Spring 2015

Design And Implementation Of An Autonomous Wireless Sensor-Based Smart Home

Christopher E. Osiegbu
North Carolina Agricultural and Technical State University

Design and Implementation of an Autonomous Wireless Sensor-based Smart Home

Christopher E. Osiegbu

North Carolina A&T State University

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department: Electrical and Computer Engineering

Major: Electrical Engineering

Major Professor: Dr. Fatemeh Afghah

Greensboro, North Carolina

2015

The Graduate School
North Carolina Agricultural and Technical State University

This is to certify that the Master's Thesis of

Christopher E. Osiegbu

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2015

Approved by:

Dr. Fatemeh Afghah
Major Professor

Dr. Robert Li
Committee Member

Dr. Abdollah Homaifar
Committee Member

Dr. John Kelly
Department Chair

Dr. Sanjiv Sarin
Dean, The Graduate School

Biographical Sketch

Christopher E. Osiegbu earned his Bachelor of Engineering degree in Electrical Electronics Engineering from the University of Port Harcourt in 2008. He joined the Masters of Science Electrical Engineering program at North Carolina A&T State University in 2014. During the interval between his bachelors and master's program, he gained four years working experience as an Electrical Engineer in the Energy (Oil and Gas) and Telecommunication industries. He interned for IBM in the summer of 2014 as a software test Engineer. He is a member of IEEE with certifications as a Cisco certified Network administrator and a Microsoft certified system administrator.

Dedication

I dedicate my thesis to my family and many friends. A special feeling of gratitude to my loving parents, Sir Professor and Lady Osiegbu whose guidance and words of encouragement were vital to my success. My siblings Stella, Charles, Patrick, Stephen and their spouses who have never left my side and are very special. I also dedicate this dissertation to my many friends who have supported me throughout the process. I will always appreciate all they have done. I dedicate this work with a grateful and thankful heart to God, one above all else.

Acknowledgments

I cannot express enough thanks to my committee for their continued support and encouragement: Dr. Fatemeh Afghah, my committee chair, who spent countless hours encouraging and guiding me from the beginning through to the completion of this thesis; Dr. Abdollah Homaifar; and Dr. Robert Li. I offer my sincere appreciation for the learning opportunities provided by my committee. My completion of this project could not have been accomplished without the support of my classmates, especially Seifemichael Amsalu whose machine learning tutorial cannot go unmentioned.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3G	Third generation of mobile telecommunications technology
ABI	Allied Business Intelligence
ADC	Analog to digital converter
BPSK	Binary phase shift keying
CMOS	Complementary metal-oxide semiconductor
CSMA	Carrier sense multiple access
CSMA/CA	Carrier sense multiple access with collision avoidance
DBN-ANN	Dynamic Bayesian network and Artificial neural network
DBN-R	Dynamic Bayesian network and reinforced learning
DSN	Distributed sensor network
DSSS	Direct sequence spread spectrum
EEPROM	Erasable programmable read-only memory
EM	Electromagnetic
FHSS	Frequency-hopping spread spectrum
FSK	Frequency shift keying
GFSK	Gaussian frequency-shift keying
GPRS	General Packet radio service
GSM	Global System for Mobile
IDE	Integrated development environment
IEEE	Institute of Electrical and Electronics Engineers
IOT	Internet of things
Irda	Infrared data association

ISM	The industrial, scientific and medical
M2M	Machine-to-Machine
MAC	Media access control
MATLAB	Matrix laboratory
MCU	Microcontroller Unit
MEMS	Micro electro mechanical systems
NFC	Near field communication
O-QPSK	Offset quadrature phase shift keying
PAN	Personal Access Network
PDA	Personal data access
PIR	Passive infrared
QPSK	Quadrature phase shift keying
RAM	Random access memory
RF	Radio frequency
RFID	Radio frequency identification
ROM	Read-only memory
RSSI	Radio signal strength Indicator
SOSUS	Sound surveillance system
SVM	Support vector machine
UWB	Ultra wide band
WPAN	Wireless personal area network
WSAN	Wireless sensor actuator network
WSN	Wireless sensor network

Abstract

The Smart home has gained widespread attentions due to its flexible integration into everyday life. This next generation of green home system transparently unifies various home appliances, smart sensors and wireless communication technologies. It can integrate diversified physical sensed information and control various consumer home devices, with the support of active sensor networks having both sensor and actuator components. Although smart homes are gaining popularity due to their energy saving and better living benefits, there is no standardized design for smart homes. In this thesis, a smart home design is put forward that can classify and predict the state of the home utilizing historical data of the home. A wireless sensor network was setup in a home to gather and send data to a sink node. The collected data was utilized to train and test a classification model achieving high accuracy with Support Vector Machine (SVM). SVM was further utilized as a predictor of future home states. Based on the data collection, classification and prediction models, a system was designed that can learn, run with minimal human supervision and detect anomalies in a home. The aforementioned attributes make the system an asset for senior care scenarios.

CHAPTER 1

Introduction

Smart home is the term commonly used to describe a home with devices that are capable of communicating with each other and can be controlled remotely by the user. The primary objective of a smart home is to enhance comfort, energy saving, and security for the residents [1].

The smart home is a great asset when it comes to senior care. In 2010 the population of senior citizens (65 and older) globally was 470 million and by 2025, that number is projected to be 820 million [2]. This triggers an effect in all aspects of our society – business, healthcare, policy and technology. As our aging population continues to grow, so does the demand for both in-home care and residential care facilities. This is a niche for the smart home. According to a 2012 survey by Genworth, assisted living, semi-private and private nursing homes cost up to $32,568, $65,160 and $73,800 annually respectively in the US. Projected rise in the elderly population will lead to greater demand and therefore the cost can only increase. In addition to cost saving, the smart home promotes independent living and social interaction.

Still, the long term adoption and commercialization of smart homes is hindered by unanswered questions in terms of reliability, implementation cost, standardization and security. Smart homes need to be reliable so as not to fail when users become reliant on their services. In a home that relays real time patient information to a hospital, system failure can lead to misdiagnosis. Standardization of smart homes will lead to reduction in the cost of implementation. Being that data mining is integral to the learning process of smart homes, data security is also very important.

The United States department of energy stated in 2012 the average person spent $3,052 on energy within the US [3]. This can be reduced by a smart home equipped with motion, glare and luminosity sensors, constantly providing the actuating system with data leading to situational awareness. This can drastically reduce the energy bill by controlling the utilities in the home in an energy efficient manner.

The smart home can provide endless security solutions, depending on the requirements of the user. Common solutions tend to emphasize on intruder alert and gas detection. These are key security requirements given that on the average, about 170 people in the United States die every year from carbon monoxide (a deadly, colorless, odorless, poisonous gas) produced by non-automotive consumer products [4]. These products include malfunctioning fuel-burning appliances such as furnaces, water heaters and room heaters; engine powered equipment such as portable generators; fireplaces; and charcoal that is burned in homes and other enclosed areas. Also, the United States leads the world in burglary cases. About 4 burglaries occur every minute; that's one every 15 seconds. Around 60% of burglars used forcible entry to gain access to homes and 30% entered homes through an unlocked door, window or other opening without resorting to force [5]. With gas sensors, smart locks and passive infrared (PIR) sensors interfaced with a smart control system and efficient situational reactive protocols (which can include contacting emergency services) these statistics can be slashed.

Data gathering is the foundation of every smart home's successful implementation. This is commonly performed through wireless sensor networks (WSN). With the recent advances in micro electro-mechanical systems (MEMS) technology, wireless communications, and digital electronics, the design and development of low-cost, low power, multifunctional sensor nodes that are small in size and can communicate untethered in short distances have become feasible.

In general, the WSNs may consist of different types of sensors such as seismic, magnetic, thermal, visual, infrared, acoustic, and radar. These sensors are spatially distributed to monitor physical or environmental conditions that could include the following temperature, humidity, pressure, speed, pressure, direction, movement, light, soil makeup, noise levels, the presence or absence of certain kinds of objects, and mechanical stress levels on attached objects [6].

Nevertheless, even with reliable data gathering there are design questions that remain regarding the smart home such as (1) how reliable is it (2) how much autonomy will the system have? These are some of the questions that the current research in this field aims to answer. Other researchers have put forward models based on machine learning algorithms, which studied human behavior (routines) and used the data to predict future states in the home. They Postulate that this approach will increase reliability and the autonomy of the smart home system. While these solutions are suited for permanent residents and will be effective after the system has been in place for a period of time, they do not cater to commercial buildings with frequently changing residents.

In contrast, my design takes into account activities within the home regardless of the routine or schedule of the home resident. For example in a smart space that has labeled states such as sleeping, reading, grooming, toileting and entertainment as classes recognized by the smart home, mapping these activities to an actuating system will produce smart reactions.

In this work, a comparative model was used to prompt the retraining of the system, displaying the system’s learning ability in the case a room is re-purposed. This also gives the system commerciality, as it can be deployed in multiple residents and commercial buildings.

To implement this system, a WSN was installed in a one bedroom apartment to gather data. The data was used to train a classification model. With the classification model, the real time state of the home was able to be determined and compared to the home's predicted state. If the two states matched, the output was sent to an actuating system. Otherwise, the real time state was sent to the actuating system. In the case the real-time data is not recognized as a classified state and it surpasses a threshold of occurrence (meaning the state is not an anomaly it is an unclassified state), this data was sent back to the database for retraining of the models.

The smart home is the next generation of building automation, which can easily be interfaced with the emerging Internet of Things (IoT) as briefly described below.

1.1 Internet of Things

The term IoT was initially proposed to refer to uniquely identifiable interoperable connected objects with radio-frequency identification (RFID) technology. The acceptable definition of IoT as a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual 'Things' have identities, physical attributes, and virtual personalities and use intelligent interfaces, and seamlessly are integrated into the information network [7]. The basic idea of this concept is the pervasive presence around us of a variety of things or objects – such as RFID tags, sensors, actuators, mobile phones, etc. –, which through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals [8]. RFID is at the forefront of the technologies driving the vision of the IoT. This is because of the RFID maturity, low cost, and strong support from the business community. However, [8] states that a wide portfolio of device, network, and service technologies will eventually build up the IoT. Near Field Communications (NFC) and Wireless Sensor and Actuator Networks (WSAN) together with RFID are recognized as ‘‘the atomic components that will link the real world with the digital world” [8] .

According to the authors in [9], we will have hundreds of billions of RFID-tagged objects at approximately five cents per tag by 2015. In this type of RFID system, each physical object is accompanied by a rich, globally accessible virtual object that contains both current and historical information on that object's physical properties, origin, ownership, and sensory context (for example, the temperature at which a milk carton is being stored). When ubiquitous and available in real time, this information can dramatically streamline how we manufacture distribute, manage, and recycle our goods. It can also transform the way we perform everyday activities by giving applications current and detailed knowledge about physical events. This "real-life" context can unlock the door to various business, environmental, personal, and social contexts hitherto inaccessible to Internet applications. The incredible amount of information captured by a trillion RFID tags will have a tremendous impact on our lives.

1.2 Benefits of Smart Homes

The smart home presents a vast majority of benefits but the key benefits of the smart home can be categorized as follows:

1.2.1 Convenience

With the touch of a button on a smart home controller or by using voice commands, a scene can be set. Dim lights, TV on and conducive temperature. At night rather than secure your home, this can be done by your home. Using cameras, the home can be monitored remotely. Also, datasets can be monitored to ensure residents did not leave the gas on erroneously and if so an actuator system can shut off the gas before the indoor atmosphere approaches danger levels. Home automation is a tool through which an individual can schedule automatic tasks such as, watering the lawn, switching on the dish washer, activating the security system etc. removing the need to perform these labor-intensive tasks on a regular basis.

1.2.2 Telemedicine

The smart home can serve as a luxury for some users but can also serve as a necessity especially for users in need of health care. Telemedicine can be defined as the use of telecommunication technologies to provide medical services. The key aspect of telemedicine is the use of electronic signals to transfer information from one place to another [10]. The smart home can bridge the gap between doctors and their patients. From the comfort of patient's homes with the aid of sensors or robots equipped with sensors, their vital signs such as body temperature, pulse rate and respiration rate (rate of breathing) can be monitored and even blood pressure measured. Movement of the patient can also be monitored in the case they have been lying still for an unusually long period of time, or may have taken a tumble.

1.2.3 Security

The smart home can present many security advantages. Smart homes include advanced security systems with cameras, motion sensors and a link to the local police station or a private security company. Smart homes may also use key cards or fingerprint identification in place of conventional locks, making it difficult for intruders to break in [11].

1.2.4 Conservation of energy

Smart homes offer enhanced energy-efficiency. Lights can shut off automatically when presence is undetected in a room, and the thermostat can be set to let the indoor temperature drop during the day before returning it to a more comfortable level just before residents arrives in the evening. A smart home can provide detailed real-time data on distributed energy consumption. This information allows the utility operators to improve their system by refining energy distribution. This in turn can help the user save cost on electricity, water and natural gas, and also reduce the strain on natural resources.

1.2.5 Home troubleshooting

A smart home properly equipped with sensors at the right places, can effectively monitor the structure of the house. For example, flow sensors on pipes can monitor water and gas pressure. This will simplify the troubleshooting of loss in pressure. Facilities can also be monitored, and faults anticipated through preventive maintenance. With the appropriate WSAN the electrical mains on the house can be controlled by a smart controller, this will become important in the case of electrical fires.

1.3 Design Objectives of a Smart Home

The challenges involved in the design of a smart home are relative to the complexity of the smart home solution. A smart home system needs to provide information in a suitable manner taking into consideration communication medium, reaction time, ease of understanding of implicit and explicit user actions, identity of different users, privacy concerns, etc. If this is not the case, users may feel inconvenienced and deem the system not worth the trouble [12]. This means when designing a smart home system, the designer needs to take into account the preference of a user and utilize this information in the seamless control of the home. The categories below give more insight to the design objectives of a smart home.

1.3.1 Optimum communication system

From the definition of a smart home, machine-to-machine (M2M) communication is a requirement. In order to achieve a robust network, a heterogeneous communication network will be better suited for smart homes, capitalizing on each protocol’s strength.

For M2M devices, the communication protocol suitable will be the IEEE 802.15.4 (ZigBee/6LoWPAN) protocol. This protocol is well suited for low power/low data rate applications such as heating, ventilating and air conditioning (HVAC) control and appliances.

The IEEE 802.11 (Wi-Fi) protocol is better suited for high data rate applications such as audio and video streaming. Cellular communication (GPRS/3G) is best utilized for applications that need to roam in and out of the home network. The Bluetooth protocol is well suited for low data rate communications such as audio connections and file transfer [13].

1.3.2 User heterogeneity

Users are diverse and inadvertently their preferences on what they deem fit, as the ideal capability of their smart home will differ. This poses challenges in coming up with a basic solution for the average user. For example in telecommunications, there are high tech phones, but the basic feature required of a digital phone taking the world into context is a device which can make calls and download or upload data. The camera pixel, memory, multitasking etc. are all add-ons, which depends on the financial investment the user is willing to make on a phone. Taking into account the different cultures around the world, it is difficult to pinpoint what the basic requirement of a smart home should be. This means every solution implemented will have to be customized to a user's preference and commercially this might not be viable. Hence, a smart home should be able to learn patterns in the home and make decisions as a result of this knowledge. This is where machine learning becomes a requirement.

1.3.3 Learning

The environment present unique challenges to machine learning and decision-making algorithms deployed for smart homes. One is the curse of dimensionality. The state space of an intelligent environment is quite large. For example, if we were to examine a very small environment with 10 motion sensors and 5 lights for a total of 15 binary-state objects, the model would encompass $2^{15} = 32\ 768$ unique states. If we assume each state takes on 0.01 to reason out it would take 5.46 minutes to make a decision [14]. Hierarchical models improve the performance of learning algorithms by facilitating the reuse of portions of the learnt task for new problems.

1.3.4 Cost

When designing a smart home, low cost is a key objective. This can be achieved with WSN in comparison to home care robots. WSN are scalable, cheap and can easily be re-purposed. Standardization in the smart home industry can lead to lower implementation cost. When considering cost it is also wise to consider maintenance cost. Sensors require batteries and the capacity of these batteries determine the life span of the network. A suitable power plan for the sensor nodes will reduce the need for frequent network maintenance. Therefore, relatively low implementation and maintenance costs are key objectives to meet in the design of smart homes before commercialization can be achieved.

