T.Y. B.Sc.

(Computer Science)

SEMESTER - VI (CBCS)

WIRELESS SENSOR NETWORKS AND MOBILE COMMUNICATION

SUBJECT CODE - USCS601

Prof.(Dr.) D. T. Shirke
Offg. Vice-Chancellor,
University of Mumbai,

Prin. Dr. Ajay Bhamare
Offg. Pro Vice-Chancellor,
University of Mumbai,

Prof. Prakash Mahanwar
Director,
IDOL, University of Mumbai,

Programme Co-ordinator : **Shri. Mandar Bhanushe**
Head, Faculty of Science and Technology
IDOL, University of Mumbai, Mumbai

Course Co-ordinator : **Ms. Mitali Vijay Shewale**
Doctoral Researcher,
Veermata Jijabai Technological Institute
Mumbai

Editor : **Mr. Anish Raut**
Assistant Manager,
Dahua Technology India Pvt. Ltd

Course Writers : **Ms. Zainab khan**
Assistant Professor,
Thakur shyamnarayan degree college

: **Ms. Mitali Vijay Shewale**
Doctoral Researcher,
Veermata Jijabai Technological Institute
Mumbai

June 2023, Print - I

Published by : Director
Institute of Distance and Open Learning ,
University of Mumbai,Vidyanagari, Mumbai -400 098.

DTP Composed : Mumbai University Press
Printed by Vidyanagari, Santacruz (E), Mumbai - 400 098

CONTENTS

SEMESTER VI

THEORY

Course: USCS601	TOPICS (Credits : 03 Lectures/Week: 03) Wireless Sensor Networks and Mobile Communication	
Objectives: In this era of wireless and adhoc network, connecting different wireless devices and understanding their compatibility is very important. Information is gathered in many different ways from these devices. Learner should be able to conceptualize and understand the framework. On completion, will be able to have a firm grip over this very important segment of wireless network. **Expected Learning Outcomes:** After completion of this course, learner should be able to list various applications of wireless sensor networks, describe the concepts, protocols, design, implementation and use of wireless sensor networks. Also implement and evaluate new ideas for solving wireless sensor network design issues.		
Unit I	**Introduction:** Introduction to Sensor Networks, unique constraints and challenges. Advantage of Sensor Networks, Applications of Sensor Networks, Mobile Adhoc NETworks (MANETs) and Wireless Sensor Networks, Enabling technologies for Wireless Sensor Networks. **Sensor Node Hardware and Network Architecture:** Single-node architecture, Hardware components & design constraints, Operating systems and execution environments, introduction to TinyOS and nesC. Network architecture, Optimization goals and figures of merit, Design principles for WSNs, Service interfaces of WSNs, Gateway concepts.	**15L**
Unit II	**Medium Access Control Protocols:** Fundamentals of MAC Protocols, MAC Protocols for WSNs, Sensor-MAC Case Study. **Routing Protocols** : Data Dissemination and Gathering, Routing Challenges and Design Issues in Wireless Sensor Networks, Routing Strategies in Wireless Sensor Networks. **Transport Control Protocols :** Traditional Transport Control Protocols,	**15L**

	Transport Protocol Design Issues, Examples of Existing Transport Control Protocols, Performance of Transport Control Protocols.	
Unit III	**Introduction, Wireless Transmission and Medium Access Control:** Applications, A short history of wireless communication. **Wireless Transmission:** Frequency for radio transmission, Signals, Antennas, Signal propagation, Multiplexing, Modulation, Spread spectrum, Cellular systems. **Telecommunication, Satellite and Broadcast Systems: GSM:** Mobile services, System architecture, Radio interface, Protocols, Localization And Calling, Handover, security, New data services; DECT: System architecture, Protocol architecture; ETRA, UMTS and IMT- 2000. Satellite Systems: History, Applications, Basics: GEO, LEO, MEO; Routing, Localization, Handover.	**15L**

Textbook(s):

1) Protocols and Architectures for Wireless Sensor Network, Holger Kerl, Andreas Willig, John Wiley and Sons, 2005
2) Wireless Sensor Networks Technology, Protocols, and Applications ,Kazem Sohraby, Daniel Minoli and TaiebZnati, John Wiley & Sons, 2007
3) Mobile communications, Jochen Schiller,2nd Edition, Addison wisely , Pearson Education,2012

Additional Reference(s):

1) Fundamentals of Wireless Sensor Networks, Theory and Practice, Waltenegus Dargie, Christian Poellabauer , Wiley Series on wireless Communication and Mobile Computing, 2011
2) Networking Wireless Sensors, Bhaskar Krishnamachari , Cambridge University Press, 2005

1

INTRODUCTION

Unit Structure

ABSTRACT

Wireless sensor network is a type of wireless network consist a collection of tiny device called sensor node. Sensor node has a resource constraint means battery power, storage and communication capability. These sensor nodes are set with radio interface with which they communicated with one another to form a network. Wireless sensor network has very necessary application like remote has remote environmental monitoring and target tracking. The goal of our survey is to present a comprehensive review of the recent literature on various aspects of wireless sensor networks and also discuss how wireless sensor network works and advantages and disadvantages over the traditional network. Wireless sensor networks are networks composed of a number of sensor nodes that communicate wirelessly. It's utilized over a wide range of applications. This paper looks at the wireless sensor networks from the applications point of view and surveyed different application areas where the use of such sensor networks and their specifications, capabilities.

keywords: wireless sensor networks.

1.0 OBJECTIVES

- In this lesson, you will be introduced for the types of applications for which wireless sensor networks are intended and a first intuition about the types of technical solutions that are required, both in hardware and in networking technologies. Also, able to understand the capabilities and limitations of the nodes in a sensor network and principles options on how individual sensor nodes can be connected into a wireless sensor network.
- The objective of this chapter is to provide an up-to-date treatment of the fundamental techniques, applications, taxonomy, and challenges of wireless sensor networks.
- Wireless sensor networks aim to gather environmental data and the node devices placement may be known or unknown a priori. Network nodes can have actual or logical communication with all devices; such a communication defines a topology according to the application.
- communication technologies continue to undergo rapid advancement. In recent years, there has been a steep growth in research in the area of wireless sensor networks (WSNs). In WSNs, communication takes place with the help of spatially distributed, autonomous sensor nodes equipped to sense specific information. WSNs can be found in a variety of both military and civilian applications worldwide. Examples include detecting enemy intrusion on the battlefield, object tracking, habitat monitoring, patient monitoring and fire detection. Sensor networks are emerging as an attractive technology with great promise for the future. However, challenges remain to be addressed in issues relating to coverage and deployment, scalability, quality-of-service, size, computational power, energy efficiency and security. This paper presents an overview of the different applications of the wireless sensor networks and various security related issues in WSNs. in the

last two to three years a number of theoretical and/or simulation studies were done on the topic of object-tracking. while these studies are useful, they are too general and provide little guidance for the actual deployment of sensor networks for real- life location-tracking of an enemy. this thesis focuses on developing an object-tracking application and prescribes sensor network configurations that work well with our algorithms. we implement our software using crossbow hardware technology. the major issues addressed in this projectare the evaluation and efficient use of a wireless sensor network product with no changes, in a real-world application, and efficient ways to algorithmically analyze the collected raw data from the specific wireless sensor networks product. although the focus is the development of a real-world application using wireless sensor networks, it also provides be a great opportunity to explore the new area of wireless communication overall.

1.1 INTRODUCTION

1.1.1 Introduction to Wireless Sensors Networks

- wireless sensor network is a wireless network consisting of spatially distributed autonomous devices that use sensors to monitor physical or environmental conditions. These autonomous devices, or nodes, combine with routers and a gateway to create a typical WSN system. The distributed measurement nodes communicate wirelessly to a central gateway, which provides a connection to the wired world where you can collect, process, analyze, and present your measurement data. To extend distance and reliability in a wireless sensor network, you can use routers to gain an additional communication link between end nodes and the gateway. Currently, wireless sensor networks are beginning to be deployed at an accelerated pace. It is not unreasonable to expect that in 10-15 years that the world will be covered with wireless sensor networks with access to themvia the Internet (Figure-1). This can be considered as the Internet becoming a physical network. This new technology is exciting with unlimited potential for numerous application areas including environmental, medical, military, transportation, entertainment, crisis management, homeland defense, and smart spaces.

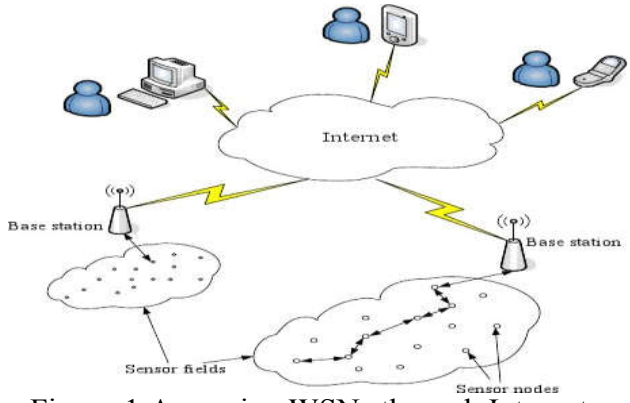


Figure-1 Accessing WSNs through Internet.

- The major challenges to be addressed in WSNs are coverage and deployment, scalability, quality- of- service, size, computational power, energy efficiency and security. Among these challenges, security is a major issue in wireless sensor networks. Most of the threats and attacks against security in wireless networks are almost similar to their wiredcounterparts while some are exacerbated with the inclusion of wireless connectivity. In fact, wireless networks are usually more vulnerable to various security threats as the unguided transmission medium is more susceptible to security attacks than those of the guided transmission medium. The broadcast nature of the wireless communication is a simple candidate for eavesdropping. In this paper we present an overview of the applications and security issues relating to Wireless Sensor Networks(WSNs).

- Vision of Ambient Intelligence

- The most common form of information processing has happened on large, general-purpose computational devices, ranging from old-fashioned mainframes to modern laptops or palmtops. In many applications, like office applications, these computational devices are mostly used to process information that is at its core centered around a human user of a system, but is at best indirectly related to the physical environment.

- In another class of applications, the physical environment is at the focus of attention. Computation is used to exert control over physical processes, for example, when controlling chemical processes in a factory for correct temperature and pressure. Here, the computation is integrated with the control; it is embedded into a physical system. Unlike the former class of systems, such embedded systems are usually not based on human interaction but are rather required to work without it; they are intimately tied to their control task in the context of a larger system. Such embedded systems are a well-known and long-used concept in the engineering sciences (in fact, estimates say that up to 98% of all computing devices are used in an embedded context. Their impact on everyday life is also continuing to grow at a quick pace. Rare is the household where embedded computation is not present to control a washing machine, a video player, or a cell phone. In such applications, embedded systems meet

- human-interaction-based systems. Technological progress is about to take this spreading of embedded control in our daily lives a step further. There is a tendency not only to equip larger objects like a washing machine with embedded computation and control, but also smaller, even dispensable goods like groceries; in addition, living and working spaces themselves can be endowed with such capabilities. Eventually, computation will surround us in our daily lives, realizing a vision of “Ambient Intelligence” where many different devices will gather and process information from many different sources to both control physical processes and to interact with human users.

- To realize this vision, a crucial aspect is needed in addition to computation and control: communication. All these sources of information have to be able to transfer the information to the place where it is needed – an actuator or a user – and they should collaborate in providing as precise a picture of the real world as is required. For some application scenarios, such networks of sensors and actuators are easily built using existing, wired networking technologies. For many other application types, however, the need to wire together all these entities constitutes a considerable obstacle to success: wires constitute a maintenance problem; wires prevent entities from being mobile; and wires can prevent sensors or actuators from being close to the phenomenon that they are supposed to control. Hence, wireless communication between such devices is, in many application scenarios, an inevitable requirement. Therefore, a new class of networks has appeared in the last few years: the so- called Wireless Sensor Network (WSN). These networks consist of individual nodes that areable to interact with their environment by sensing or controlling physical parameters; these nodes have to collaborate to fulfill their tasks as, usually, a single node is incapable of doing so; and they use wireless communication to enable this collaboration. In essence, the nodes without such a network contain at least some computation, wireless communication, and sensing or control functionalities. Despite the fact that these networks also often include actuators, the term wireless sensor network has become the commonly accepted name. Sometimes, other names like "wireless sensor and actuator networks" are also found.

- These WSNs are powerful in that they are amenable to support a lot of very different real- world applications; they are also a challenging research and engineering problem because of this very flexibility. Accordingly, there is no single set of requirements that clearly classifies all WSNs, and there is also not a single technical solution that encompasses the entire design space. For example, in many WSN applications, individual nodes in the network cannot easily be connected to a wired power supply but rather have to rely on onboard batteries. In such an application, the energy efficiency of any proposed solution is hence a very important

- figure of merit as a long operation time is usually desirable. In other applications, power supply might not be an issue and hence other metrics, for example, the accuracy of the delivered results, can become more important. Also, the acceptable size and costs of an individual node can be relevant in many applications. Closely tied to the size is often the capacity of an onboard battery; the price often has a direct bearing on the quality of the node's sensors, influencing the accuracy of the result that can be obtained from a single node. Moreover, the number, price, and potentially low accuracy of individual nodes is relevant when comparing a distributed system of many sensor nodes to a more centralized version with fewer, more expensive nodes of higher accuracy. Simpler but numerous sensors that are close to the phenomenon under study can make the

architecture of a system both simpler and more energy efficient as they facilitate distributed sampling – detecting objects, for example, requires a distributed system.

- Realizing such wireless sensor networks is a crucial step toward a deeply penetrating Ambient Intelligence concept as they provide, figuratively, the "last 100 meters" of pervasive control. To realize them, a better understanding of their potential applications and the ensuing requirements is necessary, as is an idea of the enabling technologies. These questions are answered in the following sections; a juxtaposition of wireless sensor networks and related networking concepts such as fieldbuses or mobile ad hoc network is provided as well.

1.1.2 **Overview of Wireless Sensor Networks**

- The progress in wireless communications, digital electronics, and micro systems has enabled the development of small-size, low-cost, power-efficient multifunctional sensors. Moore's law predicts a great future for this technological field. In the future the typical sensor nodes the size "of a 35 mm film canister" (Wikipedia, Wireless Sensor Network Webpage, 2005), and their development cost will be drastically reduced, generating an explosion in the wireless sensor network usage.

- Wireless sensor networks (WSN) is a rich domain that involves both hardwareand system design. It consists of sensor devices that are "small in size and able to sense, process data, and communicate with each other, typically over an RF (radio frequency)

- Channel" (Haenggi, 2005). Their purpose is to collect and process data from the environment, produce a detection event and then forward the information to a specific destination.

- Wireless sensor networks are a specialization of the wireless ad-hoc mesh networks. They inherit all the ad-hoc and mesh characteristics described above. They are wireless self-organizing, self-healing, and adaptive networks. They contain a large number of small, inexpensive, low-power nodes and use specialized communication techniques and routing, like "an asymmetric many-to-one data flow" (Carle & Simplot- Ryl, 2004) to communicate. Nodes' characteristics (size, lifetime, computational power), system's architecture, and protocols enable WSN to be deeply embedded into the environment. If these capabilities will be combined with the Internet, an "embedded Internet" (Culler & Hong, 2004) will be produced. Zhao and Guibas (2004) accent that "sensor networks extend the existing Internet deep into the physical environment. The resulting network is orders of magnitude more expansive and dynamic than the current TCP/IP network." Figure 3 provides an illustration of a sensor network and Internet integration. A complete WSN implementation is a "macroscopic view" (Carle & Simplot- Ryl, 2004) of the environment, implementing pervasive computing, it "enable us to

observe and interact with physical phenomena in real time at a fidelity that was previously unobtainable." (Carle, & Simplot-Ryl, [2004]). WSN is a new, interesting, and active research area; it introduces various challenges and concerns; the following section highlights some of them.

- Figure 3. Integration of a Wireless Sensor Network and the Internet (Zhao & Guibas, 2004).

1.2 UNIQUE CONSTRAINTS AND CHALLENGES

- ✓ Self-organizing capabilities
- ✓ Short-range broadcast communication and multi-hop routing Dense deployment and cooperative effort of sensor nodes Frequently changing topology due to fading and node failures
- ✓ Limitations in energy, transmitted power, memory, and computing power.
- ✓ They also highlights that the WSN differ from the wireless ad-hoc mesh networks in the latter three characteristics. Zhao and Guibas (2004) identify "limited hardware," "limited support for networking," and "limited support for software development" as general WSN design and implementation challenges. Wang, Hassanein, and Xu (2005) add "data redundancy," the diversity of the possible application, and security and privacy concerns. Before some of the above concerns are analyzed further, we will discuss in the following section the important WSN applications that set the requirements and drove a WSN development.
- ✓ Unlike a centralized system, a sensor network is subject to a unique
- ✓ Set of resource constraints such as finite on-board battery power
- ✓ And limited network communication bandwidth. In a typical sensor
- ✓ Network, each sensor node operates untethered and has a micropro-
- ✓ Cessor and a small amount of memory for signal processing and task
- ✓ Scheduling. Each node is also equipped with one or more sensing
- ✓ Devices such as acoustic microphone arrays, video or still cameras,
- ✓ Infrared (IR), seismic, or magnetic sensors. Each sensor node com-
- ✓ Municates wirelessly with a few other local nodes within its radio
- ✓ Communication range.
- ✓ Sensor networks extend the existing Internet deep into the physi-
- ✓ Cal environment. The resulting new network is orders of magnitude

- ✓ More expansive and dynamic than the current TCP/IP network and
- ✓ Is creating entirely new types of traffic that are quite different from
- ✓ What one finds on the Internet now. Information collected by and
- ✓ Transmitted on a sensor network describes conditions of physical
- ✓ Environments—for example, temperature, humidity, or vibration—
- ✓ And requires advanced query interfaces and search engines to effec-
- ✓ Tively support user-level functions. Sensor networks may inter-
- ✓ Network with an IP core network via a number of gateways, as in
- ✓ Figure 1.1. A gateway routes user queries or commands to appropriate
- ✓ Nodes in a sensor network. It also routes sensor data, at times aggre-
- ✓ Gated and summarized, to users who have requested it or are expected
- ✓ To utilize the information. A data repository or storage service may
- ✓ Be present at the gateway, in addition to data logging at each sensor.
- ✓ Advantages of Sensor Networks
- ✓ Networked sensing offers unique advantages over traditional cen-
- ✓ tralized approaches. Dense networks of distributed communicating
- ✓ sensors can improve signal-to-noise ratio (SNR) by reducing average distances from sensor to
- ✓ source of signal, or target. Increased
- ✓ energy efficiency in communications is enabled by the multihop
- ✓ topology of the network [184]. Moreover, additional relevant information from other sensors can be aggregated during this multihop transmission through in-network processing [104]. But perhaps the
- ✓ greatest advantages of networked sensing are in improved robustness

1.3 ADVANTAGES OF SENSOR NETWORKS

1. It is scalable and hence can accommodate any new nodes or devices at any time.
2. It is flexible and hence open to physical partitions.
3. All the WSN nodes can be accessed through centralized montoring system.
4. As it is wireless in nature, it does not require wires or cables. Refer difference between wired network vs wireless network.

