches are mostly not preferred for WSNs due to their dependency on training data. Non-trained approaches are based upon voting and linear aggregation methodologies. Among these, minimum selection, product aggregation, maximum selection, sum aggregation, majority voting, weighted averaging, linear combination, and concatenation are highly utilized ones. The success of the depicted methodologies lies in their operation simplicity.

There are three widespread operation architectures (modes) of a fusion system. These are: serial architecture, parallel architecture, or hybrid, also referred to as hierarchical, architectures. Each depicted architecture defines the way in which available sources utilized in a fusion. In a serial architecture, sources are combined incrementally while in a parallel architecture they are combined at once (in parallel fashion). In hybrid implementations, sources may be divided into layers (levels) and at each layer, selected sources are utilized in parallel, serial, or in a combined manner of these two modes. Rather detailed explanations about these architectures are as follows:

- The combiner is utilized in more than one step in a serial operation mode. In any step, one remaining source from the available inputs is fused with the output of the former step. The result of a former step is typically utilized in order to straighten the number of possible outputs prior the utilization of another source.

- The combiner is utilized at once in a parallel operation architecture. There exists only a single fusion step. For this reason, fusion is performed over available source without the need to wait for other sources, if exists, to arrive. The best case where a parallel scheme can be utilized is the case when all input modalities are accessible at once at the time of fusion.

- Both aforementioned architectures are employed in a hybrid implementation as the name suggests. For this reason, there surely exists mode than one fusion step in this type of schemes. In each step, one or some of the sources are fused in parallel and the output is fed forward to the following step.

2.5.2.2 Fusion Design

Apart from the fusion methodology, there exists a categorization of fusion systems with respect to the time where the classification of a phenomenon is conducted: *Early Fusion* and *Late Fusion*.

Typically, the fusion conducted by a system is referred to as *early* when the fusion algorithm is applied prior the classification process. The reason behind early fusion is that the obtainable information prior to classification is considerably more than late fusion due to the unprocessed or unclassified data access. Although it is more feasible to uncover hidden relations existing among different sources (modalities), it is also a computation-intensive task which consumes much more energy. Additionally, a great amount of training data is required in order to train such a system. Therefore, early fusion is commonly known as an effective but also a complex type of fusion [54], [59]. A representation of an early fusion structure is presented in Fig. 2.3. In the figure initially, feature extraction from the sample is performed. Then, extracted features are utilized directly in the fuser (combiner) without any prior classification process. Finally, a learning methodology is performed right after the fusion of inputs.

Furthermore, the fusion, when performed after the classification process, is referred to as *late* fusion. Unlike early fusion, this type of fusion is uncomplicated since results of prior classification processes are utilized, which is much more interpretable. On the other hand, available information is much restricted than early fusion due to unavailability of raw data. Moreover, the chances in order to effectively correlate results are low. Therefore, this type of fusion is computationally less complex but

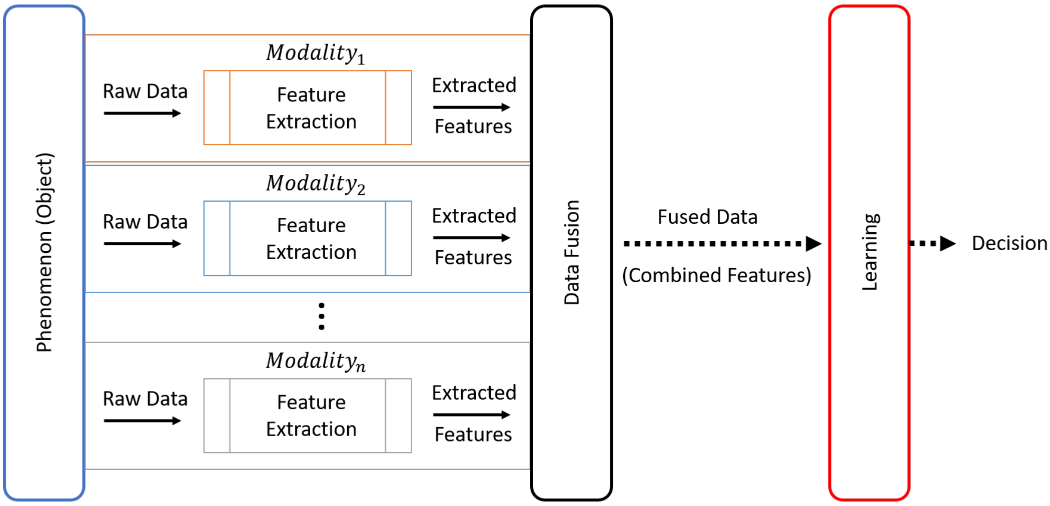

Figure 2.3: A representation of an early fusion structure.

also less effective in the overall result [54]. A representation of a late fusion scheme is demonstrated in Fig. 2.4. As opposed to an early fusion scheme presented in Fig. 2.3, the final decision of a system can only be made either by a simple aggregation approach or by an higher-level learning methodology [60].

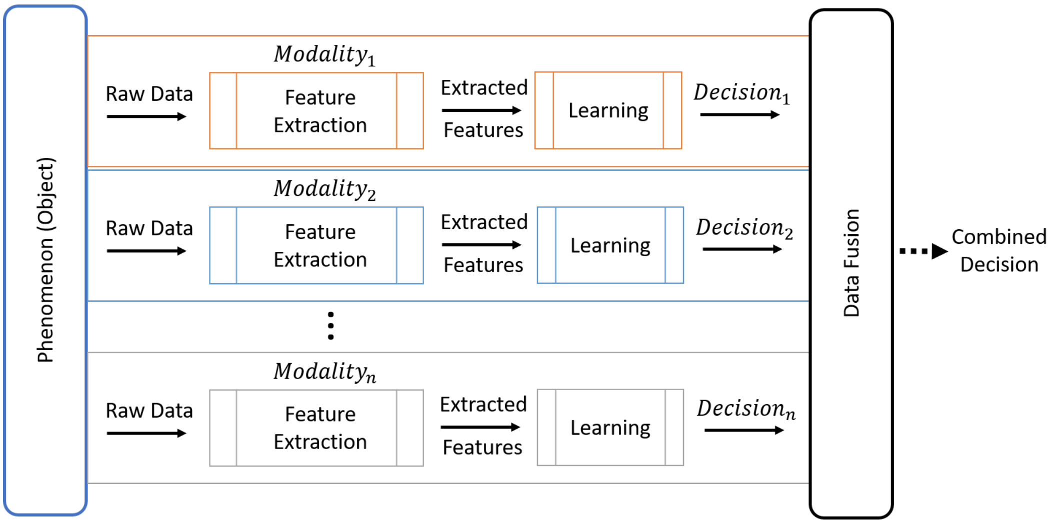

Figure 2.4: A representation of a late fusion scheme.

According to the clarifications done, early fusion can also be referred to as fusion of values whereas late fusion can be referred to as fusion of decisions. Each approach has its own pros and cons and has distinct application areas. It should also be noted that the overall computational complexity of a late fusion architecture can also be quite high depending on the computational complexities of each classifier utilized in the process. Although the implementation of fusion is considerably easy, the depicted computational complexity problem may deter the operation of a late fusion scheme.

In the light of aforementioned reasons, it is valid to state that transmission data volume can be reduced with the use of fusion techniques in data aggregation scenarios. In accordance with the data volume reduction, energy consumption can also be decreased on a large scale. Moreover, as a direct effect, total network lifespan increases because there exists a larger number of operating nodes at some time t at the network when it is compared against fusion-less data gathering methodologies.

Moving object detection, event recognition, clustering or classification of objects are among the widespread utilization fields of fusion. A sample study utilizing fusion for detecting targets in a distributed manner resides in [61]. In the study, a description of a Fusion Center (FC) is provided and outputs produced at sensor-level are fused at the depicted FC. As depicted in the study of [62], cluster-based fusion, where nodes form clusters and data is fused at cluster heads (CHs), is an important fusion methodology.

In the literature, the effect of data fusion over detection delay in real-world scenarios is studied and presented in [63], which depicts the improvement considering the performance requirements such as short detection delay in real time scenarios. There are various approaches regarding where and how to apply the fusion process and [64] presents aggregation tree based methodologies together with related processing techniques among these approaches. A detailed classification of aggregation approaches in the domain of WSNs and data fusion can also be found in [65] and [66], respectively.

To summarize, the background information covered in the sections of this chapter directly relates to the proposed methodologies covered in the upcoming chapters. This concludes the framework and the big picture of this dissertation.

CHAPTER 3

A TWO-TIER DISTRIBUTED FUZZY LOGIC BASED PROTOCOL (TTDFP) FOR DATA AGGREGATION

In this chapter, a two-tier distributed fuzzy logic based protocol in order to improve the efficiency of data aggregation operations in multi-hop WSNs is proposed. Clustering is utilized for efficient aggregation requirements in terms of consumed energy. In a clustered network, member (leaf) nodes transmit obtained data to cluster-heads (CH) and CHs relay received packets to the sink. This CH-generated transmission occurs over other CHs in multi-hop wireless networks. Due to the adoption of a multi-hop topology, hotspots and/or energy-hole problems may arise.

Although the main focus of many studies in the literature is on energy-efficient clustering, to the best of knowledge, none takes the efficiency of clustering and routing phases jointly into account. The proposed protocol is distribution-adaptive that runs and scales efficiently for sensor network applications. Additionally, along with the two-tier fuzzy logic based protocol, an optimization framework is utilized to tune the parameters used in the fuzzy clustering tier in order to optimize the performance of a given WSN.

This chapter also includes performance comparisons and experimental evaluations with selected state-of-the-art studies. Obtained experimental results reveal that proposed protocol performs better than any of the other protocols under the same network setup considering metrics used for comparing energy-efficiency and network lifespan of the protocols.

3.1 Overview

In clustering, efficiently selecting cluster leaders can significantly decrease consumed energy. For this reason, there is an ongoing thorough study concerning selection mechanisms in the literature. The most common point of the proposed solutions is the utilization of a two-stage process: in the first stage a CH with more energy is selected, and then in the second stage leadership is transferred among member nodes with the aim of balancing consumed energy. This common point actually retains two crucial hidden pieces of knowledge:

- Some of these methodologies like [14] and [16], which will be detailed in the next section, lack other necessary relevant information such as the location or connectivity of a node that can be utilized as input parameters in the clustering process or possess a centralized operation architecture. For this reason, the final clusters cannot meet the expected efficiency demand. Additionally, because of not taking the position of each node into account, the hotspots or the energy hole problems may arise in WSNs. The hotspots problem is related to the premature death of the leaders that are around the base station or on busy routes because of the intense inter-cluster relay. Similarly, the energy hole problem is related with the early energy depletion of some close nodes that are located in an area which degrades or sometimes completely impedes the transmission of the relayed data to the intended location. Besides, the energy hole problem may also occur in evolving networks since initial deployment location of nodes may change drastically. This variable node location situation has a great influence on the density and the connectivity of the nodes over time. This differentiation in node location is particularly important in heterogeneous or multi-hop sensor networks when compared to homogeneous or single-hop networks.

- The available solutions try to master *only* in clustering to prolong the lifespan of WSNs. On the contrary, not only clustering but also routing phase should pursue an energy-efficient operation in multi-hop environments. Otherwise, most of the energy preserved from clustering would be overspent at the routing phase. This situation can also overbalance the energy distribution of the network which has a negative effect on clustering. It is observed that the utilization of a multi-hop routing methodology on its own instead of direct communication is not enough. For this reason, exploiting the power of fuzzy logic in the routing phase over its crisp counterpart becomes a wise choice.

As the main contribution of this chapter, a Two-Tier Distributed Fuzzy Logic Based Protocol (TTDFP) for the purpose of addressing data aggregation problems of multi-hop wireless sensor networks is proposed. In the first tier (Tier I), TTDFP decides upon the final CHs through an energy-based competition of provisional leaders, which are primarily chosen by means of a probabilistic model. TTDFP is a competitive, fully-distributed, and an optimized protocol considering the lifetime requirements of the WSNs. TTDFP does not require the inclusion of a central decision-making point during any of its phases. This distributed operation architecture protects the protocol from the single-point-of-failure situations. Fuzzy clustering phase handles uncertainty occurring in the clustering phenomena more efficiently than its crisp and fuzzy counterparts. The optimization framework [67] is also utilized to tune the two parameters in this tier, which are the *Maximum Competition Radius* and *Threshold*, rather than using a trial-and-error approach to find the right blend of these parameters. The optimization framework employs the Simulated Annealing algorithm to tune the aforementioned pair of parameters with the aim of optimizing the performance metrics of WSNs. Additionally, fuzziness in the second tier (Tier II) is a novelty which also enhances routing performance when compared to its crisp counterpart. Bearing in mind these contributions, it is valid to state that TTDFP is a candidate approach to be utilized in mission-critical real-life applications where utilization context may be updated with respect to the application domain.

In the clustering phase, TTDFP uses three fuzzy parameters. These are *relative node connectivity*, *distance to the base station*, and *remaining node energy*. In the routing phase, TTDFP proliferates the operation architecture of a known multi-hop routing methodology with the use of fuzzy logic. In this second tier, TTDFP utilizes two fuzzy parameters, *average link residual energy* and *relative distance* to determine an efficient routing path. In order to observe the competency of the proposed solution, it is experimentally evaluated against a selected distinctive set of the existing clustering mechanisms such as LEACH [14], CHEF [17], EEUC [18], and MOFCA [20]. Several experiments are conducted under various operating conditions. Performance results demonstrate that TTDFP is an efficient protocol that outperforms the other algorithms it is compared against under the tested experimental setups.

Here, it is wise to state that there are approaches (i.e. [19]) that employ optimization for the clustering phase *itself* in order to find an optimal solution. However, as far as known, the proposed protocol is the first study that employs fuzzy logic not only for clustering and routing jointly, but also it utilizes an optimization technique for tuning the parameters in clustering of WSNs.

Although there are various optimization algorithms, here Simulated Annealing (SA) algorithm is depicted in short since it is the approach that is utilized in TTDFP. SA depends on a probabilistic mechanism in order to approximate the optimum of a given function. Specifically, it is meta-heuristic utilized in large search spaces. It is often employed in discrete search spaces. However, it can also be utilized in continuous search spaces. SA solely requires a single initial point as a starting point and a search operation [68]. Since optimization is a vast topic in itself, interested readers may refer to the literature for more details.

3.2 System Model

The proposed solution is built upon the characteristics of the system model given below. This system model is also utilized for experimental evaluation. In the model, member nodes send data to their CHs in direct (single-hop) communication method. CHs accumulate these data, aggregate them and transmit to the sink by employing a multi-hop routing methodology. The representative snapshot of the utilized model is presented in Fig. 3.1.In the figure, the sink is depicted with a larger black dot which resides in the service area, CHs are marked with red dots, member nodes are plotted with blue dots, and competition radii of final CHs are marked with dotted circles. Lines colored in black depict direct communication channel and lines colored in green depict the multi-hop communication channel.

The assumptions in the system model can be described as follows:

- Nodes are the same considering their hardware components.

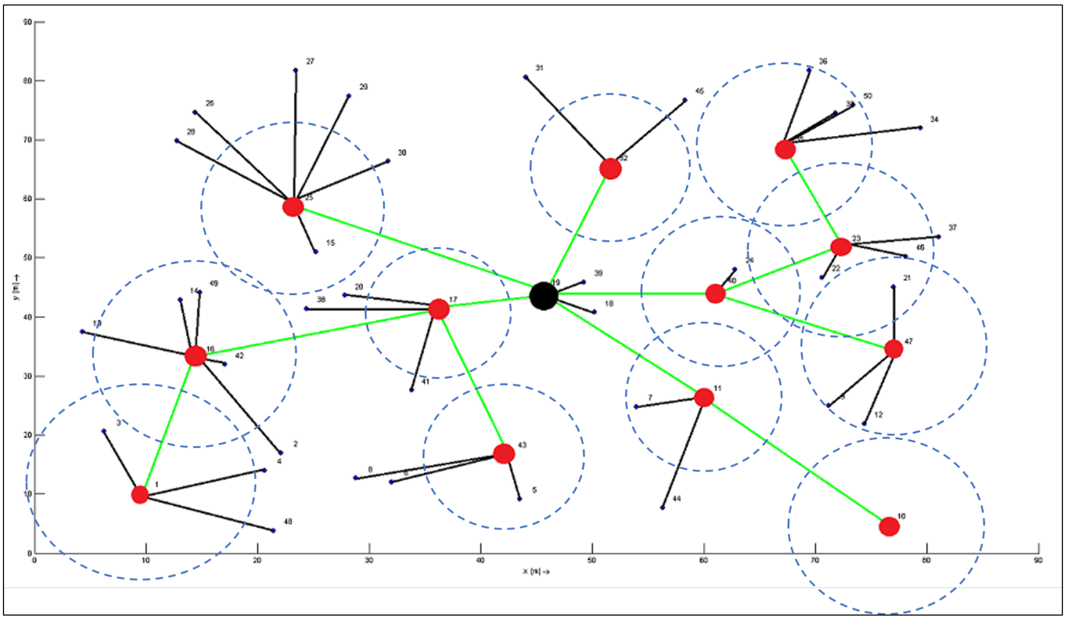

Figure 3.1: The representative snapshot of the utilized model.

- Nodes are deployed into the area randomly, following uniform distribution model and the base-station can be anywhere.
- Nodes may change their initial locations and the displacement of nodes does not consume energy.
- Nodes have the same amount of energy at the deployment phase, which is equal to 1 joule (J).
- Transmission power can be adjusted considering the distance of the receiver.
- Received Signal Strength (RSS) is employed to compute the distance between peers.

Deployed wireless nodes are modeled as a W-vertex graph G, which consists of N vertices and P edges as depicted in Eq. 3.1 and Eq. 3.2, respectively.

$$G = (N, P) \tag{3.1}$$

$$N = \{1...W\} \tag{3.2}$$

Energy consumption is represented by the First Order Radio Model as exploited in [14]. Discharged energy calculation in transceiving b bits to or from a distance of d can be measured as in Eq. 3.3 and Eq. 3.4, respectively. In the equations, E_{elec} = 50nJ/bit, ε_{fs} = 10pJ/bit/m^2, ε_{mp} = 0.0010pJ/bit/m^4 and d_0 = 20 m. E_{elec} depicts consumed energy per bit in the transceiver circuits and ε_{mp} depicts the energy consumption of a bit in RF amplifier.

$$E^{TX}(b,d) = \begin{cases} bE_{elec} + b\varepsilon_{fs}d^2, d < d_0 \\ bE_{elec} + b\varepsilon_{mp}d^4, d \geq d_0 \end{cases} \tag{3.3}$$

$$E^{RX}(b) = E^{RX-elec}(b) = bE_{elec} \tag{3.4}$$

3.3 Operation Architecture

The proposed protocol (TTDFP) includes two tiers. In the first tier, the proposed fuzzy clustering algorithm elects the set of CHs that maximize the network energy efficiency. Then, in the second tier, the optimal routing path from CH to the sink is sought by the developed fuzzy routing procedure. Finally, these two architectures are unified into a two-tier protocol to provide an efficient data aggregation methodology.

3.3.1 Tier I: Distributed Fuzzy Clustering Phase

Distributed Fuzzy Clustering Phase is devised by considering three crucial elements. The first one is energy-efficiency, the second is distributed operation requirements which provide its scalability and finally optimized run-time configuration. The clustering approach is a fuzzy distributed unequal clustering approach that utilizes local commitments in the specification of radius values and CH election. Thus, the inclusion of the base station during the election process is not required. The reason behind devising the protocol is to cope with the energy hole and the hotspots problems of WSNs in an optimized manner while still pursuing a distributed efficient operation.

The proposed distributed fuzzy clustering phase considers three parameters, which are remaining node energy, relative connectivity of a node, and distance to the base station with the goal of determining the value of competition radius for provisional CHs. Competition radius of a node is the radius which defines the area where the node competes with the other nodes located in the area prior to final Cluster Head (CH) selection. Additionally, fuzzy logic is utilized in the computation of the competition radius values. Fuzzy variables handle inexactness efficiently. The proposed protocol utilizes a probabilistic model in the selection of provisional CHs and employs randomized recurrent spinning as in [18] or [20]. The proposed approach is *fully-distributed* since the inclusion of the base station is not required in any decision stage. Additionally, there are no manually assigned algorithm-dependent parameter values in the proposed protocol. This is accomplished through the use of an optimization methodology employed in the clustering tier. The exchanged messages in the clustering tier are depicted in Algorithm 1. E_i, $Comp_i$, and S_i stand for the remaining energy, competition radius, and status of a specific node i, respectively.

When a clustering round starts, a random number (μ) is generated by each node within the interval [0,1]. Whenever μ of a specific node is smaller than the optimized threshold value (Th), then that node (i) turns into a provisional CH. Here, the optimized threshold value depicts the ratio of desired provisional CHs. The distributed fuzzy clustering phase employs relative node connectivity, distance to the base station, and remaining energy values. In this phase, competition radius values are modified in each round. The radius calculation is performed by the fuzzy inference engine. This operation is indicated in line number 7 in Algorithm 1. The employed fuzzy rules for the fuzzy inference engine are depicted in Table 3.1. For the rule evaluation process, the Mamdani Controller [38] fuzzy inference method together with the Center of Gravity (COG) technique is utilized in the competition radius defuzzification process.

$$CoG = \frac{\int_{x_{min}}^{x_{max}} f(x) * x dx}{\int_{x_{min}}^{x_{max}} f(x) dx} \tag{3.5}$$

In the defuzzification, fuzzy logic controller initially computes the area within the boundaries of the output descriptor under the membership functions, thereafter calculates the geometrical center of this area by using Eq. 3.5, where CoG is the calculated

Algorithm 1: TTDFP Tier-I: Fuzzy Clustering Algorithm

Input: A Non-Clustered Network

Output: A Fuzzy-Clustered Network

$Th \leftarrow$ Optimized value using SA
$S_i \leftarrow$ CLUSTERMEMBER
clusterMembers $\leftarrow$ NULL
myCH $\leftarrow$ This (self)
beProvisionalCH $\leftarrow$ TRUE
if *(μ <Th)* **then**
 By utilizing fuzzified input descriptors, form $Comp_i$
 Communicate Candidate (Id, $Comp_i$, E_i, C_i)
 On the reception of Candidate from Node j
 if *($E_i < E_j$)* **then**
  beProvisionalCH $\leftarrow$ FALSE
  Communicate CeaseElection(Id)
 else if *(($E_i = E_j$) and ($C_i \leq C_j$))* **then**
  beProvisionalCH $\leftarrow$ FALSE
  Communicate CeaseElection(Id)
if *(beProvisionalCH = TRUE)* **then**
 Communicate CHMessage(Id)
 $S_i \leftarrow$ CLUSTERHEAD
 On receiving JoinCH(Id) from Node j
 clusterMembers $\leftarrow$ ADD(j)
 EXIT
else
 On the reception of all CHMessages
 myCH $\leftarrow$ The closest CH
 Communicate JoinCH(Id) to the closest CH
 EXIT

center of gravity, x is the linguistic variable value, and x_{min} and x_{max} stand for range boundaries.

For the calculation of the competition radius of a specific CH, a total of three fuzzified input parameters are utilized. The first descriptor is the *Distance to the Base Station*. This fuzzified descriptor is depicted in Fig. 3.2. *Far*, *medium* and *close* are the selected linguistic variables of this fuzzy parameter.

The second descriptor is the *Node Remaining Energy*. This fuzzified descriptor is depicted in Fig. 3.3. *High*, *medium* and *low* are the selected linguistic variables of this fuzzy parameter.

The third descriptor is the *Relative Node Connectivity* of a specific node. The fuzzy set which characterize the relative node connectivity is indicated in Fig. 3.4. *High*, *medium* and *poor* are the selected linguistic variables here. This fuzzy variable is utilized with the same purpose of the density parameter as in [20]. However, in contrast to the density parameter, the calculation of relative connectivity is done in a fully-distributed manner. Relative connectivity of a provisional CHs (C_i) is computed as in Eq. 3.6. With the help of the exchanged messages, a node is already aware of its own connectivity value. Then, the maximum connectivity in its radius is used as a normalization factor as depicted in the denominator of Eq. 3.6.

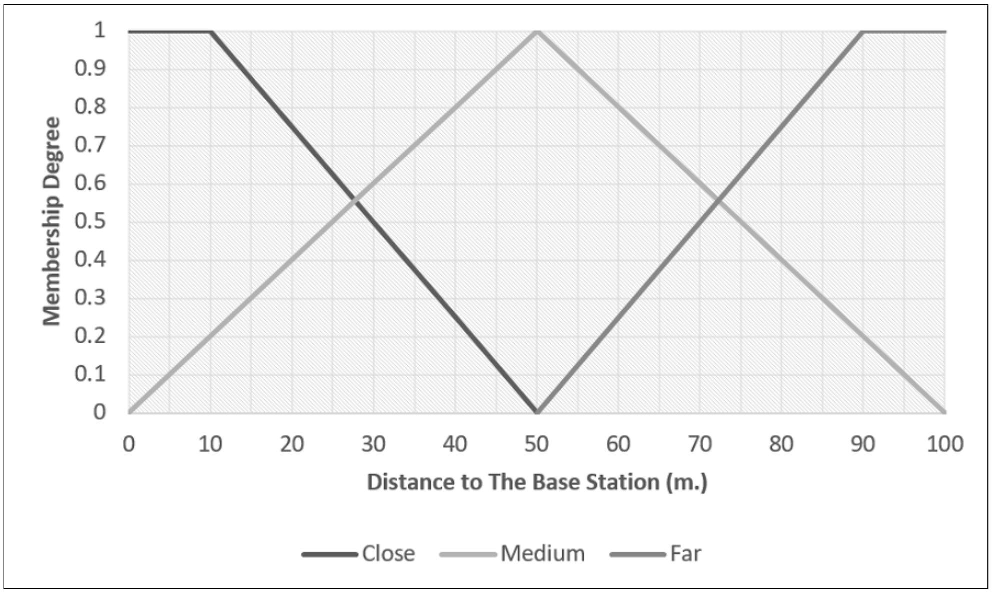

Figure 3.2: The fuzzified parameter Distance to The Base Station.