1.3.5 Scalability

Smart homes need to be designed with the knowledge that technology is forever developing. Users will require additional features, which means deploying more sensors. Hence, the design of a smart home needs to be adaptive and scalable. There has to be room to add new sensing/surveillance devices to address future applications. This will require additional software reliability. A software upgrade will be less costly than a hardware revamp of an existing system.

1.3.6 Security/Safety

This is a key objective in smart home design. Security refers to data and safety refers to the user's physical well-being. The smart home collects data, controls appliances and utilities. The communication is wireless, these means eavesdropping is a possibility. Therefore, security protocols are integral to the commercialization of smart homes. Using encrypted point to point radio protocols as the personal area network (PAN) communication protocol in the WSN and cellular communication protocol when sending data to a remote site will improve the security of the smart home. The smart home also promises safety as one of its benefits. Distress buttons and safety protocols are options to explore in the design of a smart home. The home sending a distress signal to emergency services when the user experiences heightened pulses and irregular heart rate and rhythm can be beneficial to victims of a heart attack.

1.4 Implementations and Public Usage

Smart homes can be implemented in different ways, due to lack of standardization in the industry there is not an ideal method. For this reason, the public is hesitant in embracing smart home technologies. As studied in [15], reliability, data privacy, and costs of smart home technologies also serve as a barrier in the adoption of smart homes by the public. Currently, less than 1% of homes employed full smart home technology. However, by 2018, HIS Technology, a research firm, predicts that 45 million smart home devices will be installed, and the annual business volume will be grown to $12 billion dollars. ABI Research predicts growth to $14.1 billion by 2018. The market research firm Allied Market Research projects that the global smart homes and buildings market will grow at a compound annual growth rate of 29.5% through 2020, at which point the market will be worth $35.3 billion. Another even more optimistic report from Juniper Research, predicts that the market will grow to $71 billion by 2018 [16].

Several startup and mainstream companies have different implementation of smart homes most of which are on the border of home automation. Common smart home implementations include the use of a central hub or robots to control the home.

1.4.1 Central hub

This is currently the prevalent and affordable implementation of a smart home. In this method a central hub or router (computer) is the brain controlling appliances connected to adapters which are smart (example of such an adapter is the smart outlet, with a smart relay circuitry and RF communication with the hub) the user can connect to the hub remotely using electronic communication device (smart phones, tablets or computers etc.) and switch electronics devices on or off. Users can also monitor other sensor readings such as oxygen and carbon monoxide levels, PIR readings indicating breach of property etc.

1.4.2 Robot

Just as robots are vital in the manufacturing industries, robots are also seen as vital in an alternative implementation of the smart home. The robot assumes the role of the smart home controller. Robots are mostly seen as a necessary smart home implementation when it comes to senior care in order to prevent isolation and loneliness, offering stimulating activities whilst respecting autonomy and independence. For this reason, research in robotics for the smart home is geared towards a robot capable of emotion and memory, this is theoretically investigated in [17]. Researchers at the University of Hertfordshire went a step further and developed a prototype of a social robot which supports independent living for the elderly with emotion and memory [18]. Smart home implementations are covered in chapter 2.

1.5 Problem Statement

As technology evolves the key feature for modern day technology is automation. We all desire our equipment, appliances etc. to have some sort of automation. Why stop there, why can’t our home devices be smart. Learn to assimilate information and produce appropriate results. This is the logical future of the smart home. Presently, there are few prevailing concepts of the smart home due to lack of standardization in the industry there is not one mainstay design of the smart home.

Most present day smart homes implementation has a central node which monitors/controls the homes, this can be as simple as a router or complex as a robot or an electronic device. This architecture leaves the user of this system with a complex graphical user interface to operate in order to get simple tasks done. My proposed system is geared towards the elderly or people with disabilities and taking into the account the technological awareness of this group, this control method is not ideal. This is a niche my system wishes to explore.

In general, the system can be trained to use the data it collects about activities within the home and use this data to control the home.

1.6 Objectives

Smart homes are not standardized consequently there is not a universal system design to build upon. A system model will have to be designed to cater to the goals of my thesis; system autonomy, anomaly detection, minimal human supervision and cost effective. As earlier explained, a relatively cheap method of extracting data from a home environment is through WSN. The data collected needs to be analyzed (labeled) and models need to be trained in order to create classes and predict home states. Therefore the objective of my thesis can be summarized as follows:

- Design a system model.
- Collect data through a WSN setup in a home.
- Analyze the retrieved data.
- Train classification and prediction models.

- Test the trained models.

1.7 Test Bed

In order to achieve the aforementioned objectives, I required equipment to implement a wireless sensor network and a software to analyze my data and produce results.

For the WSN, Waspmotes and Meshlium were used to collect and store data respectively. The Waspmote is a programmable board which when interfaced with a sensor board and sensors can collect data from the environment and could transmit it to a central node or to other Waspmotes. The Meshlium is a Linux router that can serve as a central server in a WSN.

For the data analysis, the Statistics Toolbox of MATLAB was used to implement a machine learning technique called Support Vector Machine (SVM) to classify and predict home states. SVM is a binary classifier which can classify data using two methods. One-vs-One or One-vs-Others. One-vs-One was used in this thesis. Classification enables us determine the state of the home given WSN data at $t = \tau$. For example if given the sensor data of oxygen, carbon monoxide and carbon dioxide we will be able to determine if the home atmosphere is safe or not through a classified label rather than analyzing individual sensor values.

1.8 Summary of Results

This section provides a summary of the results achieved in this thesis. A WSN was set up in a one bedroom apartment with one resident and data was gathered for two months (8 weeks). The data produced from the WSN was classified using SVM and a 99.86% accuracy was achieved. This meant the classification model had a high accuracy in classifying activities in the home according to received data sets. For instance, if the model received data values from sensors it could determine the state of the home.

In order to predict home states, SVM was utilized for regression. Feature vectors were extracted from the historical data of the home. The sequence of the home states were labeled and utilized in order to predict future states. For example, given seven home states and a classification model capturing the state of the home every ten minutes, for two and a half hours, the feature vector will be (11122112231444) capturing fifteen states. Using historical sequence of data, what will be the next state? This is the question answered by the SVM prediction model. When tested against 3 days of real time WSN data, a 96.59% accuracy was achieved.

The WSN, classification and prediction models were utilized in designing a system that could be used as a standard for smart homes.

1.9 Plan of Development

This thesis is organized as follows. Chapter 2, will provide a literature review of the smart home and its building blocks. Chapter 3 will focus on a detailed technical discussion on WSNs. Chapter 4 will discuss system design, the results and concluding remarks are presented in chapter 5 and 6 respectively.

CHAPTER 2

Literature Review

The current trend in smart home research combines the fields of WSN, data analysis and actuator systems. This work is dedicated to designing a system model that can extract data from a home environment, classify this data into states and predict future states in the home. The scope of my research covers WSN, its communication protocols and data analysis which will be done through machine learning. Therefore, in this chapter, smart homes and its building blocks within the scope of this thesis will be reviewed. Starting with a historical overview of smart homes and its current applications leading to a review of related work in the field of WSN and machine learning in smart homes.

2.1 Historical Overview of Smart Homes

The origin of smart homes can be attributed to science fiction writers, they depicted the smart home as a theoretical and visual concept before they became practical, ideally in the early 20th century.

Inventions such as the vacuum cleaner, dryer, washing machine, iron, and toaster, were all created for the purpose making life easier. Ideas similar to modern home automation systems originated during the World's Fairs of the 1930s [19].

The term "smart house" was first coined by the American Association of House builders in 1984 [20]. In 1996 Jim Sutherland developed ECHO IV, it could turn home appliances on and off and control home temperatures. With the invention of the microcontroller, the cost of electronic control fell rapidly. Remote and intelligent control technologies were adopted by the building services industry and appliance manufacturers. By the end of the 1990s, "domotics" was commonly used to describe any system in which informatics and telematics were combined to support activities in the home [21].

In 1998 the INTEGER Millennium House was built in Watford, England partially to showcase a variety of intelligent home automation technologies, including a building management system. The system could optimize the performance of the heating system and automate the garden irrigation system that could sense soil humidity conditions and water accordingly. The management system also boasts of an intelligent security system, lighting that could be set to one of four predefined moods and microchip-embedded programmable door keys [22].

While there is still much room for growth before the commercialization of smart homes, solutions that provide automation of home appliances do exist. In the next section these solutions will be reviewed.

2.2 Smart Home Commercial Implementations

Construction and architectural firms offer custom-built smart home solutions to clients. These solutions are usually expensive however, there are less expensive smart home implementations offered by security, software and electronics firms. Below are some commercial implementations of the smart home.

2.2.1 Data collection and monitoring implementation

This service is already live and presently being marketed by major companies such as AT&T, Comcast, and Time Warner Cable. Video camera(s) and sensors are used to collect data and send to a sink node. The data can be accessed via a web browser or a phone application by the user. The data monitored can include but is not limited to a live video stream, temperature, gas levels and water levels.

Time Warner Cable provides a mobile application or in-home touch screen through which the user can arm or disarm the smart home system, view live video from indoor/outdoor cameras or even set preferred light intensity and indoor temperature. The WSN in the home communicates with Time Warner cable servers via Wi-Fi while the indoor sensors are linked with the controller through Bluetooth and ZigBee. Figure 1 depicts a smart home with different sensors and an actuator system controlled by a monitoring panel.

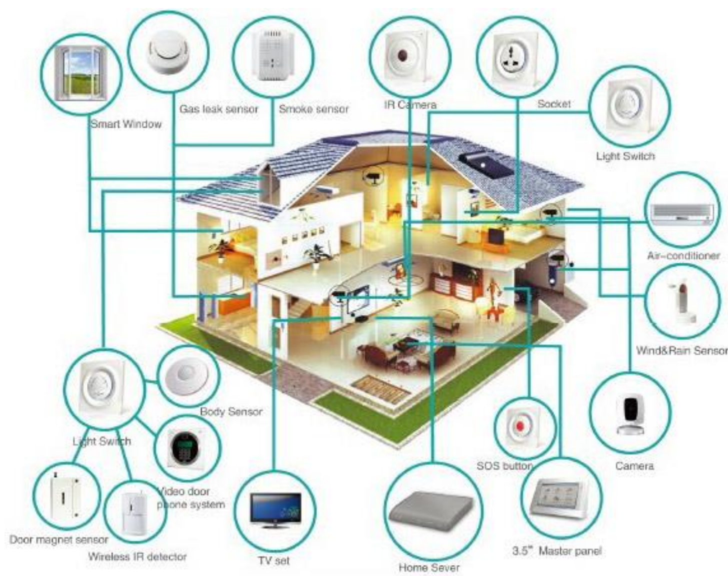

Figure 1 A smart home

2.2.2 Robots

Many experts and futurists predict that in the next several decades, robots will be in every household. Whether these are humanoid or more specifically functional robots, their integration into the smart home of the future is a near-certainty. Robots will likely be fully integrated into the smart home operating system and help manage it, also assisting with manual tasks. Robotics capabilities such as sensing, learning, and adapting are crucial to enhancing the underlying functions of the smart home.

An international research project from Orebro University in Sweden, is rolling out a special eldercare telepresence robot, called GiraffPlus. GiraffPlus uses WSN in the home to monitor the user's health. The sensors can measure blood pressure and body temperature, register movements as well as detect if someone is lying still for an unusually long period of time, or takes a tumble. The key however, is a remote controlled telepresence blue mobile robot, Giraff, which comes with a large display and speakerphone. Through the Giraff robot, caregivers can see, monitor and discuss the health of patients based on the information registered through the sensors by the GiraffPlus system [2]. Figure 2 is a picture of the GiraffPlus.

Figure 2 Giraff Plus

The project undertook in [18] was called Acceptable Robotics Companions for Ageing Years (ACCOMPANY). The project team carried out a wide range of studies which included, detecting the activity and status of people in a smart-home environment as well as focusing on robots' ability to remember and recall. Developments culminated into three interaction scenarios, which were subsequently evaluated by involving elderly people and their formal/informal care givers across France, the Netherlands and the UK.

ACCOMPANY's results demonstrated that a social robot can potentially help to prevent isolation and loneliness, highlighted different important aspects such as empathy, emotion, social intelligence as well as ethics and its norm surrounding technology for independent living [18]

This implementation however, is expensive and can lead to the robot potentially hurting the patient, through trip and fall or by other means. The components that make up the smart home belongs to different fields of research. These components are reviewed in the next section.

2.3 Components of a Smart Home

A typical smart home network consist of the following

- Sensors
- Communication network
- Actuators
- Central Controlling Unit (CCU)

The CCU can contain a built-in database for smart equipment integration, for creating and implementing preprogrammed complex scenarios. The CCU database can store an up-to-date list of all connected devices. Any new integrated device is synchronized upon connection.

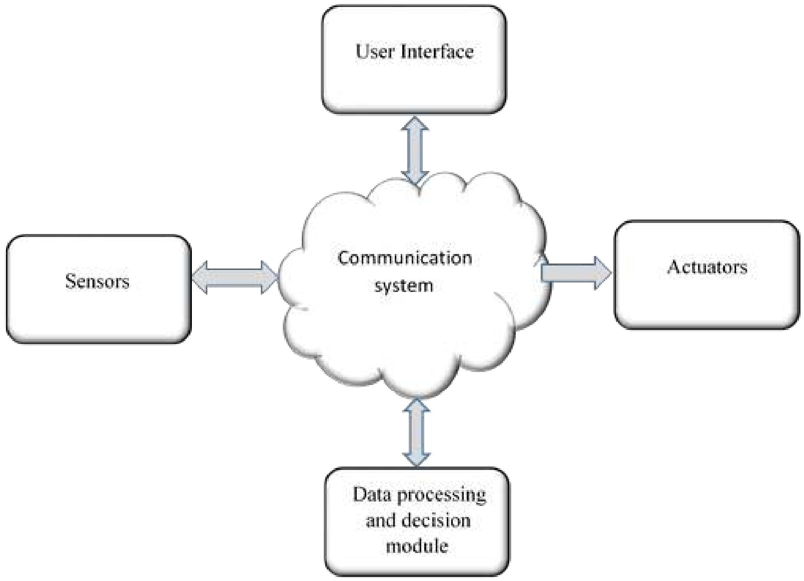

Figure 3 Architecture of a smart home

The architecture in figure 3 is a general illustration that covers different smart home installations. Sensors gather information from the environment and sends it to the data processing module (this can include sink nodes, hubs, computers or artificial intelligence) through a communication medium (wireless or wired). The data processing module can send information back to the sensors or the actuating system. The overall system can be controlled through the user interface.

The scope of this thesis covers the sensor nodes and the communication protocol which make up the WSN, and data processing (classification and prediction). In line with this scope, related works from other researchers will be reviewed in the next section.

2.4 Related Work

2.4.1 Related work on WSN

The authors in [23] addressed the different issues faced when designing ultra-low-power WSN. They focused on the transceiver and took into consideration the low-power and low-voltage constraints. Effects of simple binary modulations such as On-Off-Keying, Frequency-Shift Keying (FSK) or Binary Phase Shift Keying (BPSK) were discussed in this work. These modulation schemes lead to simple analog demodulation without the need for Analog-Digital Converters (ADC) in the receive chain therefore they require less power in contrast to phase modulation schemes such as Offset-Quadrature Phase Shift Keying (O-QPSK). O-QPSK produces higher data rates and improved spectral efficiency with a trade-off in power consumption because they require an ADC in the receive chain.

Zhang et al [24] proposed the effectiveness of different WSN topologies with varying node scales and loads. A small node scale with light loads and disabled acknowledgment (ACK), suited the star topology. When the load and nodes were large, mesh topology was more efficient.

Byun et al [25] proposed a ZigBee based intelligent self-adjusting sensor (ZiSAS) that can autonomously reconfigure middleware, network topology, sensor density and sensing rate based on the environmental situation. In essence, the sensors self-adjustment property saves power as well as bandwidth when the information being sensed is repetitive.

2.4.2 Related work on machine learning

Roy et al [26] utilized information theory and proposed a framework for asymptotic equipartition property to predict inhabitant's routes. Inhabitant's routes were predicted in order to automate device control and proactively reserve resources. For example, electrical energy and other utilities or scarce wireless bandwidth for mobile multimedia applications along most probable locations and routes were reserved for the inhabitant.