5. Wireless can be applied on large scale and in various domains such as mines, healthcare, surveillance, agriculture etc.

6. It uses different security algorithms as per underlying wireless technologies and hence provide reliable network for consumers or users.

1.3.1 Disadvantages of WSN (wireless sensors networks)

Following are the drawbacks or disadvantages of WSN:

1. As it is wireless in nature, it is prone to hacking by hackers.

2. It cannot be used for high-speed communication as it is designed for low-speed applications.

3. It is expensive to build such network and hence cannot be affordable by all.

4. There are various challenges to be considered in WSN such as energy efficiency, limited bandwidth, node costs, deployment model, Software/hardware design constraints and so on.

5. In star topology based WSN, failure of central node leads to whole network shutdown.

1.4 APPLICATIONS OF SENSOR NETWORKS

1.4.1 Industrial Control and Monitoring

The deployment of wireless network sensors in the industrial control-and-monitoring field seems very prominent. Normally, a factory has a control room to monitor and control the state of the plant and the condition of the equipment. Specific critical values, like temperature or pressure, are collected from the plant or the equipment. The values describe the plant's or the equipment's condition, which is then forwarded to the control room where it is evaluated. Traditionally, industrial control and monitoring requires the deployment of a complex, expensive wired network. Sensor networks can replace the wired network, providing reliable data transfer and reducing the initial deployment and maintenance cost.

Lighting, ventilation and air-conditioning are other possible areas for wireless sensors. WSN provide the flexibility to support dynamic changes in the environment. This is also enhanced by the WSN programming feature, which offers secure and balanced services (e.g., balanced heating and air conditioning). When used to control and monitor complex equipment like robots, or other rotating and moving equipment, WSN provide the necessary flexibility. Thus, the system's reliability is increased, because damage caused by the machinery's movement is avoided. In addition, small-size sensing nodes can be used where wired implementations are impossible.

1.4.2 Home Applications

Home automation is another large application area for wireless sensor networks. The uses in the industrial applications field described above also apply to home implmentations. Centralized control of home appliances has already been implemented by using wired solutions or other wireless technology solutions. Their replacement by a wireless sensor network provides a development and maintenance cost reduction, system flexibility, and stretch ability. WSN also provides total, and secure control of the home devices. Another area for the use of WSN that is relevant to home application is the toy industry, a large market. The nature of wireless networks enable toys to behave incomplex and logical ways at a reasonable cost..

1.4.3 Environmental and Agricultural Monitoring

environmental monitoring of WSN implementations as pioneers in this technology. Wireless networks can be used for habitat monitoring and ecosystem measurements. Haenggi (2004) finds that seismic activity, forest fire, floods and water quality also can be detected and localized by the use of WSNs. Culler and Hong (2004) claim that the outdoor deployment, low power operation, fault tolerance, data quality, and networking characteristics of WSNs are ideal for environmental applications. Moreover, given those characteristics, WSNs can be used for agricultural purposes. Better knowledge of the agricultural environment enables the more precise control of fertilizers, water management cost reduction, quality maximization and environment protection.

1.4.4 Military and Security Applications

As with almost any new technology military and security application are recommended uses for wireless sensor networks. WSNs can assist or replace quards around a building or camp perimeter. Target localization and identification is another potential use, whereby friendly troops use WSNs to identify themselves (Callaway,2004). Haenggi (2005) finds that such implementation can improve "military command, control, communication and computing (C4)" schema. Additionally, he describes an application for "surveillance and battle-space monitoring" in which the proper sensors are deployed in the ground or are carried by unmanned vehicles to monitor opposing forces. Haenggi (2005) mentions other potential uses in an "urban warfare" field: "to prevent reoccupation" of buildings that have already been cleared; and for "self-healing minefields," where, instead of a "static complex obstacle," the WSNs provide "an intelligent, dynamic obstacle that senses related positions and responds to an enemy breaching attempt by physical reorganization."

1.4.5 Asset Tracking

Among the potential uses of wireless sensor networks, asset tracking is also a large area of interest for military and commercial application. Calllaway (2004) describes a possible use: for tracking "shipping containers both in a port and on a ship. By placing WSN nodes inside each

container, it and its content become recognizable from a distance. An exact knowledge of the container's type and position can save handlers a great amount of time by preventing unnecessary errors. The WSNs provide a cost-effective way to increase the "shipper's productivity."

1.4.6 Heath Monitoring

identifies two different wireless sensor network medical applications that are expected to rapidly increase. First, he mentions "medical sensing" in which data such as "body temperature, blood pressure, and pulse," collected from the system, can be transmitted to a local or remote computer for health monitoring uses. Additionally, WSNs can be used in the "micro-surgery" field, where tiny medical instruments are used to perform " microscopic and minimal invasive surgery."

Application Categories

The above applications show that, among the WSN applications, there are some common features. Holger and Willig (2005) identify the existence of data "sources" and "sinks" in most of the WSN applications in which the "sources" are the nodes that sense the data from the environment and the "sinks" are the nodes where the data arrived, like gateways. The "sinks" can be WSN components or they can sit outside the system. Holger and Willig (2005) place the applications based on the sources-and-sinks interaction in four categories. The first category is "event detection," is which the sources, when they detect an event send messages to the sinks. An event could be a single value, for example, an above threshold humidity, or a complicated type. Holger and Willig's second category is "periodic measurements," in which the sources periodically send messages to the sinks. The third category comprises "function approximation and edge detection" in which the WSN system, based on specific finite values, approximates an "unknown function." The final category is "tracking" in which the event producer is mobile, and thus a WSN is used to detect the object's position and possibly its speed and direction.

The preceding section included categories and possible implementations of wireless sensor networks. According to Haenggi (2005), the opportunities for the WSNs are "ubiquitous." Zhao and Guibas (2004) find that "the main long-term will be the increase in the number of sensors per application and the increase in the decentralization of sensor control and processing." However, the relevant constraints and challenges, that

are mentioned above will be further analyzed in the next sections. They must be addressed for easier and faster deployment of the wireless sensor network applications.

- The major applications of WSNs
- Application of WSNs

1. **Logistics**

Logistics is a multi-player business which has changed significantly in the last decade. E.g. transport of food. Figure 9 shows one application scenario, where wireless sensor network nodes are connected to goods (mostly food because of their perishable nature). The goods are loaded from a storehouse or warehouse to a good carrier vehicle, in which their nodes need to be self-organize and form a network of nodes, which can forward information of the goods' from one state to the outside world using a gateway (e.g. a telematics unit).

Logistics benefits clearly from Wireless Sensor Networks. However, the requirements of logistics for applicable WSNs are challenging.

2. **Environmental monitoring**

Simple computations and to send/receive data performance done by the sensor nodes. These nodes are small in size and are embedded into devices. Data collection is the typical usage where data collected from the surrounding environment via sensors. Environment monitoring has become an important field of control and protection, providing real-time systems and control communication with the physical world. During data collection sensor nodes

Monitor and manage air quality,

Monitor and manage conditions of traffic,

Monitor and manage weather situations.

Characteristics of an environmental monitoring system

Autonomy. Batteries must be able to power the weather stations during the whole deployment.

Reliability. The network has to perform simple and predictable operations, to prevent unexpected crashes.

Robustness. The network must account for a lot of problems such as poor radio connectivity (e.g., in case of snow fall) or hardware failures.

Flexibility. One must be able to quickly add, move, or remove stations at any time depending on the needs of the applications.

3. Industrial supervision

The advances in wireless communication, microelectronics, digital electronics, and highly integrated electronics and the increasing need for more efficient controlled electric systems make the development of monitoring and supervisory control tools the object of study of many researchers.

4. Intelligent buildings

Wireless Sensor Networks (WSN) has become cardinal towards the implementation of smart homes, and they are proved to be a permitting technology for assisted living. WSNs are deemed appropriate for placement in home environments for diverse applications.

Military applications

WSNs consist of a large number of small sensor nodes. Costing of small nodes is also less expensive. In military operations, there is always a threat or security challenges of being attacked by enemies. So if regular use of small nodes which is less expensive help to reduce the loss.

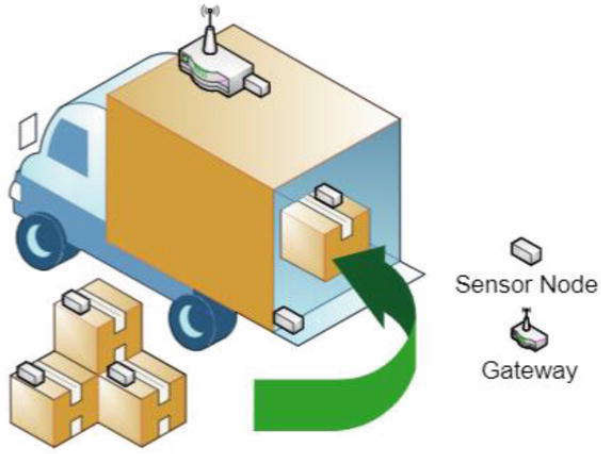


Figure 9.

Wireless sensor network for logistics.

Figure 10 shows wireless sensor networks for military application. This application provides suitable sensors which can be used in top secret missions. These sensors can detect, identify and classify threads based on the count, number, whether it is armored vehicles or men in foot, type and amount of weapons they carry, etc., can be detected in advance. This application provides reliable real time war pictures and better situational awareness.

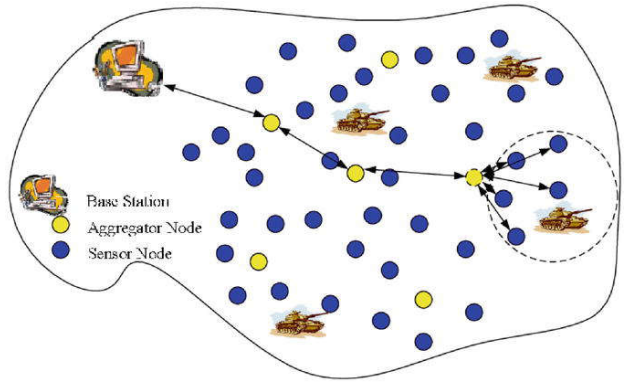


Figure 10.

Wireless sensor network for military application.

TYPES OF APPLICATIONS *(Cont. - Types of wireless sensor networks through)*

Many of these applications share some basic characteristics. In most of them, there is a clear difference between sources of data – the actual nodes that sense data – and sinks – nodes where the data should be delivered to. These sinks sometimes are part of the sensor network itself; sometimes they are clearly systems "outside" the network (e.g. the firefighter's PDA communicating with a WSN). Also, there are usually, but not always, more sources than sinks and the sink is oblivious or not interested in the identity of the sources; the data itself is much more important.

The interaction patterns between sources and sinks show some typical patterns. The most relevant ones are:

Event detection Sensor nodes should report to the sink(s) once they have detected the occurrence of a specified event. The simplest events can be detected locally by a single sensor node in isolation (e.g. a temperature threshold is exceeded); more complicated types of events require the collaboration of nearby or even remote sensors to decide whether a (composite) event has occurred (e.g. a temperature gradient becomes too steep). If several different events can occur, event classification might be an additional issue.

Periodic measurements Sensors can be tasked with periodically reporting measured values. Often, these reports can be triggered by a detected event; the reporting period is application dependent.

Function approximation and edge detection The way a physical value like temperature changes from one place to another can be regarded as a function of location. A WSN can be used to approximate this unknown function (to extract its spatial characteristics), using a limited number of samples taken at each individual sensor node. This approximate mapping should be made available at the sink. How and when to update this mapping depends on the application's needs, as do the approximation accuracy and the inherent trade-off against energy consumption.

Similarly, a relevant problem can be to find areas or points of the same given value. An example is to find the isothermal points in a forest fire application to detect the border of the actual fire. This can be generalized to finding "edges" in such functions or to sending messages along the boundaries of patterns in both space and/or time.

Tracking The source of an event can be mobile (e.g. an intruder in surveillance scenarios). The WSN can be used to report updates on the event source's position to the sink(s), potentially with estimates about speed and direction as well. To do so, typically sensor nodes have to cooperate before updates can be reported to the sink. These interactions can be scoped both in time and in space (reporting events only within a

given time span, only from certain areas, and so on). These requirements can also change dynamically overtime; sinks have to have a means to inform the sensors of their requirements at runtime. Moreover, these interactions can take place only for one specific request of a sink (so-called "one-shot queries"), or they could be long-lasting relationships between many sensors and many sinks.

The examples also have shown a wide diversity in deployment options. They range from well planned, fixed deployment of sensor nodes (e.g. in machinery maintenance applications) to random deployment by dropping a large number of nodes from an aircraft over a forest fire. In addition, sensor nodes can be mobile themselves and compensate for shortcomings in the deployment process by moving, in a post deployment phase, to positions such that their sensing tasks can be better fulfilled. They could also be mobile because they are attached to other objects (in the logistics applications, for example) and the network has to adapt itself to the location of nodes.

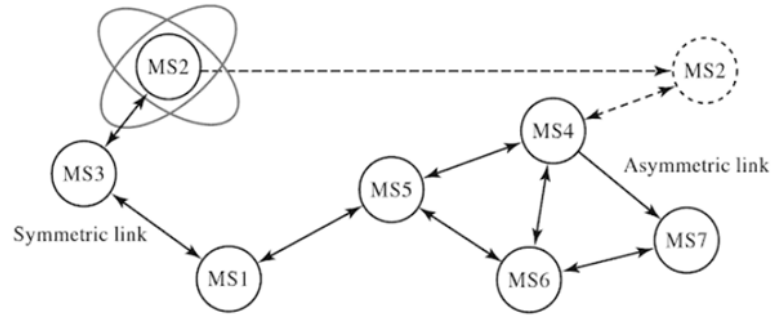
Closely related to the maintenance options are the options for energy supply. In some applications, wired power supply is possible and the question is mute. For self-sustained sensor nodes, depending on the required mission time, energy supply can be trivial (applications with a few days of usage only) or a challenging research problem, especially when no maintenance is possible but nodes have to work for years. Obviously, acceptable price and size per node play a crucial role in designing energy supply.

1.5 MOBILE ADHOC NETWORKS (MANETS) AND WIRELESS SENSOR NETWORKS

Mobile Adhoc Network (MANETs)

- A MANET consists of a number of mobile devices that come together to form a network as needed, without any support from any existing internet infrastructure or any other kind of fixed stations.
- A MANET can be defined as an autonomous system of nodes or MSs(also serving as routers) connected by wireless links, the union of which forms a communication network modeled in the form of an arbitrary communication graph.
- This is in contrast to the well-known single hop cellular network model that supports the needs of wireless communication between two mobile nodes relies on the wired backbone and fixed base stations.
- In a MANET, no such infrastructure exists and network topology may be changed dynamically in an unpredictable manner since nodes are free to move and each node has limiting transmitting power, restricting access to the node only in the neighboring range.
- MANETs are basically peer-to-peer, multi-hop wireless networks in which information packets are transmitted in a store and forward

manner from a source to an arbitrary destination, via intermediate nodes as given in the figure:



- As nodes move, the connectivity may change based on relative locations of other nodes. The resulting change in the network topology known at the local level must be passed on to other nodes so that old topology information can be updated.
- For example, as MS2 in the figure changes its point of attachment from MS3 to MS4, other nodes that are part of the network should use this new route to forward packets to MS2. In the figure, we assume that it is not possible to have all nodes within each other's radio range. In case all nodes are closed by within each other's radio range, there are no routing issues to be addressed.
- In figures raise another issue, that of symmetric and asymmetric (bidirectional) and asymmetric (unidirectional) links. Consider symmetric links with associative radio range; for example, if MS1 is within radio range of MS3, then MS3 is also within radio range of MS1. The communication links are symmetric. This assumption is not always valid because of differences in transmitting power levels and the terrain. Routing in asymmetric networks is relatively hard task. In certain cases, it is possible to find routes that exclude asymmetric links, since it is cumbersome to find the return path. The issue of efficient is one of the several challenges encountered in a MANET.
- The other issue is varying the mobility patterns of different nodes. Some other nodes are highly mobile, while others are primarily stationary. It is difficult to predict a node's movement and direction of movement and numerous studies have been performed to evaluate their performance using different simulators.

Characteristics of MANET

Some characteristics of adhoc network are as follows:

- **Dynamic topologies:** nodes are free to move arbitrarily; thus the network topology may be changed randomly and unpredictably and primarily consists of bidirectional links. In some cases where the transmission power of two nodes is different, a unidirectional link may exist.

- **Bandwidth-constrained and variable capacity links:** wireless links continue to have significantly lower capacity than infrastructure networks.