Table 3.1: Fuzzy Rules in Tier-I: Distributed Fuzzy Clustering Phase

Distance to the Sink	**Remaining Energy**	**Relative Node Connectivity**	**Competition Radius**
Close	Low	High	CR_1
Close	Low	Medium	CR_2
Close	Low	Poor	CR_3
Close	Medium	High	CR_4
Close	Medium	Medium	CR_5
Close	Medium	Poor	CR_6
Close	High	High	CR_7
Close	High	Medium	CR_8
Close	High	Poor	CR_9
Medium	Low	High	CR_{10}
Medium	Low	Medium	CR_{11}
Medium	Low	Poor	CR_{12}
Medium	Medium	Poor	CR_{13}
Medium	Medium	Medium	CR_{14}
Medium	Medium	High	CR_{15}
Medium	High	Poor	CR_{16}
Medium	High	Medium	CR_{17}
Medium	High	High	CR_{18}
Far	Low	Poor	CR_{19}
Far	Low	Medium	CR_{20}
Far	Low	High	CR_{21}
Far	Medium	Poor	CR_{22}
Far	Medium	Medium	CR_{23}
Far	Medium	High	CR_{24}
Far	High	Poor	CR_{25}
Far	High	Medium	CR_{26}
Far	High	High	CR_{27}

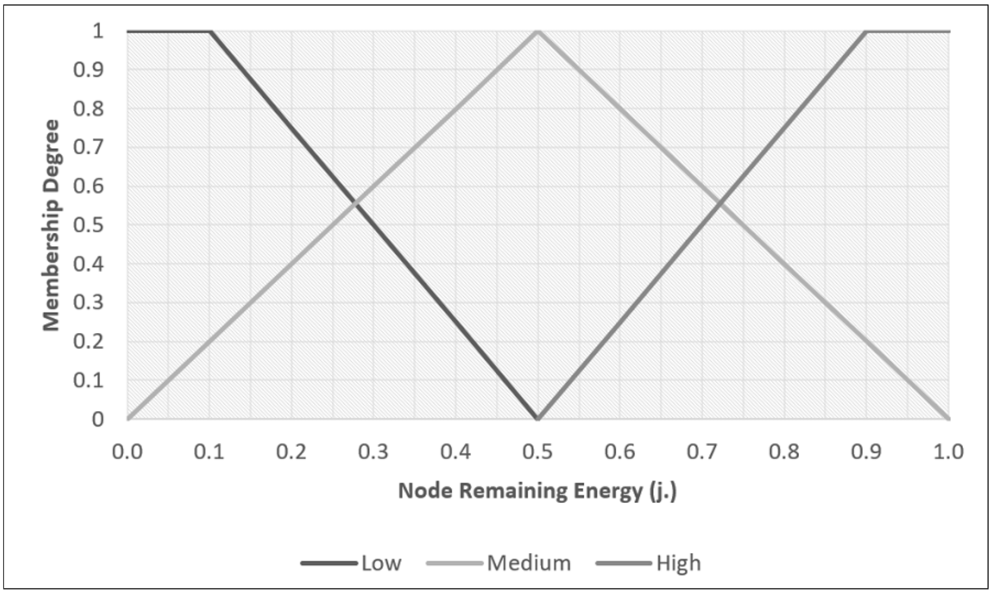

Figure 3.3: The fuzzified parameter Node Remaining Energy.

$$C_i = \frac{Connectivity_i}{max(Connectivity)inRadius} \tag{3.6}$$

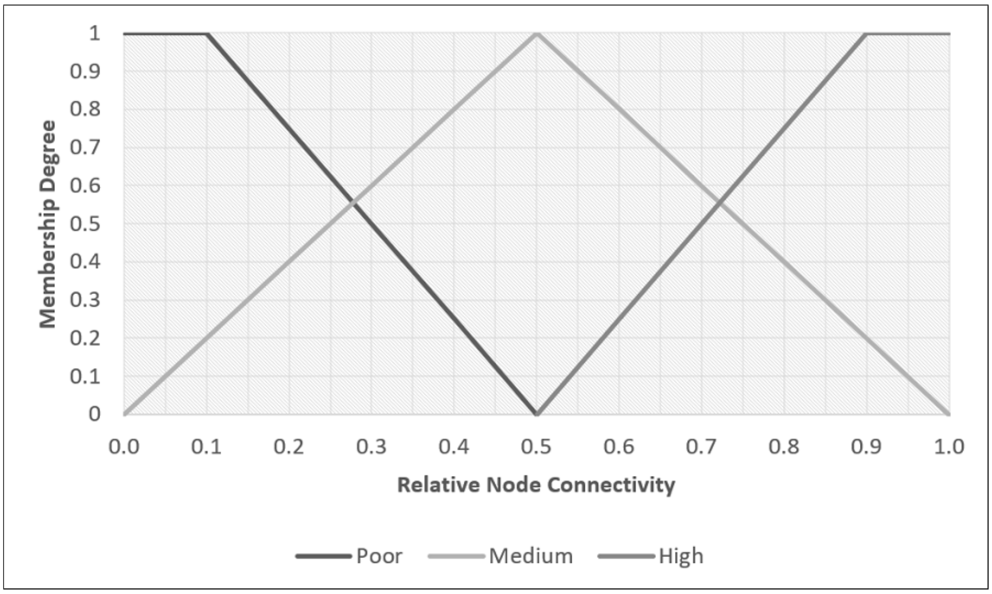

Figure 3.4: The fuzzified parameter Relative Node Connectivity.

There is only a single fuzzified output descriptor named as the competition radius of a provisional CH. The fuzzy set of this fuzzified output descriptor is presented in Fig. 3.5. There exist 27 linguistic variables denoted as CRi. Only $CR1$ and $CR27$ have trapezoidal membership functions, the rest of the linguistic variables are described by triangular functions. In order to compare the proposal with MOFCA, the same fuzzy output function for competition radius is utilized as in [20].

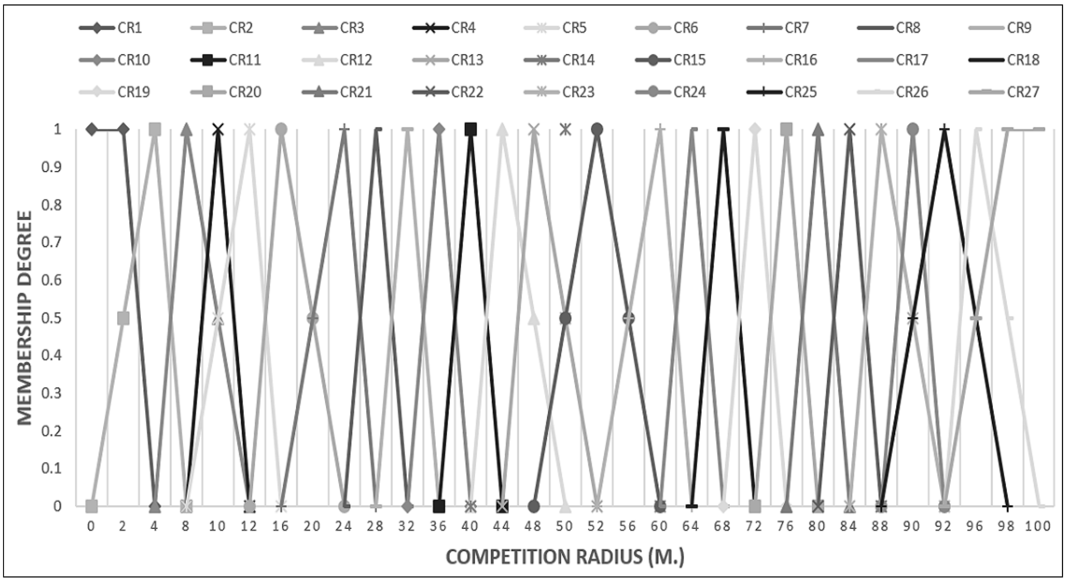

Figure 3.5: The fuzzified output descriptor Competition Radius.

When competition radius determination of provisional CHs finishes, the actual competition starts. Every provisional CH communicates CandidateCH Message to start the competition with remaining provisional CHs. CandidateCH Message is sent to the provisional CHs which situated in the maximum competition radius value obtained by using the optimization approach. The message consists of the identifier (*Id*), competition radius ($Comp_i$), remaining energy (E_i), and relative connectivity (C_i) of the source node. The first utilized and key parameter in CH competition is the remaining energy of sensor nodes. If a provisional CH gets a CandidateCH Message from another provisional CH that resides in its competition range and the amount of remaining energy in the source node is greater than the amount of remaining energy in the destination node, then the destination node leaves the competition and communicates a CeaseElection Message. If there happens a tie considering the remaining energies of the nodes, it is broken by the employment of the computed relative node connectivity values. If a specific provisional CH owns the highest remaining energy level among the provisional CHs which it gets CandidateCH Message from, or if it possesses the greatest relative connectivity inside its competition range among equal-energy provisional CHs, then it turns into a final CH.

The above-mentioned competition assures that there exists a single CH in the competition range of a specific CH and energy consumption is distributed evenly over the network. The aim of this competition is to diminish intra-cluster occupation and increase inter-cluster work, through assignment of narrow cluster sizes to the CHs that are closer to the sink since these CHs are exposed to higher inter-cluster relay (due to their position on the relay path) and CHs that relatively further have lower inter-cluster relay. As can be seen from the system model and the operation of the clustering, there are no unrealistic assumptions, like the use of infinite transmission range, in the proposed protocol which distinguish the procedure from some other proposals in the literature.

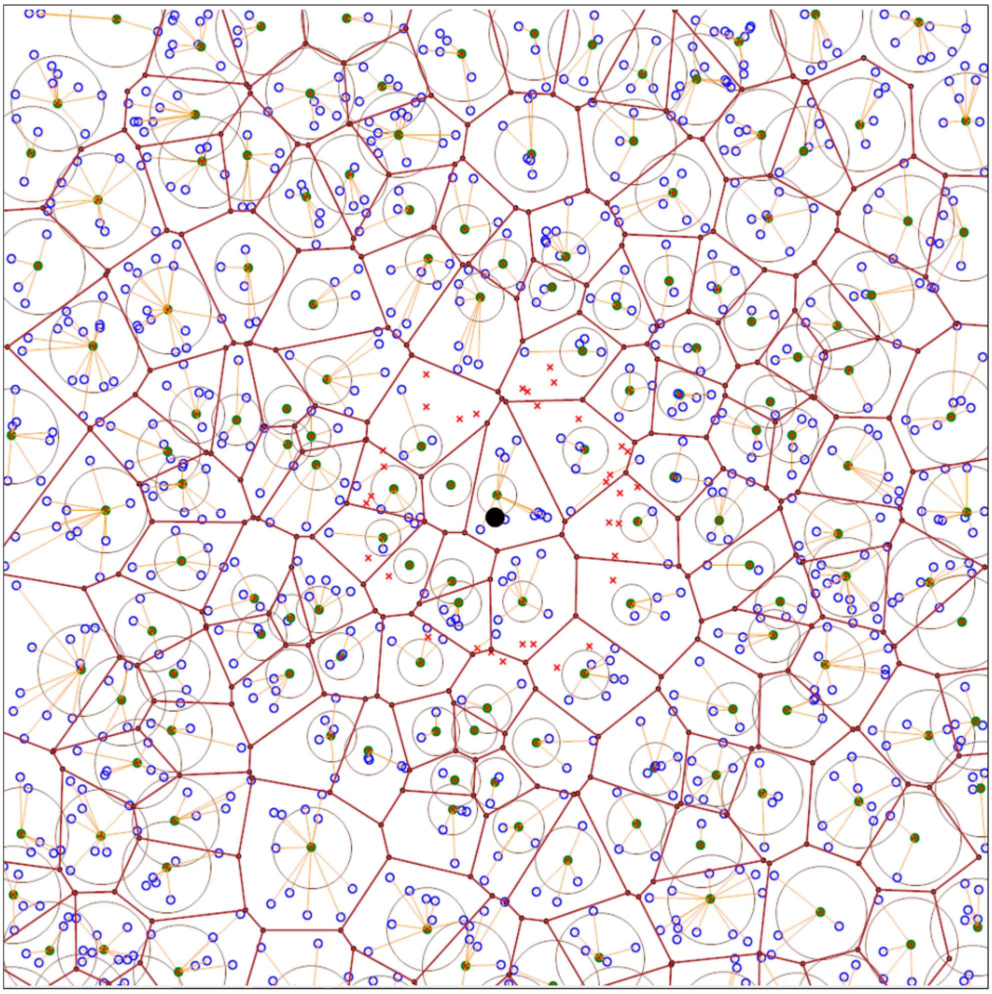

Figure 3.6: A Voronoi graph snapshot of a scene from a WSN clustered by TTDFP.

Fig. 3.6 depicts a representative scene from a WSN clustered by TTDFP. In the figure, 1000 nodes are deployed randomly to the 1000m x 1000m service area. The sink has the black color and a greater size in contrast to the other nodes, elected CHs have the red color, and the other remaining (member) nodes have the blue color. Service areas of generated clusters are depicted as green circles, transmission path from a member node to its cluster head is highlighted as an orange line, and dead nodes are marked with two crossed red lines (x). A Voronoi graph is also used here to differentiate between cluster ranges.

3.3.2 Optimization Approach using Simulated Annealing

In the literature, there are various manually value-assigned parameters employed in numerous proposed clustering/routing protocols. These parameters are assigned certain values after trial-and-error processes most of the time. In the study [67], several experiments are conducted to highlight the importance of tuning such parameters as they significantly affect network lifetime and the overall performance of the WSN. When there are multiple parameters of a given protocol with wide ranges of possible values, the search space and the computational complexity of finding the optimal blend of these values increase asymptotically. Additionally, when these values are assigned *manually*, it may cause a best-case scenario for the proposed protocol which generally biases the final comparison result in the experimental evaluations. Therefore here, an efficient framework is utilized to tune such parameters in order to optimize the performance metrics of WSNs, where performance metrics can vary from network lifetime to a number of packets successfully received by the sink [67].

In the aforementioned study, a framework that uses local search algorithms, simulated annealing (SA) and k-beams, is presented along with a discrete-event system simulator, OMNET++ with Castalia framework, to tune the parameters of clustering/routing protocols presented in the literature. The local search algorithm explores the search space of the possible combinations of parameters' values then invokes the simulator to assess the performance of a WSN under a given combination of parameter values. The use of a simulator to measure the fitness of combinations of parameter values is needed given the absence of a closed-form equation that calculates different performance metrics of WSNs given a set of parameters' values.

In this subsection, inspiration is taken from the study [67] and then adapted to this study as follows; first, rather than using OMNET++ with Castalia framework [69], the experiments are conducted on the simulation environment which was presented and described in detail in [20]. Second, the optimization is chosen to be performed solely using the SA algorithm the details of which are available in [70].

Algorithm 2: TTDFP Optimization Framework

Input: $(MaxCompRad, Th)_{initial}$

Output: $(MaxCompRad, Th)_{optimal}$

Pass $(MaxCompRad, Th)_{initial}$ to SA
Pass Operating Conditions to DESS
Pass Performance Metric to DESS
FZA: Trigger Optimization Framework
    Simulated Annealing: Explore Search Space
        Trigger DESS to evaluate a new pair
        Return fitness score
    Return $(MaxCompRad, Th)_{optimal}$

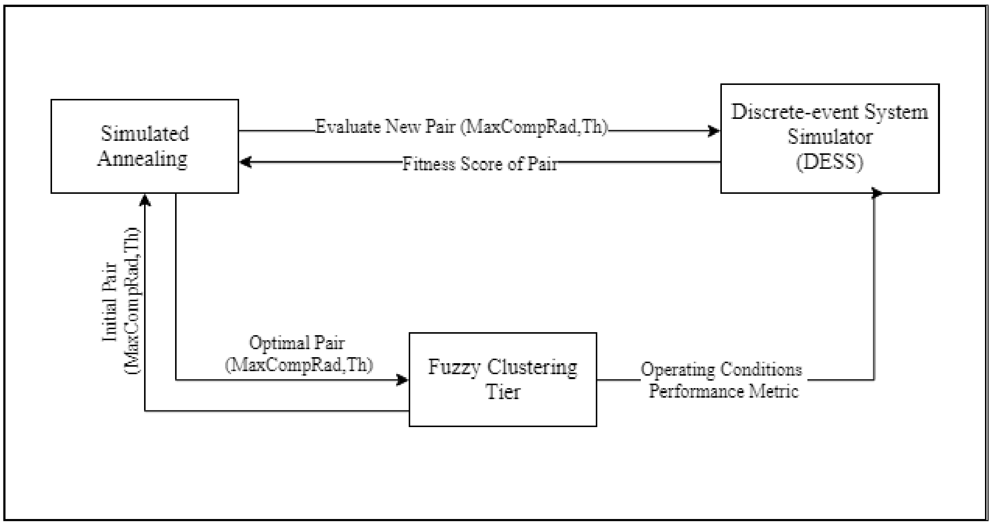

Figure 3.7: Optimization Framework Internals.

In the MOFCA [20] case, there are only two manually value-assigned parameters which are *maximum competition radius* (MaxCompRad) and *threshold* (Th). Thus, the modified version of the optimization framework in [67] can also be employed to tune this pair of parameters. Experimental results in the following sections highlight the effect of optimizing MOFCA when compared to its original version. For TTDFP, the same pair of parameters, *maximum competition radius* and *threshold*, is utilized in the fuzzy clustering tier in an optimized manner that differs from the previous studies as these inputs do not possess estimated fixed values. In addition, there are no manually value-assigned parameters in the second tier of the TTDFP protocol, thus the optimization framework is utilized only in the first tier. A general diagram showing the different blocks of the optimization framework and their corresponding interactions with the protocol itself is presented in Figure 3.7. Furthermore, algorithm 2 depicts the flow of the optimization framework as well as how the two units of the optimization framework interact and collaborate to achieve the predefined goal of parameter tuning.

It is important to emphasize that these parameters are not treated independently during optimization, the pursued approach optimizes them jointly targeting the maximum network lifetime, where this lifespan can be defined as the round when the First Node Dies (FND), Half of the Nodes Die (HND), or Last Node Dies (LND). For this reason, the selected parameters are optimized considering the depicted efficiency metrics at the start of the fuzzy clustering tier for both MOFCA [20] and TTDFP protocols using the optimization approach, the details of which are presented in [67]. The whole operation procedure of the methodology is not presented here in order not to degrade the readability of the chapter. However, the original study consists of a more detailed explanation of the operation.

This concludes fuzzy clustering phase and optimization essentials implemented in this tier. In the next subsection, the second tier which consists of the details of fuzzy routing is presented.

3.3.3 Tier II: Fuzzy Routing Phase

Two crucial elements are taken into account when devising the fuzzy routing tier:the first element is energy-efficiency in the tier, which is a must for the overall efficiency of TTDFP, and the other is lightweightness in the computational aspect. Like the previous tier, this tier also uses a distributed approach since the sink is not included in the Routing Route (*RR*) selection procedure. There is only a single point where the base station is included, which is the route set-up procedure initiated by the source node. There is no escape from this inclusion since the base station is the destination node. The rationale for devising this fuzzy routing tier is to decrease and balance the consumed energy over the whole network by a less number of transceiver unit activations through an efficient route selection. The fuzzy routing scheme extends the multi-hop routing approach defined in [18] with two fuzzified input parameters, which are *average link residual energy* and *relative distance*, one of which is already utilized in [71] in the selection of a routing path. However, in this study relative distance is employed instead of hop count in order to obtain better accuracy. The operation of the fuzzy routing phase is explained in Algorithm 3. $ALRE_r$, RD_r, Ch_r stand for average link residual energy, relative distance, and the chance value of a specific route r, respectively.

In search of a routing route, a source CH sends an RREQ packet addressed to the base station and starts collecting the RREP packets which are broadcast and addressed to itself by the sink. These packets arrive from different transceiver nodes and are control packets which do not consume as much energy as data packets. In normal operation, whenever a CH requests to send data to the base station, it sends an RREQ destined to the base station and waits for the RREP packets to arrive. According to the received RREP packets, it sets up a route considering the fuzzified input parameters and after generation of the chance value. When the RR is determined, data transmission over the path is initiated. In the routing route selection procedure, the source CH generates a value (*Min*), which is assigned as 1. This *Min* value only serves as a comparison value and is employed in the Routing Route election procedure. If there are no possible routes, the CH quits transmitting data. If the count of possible routes is greater than this *Min* value (line number 10), then fuzzy routing process (line 11-20) is initiated. Since the fuzzy routing tier employs *relative distance* together with the *average link residual energy* for computing the chance values of available paths, the chance value of each route changes dynamically as rounds pass by. This dynamism is the result of varying average link energies caused by energy depletion of nodes that are on the path. Computation of the chance values is done by the fuzzy inference engine which utilizes the fuzzy rule base to handle uncertainty.

The fuzzy operation starts at line number 12 in Algorithm 3 and *Candidate Routes*

Algorithm 3: TTDFP Tier-II: Fuzzy Routing Algorithm

Input: Route/s (RREP packet/s)

Output: The Routing Route (RR)

$Min \leftarrow 1$
$RR \leftarrow$ NULL
$Enum(Routes)$
if $(Count(Routes) < Min)$ **then**
    `/* No possible route (not connected) */`
    $RR \leftarrow NULL$
    EXIT
else if $((Count(Routes) = Min$ **then**
    `/* use the only possible route */`
    $RR \leftarrow OnlyRoute$
    EXIT
else
    foreach $route\ r \in Routes$ **do**
        By using fuzzified inputs, form Ch_r
        $CandidateRoute_r\ (Id_r, Ch_r, ALRE_r)$
    RR $\leftarrow CandidateRoute_1$
    for $i = 2\ to\ Count(Routes)$ **do**
        if $(Ch_i < Ch_{RR})$ **then**
            `/* RR does not change */`
        else if $((Ch_i = Ch_{RR})\ and\ (ALRE_i \leq ALRE_{RR}))$ **then**
            `/* RR does not change */`
        else
            RR $\leftarrow CandidateRoute_i$
    EXIT

are produced accordingly. The Fuzzy rules utilized in this tier are given in Table 3.2. Rules are evaluated by using the Mamdani Controller as an inference methodology and the Center of Gravity (COG) technique for the defuzzification of the output variable value.

Table 3.2: Fuzzy Rules in Tier-II: Fuzzy Routing Phase

Relative Distance	**Average Link Residual Energy**	**Chance Value**
Far	Low	Very Low
Far	Medium	Extra Low
Far	High	Moderately Low
Regular	Low	Low
Regular	Medium	Normal
Regular	High	High
Close	Low	Moderately High
Close	Medium	Extra High
Close	High	Very High

To be able to compute the route chance values, a total of two fuzzified input parameters is employed in this tier. The first fuzzified parameter is Average Link Residual Energy (*ALRE*) and Fig. 3.8 depicts the fuzzy set of this parameter. The chosen linguistic variables of this set are *low*, *medium* and *high*.

The second fuzzified parameter is the Relative Distance (length) of a route. The fuzzy set of this parameter is depicted in Fig. 3.9. *Close*, *regular* and *far* are the chosen linguistic variables of this fuzzy set.

Computations of the values of Relative Distance (RD) and Average Link Residual Energy (ALRE) of a specific route r are presented in Eq. 3.7 and Eq. 3.8, respectively. The usage of the relative and average values does not necessitate the inclusion of the sink in the decision process. As in the previous phase, this established aspect promotes the fuzzy routing phase to be a distributed promising route election methodology. $max(Distance)$ value in Eq. 3.7 can be deduced from the received

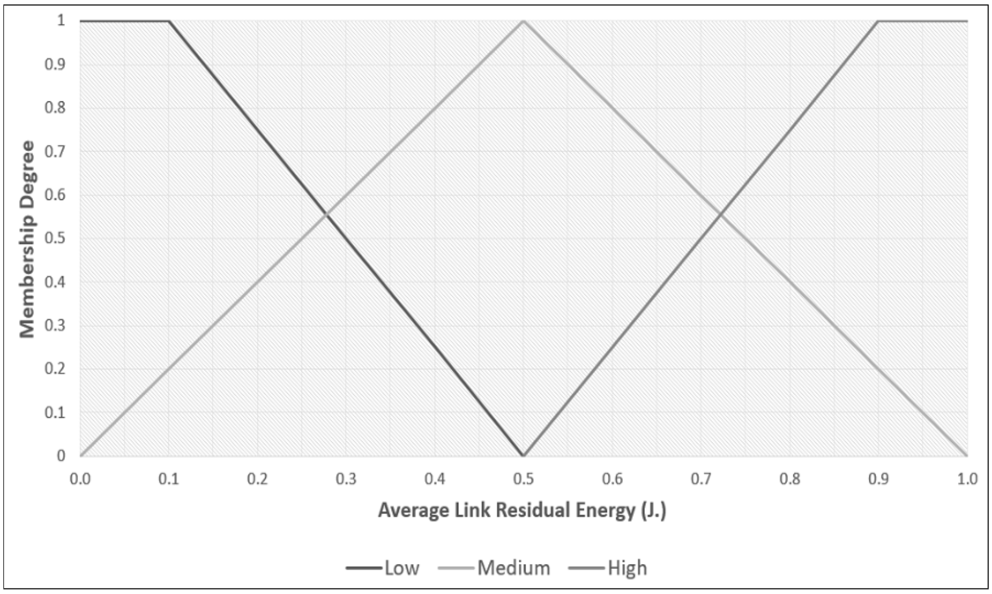

Figure 3.8: The fuzzy set of the first input parameter (*ALRE*).