In [27] two prediction algorithms namely, dynamic Bayesian network-Artificial neural networks (DBN-ANN) and dynamic Bayesian network-Reinforcement learning (DBN-R) based on the deep learning framework were used to predict various activities in a home. Using home activity datasets, these algorithms were compared to existing methods such as nonlinear SVM and k-means, in terms of prediction accuracy of newly activated sensors. The goal of the research was to predict human behavior in order to proactively provide service to the users.

A prediction algorithm, called Sequence Prediction via Enhanced Episode Discovery (SPEED), was introduced by the authors in [28]. This algorithm was used to predict inhabitant's activity in smart homes. SPEED is a variant of the sequence prediction algorithm. It works with the episodes of smart home events that have been extracted based on the on -off states of home appliances. An episode is a set of sequential user activities that periodically occur in smart homes. The extracted episodes are processed and arranged in a finite-order Markov model. A method based on prediction by partial matching (PPM) algorithm is applied to predict the next activity from the previous history.

Youngblood et al [14] introduced a data-driven approach for building a hierarchical model of inhabitant activities in an intelligent environment. They designed the ProPHeT decision-learning algorithm that learns a strategy for controlling a smart environment based on sensor observation, power line control, and the generated hierarchical model as validation for their hypothesis. Thus demonstrating that data-mining techniques provide a useful mechanism for generating abstract nodes in the hierarchy, and reinforcement learning can be applied to the model in order to learn a policy to control the environment.

Most of the related work in this section studied human behavior in a smart home and used the results to predict activities in the home. The predictions are mostly based on frequency, hence these models are suitable for permanent residents with strict schedules. However, they do not present reliable solutions for homes with spontaneous residents or frequently changing occupants such as hotel guests.

Alternatively, in this thesis the smart home learning problem is studied from a different perspective. Rather than taking individual user actions into consideration, activities which generally go on within the home are taken focused on. Possible activities that can take place in a given space, for instance a room are classified as states. Therefore at a given time WSN $t=\tau$, the state of the home can be determined and at WSN $t = t + \tau$ the state of the home can be predicted. This will be explained further in chapter 4.

CHAPTER 3

Wireless Sensor Networks

Data gathering is the foundation of smart homes. In this thesis a WSN is used for that purpose. Therefore, WSN will be the focus of this chapter. Firstly, a historical overview is presented leading to the current state of WSN, its applications and architectural makeup of the sensing nodes. The challenges of WSN will also be discussed, rounding off with network topologies and communication protocols. Data fusion and mining will be highlighted because they are key to the data analysis module of the system.

3.1 Historical Overview of Wireless Sensor Networks

The origin of WSNs can be seen in military and heavy industrial applications, far removed from the light industrial and consumer WSN applications that are prevalent today. The first wireless network that bore any real resemblance to a modern WSN is the Sound Surveillance System (SOSUS), developed by the United States Military in the 1950s to detect and track Soviet submarines [29]. This network used submerged acoustic sensors – hydrophones – distributed in the Atlantic and Pacific oceans. This sensing technology is still in service today, notwithstanding serving more peaceful functions of monitoring undersea wildlife and volcanic activity.

The United States Defense Advanced Research Projects Agency (DARPA) started the Distributed Sensor Silicon Laboratories, Inc. Rev 1.0 2 Network (DSN) program in 1980 to formally explore the challenges in implementing distributed/wireless sensor networks. Smart Dust was one of the earliest applications of the WSN phenomenon. The main goal of this DARPA funded project was to provide technologies for sensor networks that will be used for military operations in hostile environments. The use of WSNs extend the reach of users to areas where it was too dangerous for humans to operate continuously [30]. With the birth of DSN and its progression into academia through partnerships with universities such as Carnegie Mellon University and the Massachusetts Institute of Technology Lincoln Labs, provided WSN technology with a home in academia and civilian scientific research [29]. Governments and universities eventually began using WSNs in applications such as air quality monitoring, forest fire detection, natural disaster prevention, weather stations and structural monitoring. The military, science/technology and heavy industrial applications of previous decades were all based on bulky, expensive sensors and proprietary networking protocols. Market demanded WSNs placed a premium on functionality and performance, while other factors such as hardware and deployment costs, networking standards, power consumption and scalability fell to the wayside.

Although the technology for large-volume industrial and consumer applications did not exist in the 20th century, both academia and industry recognized the potential for such networks and formed joint efforts to solve the engineering challenges.

Examples of these academic/industrial initiatives include:

- UCLA Wireless Integrated Network Sensors (1993)
- University of California at Berkeley Pico Radio program (1999)
- μ Adaptive Multi-domain Power Aware Sensors program at MIT (2000)
- NASA Sensor Webs (2001)
- ZigBee Alliance (2002)
- Center for Embedded Network Sensing (2002).

The goal of many of these research initiatives was to enable high-volume deployment of WSNs in light industrial and consumer applications by reducing the cost and energy per sensor, while simplifying development and maintenance tasks. Reducing WSN deployment costs while increasing functionality involves major advances in four key technology areas: sensors, CMOS-based semiconductor devices, networking protocols and energy storage/generation technology [29].

3.2 Modern Wireless Sensor Networks

The emergence of the WSN paradigm has triggered extensive research on many aspects of it. WSNs may consist of many different types of sensors including seismic, magnetic, thermal, visual, infrared, acoustic, and radar, which are spatially distributed to monitor physical or environmental conditions that include the following temperature, humidity, pressure, speed, pressure, direction, movement, light, soil makeup, noise levels, the presence or absence of certain kinds of objects, and mechanical stress levels on attached objects [30]. Each node in a WSN is furnished with a radio transceiver, a microprocessor, one or more sensors and an electronic circuit for interfacing the sensors with an energy source, usually a battery or solar panel. Though these sensors have limited data gathering and processing capabilities, through on-board microprocessors, sensor nodes can be programmed to accomplish complex tasks. Transceivers are used in providing wireless connectivity to communicate the observed phenomena of interest to a sink node.

3.3 Wireless Sensor Networks Applications

Wireless sensor networks have the potential to revolutionize communications in a way similar to what we call the Internet of things by offering a wide range of different applications some of which remain to be discovered. Sensor networks are currently being employed and have a huge potential for applications in various fields, including:

- Environment (soil, ocean, atmospheric gases, seismic)
- Health/Fitness (Collecting patient information)

- Management of critical industrial areas: monitoring of oil containers, checking the concentration of chemicals and gases, pharmaceutical process monitoring etc.
- Warehouse management and supply chain monitoring and historical states of the goods with the conditions of critical conservation.
- Military applications: surveillance and recognition.
- Building automation.
- Power (Smart Grids).

A sensor node may vary in size from that of a shoebox down to the size of a grain of dust, although functioning "motes" of genuine microscopic dimensions are yet to be created to my knowledge. The cost of sensor nodes varies from a few to hundreds of dollars based on the complexity and size of individual nodes which results to constraints on resources such as energy, memory, computational speed and communications bandwidth.

3.4 Architecture of a Wireless Sensor Node

A sensor node is a small battery powered device that is capable of processing sensory information and communication with other nodes or a sink node. A general architecture of a sensory node can be seen in figure 4.

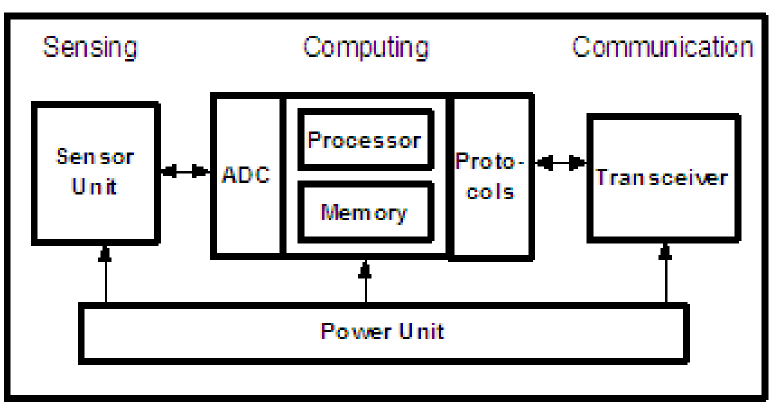

Figure 4 Architecture of a sensor node

A node usually consists of a transceiver, microcontroller, power source, memory unit, and may contain a few sensors. These components will be further explained in subsequent sections.

3.4.1 Transceiver

This is a communication device through which the sensor sends information within the network. The transceivers mostly operate in three modes;

- Transmission
- Receiving
- Sleeping

Power consumption by a transceiver is at its highest during transmission mode, followed by receiving or idle mode, then sleep mode. Transceivers can communicate within the network using RF, infrared, optical, acoustic or magneto-inductive links. The use of communication links depend largely on the purpose of the network. Figure 5 shows an RF transceiver called Xbee, which is used by the nodes in this thesis to send information.

Figure 5 Xbee transceiver

A popular option for transceiver communication is RF. Radio links use the ISM bands, which offer license-free communication in most countries transmitting mostly using 868MHz, 915 MHz and 2.4 GHz.

The main advantages of using the ISM bands are the free radio, huge spectrum allocation, and global availability. Communication in the ISM band is not bound to a particular standard, thereby giving more freedom for the implementation of energy-efficient networking protocols for WSNs. The ISM band is not regulated or assigned to a particular type of user, it can be used by any wireless network. This increases the chance of interference in WSNs, which typically use low-power communication techniques in this spectrum band.

Another possible mode of internode communication in WSN is infrared. Infrared communication is license free and robust to interference from electrical devices. Infrared-based transceivers are cheaper and easier to build. Many of today's laptops, PDAs, and cell phones offer an Infrared Data Association (IrDA) interface. On the other hand, its main drawback is it requires a line of sight between the sender and the receiver. This makes infrared a reluctant choice as the transmission medium [30].

3.4.2 Microcontroller

A microcontroller (sometimes abbreviated µC, uC or MCU) is a small computer on a single integrated circuit containing a processor core, memory, and programmable input/output peripherals. An ATmel microcontroller which is used by the Waspmote is shown in figure 6.

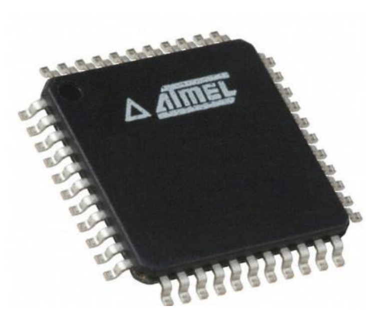

Figure 6 an ATmel Microcontroller

The microcontroller executes instructions, processes data and controls the correct operation of other peripheral devices in the sensor node. A microcontroller is used in many embedded systems including sensor nodes due to low cost, easy programming, ease of interfacing with peripherals and low power consumption. Microcontrollers are designed to perform specific tasks. Specific means applications where the relationship of input and output is defined. Microcontroller has a CPU, in addition with a fixed amount of RAM, ROM and other peripherals all embedded on a single chip.

3.4.3 Sensors

Sensors are physical devices that produce electrical signals in response to physical changes in the surrounding environment. They measure environmental parameters such as temperature, humidity, gas concentration and light intensity. In general, most sensors fall into one of two categories analog and digital sensors. Figure 7 depicts two sensors used in this thesis, a humidity and a PIR sensor that senses relative humidity and presence respectively:

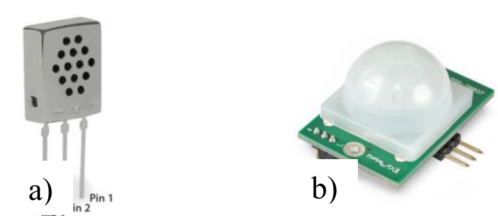

Figure 7 a) Humidity b) PIR sensor

The Humidity sensor is an analog sensor and the PIR sensor is a digital sensor

3.4.3.1 Analog Sensors: They have an output within a voltage range, usually ranging from 0 to 5 volts. Although in latter chapters it will be explained the analog sensors used in this thesis range from 0 – 3.3 Volts. The value of the sensor output can exist anywhere within the range. This output is converted by an analog to digital converter in order for the microcontroller to utilize.

3.4.3.2 Digital Sensors: These on the other hand, generate what is called a 'Discrete Signal'. This means that there is a range of values the sensor can produce as output, but the value must increase in steps. There is a known relationship between any value and the values preceding and following it. The measuring signal is directly converted into a digital signal inside the sensor

3.4.4 Memory

Sensor nodes are also constrained in terms of processing and memory. Most of the sensor nodes rely on on-chip memory inside the microcontroller due to energy concerns. The most popular memories used are flash and Electrically Erasable Programmable Read-Only Memory (EEPROM).

EEPROM is a type of non-volatile memory used in computers and other electronic devices to store small amounts of data that must be saved when power is removed.

Flash memory is an electronic non-volatile computer storage medium that can be electrically erased and reprogrammed. Flash memories are used due to their cost and storage capacity.

Sensor nodes can also have external memory cards such as micro SD cards, these can be used to store data. Memory requirement of a sensor node is entirely application dependent. Memories are used for two purposes: storing application related or personal data and program memory used for programming the sensor node.

3.4.5 Power source

In a sensor node, power is consumed for sensing, data processing and communication. Most of the power is consumed in data communication. The most popular sources of power for sensor nodes are rechargeable batteries. Recently solar panels are being used with rechargeable batteries to eliminate the need for replacing batteries.

3.5 Challenges of Wireless Sensor Network

The major challenge for the proliferation of WSNs is energy [30]. This challenge adversely affects the WSN. Extremely energy-efficient solutions are required for each aspect of WSN design to deliver the potential advantages of the WSN phenomenon. Though energy is the main challenge, WSN still have other challenges such as limited storage and processing capabilities thus each sensor is not equipped to have the whole software/middleware component therefore unable to provide complex services. WSN also have short radio ranges this mandates the development of efficient multi-hop communication protocols, which can also effectively operate in densely deployed sensor networks.

Interference can come from the same network if the underlying medium access technology does not schedule contention-free communications. This is particularly problematic if the two transmitters can hear the receiver, but not hear each other – this is known as the hidden node problem. Interference can also come from another network operating in the same radio space, or from a different radio technology using the same frequency band. The latter, known as "external" interference, is especially present in unlicensed bands such as the 2.400 to 2.485 GHz ISM band, crowded with Wi-Fi, Bluetooth and 802.15.4.

A second phenomenon, multipath fading which is the constructive and deconstructive interference of waves of a single pulse resulting in attenuation of the signal, can prevent a transmitted packet from reaching a receiver. Often described as "self-interference," this occurs when the recipient receives both the signal traveling over the line of-sight path from the transmitter as well as "echoes" of the same signal that have bounced off objects in the environment (floors, ceilings, doors, people, etc.). Since those copies travel different distances, they reach the receiver at different times, potentially interfering destructively. Fades of 20 to 30 dB are not uncommon.

Wireless network bandwidth is scarce relative to node computation, either because the transmission bandwidth is small or because it consumes a large amount of energy to transmit packets on the wireless link, therefore bandwidth management is key. Depending on the service scenarios, sending data from all sensors at regular intervals is an unnecessary operation. For example sending gas data at regular intervals in a home is not efficient. Nodes should only send data when there is a significant change in values.

In conclusion, most of the challenges in employing WSN boil down to energy. The common communication protocol used in WSN is ZigBee due to its energy efficiency, this does not bold well when retrieving real time multimedia data. Hence, heterogeneous communication protocols and sensor nodes with varying sensing capabilities are instrumental in any WSN setup. Energy is most expelled when transmitting so nodes transmitting large packets should be as close to the destination as possible or through efficient multi-hop protocols. The logical arrangement of these nodes is called the network topology.

3.6 Wireless Sensor Network Topology

The topology of WSNs is a crucial element which plays an important role in minimizing various constraints like limited energy, latency, computational resource crisis and quality of communication. The following are commonly used topologies

I. Point to Point Topology

II. Star Topology.

III. Mesh Topology.

IV. Tree Topology.

3.6.1 Point to point Topology

This topology consists of a dedicated long-range, high-capacity wireless link between two sensor nodes. Switched point-to-point topologies are the basic model of conventional networks. Figure 8 shows the point-to-point topology.