- **Energy-constrained operation:** some or all of the MSs in a MANET may rely on batteries or other exhaustible means for their energy. For these nodes or devices, the most important system design optimization criteria may be energy conservation.

- **Limited physical security:** MANETs are generally more prone to physical security threats than wire line networks. The increased possibility of eavesdropping, spoofing, and denial of services (DoS) attacks should be considered carefully. To reduce security threats, many existing link security techniques are often applied within wireless networks.

Applications of MANET

Some specific applications of ad hoc networks include industrial and commercial applications involving cooperative mobile data exchange. There are many existing and future military networking requirements for robust, IP-compliant data services within mobile wireless communication networks, with many of these networks consist of highly dynamic autonomous topology segments. Advanced features of Mobile ad hoc networks, including data rates compatible with multimedia applications global roaming capability, and coordination with other network structures are enabling new applications.

- **Defense applications:** Many defense applications require on the fly communications set-up, and ad hoc/sensor networks are excellent candidates for use in battlefield management.

- **Crisis management applications:** These arise, for example, as a result of natural disasters in which the entire communication infrastructure is in disarray. Restoring communications quickly is essential.

- **Telemedicine:** The paramedic assisting the victim of a traffic accident in a remote location must access medical records (e.g. X-rays) and may need video conference assistance from a surgeon for an emergency intervention. In fact, the paramedic may need to instantaneously relay back to the hospital the victim's X-rays and other diagnostic tests from the site of the accident.

- **Tele-geoprocessing application:** The combination of GPS, GIS (Geographical Information Systems), and high-capacity wireless mobile systems enables a new type of application referred to as tele-geo processing.

- **Virtual Navigation:** A remote database contains the graphical representation of building, streets, and physical characteristics of a large metropolis. They may also "virtually" see the internal layout of

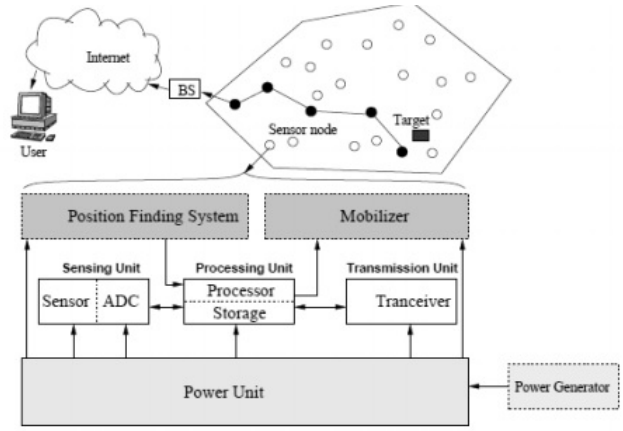
buildings, including an emergency rescue plan, or find possible points of interest.

- **Education via the internet:** educational opportunities available on the internet or remote areas because of the economic infeasibility of providing expensive last-mile wire line internet access in these areas to all subscribers.
- **Vehicular area network:** This a growing and very useful application of adhoc network in providing emergency services and other information. This is equally effective in both urban and rural setup. The basic and exchange necessary data that is beneficial in a given situation.

wireless sensor networks

Wireless sensor networks (WSNs) have the power of distributed communication, computing, and sensing features. They are characterized as infrastructure less, fault tolerant and self-organizing networks which provide opportunities for low-cost, easy-to-apply, rapid and flexible installations in an environment for various applications

The wireless sensor and the sensor node architecture are given in the diagram below-



Characteristics of WSN

The characteristics of WSN are as follows −

- **Resource constraints** − Nodes of WSN are smaller in size and get power from the batteries. It justifies that service provided by the nodes like communication and computation amount of memory is very limited.
- **Communication paradigm** − The data centric feature of WSN explains its data centric nature and justifies that the communication is restricted to nodes.
- **Application specific design** − WSN is application specific i.e. the architecture of WSN is based on application.

- **Node failure and unreliable communication** − Various factors like harsh operating conditions leading to instability, unpredictability, nodal mobility, environmental interferences makes typical WSN nodes to be error-prone.
- **Scalability and density** − The number of nodes in WSNs may be large and densely deployed to a higher degree in various applications.
- **Dynamic Topologies** − Nodes are free to travel randomly at different speeds in few applications and sometimes may fail to operate, to add or to replace. So there can be different network topology.
- **Communication models** − WSNs use different communication models − Flat/ hierarchical /distributed WSNs; or homogeneous/ heterogeneous WSNs.

Operating Environment

The WSNs are mostly deployed in remote and hazardous locations for unattended operations because of their ability to withstand harsh environmental conditions.

Requirements of WSN

The requirements of WSN are explained below:

- **Flexibility** − The architecture of WSN is not fixed. Rather it varies from application to application which justifies that the protocols and algorithms have the characteristics of self-organization.
- **Fault tolerance** − The nodes in WSNs have the capability to sustain the functions carried out in the network even in situations like limited battery power, interference from external sources, failure rate of nodes, harsh environmental conditions.
- **Lifetime** − The two major factors that should be taken into consideration are load balancing and energy saving. These two factors can enhance the lifetime of the WSN architecture as long as possible.
- **Scalability** − The number of nodes in a WSN network can be large. Accordingly WSN architecture and protocols should be designed.
- **Real-time** − The Various capabilities like sensing, processing and communication of WSN are used in various real-world problems so should follow stringent time.
- **Security** − For example in health care data and military data, the data offered by WSN network are private which are sensitive in nature. So security is evident in such architectures.
- **Production cost** − The cost of nodes in WSN network has to be low as once the nodes run out of the energy it has to be replaced by newer nodes.
- **Deployment** − In large-scale WSNs, there is random deployment of nodes whose maintenance and replacements are not practically possible. So there is a huge requirement of re-configuration and re-programming.

- **Dependability** − One can rely on WSN as the architectural design is robust that leads to secure collection of data and reliable delivery with no loss.

A **MANET** is a mobile ad-hoc network that contains wireless links and nodes. It is an infrastructure-less network, and it can change its topology and configure itself on the fly, it can communicate via multiple hops. Whereas a **Wireless Sensor Network (WSN)** is a set of spatially distributed and dedicated sensors that are interlinked via the wireless medium for monitoring and recording the physical conditions of the environment and organizing the collected data at a central location.

Let's look at the similarities between MANET and WSN

1. Both are infrastructure-less, distributed wireless networks
2. Routing Techniques are more or less the same
3. Both are Ad-hoc networks
4. Topology can change over a period
5. Nodes can be operated on a battery
6. Both wireless channels use unlicensed spectrum (cause of interference)

What makes them different?

1. The data rate of MANETs is more than WSN
2. The number of nodes in the WSN is more than MANETs
3. Mobility is very high in MANETs(since nodes are less) than WSN
4. Sensor nodes of WSN are generally static and cooperate together to transfer the sensed data
5. Sensor nodes usually consume less energy than MANET's nodes
6. MANETs are usually close to civilization
7. Public-key cryptography is used in MANETs whereas symmetric key cryptography used in WSNs for security purposes
8. Compared to MANETs, WSNs are smaller, more powerful, and more memory-constrained
9. Mostly, MANETs are used for distributed computing whereas WSNs are used for information gathering from the environment
10. WSNs are more prone to failures than MANETs.

1.6 ENABLING TECHNOLOGIES FOR WIRELESS SENSOR NETWORKS

Building such wireless sensor networks has only become possible with some fundamental advances inenabling technologies.

First technology is the miniaturization of hardware. Smaller feature sizes in chips have driven down the power consumption of the basic components of a sensor node to a level that the constructions of WSNs can be planned. This is particularly relevant to microcontrollers and

memory chips and the radio modems which are responsible for wireless communication have become much more energy efficient. Reduced chip size and improved energy efficiency is accompanied by reduced cost.

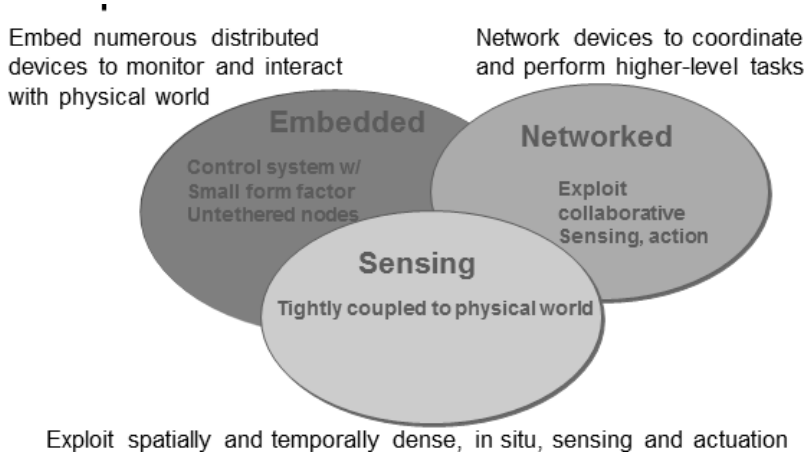


Figure 1.2: Enabling Technologies

Second one is processing and communication and the actual sensing equipment is the third relevant technology. Here, however, it is difficult to generalize because of the vast range of possible sensors.

These three basic parts of a sensor node have to accompanied by power supply. This requires, depending on application, high-capacity batteries that last for long times, that is, have only a negligible self- discharge rate, and that can efficiently provide small amounts of current. Ideally, a sensor node also has a device for energy scavenging, recharging the battery with energy gathered from the environment – solar cells or vibration-based power generation are conceivable options. Such a concept requires the battery to be efficiently chargeable with smallamounts of current, which is not a standard ability. Both batteries and energy scavenging are still objects of ongoing research.

The counterpart to the basic hardware technologies is software. This software architecture on a single node has to be extended to a network architecture, where the division of tasks between nodes, not only on a single node, becomes the relevant question-for example, how to structure interfaces for application programmers. The third part to solve then is the question of how to design appropriate communication protocols.

Building such wireless sensor networks has only become possible with some fundamental advances in enabling technologies. First and foremost among these technologies is the miniaturization of hardware. Smaller feature sizes in chips have driven down the power consumption of the basic components of a sensor node to a level that the constructions of WSNs can be contemplated. This is particularly relevant to microcontrollers and memory chips as such, but also, the radio modems, responsible for wireless communication, have become much more energy efficient. Reduced chip size and improved energy efficiency is accompanied by reduced cost, which is necessary to make redundant deployment of nodes affordable.

Next to processing and communication, the actual sensing equipment is the third relevant technology.

These three basic parts of a sensor node have to accompanied by power supply. This requires, depending on application, high-capacity batteries that last for long times, that is, have only a negligible self-discharge rate, and that can efficiently provide small amounts of current. Ideally, a sensor node also has a device for energy scavenging, recharging the battery with energy gathered from the environment – solar cells or vibration-based power generation are conceivable options. Such a concept requires the battery to be efficiently chargeable with small amounts of current, which is not a standard ability. Both batteries and energy scavenging are still objects of ongoing research.

The counterpart to the basic hardware technologies is software. The first question to answer here is the principal division of tasks and functionalities in a single node – the architecture ofthe operating system or runtime environment. This environment has to support simple retasking, cross-layer information exchange, and modularity to allow for simple maintenance. This software architecture on a single node has to be extended to a network architecture, where the division of tasks between nodes, not only on a single node, becomes the relevant question – for example, how to structure interfaces for application programmers. The third part to solve then is the question of how to design appropriate communication protocols.

1.7 LIST AND REFERENCES

- Wireless Sensor Networks Technology, Protocols, and Applications , Kazem Sohraby, Daniel Minoli and TaiebZnati, John Wiley & Sons, 2007.
- Fundamentals of Wireless Sensor Networks, Theory and Practice, Waltenegus Dargie, Christian Poellabauer , Wiley Series on wireless Communication and Mobile Computing,2011
- Internet Resources

1.8 CONCLUSION

- WSN follows different topologies such as star, tree, mesh, hybrid etc. Hence one can understand pros and cons of these topologies to derive advantages of WSN and disadvantages of WSN. Moreover WSN uses different underlying wireless technologies. Hence one can also refer advantages and disadvantages of Zigbee, Z-wave, WiFi, and WiFi6 et.
- Each such sensor network node typically has many parts: a radio transceiver with an internal antenna or connection to an external antenna, a microcontroller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery or an embedded form of energy harvesting.
- A sensor node might vary in size and size can be a size of a grain of dust.

- Sensor collects the analog data from the physical world and an ADC converts this data to digital data.
- When a large number of sensor nodes are deployed in a large area to cooperatively monitor a physical environment, the networking of these sensor node is equally important
- In flat architecture, the base station sends commands to all the sensor nodes.
- In hierarchical architecture, a group of sensor nodes are formed as a cluster and the sensor nodes transmit data to corresponding cluster heads.
- Wireless sensor network mainly consists of sensor nodes. A wireless sensor network consists of many different components.
- The static parts would be connected to the constant power supply, so that wireless parts can use low power to communicate to them and also nodes can go in the standby mode from time to time.
- A dynamic maintenance approach works as an 'on-the-fly'-based triggering technique that creates a new topology when the current one is no longer optimal.
- MANET stands for Mobile ad-hoc Network also called as wireless ad hoc network or ad hoc wireless network that usually has a routable networking environment on top of a Link Layer ad hoc network.
- The key to achieving a longer lifetime for WSN is to design wireless sensor networks that minimize power consumption of wireless sensor devices, hence the name "low power".
- Challenges in wireless sensor node in various ways for an application.
- The major applications of WSNs
- The goods are loaded from a warehouse to a freight vehicle.
- Their typical usage is to gather information about their environment via sensors, to potentially pre-process these data, and to finally transmit them.
- Characteristics of an environmental monitoring system

2

SENSOR NODE HARDWARE AND NETWORK ARCHITECTURE

Unit Structure

2.0 OBJECTIVES

In WSN, the main task of a sensor node is to sense data and sends it to the base station in multi hop environment for which routing path is essential. For computing the routing path from the source node to the base station there is huge numbers of proposed routing protocols exist (Sharma et al., 2011).

Currently, WSN (Wireless Sensor Network) is the most standard services employed in commercial and industrial applications, because of its technical development in a processor, communication, and low-power usage of embedded computing devices. The wireless sensor network architecture is built with nodes that are used to observe the surroundings like temperature, humidity, pressure, position, vibration, sound, etc. These nodes can be used in various real-time applications to perform various tasks like smart detecting, a discovery of neighbor nodes, data processing and storage, data collection, target tracking, monitor and controlling, synchronization, node localization, and effective routing between the base station and nodes. Presently, WSNs are beginning to be organized in an enhanced step. It is not awkward to expect that in 10 to 15 years that the world will be protected with WSNs with entree to them via the Internet. This can be measured as the Internet becoming a physical n/w. This technology is thrilling with infinite potential for many application areas like medical, environmental, transportation, military, entertainment, homeland defense, crisis management, and also smart spaces.

2.1 SENSOR NODE HARDWARE AND NETWORK ARCHITECTURE

2.1.1 Key Definitions Of Sensor Networks :

Definition: A Sensor Network is composed of a large number of sensor nodes, which are tightly positioned either inside the phenomenon or very close to it.

Sensor networks have the contribution from signal processing, networking and protocols, databases and information management, distributed algorithms, and embedded systems and architecture.

A wireless sensor network (WSN) can be defined as a network of low-size and low-complex devices denoted as nodes that can sense the environment and communicate the information gathered from the monitored field through wireless links.

The following are the Key terms and concepts that will be used in sensor network development techniques.

- **Sensor:** A transducer that converts a physical phenomenon such as heat, light, sound, or motion into electrical or other signals that may be further operated by other apparatus.
- **Sensor node:** A basic unit in a sensor network, with on-board sensors, processor, memory, wireless modem, and power supply. It is often abbreviated as node. When a node has only a single sensor on board, the node is sometimes referred as a sensor.
- **Network topology:** A connectivity graph where nodes are sensor nodes and edges are communication links. In a wireless network, the link represents a one-hop connection, and the neighbors of a node are those within the radio range of the node.
- **Routing:** The process of determining a network path from a packet source node to its destination.
- **Date-centric:** Approaches that name, route, or access a piece of data via properties, such as physical location, that are external to a communication network. This is to be contrasted with addresscentric approaches which use logical properties of nodes related to the network structure.
- **Geographic routing:** Routing of data based on geographical features such as locations or regions. This is an example of datecentric networking.
- **In-network:** A style of processing in which the data is processed and combined near where the data is generated.
- **Collaborative processing:** Sensors cooperatively processing data from multiple sources in order to serve a high-level task. This typically requires communication among a set of nodes.
- **State:** A snapshot about a physical environment (e.g., the number of signal sources, their locations or spatial extent, speed of movement), or a snapshot of the system itself (e.g.,the network state).
- **Uncertainty:** A condition of the information caused by noise in sensor measurements, or lack of knowledge in models. The uncertainty affects the system‘s ability to estimate the state accurately and must be carefully modeled. Because of the ubiquity of uncertainty in the data, many sensor network estimation problems are cast in a statistical framework. For example, one may use a covariance matrix to characterize the uncertainty in a Gaussian-like process or more general

probability distributions for non-Gaussian processes.Task: Either high-level system tasks which may include sensing, communication, processing, and resource allocation, or application tasks which may include detection, classification, localization, or tracking.