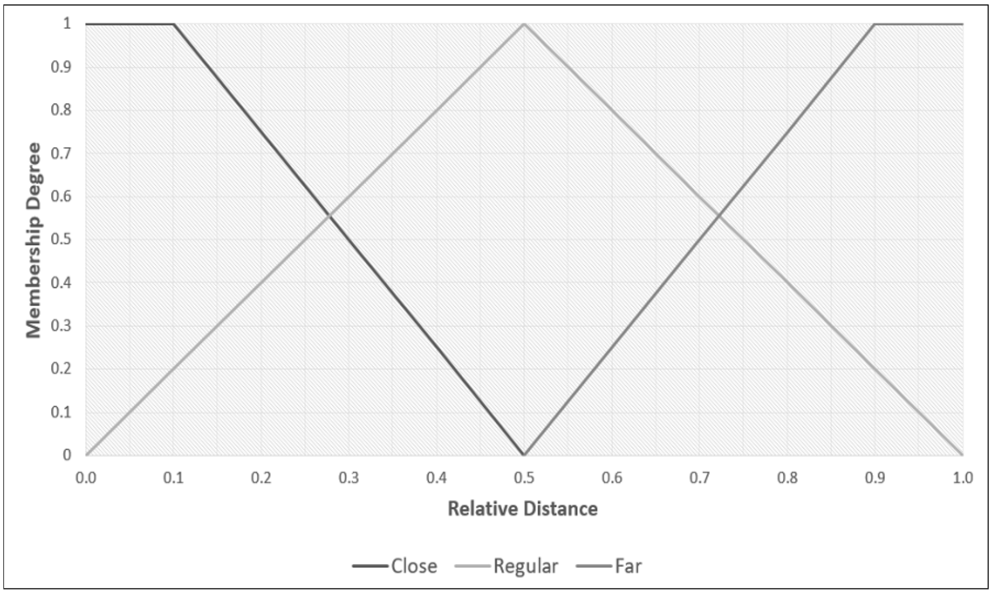

Figure 3.9: The fuzzy set of the second input parameter (*RD*).

RREP packets and utilized in the normalization procedure of varying link distances.

$$RD_r = \frac{Distance_r}{max(Distance)} \tag{3.7}$$

$$ALRE_r = \frac{\sum_{i=1}^{m} RE_i}{m} \tag{3.8}$$

The only fuzzified output descriptor is the chance value of a received route. The set utilized for this fuzzified output variable is delineated in Fig. 3.10. There are a total of nine chosen linguistic variables. There is no particular reason behind choosing this function other than its satisfactory results.

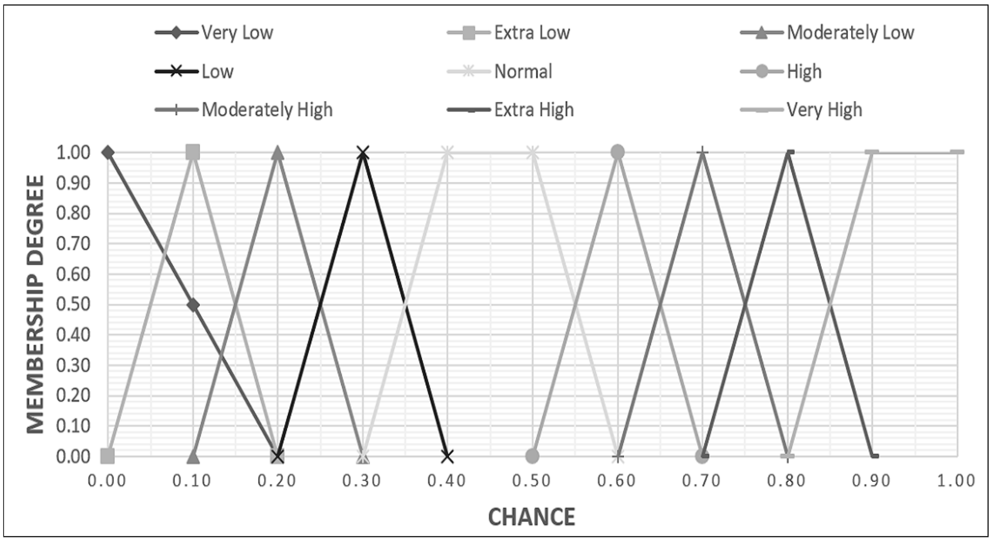

Figure 3.10: The fuzzy set of the output parameter (Chance).

The employed fuzzy rules in this tier are in the form of $IF\ a$ and $b\ THEN\ c$ as in the previous tier, which means that for example IF a specific route has HIGH *ALRE* and has CLOSE *RD*, THEN the *CHANCE* value of that route is VERY HIGH. It is also noteworthy to state that although the fuzzy approach extends its crisp counterpart with two additional fuzzy parameters, it is actually based upon the counterpart and consists of the basic principles of the underlying architecture inherently. For instance, if the crisp routing protocol does not generate some path i because of its congestion or resource values, the routing tier will not consider that path. This concludes the operation architecture of the fuzzy routing phase.

Tier-I on the basis and Tier-II on top of it make up the TTDFP architecture. Since energy consumption is directly proportional to the square of the transmitted distance, distance has a crucial impact over the network lifetime. Decreasing the transmission distance in both tiers helps in maintaining an energy-efficient operation. It is a fully-distributed, fuzzy logic-based and an optimized protocol which targets inefficient data aggregation scenarios of wireless sensor networks.

3.4 Experimental Evaluation

The obtained performance results of the TTDFP protocol are presented in this section. For the performance evaluation of TTDFP, it is compared against the selected distinctive protocols existing in the literature, which are LEACH, CHEF, EEUC, MOFCA-Original, and MOFCA-Optimized using two different scenarios which have two separate cases. These scenarios are depicted so as to quantify the contribution of each tier individually, which means that not only the performance of fuzzy clustering tier in TTDFP against selected clustering protocols but also the performance of fuzzy routing tier in TTDFP against its crisp counterpart are compared. MOFCA-Optimized is the tuned version of the original MOFCA protocol, where the threshold and maximum competition radius values are obtained from the optimization framework specifics of which are presented in detail in the previous section. In the evaluated scenarios, the location of the sink is stationary and can either be in the service area or out of it. Nodes are deployed randomly following the uniform distribution properties.

In Scenario 1, CHs of the LEACH protocol relay the aggregated data to the base station by using direct transmission and the aspired percentage of CHs for LEACH is 10%. The rest of the compared protocols, including TTDFP, utilize the same routing scheme, which is multi-hop routing as defined in [18], in both cases of Scenario 1. In Scenario 2, compared protocols either employ the routing scheme as defined in [18] in Case A or the proposed fuzzy routing scheme in Case B. CHEF α value is set to 2.5 as in [17] and the threshold optimal value is computed with the equations defined in that study and set as approximately 0.3 and 0.22 for 100 and 1000 nodes, respectively. Coefficient and threshold values of EEUC and MOFCA-Original are set to the same values as in the original studies [18] and [20], respectively. For MOFCA-Optimized and TTDFP, these values are obtained using the Simulated Annealing (SA) algorithm. In the discussion subsection, there is also a corroborative scenario which targets heterogeneous networks that consist of nodes with different initial energy levels. For this reason, the assumption that nodes possess equal initial energy is broken in the case of that specific scenario. Tested scenarios are sketched as follows:

- Scenario 1 targets the exploitation of the contribution only in fuzzy clustering tier. This scenario has two cases:

  - In Case A, the sink is out of the service area,
  - In Case B, the sink is in the service area.
- Scenario 2 targets the exploitation of the contribution only in fuzzy routing tier. This scenario has also two cases:
  - In Case A, the compared protocols employ the multi-hop routing scheme as defined in [18] (which is referred as the crisp counterpart in this study),
  - In Case B, the same protocols employ the proposed fuzzy routing scheme.

In the of depicted scenarios, clusters are reformed at every single round. Every member node sends 4000 bits data to the CH it belongs. Every CH compresses the received data with a ratio of 10% as employed in [17] prior to relaying towards the sink. Elaborations related to the calculation of this ratio and the final length of data of a CH after compression resides in [20].

Various experiments are done by using a WSN simulator for the purpose of testing the efficiency of TTDFP. The simulator is a discrete event simulator and it is able to simulate the compared algorithms under the same conditions. The details of the simulator can be found in [20]. However, MATLAB or Castalia [69] platforms can also be used for the simulation of the scenarios above. Wireless sensor nodes are randomly located into a 1000m x 1000m area and the full battery power of a node is set as 1 Joule (J). In the experiments, the sensor network is deployed so as to form a rectangular area-of-interest (service area). This is due to its easiness and effectiveness in simulations as it does not have any best or worst case effect on any architecture. The reason behind changing the location of the base station in cases is to target different WSN application types such as environmental monitoring, path or trail surveillance, and etc. Experiments are carried out on a 2.70 GHz quad-core workstation with 32 GB DDR4 2133 Mhz RAM and 512 GB SSD Drive.

In the experiments, three metrics for evaluation of the energy-efficiency of the depicted protocols are utilized. These metrics are First Node Dies (FND), Half of the Nodes Die (HND), and Total Remaining Energy (TRE). *Node Dies (ND)* metrics, in general, depict estimated values for the round in which the event, like the death of the first node or the half of the nodes, is generated. As done in some studies in the literature, the approach also does not consider LND metric for comparisons since most sensor networks require a portion of nodes to be alive in order for the network to be in fully-alive operation. Bearing in mind this restriction, FND, HND, and TRE metrics are chosen with the aim of evaluating the energy-efficiency of the crosschecked protocols.

3.4.1 Scenario 1

In this scenario, the performance of *fuzzy clustering tier* is assessed independently. For this reason, LEACH employs direct transmission and all the remaining protocols, including TTDFP, employ the multi-hop routing scheme as defined in [18]. The sink is either out of the service area as in Case A or in the service area as in Case B, and the nodes are deployed randomly into the service area.

3.4.1.1 Case A

An illustrative capture from the experimental environment in this case is depicted in Fig. 3.11 and the configuration of this case in Scenario 1 is delineated in Table 3.3. The maximum competition radius for the EEUC and MOFCA-Original protocols are set as 55m and 70m, respectively. The threshold (Th) and maximum competition radius ($Comp$) values for MOFCA-Optimized and TTDFP protocols are optimized using the Simulated Annealing (SA) approach and obtained as 0.15 (Th) and 75m ($Comp$) for MOFCA-Optimized, and 0.2 (Th) and 80m ($Comp$) for TTDFP. In addition, TRE is measured at the round 20 in this case.

The experimental results for Case A in Scenario 1 are presented in Table 3.4.

In this case, the proposed TTDFP approach outperforms all compared algorithms considering the utilized metrics. The performances of the compared protocols except TTDFP are the same according to the results of the FND metric. However, the performances of the protocols differ if the HND metric is considered. The number of alive nodes in connection with the number of rounds in Case A in Scenario 1 is

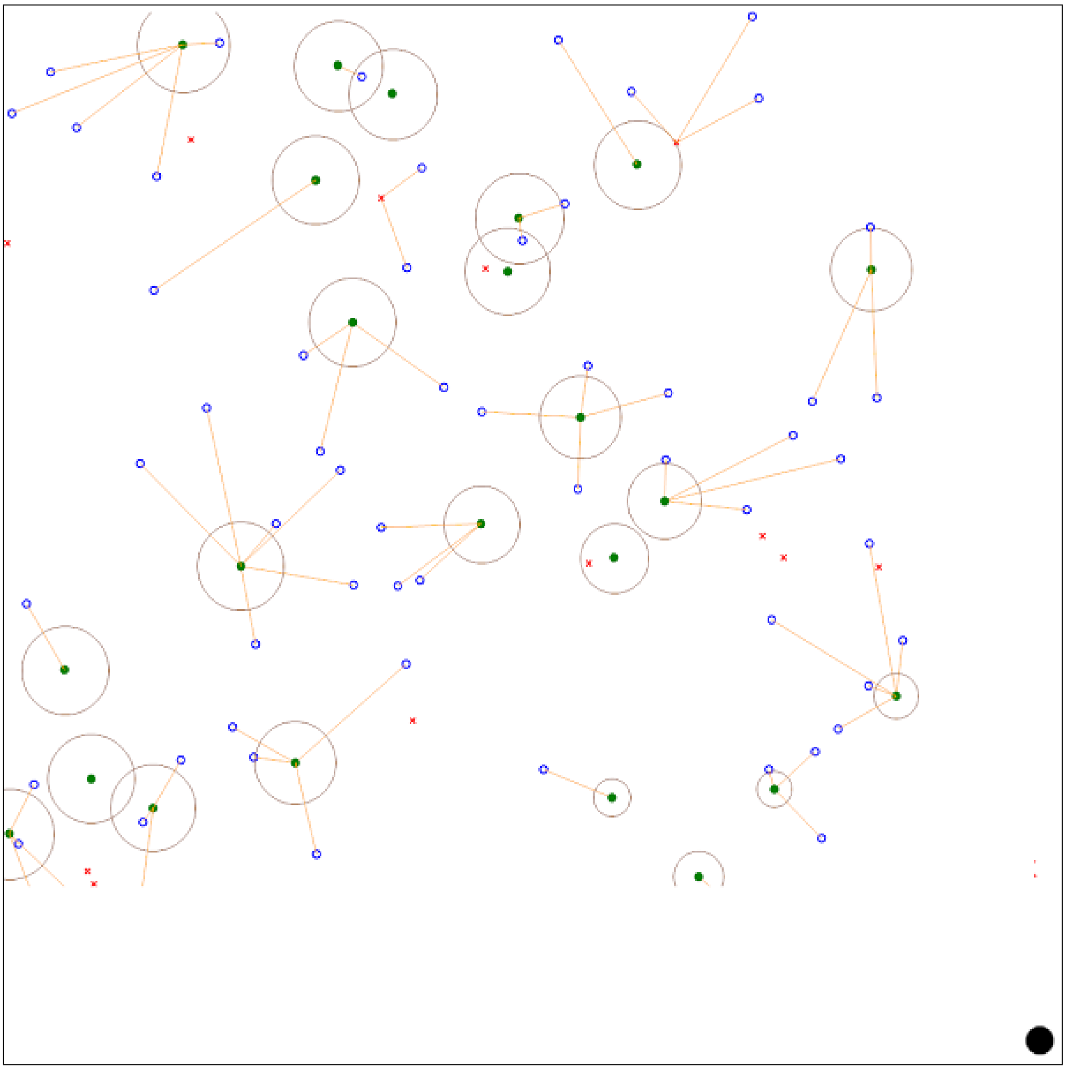

Figure 3.11: A scene captured from the WSN in Case A Scenario 1.

Table 3.3: The Configuration of Case A Scenario 1

Parameter	**Value**
Service Area	1000m x 1000m
Location of the sink	(1250,1250)
Number of deployed nodes	100
Data packet size	4000 bits
ε_{mp}	0.0010pJ/bit/m^4
E_{elec}	50nJ/bit
Aggregation ratio	10%

depicted in Fig. 3.12. As can be seen from the figure, the starting point for the death of the deployed nodes in TTDFP occurs at the round 3, which means that it has a longer fully-alive operation time than all other protocols. Although the performances of all other protocols initially look similar, EEUC performs better than both LEACH and CHEF when considering the HND metric, and MOFCA-Original outperforms these three protocols. Also, the performance of MOFCA-Original is close to MOFCA-Optimized, and they pursue a more or less similar energy consumption model. Moreover, the tuned version MOFCA-Optimized depletes remaining energy slower than its original version, which means that the optimization strategy followed generates a better combination of threshold and competition radius values than when these values are assigned manually. TTDFP preserves its efficient operating model for the HND metric as it performs better than the others, as in the case for the FND metric. In Case A, the TTDFP clustering phase performs 96.8% better than LEACH, nearly 45.8% better than CHEF and EEUC, 10.4% better than MOFCA-Original, and 6.2% better than MOFCA-Optimized when considering the TRE metric, which means that the energy depletion occurring in the TTDFP clustering phase balances energy consumption over the network better as well.

Table 3.4: Obtained Results of Case A Scenario 1

Algorithm	**FND**	**HND**	**TRE**
LEACH	2	14	1.24
CHEF	2	17	21.62
EEUC	2	18	21.64
MOFCA	2	25	35.74
MOFCA-Optimized	2	28	37.43
TTDFP	3	32	39.91

3.4.1.2 Case B

An illustrative capture from the experimental environment in this case is depicted in Fig. 3.13. The reason behind choosing this case for evaluation is to test the impact of the location of sink over the compared clustering protocols. The configuration of this case is the same as the previous case except for the location of the base station,

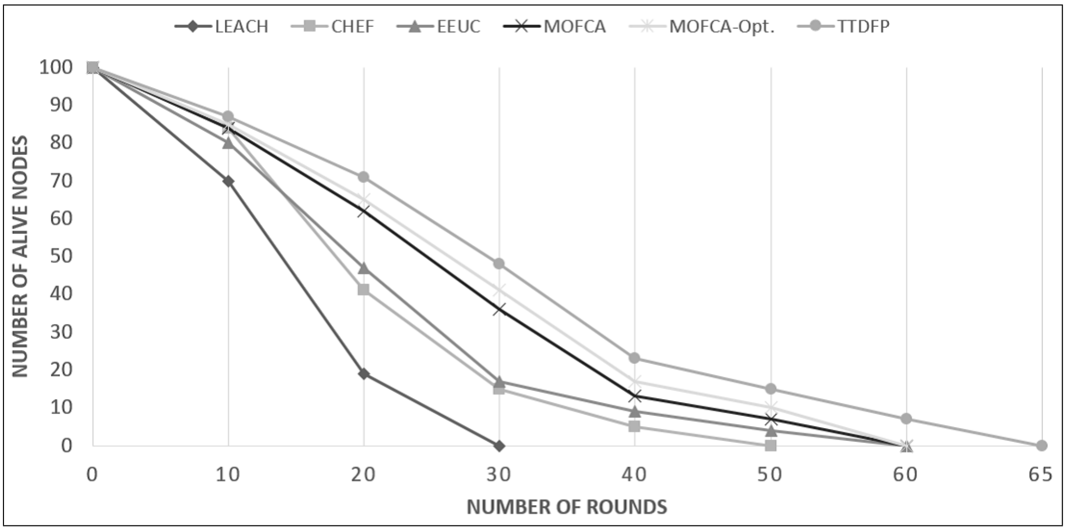

Figure 3.12: Number of alive nodes in connection with the number of rounds in Case A Scenario 1.

which is at (500,500), and the number of deployed nodes, which is 1000. Maximum competition radius values for EEUC and MOFCA are set as 70m and 80m, respectively. Threshold and maximum competition radius values for MOFCA-Optimized and TTDFP are optimized using SA and the values obtained are 0,3 Th and 65m $Comp$ for MOFCA-Optimized, and 0,23 Th and 75m $Comp$ for TTDFP. As can be seen from Table 3.5, clustering tier of the proposed TTDFP approach outperforms all of the compared approaches when considering all three metrics as in the previous case of this scenario. In this case, TRE is measured at the round 100.

Table 3.5: Obtained Results of Case B Scenario 1

Algorithm	FND	HND	TRE
LEACH	13	76	151.76
CHEF	9	135	380.25
EEUC	11	147	469.74
MOFCA-Original	14	152	486.14
MOFCA-Optimized	15	164	501.47
TTDFP	17	176	561.77

The number of alive nodes in connection with the number of rounds in this case is depicted in Fig. 3.14. According to the obtained results, in this case, the perfor-

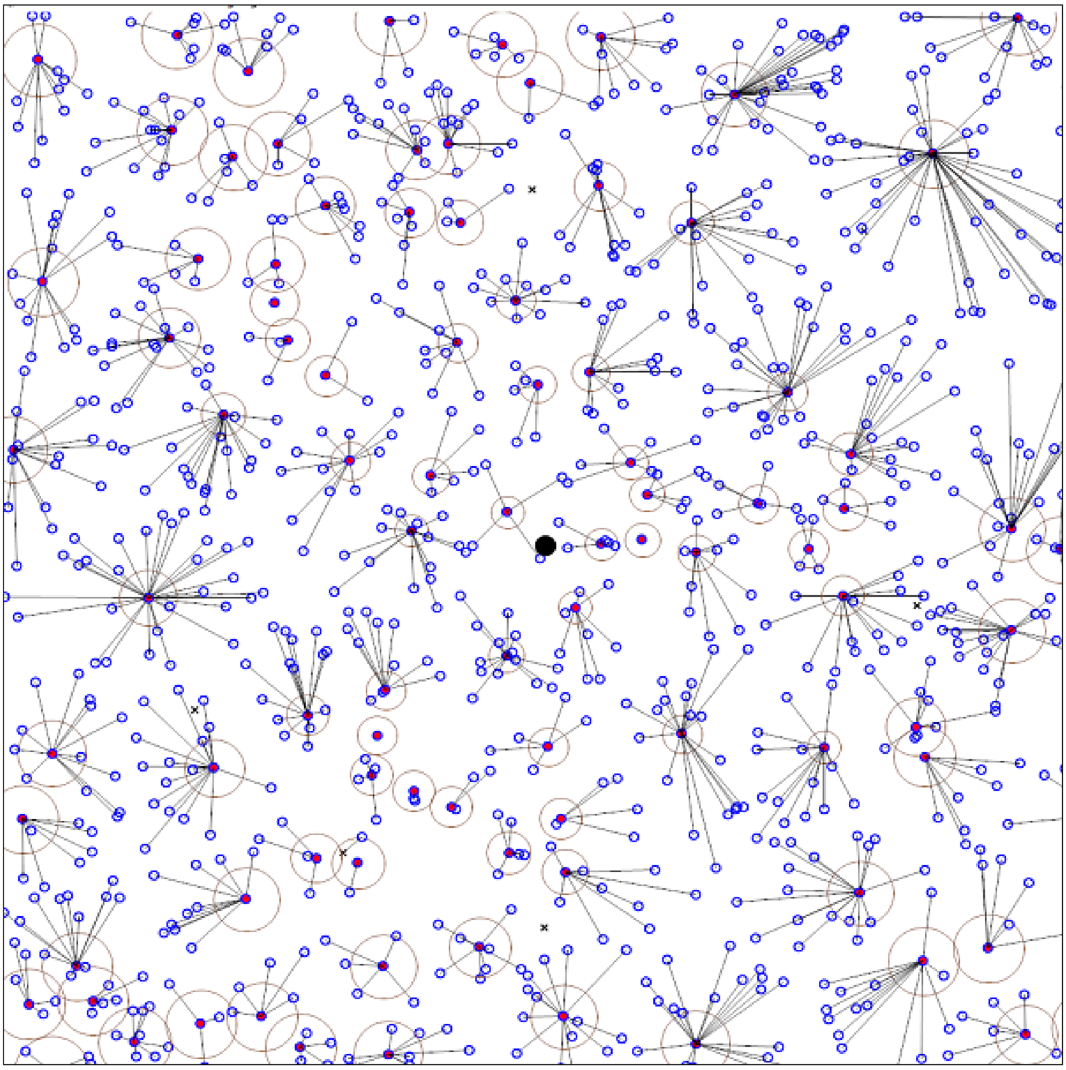

Figure 3.13: A scene captured from the WSN in Case B Scenario 1.

mance of TTDFP is nearly 23.5% more efficient than LEACH, 47% more efficient than CHEF, 35.2% more efficient than EEUC, 17.6% more efficient than MOFCA, and 11.7% more efficient than MOFCA-Optimized for the FND metric. CHEF performs the worst among all when considering the FND metric, but it performs better than LEACH for the HND metric. Although LEACH performs better than CHEF for the FND metric, efficiency is not preserved against CHEF as it depletes remaining energy of nodes faster when the death of the nodes commences, as can be seen from the HND and TRE metrics. Again, TTDFP performs the best among all protocols in this case with respect to all considered metrics. This situation corroborates that the proposed TTDFP protocol clustering tier maintains its efficient operation fashion under various operating conditions.

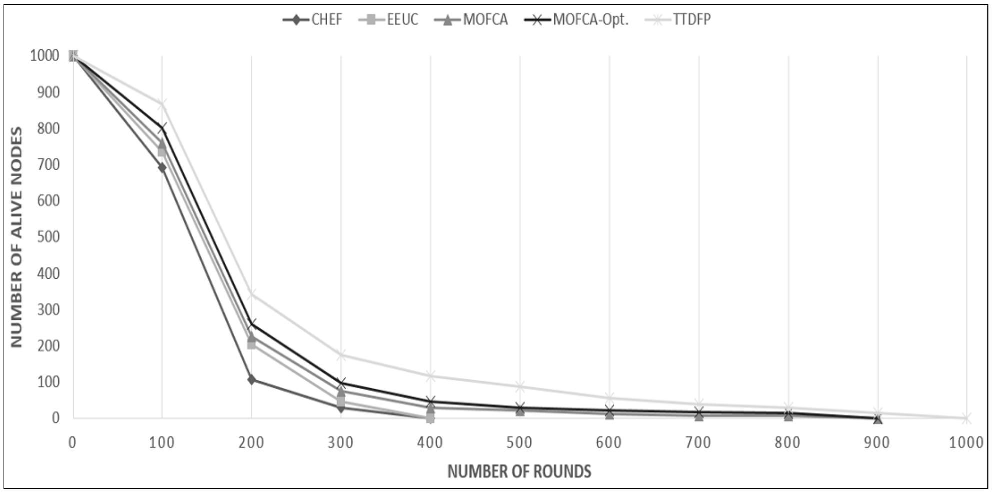

Figure 3.14: Number of alive nodes in connection with the number of rounds in Case B Scenario 1.

Scenario 1, with its two cases, independently assesses the performance of clustering tier, under which all protocols except LEACH follows the same routing scheme, it is valid to state that TTDFP clustering tier outperforms compared clustering protocols in its sole aspect.

3.4.2 Scenario 2

In this scenario, the performance of *fuzzy routing tier* is assessed independently. For this reason, LEACH is not included in the scenario since it is specifically designed for direct transmission. CHEF, EEUC, MOFCA with its two versions, and TTDFP employ the multi-hop routing scheme as defined in [18] in Case A and the proposed fuzzy routing scheme in Case B.