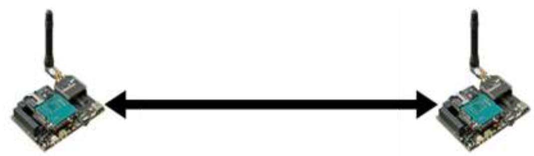

Figure 8 Point to Point topology

This topology was the most common traditional topology used in the wireless sensor networks. The main advantage and disadvantage of this topology is single data communication channel. Resources can be applied to this single channel, hence it can be made secure but failure of this channel through jamming or deep fading will interrupt communication between nodes.

3.6.2 Star topology

Similar to a computer network, star topology as can be seen in figure 9 has multiple communication channels all leading to a central coordinating unit (sink node). The coordinator is responsible for building the network, every message passes through the coordinator. The nodes cannot directly send messages to one another. This is the topology applied in this thesis because having all the data sent to a single database was a requirement in the system design.

Advantages of star topology

- The topology is scalable.
- Low power consumption relative to mesh topologies.

Disadvantage of star topology

- Single point of failure. The network is thrown into chaos if the coordinator goes down.

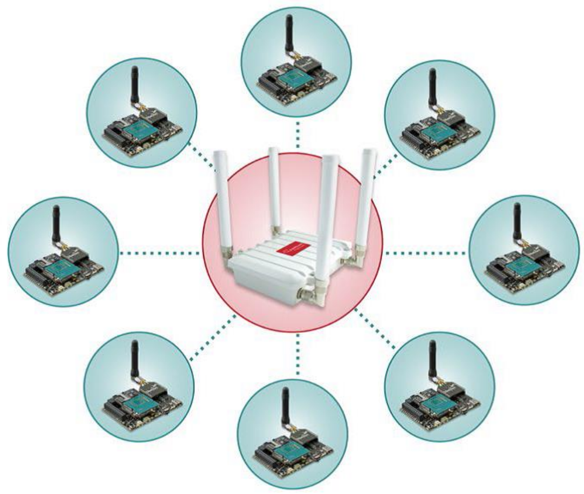

Figure 9 Star topology

3.6.3 Mesh topology

It is a multi-hoping system in which each node can communicate directly with another node based on the proximity of nodes. This can be seen in figure 10, all nodes cooperate in the distribution of data in the network using multi-hoping routing technique. Sent packets hop from node to node until the destination is reached.

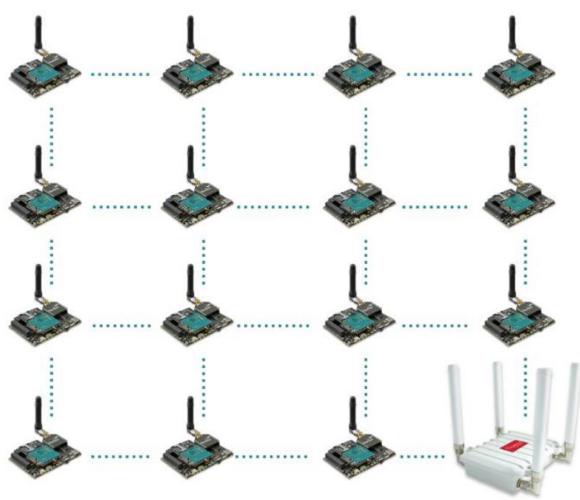

Figure 10 Mesh topology

Advantages of mesh topology

- No single point of failure.

- Alternate paths to destination nodes, which leads to a higher probability of data recovery.

Disadvantages of mesh technology

- Greater power consumption relative to star topology.
- Increased latency.
- Redundant paths.

3.6.4 Tree topology

In this topology, we have two types of nodes; parent and leaf nodes. Data packets are passed from a leaf to its parent node. This is evident in figure 11. The coordinating node only communicates with the parent nodes.

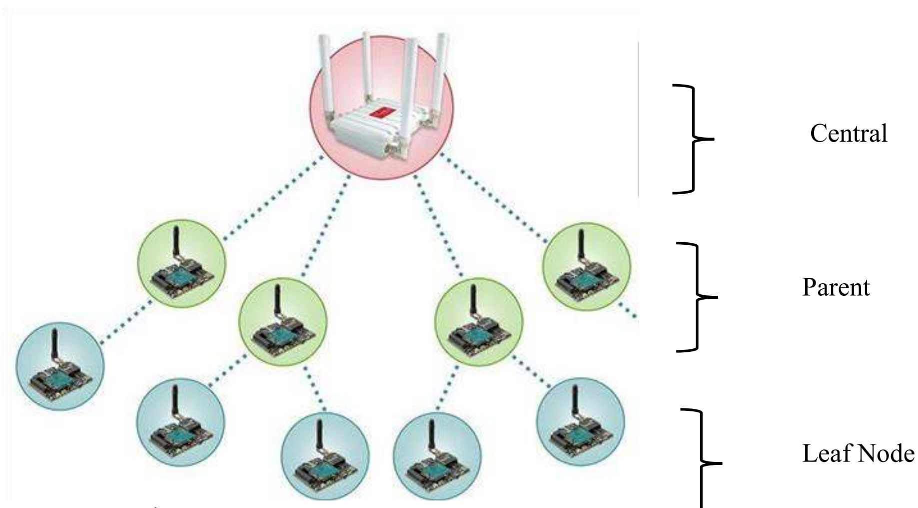

Figure 11 Tree topology

Advantages of tree topology

- Its aggregate power consumption is low relative to star and mesh topologies.
- Relatively easy to troubleshoot

Disadvantages of tree topology

- This topology is time consuming and costly as all cost depends upon the formation of the tree.

- If a parent node fails, then its entire sub-tree is cut off from the network
- The nodes closer to the coordinator consumes a lot of power in forwarding packets from all the nodes in their sub-parts
- Real time delays can occur due to the time taken in sending data from leaf to the coordinator node as data will go through all the parent node of that group.

After the topology has been selected the nodes have to communicate. This is done via communication protocols.

3.7 WSN Communication Protocols

The successful operation of a WSN relies on reliable communication between the nodes in the network. In a multi-hop sensor network, nodes can communicate through a wireless medium creating links between each other. The transceiver of the node is the physical layer component used to exchange data between individual nodes or with the coordinator. Radio frequency-based communication fits the requirements of most wireless sensor applications because it provides relatively long range and high data rates, acceptable error rates at reasonable energy expenditure, and does not require line of sight between sender and receiver.

Most of the current hardware for sensor nodes is based on RF circuit design. RF communication takes place through electromagnetic (EM) waves that are transmitted on the RF bands, which span the 3 Hz to 300GHz spectrum [30]. The 915 MHz and 2.4 GHz industrial, scientific and medical (ISM) band has been widely suggested for sensor networks [30]. There are different kinds of radio frequency communication such as:

I. Zigbee

II. Bluetooth

III. Wi-Fi

IV. GPRS

3.7.1 Zigbee

ZigBee (IEEE 802.15) is a short-distance, two-way wireless communication technology with low complexity, low power consumption, low data rate and low cost [31]. Zigbee consists of link and MAC layer protocols that are compliant with the IEEE 802.15.4 standard, as well as higher layer protocols for ad-hoc networking (mesh, star, or tree topologies), power management, and security. Zigbee supports data rates up to 250 Kbps with Phase shift keying modulation and DSSS.

3.7.2 Bluetooth

Originally standardized as IEEE 802.15.1 protocol, Bluetooth is a wireless technology standard for exchanging data over short distances using short-wavelength UHF radio waves in the ISM band from 2.4 to 2.485 GHz from fixed and mobile devices, and building personal area networks (PANs) [32]. They provides up to 1 Mbps data rate, including three guaranteed low latency voice channels, using Gaussian frequency shift keying (GFSK) modulation and Frequency-hopping spread spectrum (FHSS) as the method of transmitting radio signals. Bluetooth normally transmits at a power of 1 mW with a transmission range of 10 m, although this can be extended to 100 m by increasing the transmit power to 100 mW. Time division is used for channel access, with the master node coordinating the Frequency Hopping sequence and synchronization with the slave nodes [33].

3.7.3 Wi-Fi

Wi-Fi is an 802.11 a/b/g/n IEEE standard protocol that allows electronic devices to exchange data using 2.4 GHz UHF and 5 GHz SHF radio waves. It uses WEP, WPA and WPA2 to encrypt transmitted data. Wi-Fi networks have limited range. A typical wireless access point using 802.11b or 802.11g with a stock antenna might have a range of 35 m (115 ft.) indoors and 100 m (330 ft.) outdoors. IEEE 802.11n, however, can more than double the range. Wi-Fi has fairly high power consumption compared to some other standard technologies such as Bluetooth and ZigBee.

3.7.4 GPRS

General packet radio service (GPRS) is a packet oriented mobile data service on the 2G and 3G cellular communication system's global system for mobile communications (GSM).

GPRS is a best-effort service, implying variable throughput and latency that depend on the number of other users sharing the service concurrently. A maximum data rate of 171.2 Kbps is possible with GPRS when all 8 timeslots of a GSM frame. The data rates of GPRS can be further enhanced through variable-rate modulation and coding [34]

3.8 Sensor Data Fusion

In measuring a particular property of the environment, we often use multiple sensors (both homogenous and heterogeneous) to increase the reliability of the measurement of the property. Multisensory data fusion fuses the output from two or more devices that contain sensor or sensor groups and retrieve one or more particular properties of the environment [35]

This is the combining of sensory data or data derived from sensory data from disparate sources such that the resulting information is in some sense more accurate or complete than would be possible when these sources were used individually. For example, I used multiple PIR sensors and consolidated their data to determine presence in a room. If I used just one, because of the short range of 7 meters, the produced result will not be an accurate representation of the room properties.

3.9 Data Mining

Data mining is the process of transforming knowledge from data format into some other human understandable format like rule, formula, theorem, etc [36]. In order to keep the knowledge unchanged in a data mining process, the knowledge properties should be kept unchanged during a knowledge transformation process. Data mining is a multi-disciplinary field that combines results, for example, from statistics, artificial intelligence, pattern recognition, machine learning, information theory and data visualization. The mined information can be for example correlations, patterns, trends or groups [37]. In this thesis data mining was used to find activity patterns from the WSN data which was then used to build a supervised vector machine (SVM) model.

In data mining supervised and unsupervised techniques are used. In this thesis supervised technique was used. In chapter 4, this topic will be expatiated on.

CHAPTER 4

Methodology

In this chapter, a detailed discussion is provided on the method used to carry out this research. The aim of this work was to design a system model that can serve as a standard for smart homes. In order to design this system, supporting blocks that can aid its understanding of the environment has to be implemented. Therefore, this research work was divided into two parts

- Hardware
- Software

The hardware part involved the collection of data through a WSN, which is shown in figure 12.

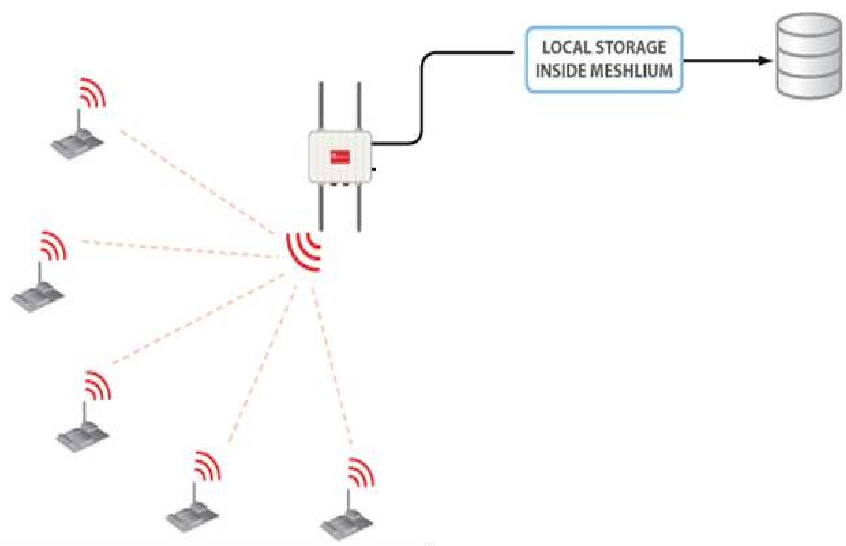

Figure 12 WSN architecture

Figure 12 shows the architecture of the WSN. The Waspmotes are programmed to send data to the Meshlium, which is the sink node.

The software part involved data preprocessing and model training for classification and prediction of home states.

4.1 Test Bed

A WSN test bed was used for data gathering in this research. The Waspmote and Meshlium were used as sensor and sink nodes respectively. These devices are made by Libelium which is an award winning IoT Company based in Spain. They deliver modular, programmable open source sensor platform for the Internet of Things, enabling system integrators to implement reliable Smart Cities and machine-to-machine solutions. Their solutions serve different applications such as smart cities, smart water, smart parking and smart homes, all of which use WSN to collect real time data for varied use.

4.2 Meshlium

Meshlium is a Linux router which contains 6 different radio interfaces. It supports a number of communication protocols (Wi-Fi, 3G/GPRS, Bluetooth, ZigBee, DigiMesh and 802.15.4) at different frequencies (5GHz, 2.4GHz, 900MHz and 868MHz). The Waspmote, equipped with the appropriate transceiver module can communicate with the Meshlium via the aforementioned protocols. A picture of the Meshlium can be seen in figure 13.

Figure 13 the Meshlium

4.3 Waspmote

Waspmote is a modular sensor node device, equipped with an ATmega1281 microcontroller, it supports a number of communication protocols and standards (ZigBee, Bluetooth, Wi-Fi and

GPRS) at varied frequencies (2.4GHz, 900MHz and 868MHz). It counts with a hibernation mode of 0.6uA that allows it to save battery power when it is not transmitting. Waspmote modular architecture enables parts to be plugged in and removed based on the objective of the Waspmote [38]. The Waspmote does not come preloaded with an operating system on its microcontroller, it has a bootloader which loads programmed codes during run time.

The Waspmote communicates with other nodes using the Xbee. Xbee is a microcontroller which uses the Zigbee/Digimesh/802.15.4 protocols. The Xbee is plugged in the radio socket as indicated on the Waspmote board in figure 13.

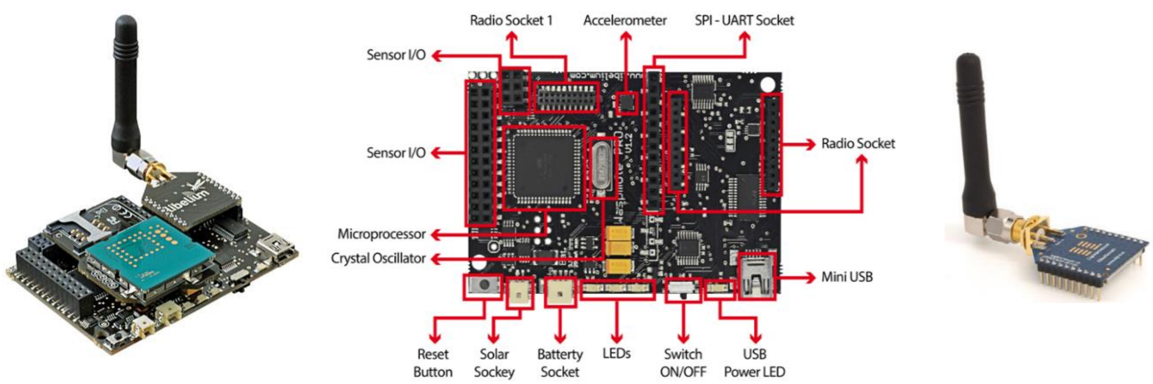

Figure 14 A) Waspmote B) Waspmote board C) Xbee Module

Table 1

Xbee information

Module	Frequency	TX power	Sensitivity	Channels	Bandwidth	Distance
PRO	2.4 – 2.46GHz	63.1mW	-100dBm	12	5MHz	7000m

The frequency used for communication in this thesis, was the free band of 2.4GHz. Channel C (2.405 – 2.410 GHz). The bandwidth was 5MHz.

4.4 Waspmote Programming

When implementing a WSN each node has to be programmed with a source code in order to carry out its task. This programming is done via the Waspmote pro integrated development environment (IDE). The structure of the code is divided into 2 basic functions: setup and loop. The setup is the first part of the code, which is only run once when the Waspmote is initialized or reset. The loop on the other hand just like its name implies, runs continuously forming an infinite loop. The goal of this function is to measure, send the information and save energy by entering a low consumption state (sleep state).