- **Detection:** The process of discovering the existence of a physical phenomenon. A threshold- based detector may flag a detection whenever the signature of a physical phenomenon is determined to be significant enough compared with the threshold.
- **Classification:** The assignment of class labels to a set of physical phenomena being observed.
- **Localization and tracking:** The estimation of the state of a physical entity such as a physical phenomenon or a sensor node from a set of measurements. Tracking produces a series of estimates over time.
- **Value of information or information utility:** A mapping of data to a scalar number, in the context of the overall system task and knowledge. For example, information utility of a piece of sensor data may be characterized by its relevance to an estimation task at hand and computed by a mutual information function.
- **Resource:** Resources include sensors, communication links, processors, on-board memory, and node energy reserves. Resource allocation assigns resources to tasks, typically optimizing some performance objective.
- **Sensor tasking:** The assignment of sensors to a particular task and the control of sensor state (e.g., on/off, pan/tilt) for accomplishing the task.
- **Node services:** Services such as time synchronization and node localization that enable applications to discover properties of a node and the nodes to organize themselves into a useful network.
- **Data storage:** Sensor information is stored, indexed, and accessed by applications. Storage may be local to the node where the data is generated, load-balanced across a network, or anchored at a few points (warehouses).
- **Embedded operating system (OS):** The run-time system support for sensor network applications. An embedded OS typically provides an abstraction of system resources and a set of utilities.
- **System performance goal:** The abstract characterization of system properties. Examples include scalability, robustness, and network longevity, each of which may be measured by a set of evaluation metrics.
- **Evaluation metric:** A measurable quantity that describes how well the system is performing on some absolute scale. Examples include packet loss (system), network dwell time (system), track loss (application), false alarm rate (application), probability of correct association

(application), location error (application), or processing latency (application/system). An evaluation method is a process for comparing the value ofapplying the metrics on an experimental system with that of some other benchmark system.

2.2 SINGLE-NODE ARCHITECTURE

2.2.1 Hardware Components & Design Constraints

2.2.2 HARDWARE COMPONENTS:

Choosing the hardware components for a wireless sensor node, obviously the applications has to consider size, costs, and energy consumption of the nodes. A basic sensor node comprises five main components such as Controller, Memory, Sensors and Actuators, Communication devices and Power supply Unit.

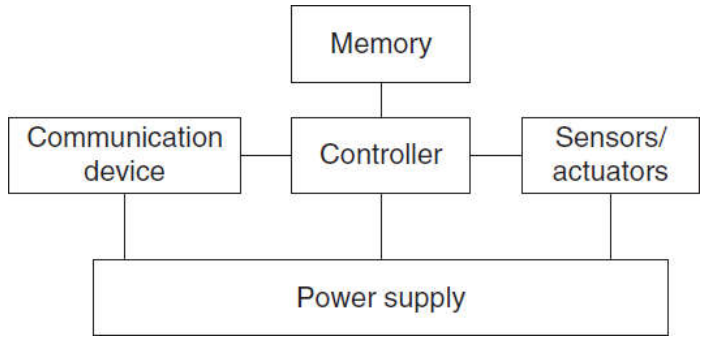


Figure 1.3: Sensor node Hardware components

2.2.3 Controller:

A controller to process all the relevant data, capable of executing arbitrary code. The controller is the core of a wireless sensor node. It collects data from the sensors, processes this data, decides when and where to send it, receives data from other sensor nodes, and decides on the actuator‘s behavior. Ithas to execute various programs, ranging from time- critical signal processing and communication protocolsto application programs; it is the Central Processing Unit (CPU) of the node.

For General-purpose processors applications microcontrollers are used. These are highly overpowered, and their energy consumption is excessive. These are used in embedded systems. Some of the key characteristics of microcontrollers are particularly suited to embedded systems are their flexibility in connecting with other devices like sensors and they are also convenient in that they often have memory built in.

A specialized case of programmable processors are Digital Signal Processors (DSPs). They are specifically geared, with respect to their architecture and their instruction set, for processing large amounts of vectorial data, as is typically the case in signal processing applications. In a wireless sensor node, such a DSP could be used to process data coming

from a simple analog, wireless communication device to extract a digital data stream. In broadband wireless communication, DSPs are an appropriate and successfully used platform.

An FPGA can be reprogrammed (or rather reconfigured) —in the field to adapt to a changing set of requirements; however, this can take time and energy – it is not practical to reprogram an FPGA at the same frequency as a microcontroller could change between different programs.

An ASIC is a specialized processor, custom designed for a given application such as, for example, high-speed routers and switches. The typical trade-off here is loss of flexibility in return for a considerably better energy efficiency and performance. On the other hand, where a microcontroller requires software development, ASICs provide the same functionality in hardware, resulting in potentially more costly hardware development.

Examples: Intel Strong ARM, Texas Instruments MSP 430, Atmel ATmega.

2.2.4 Memory:

Some memory to store programs and intermediate data; usually, different types of memory are used for programs and data. In WSN there is a need for Random Access Memory (RAM) to store intermediate sensor readings, packets from other nodes, and so on. While RAM is fast, its main disadvantageis that it loses its content if power supply is interrupted. Program code can be stored in Read-Only Memory (ROM) or, more typically, in Electrically Erasable Programmable Read-Only Memory (EEPROM) or flash memory (the later being similar to EEPROM but allowing data to be erased or written in blocks instead ofonly a byte at a time). Flash memory can also serve as intermediate storage of data in case RAM is insufficient or when the power supply of RAM should be shut down for some time.

2.2.5 Communication Device:

Turning nodes into a network requires a device for sending and receiving information over a wireless channel.

Choice of transmission medium: The communication device is used to exchange data between individual nodes. In some cases, wired communication can actually be the method of choice and is frequently applied in many sensor networks. The case of wireless communication is considerably more interesting because it include radio frequencies. Radio Frequency (RF)- based communication is by far the most relevant one as it best fits the requirements of most WSN applications.

Transceivers: For Communication, both transmitter and receiver are required in a sensor node to convert a bit stream coming from a microcontroller and convert them to and from radio waves. For two tasks a combined device called transceiver is used.

Transceiver structure has two parts as Radio Frequency (RF) front end and the baseband part.

1. The radio frequency front end performs analog signal processing in the actual radio frequency Band.

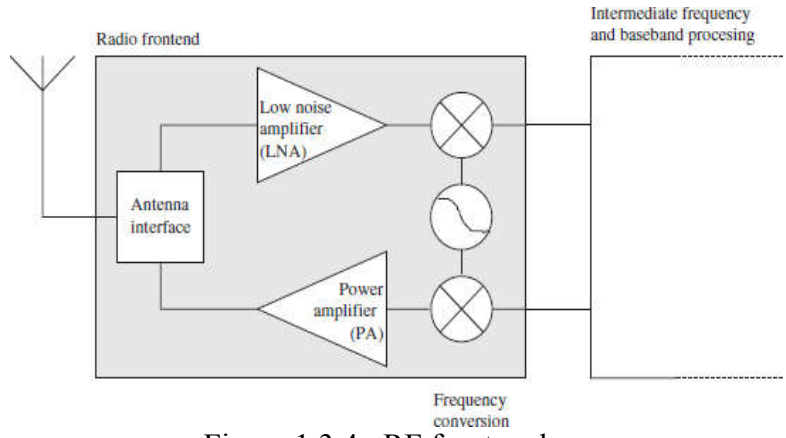


Figure 1.3.4 : RF front end

2. The baseband processor performs all signal processing in the digital domain and communicates with a

sensor node‘s processor or other digital circuitry.

a. The Power Amplifier (PA) accepts upconverted signals from the IF or baseband part and amplifiesthem for transmission over the antenna.

b. The Low Noise Amplifier (LNA) amplifies incoming signals up to levels suitable for further processing without significantly reducing the SNR. The range of powers of the incoming signals

varies from very weak signals from nodes close to the reception boundary to strong signals from nearby nodes; this range can be up to 100 dB.

c. Elements like local oscillators or voltage-controlled oscillators and mixers are used for frequency conversion from the RF spectrum to intermediate frequencies or to the baseband. The incoming signal at RF frequencies f_{RF} is multiplied in a mixer with a fixed- frequency signal from the local oscillator (frequency f_{LO}). The resulting intermediate-frequency signal has frequency $f_{LO} - f_{RF}$. Depending on the RF front end architecture, other elements like filters are also present.

Transceiver tasks and characteristics:

□ Service to upper layer: A receiver has to offer certain services to the upper layers, most notably to the Medium Access Control (MAC) layer. Sometimes, this service is packet oriented; sometimes, a

transceiver only provides a byte interface or even only a bit interface to the microcontroller.

- Power consumption and energy efficiency: The simplest interpretation of energy efficiency is the energy required to transmit and receive a single bit.
- Carrier frequency and multiple channels: Transceivers are available for different carrier frequencies; evidently, it must match application requirements and regulatory restrictions.
- State change times and energy: A transceiver can operate in different modes: sending or receiving, use different channels, or be in different power-safe states.
- Data rates: Carrier frequency and used bandwidth together with modulation and coding determine the gross data rate.
- Modulations: The transceivers typically support one or several of on/off-keying, ASK, FSK, or similar modulations.
- Coding: Some transceivers allow various coding schemes to be selected.
- Transmission power control: Some transceivers can directly provide control over the transmission power to be used; some require some external circuitry for that purpose. Usually, only a discrete number of power levels are available from which the actual transmission power can be chosen. Maximum output power is usually determined by regulations.
- Noise figure: The noise figure NF of an element is defined as the ratio of the Signal-to- Noise Ratio (SNR) ratio SNR_I at the input of the element to the SNR ratio SNR_O at the element's output: $NF = \frac{SNR_I}{SNR_O}$. It describes the degradation of SNR due to the element's operation and is typically given in dB: $NF\,dB = SNR_I\,dB - SNR_O\,dB$.
- Gain: The gain is the ratio of the output signal power to the input signal power and is typically given in dB. Amplifiers with high gain are desirable to achieve good energy efficiency.
- Power efficiency: The efficiency of the radio front end is given as the ratio of the radiated power to the overall power consumed by the front end; for a power amplifier, the efficiency describes the ratio of the output signal's power to the power consumed by the overall power amplifier.
- Receiver sensitivity: The receiver sensitivity (given in dBm) specifies the minimum signal power at the receiver needed to achieve a prescribed E_b/N_0 or a prescribed bit/packet error rate.
- Range: The range of a transmitter is clear. The range is considered in

absence of interference; it evidently depends on the maximum transmission power, on the antenna characteristics.

- Blocking performance: The blocking performance of a receiver is its achieved bit error rate in the presence of an interferer.
- Out of band emission: The inverse to adjacent channel suppression is the out of band emission of a transmitter. To limit disturbance of other systems, or of the WSN itself in a multichannel setup, the transmitter should produce as little as possible of transmission power outside of its prescribed bandwidth, centered around the carrier frequency.Carrier sense and RSSI: In many medium access control protocols, sensing whether the wireless channel, the carrier, is busy (another node is transmitting) is a critical information. The receiver has to be able to provide that information. the signal strength at which an incoming data packet has been received can provide useful information a receiver has to provide this information in the Received Signal Strength Indicator (RSSI).
- Frequency stability: The frequency stability denotes the degree of variation from nominal center frequencies when environmental conditions of oscillators like temperature or pressure change.
- Voltage range: Transceivers should operate reliably over a range of supply voltages.

Otherwise, inefficient voltage stabilization circuitry is required.

2.2.6 Sensors and actuators:

The actual interface to the physical world: devices that can observe or control physical parameters of the environment.

Sensors can be roughly categorized into three categories as

2.2.6.1 Passive, omnidirectional sensors:

These sensors can measure a physical quantity at the point of the sensor node without actually manipulating the environment by active probing – in this sense, they are passive. Moreover, some of these sensors actually are self-powered in the sense that they obtain the energy they need from the environment – energy is only needed to amplify their analog signal.

2.2.6.2 Passive, narrow-beam sensors These sensors are passive as well, but have a well- defined notion of direction of measurement.

2.2.6.3 Active sensors This last group of sensors actively probes the environment, for example, a sonar or radar sensor or some types of seismic sensors, which generate shock waves by small explosions. These are quite specific – triggering an explosion is certainly not a lightly undertaken action – and require quite special attention.

Actuators: Actuators are just about as diverse as sensors, yet for the purposes of designing a WSN that converts electrical signals into physical phenomenon.

2.2.7 Power supply:

As usually no tethered power supply is available, some form of batteries are necessaryto provide energy. Sometimes, some form of recharging by obtaining energy from the environment is available as well (e.g. solar cells). There are essentially two aspects: Storing energy and Energy scavenging. Storing energy: Batteries

Primary batteries			
Chemistry	Zinc-air	Lithium	Alkaline
Energy (J/cm^3)	3780	2880	1200
Secondary batteries			
Chemistry	Lithium	NiMHd	NiCd
Energy (J/cm^3)	1080	860	650

2.2.7.1 Traditional batteries:

The power source of a sensor node is a battery, either non- rechargeable (—primary batteries‖) or, if an energy scavenging device is present on the node, also rechargeable(—secondary batteries‖).

TABLE 1.1: Energy densities for various primary and secondary battery types

Upon these batteries the requirements are

2.2.7.2 Capacity:

They should have high capacity at a small weight, small volume, and low price.

The main metric is energy per volume, J/cm^3.

2.2.7.3 Capacity under load:

They should withstand various usage patterns as a sensor node can consume quite different levels of power over time and actually draw high current in certain operation modes.

2.2.7.4 Self-discharge:

Their self-discharge should be low. Zinc-air batteries, for example, have only a very short lifetime (on the order of weeks).

2.2.7.5 Efficient recharging:

Recharging should be efficient even at low and intermittently available recharge power.

2.2.7.6 Relaxation:

Their relaxation effect – the seeming self-recharging of an empty or almost empty battery when no current is drawn from it, based on chemical diffusion processes within the cell – should be clearly understood. Battery lifetime and usable capacity is considerably extended if this effect is leveraged.

2.2.7.7 DC–DC Conversion:

Unfortunately, batteries alone are not sufficient as a direct power source for a sensor node. One typical problem is the reduction of a battery‘s voltage as its capacity drops. A DC – DC converter can be used to overcome this problem by regulating the voltage delivered to the node‘s circuitry. To ensure a constant voltage even though the battery‘s supply voltage drops, the DC – DC converter has to draw increasingly higher current from the battery when the battery is already becoming weak, speeding up battery death. The DC – DC converter does consume energy for its own operation, reducing overall efficiency.

Energy scavenging: Depending on application, high capacity batteries that last for long times, that is, have only a negligible self-discharge rate, and that can efficiently provide small amounts of current. Ideally, a sensor node also has a device for energy scavenging, recharging the battery with energy gathered from the environment – solar cells or vibration-based power generation are conceivable options.

2.2.7.8 Photovoltaics:

The well-known solar cells can be used to power sensor nodes. The available power depends on whether nodes are used outdoors or indoors, and on time of day and whether for outdoor usage. The resulting power is somewhere between 10 $\mu W/cm^2$ indoors and 15 mW/cm^2 outdoors. Single cells achieve a fairly stable output voltage of about 0.6 V (and have therefore to be used in series) as long as the drawn current does not exceed a critical threshold, which depends on the light intensity. Hence, solar cells are usually used to recharge secondary batteries.

2.2.7.9 Temperature gradients:

Differences in temperature can be directly converted to electrical energy.

2.2.7.10 Vibrations:

One almost pervasive form of mechanical energy is vibrations: walls or windows in buildings are resonating with cars or trucks passing in the streets, machinery often has low frequency vibrations. both amplitude and frequency of the vibration and ranges from about 0.1 $\mu W/cm^3$ up to 10, 000 $\mu W/cm^3$ for some extreme cases. Converting vibrations to electrical energy can be undertaken by various means, based on electromagnetic, electrostatic, or piezoelectric principles.

2.2.7.11 Pressure variations:

Somewhat akin to vibrations, a variation of pressure can also be used as a power source.

2.2.7.12 Flow of air/liquid:

Another often-used power source is the flow of air or liquid in wind mills orturbines. The challenge here is again the miniaturization, but some of the work on millimeter scale MEMS gas turbines might be reusable.

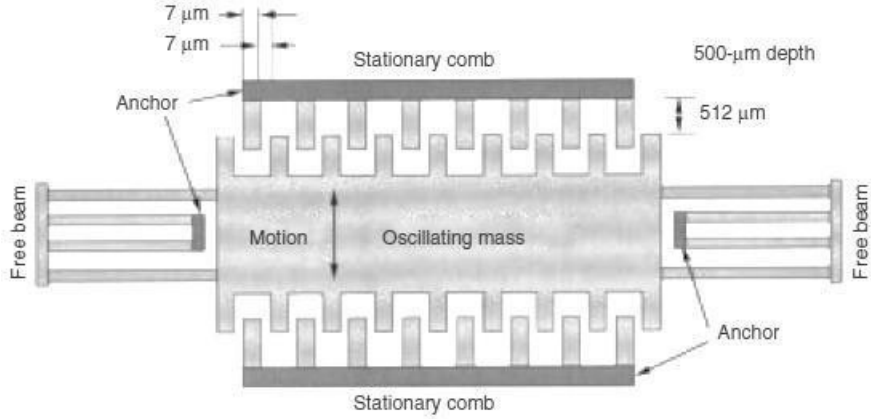


Figure 1.5 A MEMS device for converting vibrations to electrical energy, based on a variable

capacitor

Energy source	Energy density
Batteries (zinc-air)	1050–1560 mWh/cm^3
Batteries (rechargeable lithium)	300 mWh/cm^3 (at 3–4 V)
Energy source	**Power density**
Solar (outdoors)	15 mW/cm^2 (direct sun) 0.15 mW/cm^2 (cloudy day)
Solar (indoors)	0.006 mW/cm^2 (standard office desk) 0.57 mW/cm^2 (<60 W desk lamp)
Vibrations	0.01–0.1 mW/cm^3
Acoustic noise	$3 \cdot 10^{-6}$ mW/cm^2 at 75 dB $9,6 \cdot 10^{-4}$ mW/cm^2 at 100 dB
Passive human-powered systems	1.8 mW (shoe inserts)
Nuclear reaction	80 mW/cm^3, 10^6 mWh/cm^3

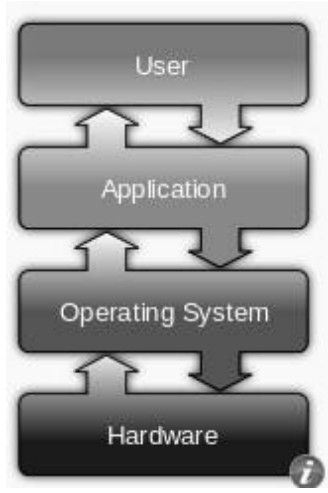
TABLE 1.2: Comparison of energy sources

2.3 OPERATING SYSTEMS AND EXECUTION ENVIRONMENTS

- ✓ An operating system (OS) is system software that manages computer hardware and software resources and provides common services for computer programs.