The configuration of this scenario is as presented in Table 3.6. Since the only difference is in the underlying routing scheme between cases, the configurations of Case A and Case B are not presented separately. The maximum competition radius for the EEUC and MOFCA algorithms in this scenario are set as 70m and 80m, respectively. The threshold and maximum competition radius values for MOFCA-Optimized and TTDFP are optimized using SA and obtained as 0.2 (Th) and 65m ($Comp$) for MOFCA-Optimized, and 0.2 (Th) and 80m ($Comp$) for TTDFP. TRE is measured at rounds 150 in both cases of this scenario.

Table 3.6: The Configuration of Scenario 2

Parameter	Value
Service Area	1000m x 1000m
Location of the sink	(500,500)
Number of deployed nodes	1000
Data packet size	4000 bits
ε_{mp}	0.0010pJ/bit/m^4
E_{elec}	50nJ/bit
Aggregation ratio	10%

First, the obtained performance results of compared approaches under different routing schemes in two different cases are presented and then the performance gain by comparing the results of the same protocol in different cases are delineated in order to highlight the performance of fuzzy routing.

3.4.2.1 Case A

The obtained experimental results for the compared protocols of Case A (crisp multi-hop routing) in Scenario 2 are depicted in Table 3.7.

Table 3.7: Obtained Results of Case A Scenario 2

Algorithm	FND	HND	TRE
CHEF	9	135	156.94
EEUC	11	147	216.65
MOFCA-Original	14	152	230.24
MOFCA-Optimized	15	164	240.36
TTDFP	17	176	254.03

3.4.2.2 Case B

The obtained experimental results for Case B Scenario 2 are depicted in Table 3.8.

Table 3.8: Obtained Results of Case B Scenario 2

Algorithm	FND	HND	TRE
CHEF	10	142	161.17
EEUC	12	155	223.49
MOFCA-Original	17	158	235.74
MOFCA-Optimized	17	169	244.81
TTDFP	18	181	262.92

To be able to effectively compare the performance of fuzzy routing over its crisp counterpart, each protocol can be investigated under different routing schemes as can be seen from Tables 3.7 and 3.8. There is little gain for all protocols if the routing scheme is updated to fuzzy routing when considering the FND metric. However, the gain due to fuzzy routing increases as the rounds pass by. This situation can be corroborated from the HND metric. CHEF performs 4.9%, EEUC performs 5.1%, MOFCA-Original performs 3.7%, MOFCA-Optimized performs 2.9%, and TTDFP performs 2.7% better under fuzzy routing conditions considering the HND metric. Since the TRE is measured at the same round under both routing schemes, it is obvious that energy is preserved more under fuzzy routing conditions than crisp routing conditions for all protocols. Moreover, although fuzzy routing brings an overhead in the packet structure and computational aspects, this overhead is tolerable and insignificant. The operation of fuzzy routing when compared to its crisp counterpart also ensures that energy is balanced throughout the network for all protocols.

3.4.3 Discussion

In order to grasp the essential properties of the proposed protocol and the theoretical advantages of the designed method over compared algorithms, it is beneficial to conceive common and discriminating features of evaluated algorithms. Therefore, common and discriminating features of the compared protocols are depicted in Table 3.9. As can be grasped from the table, TTDFP is the *only* protocol which includes fuzzy clustering, improved by the optimized parameters, jointly with fuzzy routing. Also, it is the only protocol that evaluates multiple routes in fuzzy fashion prior to

Table 3.9: Common and Discriminating Features of Evaluated Algorithms

Clustering Approach	**Fuzzy Computation**	**Transmission Scheme**	**Location Consideration**	**Size of Clusters**	**Optimized Parameters**	**Efficient Routing**
LEACH [14]	×	Direct	×	Equal	×	×
CHEF [17]	✓	EEUC	×	Equal	×	×
EEUC [18]	✓	EEUC	✓	Unequal	×	×
MOFCA [20]	✓	EEUC	✓	Unequal	×	×
MOFCA-Opt	✓	EEUC	✓	Unequal	✓	×
TTDFP	✓	Fuzzy	✓	Unequal	✓	✓

relaying data to the sink.

In fact, the design of TTDFP makes it perform better than the compared algorithms due to the following reasons: In the first tier, optimized parameters provide a more efficient run-time for the fuzzy clustering algorithm than that of compared protocols. Moreover, fuzzy routing tier balances energy consumption over relay paths more efficient than probabilistic crisp multi-path routing approach due to its fuzzy computation. Finally, the overhead of fuzzy computation is insignificant which can be corroborated by the obtained experimental results.

According to the efficiency comparisons done, TTDFP outperforms the compared protocols. However, not only efficiency but also scalability and computational complexity of a protocol are the crucial features for it to be utilized in large-scale networks, which is very important for today's most WSNs. For this reason, conducted experiments are run over the same WSN under varying number of nodes in the presented scenarios. As for the scalability analysis of the protocol, the test cases with 1000 nodes are vital, since they consist of a large number of nodes when compared to the remaining cases. TTDFP is significantly more energy-efficient in each of its tiers than all of the compared protocols as demonstrated by the results of the conducted experiments. In addition, as the number of nodes increases, it starts to deviate from the compared approaches considerably, which corroborates the scalability aspect of the proposal.

As for the computational complexity of the protocol, the clustering tier in TTDFP is O(n^2) because (n^2 - n) number of comparisons are done in the worst case to elect CHs as in [20]. Additionally, since the optimization approach is triggered only once in the base station prior to the deployment phase, where there is generally no restriction on computational resources in real-life situations, the computational complexity of the optimization part can be ignored. For the fuzzy routing phase, each elected CH makes at most (n^2 - n) number of comparisons among possible routes, which is also O(n^2). For this reason, when these two-tiers are considered jointly, the overall computational complexity of TTDFP is O(n^2), which makes it a prominent approach.

It is valid to state that each tier in TTDFP is theoretically feasible in its sole aspect. There is no repetition condition (loop) which may prevent convergence of the first tier since the clustering algorithm is executed once for each round. Although there is a single loop in the SA approach, it is controlled by the maximum number of steps and improvement threshold values which terminate the loop when the efficiency saturates (stabilization over time takes place). Fuzzy computation in the routing tier occurs only in the route determination process and does not consist of a loop as in the first tier. For these reasons, it is valid to state that TTDFP is a theoretically feasible approach.

Although the performance of each tier in the proposed protocol is assessed independently, it is done to refrain from bias and to highlight the gain resulted in each tier. However, it can be more clarifying to study how these protocols run as they are (with their original clustering and routing schemes). The performance gain resulted from the proposed protocol with respect to the compared counterparts is depicted in Table 3.10. In this experiment, the sink is located at (0,500), a total number of 200 nodes are deployed, and TRE is measured at the round 50.

Table 3.10: The Overall Performance Comparison of TTDFP

Algorithm	**FND**	**HND**	**TRE**
LEACH	4	34	12.67
CHEF	6	42	32.23
EEUC	2	27	12.73
MOFCA-Original	4	43	35.42
MOFCA-Optimized	4	47	36.32
TTDFP	6	59	44.86

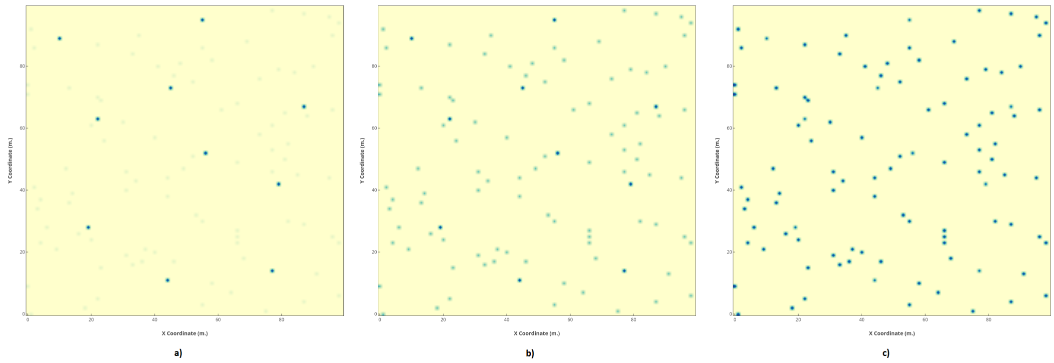

Figure 3.15: Heat maps for the operation architecture of TTDFP in heterogeneous networks.

It is also noteworthy to consider the stated assumptions. The mentioned assumptions may be restrictive for heterogeneous networks, which have different types of nodes when considering their battery sizes. The operation of the algorithm can be generalized for the non-restricted case as follows: "At the initial stages (rounds) of the operation, if there are nodes at different energy levels, these high-energy nodes are selected as final Cluster Heads most of the time until their energy deplete and become more or less the same as other (low-energy) nodes as the rounds pass by. If this is the case, such as in heterogeneous wireless networks, the real final CH competition commences at these energy-balanced rounds." In order to clarify this operation architecture and demonstrate the applicability of proposal in heterogeneous wireless networks, here an extra scenario is depicted. Generated heat maps for the operation architecture of TTDFP in this example scenario are provided Fig. 3.15.

In this heterogeneous network scenario nodes have varying initial energy levels such that normal nodes (90% of the total number) possess 1 joule of initial energy, whereas remaining nodes, which may be called super-nodes, (10% of the total number) possess 10 joules of initial energy. A total of 100 nodes are deployed to the 100x100 m. area in this experiment and the compared protocols run as they are (with their original clustering and routing schemes). The plot in part *a* (the left-most) depicts the heat-map of the network at the initial deployment time, the plot in part *b* (the middle) depicts the heat-map of the network at the round 500, and the plot in part *c* (the right-most) depicts the heat-map of the same network at some round *r*, respectively. All heat-maps are interpolated and smoothed with respect to the maximum energy of the deployed nodes. Normal nodes in the (a) plot are barely visible since the initial energy of each (1 j) is significantly less than the initial energy of super-nodes (10 j). However, they become more visible (clearer) as the rounds pass by. This visibility is not the result of an increase in their energy rather the result of a significant decrease in the energy levels of super-nodes since this maximum energy of a node in the network is utilized in the smoothing operation. As can be seen from the evolving nature of the heat maps, it is verified that super-nodes in TTDFP are elected as CHs initially more when compared to normal nodes until some round r. When the energy-levels of super-nodes and normal nodes start to balance, normal nodes are also elected as final CHs as much as super-nodes.

The decrease in TRE of each protocol under heterogeneous network scenario is presented in Fig. 3.16. Under this heterogeneous environment, the proposal maintains its efficient operation manner and consumes TRE much slower than the compared algorithms. The performances of the compared protocols except LEACH are more or less the same when compared to the homogeneous network environment. This is the result of their energy-based operation architectures. However, periodical rotation of nodes without considering their energy levels in LEACH makes the network consume TRE faster and more vulnerable to death of the nodes.

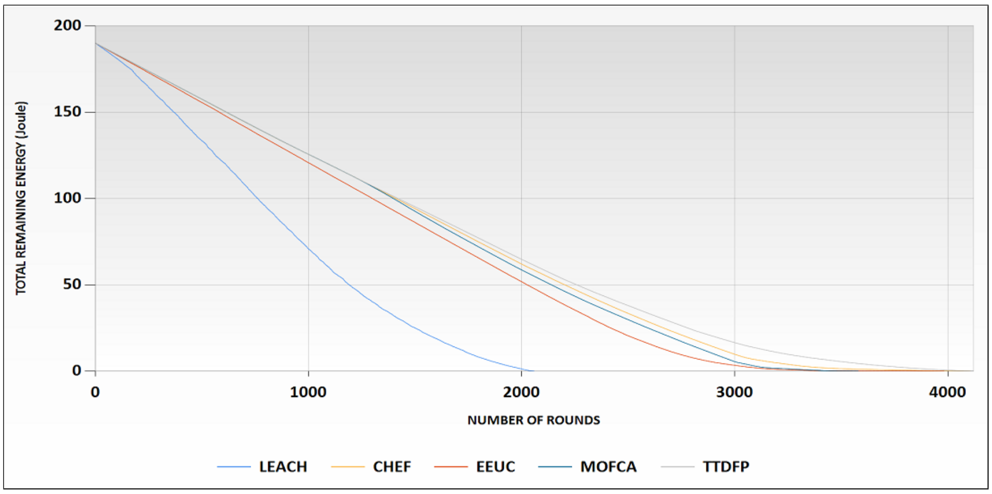

Figure 3.16: The decrease in TRE of each protocol under heterogeneous networks.

In summary, the contributions in this chapter by devising TTDFP protocol are two folds: First, the fuzzy clustering phenomena is improved. In Tier I of the proposal, a more energy-efficient operation is pursued. Moreover, in Tier II, there is a novelty such that a crisp routing protocol is extended with two fuzzy descriptors. This is done to overcome insufficiency resulted from crisp routing and to balance energy consumption over relay paths.

3.5 Remarks

In this chapter, a TTDFP (Two-Tier Distributed Fuzzy Protocol) which is an energy-efficient protocol for data aggregation in multi-hop WSNs is proposed. The proposal considers *relative node connectivity*, *distance to the base station*, and *remaining node energy* parameters in the election of CHs; and *average link remaining energy* together with *relative distance* parameters in the selection of *RR* while employing fuzzy logic in order to handle the occurring inexactness in these phases.

It is conceivable from the obtained results that TTDFP is a fully-distributed, fuzzy logic based and optimized protocol which targets inefficient data aggregation problems in WSNs. According to the results of the tested scenarios, the performance of each tier in TTDFP is much better than the compared protocols. Considering the experimental results done throughout the study, TTDFP preserves its scalability as the number of nodes increase. Furthermore, the TTDFP protocol has no manually-value-assigned parameters as it uses an optimization framework to avoid the inefficient trial-and-error approach as well as any bias resulting from the best-case behavior.

For future work, optimization of the other algorithm-dependent parameters using some other methodologies such as Particle Swarm Optimization (PSO) can be considered. Additionally, exploring the behavior and testing the performance of TTDFP under WSNs coupled with various node movement strategies can also be investigated.

CHAPTER 4

CLONALG-M TO IMPROVE ENERGY-EFFICIENCY OF RULE-BASED FUZZY CLUSTERING ALGORITHMS

In clustered networks, clustering can be done either following a crisp approach or a fuzzy one. Fuzzy clustering methodologies are found to be superior to crisp clustering counterparts when the boundaries among clusters are uncertain. As a result of this, a significant number of studies have proposed fuzzy-based solutions to the clustering problem. Most rule-based fuzzy systems determine and tune the fuzzy rules and the shapes of the output membership functions by employing some field experts in trial and error processes; thus, a considerable amount of time is dedicated to obtain and tune these functions, and it is almost impossible or impractical to contrive a fuzzy system that possesses the optimality property.

In this chapter, a Modified Clonal Selection algorithm (CLONALG-M) is proposed with the purpose of increasing the energy-efficiency of rule-based fuzzy clustering algorithms. Although there exist some studies in the literature focusing on fuzzy optimization in general as stated in the literature survey chapter of this dissertation, to the best of knowledge, none takes the performance improvement of rule-based fuzzy clustering algorithms into consideration. The CLONALG-M algorithm gets inspiration from Clonal Selection Principle, which is employed to elucidate the basic principles of an adaptive immune system. In the study, this principle is applied to identify the approximate placement of output membership functions which boosts the performance of rule-based fuzzy clustering algorithms, whose rule base and shape of membership functions are known a priori. Experimental analysis and evaluations of the proposed approach are done over fuzzy rule-based clustering methodologies and the results reveal that proposed approach performs and scales well for fuzzy functions.

The operation of CLONALG (a representative of CLONal Selection ALGorithms) is modified and the application of the modified version (CLONALG-M) is proposed to extend the lifespan of WSNs through determining the approximate placement of output membership functions in rule-based fuzzy clustering algorithms whose rule bases and shapes of membership functions are known a priori. CLONALG-M takes the initially defined fuzzy output function and approximates it based on the principles of CLONALG-M. To the best of knowledge, there is no existing study that utilizes the original or any modified version of CLONALG on output function approximation in the fuzzy clustering context. In order to evaluate the performance of the proposal, it is applied to the selected fuzzy clustering mechanisms such as CHEF and MOFCA and also compared with a standard Genetic Algorithm (GA) implementation. Various experiments are conducted under defined scenarios. The obtained results reveal that CLONALG-M is an encouraging approximation approach to be utilized for fuzzy output functions in clustering applications of WSNs.

4.1 Overview

An ever-increasing portion of clustering methodologies utilizes fuzzy logic in order to cope with uncertainty problems in WSNs. Therefore, they are categorized as fuzzy methodologies in general. In most available approaches, fuzzy sets and rules are exploited in order to obtain a preferable blend of applicable input parameters that can generate an optimal output. The parameters that define the shape of fuzzy output functions in rule-based clustering algorithms are usually obtained from a field expert or generated automatically. Therefore, it consumes a considerable amount of time to determine the shape of these functions together with the tuning operations, and it is often impractical to design the optimal fuzzy system in detail.

Designing a fuzzy logic-based system covers the definition of fuzzy sets that are generally depicted by membership functions and rules. As soon as the fuzzy system is designed, one of the foremost problems to handle is to establish the optimal placement of these described membership functions. In general, most membership functions in fuzzy control systems are presumed to be triangular or trapezoidal in shape and linear in a computational manner. For this reason, the actual problem is to decide on the value of the variables that characterize the shapes of these functions. If a suitable representation can be chosen, then the membership function approximation is reduced to a discrete optimization problem that could be modeled using parameters in general.

In the upcoming sections of this chapter, the proposed methodology is presented in detail and extensively experimented.

4.2 Methodology

Before detailing the proposed methodology, the basic principles of CLONALG are discussed initially.

4.2.1 CLONALG

Castro and Zuben proposed CLONALG to integrate main processes contained in clonal selection principle. CLONALG is primarily proposed for pattern recognition and machine-learning tasks, and afterward transcribed to be implemented to optimization domain. The proposed steps in the original CLONALG algorithm for an optimization task is given below [50]:

1. "Generate j antibodies randomly."
2. "Repeat until a stopping condition:"
   (a) "Calculate the affinities of antibodies. These affinities comply with the evaluation of a fitness function."
   (b) "Choose the n highest affinity antibodies."
   (c) "The chosen n antibodies is cloned to their affinities proportionally, forming a repertory C of clones: The greater the affinity, the greater number of clones formed and vice-versa."
   (d) "The clones from C are subject to hypermutation process inversely proportional to their antigenic affinity. The greater the affinity, the less mutation, and vice-versa."

(e) "Calculate the affinities of the C mutated clones."

(f) "Using this set C of clones and antibodies, choose the j greatest affinity clones to constitute the renewed antibodies population."

(g) "Substitute the d lowest affinity antibodies by newly formed individuals randomly."

3. "End repeat."

According to the CS theory, it is suggested commencing with an initial repertory j of immune cells. Then, the immune system can transform itself in response to various experiences with the surroundings. Through selection and cumulative variation on generations of large-scale cells, the immune system is able to obtain the required information to defend the host structure against pathogenic threats of the surroundings.

Although the utilization of CLONALG for optimization problems is common, the utilization of the same approach in the fuzzy context is rare. The feasibility of finding the most convenient designation of membership functions for a Multiple-Input Multiple-Output system using CLONALG for unconstrained optimization problems is presented in [72]. In the study, authors depict how the fuzzy membership optimization problem is converted to the parameter optimization problem and utilize CLONALG for the problem at hand. However, the CLONALG algorithm is modified to be applied for the constrained approximation problems in the following subsection, which is mostly the case for the output membership functions of rule-based fuzzy clustering algorithms in WSNs.

4.2.2 CLONALG-Modified (CLONALG-M)

Applied modifications to CLONALG which make up CLONALG-M are as follows:

- Generate j antibodies randomly such that $j \geq NumberOfRules$ (in Step 1),
- Each antibody (possible solution) should perform a feasibility check prior to evaluating the objective function in order to determine whether it is a valid fuzzy output function (in Step 2.a),

- Mutation operates in a constrained fashion (in Step 2.d),
- Mutated clones should perform feasibility checks prior to determination of affinity whether they still preserve fuzzy function validity measures (in Step 2.e).

The operation of CLONALG-M on a fuzzy output function with pseudo-code is presented in Algorithm 4. CLONALG-M takes initially an expert defined fuzzy output function together with $Size_{pop}$, $Size_{sel}$, $Size_{prob}$, $\#_{randcells}$, $Rate_{clone}$, and $Rate_{mutation}$ parameters and returns the approximated fuzzy output function. These parameters denote *Population Size*, *Selection Size*, *Problem Size*, *Number of Random Cells*, *Clone Rate*, and *Mutation Rate*, respectively.

First, the $Population$ is assigned the initially created cells and then starts iterating over the search space while stopping condition, which is here depicted as 1% improvement, is not reached. Since CLONALG-M checks the feasibility of each P_i prior to affinity calculation and regenerates or remutates if necessary, every generated P_i in CLONALG-M satisfy the fuzzy function validity measures. It is wise to check feasibility while these P_i are changing. If this check is not performed, the generated P_i may not satisfy the fuzzy validity measures. These feasibility checks in CLONALG-M are depicted in lines 4 & 16 of Algorithm 4. Additionally, mutation operates in a controlled fashion such that mutated clones are unique in CLONALG-M.

The following subsection elaborates into the details of CLONALG-M operation on a generic fuzzy output function in order to grasp the computational aspects and essential features of the proposed approach.

4.2.3 Example Operation of CLONALG-M on a Generic Fuzzy Function

The output membership function of a generic rule-based fuzzy system on which computational aspects of CLONALG-M are explained on consists of four rules as depicted in Fig. 4.1. As can be seen in the figure, there are both trapezoidal and triangular membership functions.

Triangular functions are exemplified by three points (a, b, c) which denote the start-

Algorithm 4: CLONALG-M Algorithm

Input: An Initial Expert-Defined Fuzzy Output Function, Params

Output: *Population*: An Approximated Fuzzy Output Function

$Population \leftarrow$ CreateInitialCells($Size_{pop}$,$Size_{prob}$)
while $\neg$ *StopCondition()* **do**
    for $P_i \in Population$ **do**
        CheckFeasibility(P_i);
        if $P_i \neq Feasible$ **then**
            repeat
                Regenerate(P_i);
            until $P_i = Feasible$;
        Affinity(P_i);
    $Sel_{pop} \leftarrow$ Select($Population$,$Size_{sel}$)
    $Clone_{pop} \leftarrow \emptyset$
    for $P_i \in Sel_{pop}$ **do**
        $Clone_{pop} \leftarrow$ Clone(P_i,$Rate_{clone}$)
    for $P_i \in Clone_{pop}$ **do**
        ControlledMutation(P_i,$Rate_{mutation}$)
        CheckFeasibility(P_i);
        if $P_i \neq Feasible$ **then**
            repeat
                Remutate(P_i);
            until $P_i = Feasible$;
        Affinity(P_i);
    $Population \leftarrow$ Select($Population$,$Clone_{pop}$,$Size_{pop}$)
    $Rand_{pop} \leftarrow$ CreateRandCells($NumberOf_{randcells}$)
    Replace($Population$,$Rand_{pop}$)
Return($Population$)

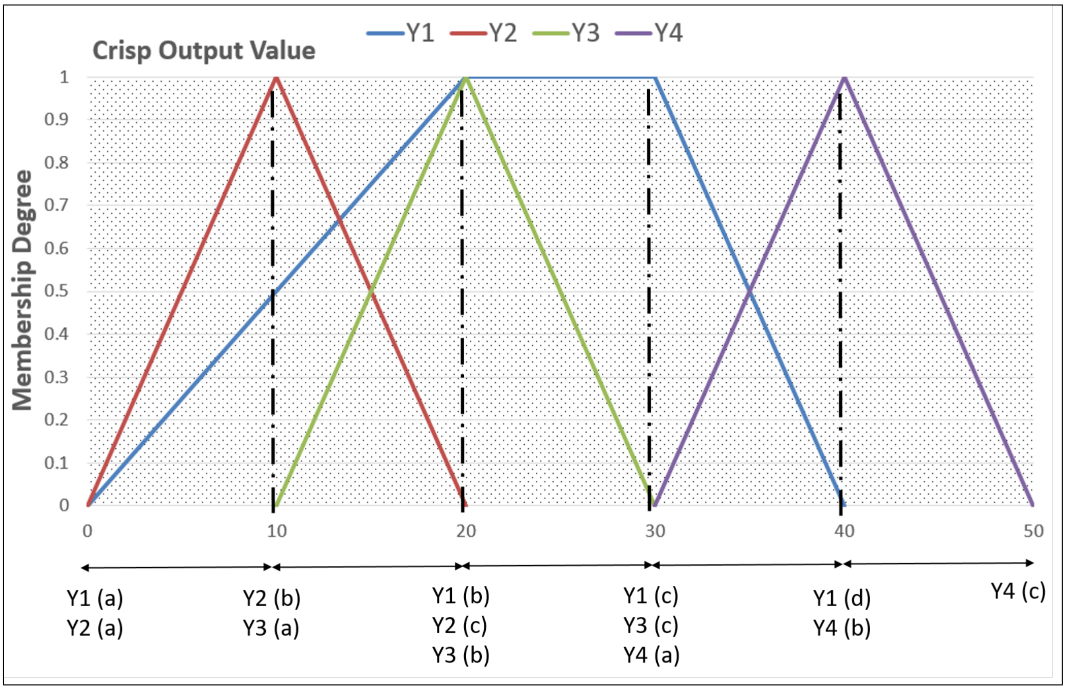

Figure 4.1: The output membership function of a generic fuzzy system.

ing, top, and ending points of triangles corresponding to the crisp output values. However, trapezoidal membership functions are exemplified by four points (a,b,c,d) in the population representation. What is expected from CLONALG-M is to obtain the approximate values of these points which improve the performance of the fuzzy clustering algorithm near to the optimality, when the initially expert-defined output membership function is given as input.