Below is a typical program skeleton.

```
// 1. Include Libraries
// 2. Definitions
// 3. Global variables declaration
void setup()
{
// 4. Modules initialization
}
void loop()
{
// 5. Measure
// 6. Send information
// 7. Sleep Waspmote}
```

The Waspmotes were programmed and spatially placed in a one bedroom apartment to collect data. The schematic of the apartment and can be seen in figure 15.

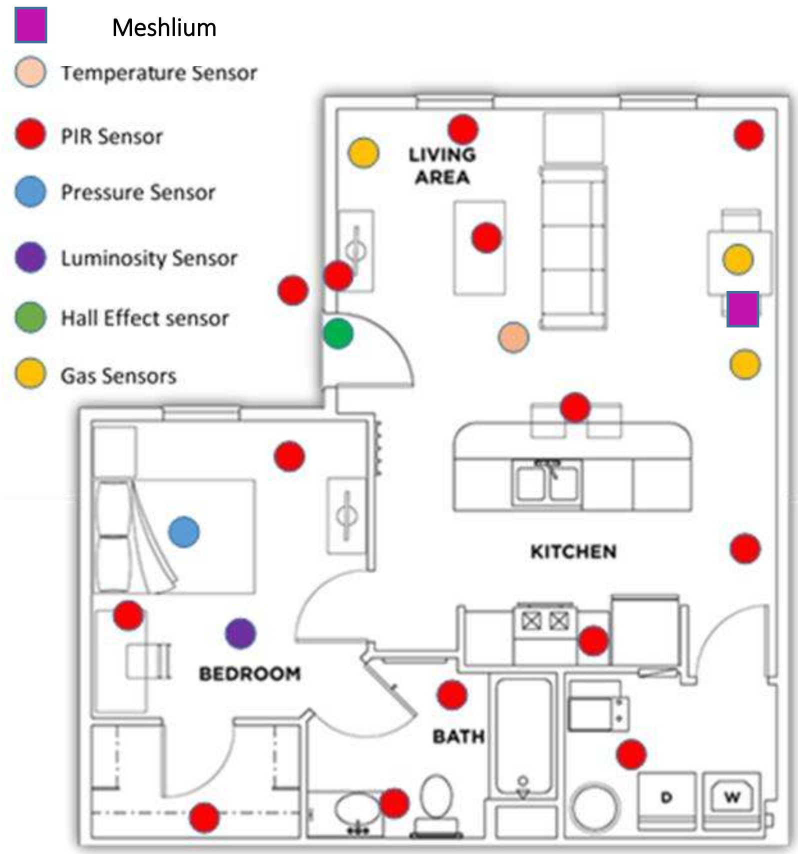

Figure 15 One bedroom floor plan with sensor arrangement

Each sensor was placed on a sensor board and interfaced with a Waspmote node. A measure of the power level received by the Meshlium from Waspmotes were taken at varied distances in the apartment. This measurement of the received power level is called the received signal strength indicator (RSSI). This was done to ensure the RSSI was above a mean level (-82dbm) required for the successful delivery of packets with minimum errors (this was determined through testing with a software called XCTU). If at a certain distance from Meshlium, the RSSI was less than -82dbm, the transmission power had to be increased through the reconfiguration of the RF transceiver (Xbee). However, increased transmission power is detrimental to battery life.

Therefore, nodes closer the Meshlium transmitted with less power. The Xbee has five power parameters/levels, their corresponding power values can be seen in table 2.

Table 2

Power Output level [39]

Parameter	Xbee transmission power
0	-10dBm
1	-6dBm
2	-4dBm
3	-2dBm
4	0dBm

After testing the signal strength of the Waspmote at varied distances from the Meshlium, the power level of the nodes in the Living room were configured to 0, the nodes in the bedroom open area were configured to 2 and nodes in the bath and closet were configured to 3. This meant the nodes were transmitting at 0.1mW, 0.39mW and 0.63mW respectively using the formula below.

$$mW = 10^{\left(\frac{dBm}{10}\right)} \tag{4.0}$$

4.5 Sensors

Sensors are the foundation of smart home architecture because they collect data which are used for monitoring and automation. The sensors are interfaced with the Waspmote board via sensor boards. The Waspmote boards were programmed using the Waspmote pro IDE, to collect data and send it to the Meshlium every ten minutes depending on the situation. The following subsection elaborates on sensors used in this research.

4.5.1 PIR sensor

These sensors are indicated with red dots on figure 15, they are pyroelectric sensors mainly consisting of an infra-red receiver and a focusing lens that bases its operation on the monitoring of the variations in the levels of reception of detected infra-reds, reflecting this movement by setting its output signal high. With a 6-7m range and spectral range of 10µm, these sensors were arranged so as to always sense presence. The PIR sensors best work when placed perpendicular. One PIR sensor was insufficient to sense presence in a large room due to its reliance on line of sight therefore, more than one sensor was required so as not to misrepresent the information in the home. This sensor sent a message to the Meshlium any time it sensed presence.

4.5.2 Temperature sensor (MCP9700A)

This is an analog sensor that converts temperature values into proportional analog voltage. The range of output voltages is between 100mV (-40°F) and 1.75V (257°F) [40]. This sensor sent a message to the Meshlium every 10 minutes.

4.5.3 Hall Effect sensor (PLA41201)

This is a magnetic sensor based on the Hall Effect. This was used to detect the opening and closing of doors. This sensor switch remains closed in the presence of a magnetic field and open in the absence of one. This sensor together with PIR sensor was used to detect the arrival and departure of the resident. It is an interrupt sensor, so it sent information to the Meshlium when its value changed, which was between 1 and 0 indicating open and closed respectively.

4.5.4 Pressure sensor (Flexi force PS-02)

This is a sensitive resistive sensor. Resistance between the two ends of the terminals varies depending on the force exerted on the end of the sensor. This was used to indicate the resident was on the bed. A baseline pressure of 7.2pa was ascertained as the minimum pressure measured when the resident was on the bed. This sensor sent information to the Meshlium every 10 minutes.

4.5.5 Luminosity sensor (LDR)

This is a resistive sensor whose conductivity varies depending on the intensity of light received on its photosensitive part. The measurable spectral range (400nm – 700nm) coincides with the human visible spectrum so it can be used to detect light/ darkness in the same way that a human eye would [40]. This sensor sent its readings to the meshlium every 10 minutes.

4.5.6 Gas sensors

This comprised of the carbon monoxide (CO), humidity, oxygen (O_2) and carbon dioxide (CO_2) sensors. The CO sensor is a resistive sensor sensitive to changes in concentration of CO. This is an important sensor for the home because of the health threat from exposure to high levels of CO. Exposure is particularly detrimental to people with heart diseases [41]. Ideal indoor CO levels are between 0-9ppm indoor [41]. The humidity, O_2 and CO_2 sensors are analog sensors which provides voltage output proportional to the relative humidity, O_2 and CO_2 concentration in the atmosphere respectively. Ideal humidity is generally described to be between 30% and 60% [41]. If seniors or infants are exposed to humidity outside these range for long periods, this can lead to dry skin or respiratory issues [41]. Occupation safety and health administration (OSHA) defines an atmosphere as O_2 deficient if it contains less than 19.5% of O_2 and an atmosphere as O_2 enriched if it contains greater than 22% of O_2. O_2 deficient environments can lead to suffocation and loss of consciousness, while enriched environments can lead to fire and explosion risk. Properly ventilated buildings should have CO_2 levels between 0.1% and 0.6% [41] . When CO_2 levels increase in a home, this leads to a decrease in O_2 which can cause dizziness and loss of consciousness. These gas sensors were adequately calibrated before use. They were not placed too close to the kitchen so as not to indicate false readings during cooking. Sensors are interfaced with the motes using sensor boards. Figure 16 shows the aforementioned sensors.

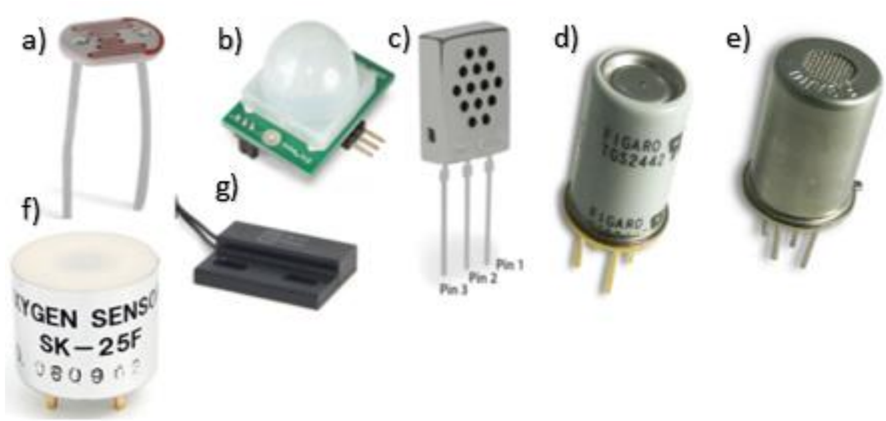

Figure 16 a) Luminosity b) PIR c) Humidity d) CO e) CO_2 f) O_2 g) Hall Effect sensors

The goal of each node was to measure physical phenomenon and send this data encapsulated within a frame to a sink node. In the next section we discuss the frame.

4.6 Waspmote Frame

A frame is a digital data transmission unit in computer networking and telecommunication. It is a data packet in layer two of the OSI model. After the sensors are read and mathematical applications are made such as in the calibration of the gas sensor, a frame is created. The Meshlium recognizes two frame types American Standard Code for Information Interchange (ASCII) and binary. For this thesis I used the ASCII frame due to its simplicity when reading the payload [42].

4.6.1 American Standard Code for Information Interchange

ASCII is a character-encoding scheme. Originally based on the English alphabet, it encodes 128 specified characters into 7-bit binary integers. The structure of the ASCII frame is made up of two parts, the header and the payload.

The header contains information used for synchronization by the Meshlium. The payload contains the sensor values.

Table 3

ASCII frame

Header								**Payload**			
<= = >	0x80	0x03	#	3569028 4	#	NODE_ 001	#	214	#	TEMP	#
A	B	C	D	E	D	F	D	G	D	Sensor1	D

A→ Start Delimiter [3 Bytes]: It is composed by three characters: "<=>". This is a 3-Byte field and it is necessary to identify the start of each frame.

B→ Frame Type Byte [1 Byte]: This field is used to determine the frame type. There are two types of frames: Binary and ASCII. But it also defines the aim of the frame such as event frames or alarm frames.

C →Number of Fields Byte [1 Byte]: This field specifies the number of sensor fields sent in the frame. This helps to calculate the frame length.

D →Separator [1 Byte]: The '#' character defines a separator and it is put before and after each field of the frame.

E → Serial ID [10 Bytes]: This field identifies each Waspmote device uniquely. The serial ID is got from a specific chip integrated in Waspmote that gives a different identifier to each Waspmote device. It is read only.

F → Waspmote ID [0Byte-16Bytes]: This is a string defined by the user which may identify each Waspmote inside the user's network. The field size is variable [from 0 to 16Bytes]. This field can be modified.

G → Frame sequence [1Byte-3Bytes]: This field indicates the number of sequence frame. This counter is 8-bit, so it goes from 0 to 255. However, as it is an ASCII frame, the number is converted to a string so as to be understood. This is the reason the length of this field varies between one and three bytes. Each time the counter reaches the maximum 255, it is reset to 0. This sequence number is used in order to detect loss of frames. [42]

The Waspmote and the Meshlium need to be in the same network in order to send and receive frames. The next section presents the required network credentials.

4.7 Waspmote Networking Overview

In order for a Waspmote to be part of a network, it must synchronize with the coordinating node in this case the Meshlium. It synchronizes with the coordinator by providing network credentials. These credentials are embedded in the source code. Network credentials are placed in the setup part of the source code.

These credentials include:

PAN ID*:* This can be considered a network ID, all nodes in the same network must have the same ID. Depending on the network communication protocol, this ID varies in bit size. 64bit is used for ZigBee and 24bit is used for DigiMesh.

Encryption: Encryption libraries are designed to add to the Waspmote sensor platform the capabilities necessary to protect the information gathered by the sensors. This is done by providing a common pre shared key which is used to encrypt the information using AES 128. Pre shared in the sense that, this key is also entered into the Meshlium management system as a parameter.

Meshlium MAC address: This is a 64 bit destination MAC address. This is the address the unicast packet is sent to from the Waspmote.

4.8 Waspmote Communication

The Waspmote can communicate with the Meshlium using communication protocols such as ZigBee, DigiMesh, Wi-Fi, Bluetooth and GPRS. These are all wireless technology and their signals can be influenced by multipath and shadowing fading

Multipath fading is the constructive and destructive interference of multipath components of a transmitted pulse which attenuates the signal strength.

Shadowing fading is the attenuation of a signal due to physical obstructions in the path of a transmitted signal as it propagates from the transmitter to the receiver.

In order to combat this phenomena as a thumb rule in telecommunications it is wise to measure the signal strength from each node to the receiver and ensure a 10 dB margin above the mean RSSI [43].

4.8.1 Selecting a wireless communication standard

Wireless systems have their limitations and disadvantages. Wireless systems have the following disadvantages

- Security vulnerabilities
- High cost of setting up the infrastructure

- Signals are influenced by physical obstructions, climatic conditions, interference from other wireless devices and jamming.
- Communication range is limited by transmission medium (attenuation)

A wireless communication standard that could mitigate or minimize these disadvantages was needed for the WSN. Radio waves are ideal when large areas need to be covered and obstacles exist in the transmission path. Taking into consideration the disadvantages of wireless communication networks and the need for radio waves, DigiMesh was best suited for the communication purposes of the WSN, due to the following properties:

- Low data rate transmission (250Kps)
- Safety and reliability
- Low costs
- Low power consumption.

4.8.2 DigiMesh

DigiMesh is a proprietary peer to peer wireless communication protocol developed by Digi. It is built on 802.15.4 [44]. This communication protocol offers only one node type therefore there are no parent-child relationships. All nodes can be configured to sleep when there are no events. This is ideal for energy conservation. DigiMesh offers different RF data rate options 900 MHz (10, 125, 150 Kbps), 2.4 GHz (250 Kbps). DigiMesh offers one layer of addressing, a 64bit MAC address. This simplified addressing method improves network setup and troubleshooting.

DigiMesh combats interference by using Direct-Sequence Spread Spectrum (DSSS) at 2.4 GHZ and Frequency-Hopping Spread Spectrum (FHSS) at 900MHz. It can send a payload amounting to 256 bytes. DigiMesh at 2.4GHz was used in this research.

The essence of the Waspmote is to sense physical phenomenon encapsulate its information in a packet and send it to the Meshlium. The process of sending this packet starts with the physical layer.

4.9 Physical Layer

The physical (PHY) layer is responsible for the conversion of bit streams into signals that are best suited for communication across the wireless channel. More specifically, the physical layer is responsible for frequency selection, carrier frequency generation, signal detection, modulation and data encryption. In general, wireless links can be formed by RF, optical, acoustic, or magnetic induction techniques. The communication in smart homes is mostly performed via RF links.

The main types of technologies used for RF communication in WSNs can be classified into three as narrow-band, spread-spectrum, and ultra-wide-band (UWB) techniques. Narrow-band technologies aim to optimize bandwidth efficiency by using M-ary modulation schemes in a narrow band. Spread spectrum and UWB, on the other hand, use a much higher bandwidth and spreads the information onto this higher bandwidth. Spread-spectrum techniques use chip codes of higher rate for spreading the spectrum. On the other hand, UWB accomplishes communication by relative positioning of UWB pulses with respect to a reference time [6]. In this work spread spectrum was utilized to enhance the system performance is the presence of noise and interference as described in the next section.

4.10 Spread Spectrum

Spread-spectrum techniques have recently been used for RF communication to improve the data rate and resistance to interference. There are two types of spread-spectrum techniques: frequency hopping spread spectrum (FHSS) and direct sequence spread spectrum (DSSS) [6].

IEEE 802.15.4 which is the MAC layer of DigiMesh, uses DSSS to modulate the information being sent to the physical layer, this process causes the total information transmitted, to occupy a larger bandwidth. It uses a lower spectral power density for each signal.