- ✓ For hardware functions such as input and output and memory allocation, the operating system acts as an intermediary between programs and the computer hardware.
- ✓ An embedded system is some combination of computer hardware and software, either fixed in capability or programmable, that is specifically designed for a particular function.
- ✓ Embedded operating systems are designed to be used in embedded computer systems. They are able to operate with a limited number of

resources. They are very compact and extremely efficient by design.



2.4 INTRODUCTION TO TINYOS AND NESC

The use of an event-based programming model as the only feasible way to support the concurrency required for sensor node software while staying within the confined resources and running on top of the simple hardware provided by these nodes. The open question is how to harness the power of this programming model without getting lost in the complexityof many individual state machines sending each other events. In addition, modularity should be supported to easily exchange one state machine against another. The operating system TinyOS, along with the programming language nesC, addresses these challenges.

TinyOS supports modularity and event-based programming by the concept of components. A component contains semantically related functionality, for example, for handling a radio interface or for computing routes. Such a component comprises the required state information in a frame, the program code for normal tasks, and handlers for events and commands. Both events and commands are exchanged between different components. Components are arranged hierarchically, from low-level components close to the hardware to high-level components making up the actual application. Events originate in the hardware and pass upward from low-level to high-level components; commands, on the other hand, are passed from high-level to low-level components.

Figure shows a timer component that provides a more abstract version of a simple hardware time. It understands three commands ("init", "start", and "stop") and can handle one event("fire") from another component, for example, a wrapper component around a hardware timer. It issues "setRate" commands to this component and can emit a "fired" event itself.

The important thing to note is that, in staying with the event-based paradigm, both command and event handlers must run to conclusion; they are only supposed to perform very simple triggering duties. In particular, commands must not block or wait for an indeterminate amount of time; they are simply a request upon which some task of the hierarchically lower component has to act. Similarly, an event handler only leaves information in its component's frame and arranges for a task to be executed later; it can also send commands to other components or directly report an event further up.

The actual computational work is done in the tasks. In TinyOS, they also have to run to completion, but can be interrupted by handlers. The advantage is twofold: there is no need for stack management and tasks are atomic with respect to each other. Still, by virtue of being triggered by handlers, tasks are seemingly concurrent to each other.

The arbitration between tasks – multiple can be triggered by several events and are ready to execute – is done by a simple, power-aware First In First Out (FIFO) scheduler, which shuts the node down when there is no task executing or waiting.

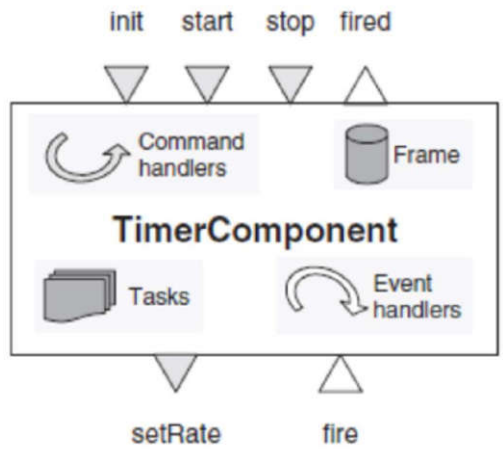


Figure 2.9 Example Timer component (adapted from references [285, 353])

With handlers and tasks all required to run to completion, it is not clear how a component could obtain feedback from another component about a command that it has invoked there – for example, how could an Automatic Repeat Request (ARQ) protocol learn from the MAC protocol whether a packet had been sent successfully or not? The idea is to split invoking such a request and the information about answers into two phases: The first phase is the sending of the command, the second is an explicit information about the outcome of the operation, delivered by a

separate event. This split-phase programming approach requires for each command a matching event but enables concurrency under the constraints of run-to- completion semantics – if no confirmation for a command is required, no completion event is necessary.

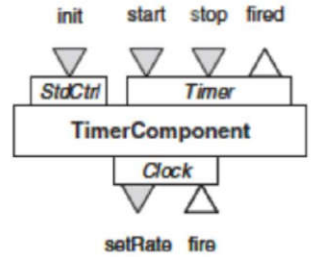


Figure 2.10 Organizing the Timer component using interfaces [285, 353]

Listing 2.1: Defining modules and interfaces [285]

```
interface StdCtrl {
  command result_t init();
}

interface Timer {
  command result_t start (char type, uint32_t interval);
  command result_t stop ();
  event result_t fired();
}

interface Clock {
  command result_t setRate (char interval, char scale);
  event result_t  fire ();
}

module TimerComponent {
  provides {
    interface StdCtrl;
    interface Timer;
  }
  uses interface Clock as Clk;
}
```

Having commands and events as the only way of interaction between components (the frames of components are private data structures), and especially when using split-phase programming, a large number of commands and events add up in even a modestly large program. Hence, an abstraction is necessary to organize them. As a matter of fact, the set of commands that a component understands and the set of events that a component may emit are its interface to the components of a hierarchically higher layer; looked at it the other way around, a component can invoke certain commands at its lower component and receive certain events from it. Therefore, structuring commands and events that belong together forms an interface between two components.The nesC language formalizes thi s intuition by allowing a programmer to define interface types that define commands and events that belong together. This allows to easily express split-phase programming style by putting commands and their corresponding completion events into the same interface. Components then provide certain interfaces to their users and in turn use other interfaces from underlying components.

Figure shows how the Timer component of the previous example can be reorganized into using a clock interface and providing two interfaces StdCtrl and Timer. The corresp onding nesC code is shown in Listing 1.

Note that the component TimerComponent is defined here as a module since it is a primitive component, directly containing handlers and tasks.

Such primitive components or modules can be combined into larger configurations by simply "wiring" appropriate interfaces together. For this wiring to take place, only components that have the correct interface types can be plugged together (this is checked by the compiler). Figure shows how the TimerComponent and an additional component HWClock can be wired together to form a new component CompleteTimer, exposing only the StdCtrl and Timer interfaces to the outside; Listing 2 shows the correspondin g nesC code. Note that both modules and configurations are components.

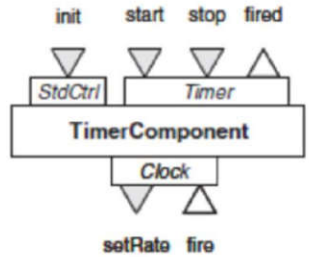


Figure 2.11 Building a larger configuration out of two components [285, 353]

Listing 2.2: Wiring components to form a configuration [285]

```
configuration CompleteTimer {
  provides {
    interface StdCtrl;
    interface Timer;
  }
  implementation {
    components TimerComponent, HWClock;
    StdCtrl = TimerComponent.HWClock;
    Timer = TimerComponent.Timer;
    TimerComponent.Clk = HWClock.Clock;
  }
}
```

Using these component definition, implementation, and connection concepts, TinyOS and

nesC together form a powerful and relatively easy to use basis to implement both coreoperating system functionalities as well as communication protocol stacks and application functions. Programmers do use these paradigms and arrive at relatively small, highly specialized components that are then combined as needed, proving the modularity claim. Also, code size and memory requirements are quite small.

Overall, TinyOS can currently be regar ded as the standard implementation platform for WSNs. It is also becoming available for an increasing number of platforms other than the original "motes" on which it had been developed. On top of the TinyOS operating system, a vast range of extensions, protocols, and applications have been developed. A virtual machine concept describes on top of TinyOS that provides a high-level interface to concisely represent programs; it is particularly beneficial

for over-the-air reprogramming and retasking of an existing network. Conceiving of the sensor network as a relational database is made possible by the TinyDB project.

Other examples

Apart from TinyOS, there are a few other execution environments or operating systems for WSN nodes. One example is Contiki10, which has been ported to various hardware platforms and actually implements a TCP/IP stack on top of a platform with severely restricted resources. Other examples are ecos and the Mantis project.

Some Examples of Sensor Nodes

There are quite a number of actual nodes available for use in wireless sensor network research and development. Again, depending on the intended application scenarios, they have to fulfill quite different requirements regarding battery life, mechanical robustness of the node's housing, size, and so on.

The "Mica Mote" family

Starting in the late 1990s, an entire family of nodes has evolved out of research projects at the University of California at Berkeley, partially with the collaboration of Intel, over the years. They are commonly known as the Mica motes11, with different versions (Mica, Mica2, Mica2Dot) having been designed. They are commercially available via the company Crossbow12 in different versions and different kits. TinyOS is the usually used operating system for these nodes.All these boards feature a microcontroller belonging to the Atmel family, a simple radio modem (usually a TR 1000 from RFM), and various connections to the outside. In addition, it is possible to connect additional "sensor boards" with, for example, barometric or humidity sensors, to the node as such, enabling a wider range of applications and experiments. Also, specialized enclosures have been built for use in rough environments, for example, for monitoring bird habitats. Sensors are connected to the controller via an I2C bus or via SPI, depending on the version.

2.5 NETWORK ARCHITECTURE

The architecture of wireless sensor networks draws upon many sources. Historically, a lot of related work has been done in the context of self-organizing, mobile, ad hoc networks. While these networks are intended for different purposes, they share the need for a decentralized, distributed form of organization. From a different perspective, sensor networks are related to real-time computing and even to some concepts from peer-to-peer computing, active networks, and mobile agents/swarm intelligence.

NETWORK ARCHITECTURE:

It introduces the basic principles of turning individual sensor nodes into a wireless sensor network. In this optimization goals of how a network should function are discussed as

2.5.1. Sensor network scenarios

2.5.2 Optimization goals and figures of merit

2.5.3 Gateway concepts

2.5.1 Sensor Network Scenarios :

2.5.2 Types of sources and sinks:

Source is any unit in the network that can provide information (sensor node). A sink is the unit where information is required, it could belong to the sensor network or outside this network to interact with another network or a gateway to another larger Internet. Sinks are illustrated by Figure 1.11, showing sources and sinks in direct communication.

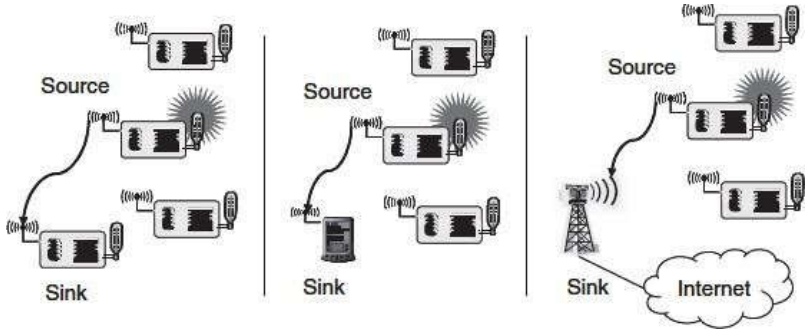


Figure 1.11 Three types of sinks in a very simple, single-hop sensor network

2.5.3 Single-hop versus multi-hop networks:

Because of limited distance the direct communication between source and sink is not always possible. In WSNs, to cover a lot of environment the data packets taking multi hops from source to the sink. To overcome such limited distances it better to use relay stations, The data packets taking multi hops from source to the sink as shown in Figure 1.12, Depending on the particular application of having an intermediate sensor node at the right place is high.

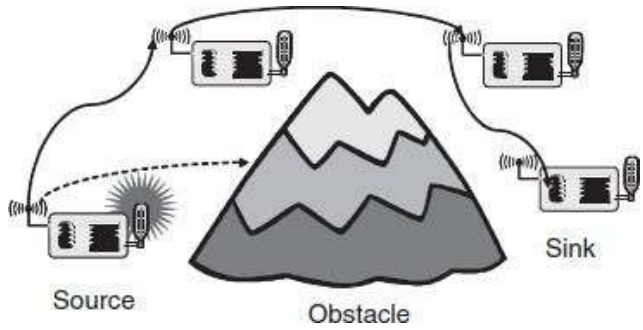


Figure 1.12 Multi-hop networks: As direct communication is impossible because of distance and/or

obstacles

Multi-hopping also to improves the energy efficiency of communication as it consumes less energy to use relays instead of direct communication, the radiated energy required for direct communication over a distance d is cd^{α} (c some constant, $\alpha \geq 2$ the path loss coefficient) and using a relay at distance d/2 reduces this energy to $2c(d/2)^{\alpha}$

This calculation considers only the radiated energy. It should be pointed out that only multi- hop networks

operating in a store and forward fashion are considered here. In such a network, a node has to correctly receive a packet before it can forward it somewhere. Cooperative relaying (reconstruction in case of erroneous packet reception) techniques are not considered here.

2.5.4 Multiple sinks and sources:

2.5.4.1 In many cases, multiple sources and multiple sinks present. Multiple sources should send information to multiple sinks. Either all or some of the information has to reach all or some of the sinks. This is illustrated in figure 1.13.

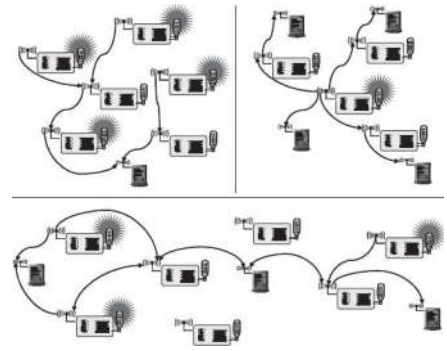


Figure 1.13 Multiple sources and/or multiple sinks.

Note how in the scenario in the lower half, both sinks and active sources are used to forward data to the sinks at the left and right end of the network.

Three types of mobility: In the scenarios discussed above, all participants were stationary. But one of the main virtues of wireless communication is its ability to support mobile participants In wireless sensor networks, mobility can appear in three main forms

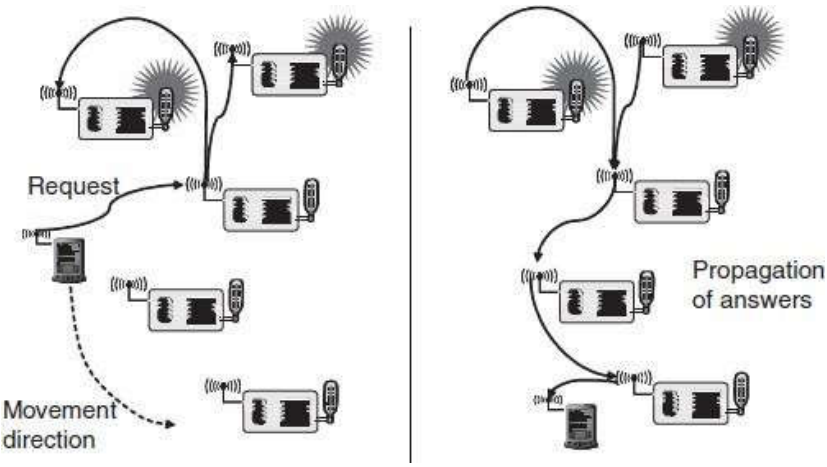
a. Node mobility

b. Sink mobility

c. Event mobility

(a) Node Mobility: The wireless sensor nodes themselves can be mobile. The meaning of such mobility is highly application dependent. In examples like environmental control, node mobility should not happen; in livestock surveillance (sensor nodes attached to cattle, for example), it is the common rule. In the face of node mobility, the network has to reorganize to function correctly.

(b) Sink Mobility: The information sinks can be mobile. For example, a human user requested information via a PDA while walking in an intelligent building. In a simple case, such a requester can interact with the WSN at one point and complete its interactions before moving on, In many cases, consecutive



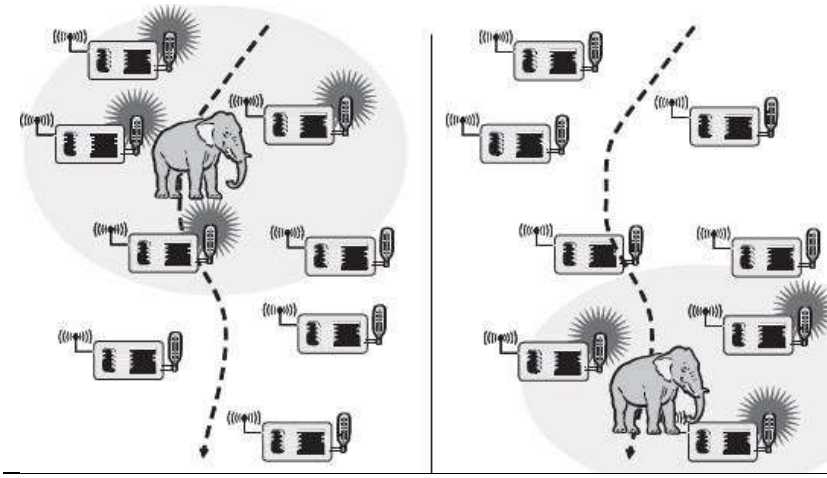
interactions can be treated as separate, unrelated requests.

Figure 1.14

Sink mobility: A mobile sink moves through a sensor network as information is being retrieved on its behalf(c) Event Mobility: In tracking applications, the cause of the events or the objects to be tracked can be

mobile. In such scenarios, it is (usually) important that the observed event is covered by a sufficient number of sensors at all time. As the event source moves through the network, it is accompanied by an area of activity within the network – this has been called the frisbee model. This notion is described by Figure 1.15, where the

task is to detect a moving elephant and to observe it as it moves around Figure 1.15 Area of sensor nodes detecting an event – an elephant– that moves through the network along with the event source (dashed line indicate the elephant's trajectory; shaded ellipse the activity area following or even preceding the elephant)



2.6 OPTIMIZATION GOALS AND FIGURES OF MERIT

For all WSN scenarios and application types have to face the challenges such as

- ✓ How to optimize a network and How to compare these solutions?
- ✓ How to decide which approach is better?
- ✓ How to turn relatively inaccurate optimization goals into measurable figures of merit? For all the above questions the general answer is obtained from

- ❖ Quality of service
- ❖ Energy efficiency
- ❖ Scalability
- ❖ Robustness

2.6.1 Quality of service:

WSNs differ from other conventional communication networks in the type of service they offer. These networks essentially only move bits from one place to another. Some generic possibilities are

Event detection/reporting probability- The probability that an event that actually occurred is not detected or not reported to an information sink that is interested in such an event For example, not reporting a fire alarm to a surveillance station would be a severe shortcoming.