The number of created initial cells should be at least greater than or equal to the rule number. A modification is proposed in order for the immune system to defend against early convergence situations. Taking into account this modification, initial cells (P_i) are generated and assigned to the population. Then, a feasibility check is performed for each cell in the population. This feasibility check measures if the generated cell satisfies the validity measure of a fuzzy function or not. The validity measure is utilized as the covering of the whole range of the crisp output value. In order to test the validity, first, the range of the crisp output value is assigned to a set. Then, for each membership function, ranges that reside in a to c are removed from the set if the function is triangular; otherwise, ranges that reside in a to d are removed from the set if the function is trapezoidal. Finally, a check is performed to test whether the set is empty or not. If empty, the generated cell P_i is accepted; otherwise, regeneration of the cell that does not satisfy the validity measure is performed as depicted in line 7 of Algorithm 4. After all generated cells become feasible, the affinity of each cell is calculated.

The affinity of each cell is obtained from the HNA metric which depicts the round in the WSN at which half of the deployed wireless nodes die. It is a widely used metric in the context of efficiency comparison of clustering protocols. The corresponding HNA metric value is obtained by utilization of the First Order Radio Model, details of which is given in the experimental evaluation section. Since *Selection* and *Cloning* processes are done exactly in the same manner as the original CLONALG, the details of these operations are not explained here in order not to degrade the readability of the proposal. However, interested readers may refer to the original study [50] for a more comprehensive explanation of these operations.

$$NoC = \sum_{i-1}^{n} round(\beta \frac{N}{i}) \tag{4.1}$$

Representation of base lengths of membership functions depends on the range of the crisp output value. If the example given in this generic fuzzy output function, where the range is 50, is considered, the representation can be done by using 5 bits. The actual representation bit number varies according to the fuzzy output function domain interval values. After obtaining values of affinities, they are put in descending order. If it is presumed that the parameter $Size_{sel}$ is given as five (5), then the five (5) greatest affinity cells are chosen. The Number of Clones (*NoC*) of P_i is determined according to the Eq. 4.1. This equation employes the principles of affinity proportional reproduction stated in the original CLONALG.

One discriminating property of CLONALG-M is the mutation operator controlled by the employed constraints. Since changing the total number of rules is not intended and beyond the scope of this dissertation, which is also known as the compactification of the rule-base, the mutation (including the remutation) operates under the following constraints:

1. $\forall TRG \in MF \nexists i, j$ such that

    (a) $(a_i = a_j \wedge b_i = b_j \wedge c_i = c_j)$

    (b) $(a_i = b_i = c_i)$

    (c) $(a_i = c_i)$

2. $\forall TRP \in MF \nexists i, j$ such that

    (a) $(a_i = a_j \wedge b_i = b_j \wedge c_i = c_j \wedge d_i = d_j)$

    (b) $(a_i = b_i = c_i = d_i)$

    (c) $(b_i = c_i)$

    (d) $(a_i = d_i)$

In the constraints, MF stands for Membership Function, TRG stands for Triangular Membership Function, and TRP stands for Trapezoidal Membership Function. Seeing that the compactification of the rule-base is beyond the scope of this study, these constraints assure that generated MFs (both TRGs and TRPs) are unique. This controlled mutation is depicted as line 15 in Algorithm 4.

These elaborations conclude the example operation of CLONALG-M on a generic fuzzy output function and present the details of computational aspects. In the next section, the proposal is experimentally evaluated under depicted scenarios.

4.3 Experimental Evaluation

The details about the properties of utilized system model are elaborated for clarification purposes prior to presenting the obtained results of performance tests.

4.3.1 System Model

The system model and the assumptions that are utilized for the experimental evaluation in this chapter are the same as the corresponding ones which are explained in the previous chapter. Therefore, a graphical exemplification of the utilized model is not presented here again. However, there is a single additional assumption which is only valid for the experimental evaluations in this chapter. The additional assumption is as follows: The number and shapes of fuzzy membership functions are known a priori and do not change throughout the operation.

4.3.2 Performance Results

For the efficiency evaluation of the proposal, it is compared with the elaborated standard GA implementation on the selected fuzzy clustering protocols, which are CHEF and MOFCA, in three depicted scenarios.

In the scenarios, the protocols utilize a multi-hop routing methodology presented in detail in [18]. The CHEF α value is assigned as 2.5 and the threshold optimal value is determined by the use of equations in [17] and assigned as roughly 0.3, 0.2 for 100, 1000 nodes, respectively. The threshold value of the MOFCA algorithm is set to 0.3 as depicted in [20]. Sketch of the depicted scenarios is as follows:

- In Scenario 1, network boundaries are 100x100 m. and 100 nodes are randomly scattered into the network boundaries.
- In Scenario 2, network boundaries are 1000x1000 m. and 100 nodes are randomly scattered into the network boundaries.
- In Scenario 3, network boundaries are 1000x1000 m. and 1000 nodes are randomly scattered into the network boundaries.

The features of the GA implementation against which CLONALG-M is compared against are as follows:

- The implementation consists of 3 fundamental operations that are selection, genetic operations and replacement, which are the minimum requirements of a simple GA.
- The initial population is constructed at the size of 50 with random construction.
- Chromosome encoding follows the same principles depicted for TRGs and

TRPs in CLONALG-M computation. Fuzzy function related information is encoded in the chromosome.

- The fitness (evaluation) function that is utilized for the calculation is the same as in CLONALG-M, which is the HNA metric result that is calculated through the First Order Radio model.
- Parent selection is done proportionally. Matching in crossover is implemented randomly, crossover points are determined at the MF level so as to replace the depicted MF completely.
- Bit mutation in CLONALG-M computation follows the same principles depicted for TRGs and TRPs.

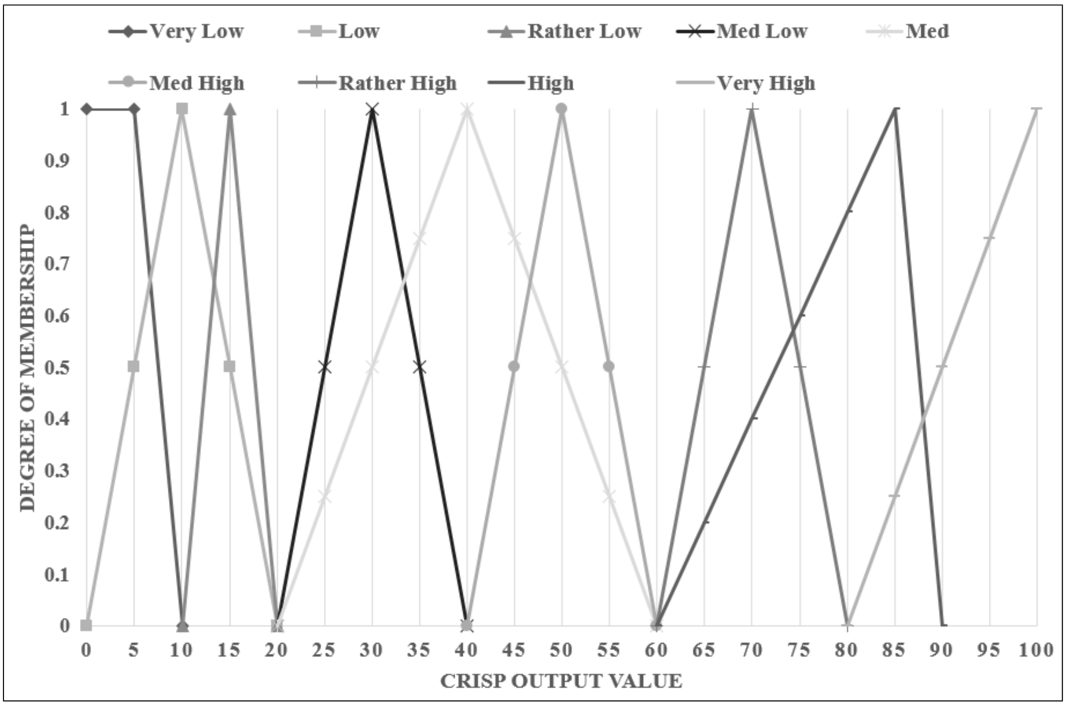

Figure 4.2: Initially defined fuzzy output function for CHEF.

In the tested scenarios, clustering has two distinctive stages: set-up and steady. Clusters are reformed at each round at the set-up phase. Then any member node transmits acquisition to its CH. The size of transmitted data is 4000 bits per round. Received data is aggregated with the defined ratio R_{agg} prior to relaying towards the base station. The assigned value of the aggregation ratio is 10% as used in [20].

Initially defined fuzzy output functions for the CHEF and MOFCA protocols are

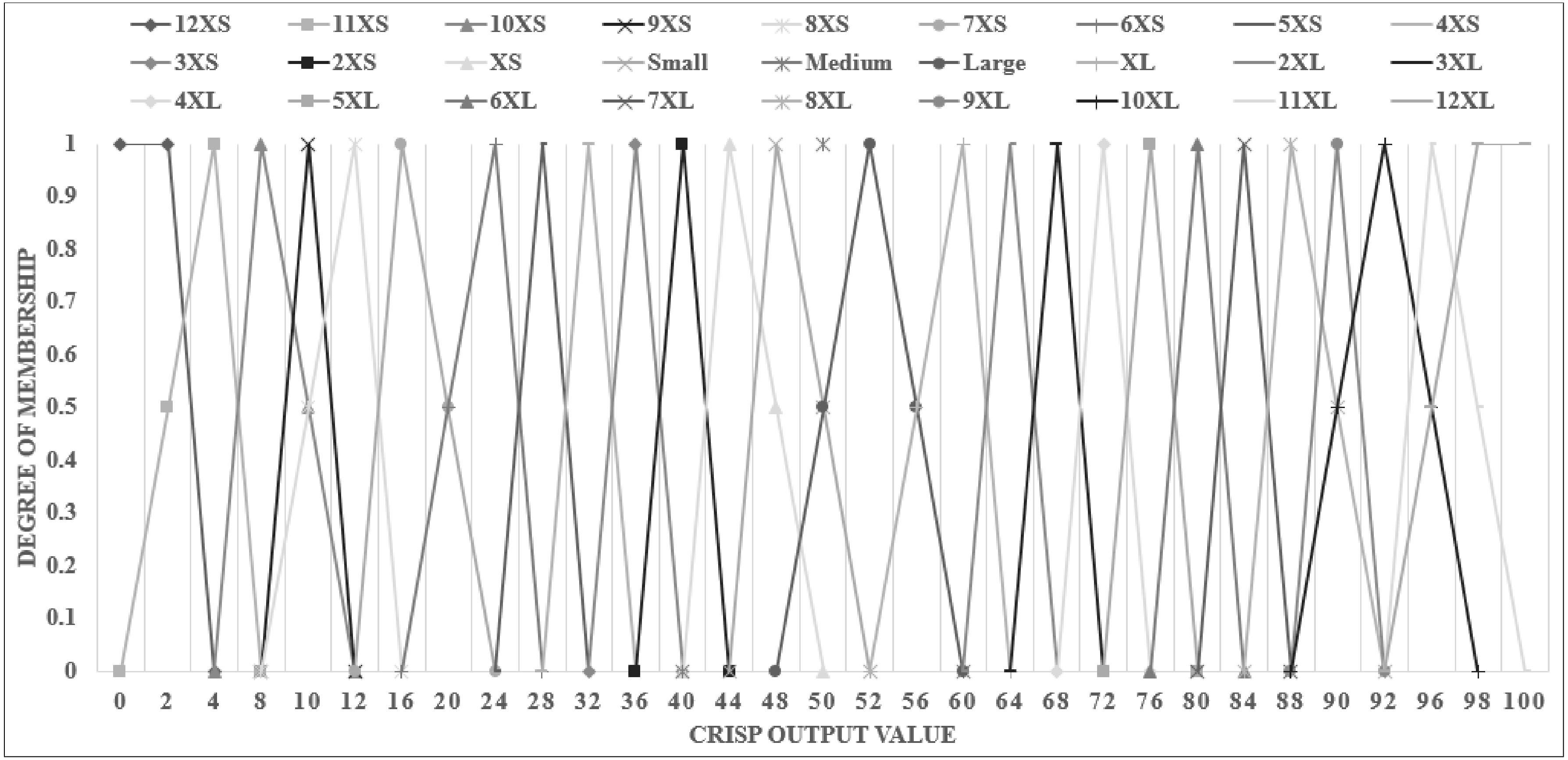

Figure 4.3: Initially defined fuzzy output function for MOFCA.

given in Fig. 4.2 and Fig. 4.3, and modified versions of these functions which are approximated to optimality using CLONALG-M are presented in Fig. 4.4 and Fig. 4.5, respectively. As can be seen from the figures, obtained fuzzy output functions conform to the depicted fuzzy validity measures and are not symmetric in form. In the remaining part of this subsection, the efficiency of the selected algorithms with respect to the applied approximation methodology is compared considering WSN performance metrics.

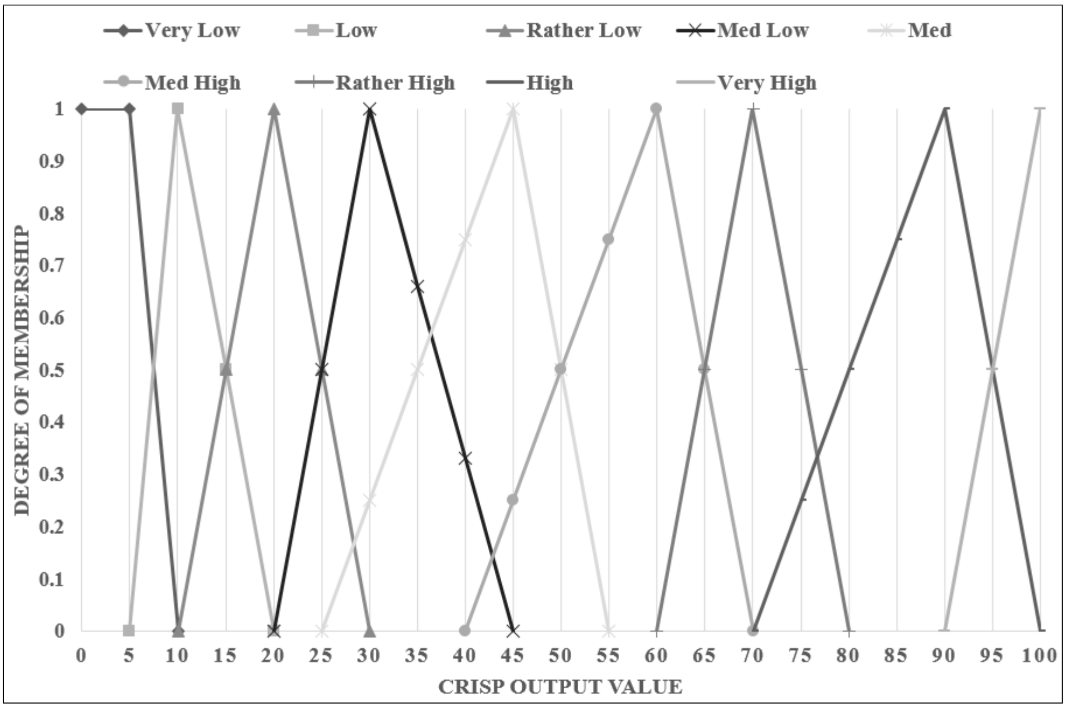

Figure 4.4: Fuzzy output function for CHEF obtained using CLONALG-M.

Commonly considered metrics in order to assess the lifespan of WSNs and efficiency perspective of protocols are presented in detail in the previous chapter. Therefore, the same metrics are employed to evaluate the performance of the methodologies.

A series of experimentation are run by utilizing the WSN simulator presented in [20] to be able to evaluate the validity of the proposal. The simulator is a Discrete Event System Simulator (DESS) and has an ability to simulate the chosen algorithms under the same environment. Conducted experiments are run on an eight-core 2.7 GHz Intel Core i7 workstation under the Windows 10 operating system. Each depicted scenario is run 20 times to acquire more stable and reliable metric results and the mean of the results are depicted in this section.

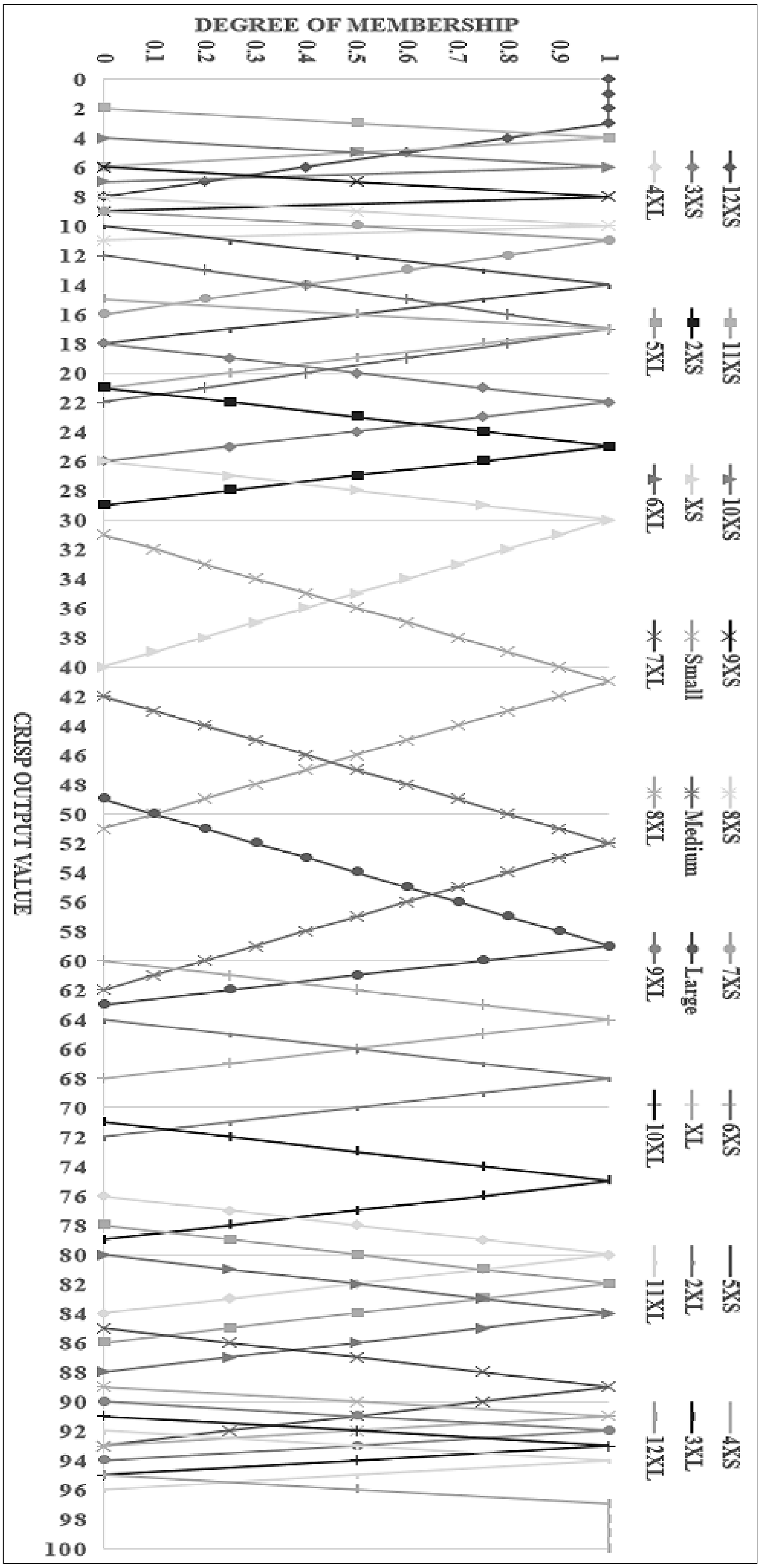

Figure 4.5: Fuzzy output function for MOFCA obtained using CLONALG-M.

4.3.2.1 Scenario 1

This is the base scenario where the preliminary comparison of depicted methodologies is performed. In this scenario, network boundaries are 100x100m. and nodes are deployed randomly within the boundaries of the network. Selected CHs of CHEF and MOFCA relay the aggregated data to the sink using multi-hop routing. The configuration applied in this scenario is depicted in Table 4.1.

Table 4.1: The Applied Configuration for Scenario 1

Parameter	**Value**
Network Size	100m x 100m
Location of the sink	(50,50)
Number of deployed nodes	100
Data packet size	4000 bits
ε_{mp}	0.0010pJ/bit/m^4
E_{elec}	50nJ/bit
Aggregation ratio	10%

Table 4.2: Simulation Results for Scenario 1

METHODOLOGY	**FND**	**HNA**	**TRE(j)**
CHEF	1197	1608	10.27
CHEF_GA	1314	1642	11.23
CHEF_CLONALG-M	1338	1616	12.17
MOFCA	1201	1618	11.59
MOFCA_GA	1346	1628	12.10
MOFCA_CLONALG-M	1359	1633	12.36

The obtained results in Scenario 1 are presented in Table 4.2. The maximum competition radius for the MOFCA algorithm in this scenario is set as 40 m. As can be seen from Table 4.2, the proposed approximation approach outperforms the original algorithm and the GA-approximated versions of both CHEF and MOFCA when considering the FND metric. In the scenario, TRE is measured in the round 1500. Performances of the original versions of both protocols are the worst when considering all metrics when compared to approximated versions which corroborate the gain pursued by optimization methodologies, either GA or CLONALG-M. CLONALG-M performs better than GA if the FND and TRE metrics are considered for both protocols. The number of alive nodes in connection with the number of rounds of this scenario is presented in Fig. 4.6. As grasped from the figure, the inception points for the death of the sensor nodes in both clustering algorithms that are approximated by CLONALG-M occur after original or GA approximated versions. Performances of CLONALG-M and GA initially look similar, however, regarding the HNA metric, GA performs better than CLONALG-M for both clustering methodologies. Although the performances of GA-approximated versions are better than CLONALG-M modified versions, GA-approximated versions deplete TRE faster CLONALG-M modified versions. In this scenario, CLONALG-M modified versions perform 10.5% better than the compared versions on average when considering the HNA metric, and 11.3% better than the compared versions on average when considering the TRE metric. Although the performance gain is nearly insignificant when it is compared against GA, the gain is crucial when it is compared against the original versions (CHEF and its versions).

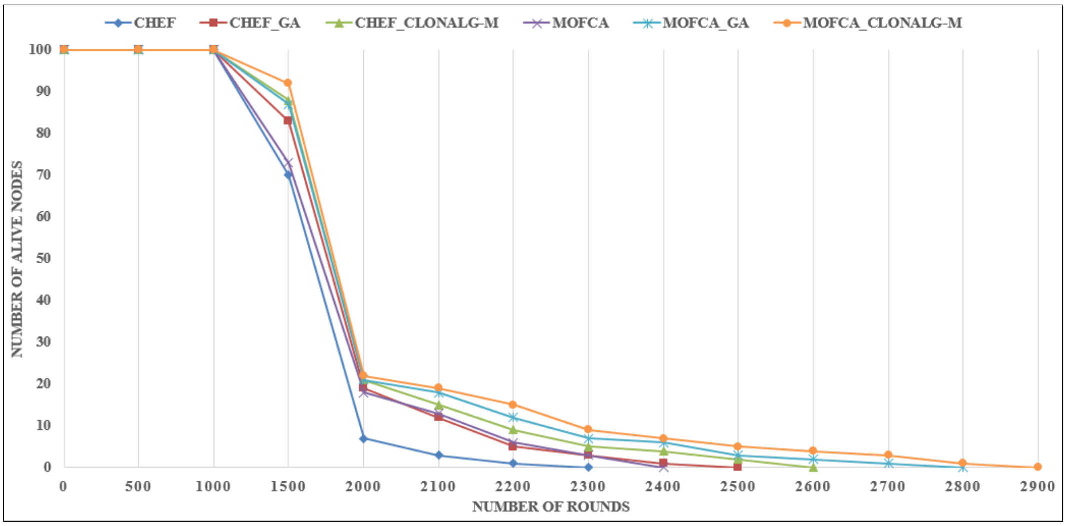

Figure 4.6: The number of nodes to the number of rounds in Scenario 1.

The number of clusters to the number of rounds and overall approximation impact on clustering algorithms in Scenario 1 are given in Fig. 4.7 and Fig. 4.8, respectively. According to Fig. 4.7, all versions of selected clustering algorithms except CLONALG-M approximated versions suffer initially from not creating the necessi-

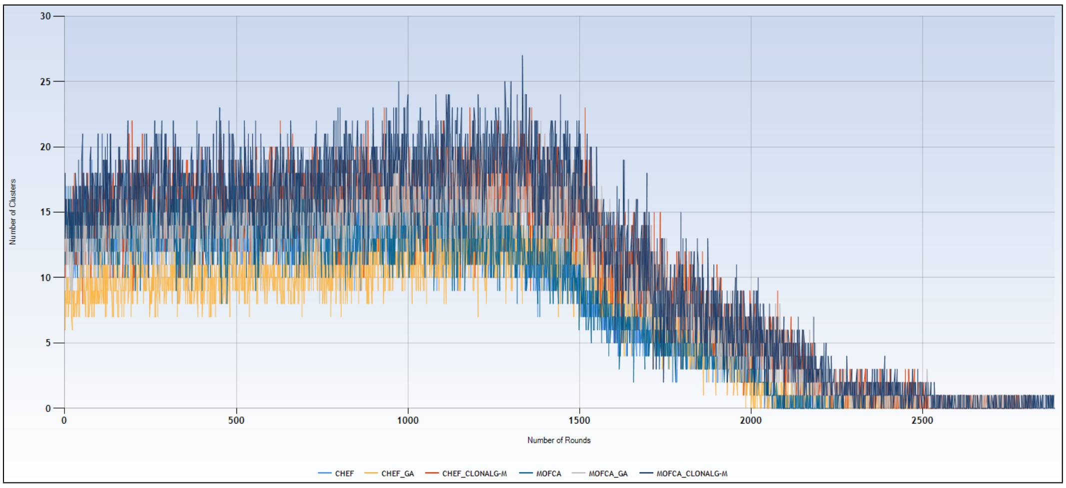

Figure 4.7: The number of clusters to the number of rounds in Scenario 1.

tated number of clusters. This situation forces those methodologies to consume more than necessary energy and, as a result, energy depletion balance decreases. However, as the rounds go by, the GA-approximated version improves genetically and maximizes the network lifespan when compared to CLONALG-M approximated versions. However, it cannot perform as efficient as CLONALG-M, which can be analyzed from the TRE metric result. According to Fig. 4.8, it is valid to state that approximated versions perform better than original versions in this scenario which corroborate the need for approximation approaches applied to these clustering algorithms.