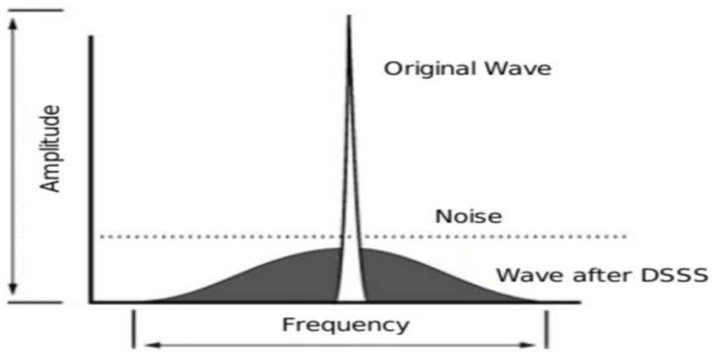

Figure 17 Direct sequence spread spectrum

DSSS is based on pseudo-noise (PN) codes that are called chips. A stream of chips is used to modulate the information bits to be sent. The chips have a lesser duration compared to a bit. Hence, each bit is modulated with a number of chips. Since the chip rate is much higher than the bit rate, the narrow-band information is again spread over a much larger bandwidth. By communicating the sequence of PN codes to the receiver, a DSSS signal can be decoded.

DSSS can provide a better packet error rate if equal bandwidth occupancies are considered for binary modulation. It can be observed that spread-spectrum techniques provide a fine balance between system complexity and interference mitigation compared to other RF communication techniques [30].

There are different DSSS modulation techniques depending on the hardware, physical limits of the circuit and number of symbols which can be processed at a given time [45]. Binary Phase Shift Keying (BPSK), Offset Quadrature Phase Shift Keying (O-QPSK) and Parallel Sequence Spread Spectrum (PSSS) let communication from bandwidths from 20Kb/s to 250Kb/s. The modulation scheme used by Waspmote nodes is offset quadrature phase shift keying (O-QPSK).

4.11 Offset Quadrature Phase Shift Keying

Modulation basically is used to convert information (digital bit stream or analog audio signal) in the form that can be physically transmitted over the air or cable.

$$S(t) = A(t)cos\theta(t) \tag{4.1}$$

$$S(t) = A(t)\cos[\,\omega_0 t + \varphi(t)] \tag{4.2}$$

ω_0 – Carrier frequency

φ – Phase

O-QPSK modulation scheme has been adopted by the IEEE 802.15.4 standard to modulate the chips sent for each bit as a part of the DSSS scheme. O-QPSK is a digital modulating scheme with a maximum phase shift of about +/- 90 degrees. In O-QPSK, after splitting the bit stream into odd and even, one bit stream is made offset by 1 bit period with respect to the other. After this, the direct and shifted bit streams are fed to the mixers.

The modulation structure shown in Figure 18, consists of the conversion of bits to symbols, conversion of symbols to chips and O-QPSK modulation of the chips. In the IEEE 802.15.4 standard, a byte is represented by two symbols with 4 bits per symbol, which results in a 16-ary modulation. Each 16 symbol is represented by a combination of 32 chips. Then, each chip is transmitted using O-QPSK modulation.

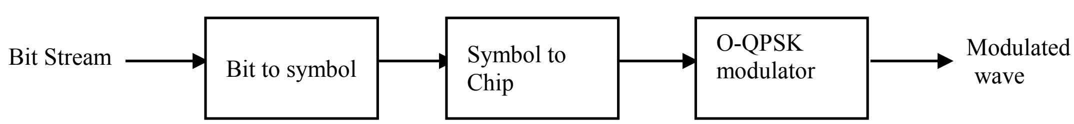

Figure 18 The modulation structure in IEEE 802.15.4 *[6]*

The device responsible for the modulation operation is called the modulator and the resulting waveform or output of modulation is called a modulated wave. The bit error probability of O-QPSK is given by [46]

$$p_b^{OQPSK} = Q\sqrt{({}^{E_b}/_{N_0})_{DS}} \tag{4.3}$$

Where

$$({}^{E_b}/_{N_0})_{DS} = \frac{2N\,x\,E_b/N_0}{N+4(E_b/N_0)(K-1)/3} \tag{4.4}$$

Where N is the number of chips per bit and K is the number of simultaneously transmitting users. After the data has been converted to bits and modulated the data is transmitted over a wireless communication channel

4.12 Node Transmission

The wireless channel exhibits a broadcast nature, were the transmission of a sensor node can be received by multiple sensor nodes. This nature results in each sensor node sharing the wireless channel with the nodes in its transmission range. Since the communication has to be performed through this broadcast wireless channel, the design of Medium Access Control (MAC) protocols is of crucial importance in WSNs. Most of the MAC protocols proposed for WSNs, rely on a conventional medium access scheme called carrier sense multiple access (CSMA) [30].

IEEE 802.15.4 uses two techniques to avoid all nodes emitting at the same time Carrier Sense Multiple Access-Collision Avoidance (CSMA-CA) and Guarantee Time Slots (GTS). The Waspmote uses CSMA-CA.

4.12.1 Carrier sense multiple access-collision avoidance

CSMA is a listen-before-transmit method. The Waspmote uses CSMA-CA method due to the fact that it is inherent to each mote and does not depend on a central server. Each node listens to the channel prior to transmitting. If the energy found is higher than a specific level, the node waits during a random time and tries again. The amount of time specified to scan each channel may affect the detected energy therefore selecting the minimum energy value maybe insufficient in detecting channel energy, so a higher value is recommended. Usually, energy value for a free channel is around -84dBm and -37dBm for an occupied one [39]. In this thesis, using the XCTU software to configure the transceiver, the clear channel assessment threshold value for a free channel was configured to -44dBm.

4.12.2 Routing

The routing protocol used by DigiMesh is Ad-hoc On-demand distance vector algorithm (AODV). This is a pure on-demand route acquisition algorithm. The AODV Routing Protocol uses an on-demand approach for finding routes. This means a route is established only when it is required by a source node for transmitting data packets. Using this method reduces routing overhead.

4.13 System Software

This section expatiates on the software part of the system. The software used was MATLAB. Data mining and machine learning will be covered. General definitions and classifications of machine learning systems are presented, with a detailed look at SVM, which was used in this thesis.

4.14 Data Mining

Learning techniques have a significant role in data mining. Understanding climate change, finding alternative energy sources and preserving the health of an ageing population are all cross-disciplinary problems that require high-performance data storage, smart analytics, transmission and mining to solve [47].

In this thesis, data mining was used to find activity patterns from the WSN data. The data was then used as inputs to the SVM classification and prediction models. In general data mining techniques can be divided into supervised, unsupervised and semi-supervised techniques.

4.14.1 Data mining techniques

4.14.1.1 Supervised (e.g., discriminant analysis) is a technique in which the input pattern is identified as a member of a predefined class. In supervised technique, the model defines the effect one set of observations called inputs, has on another set of observations called outputs. The goal is to learn a mapping from learning x to y, given a training set made of pairs (x_i, y_i).

4.14.1.2 Unsupervised (e.g., clustering) is a technique in which the pattern is assigned to a hitherto unknown class. This is a learning method in which training samples are unlabeled. No information about the input is given to the model thus the system cannot know the correctness of the outcome during the training. The goal of unsupervised learning is to find interesting structure in the data X.

Note that the recognition problem here is being posed as a classification or categorization task, where the classes are either defined by the system designer (in supervised classification) or are learned based on the similarity of patterns (in unsupervised classification) [18].

4.14.1.3 Semi-supervised learning (SSL) is halfway between supervised and unsupervised classification. In addition to unlabeled data, the algorithm is provided with some supervision information but not necessarily for all examples. Hence new patterns are discovered. In this case, the data of SSL set $X = (x_i)_{i\in[n]}$ can be divided into two parts: the points $X_l = (X_1, . . . , X_l)$, for which labels $Y_l = (Y_1, . . . Y_l)$ are provided, and the points $X_u := (X_{l+1}, . . . , X_{l+u})$, the labels of which are not known.

Larger and more complex dataset are better learnt using unsupervised learning relative to supervised learning. This however, was not suited to my research at this stage.

14.15 Data Labeling

Approximately, eight weeks of data was collected. The labeling exercise captured seven (7) states in the home. When labeling the data, safety was considered integral to the smart home design. Consequently, the gas levels were taken into consideration when labeling the home states, except for the arrival and departure states. The labeled states are listed and explained below.

1. Safe sleeping (SS): The resident was on the bed sleeping and the gas levels were within acceptable levels.
2. Safe Away (SA): The resident was away from the home and gas levels were within adequate levels.
3. Safe Home Activity (SH): The resident was active in the home and the gas levels were adequate.
4. Alarm (AL): The gas levels exceeded recommended indoor levels but were not up to emergency levels but further exposure may have been dangerous.
5. Emergency (EMR): The gas levels have reached emergency levels.
6. Departure (DP) : This state indicated the departure of the resident
7. Arrival (AR): This state indicated the arrival of the resident.

The rules guiding the labeling can be found in table 3.

Table 4

Classification table

Sensor	SS	SH	SA	AR	DP	AL	EMR
O2 (%)	>20 & <25	>20 & <25	>20 & <25	-	-	>8, ≤ 17 & >25, <30	≤ 8 & >30
CO2 (%)	≥0.03 & ≤0.1	≥0.03 & ≤0.1	≥0.03 & ≤0.1	-	-	>0.1 & <0.15	≥0.15
CO (ppm)	≥0 & ≤9	≥0 & ≤9	≥0 & ≤9	-	-	≥9 & ≤15	>15
Humidity (%RH)	≥30 & ≤60	≥30 & ≤60	≥30 & ≤60	-	-	≤20 & >60	≤10 & ≥70
PIR Living room	-	1	0	1	0	1	1
PIR Bedroom	1	-	0	-	0	-	-
PIR Outside	-	-	0	1	1	-	-
Luminosi ty	≥300& ≤600	≥2 & ≤600	-	-		-	-
Free Fall	-	-	-	-		-	1
Pressure (Pa)	>7	<7	-	-		-	-
Hall effect	0	0	0	1	1	-	-

Table 4 can be read as follows, using the SH class as an example. To classify a data-set as SH the following conditions apply.

1. The O_2 level will have to be greater than 20% but less 25%,
2. The CO_2 level will have to be equal to or between 0.03% and 0.1%
3. The CO level will have to be equal to or between 0 and 9ppm,
4. The relative humidity will have to be equal or between 30 and 60%

5. The PIR living room (PIRL) will have to be equal to 1
6. The luminosity (LUM) will have to be equal to or between 2 and 60
7. The pressure will have to be less than 7 Pa
8. The Hall Effect value will have to be equal to zero.

An alternate view of the sensors that were instrumental in state classification is shown in figure 19.

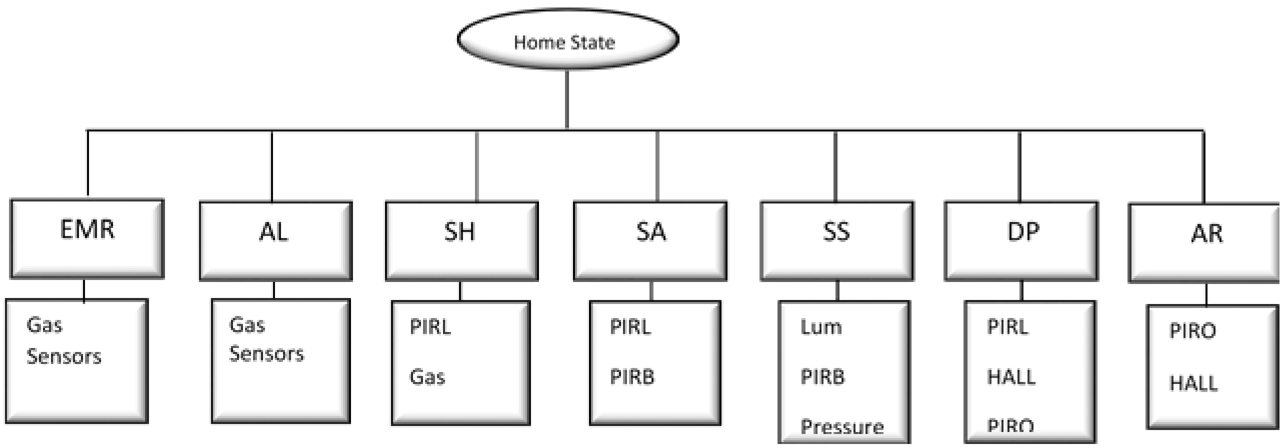

Figure 19 Labels make-up sensors

After labeling, the data was ready to be used to train a classification model. This was necessary so that at any given time, presented with sensor values, the model can determine the state of the home. Supervised machine learning was used to carry out this classification.

4.16 Supervised Machine Learning

Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data. Such algorithms operate by building a model from example inputs and using them to make predictions or decisions [48].

Supervised learning can be applied using machine learning techniques such as Naive Bayes, Decision trees, Support vector machines (SVM), artificial neural networks and Hidden Markov models [49]. SVM is utilized in this research due to its high performance in many applications such as prediction, handwritten recognition, face detection, text categorization, and speech recognition. A comprehensive survey on the applications of the SVMs in pattern recognition is presented in [50]. In most of these applications SVM outperforms other machine learning techniques in generalizing the pattern recognition problem.

4.16.1 Support vector machine

In machine learning, support vector machines (SVM) are supervised learning models with associated learning algorithms that analyse data and recognize patterns, used for classification and regression analysis [51].

SVMs are linear discriminant functions that map the feature space to a higher dimension. Training an SVM causes it to find the optimal hyper plane which has the maximum distance from the nearest training patterns, called support vectors. This is expected to give more generalization capabilities to the SVM [49]. SVM can perform linear and nonlinear classifications.

A basic classification of linear SVM finds a linear boundary by taking a set of input data and predicts two possible classes from the input. This is depicted in figure 20.

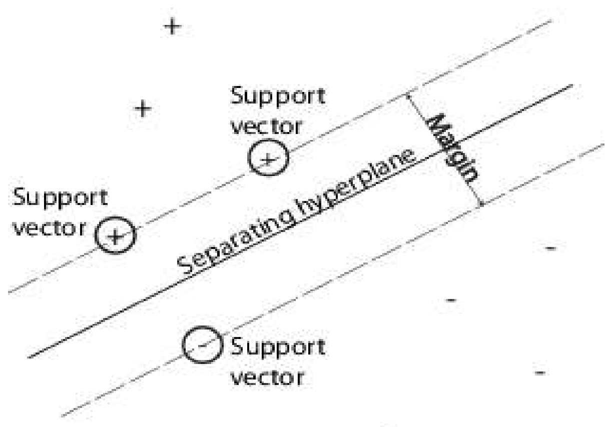

Figure 20 SVM linear classification

When faced with a nonlinear classification, SVM uses kernel functions to separate the input data into two possible classes by mapping input data into high dimensional feature spaces as shown in figure 21.

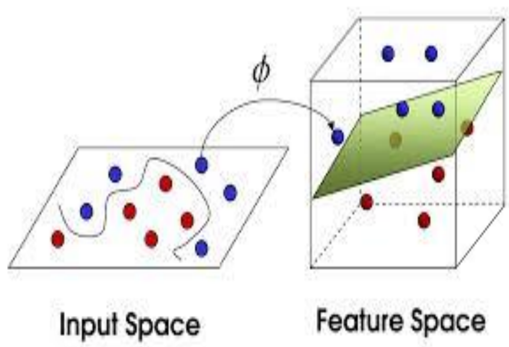

Figure 21 Nonlinear SVM classification

Training an SVM requires choosing the functions that map the input to a higher dimensional space and using, for example, quadratic programming optimization techniques to find the optimal hyper-plane. The functions are often chosen using the designer's knowledge of the problem domain. Training an SVM is quite efficient and they can represent complex nonlinear functions [49].

4.17 Classification

The smart home problem is a multi-class classification problem (Safe-Sleeping, Safe Home Activity, Safe Away, Alarm Home, Emergency, Departure and Arrival), the SVM cannot be applied directly. There are different methods to apply SVM for multi-class classification. The most common ones are one-vs.-one and one-vs.-the-rest [52].

The one-vs.-one trains a classifier for each possible pair of classes. For M classes, this results in $(M-1)M/2$ binary classifiers. When SVM tries to classify a test pattern, it evaluates all the binary classifiers, and classifies according to which of the classes gets the highest number of votes.