Event classification error- If events are not only to be detected but also to be classified, the error in classification must be small

Event detection delay -It is the delay between detecting an event and reporting it to any/all interested sinks

Missing reports -In applications that require periodic reporting, the

probability of undelivered reportsshould be small

Approximation accuracy- For function approximation applications, the average/maximum absolute or relative error with respect to the actual function.

Tracking accuracy Tracking applications must not miss an object to be tracked, the reported position should be as close to the real position as possible, and the error should be small.

2.6.2 Energy efficiency:

Energy efficiency should be optimization goal. The most commonly consideredaspects are:

Energy per correctly received bit-How much energy is spent on average to transport one bit ofinformation (payload) from the transmitter to the receiver.

Energy per reported (unique) event-What is the average energy spent to report one event

Delay/energy trade-offs-–urgent events increases energy investment for a speedy reporting events. Here, the trade-off between delay and energy overhead is interesting

Network lifetime The time for which the network is operational

Time to first node death-When does the first node in the network run out of energy or fail and stopoperating?

Network half-life-When have 50 % of the nodes run out of energy and stopped operating

Time to partition-When does the first partition of the network in two (or more) disconnected partsoccur?

Time to loss of coverage the time when for the first time any spot in the deployment region is nolonger covered by any node‘s observations.

Time to failure of first event notification A network partition can be seen as irrelevant if theunreachable part of the network does not want to report any events in the first place.

Scalability: The ability to maintain performance characteristics irrespective of the size of the network is referred to as scalability. With WSN potentially consisting of thousands of nodes, scalability is an obviously essential requirement. The need for extreme scalability has direct consequences for the protocol design. Often, a penalty in performance or complexity has to be paid for small networks. Architectures and protocols should implement appropriate scalability support rather than trying to be as scalable as possible. Applications with a few dozen nodes might admit more-efficient solutions than applications with thousands of nodes.

Robustness: Wireless sensor networks should also exhibit an appropriate robustness. They should not fail just because a limited number of nodes run out of energy, or because their environment changes and severs existing radio links between two nodes. If possible, these failures have to be compensated by finding other routes.

Gate way concepts:

For practical deployment, a sensor network only concerned with itself is insufficient.

The network rather has to be able to interact with other information devices for example to read the temperature sensors in one‘s home while traveling and accessing the Internet via a wireless.

Wireless sensor networks should also exhibit an appropriate robustness

They should not fail just because of a limited number of nodes run out of energy or because of their environment changes and breaks existing radio links between two nodes.

If possible, these failures have to be compensated by finding other routes.

Figure 1.16 shows this networking scenario, The WSN first of all has to be able to exchange data with such a

mobile device or with some sort of gateway, which provides the physical connection to the Internet. The WSN support standard wireless communication technologies such as IEEE 802.11. The design of gateways becomes much more challenging when considering their logical design. One option is to regard a gateway as a simple router between Internet and sensor network.

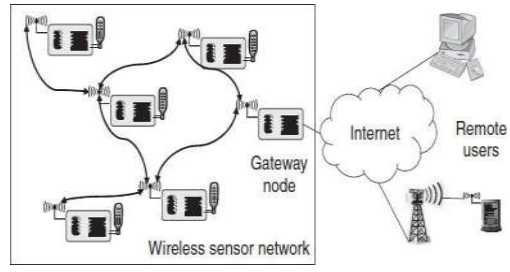


Figure 1.16 A wireless sensor network with gateway node, enabling access to remote clients via the

Internet

WSN to Internet communication: Assume that the initiator of a WSN – Internet communication resides in the WSN.

For example, a sensor node wants to deliver an alarm message to some Internet host.

The first problem to solve is how to find the gateway from within the network

Basically, a routing problem to a node that offers a specific service has to be solved, integratingrouting and service discovery

If several such gateways are available, how to choose between them?

In particular, if not all Internet hosts are reachable via each gateway or at least if some gateway shouldbe preferred for a given destination host?

How to handle several gateways, each capable of IP networking, and the communication amongthem?

One option is to build an IP overlay network on top of the sensor network

How to map a semantic notion (—Alice) to a concrete IP address?

Even if the sensor node does not need to be able to process the IP protocol, it has to include sufficient information (IP address and port number, for example) in its own packets;

the gateway then has to extract this information and translate it into IP packets.

An ensuing question is which source address to use here – the gateway in a sense has to perform tasks similar to that of a Network Address Translation (NAT) device.

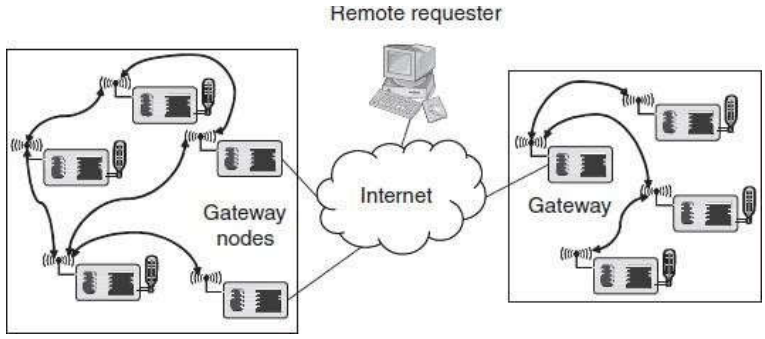


Figure 1.17: A wireless Sensor Network with gateway node, enabling access to remote

clients via the WSN

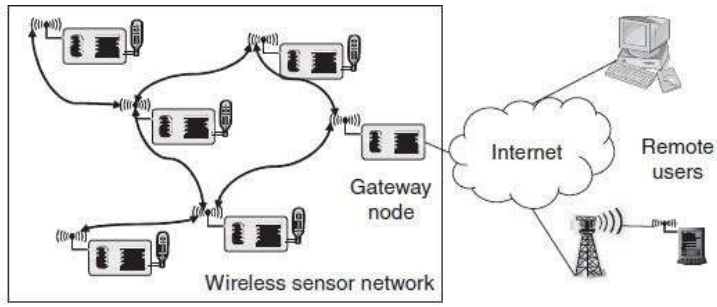
Internet to WSN communication: The case of an Internet-based entity trying to access services of aWSN is even more challenging.

This is fairly simple if this requesting terminal is able to directly communicate with the WSN.

The more general case is, however, a terminal —far away requesting theservice, not immediatelyable to communicate with any sensor node and thus requiring the assistance of a gateway node

- ✓ First of all, again the question is how to find out that there actually is a sensor network in the desired location, and how to find out about the existence of a gatewaynode?

- ✓ Once the requesting terminal has obtained this information, how to access the actualservices.
- ✓ The requesting terminal can instead send a properly formatted request to this gateway, whichacts as an application-level gateway
- ✓ The gateway translates this request into the proper intra sensor network protocolinteractions
- ✓ The gateway can then mask, for example, a data-centric data exchange within thenetwork behind an identity-centric exchange used in the Internet
- ✓ It is by no means clear that such an application-level protocol exists that represents an actual simplification over just extending the actual sensor network protocols to the remote terminal
- ✓ In addition, there are some clear parallels for such an application-level protocol with so- called Web Service Protocols, which can explicitly describe services and the way they can be accessed



A wireless sensor network with gateway node, enabling access to remote clients via the Internet

Figure 1.18: A wireless Sensor Network with gateway node, enabling access to remote

clients via the internet

2.7 DESIGN PRINCIPLES FOR WSN

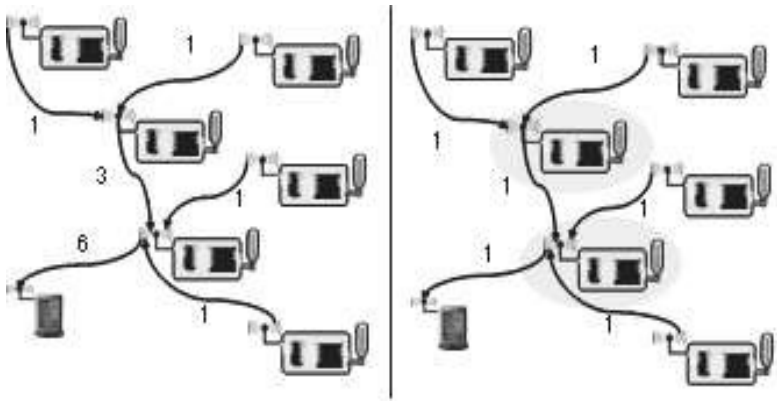
2.7.1 Distributed Organization

- Both the scalability and the robustness optimization goal are required to organize the network in a distributed fashion.
- When organizing a network in a distributed fashion, it is necessary to know potential shortcomings of this approach
- In many cases, a centralized approach can produce solutions that perform better or require fewer resources.
- One possibility is to use centralized principles in a localized fashion by electing, out of set of equal nodes.

- Such elections result in a dynamic hierarchy.
- The election process should be repeated continuously until the elected node runs out of energy

2.7.2 In Network Processing Techniques

1. Aggregation: The simplest in-network processing technique is aggregation. The term aggregation means that information is aggregated into a condensed form in nodes intermediate between sources and sinks out of information provided by nodes further away from the sink. The aggregation function must be applied in the intermediate nodes as shown in Figure 2.5.



2. Distributed Source Coding and Distributed Compression:

- The objective is to encode the information provided by several sensors by using traditional coding schemes, which may be complex for simple sensor nodes.
- The readings of adjacent sensors are going to be quite similar and correlated. Such correlation can be exploited instead of sending the sum of the data so that the overhead can be reduced.

3.Distributed and collaborative signal processing: When complex computations on a certain amount of

data is to be done, it can be more energy efficient to compute these functions on the sensor nodes using Fast

Fourier Transform (FFT). In principle, this is similar to algorithm design for parallel computers. However the

energy consumption of communication and computation are relevant parameters to decide between various algorithms.

3. Mobile code/Agent-based networking: The idea of mobile code is to have a small, compact representation of program code to be sent from node to node. This code is executed locally for collecting measurements and then decides where to be sent next. This idea has been used in various environments.

2.7.3 Adaptive Fidelity & Accuracy

- The idea of making fidelity of computation depends upon the amount of energy available for that particular computation.
- This concept can be extended from a single node to an entire network. As an example, consider a function approximation application.
- When more sensors participate in the approximation, the function is sampled at more points and the approximation is better. But more energy has to be invested.
- Hence, it is up to an application to define the degree of accuracy of the results and the task of the communication protocols to achieve this accuracy.

2.7.4 Data Eccentricity

- In traditional communication networks, the focus will beon the pair of communicating peers, the sender and the receiver of data.
- In a wireless sensor network, the interest of an application is actual information reported about the physical environment. This is applicable when a WSN is redundantly deployed such that any given event can be reported by multiple nodes.
- This method of concentrating on the data rather than identity of nodes is called data-centric networking.
- For an application, this means that an interface is exposed by the network where data only is addressed inrequests.

2.7.5 Exploit Local Information

- Another useful technique is to exploit location information in the communication protocols when-ever such information is present.
- Since the location of an event is crucial information for many applications, mechanisms must be available to determine the location of sensor nodes.
- It can simplify the design and operation of communication protocols and can improve their energy efficiency.

2.7.6 Exploit Activity Patterns

- Activity patterns in a wireless sensor network are quite different from that of traditional networks.
- The data rate averaged over a long time can be very small.
- This can be detected by a larger number of sensors,breaking into a frenzy of activity, causing a well-known event shower effect.
- Hence, the protocol design should be able to handle such bursts of traffic by switching between modes of quiescence and of high activity.

2.7.7 Exploit Heterogeneity

- Sensor nodes can be heterogeneous by constructions, that is, they have larger batteries, farther-reaching communication devices, or more

processing power.

- They can also be heterogeneous by evolution, that is, they started from an equal state, but scavenge energy from the environment due to overloading.
- Heterogeneity in the network is both a burden and an opportunity.
- The opportunity is an asymmetric assignment of tasks, giving nodes with more resources or more capabilities the more demanding tasks.
- The burden is asymmetric task assignments cannot be static but have to be reevaluated.

2.7.8 Component Based Protocol Stacks

- The concept is a collection of components which can form a basic "toolbox" of protocols and algorithms to build upon.
- All wireless sensor networks will require some form ofphysical, MAC, Link layer protocols, routing and transport layer functionalities.
- Moreover, "helper modules" like time synchronization, topology control can be useful.
- On top of these basic components, more abstract functionalities can then be built.
- The set of components active on a sensor node can be complex and will change from application to application.
- Protocol components will also interact with each othereither by using simple exchange of data packets or by exchange of cross-layer information.Services Interfaces of WSN:

2.8 LIST OF REFERENCES

- Protocols and Architectures for Wireless Sensor Network, Holger Kerl, Andreas Willig, John Wiley and Sons, 2005
- Internet References

2.9 SUMMARY

- The separation of functionalities is justified from the hardware properties as is it supported by operating systems like TinyOS. These trade-offs form the basis for the construction of networking functionalities, geared toward the specific requirements of wireless sensor network applications.
- The wireless sensor networks and their networking architecture will have many different guises and shapes. For many applications, but by no means all, multihop communication is the crucial enabling technology, and most of the WSN research as well as the following part of this book are focused on this particular form of wireless networking. Four main optimization goals – WSN-specific forms of quality of service support, energy efficiency, scalability, and robustness – dominate the requirements for WSNs and have to be carefully

arbitrated and balanced against each other. To do so, the design of WSNs departs in crucial aspects from that of traditional networks, resulting in a number of design principles. Most importantly, distributed organization of the network, the use of in-network processing, a data-centric view of the network, and the adaptation of result fidelity and accuracy to given circumstances are pivotal techniques to be considered for usage.

- The large diversity of WSNs makes the design of a uniform, general-purpose service interface difficult; consequently, no final solutions to this problem are currently available. Similarly, the integration of WSNs in larger network contexts, for example, to allow Internet- based hosts a simple access to WSN services, is also still a fairly open problem. The physical layer is mostly concerned with modulation and demodulation of digital data; this task is carried out by so-called transceivers. In sensor networks, the challenge is to find modulation schemes and transceiver architectures that are simple, low cost, but still robust enough to provide the desired service.

2.10 UNIT END EXERCISES

1. Discuss the 4 different types of controllers.
2. State and explain any 5 characteristics of Transceiver. 10. What are the transceiver operational states?
3. In Wireless Sensor Networks, state the three types of Mobility. 12. Write a short note on 4 aspects of optimization goals?
4. List and explain any 5 basic principles for designing network protocols.
5. What are the requirements for WSN service interfaces?
6. State the reasons why gateways are needed in WSN.
7. Explain Single-node Architecture in Detail?
8. Explain network Architecture?

3

MEDIUM ACCESS CONTROL PROTOCOLS

Unit Structure

3.0 OBJECTIVES

- To understand fundamental and performance requirements of MAC protocols
- To get familiar with some of the MAC protocols along with the case study

3.1 INTRODUCTION

WSNs are often made up of numerous cheap, low-power, multifunctional wireless devices that are randomly and hastily placed around a geographic area. Because to resource constraints, sensing devices can only process and communicate a finite quantity of data at a time. Yet, it is the combined effort of these sensing devices that holds out hope for a large impact on a variety of applications in many different sectors, such as science and engineering, military scenarios, protecting key infrastructure, and environmental monitoring.

A high degree of self-organization and coordination between the sensors is necessary to carry out the duties necessary to support the underlying application in order to fully utilize the potential benefits of WSNs. The requirement for the wireless sensor nodes to self-organize into a multi-hop wireless network lies at the core of this cooperative endeavor to accomplish communications. To properly complete the task for which they are deployed, wireless sensor nodes must therefore be equipped with effective communications and network protocols.

Communication linkages must be established between nearby sensor nodes in order to construct a multi-hop wireless network infrastructure for data transfer. Nevertheless, communication in wireless networks is accomplished via electromagnetic signal transmission in the air, as opposed to communication over a directed media in wired networks. So, all sensor network nodes must fairly share this shared communication medium. A medium access control protocol must be used to accomplish this. The primary determinant of WSN performance is the selection of the medium access control protocol.

3.2 FUNDAMENTALS OF MAC PROTOCOLS

The spatial dispersion of the communicating nodes presents a significant challenge in developing efficient MAC techniques for shared access media. The nodes must exchange a certain amount of coordinating information in order to agree on which node can use the communication channel at any given moment. However, the utilization of the communication channel itself is often necessary for the sharing of this information. The complexity of the access control protocol is increased by the multiaccess medium problem's recursive nature, which also raises the administrative burden needed to control access among the contending nodes. Moreover, due to spatial dispersal, a given node cannot instantly know the status of other nodes in the network.Every data collected by a node, whether explicitly or implicitly, is at least as old as the time it took for it to travel over the communication channel.

The aggregate behavior of a distributed multipleaccess protocol is influenced by two key variables: the overhead involved and the intelligence of the decision made by the access protocol. These two elements are inextricably linked. Making an effort to raise decision quality does not always result in cost savings.On the other side, lowering the

overhead is probably going to result in a lower-quality decision. As a result, these two aspects must be traded off.