4.3.2.2 Scenario 2

In this scenario, network boundaries are 1000x1000m. and nodes are randomly scattered within the network boundaries. Applied configuration in this scenario is depicted in Table 4.3. The reason for selecting this scenario is to exploit the effect of approximation methodology over the fuzzy clustering algorithms in sparsely deployed WSNs.

The obtained performance results of Scenario 2 are given in Table 4.4. The maximum competition radius for MOFCA is set as 60 m. In this scenario, TRE is obtained at the round 50. As can be corroborated from Table 4.4, the proposed approximation methodology performs better than all compared approaches when considering all metrics.

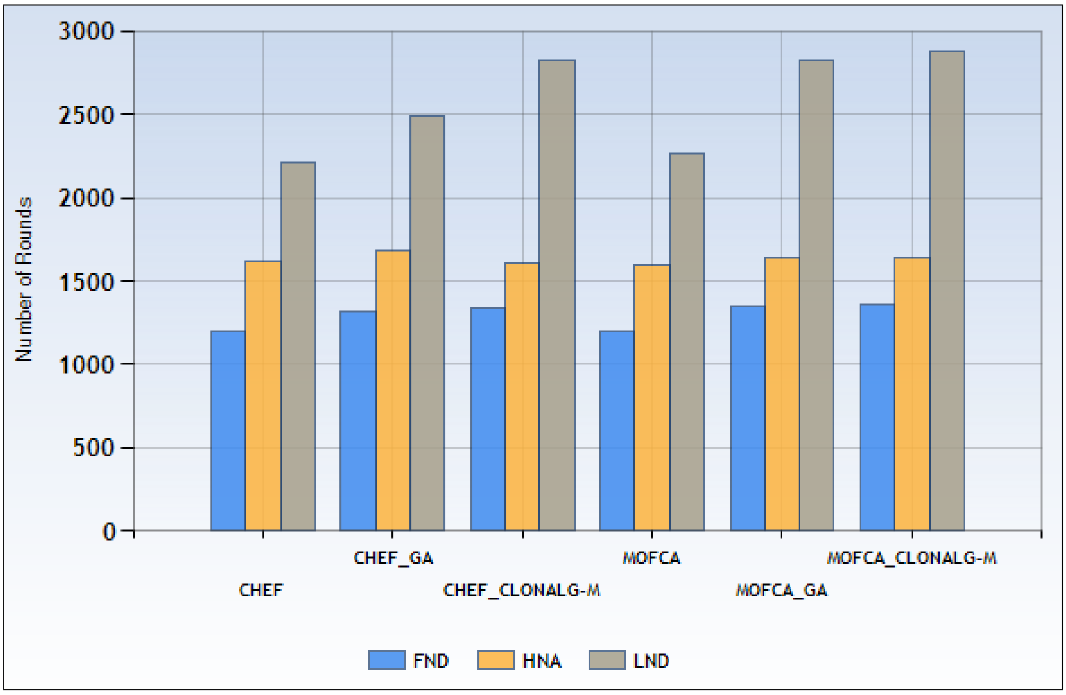

Figure 4.8: Overall approximation impact on clustering algorithms in Scenario 1.

Table 4.3: Configuration for Scenario 2

Parameter	**Value**
Network Size	1000m x 1000m
Location of the sink	(500,500)
Number of deployed nodes	100
Data packet size	4000 bits
ε_{mp}	0.0010pJ/bit/m^4
E_{elec}	50nJ/bit
Aggregation ratio	10%

Here, performances of the GA-approximated versions are the poorest, most probably because of the generated final fuzzy functions. Although this is valid, they do not cover the crisp range effectively. CHEF and MOFCA original versions outperform the GA-approximated versions when considering all of the metrics, which is different from the obtained results of Scenario 1. Expected from the GA was to approximate these fuzzy functions to optimality, however, it fails for sparsely deployed WSNs. In this scenario, CLONALG-M approximated versions perform 43.6% better than the compared versions in average when considering the FND metric, 10.4% better

Table 4.4: Simulation Results for Scenario 2

METHODOLOGY	FND	HNA	TRE(j)
CHEF	10	50	17.39
CHEF_GA	8	43	12.40
CHEF_CLONALG-M	16	52	18.93
MOFCA	18	46	18.29
MOFCA_GA	8	44	16.72
MOFCA_CLONALG-M	26	53	19.86

than the compared versions in average when considering the HNA metric, and 14.3% better than the compared versions in average when considering the TRE metric.

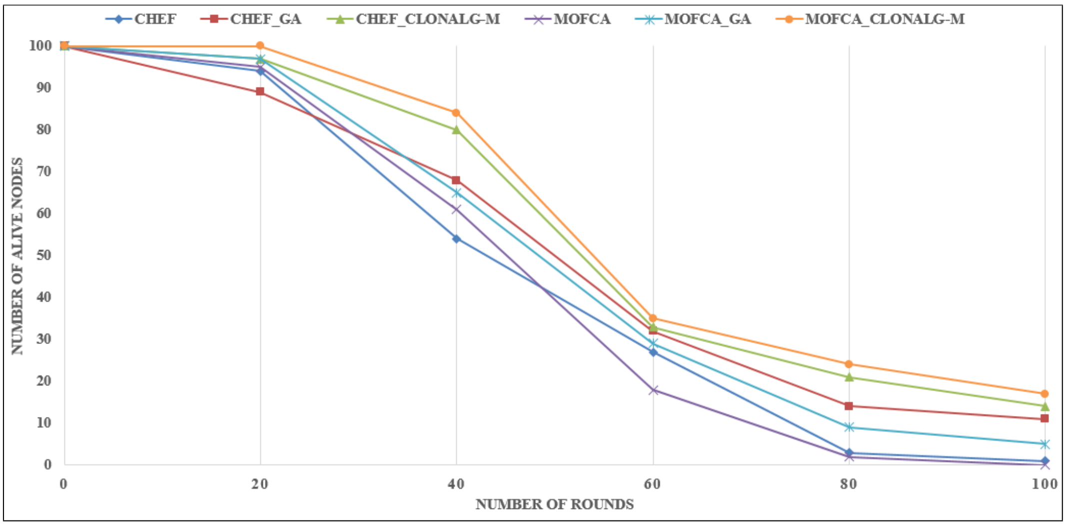

Figure 4.9: The number of nodes to the number of rounds in Scenario 2.

The number of alive nodes in connection with the number of rounds in Scenario 2 is depicted in Fig. 4.9. According to the obtained result, it is valid to state that the sparse deployment type of the network has an uttermost impact on the efficiency of GA since this is both true for CHEF and MOFCA GA-approximated versions in this scenario.

The number of clusters in consideration with the number of rounds and overall approximation impact on clustering algorithms in Scenario 2 are given in Fig. 4.10 and Fig. 4.11, respectively. As depicted in Fig. 4.10, GA-approximated versions of

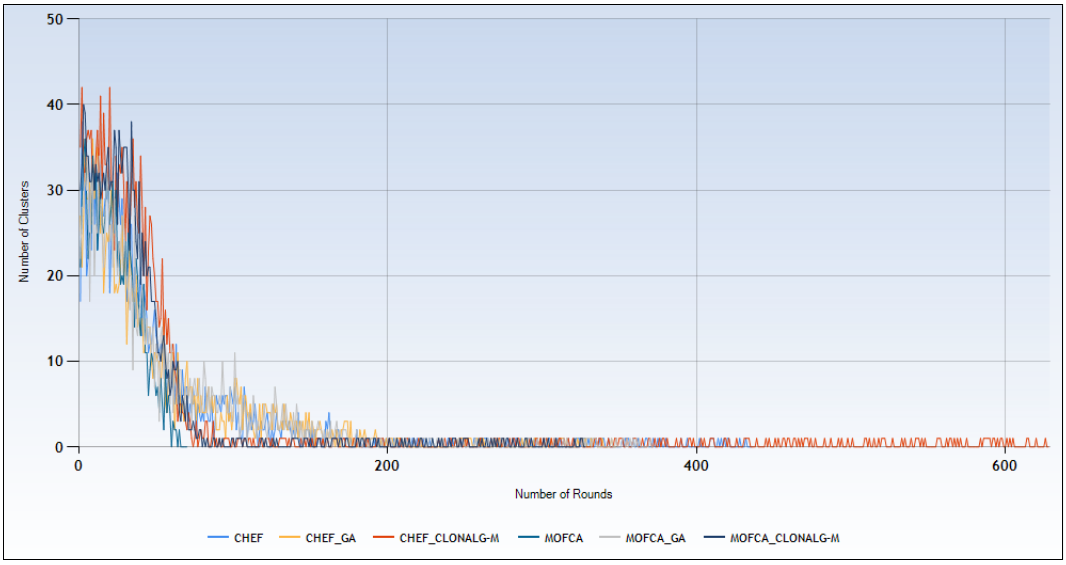

Figure 4.10: The number of clusters to the number of rounds in Scenario 2.

CHEF and MOFCA again suffer from not creating the required number of clusters in this scenario. Moreover, different from the previous scenario, GA-approximated versions cannot acquire this property throughout the run-time of the protocols. According to Fig. 4.11, the performance of CLONALG-M is significant if LND metric is paid attention. However, as stated in the previous subsection, this metric is not considered since it is the final round of operation in WSNs. In this figure, results of the FND and HNA metric both corroborate presented results of Table 4.4.

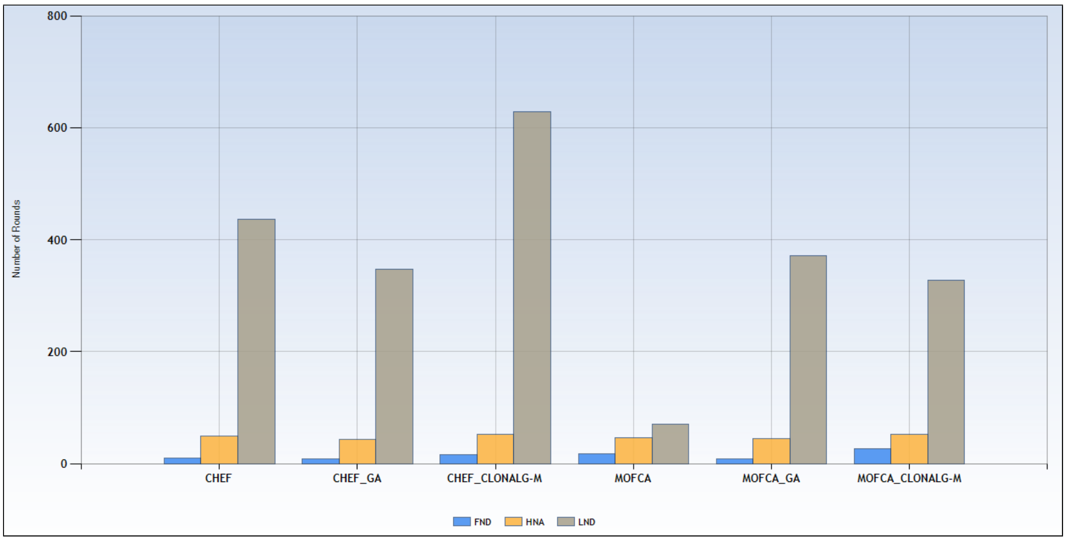

Figure 4.11: Overall approximation impact on clustering algorithms in Scenario 2.

4.3.2.3 Scenario 3

In this scenario, network boundaries are 1000x1000 m. The applied configuration of this scenario is similar to Scenario 2, except the number of deployed nodes, so it is not given here again due to page restrictions. The motivation behind choosing this scenario is to highlight the effect of applied methodologies over selected fuzzy algorithms in densely deployed WSNs. Additionally, this scenario serves to test and measure the scalability properties of the proposed approach.

Table 4.5: Simulation Results for Scenario 3

METHODOLOGY	FND	HNA	TRE(j)
CHEF	6	129	376.99
CHEF_GA	15	133	456.99
CHEF_CLONALG-M	16	144	474.72
MOFCA	11	122	447.62
MOFCA_GA	12	135	477.08
MOFCA_CLONALG-M	13	142	490.06

The obtained performance results of Scenario 3 are given in Table 4.5. The maximum competition radius for MOFCA is set as 70 m. In the scenario, TRE is measured in the round 100. As can be seen from Table 4.5, CLONALG-M approximated versions of both protocols performs better than the original and GA-approximated versions of CHEF and MOFCA in this scenario when considering all metrics. However, the operation of GA-approximated versions are similar to CLONALG-M approximated versions for FND metric. GA or CLONALG-M can be applied in place of each other. Although, there is no significant performance gain between methodologies considering the FND, the difference increases as the rounds go by. In this scenario, the CLONALG-M approximated versions perform 6.6% better when considering the FND metric, 6.2% better when considering the HNA metric, and 3.1% better when considering the TRE metric than the GA-approximated versions. The performance gain increases if the original versions are taken into account in the comparisons.

The number of alive nodes with respect to the number of rounds in this scenario is depicted in Fig. 4.12. According to Fig.4.12, the decrease in the number of alive

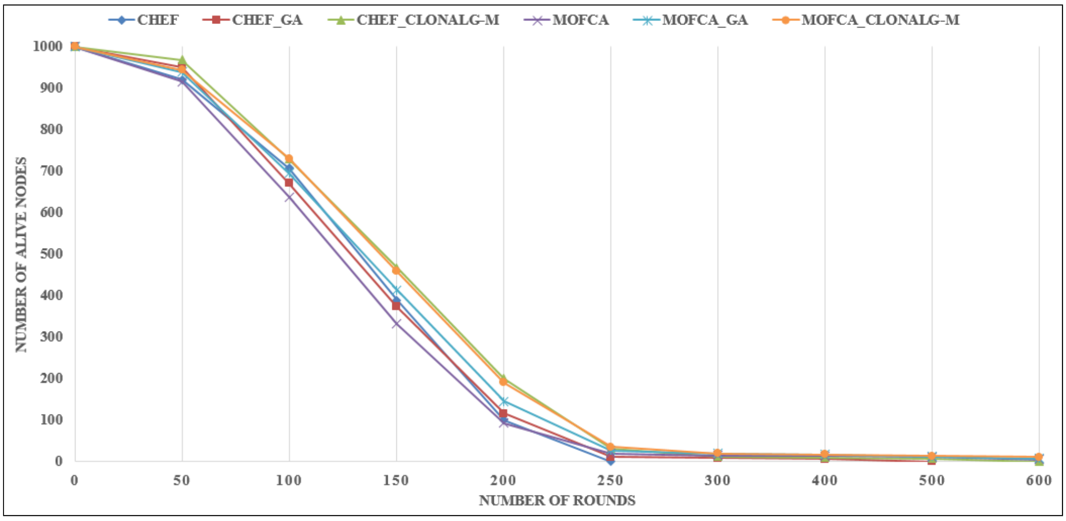

Figure 4.12: The number of nodes to the number of rounds in Scenario 3.

nodes for all compared methodologies are more or less follows the same routine in this scenario. They pursue similar energy consumption patterns. However, as can easily be seen from the figure, CHEF_CLONALG-M and MOFCA_CLONALG-M versions outperform the compared methodologies significantly until round 300. From this round on, the compared approaches catch up CLONALG-M approximated versions. However, CLONALG-M versions maintain their efficient operation manner throughout the lifespan of the network.

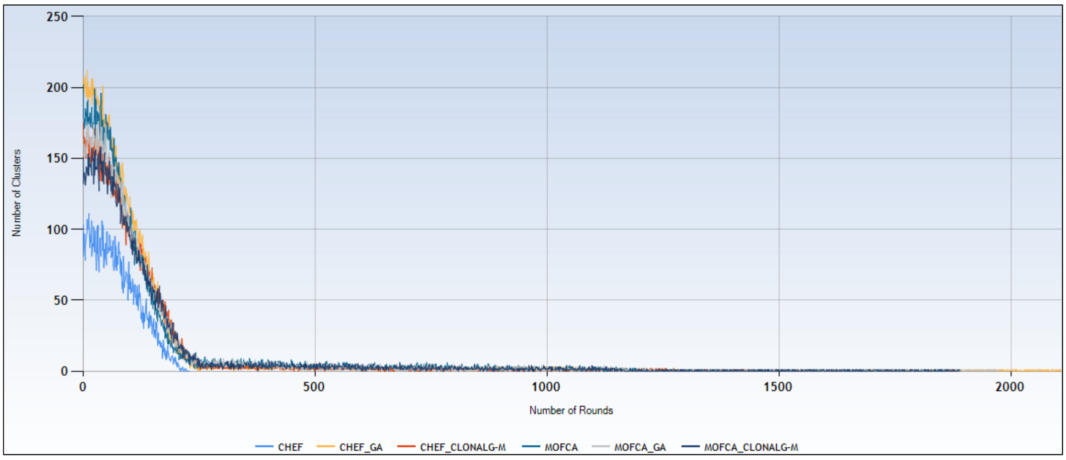

Figure 4.13: The number of clusters to the number of rounds in Scenario 3.

The number of clusters with respect to the number of rounds and overall approximation impact on clustering algorithms in Scenario 3 are given in Fig. 4.13 and Fig. 4.14, respectively. In this scenario, efficient operation of CLONALG-M approxi-

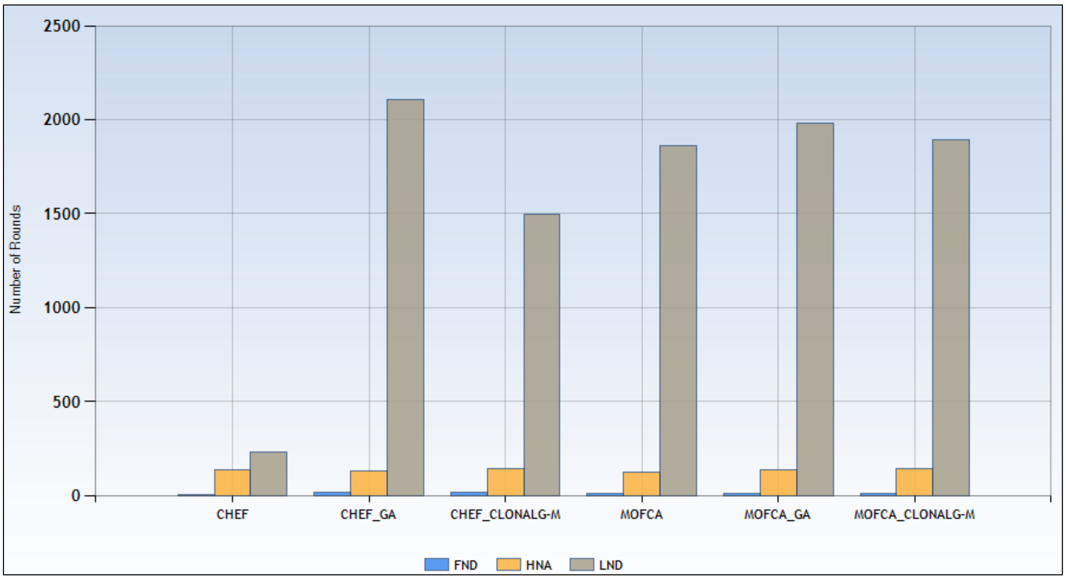

Figure 4.14: Overall approximation impact on clustering algorithms in Scenario 3.

mated versions are maintained as in previous scenarios.

If the obtained performance results in the tested scenarios are analyzed, it can be concluded that CLONALG-M is a promising approach to be applied for fuzzy output function approximations in rule-based clustering applications in WSNs. When the impact of network size and the number of nodes in network boundaries over compared methodologies are evaluated, it is valid to state that they have insignificant distinctive values since they do not affect the way the protocols operate. However, since the efficient operation of CLONALG-M is maintained as the network size and the number of nodes increase, it is also valid to state that CLONALG-M is an efficient methodology and scales well for large-scale WSNs.

4.4 Remarks

In this chapter, a modified clonal selection algorithm (CLONALG-M) is proposed to be applied for performance improvement of rule-based fuzzy clustering algorithms in WSNs. The proposed CLONALG-M algorithm modifies the basic principles of the original CLONALG approach while considering the fuzzy validity measures in the mutation and generation phases of the possible solutions.

According to the conducted experimental evaluations, it enhances the performance of the fuzzy clustering algorithms and approximates the fuzzy output functions to optimality. Experimental analysis conducted on the depicted scenarios clearly shows that the efficiency of CLONALG-M is fairly better than the compared standard GA implementation in all experiments run throughout this chapter.

In this chapter, fuzzy output function approximation using CLONALG-M is done offline because of its computation-intensive nature, which is also the case for GA approximations in the conducted tests.

This concludes the approximation of fuzzy output functions proposed for utilization in rule-based fuzzy clustering systems with the purpose of increasing the system performance.

CHAPTER 5

AN EMPIRICAL STUDY ON MULTI-MODAL OBJECT CLASSIFICATION PROBLEM IN WIRELESS SENSOR NETWORKS

In this chapter, an empirical study on multi-modal object classification problem in wireless sensor networks is conducted in order to corroborate whether or not the proposed intelligent fuzzy clustering methodology,while being energy-efficient, maintains object classification accuracy when its compared against the selected clustering algorithms.

The presented empirical study is based on the related work described in the previous chapters and the research project supervised by The Scientific and Technological Research Council of Turkey under Grant No. 114R082. The analysis helps to comprehend how a multi-modal object classification performance of a WSN can be augmented by utilizing the proposed intelligent fuzzy clustering methodology. For this reason, required explanations are elaborated upon completing the brief description.

5.1 Brief Description

In this thesis, it is implicitly asserted that the proposed intelligent methodology not only improves energy-efficiency but also maintains the classification accuracy when utilized in multi-modal object classification problems in wireless sensor networks. According to this assertion, new approaches for fuzzy clustering and fuzzy optimization are proposed and extensive experimental analysis is conducted in previous chapters. Although performed experimentation is enough to prove the benefits of the proposed methodologies considering the energy-efficient operation requirements of a protocol, conducting an empirical study on multi-modal object classification problem on wireless multimedia sensor networks can highlight the positive features of the proposals also considering the classification accuracy.

5.2 Multi-Modality Revisited

As delineated in the previous chapters about data fusion domain, the multi-modality issue not only occurs in multimedia processing but also in multi-sensor data processing. An ever-increasing number of applications deployed on WSNs tries to discriminate between detected objects based upon their characteristics. Most studies in the literature treat data acquired from different types of sensors as a different modality for fusion in order to handle inefficiency occurring in the classification process.

In terms of explaining the basics of the conducted empirical study, setups in which an object (phenomenon) is observed using multiple sensors, each of which is a member of a different type such as PIR, acoustic, imaging, microphone, or seismic, are considered. In this type of a setup, each acquired data from a sensor-type is described as a separate modality. If these modalities are not correlated in time or space, then they are associated with one data set. However, when necessary correlations occur through fusion, the whole data is regarded as multi-modal. A wireless node can access the multi-modal data if the node consists of different types of sensors. The multi-modal data access is also valid for a CH since different modalities at each wireless node are transferred to the CH for fusion and aggregation purposes. A vital feature of multi-modal data, as stated in [73], is its *complementarity* in the sense that several kinds of value are added to the whole, acquisition of which is generally impossible from any remaining modality in the environment.

5.2.1 Fusion System

Data fusion mainly struggles to make a more efficient and/or accurate decision by the combination of available input sources. In other words, fusion methodologies and aggregation techniques acquire the most inclusive and identifying datum related to an entity by the utilization of various correlated data that are supplied from different sources, therefore facilitating the end user to increase accuracy, decrease noise, and extract semantic information for decision making. For this reason, fusion is implemented in this empirical study in the object classification process considering the light-weight and accurate operation requirements of WSNs. To emphasize again, the following are the main points of discussion for not sticking to and counting upon a single source:

- Fusion of supplementary data ensures a more descriptive exemplification of the subject matter. Additionally, if different fusion sources consist of redundant data, then inexactness or vagueness of a decision decreases. Thus, robustness and reliability of a possible decision increases [54],
- One of the functional benefits of data fusion resides in its unreliable source filtering property. In real life applications, it is often unknown to deduce the performance or added value that is brought by additional modality at the design time. Not only the performance but also the reliability of the modality cannot be predicted correctly in real-world environments. From this perspective, dependency on any of the available sources or their weights on the overall result can together be decreased [54],
- Noise in the data is an undesirable property which may cause incorrect classification results. Multiple sources in this respect can operate as a filter and decrease the effect of noise on the result [51],
- While no source is impeccable, the collaborative effort among them generally produces better results hence improves usability [55],
- Each input has an implicit upper bound on the overall result. Therefore, no system performance can be repeatedly tuned by modifying its components or operation steps [56].

5.2.2 Design Principles

Although there exist various expectations from a real-life fusion system design, effectiveness, efficiency, and feasibility are placed among the top third. These expectations can be reported as but not limited to:

- An output obtained from a fusion system could be operated at least as efficiently as any of its inputs [54],
- Performance and robustness of classification should be increased by the utilization of fusion [54], [74], and[75],
- A true fusion system should be feasible in the sense that the design principles could easily be implemented in real-world scenarios,
- Flexibility and speed of operation are among the desired properties of a fusion system [76].