The one-vs.-the-rest gets M-class classifiers, it constructs a set of M binary classifiers each trained to separate one class from the rest, and combines them by performing the multi-class classification according to the maximal output. In this work, one-vs-one multi-class SVM classification method is used to estimate the state of the smart home.

The processed features of the sensor data at $t = \tau$ are presented to the multi-class SVM classifier. In the classification stage, the dataset (the feature vectors with their corresponding labels) is divided into 70% training and 30% testing set. The training set is used to build the Multi-class SVM classifier and the testing set is used to evaluate the classifier's performance. The flowchart of the pattern recognition algorithm based on Multi-class SVM is shown in figure 22.

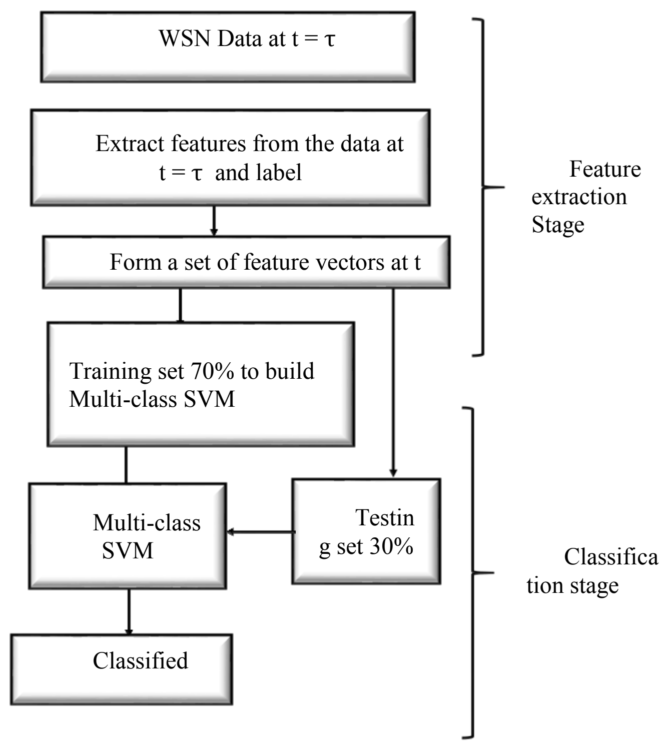

Figure 22 Classification flow chart

4.18 Prediction of States

The classification model has be trained and can now classify home states. The next phase of this work involves the prediction of future states in the home. At this stage because the data was already classified, what was being predicted was the next state in the home at a particular time not individual sensor readings. For instance, a week or three hours from a given time what will be the state of the home? The EMR and AL states were anomalies because they were not regular occurrences in the home. Therefore the data was normalized by switching those labels to the previous label before their occurrence. For example, if the state of the home was SS before AL occurred, the label was returned to SS.

SVM was utilized for regression, extracting and labeling sequential home states from the historical data. The time duration of a sequence was varied from weeks to hours, eventually a two and a half hour window was used. As earlier mentioned, the state of the home was captured every ten minutes, therefore a two and a half hour window will consist of fifteen determined states. In order to determine a given state at time t = t + τ, the previous 14 states were used.

$$f(t) = [D(t-l), D[t-l+1] \ldots \ldots \ldots D(t-l+l)] \quad (4.5)$$

The feature vector $f(t)$ contains past states, with l being the number extracted features within a window, in this case fifteen. For a given sequence SVM provides a label;

$$L(t) = D(t+1) \quad (4.6)$$

For instance, given five home states and a classification model capturing the state of the home every ten minutes, for two and a half hours, the feature vector will be [11122112231444] capturing fifteen states. Using historical sequence of data, what will be the next state, if in the future we have [11122211223144]? This is the question answered by the SVM prediction model, as it will predict state 4.

The prediction flow chart bears similarities to the classification flow chart with notable changes as shown in figure 23.

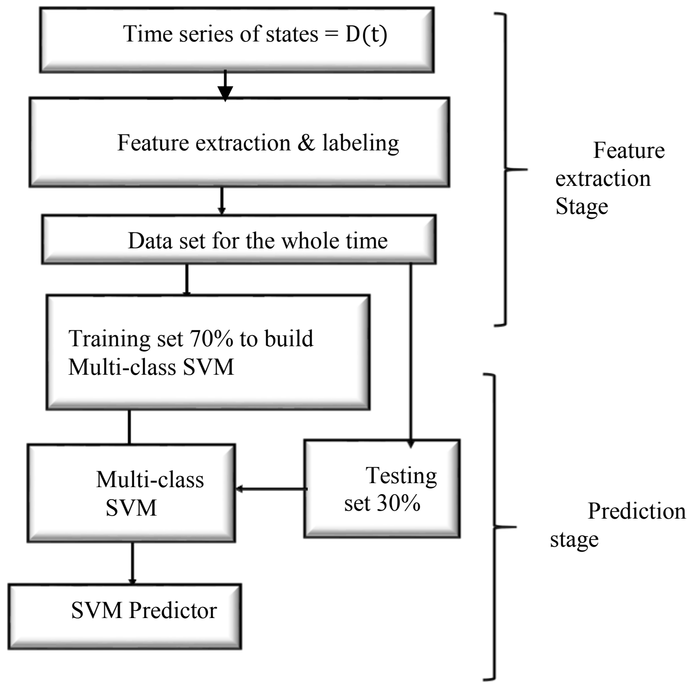

Figure 23 Predictor flow chart

After the feature extraction and labeling, the whole data set is used to train the multiclass SVM, and 30 percent of the data is used to test the multiclass SVM, as shown in figure 23.

CHAPTER 5

Results

In this chapter, results attained from the implemented test-bed and data analysis are presented and analyzed. Finally, the designed system is presented.

5.1 Data Gathering

The test-bed was a one bedroom apartment building equipped with thirteen PIR, one temperature, one luminosity, one pressure, one of each gas and one Hall Effect sensor. The data was collected for two months, approximately eight weeks. When setting up the WSN, as earlier discussed the RSSI was measure at varied distances from the Meshlium. Figure 24 and figure 25 show the closest and farthest nodes from the Meshlium respectively and their range test results.

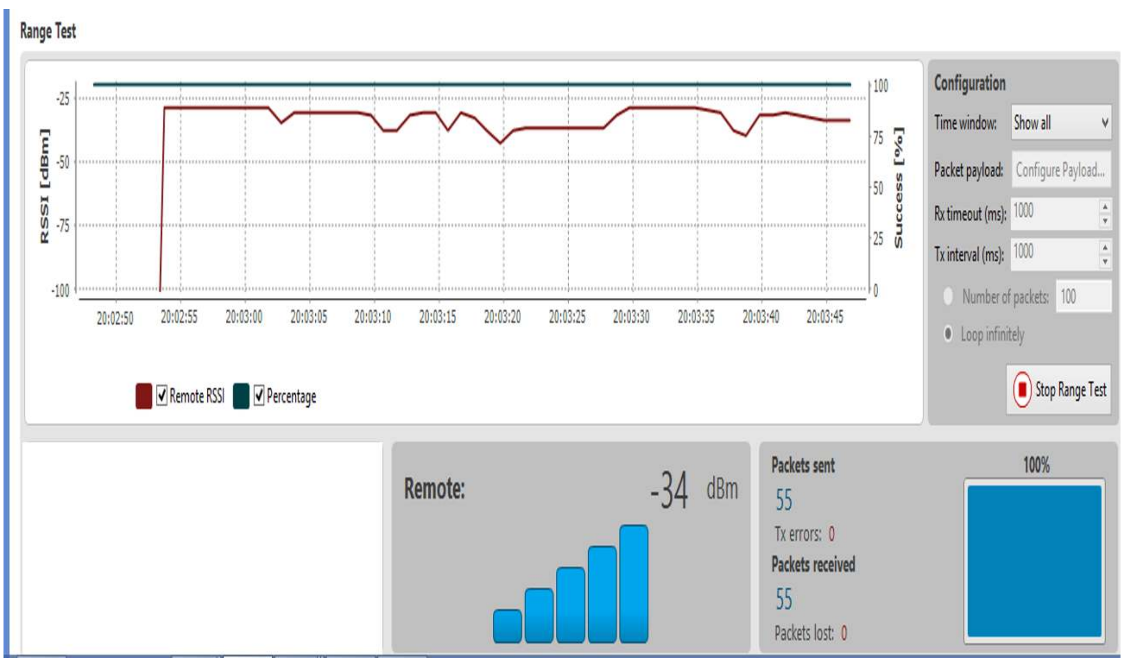

Figure 24 Closest node to Meshlium

Figure 25 Furthest node from Meshlium

Figure 24 and 25 are range test results. They have 2 Y axis, one graded to dBm representing received power determined by the previously sent packet and the other representing the percentage of successfully transmitted packets. The X axis represents time. The network was synchronized before testing began so if packets were lost, it was due to the channel and not handshaking during synchronization. The closest node to the Meshlium was transmitting at -10dBm and the furthest was transmitting at -4dBm. As indicated on the figures by the green lines, both had 100% packet delivery. The red line indicates the strength of the signal (RSSI), measured in dBm. In figure 24, the signal strength is high but in figure 25 it is low. In the case of fading, the network may have suffer from loss of packets from the furthest node due to its low RSSI, so its transmitting power was increased to -2dBm and its range test showed significant in its RSSI.

5.2 Data Classification

The one-vs.-one method was used to train the classifier with 70% of the data set. Using equation 4.7 to calculate the amount of classifiers.

$$C = \frac{M(M-1)}{2} \tag{4.7}$$

Where M represents the seven states, the number of classifiers used was twenty one. These classifiers are shown in table 5 with each digit representing one of the states in the home.

Table 5

SVM multiclass classifiers

States	1	2	3	4	5	6	7
1		x	x	x	x	x	x
2			x	x	x	x	x
3				x	x	x	x
4					x	x	x
5						x	x
6							x
7							

Table 6

Data set of sensor values

O2	CO2	CO	Huma	PIR O	PIR R	PIR L	LUM	TEMP	PRE	HE
21	335.42	0.0002	33.8	0	1	0.00	500	68.00	7.08	0.00

Given a data set, consisting of sensor values at a given time as shown in table 6, the data is tested against each classifier in table 5, scoring a vote to a given class that matches the dataset. At the end of the voting process, the data will be classified under the class with the highest vote. If there is a tie, this will lead to misclassification.

Seventy percent of the dataset was used to train the model. The performance of the model was evaluated using the remaining 30% of the dataset. The result is presented on the confusion Matrix (table 7) where each column represents the number of instance in a predicted class and each row represents the number of instances in an actual class. The testing phase had an accuracy of 99.72%.

Table 7

Confusion matrix after testing with 30% of the data set

Actual State at t = τ	Classified state at t=τ						
	SS	SA	SH	DP	AR	AL	EMR
SS	963	0	0	0	0	0	0
SA	0	888	0	0	0	0	0
SH	4	0	932	0	0	0	0
DP	4	0	0	16	0	0	0
AR	0	0	0	0	21	0	0
AL	0	0	0	0	0	1	0
EMR	0	0	0	0	0	0	1

In the confusion matrix (table 7) it can be seen that after testing, a few data sets were misclassified. Four SH and Four DP. The rest of the data was classified appropriately. In figure 27 a plot of all the states in the home can be seen.

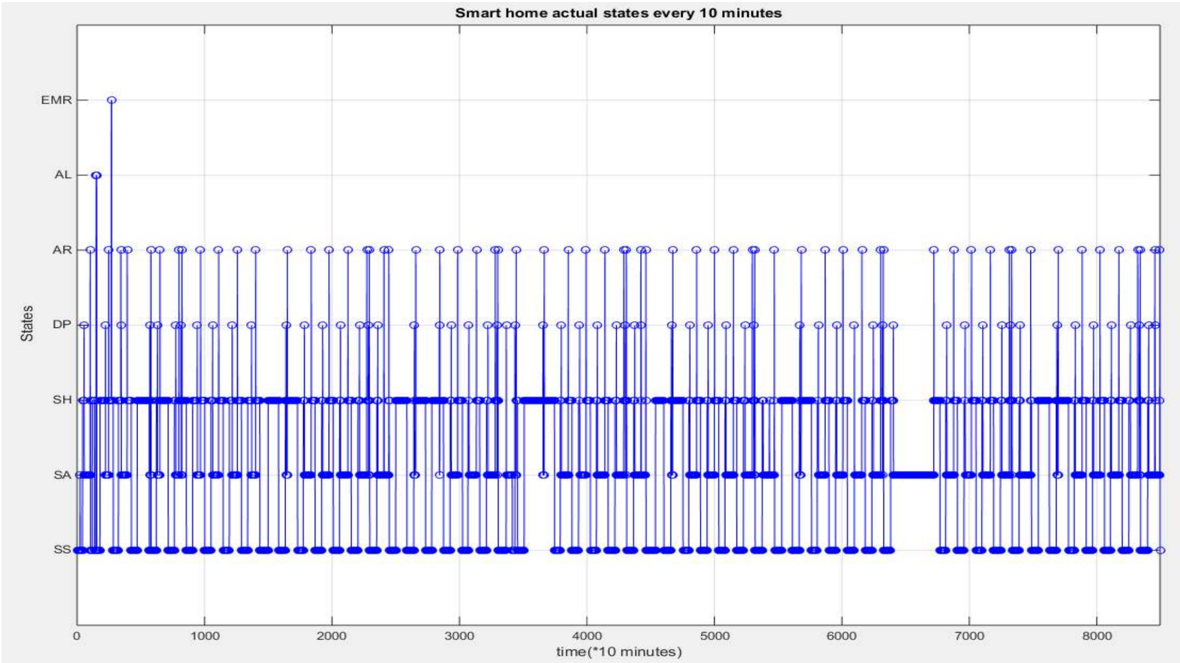

Figure 26 Plot of all the states in the home against time

5.3 State Prediction

In the prediction model, One-vs.-one Multi-class SVM was used to predict the states at WSN = t + τ. The feature vectors for the predictor are extracted from the time-series states in such a way that the previous 2.5 hours state data will predict the next state. The performance of the model was tested by varying the time (sequence length) used to predict the next sate and 2.5 hours gave the best accuracy. The performance of the prediction model is evaluated using the confusion matrix (Table 9), were each column represents the number of instances in a predicted state and each row represents the number of instances in an actual state at t = τ.

Table 8

Confusion matrix of predicted states in the home

Actual state at t = τ	Predicted state at WSN = t + τ				
	SS	SA	SH	DP	AR
SS	803	1	14	0	0
SA	11	744	2	0	0
SH	18	2	786	0	0
DP	2	0	17	0	0
AR	1	15	0	0	0

The confusion matrix shows the performance of the prediction model which had a 96.59% accuracy. Figure 27 shows the performance of the prediction model compared against real time

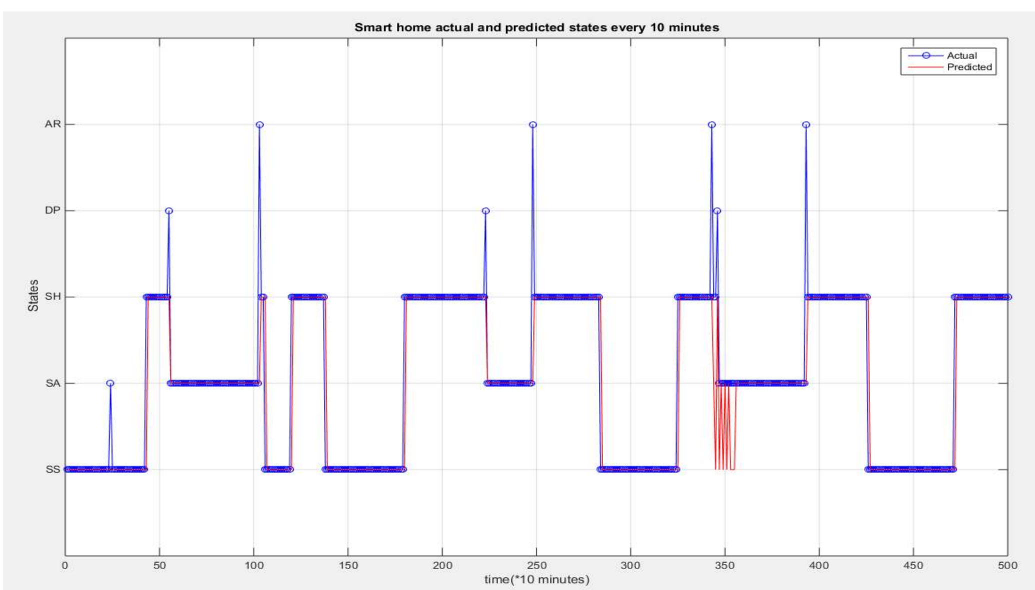

Figure 27 Prediction plot of a three day time series

data. The total input data points were 8,496. During the training of the prediction model, three days' worth of data was left out, 432 data sets. Figure 27 shows the plot of those 3 days and the system prediction

The resulting model performance shows that the AR and DP states were not accurately predicted. This was due to the fact they were few compared to the domineering activities such as SH, SS and SA. The latter activities lasted for longer time periods. The smart home design predicted dominant activities which happened within the home, hence the systems prediction performance was satisfactory.