Although challenging, figuring out the type and volume of data a distributed multiple access protocol uses could be useful. Knowing precisely what data is required could help one recognize its importance.The majority of the distributed multiple-access protocols for WSNs that have been proposed function somewhere along a spectrum of information, from no information to perfect information. Moreover, the data can be predefined, dynamic, or local.Every node involved in communication is aware of predetermined information. During protocol execution, various nodes obtain dynamic global information. Each node is aware of local information. Global information that is predetermined and dynamic may lead to effective coordination among the nodes that could even be flawless. Yet, the cost in terms of unused channel capacity is typically substantial. The usage of local information has the potential to lower the overhead needed to coordinate the competing nodes, but it may have a negative impact on the protocol's overall speed.

The majority of access approaches for shared-medium networks are built on the trade-off between the MAC protocol's efficiency and the overhead necessary to achieve it. The performance measurements for the MAC protocol are presented in the remaining portion of this section, along with the most popular methods for controlling access to the medium.

3.2.1 Performance requirements

The breadth of the research has been fairly extensive in attempting to ascertain the performance requirements of MAC protocols. Delay, throughput, robustness, scalability, stability, and fairness have historically dominated MAC protocol design. The description of these performance metrics is provided below.

1. **Delay:** The length of time a data packet spends in the MAC layer before it is successfully transferred is referred to as delay. In addition to network traffic volume, the MAC protocol's design decisions also affect delay. The MAC protocol must enable delay-bound guarantees for time-critical applications in order for those applications to comply with QoS standards. The specific QoS requirements' semantics vary depending on the application. With proper message scheduling, both locally within a communicating node and globally across all nodes in the network, guaranteed delay limitations are typically established. It is possible to distinguish between probabilistic and deterministic delay guarantees.An expected value, a variance, and a confidence interval are often used to describe probabilistic delay guarantees. A known number of state changes occur between message arrival and message transmission thanks to deterministic delay guarantees. Deterministic MAC methods thus provide an access time upper bound. In a real-time setting, where the accuracy of the application depends on the adherence of its underlying tasks to their predetermined execution deadlines, determinism is a critical need.

2. **Throughput:** The rate at which messages are handled by a communication system is known as throughput. It is typically expressed as a number of messages or bits per second. It reflects the portion of the channel capacity that is utilized for data transmission in wireless contexts. As the initial load on the communication system rises, throughput rises. The throughput stops increasing and, in certain situations, may even begin to decrease once the load hits a certain threshold. To maximize channel throughput while decreasing message delay is a key goal of a MAC protocol.

3. **Robustness:** In terms of reliability, availability, and dependability requirements, robustness measures how sensitive the protocol is to errors and false information. Error confinement, error detection and masking, reconfiguration, and restart are only a few of the multifaceted concerns that robustness must concurrently solve. It is challenging to achieve robustness in a time-varying network like a WSN since it heavily depends on the failure models of the links and communication nodes.

4. **Scalability:** The ability of a communications system to maintain its performance characteristics regardless of the size of the network or the number of competing nodes is referred to as scalability. In WSNs, the total number of sensor nodes may surpass thousands and, in some situations, millions. Scalability in these networks becomes crucial. Scalability is difficult to achieve, particularly in time-varying contexts like wireless networks. Avoiding relying on globally consistent network states is a typical strategy for achieving scalability. With the creation of hierarchical structures and information aggregation techniques, localizing interactions among the communication nodes is another strategy.For instance, creating clusters of sensor nodes enables the development of highly scalable shared medium access protocols. Similar to this, combining data from various sensors enables the creation of traffic patterns that can be effectively used to scale the MAC protocol to many sensor nodes.

5. **Stability:** The ability of a communications system to manage changes in the traffic load over extended periods of time is referred to as stability. For instance, a stable MAC protocol must be capable of handling sudden loads that are greater than the maximum sustained load, provided that the channel's maximum capacity is not exceeded by the promised long-term load. Usually, a MAC protocol's scalability is examined in terms of either delay or throughput.If the message waiting time is constrained, a MAC protocol is regarded as stable in terms of latency. There is a limited backlog of messages in the transmission queue, which can be used to identify these systems. A MAC protocol is stable in terms of throughput if the throughput does not decrease as the load being delivered rises. In time-varying large-scale WSNs, it is challenging to accommodate load changes while preserving system stability. The careful scheduling of bursty traffic is one potential method for the MAC protocol to respond to significant swings in the traffic load.

6. **Fairness:** If a MAC protocol distributes channel capacity among competing communicating nodes fairly without unnecessarily slowing down the network throughput, it is said to be fair. To ensure equitable QoS and prevent instances where certain nodes perform better than other nodes, it is desirable to achieve fairness across competing nodes. No application is therefore starved or harshly penalized. It should be noted that the definition of fairness given above makes the assumption that the channel capacity requirements of all communication nodes are equal. But, it's possible that the network may need to support a variety of traffic sources with varying traffic production patterns and a wide range of QoS requirements.Communicating nodes are given various weights to reflect their relative resource shares in order to accommodate varied resource needs. Then, based on the weights assigned, proportional fairness is attained. If it is impossible to increase any competitive node's allocation without lowering another node's service rate below its proportional fair share, a MAC protocol is said to be proportionally fair.

   While global information may be needed to coordinate access to the communication medium among all contending stations, equitable resource allocation in wireless networks can be challenging to implement. Even when a centralized resource allocation approach is utilized, it is challenging to calculate each contesting node's fair share due to the time-varying properties of wireless networks.

7. **Energy efficiency:** A sensor node has one or more integrated sensors, a small embedded processor, and short-range radio communication capabilities. These sensor nodes are powered by small-capacity batteries. Wireless sensor nodes are frequently installed in unsupervised locations, which makes it challenging to replace their batteries. Additionally, it is difficult and unstable to recharge sensor batteries using scavenged energy. The lifespan of a sensor node is directly impacted by these harsh restrictions.In order to increase the lifespan of sensor nodes in WSNs, energy conservation becomes of utmost importance. One of the most crucial considerations in the design of the MAC protocol for wireless sensor nodes is energy economy. The MAC-layer protocols' energy inefficiency is caused by a number of factors.

Energy-saving link-layer protocols decrease, if not completely eliminate, energy waste from the sources mentioned above by managing the radio. With comprehensive energy management strategies that not only pay attention to the sensor node radio but also to other sources of energy usage, more energy gains can be made.

3.2.2 Common protocols

The main determining factor in a WSN's success is the selection of the MAC technique. The shared media access issue has been addressed using a number of different tactics. These strategies make various efforts to balance the overhead required to make the best resource allocation

decision with the quality of the decision. Three broad categories: fixed assignment, demand assignment, and random assignmentcan be used to group these systems.

1. **Fixed assignment protocols:** Each node is given a specified fixed quantity of the channel resources in fixed-assignment schemes. Each node uses its allotted resources solely without engaging in inter-node competition. Frequency-division multiple access (FDMA), time-division multiple access (TDMA), and code-division multiple access are common protocols that fall under this category (CDMA).

2. **Demand assignment protocols:** Demand assignment protocols' primary goal is to increase channel usage by optimally or almost optimally assigning the channel's capacity to competing nodes. Demand assignment methods ignore idle nodes and only take into account nodes that are prepared to transmit, in contrast to fixed-assignment schemes, where channel capacity is exclusively provided to the network nodes in a predetermined manner independent of their present communication demands. The chosen node is given access to the channel for a predetermined period of time, which can range from a fixed time slot to the duration of a data packet transmission.

   The access to the channel between competing nodes must typically be arbitrated by a network control mechanism when using demand assignment protocols. Furthermore, in order for competing stations to dynamically seek access to the communication medium, a logical control channel other than the data channel may be necessary. The requirement to seek access to the channel may cause data transfer to be delayed depending on the protocol's characteristics. Demand assignment protocols can also be divided into centralized and distributed types. Whereas token- and reservation-based systems employ distributed control, polling systems are an example of centralized control.

3. **Randomassignment protocols:** Each communicating node in fixed-assignment schemes is given a frequency band in FDMA systems or a time slot in TDMA systems. Regardless of whether the node has data to send or not, this assignment is static. So, if the traffic source is bursty, these techniques may not be effective. The allocated bandwidth is lost when the node is idle and there is no data to be transferred. In order to overcome this flaw, random assignment schemes do away with pre-allocating bandwidth to communication nodes.

   Random assignment techniques have no influence over which communicating node will next be able to access the media. Additionally, no node is given a predicted or scheduled time to broadcast according to these methodologies. To access the transmission medium, all backed-up nodes must compete. When multiple nodes attempt to communicate at once, collision happens. The protocol must have a means to detect collisions and a plan for

scheduling colliding packets for later retransmissions in order to deal with collisions.

The first long radio links and satellite communications used random access techniques. One of the earliest such media access protocols was the ALOHA protocol, often known as pure ALOHA. Simply put, ALOHA enables nodes to communicate anytime they have data to send. The creation of numerous methods, such as carrier-sense multiple access (CSMA), carrier-sense multiple access with collision detection (CSMA/CD), and carrier-sense multiple access with collision avoidance (CSMA/CA), was prompted by efforts to enhance the performance of pure ALOHA.

3.3 MAC PROTOCOLS FOR WSNS

The most important consideration for designing scalable and reliable MAC layer protocols for WSNs is the requirement to conserve energy. Overhead that is too high, listening that isn't being used, packet collisions, and overhearing are some of the factors that lead to energy waste.The competition between the nodes necessitates the sharing of control and synchronization information in order to control access to the media. These control and synchronization packets might consume a lot of energy when they are explicitly exchanged. Extended durations of inactive listening might also reduce network speed and use more energy. Sometimes, the energy used by a sensor over the course of its lifespan is more than half squandered by idle listening.Another big source of energy waste is the retransmission of collision packets. The MAC-layer protocol may experience considerable performance reduction if there are many of these collisions. Similar to excessive overhearing, excessive overhearing results in a node receiving and decoding packets meant for other sensor nodes, which increases energy consumption and significantly reduces network throughput. After the node recognizes that the destination address is different from its own address, these packets are eventually dropped.

Most MAC-layer protocols' primary goal is to lessen energy loss brought on by collisions, idle listening, overhearing, and excessive overhead. Schedule- and contention-based MAC-layer protocols are two broad categories that fit these protocols. A class of deterministic MAClayer protocols known as schedule-based protocols bases channel access on a schedule. One sensor node at a time can access a channel. This is accomplished by allocating resources in advance to each sensor node. Contention-based MAC-layer methods prevent resource pre allocation to specific sensors. Instead, all nodes share a single radio channel that is made available as needed. Yet, attempts to access the communications medium simultaneously are met with collision.

In order to arrange channel access among competing sensor nodes, distributed, randomized techniques are often used to resolve collisions. As nodes become inactive, the fundamental strategy employed to reduce overhearing is to push them into a sleep state. Yet, uncoordinated slumber can make it challenging to communicate with nearby nodes. Several

MAC-layer protocols have proposed a range of less restrictive schedules to alleviate this weakness and synchronize the activities of the network sensors.

3.3.1 Schedule-based protocols

The existence of a timetable that controls access to resources to prevent contention between nodes is presummated by schedule-based MAC protocols for WSNs. Resources like time, a frequency range, or a CDMA code are typical. Schedule-based MAC protocols' primary goal is to maximize energy efficiency in order to increase the lifespan of the network. Scalability, adaptability to shifts in traffic load, and network topology are further desirable qualities. The majority of scheduled-based protocols for WSNs employ a kind of TDMA that divides the channel into time slots, as shown in Figure 1.

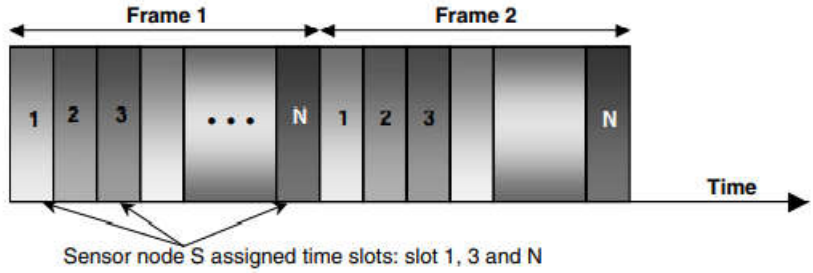


Figure 1: TDMA-based MAC protocols for wireless sensor networks

A logical frame is made up of N consecutive slots, where N is a system variable. This logical structure keeps coming back in cycles. Each sensor node is given a set of precise time slots for each logical frame. The schedule that the sensor node follows for each logical frame is made up of this set. The schedule can either be hybrid, where the structure varies over various time scales and sensor behavior, or fixed, where it is built on demand by the base station on a per-frame basis to represent the current requirements of sensor nodes and traffic pattern.

A sensor switches between its active mode and sleep mode in accordance with its assigned schedule. In the active mode, the sensor transmits and receives data frames using the slots that have been given to them within a logical frame. Sensor nodes enter sleep mode when they are not in their designated slots. In order to save energy, the sensor nodes in this mode turn off their radio transceivers.

1. **Self organizing medium access control for sensornets (SMACS):** A medium access control protocol called SMACS enables the creation of haphazard network topologies without the need for network node global synchronization. Nonsynchronous scheduled communication, a major characteristic of SMACS, enables links to be generated and planned concurrently throughout the network without the requirement for expensive global connectivity information exchange or time synchronization. Each node in the network keeps a super frame, which is TDMA-like, for communication with recognizedneighbors.A

superframe has a defined length. The superframe is additionally broken into smaller frames. Each frame's size is not constant and may change over time for a single node as well as from one node to another. Each node must conduct a neighborhood discovery operation on a regular basis in order to identify nearby nodes according to SMACS. By giving each detected neighbor a time slot, each node creates a link with them. The time slots are chosen so that at each slot, the node only converses with its neighbors. The link construction mechanism must make sure that there is no interference between neighboring links, despite the fact that a node and its neighbors are not needed to transmit at distinct slot times.This is accomplished by spreading code (CDMA) or randomly selecting a channel from a vast pool of channels for each link. Each node in the superframe structure keeps its own time slot schedules with all of its neighbors, and to communicate, nodes must set their radios to the appropriate frequency channel or CDMA code.

2. **Bluetooth:** A centralized TDMA-based protocol serves as the main media access control mechanism for the developing technology known as Bluetooth. With one common short-range radio link, Bluetooth is intended to replace cables and infrared links used to connect various electronic devices, including cell phones, headsets, PDAs, digital cameras, laptop computers, and their accessories. The ISM frequency range at 2.45 GHz is where Bluetooth operates. Its physical layer is based on a technique for allocating hopping sequences and a pseudorandom frequency-hopping scheme with a hopping frequency of 1.6 kHz. With 1-MHz spacing, a set of 79 hop carriers are defined. Each hop sequence establishes a Bluetooth channel with a 1 Mbps data rate.A piconet is a collection of devices that communicate via a single channel.

   In order to support broadcasting by a slave to all members of its piconet, Bluetooth defines four different types of communication between nodes: intra piconet unicast for slave-to-slave communication within a piconet; intra piconet broadcast; inter piconet unicast for piconet-to-piconet communications; and inter piconet broadcast for piconet-to-all scatternet node communications.

   The source slave enters its own MAC address in the data packet's equivalent field for intra-piconet unicast transmission, sets the packet's forward field to 1, and sets the destination address to the desired destination node. The master examines the forward field after receiving the message. If it is, the master sends the message to the intended slave device indicated by the destination address of the original packet, replacing the MAC address field with its own MAC address.

   The source slave writes its own MAC address, sets the forward field to 1, and sets the destination address to 000 for intra piconet broadcast communication. The forward field is already set when the master receives the message. The message is then sent to all of the nodes in

the piconet by the master, who substitutes its own address for the MAC address.

For inter piconet unicast communication, the source device sends the data packet with its own MAC address and sets the forward field to 1, the broadcast field to 1 and the destination address to the relay of the next piconet. Furthermore, the source device sets the routing vector field (RVF) of the packet to contain the logical path to the targeted destination device in the intended piconet. The RVF is a sequence of tuples of the form (LocId, Mac_Addr), where LocId represents the identity of the local master and Mac_Addr its corresponding piconet MAC address. Upon receiving the message, the master forwards it to the relay node. The relay extracts from the RVF the next pair, containing the local identity and the MAC address of the master, and sends the message to this master. This process is repeated until the RVF becomes empty, signaling that the destination device has been reached.The relay sends the message to this master by extracting the following pair from the RVF, which contains the local identity and MAC address of the master. Up until the RVF is empty, indicating that the destination device has been reached, this procedure is repeated.

The source device produces a packet with its own MAC address and sets the forward and broadcast fields to 1 and the destination address to 000 for inter-piconet broadcast communication. The master is then notified of the packet.The packet is sent to all slaves in the piconet, including relay nodes, when the master observes that the broadcast field is set to 1. Relay nodes receive broadcast packets and forward them to all connected masters save the one from which they originated.

3. **Low-Energy Adaptive Clustering Hierarchy (LEACH):** Nodes are clustered using a hierarchical approach using LEACH. Nodes take turns acting as cluster heads inside each cluster. To establish communication between nodes and their cluster head, LEACH employs TDMA. Messages from the cluster head's cluster nodes are forwarded to the base station.

A TDMA schedule is established by the cluster head node and sent to every other node in the cluster. Data message collisions are avoided by the scheduling.The nodes can utilize the schedule to identify the times when they need to be active as well. With the exception of the head cluster, this enables each cluster node to turn off its radio components until the designated time intervals. LEACH presupposes that cluster nodes begin the cluster setup phase simultaneously and maintain synchronization throughout. Sending synchronization pulses to every node from the base station is one method of synchronization that could be used.

LEACH employs a code assignment system based on transmitters to lessen inter-cluster interference. Direct-sequence spread spectrum (DSSS) is used to communicate between a node and its cluster head. Each cluster is given a specific spreading code, which is utilized by all nodes in the cluster to transmit data to the cluster head. On a first-in, first-served basis,

spreading codes are given to cluster heads starting with the first one to announce its position and moving on to succeeding cluster heads. Nodes must modify their transmission power in order to lessen interference from neighboring clusters.