The aforementioned expectations are valid for any fusion system independent of its application domain. Moreover, there are other requirements such as a light-weight operation architecture together with the accuracy of object/event detection process and energy-efficiency when the system under discussion is applied to a domain of WSNs where resource-constraints apply.

5.3 System Architecture

This conducted empirical study is based on the research project supervised by The Scientific and Technological Research Council of Turkey under Grant No. 114R082. With the project, a new approach and a framework targeting the energy-efficiency and accuracy trade-off problems in WMSNs are aimed. The focus is on increasing accuracy of transferred information as well as the wireless network energy efficiency. In the project, by using fuzzy clustering algorithms, a wireless sensor network consuming much less energy than currently used ones is constructed and realized. Thus, a WMSN framework that reduces energy consumption while at the same time preserving accuracy is developed and experimentally verified. The system architecture applied in the research project is presented in Fig. 5.1.

In the scope of the research project, a WMSN node is realized on a Raspberry Pi Model B+ board as demonstrated in Fig. 5.2, equipped with necessary sensors and is set up so that the node operates in a collaborative manner. The WSN that this empirical study conducted on consists of a total number of 100 virtual nodes that are

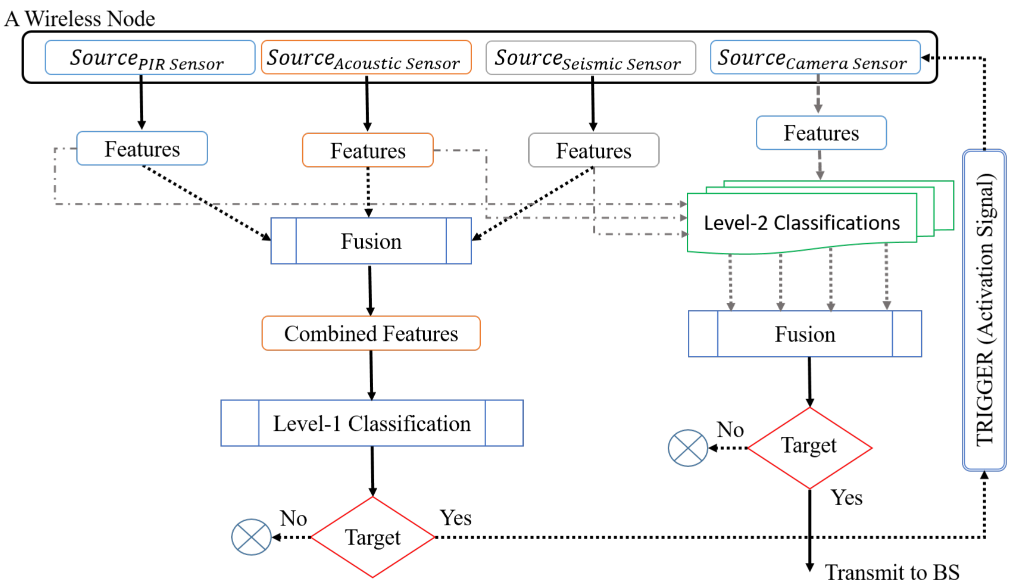

Figure 5.1: The research project system architecture.

deployed into an area size of 1000 x 1000 m, and the first order radio model is used for energy consumption measurement as described in the previous chapters.

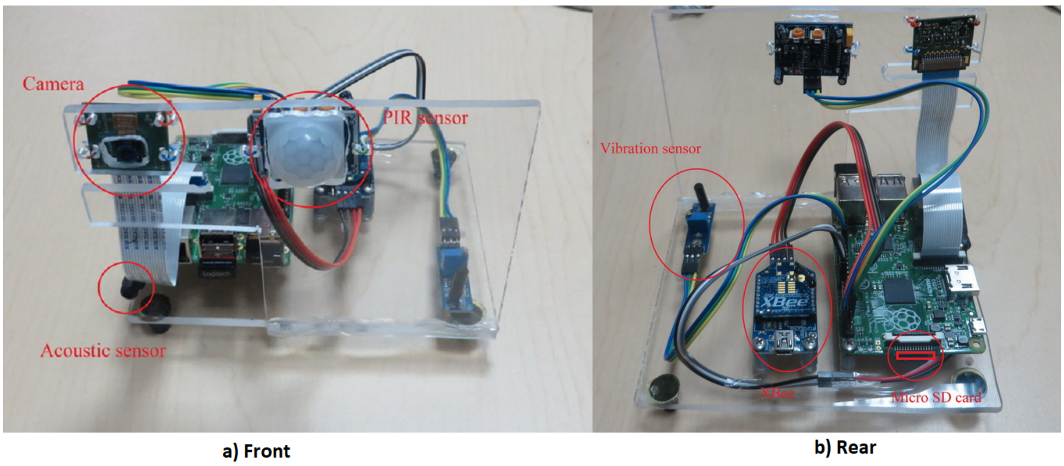

Figure 5.2: The Realized WMSN Node.

5.4 Implementation Details

Since the clustering process directly relates to paths data flows through, the following modifications are made to the system architecture presented in Fig. 5.1 in order to measure and emphasize the effect of clustering on classification:

- Each wireless node is only equipped with three sensors, one of them is PIR and the other sensors are from the seismic, acoustic, or camera sensor classes.

- The first and second level fusion schemes are both late fusions such that separate classifiers are utilized for each modality data.

- Due to the first modification, the triggering process is canceled and second level fusion and classifications are only performed on the first level CHs, by considering the timestamps and locations of the detections, of which the data acquired nodes are members.

The reasons behind the modifications are as follows:

- Firstly, all available wireless nodes in the market may not utilize camera sensors due to using these sensors as higher-order assets in some nodes,

- Secondly, with this modification, in addition to homogeneous sensor networks, heterogeneous counterparts can also be targeted,

- And finally, also as a crucial reason, impact of clustering on classification can be rendered vague if all fusion and classification operations are done following an in-node processing manner. In the in-node processing case, only final classification results are transmitted to CHs and CHs relay the acquisitions to the sink. However, with the modifications done on the architecture, it is now possible to transmit one modality representing a phenomenon to one CH and one other modality representing the same phenomenon to another CH. When this is the case, there is a chance for the fusion and classification results to change due to the availability or unavailability of a modality at a CH.

Prior to the network deployment, clock synchronization is initially performed. All clocks are set to the beginning of time, which is 00(h):00(m):00(sec):000(msec). Then, when there is no object in the range of specific sensor types, measured background noise of the utilized sensor is generated and employed as the threshold value for the detection of signals. Thereafter, created class signatures pertaining to the seismic, microphone, acoustic, and camera sensor types are stored in the wireless nodes. Finally, the simulation drawer generated data (objects) are flushed into the network considering the timestamps and locations of the objects. In the class signature preparation, seismic and acoustic data properties are employed as delineated in [77] and [78], respectively.

Classification outputs pertaining to each modality at sensor or CH level are generated as follows:

- PIR data is utilized as a boolean data which discriminate between the presence or absence of an object and associated with a class label randomly since PIR data is regarded as useless in the object classification process.

- For the seismic classification, k-Nearest Neighbor algorithm is employed. Since the utilized algorithm is a supervised approach, classifier is trained offline prior to the node deployment phase in order to discriminate between class labels.

- For the acoustic classification of the detection, signal analysis is performed in order to determine the target presence by utilizing Binary Fuzzy Classifier as described in [79]. Here, different from the original study, class labels are generated instead of evidences. There are multiple binary fuzzy classifiers utilized in an hierarchical manner and each classifier is a fuzzy rule-based classifier.

- For visual classification of the detection, *Speed* and *Shape Ratio* are extracted through the utilization of Minimum Bounding Rectangle of the detection. Elaborations related to the visual classification approach pursued for this modality reside in [80].

For the fusion of classification results, two different methodologies are applied and the performance of each is presented in the evaluation section. These methodologies are as follows:

- *Standard Weighted Averaging*: Each modality possesses an expert defined weight in the calculation of a fusion result. The assigned weight for the imaging modality is equal to 65%, is equal to 15% for the seismic modality, is equal to 15% for the acoustic modality, and is equal to 5% for the PIR modality.

- *Simulated Annealing-based Weighted Averaging*: Like the standard weighted averaging methodology, each modality possesses a weight obtained from the

Simulated Annealing algorithm. The calculation of the SA-based weights follows from the Optimization Framework, internals of which are explained in Subsection 3.3.2 of Chapter 3. The only difference lies in the utilization of different input parameters. In this specific purpose, inputs of the SA algorithm are the expert-defined weights of modalities.

It is also noteworthy to state that the standard weighted averaging is also applied in a modified manner in this empirical study due to the presence or absence of a sensor type in a wireless node. Since each wireless node is geared with three types of sensors, one of which is PIR sensor and the others are from the remaining sensor types, assigned weights are modified in favor of the sensors except the PIR in the sense that the total weight assignment in a node is still equal to 100%. For instance, if a wireless node consists of a PIR sensor together with imaging and seismic sensors, a total of 8% is assigned to the PIR modality, a total of 71% is assigned to the imaging modality, and 21% is assigned to the seismic modality. That is to say, 1/5 of the remaining weight is assigned to the PIR modality and 2/5s of the remaining weight are assigned to the other modalities.

Additionally, SA-based weights are calculated offline due to the computation-intensive nature of the operation and obtained as roughly 70% for the imaging modality, 14% for the seismic modality, 14% for the acoustic modality, and 2% for the PIR modality. The distribution of the remaining weight on a wireless node is done as in the case for the standard weighted averaging approach.

5.5 Evaluation

The graphical exemplification of the utilized system model for node communication is presented in Chapter 3; therefore, not presented here again. But, it is noteworthy to state that the member nodes perform data acquisition with the equipped sensors, fuse and classify detections, and then send the resulting data, which is the decided class of an object to their CHs utilizing a single-hop communication channel, whereas CHs accumulate the decisions of member nodes, perform higher level classification if possible and then relay the result to the destination via multi-hop routing methodology.

With the aim of evaluating the proposed intelligent fuzzy clustering approach in terms of classification accuracy, it is compared against CHEF and MOFCA in a predefined scenario. In the scenario, TTDFP utilization is flourished with intelligence in the sense that the SA procedure in TTDFP is triggered when the ratio of the alive-node count with respect to the total deployed nodes differentiates in quartiles. TTDFP utilizes its fuzzy routing methodology whereas CHEF and MOFCA employ the multi-hop routing methodology presented in [18]. The CHEF α value is given as 2.5 and the threshold optimal value is determined by the use of equations in [17] and assigned as roughly 0.3 for 100 nodes as in the original study. Threshold and coefficient values of TTDFP are obtained using SA as explained in Chapter 3 and vary at each quartile.

5.5.1 The Data Set

This empirical study is conducted on a generic data set generated by the WSN data simulator developed by Cihan Küçükkeçeci. Data Simulator concentrates on generating a graph-based big data model so as to simulate wireless sensor networks. The simulator deploys initiated sensor nodes manually on a rectangular area depicted by their latitude and longitude values and insert the obtained values as locations of the nodes into the OrientDB graph database. However, for this study, the manual deployment feature is modified as random deployment, and the data store is modified as SQL Server, which stores the locations of nodes as coordinates in 2D space in accordance with the requirements of this empirical study.

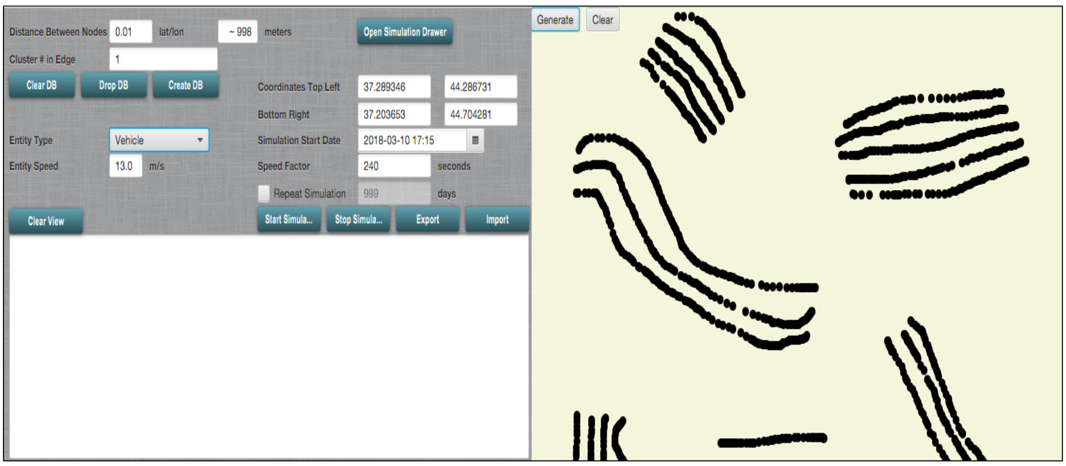

Figure 5.3: Snapshots from the original data simulator.

In order to create objects in the rectangular area, simulation drawer is employed. Snapshots from the data simulator are shown in Fig. 5.3. According to the chosen phenomenon type on the drawer and following a path on the drawer, object features corresponding to the utilized sensor types are created with their timestamps. In this empirical study, 500 human class, 500 animal class, and 500 vehicle class of objects are utilized, their relevant data is generated using the simulator, and in order to train and measure the performance of the classifiers, ten-fold cross validation is applied.

5.5.2 Performance Metrics

Similarly to TRE, HNA, and FND metrics utilized for predicting the lifespan of sensor networks and measuring the efficiency aspect of protocols, *Precision* and *Recall* metrics are utilized with the purpose of measuring the classification accuracy resulted by the use of compared protocols interchangeably in the clustering phase together with *F-Score* metric which is applied to measure the accuracy of the conducted tests.

Precision depicts the portion of relevant instances among the retrieved instances whereas *Recall* depicts the portion of relevant instances retrieved over all relevant instances, which are useful in describing the classification performance of a system.

Precision is also phrased as "positive predictive value" and recall is phrased as "sensitivity" in the literature. The performance of the classifier being tested for a classification task can be compared with reliable external judgments using the terms true positives (tp), true negatives (tn), false positives (fp), and false negatives (fn). These positive and negative terms express the classifier's estimation (expectation), and the true and false terms express whether that estimation aligns with the observation.

$$Precision = \frac{tp}{tp + fp} \tag{5.1}$$

$$Recall = \frac{tp}{tp + fn} \tag{5.2}$$

In the light of the clarifications made, precision and recall metrics can be defined in mathematical terms as in Eq. 5.1 and Eq. 5.2, respectively.

As stated, *F-Score* metric is employed with the purpose of corroborating the accuracy of the conducted tests. Since *F-Score* metric is the harmonic average of the depicted metrics in Eq. 5.1 and Eq. 5.2, it can be described formally as in Eq. 5.3.

$$F - Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{5.3}$$

Therefore, these described metrics are employed for the purpose of evaluating the classification performance of the compared algorithms in the same setup.

5.5.3 Experimental Results

A series of experimentation are run by utilizing the WSN simulator in [20] in order to simulate the clustering phase at each round. The simulator is a Discrete Event System Simulator (DESS) and can simulate the coded algorithms under the same environment. Conducted experiments are run on an eight-core 2.7 GHz Intel Core i7 workstation 20 times to acquire more stable and reliable metric results and arithmetic mean of obtained results are sighted in this subsection.

Table 5.1: Classification Results using Standard Weighted Averaging

Clustering Algorithm	Object	Precision	Recall	F-Score
CHEF	Animal	63.4%	52%	0.571
	Human	81.8%	72%	0.765
	Vehicle	71.4%	60%	0.652
	Average	72.2%	61.3%	0.662
MOFCA	Animal	69.2%	54%	0.606
	Human	88.3%	76%	0.816
	Vehicle	69.7%	60%	0.644
	Average	75.3%	63.3%	0.688
TTDFP	Animal	67.5%	54%	0.600
	Human	84.4%	76%	0.799
	Vehicle	72%	62%	0.666
	Average	74.6%	64%	0.688

The obtained classification performance results for the experimented scenario using the standard weighted averaging and the SA-based weighted averaging approaches are presented in Table 5.1 and 5.2, respectively.

Table 5.2: Classification Results using SA-based Weighted Averaging

Clustering Algorithm	**Object**	**Precision**	**Recall**	**F-Score**
CHEF	Animal	68.2%	56%	0.615
	Human	86.6%	78%	0.820
	Vehicle	70.4%	62%	0.659
	Average	75.1%	65.3%	0.698
MOFCA	Animal	67.5%	54%	0.600
	Human	84.4%	76%	0.799
	Vehicle	74.4%	64%	0.688
	Average	75.4%	64.6%	0.696
TTDFP	Animal	66.6%	56%	0.608
	Human	84.7%	78%	0.812
	Vehicle	74.4%	64%	0.688
	Average	75.2%	66%	0.703

If the impacts of SA-based weighted averaging and standard weighted averaging are compared with respect to the obtained classification performance results as presented in Table 5.1 and 5.2, it is clearly visible that SA-based weighted averaging improves the accuracy ratios for all utilized algorithms in most classes, which explicitly denotes the pros of the methodology. Moreover, since SA-based weighted averaging improves the accuracy ratios *for all* utilized algorithms and does not boost any bad performing algorithm to a better place or the vice-versa, it also implies that it has a significant distinctive value when compared to the standard weighted averaging approach.

In order to corroborate and derive the significance of the finding, an unpaired t-test is conducted on the results of TTDFP operation over the re-randomized random data. To be able to conduct this test, average classification accuracy results for the F-Score metric pertaining to standard weighted averaging approach and SA-based weighted averaging approach are obtained through the utilization of the algorithm 20 times for each approach.

According to the acquired result of the test; $t - value$ is obtained as -2.25 and the probability of this result, assuming the null hypothesis, is obtained as 0.030. The box-plot of the obtained result in the unpaired t-test is presented in Fig. 5.4. In the figure, *A* and *B* depicts the standard and SA-based weighted averaging methodologies, respectively.

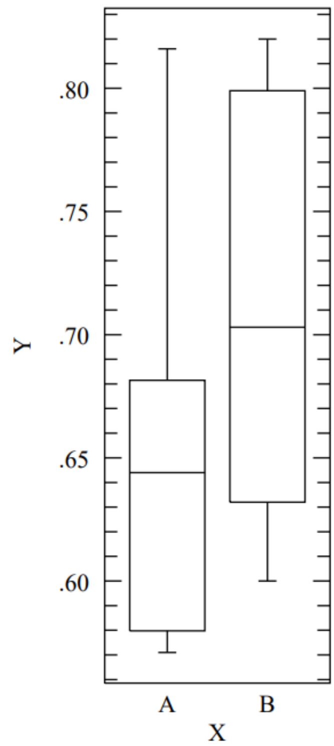

Figure 5.4: The box-plot of the obtained result in the unpaired t-test.

Although the energy-efficient operation of TTDFP is experimentally verified in the previous chapters, it is clarifying to depict the energy consumption of the compared protocols in this chapter again, since there is a difference in the transmitted data such that each leaf node transmits fixed 4000 bits of data to its CH in Chapter 3, whereas in this chapter, the nodes sense and obtain data according to their geared sensor types and the size of data differs for each different type of a node. Additionally, due to our modifications regarding intelligent input parameter tuning for clustering in the quartiles for TTDFP with the purpose of modifying the operation of TTDFP with respect to the evolving environment, it is now more comprehensible to verify and highlight the energy-efficiency aspect of the TTDFP algorithm.

The obtained energy-efficiency results for the experimented study are presented in

Table 5.3. It is wise to state that the energy consumption results are obtained from the standard weighted averaging-based fusion implementation and energy consumption because of fusion is considered insignificant due to the fact that the computation cost is insignificant when compared to the communication cost.

Table 5.3: Obtained Energy Consumption Results of the Empirical Study

Algorithm	**FND**	**HND**	**TRE**
CHEF	87	492	12.03
MOFCA	102	700	20.36
TTDFP	108	861	24.47

The variation in the number of alive nodes as the rounds pass by in the empirical study is depicted in Fig. 5.5. According to the obtained results, TTDFP preserves its efficient-operation architecture under the investigated scenario.

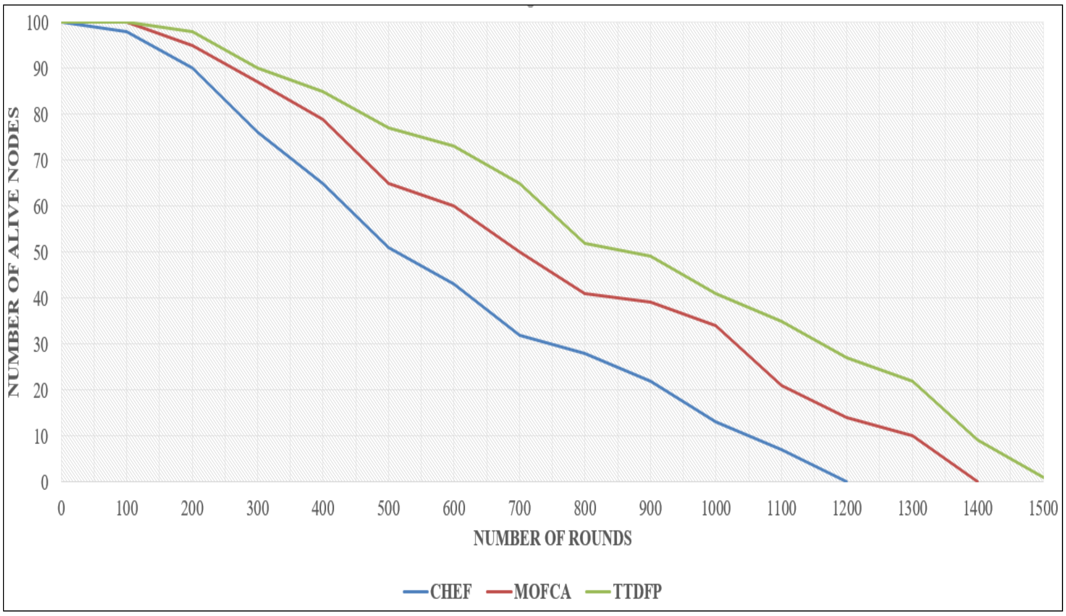

Figure 5.5: The variation in the number of alive nodes as the rounds pass by in the empirical study.

One additional test scenario is setup in accordance with the specifications of the camera (raspicam) employed in the research project. Since the camera sensor is able to record low resolution video (320 x 240 pixels) at 5 fps and each pixel bit depth can be modeled as 8 bits, the transmitted data size for a still image when an object is detected by an imaging sensor can be utilized as 600 KB (614400 bits). Additionally, initial battery charge of a wireless node is modified and utilized as 100 joules in this specific scenario. The reason behind choosing this scenario lies in the real-world usability requirements. TRE consumption of the depicted methodologies under the raspicam usability scenario is delineated in Fig. 5.6. According to the results, all protocols are affected from the increase in transmitted data size, however the pressure over CHEF is the most. MOFCA and TTDFP pursue similar energy consumption models and consumes less energy when compared to CHEF also in the case for raspicam usability scenario.

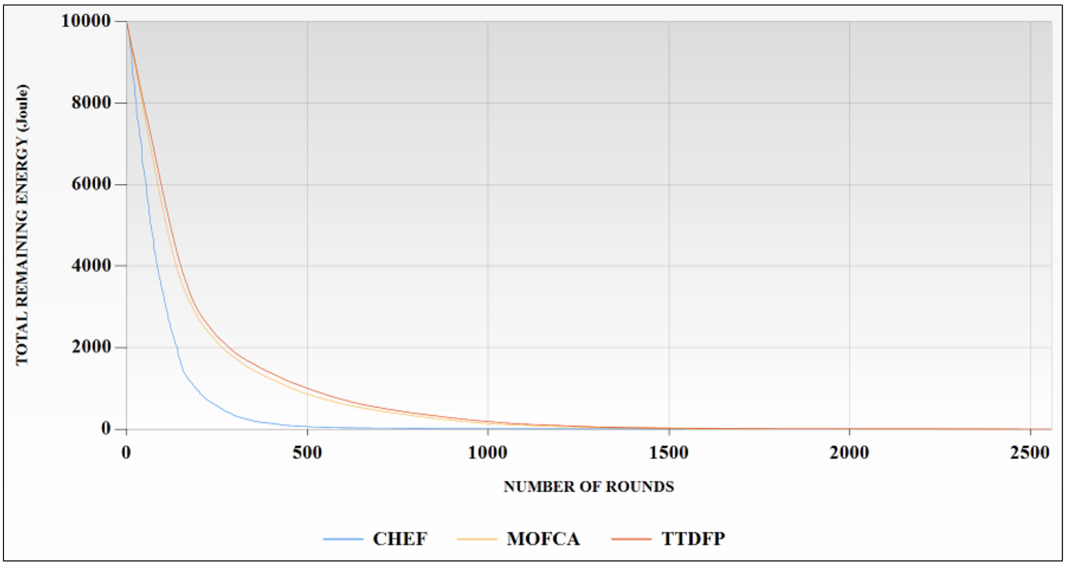

Figure 5.6: TRE consumption of the methodologies under the raspicam scenario.

5.6 Remarks

In this chapter, an empirical study on multi-modal object classification problem in wireless sensor networks is presented in order to corroborate whether or not the proposed intelligent fuzzy clustering methodology maintains object classification accuracy when its compared against selected clustering algorithms. According to the obtained classification performance results, the proposed intelligent fuzzy clustering approach, in addition to its energy-efficient operation architecture, preserves its accuracy when employed in multi-modal object classification scenarios of WMSNs.

It is also noteworthy to state that *the only* difference in the employed scenario is the compared clustering protocols which affect the path from which the data is transmitted. The remaining elements in the architecture, such as fusion and classification methodologies, are exactly the same and occur at the same locations. In this respect, it is valid to compare the impact of clustering algorithms over classification results. Consequently, it is possible to deduce that the change in transmission path directly alters available modalities for fusion at a CH for this experimented scenario, which in turn, modifies the classification outputs.