Putting together all the building blocks from this thesis the model in figure 28 is the designed system model.

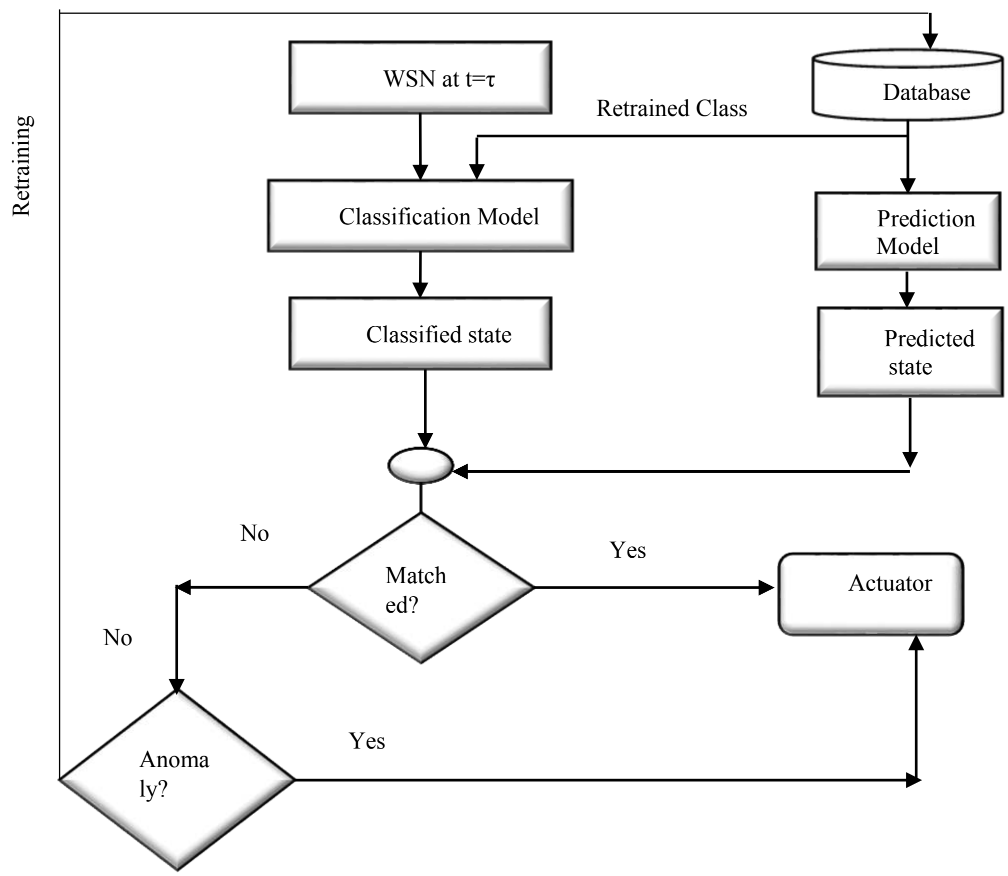

Figure 28 System flow chart

The system contains two data sources; a historical database and a WSN collecting data in real time. The information from the historical database is used to train the classification and prediction models using support vector machine (SVM).

The real time WSN data passes through the classification model so we can always have information on the state of the home. We also utilize the predicted class, by comparing it with the real time classified state. If the two states match, the output is sent to an actuating system. If they do not match the real time state will be sent to the actuating system.

When the dataset continuously produced from the WSN cannot be classified by the classification model, the dataset cannot be termed an anomaly, hence it is sent back to the database for retraining.

For instance, if the WSN from a room produces data at 3:15 PM. The context of the data prompts the class "sleeping" and the predicted state is "resident away". The smart home will implement actions for a sleeping resident such as locking the doors and switching off the TV set (just for illustration, actuator systems are not under the scope of this work). However, if a room has been re-purposed and it continuously (days) produces data with high temperatures like that of a steam room. This can no longer be considered an anomaly. Therefore, the data will be sent back to the database for retraining so this occurrence can be classified as one of the states.

In this thesis a smart home model was designed to gather data from the home, classify and predict future classes. The classification of home states was carried out with SVM and attained an accuracy of 98.6%. Utilizing this information future home states were predicted and compared against real time WSN data with satisfactory results. This smart home design can be beneficial in anomaly detection and smart home autonomy.

CHAPTER 6

Conclusion/Discussion

This work began with the goal of studying smart homes, their recent implementations and theoretical studies in a bid to make a contribution to this research area. This work though broad in scope, was essential in order to thoroughly understand smart homes current implementation and future development.

Researching smart homes exposed the gaping problem of lack of standardization. This prompted the need to create a model that can be utilized as a standard for its learning ability. Creating a system model proved daunting, as it cannot be truly tested without real world implementation. This thesis took the first step in that direction by utilizing data gathered from a dynamic environment, in order to build classification and prediction models, which are integral a systems learning ability.

The results from the classification and prediction models proved satisfactorily, though it can be argued due to the amount of states in this thesis, the robustness of the models were not thoroughly tested. This is an area which should be explored further. Increasing the home states and utilizing other machine learning techniques or perhaps a hybrid of techniques to improve the system learning ability.

However, notwithstanding the route through which classification and predictions are achieved in the home, the general system shown in figure 28 still holds. This system is capable of being fully autonomous (no human interface), as the historical state of the home can be utilized as a vital cog in the decision making module when future situations arise. Thus, leaving little room for the need of human intervention.

Furthermore, through the utilization of WSN the system is scalable to any number of sensors, in the case of structural development of the home. By comparing classified and predicted states, the system is capable of detecting anomalies in the home. If due to an unforeseen circumstance or maintenance issues the WSN is out of service, the system can control the home using its prediction model. Its retraining feature enables the system to be adaptive to new users as well as changes which occur in the home such as re-purposing of rooms.

References

[1] D. Stefanov, Z. Bien and W.-C. Bang, "The smart house for older persons and persons with physical disabilities: structure, technology arrangements, and perspectives," Neural Systems and Rehabilitation Engineering," *Neural Systems and Rehabilitation Engineering, IEEE Transactions,* vol. 12, no. 2, pp. 228,250, 2004.

[2] J. Hicks, "Forbes," Forbes, 24 April 2013. [Online]. Available: http://www.forbes.com/sites/jenniferhicks/2013/06/24/robots-and-smart-homes-for-the-silver-economy/. [Accessed 17 October 2014].

[3] Smart Home automation, "Solution Center," Smart Home, [Online]. Available: http://www.smarthome.com/sc-save-energy-the-smarthome-way. [Accessed 17 february 2015].

[4] "Carbon Monoxide Questions and Answers," Us Consumer product safety commission, [Online]. Available: http://www.cpsc.gov/en/Safety-Education/Safety-Education-Centers/Carbon-Monoxide-Information-Center/Carbon-Monoxide-Questions-and-Answers-/. [Accessed 17 February 2015].

[5] A. Chianis, "8 Surprising Home Burglary Statistics," SafeWise, 20 may 2014. [Online]. Available: http://www.safewise.com/blog/8-surprising-home-burglary-statistics/. [Accessed 17 February 2015].

[6] I. Akyildiz and M. Vuran, Wireless sensor networks, West Sussex: John Wiley & Sons Ltd, 2010.

[7] L. D. Xu, W. He and S. Li, "Internet of Things in Industries: A Survey," *Industrial Informatics, IEEE Transactions,* vol. 10, no. 4, pp. 2233 - 2243, 2013.

[8] L. Atzori, A. Lera and G. Morabito, "The Internet of Things: A survey," *Computer Networks,* no. 54, p. 2787–2805, 2010.

[9] E. Welbourne, L. Battle, G. Cole, K. Gould, K. Rector, S. Raymer, M. Balazinska and M. Borriello, "Building the Internet Using RFID : The RFID Ecosystem Experience," *Internet Computing, IEEE ,* vol. 13, no. 4, pp. 48-55, 2009.

[10] A. Moghadas, M. Jamshidi and M. Shaderam, "Telemedicine in healthcare system," in *Automation Congress, 2008. WAC 2008. World*, Hawaii, 2006.

[11] D. Hartman, "The Advantages of a Smart House," Sfgate, [Online]. Available: http://homeguides.sfgate.com/advantages-smart-house-8670.html. [Accessed 18 February 2015].

[12] K. Fujinami, "Interaction Design Issues in Smart Home Environments," in *Future Information Technology (FutureTech)*, Busan, 2010.

[13] M. Starsinic, "System architecture challenges in the home M2M network," in *Applications and Technology Conference (LISAT)*, Farmingdale, 2010.

[14] G. Youngblood and D. Cook, "Data Mining for Hierarchical Model Creation," *Systems, Man, and Cybernetics, Part C: Applications and Reviews,* vol. 37, no. 4, pp. 561 - 572, 2007.

[15] N. Balta-Ozkan, B. Boteler and O. Amerighi, "European smart home market development: Public views on technical and economic aspects across the United Kingdom, Germany and Italy," *Energy Research & Social Science,* vol. 3, pp. 65-67, 2014.

[16] Marc, "Innovations The Smart Home | Intelligent Home Automation," Entrepreneurial Insights, 22 August 2014. [Online]. Available: http://www.entrepreneurial-insights.com/smart-home-intelligent-home-automation/. [Accessed 23 February 2015].

[17] R. Valverde Ibanez, M. Keysermann and P. Vargas, "Emotional memories in autonomous robots," in *Robot and Human Interactive Communication*, Edinburgh, 2014.

[18] University of Hertfordshire, "Prototype of a robotic system with emotion and memory," Science Daily, 5 February 2015. [Online]. Available: http://www.sciencedaily.com/releases/2015/02/150205141151.htm. [Accessed 23 February 2015].

[19] J. Gerhar, Home Automation & Wiring, New York: McGraw-Hill, 1999.

[20] R. Harper, Inside the Smart Home, Springer, 2003.

[21] G. Anogianakis, C. Bühler and M. Soede, Advancement of Assistive Technology, IOS Press, 1997.

[22] Wikipedia, "Home automation," Wikipedia, 8 February 2015. [Online]. Available: en.wikipedia.org/wiki/Home_automation#cite_note-12. [Accessed 23 February 2015].

[23] C. Enz, N. Scolari and U. Yodprasit, "Ultra Low-Power Radio Design fo Wireless Sensor Networks," in *IEEE Intemnational Workshop on Radio-Frequency Integration Technology,*, Singapore, 2005.

[24] C. Zhang and W. Luo, "Topology Performance Analysis of Zigbee Network in the Smart Home Environment," College of Electrical and Information Engineering Guangxi University of Science and technology, Guangxi Zhuang, 2013.

[25] J. Byun, J. Boungju, J. Noh, Y. Kim and S. Park, "An Intelligent Self-Adjusting Sensor

for Smart Home Services based on ZigBee Communications," *Consumer Electronics, IEEE Transactions on,* vol. 58, no. 3, pp. 794 - 802, 2012.

[26] A. Roy, S. Das and K. Basu, "A Predictive Framework for Location-Aware Resource Management in Smart Homes," *IEEE transactions on mobile computing ,* vol. 6, no. 11, pp. 1270 -1283, 2007.

[27] S. Choi, E. Kim and S. Oh, "Human behavior prediction for smart homes using deep learning," in *RO-MAN*, Gyeongju, 2013.

[28] M. Alam, M. Reaz and M. Ali, "SPEED: An Inhabitant Activity Prediction Algorithm for Smart Homes," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on ,* vol. 42, no. 4, pp. 985 - 990, 2011.

[29] Silicon Laboratories, Inc, "The Evolution of Wireless Sensor Networks," Silicon Laboratories, Inc, 2013.

[30] I. Akyildiz and M. Vuran, Wireless sensor networks, West Sussex: John Wiley & Sons Ltd, 2010.

[31] H. Shang, R. Xu and J. Yuan, "Smart Home based on ZigBee and IOS Software," Electronics Engineering Department of Fudan Univerisity, Shanghai, 201.

[32] Bluetooth.com, "Bluetooth fast facts," [Online]. Available: http://www.bluetooth.com/Pages/Fast-Facts.aspx. [Accessed 11 November 2014].

[33] A. Goldsmith, "Low-Cost Low-Power Radios: Bluetooth and Zigbee," in *Wireless Communications*, Cambridge, Cambridge University Press, 2005, pp. 15-16.

[34] A. Goldsmith, "Evolution of 2G Systems," in *Wireless Communications*, Cambridge, Cambridge University Press, 2005, p. 557.

[35] G. Raju and H. Wang, "Sensor Data Fusion Using Pitman's Closeness Technique and Complete Linkage Algorithm," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, San Antonio, 1994.

[36] G. Wang, J. Hu, Q. Zhang, X. Liu and J. Zhou, "Granular Computing Based Data Mining in the Views of Rough Set and fuzzy set," in *Granular Computing*, Hangzhou, 2008.

[37] S. Stenudd, "Using machine learning in the smart environment," Julkaisija, Oulu, 2010.

[38] Libelium, "Waspmote technical guide," Libelium Comunicaciones Distribuidas S.L., 2014.

[39] Libelium, "Wapmote Networking guide," Libelium, 2014.

[40] Libelium, "Event sensor board 2.0," Libelium, 2014.

[41] "Illinois Department of Public Health Guidelines for," May 2011. [Online]. Available: http://www.idph.state.il.us/envhealth/factsheets/indoorairqualityguide_fs.htm. [Accessed 2 February 2015].

[42] Libelium, "Waspmote frame guide," Libelium, 2014.

[43] Libelium, "Waspmote programming guide," Libelium Comunicaciones Distribuidas S.L, Spain, 2014.

[44] Digi International, "Wireless Mesh Networking Zigbee vs Digimesh," Digi International Inc., Minnetonka, 2008 .

[45] D. Gascón, "Wireless sensor research group," 17 November 2008. [Online]. Available: http://sensor-networks.org/index.php?page=0823123150. [Accessed 2 February 2015].

[46] A. I and V. M, Wireless sensor networks, West Sussex: John Wiley & Sons Ltd, 2010 .

[47] M. Riedel, A. Memon and s. Memon, "High productivity data processing analytics

methods with applications," in *MIPRO*, Opatija, 2014.

[48] Wikipedia, "Machine Learning," Wikipedia, [Online]. Available: http://en.wikipedia.org/wiki/Machine_learning#cite_note-1. [Accessed 25 3 2015].

[49] S. Stenudd, "Using machine learning in the adaptive control of a smart environment," Utigivare, Vuorimiehentie, 2010.

[50] S. Lee and A. Verri, "SVM," *Springer-Verlag Berlin Heidelberg,* vol. LNCS 2388, pp. 213-236, 2002.

[51] Wikipedia, "Support vector machine," [Online]. Available: http://en.wikipedia.org/wiki/Support_vector_machine. [Accessed 24 04 2015].

[52] B. Scholkopf and A. Smol, Learning with Kernels- Support Vector Machines, Regularization, Optimization and Beyond, Cambridge: MIT Press, 2002.

[54] ABI Research, "1.5 Million Home Automation Systems Installed in the US This Year," ABI Research, 2012 November 2012. [Online]. Available: https://www.abiresearch.com/press/15-million-home-automation-systems-installed-in-th. [Accessed 2 February 2015].

[55] F. Afghah, "Advanced Wireless Network Lecture 1," North Carolina A&T State University., Greensboro, 2014.