Additionally, the utilization of the optimization framework, essentials of which is depicted in Chapter 3, in the modality weight assignment phase in the fusion process is a promising choice since it has a significant effect over the classification performance of the constructed system. Since the probability value (p) in the conducted test is obtained as 0.03, it shows that the difference between the weight calculation approaches is not as much as expected. In this respect, if the alpha (α) value is set as 0.01, it cannot be concluded that SA-based approach always improves the result.

CHAPTER 6

CONCLUSIONS

In this dissertation, an Intelligent Fuzzy Clustering Approach for Energy-Efficient Data Aggregation in Wireless Sensor Networks is proposed. The proposed approach is a distribution-agnostic approach that runs and scales efficiently for sensor network applications. Additionally, along with the proposal, an optimization framework is utilized to tune the parameters used in the fuzzy clustering process in order to optimize the performance of a given WSN. This dissertation also includes performance comparisons and experimental evaluations of the proposals with the selected state-of-the-art studies.

The experimental results reveal that the proposed protocol performs better than any of the compared protocols under the same network setup considering metrics used for comparing energy-efficiency and network lifespan of the protocols. At the same time, along with the proposed optimized fuzzy network clustering protocol, an empirical study on multi-modal object classification problem in wireless sensor networks is conducted and obtained classification performance results are also presented so as to corroborate the object classification accuracy aspect of the proposed protocol.

Although the proposed protocol fails in handling specifically engineered scenarios where node deployment occurs manually following a non-uniform distribution on the corners of a rectangular area, which is also the case for MOFCA, it is considered as a low-possibility scenario and a specific utilization aim of deploying such a network has not yet been observed.

6.1 Discussion

The conducted empirical study in this dissertation is based on the research project supervised by The Scientific and Technological Research Council of Turkey under Grant No. 114R082. With the project, a new approach and a framework targeting the energy-efficiency and accuracy trade-off problems in WMSNs are aimed. The focus is on increasing accuracy of transferred information as well as the wireless network energy efficiency. In the project, by using fuzzy clustering algorithms, a wireless sensor network consuming much less energy than currently used is constructed and realized. Thus, a WMSN framework that reduces energy consumption while at the same time preserving accuracy is developed and experimentally verified.

This dissertation starts with the preliminaries about WSNs. These initial sections mostly target an audience of people not accustomed to WSN field. Thereafter, required explanations about application challenges in WSN and multi-modal content together with streaming issues are given. Then, the problem, scope, and contributions of this dissertation are provided.

Then the big picture, background and related work in the scope of the dissertation are presented. This part covers data aggregation concept including clustering and routing aspects, fuzzy logic utilization in WSN applications, and optimization in WSNs together with object classification and related information about multi-modality and data-fusion.

For the energy-efficient clustering problem a new clustering algorithm, named Two-Tier Distributed Fuzzy Logic Based Protocol (TTDFP), is introduced and described in detail. In this sense, an overview of the methodology and system model together with detailed information about the operation architecture of TTDFP is presented. For the energy-efficient routing problem, it is tried to overcome in the designed clustering methodology as an additional tier which extends a crisp routing protocol using fuzzy variables in path determination process.

Thereafter, the designed and proposed approach (TTDFP) is extended using the proposed Clonal Selection principle-based optimization algorithm for fuzzy rule-based clustering algorithms. In this context, detailed descriptions about the proposed fuzzy optimization methodology for rule-based clustering algorithms are also provided.

The discussion of topics until Chapter 5 puts the background into words and provides required knowledge for comprehending the empirical study in this chapter. Although some parts of this chapter are still targeting the readers with restricted background on clustering, optimization and multi-modality topics in WSNs, many important details about the way they collaborate with each other are blended into a more concrete knowledge which is then utilized to corroborate the proposals done throughout this dissertation.

6.2 Future Work

Within the scope of this thesis, there are various topics to be studied on in the future. Those topics can be grouped as;

- The design of a more energy-efficient data aggregation methodology,
- Intelligent operation architecture extensions.

The elaborations related to the aforementioned topics are as follows:

- Although the proposed data aggregation methodology is verified through conducted experiments, there are several other parameters that can be considered for a more-effective and energy-efficient clustering implementation. The blend that the considered parameters fit in the clustering process needs more investigations.
- Additionally, the proposed fuzzy path determination process extends its crisp counterpart with selected fuzzy variables considering energy-efficiency and balance throughout the utilized paths. A similar extension or investigations are also necessary for utilized fuzzy parameters in order to obtain a more effective combination which eventually decreases consumed energy in transmission and balances energy more effectively.
- It is believed that there can be possible improvements and modifications that can be made on the intelligent operation behavior of the proposal. Firstly, in-

stead of SA, other optimization methodologies which do not bring much overhead into a resource-constrained environment in the obtainment of parameter values can be investigated. Then, because of its computational complexity, fuzzy function approximation in this dissertation is conducted in an offline manner. Future studies may focus on devising an online version of this methodology. And finally, the intelligence sugar blended into the architecture only modifies its operation parameters with respect to the change in the number of alive nodes in the environment. There might exist smarter choices of conducting this intelligent behavior which would resemble the way humans behave. These options should also be investigated.

In the same sense, optimization of various algorithm-dependent parameters using some other methodologies such as Particle Swarm Optimization (PSO) can be considered. Additionally, exploring the behavior and testing the performance of TTDFP under WSNs coupled with various node movement strategies can also be investigated.

Learning and classification processes performed throughout the empirical study in this dissertation are executed in accordance with the conducted research project supervised by TUBITAK. Although beyond the scope of this dissertation, these methodologies can be augmented with emerging deep learning approaches for faster and more accurate results.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey", Computer Networks, Vol. 38, No. 4, pp. 393-422, 2002.

[2] I.F. Akyildiz, T. Melodia, and K.R. Chowdury, "Wireless multimedia sensor networks: a survey", IEEE Wireless Communications, Vol. 14, No. 6, pp. 32-39, 2007.

[3] Y. Diao, D. Ganesan, G. Mathur, and P.J. Shenoy, "Rethinking data management for storage-centric sensor networks", Proceedings of the Third Biennial Conference on Innovative Data Systems Research (CIDR), Asilomar, 2007.

[4] S. Misra, M. Reisslein, and G. Xue, "A survey of multimedia streaming in wireless sensor networks", IEEE Communication Surveys & Tutorials, Vol. 10, No. 4, pp. 18-39, 2008.

[5] V.C. Gungor, and G.P. Hancke, "Industrial wireless sensor networks: challenges, design principles, and technical approaches", IEEE Transactions on Industrial Electronics, Vol. 56, No. 10, pp. 4258-4265, 2009.

[6] L. Shu, Z.B. Zhou, M. Hauswirth, D.Le Phuoc, P. Yu, and L. Zhang, "Transmitting streaming data in wireless multimedia sensor networks with holes", Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia, pp. 24-33, Finland, 2007.

[7] N. Roseveare, and B. Natarajan, "A structured approach to optimization of energy harvesting wireless sensor networks", Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC), pp. 420-425, Las Vegas, 2013.

[8] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges", IEEE Net-

work, Vol. 20, No. 3, pp. 20–25, 2006.

[9] C.E. Perkins, and E.M. Royer, "The ad hoc on-demand distance vector protocol", Ad hoc networking, Addison-Wesley Longman Publishing Co., Inc., Boston, USA, pp. 173–219, 2001.

[10] E. Belding-Royer, "Hierarchical routing in ad hoc mobile networks", Wireless Communications and Mobile Computing, Vol. 2, No. 5, pp. 515–532, 2002.

[11] M. Lotfinezhad, and B. Liang, "Effect of partially correlated data on clustering in wireless sensor networks", Proceedings of the IEEE International Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON), pp. 172-181, 2004.

[12] A.A. Abbasi, and M. Younis, "A survey on clustering algorithms for wireless sensor networks", Computer Communications, Vol. 30, pp. 2826–2841, 2007.

[13] Xuxun L, “A survey on clustering routing protocols in wireless sensor networks”, Sensors Journal, Vol. 12, No. 8, pp. 11113-11153, doi:10.3390/s120811113, 2012.

[14] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks”, Proceedings of the 33rd Hawaii International Conference on System Sciences, Vol. 8, Citeseer, pp. 802, 2000.

[15] O. Younis and S. Fahmy, “Distributed clustering in ad hoc sensor networks: a hybrid, energy-efficient approach”, Proceedings of the IEEE 23rd Joint Annual Conference of Computer and Communications Socities (INFOCOM), Hong Kong, Vol. 1, 2004; an extended version appeared in IEEE Transactions Mobile Computing, Vol. 3, No. 4, pp. 366-379, 2004.

[16] I. Gupta, D. Riordan, and S. Sampalli, "Cluster-head election using fuzzy logic for wireless sensor networks", Proceedings of the IEEE 3rd Annual Communication Networks and Services Research Conference, pp. 255–260, Canada, 2005.

[17] J. Kim, S. Park, Y. Han, and T. Chung, "CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks", Proceedings of the IEEE 10th International Conference on Advanced Communication Technology (ICACT), pp. 654–659, 2008.

[18] C.F. Li, M. Ye, G.H. Chen and J. Wu, “An energy-efficient unequal clustering mechanism for wireless sensor networks”, Proceedings of the 2nd IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS), pp. 604-611, Washington DC., 2005.

[19] H. Ali, W. Shahzad, F. Khan, “Energy-efficient clustering in mobile ad-hoc networks using multi-objective particle swarm optimization”, Applied Soft Computing, Vol. 12, pp. 1913-1928, 2012.

[20] S. A. Sert, H. Bagci and A. Yazici, “MOFCA: Multi-objective fuzzy clustering algorithm for wireless sensor networks”, Applied Soft Computing, Vol. 30, pp. 151-165, 2015.

[21] Z. Xu, Y. Yue, and J. Wang, "A Density-based energy-efficient clustering algorithm for wireless sensor networks", International Journal of Future Generation Communication and Networking, Vol. 6, No. 1, pp. 75-85, 2013.

[22] P. Sasikumar and Shilpa, “Dynamic Power Control Clustering Wireless Sensor Networks Based on Multi-Packet Reception”, Indian Journal of Science and Technology, Vol. 9, No. 37, doi:10.17485/ijst/2016/v9i37/102120, 2016.

[23] P. Nayak and A. Devulapalli, “A Fuzzy Logic-Based Clustering Algorithm for WSN to Extend the Network Lifetime”, IEEE Sensors Journal, Vol. 16, No. 1, 2016.

[24] J. N. Al-Karaki and A. E. Kamal, “Routing techniques in wireless sensor networks: a survey", IEEE Wireless Communications, Vol. 11, No. 6, pp. 6-28, 2004.

[25] Y. Wang, “Topology control for wireless sensor networks”, in Wireless Sensor Netw. Appl., pp. 113-147, Chapter 5, 2008.

[26] P. Santi, "Topology control in wireless ad hoc and sensor networks", ACM Computer Surveys, Vol. 37, No. 2, pp. 164–194, 2005.

[27] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: Recent developments and deployment challenges", IEEE Netw., Vol. 20, No. 3, pp. 20–25, May/Jun. 2006.

[28] X. Wang, M. Sheng, M. Liu, D. Zhai, and Y. Zhang, "RESP: A k-connected residual energy-aware topology control algorithm for ad hoc networks", in Proc. IEEE Wireless Commun. Netw. Conf., pp. 1009–1014, 2013.

[29] L. Li, J. Halpern, P. Bahl, Y. Wang, and R. Wattenhofer, "A conebased distributed topology control algorithm for wireless multihop networks", IEEE/ACM Trans. Netw., Vol. 13, No. 1, pp. 147–159, Feb. 2005.

[30] N. Li and J. C. Hou, "FLSS: A fault-tolerant topology control algorithm for wireless networks", *in Proc.* 10th ACM Annu. Int. Conf. Mobile Comput. Netw., pp. 275–286, 2004.

[31] Y. Chen and S. H. Son, "A fault-tolerant topology control in wireless sensor networks", *in Proc.* 3rd ACS/IEEE International Conference on Comput. Syst. Appl., p. 57, 2005.

[32] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks", Wireless Netw., Vol. 8, No. 5, pp. 481–494, 2002.

[33] Y. M. Baryshnikov, E. G. Coffman, and K. J. Kwak. "High performance sleep-wake sensor systems based on cyclic cellular automata", *in Proc.* Int. Conf. Inform. Process. Sensor Netw., pp. 517–526, 2008.

[34] N. Li, J. C. Hou, and L. Sha, "Design and analysis of an MST-based topology control algorithm", IEEE Transactions on Wireless Commun., Vol. 4, No. 3, pp. 1195–1206, May 2005.

[35] X. Li, Y. Wang, and W. Song, "Applications of k-local MST for topology control and broadcasting in wireless ad hoc networks", IEEE Transactions on Parallel

and Distributed Systems, Vol. 15, No. 12, pp. 1057–1069, Dec. 2004.

[36] H. Bagci, I. Korpeoglu, and A. Yazici, “A Distributed Fault-Tolerant Topology Control Algorithm for Heterogeneous Wireless Sensor Networks”, IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 4, pp. 914-923, Apr. 2015.

[37] Online, "Wikipedia Fuzzy Sets", URL:"http://en.wikipedia.org/wiki/Fuzzy_set", Last Accessed: March 2016.

[38] H.J. Zimmermann. Fuzzy Set Theory and Its Applications. Fourth Edition, Kluwer Academic Publishers Group, 2001.

[39] Online, URL:"WikiPedia Optimization", Last Accessed: April 2017.

[40] K. Shin, A. Abraham, and S. Y. Han, “Self Organizing Sensor Networks Using Intelligent Clustering”, International Conference on Computational Science and Its Applications - ICCSA, pp. 40-49, Glasgow, May, 2006.

[41] E. Heidari and A. Movaghar, “Intelligent Clustering in Wireless Sensor Networks”, International Conference on Networks & Communications, pp. 12-17, Dec. 2009.

[42] C. L. Karr, “Design of an Adaptive Fuzzy Controller Using a Genetic Algorithm”, Proceedings of the 4th International Conf. on Genetic Algorithms, pp. 450-457, 1991.

[43] M. A. Lee and H. Takagi, “Integrating Design Stages of Fuzzy Systems Using Genetic Algorithms”, Proceedings of the 2nd IEEE Int. Conf. on Fuzzy Systems, doi:10.1109/FUZZY.1993.327418, San Francisco, 1993.

[44] D. L. Meredith, C. L. Karr and K. Krishna Kumar, “The Use of Genetic Algorithms in The Design of Fuzzy Logic Controllers”, Proceedings of the 3rd Workshop on Neural Network, 1992.

[45] H. D. Cheng and Y. M. Lui, “Automatic Bandwidth Selection of Fuzzy Membership Functions”, Information Sciences, Vol. 103, pp. 1-27, 1997.

[46] C.-C. Yang and N. K. Bose, "Generating Fuzzy Membership Function with Self-Organizing Feature Map", Pattern Recognition Letters, Vol. 27, No. 5, pp. 356-365, 2006.

[47] A. Bagis, "Determining Fuzzy Membership Functions with Tabu Search – an application to control", Fuzzy Sets and Systems, Vol. 139, No. 1, pp. 209-225, 2003.

[48] M. Cerrada, J. Aguilar, E. Colina, and A. Titli, "Dynamical Membership Functions: an Approach for Adaptive Fuzzy Modeling", Fuzzy Sets and Systems, Vol. 152, No. 3, pp. 513–533, 2005.

[49] L. N de Castro and J. I. Timmis. Artificial Immune Systems: A New Computational Intelligence Approach. Springer-Verlag, 357 pages, London, 2002.

[50] L. N de Castro and J. V. Zuben, "Learning and Optimization Using Clonal Selection Principle", IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, Vol. 6, No. 3, pp. 239-251, 2002.

[51] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems", Pattern Recognition, Vol. 38, No. 12, pp. 2270–2285, 2005.

[52] T. Jiang and A-H. Tan, "Learning image-text associations", IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 2, pp. 161–177, 2009.

[53] P. Lamborn, and P.J. Williams, "Data fusion on a distributed heterogeneous sensor network", Proceedings of SPIE 6242, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, 62420R, DOI:10.1117/12.665850, 2006.

[54] J. Kludas, E. Bruno, and S. Marchand-Maillet, "Information fusion in multimedia information retrieval", Proceedings of 5th International Workshop on Adaptive Multimedia Retrieval (AMR), Paris, France, July 5-6, 2007.

[55] N. Poh and J. Kittler. Multimodal Information Fusion: Theory and Applications for Human-Computer Interaction. chapter 8, pages 153–169. Academic Press, 2010.

[56] L. Xu, A. Krzyzak, and C.Y. Suen, “Methods of combining multiple classifiers and their applications to handwriting recognition”, IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, No. 3, pp. 418–435, Jun 1992.

[57] Y. Wu, E. Y. Chang, K. Chen-Chuan Chang, and J. R. Smith, “Optimal multimodal fusion for multimedia data analysis”, Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 572–579, New York, 2004.

[58] L.I. Kuncheva, “Switching between selection and fusion in combining classifiers: an experiment”, IEEE Transactions on Systems, Man, and Cybernetics, Part B:Cybernetics, Vol. 32, No. 2, pp. 146–156, 2002.

[59] F. Souvannavong, B. Merialdo, and B. Huet, “Multi-modal classifier fusion for video shot content retrieval”, Proceedings of 6th Int. Workshop on Image Analysis for Multimedia Interactive Services, Montreux, Apr. 2005.

[60] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, “Early versus late fusion in semantic video analysis”, Proceedings of the 13th Annual ACM International Conference on Multimedia, pp. 399–402, New York, 2005.

[61] M. Guerriero, L. Svensson, and P. Willett, "Bayesian data fusion for distributed target detection in sensor networks", IEEE Transactions on Signal Processing, Vol. 58, No. 6, pp. 3417-3421, 2010.

[62] J. Yue, W. Zhang, W. Xiao, D. Tang, and J. Tang, "A novel cluster-based data fusion algorithm for wireless sensor networks", IEEE 7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), pp. 1-5, Wuhan, 2011.

[63] R. Tang, G. Xing, B. Liu, and J. Wang, "Impact of data fusion on real-time detection in sensor networks", IEEE 30th Symposium on Real-Time Systems (RTSS), pp. 323-332, Washington, DC, 2009.

[64] Y. Chen, J. Shu, S. Zhang, L. Liu, and L. Sun. "Data fusion in wireless sensor networks". Second International Symposium on Electronic Commerce and Security (ISECS), Vol. 2, pp. 504-509, 2009.

[65] Q.I. Tariq, S. Ahmed, and H. Zia, "An objective based classification of aggregation techniques for wireless sensor networks", Emerging Trends and Applications in Information Communication Technologies, Communications in Computer and Information Science, Vol. 281, pp. 512-523, 2012.

[66] C. Zhao, and Y. Wang, "A new classification method on information fusion of wireless sensor networks", Proceedings of the International Conference on Embedded Software and Systems Symposia (ICESS), pp. 231-236, Sichuan, 2008.

[67] A. Alchihabi, A. Dervis, E. Ever and F. Al-Turjman, “A Generic Framework for Optimizing Performance Metrics by Tuning Parameters of Clustering Protocols in WSNs", Wireless Networks, pp. 1-16, doi:10.1007/s11276-018-1665-8, 2018.

[68] R. W. Eglese, “Simulated annealing: a tool for operational research”, European Journal of Operational Research, Vol. 46, No. 3, pp. 271-281, 1990.

[69] Online. URL:"Castalia Simulation Platform", Last Accessed: June 2017.

[70] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing", Science, New Series, Vol. 220, No. 4598, pp. 671-680, 1983.

[71] S. A. Sert, C. Fung, R. George, and A. Yazici, "An Efficient Fuzzy Path Selection Approach to Mitigate Selective Forwarding Attacks in Wireless Sensor Networks", Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, doi:10.1109/FUZZ-IEEE.2017.8015552, July 2017.

[72] A. M. Acilar and A. Arslan, “Optimization of multiple input–output fuzzy membership functions using clonal selection algorithm”, Expert Systems with Applications, Vol. 38, pp. 1374-1381, 2011.

[73] D. Lahat, T. Adali, and C. Jutten, “Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects”, Proceedings of the IEEE, Vol. 103, No. 9, pp. 1449-1477, 2015.

[74] H. Chen and P. Meer, “Robust fusion of uncertain information”, IEEE Transac-

tions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 35, No. 3, pp. 578–586, 2005.

[75] D. M. J. Tax, M. van Breukelen, R. P. W. Duin, and Josef Kittler, “Combining multiple classifiers by averaging or by multiplying?”, Pattern Recognition, Vol. 33, No. 9, pp. 1475–1485, 2000.

[76] E. Bruno and S. Marchand-Maillet, “Multimodal preference aggregation for multimedia information retrieval”, Journal of Multimedia, Vol. 4, No. 5, pp. 321–329, 2009.

[77] A. E. Faghfouri and M.B. Frish, “Robust discrimination of human footsteps using seismic signals”, Proceedings of SPIE 8046, Unattended Ground, Sea, and Air Sensor Technologies and Applications XIII, 80460D, doi:10.1117/12.882726, May 2011.

[78] T. Damarla, A. Mehmood, and J. Sabatier, “Detection of people and animals using non-imaging sensors”, Proceedings of the IEEE 14th International Conference on Information Fusion, Chicago, July 2011.

[79] Q. Pan, J. Wei, H. Cao, N. Li, and H. Liu, “Improved DS acoustic-seismic modality fusion for ground-moving target classification in wireless sensor networks”, Pattern Recognition Letters, Vol. 28, No. 16, pp. 2419–2426, December 2007.

[80] H. Oztarak, T. Yilmaz, K. Akkaya, and A. Yazici, “Efficient and accurate object classification in wireless multimedia sensor networks”, Proceedings of the 21st International Conference on Computer Communications and Networks, August 2012.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name:	Sert, Seyyit Alper
Nationality:	Turkish (TC)
Date and Place of Birth:	1982, Eskisehir
Phone:	0 312 2102080
E-Mail:	alper.sert@ceng.metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
M.Sc	Computer Engineering, METU	2014
B.S.E.E	The Turkish Military Academy	2004

ACADEMIC EXPERIENCE

Date	Place	Enrollment
Sep 2014 - Jun 2018	Dept. of Computer Engineering, METU	Part-Time Instructor
Apr 2015 - Apr 2018	Dept. of Computer Engineering, METU	Project Research Scholar

PROFESSIONAL EXPERIENCE

Date	Place	Enrollment
Oct 2016 - Present	Turkish Army CIS Division	Database Administrator
Mar 2013 - Sep 2016	TRADOC Distance Learning Center	Systems Administrator
July 2010 - Mar 2013	Turkish Army CIS Division	Systems Administrator
Aug 2004 - Jun 2010	Turkish Army CIS Division	CIS Officer

RESEARCH INTERESTS

Wireless sensor networks, algorithms, optimization, and computational intelligence.

PUBLICATIONS

Journal Publications

1. **S. A. Sert**, H. Bagci and A. Yazici. *MOFCA: Multi-objective fuzzy clustering algorithm for wireless sensor networks*. Applied Soft Computing, Vol. 30, pp. 151-165, 2015.

2. T. Dokeroglu, **S. A. Sert** and M. S. Cinar. *Evolutionary Multiobjective Query Workload Optimization of Cloud Data Warehouses*. The Scientific World Journal, Vol. 2014, DOI: 10.1155/2014/435254, 16 Pages, 2014.

Conference Publications

1. **S. A. Sert**, C. Fung, R. George and A. Yazici. *An Efficient Fuzzy Path Selection Approach to Mitigate Selective Forwarding Attacks in Wireless Sensor Networks*. In Proc. of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-6, Naples, 2017.

2. **S. A. Sert**, A. Yazici and A. Cosar. *Effects of selective forwarding attacks on wireless sensor network efficiency*. In Proc. of 25th IEEE Signal Processing and Communications Applications Conference (SIU), pp. 1-4, Antalya, 2017.

3. **S. A. Sert**, A. Yazici and T. Dokeroglu. *Fuzzy processing in surveillance wireless sensor networks*. In Proc. of International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1509-1515, Vancouver, 2016.

4. T. Dokeroglu, M. S. Cinar, **S. A. Sert** and A. Cosar. *Improving Hadoop Hive Query Response Times Through Efficient Virtual Resource Allocation*. In Proc. of the Flexible Query Answering Systems (FQAS), pp. 215-225, 2016.

5. **S. A. Sert**, E. Onur and A. Yazici. *Security attacks and countermeasures in Surveillance Wireless Sensor Networks*. In Proc. of 9th International Conference on Application of Information and Communication Technologies (AICT), pp. 201-205, Rostov-on-Don, 2015.

6. **S. A. Sert**, A. Yazici and A. Cosar. *Impacts of routing attacks on surveillance wireless sensor networks*. In Proc. of the International Conference on Wireless Communications and Mobile Computing (IWCMC), pp. 910-915, Dubrovnik, 2015.

7. **S. A. Sert**, A. Yazici, A. Cosar and C. Yilmazer. *An Efficient Fuzzy Fusion-Based Framework for Surveillance Applications in Wireless Multimedia Sensor Networks*. In Proc. of the International Conference on Wireless Communications and Mobile Computing (IWCMC), pp. 767-772, Cyprus, 2014.

8. **S. A. Sert**, A. Yazici and A. Cosar. *Data Fusion and Processing in Wireless Multimedia Sensor Networks: An Analysis for Surveillance Applications*. In Proc. of 22nd IEEE Signal Processing and Communications Applications Conference (SIU), pp. 421-424, Trabzon, 2014.

9. T. Dokeroglu, **S. A. Sert**, M. S. Cinar and A. Cosar.*Designing Cloud Data Warehouses using Multiobjective Evolutionary Algorithms*. In Proc. of International Conference on Agents and Artificial Intelligence (ICAART), France, 2014.

10. E. Sen, **S. A. Sert** and A. H. Dogru. *Inconsistency Resolution of Feature Trees within Multi-level Feature Modeling*. In Proc. of Society for Design and Process Science (SDPS), Brazil, 2013.