The cluster head aggregates the data after receiving data packets from its cluster nodes and sends the data to the base station. With fixed spreading code and CSMA, a cluster head and base station can communicate with one another. The cluster head must sense the channel to make sure no other cluster heads are currently broadcasting data using the base station spreading code before it can transmit data to the base station. The cluster head delays data transmission until the channel is free of traffic if the channel is felt to be busy. The cluster head delivers the data using the base station spreading code when this incident takes place.

3.3.2 Random access-based protocols

Standard random-access Contention-based protocols, also referred to as MAC-layer protocols, don't need cooperation from the nodes accessing the channel. Nodes that have collided back off for an arbitrary period of time before trying to access the channel once more. Nevertheless, WSN environments are not well suited for these protocols.The addition of collision avoidance, request-to-send (RTS), and clear-to-send (CTS) techniques to these protocols boosts their functionality and increases their resistance to the hidden terminal issue. Yet, due to collisions, idle listening, overhearing, and significant control overhead, contention-based MAC-layer protocols continue to have low energy efficiency. The design of random-access MAC-layer protocols made an effort to address this flaw by focusing on minimizing energy waste in order to increase the network lifetime.

By using a separate signaling channel, the power aware multiaccess protocol with signaling (PAMAS) prevents overhearing between nearby nodes. In order to enable nodes to turn off their radio transceivers when they are not actively sending or receiving packets, the protocol combines the usage of a busy tone with RTS and CTS packets. Nevertheless, the protocol does not offer any techniques to cut down on energy loss brought on by inactive listening.

Latency is exchanged for energy efficiency in the sparse topology and energy management (STEM) protocol. Two radio channels: a data radio channel and a wake-up radio channelare used to accomplish this. In a STEM form, the wake-up signal is not encoded data but a busy tone. A pseudoasynchronous planned scheme is called as STEM. According to this technique, a node disables its data radio channel until it needs to communicate with another node. A node starts sending on the wake-up radio channel when it has data to send. Similar to a paging signal, the wake-up signal channel is used. This signal is transmitted for a sufficient amount of time to page all nearby nodes.A node may stay awake long enough to receive a "session" of packets after being roused from sleep.Moreover, a node may be awakened in order to receive all of its

pending packets before entering the sleep mode once more. Because it is a broad protocol, STEM can be utilized with other MAC-layer scheduling techniques. However, the approach only works in network situations when events happen seldom. If events happen often, the energy lost from sending wake-up signals continuously may equal or even be greater than the energy obtained from sleeping modes.

Using RTS and CTS packets, a variety of contention-based protocols modelled after IEEE 802.11 avoid overhearing. These protocols frequently employ the technique of forcing a contending node into sleep mode by overhearing the RTS and CTS packet exchange between two other contending nodes. To prevent idle listening, these protocols also rely on coordinated schedules between nearby nodes. Particularly when the size of the data packets is of the same order of magnitude as the size of the RTS and CTS packets, these protocols differ in how they maintain low duty cycles and achieve energy economy.

A contention-based MAC-layer protocol called the timeout-MAC (T-MAC) was created for applications with low message rates and low sensitivity to latency. T-MAC nodes communicate with one another using RTS, CTS, and acknowledgement packets to prevent collision and guarantee reliable transmission. The protocol also employs an adaptive duty cycle to lower energy usage and adjust for changes in traffic load. The T-MAC protocol's fundamental goal is to minimize idle listening by sending all messages in bursts of varying length. Between bursts, nodes are allowed to sleep. Also, based on the current load, the protocol estimates the ideal active time duration dynamically. Since messages must be buffered between active times, the buffer capacity establishes an upper limit on the maximum frame time.

A lower-power carrier-sense media access protocol for WSNs is the Berkeley media access control (B-MAC). The B-MAC protocol incorporates a modest core of media access capabilities in contrast to conventional IEEE 802.11-inspired MAC-layer protocols, which also include techniques for network organization and clustering. B-MAC employs listening for low-power communication, link-layer acknowledgments for reliability, clear channel assessment (CCA), and packet back-offs for channel arbitration. B-MAC does not directly support multipacket techniques that handle message fragmentation, deal with hidden terminal issues, or impose specific low-power policies.

3.4 SENSOR-MAC CASE STUDY

The sensor-MAC (S-MAC) protocol is specifically developed to minimize energy loss brought on by overhearing, idle listening, collision, and control overhead.The objective is to achieve high levels of reliability and scalability while improving energy efficiency. The protocol does, however, suffer some performance degradation in terms of per-hop fairness and latency. S-MAC employs a variety of methods to lower energy usage, manage overhead, and decrease latency in order to enhance application-level performance. The S-MAC-layer protocol's proposed

methods for achieving energy efficiency while maintaining low latency are covered in the sections that follow.

3.4.1 Protocol overview

The protocol design anticipates a sizable number of sensor nodes with constrained computing, communication, and storage resources. Ad hoc, self-organized, and self-managed wireless networking is how the nodes are set up. Sensor data is processed and transmitted in a store-and-forward fashion. It is expected that the applications supported by the network alternate between extended periods of inactivity, during which nothing happens, and brief active times, during which data flow towards the base station through peer sensor node messaging occurs. Additionally, it is anticipated that the applications will tolerate higher latency for a longer network lifespan.The protection of vital infrastructure as well as surveillance and monitoring of natural ecosystems are typical uses that come under this category. In these applications, the sensors must be on guard for protracted periods of time, after which they go dormant until an event happens. These events often happen orders of magnitude less frequently than it takes for a message to be transmitted across the network towards the base station.

S-MAC establishes low-duty-cycle operation on nodes in a multi-hop network and results in significant energy savings by taking advantage of the bursty characteristic of sensor applications.S-MAC nodes occasionally switch between listening and sleep modes during the protracted periods of no sensing. Every node establishes a wakeup time and a slumber period, during which its radio is off. The node resumes operation when the timer expires. The protocol makes use of synchronized sleeping amongst nearby nodes to further decrease control overhead while maintaining low message latency.Sleeping periodically results in less energy use but more delay. The requirements of the sensing application have a significant impact on how important message delay is. Applications that can tolerate latency on the order of seconds are the main emphasis of S-MAC. Yet, latency can considerably rise when nodes faithfully adhere to their schedule. S-MAC employs adaptive listening to solve this flaw and maintain message delay within the targeted-second-level latency.

As previously mentioned, the S-MAC design is concentrated on applications that work together, such as monitoring and surveillance applications. The programs work together to complete a particular objective, like safeguarding a vital infrastructure. These applications' nature makes it possible for one sensor node to have a lot of information to share with its neighbors at any given time. The idea of message passing, in which a node is permitted to deliver a lengthy message in bursts, is used by S-MAC to satisfy this need while also lowering overhead. Overhearing is reduced and avoided via message passing.

3.4.2 Periodic listen and sleep operations

By reducing idle listening, the S-MAC architecture aims to cut energy consumption. Establishing low-duty-cycle activities for sensor nodes

allows for this. Nodes periodically enter a sleep state during which their radios are fully disabled. When there is activity on the network, nodes become active.Figure 2 shows the fundamental periodic listen and sleep schedule.Each node in this system sets a wake-up timer and sleeps for the allotted amount of time. As the timer runs out, the node awakens and starts listening to see whether it needs to connect with other nodes. A frame is the collective term for the full cycle of listening and sleeping. The listening interval to frame length ratio, or duty cycle, is what distinguishes each frame. The listening interval duration can be separately chosen by sensor nodes, however for the sake of simplicity, the protocol takes the value to be the same for all nodes.

Figure 2: S-MAC period listen and sleep modes of operations

Nodes have complete control over the times they sleep and listen. To minimize the amount of management required to establish connections between these nodes, it is preferred that the schedules of nearby nodes be coordinated.S-MAC nodes build virtual clusters around schedules, but instead of communicating through a master node like a cluster head to achieve coordination, they speak directly with one another to exchange and coordinate their sleep and listen schedules.

3.4.3 Schedule selection and coordination

In order for everyone to listen and sleep at the same time, the nearby nodes coordinate their listening and sleeping schedules. Each node chooses a schedule and shares it with its neighbors during the synchronization time in order to coordinate their listening and resting. Every node has a schedule table with all of its known neighbors' schedules.

A node must first listen to the channel for a predetermined period of time, at least equal to the synchronization period, before choosing a schedule. If the node does not get a schedule from another node before the end of this waiting period, it immediately selects its own schedule. The node informs all of its neighbors of the schedule it has chosen by broadcasting a SYNC packet. The node must first carry out physical carrier sensing before broadcasting the SYNC packet, it's important to note.This lessens the possibility of SYNC packet collisions between rival nodes. The node sets its schedule to be the same as the schedule received if, during the synchronization period, it receives a schedule from a neighbor before deciding on and publishing its own schedule. The node waits until the following synchronization period to inform its adjacent nodes about the schedule.

It should be noted that after selecting and announcing its own schedule, a node can then get a different one. This might happen if channel

interference or collision contaminate the SYNC packet. The node simply discards its own schedule and adopts the new one if it has no neighbors with whom it shares a timetable. The node, on the other hand, adopts both schedules if it is aware of other nearby nodes that have already accepted it. The node must then awaken at the two adopted schedules' listen intervals. Figure 3 illustrates this.With various schedules, border nodes only need to broadcast one SYNC packet, which is advantageous. The drawback of this strategy is that border nodes use more energy because they don't spend as much time in sleep mode.

It should be observed that if a SYNC packet is delayed or lost, neighboring nodes may still be unable to find one another. S-MAC nodes must frequently discover their neighbors in order to fix this flaw. To do this, a node must periodically listen to the full synchronization time. Nodes without neighbors at the moment are anticipated to carry out neighbor discovery more frequently.

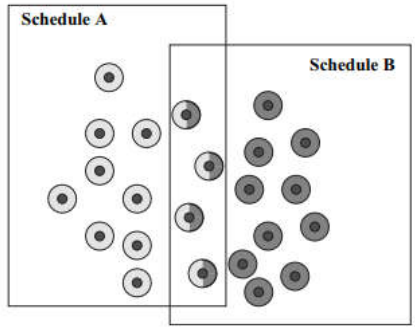


Figure 3: Border node schedule selection and synchronization

3.4.4 Schedule synchronization

To stop long-term clock drift, neighboring nodes must periodically synchronize their schedules. Sending a SYNC packet is used to update the schedule. Figure 4 illustrates how the listen interval is split into two subintervals to allow a node to receive both SYNC packets and data packets. Three cases are represented in this figure. The sender sends simply a SYNC packet in the first scenario, a data packet in the second scenario, and a SYNC packet along with the data packet in the third scenario.

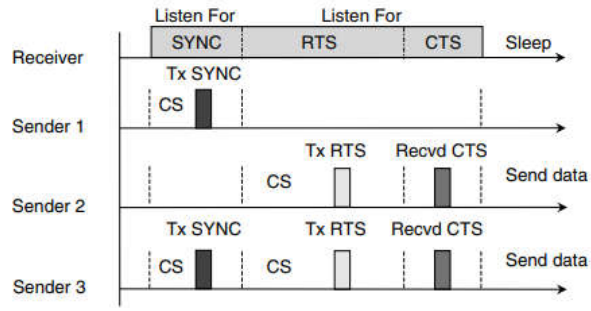


Figure 4: Timing relationship between a receiver and a variety of senders

During these subintervals, the channel access of competing nodes is controlled by a multi-slotted contention window. SYNC packet transmission takes place during the first subinterval, while data packet transmission occurs during the second subinterval. At either of these subintervals, a competing station chooses a time slot at random, conducts carrier sensing, and initiates packet transmission if it notices that the channel is empty. The RTS/CTS handshake is used during data packet transmission to guarantee exclusive access to the channel. This access method ensures that both the synchronization and data packets reach the nearby nodes.

3.4.5 Adaptive listening

As a message is stored and transferred across nearby network nodes, it may experience increased latency, according to a closer examination of the periodic listen and sleep scheme. Data packets could be delayed at each hop if a sensor must precisely adhere to its sleep pattern. The protocol makes use of an assertive method known as adaptive listening to remedy this flaw and enhance latency performance. According to this method, a node that hears, during its listen period, a neighboring node and another node exchanging a CTS or RTS packet assumes that it might be the next hop along the routing path of the overheard RTS/CTS packet, disregards its own wake-up schedule, and schedules an additional listening period around the time the transmission of the packet terminates.The duration field of the overheard CTS or RTS packet is used by the overhearing node to calculate the amount of time required to finish transmitting the packet. The node sends an RTS packet to start an RTS/CTS handshake with the overhearing node as soon as it receives the data packet. If the latter node is awake, it is ideal since then the packet forwarding process between the two nodes can start right away. The overhearing node returns to sleep until the next scheduled listen interval if it doesn't receive an RTS packet during adaptive listening.

3.4.6 Access control and data exchange

S-MAC uses a CSMA/CA-based technique, incorporating physical and virtual carrier sensing and the use of RTS/CTS handshake to lessen the impact of the hidden and exposed terminal difficulties, to govern access to the communication channel among competing sensor nodes. The network allocation vector (NAV), a variable whose value contains the amount of time left in the current packet transmission, is used to implement virtual carrier sensing. The NAV value is initially set to the value contained in the transmitted packet's duration field.As time goes on, the value decreases until it eventually equals zero. A node cannot start a transmission on its own until the NAV value is zero. When doing physical carrier sensing, the channel is listened to for signs of active transmission. To prevent collisions and hunger, carrier sensing is randomly distributed within a contention window. If the channel is clear according to both virtual and physical carrier sensing, a node is permitted to transmit.

Nodes may need to listen to every broadcast from their neighbors in order to execute virtual carrier sensing efficiently. It may be necessary for nodes to listen to packets that are meant for other nodes as a result. Overhearing packets could waste a lot of energy. S-MAC permits nodes to enter sleep mode once they hear the exchange of an RTS or a CTS packet between two other nodes in order to prevent overhearing.The node enters the sleep state until the NAV value reaches zero after initializing its NAV with the value found in the duration field of the RTS or CTS packets.The overhearing avoidance procedure may result in significant energy savings because data packets are often larger than control packets. Figure 5 depicts the collision-avoidance strategy utilized by S-MAC.

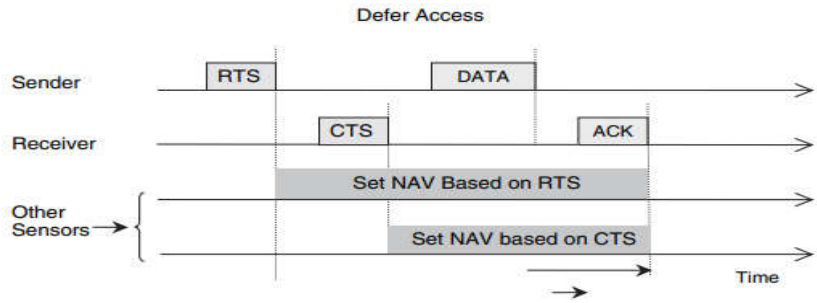


Figure 5: S-MAC collision avoidance scheme

A node must first perceive the channel before attempting to transmit a message. If the channel is congested, the node sleeps and awakens when the channel is free. When transferring a data packet across an empty channel, a node first sends an RTS packet and then waits for a CTS packet from the recipient. The node delivers its data packet after obtaining the CTS packet. After the node gets a confirmation from the recipient, the transaction is finished. It is important to note that the connecting nodes continue to exchange data packets during their regular sleep periods after the successful exchange of the RTS and CTS packets. The nodes wait until the data transfer is finished before starting their usual sleep cycle.Moreover, the exchange of the RTS and CTS packets is not necessary for the transmission of a broadcast packet, such as a SYNC packet.

3.4.7 Message passing

S-MAC introduces the idea of message forwarding, where a message is a meaningful unit of data that a node can process, to enhance application-level performance. The messages are broken up into little pieces. Then, a single burst of these fragments is delivered. A single RTS/CTS exchange is used to transport the message fragments between the transmitting and receiving nodes.The medium is set aside following this exchange for the duration required to successfully transfer the entire message. Each fragment also includes the time required to transmit all of the next pieces and their related acknowledgments in its duration field. Figure 6 illustrates this process.

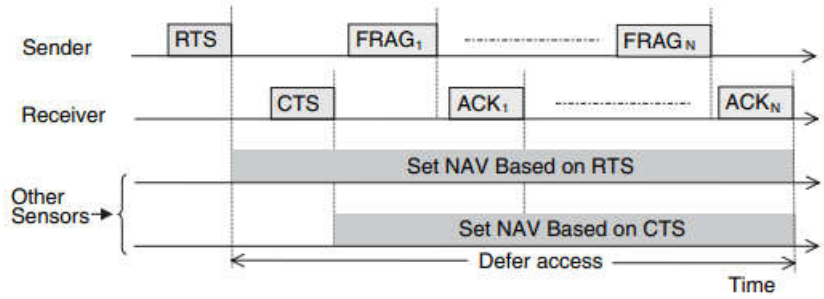


Figure 6: S-MAC message passing

The sender waits for the receiver to acknowledge the transmission of a fragment. The sender transmits the next fragment if it receives the acknowledgment. Nevertheless, if the acknowledgement is not received, the sender immediately retransmits the unacknowledged frame and extends the time needed to finish transmission of the segment to account for the time needed to transmit one more fragment and its accompanying acknowledgment. Notably, only expanded fragments or their related acknowledgements will alert sleeping nodes to this extension. The transmission extension is not heard by nodes that only heard the first RTS and CTS packet exchange. The S-MAC has the potential to significantly reduce energy consumption.It works well for situations when higher latency is acceptable and fairness is not a crucial design goal.

3.5 SUMMARY

A new technology called sensor networking has numerous potential uses, including the protection of vital infrastructure, environmental monitoring, smart cities, all-pervasive and ubiquitous healthcare, and robotic exploration. Typically, a WSN is made up of a sizable number of dispersed, battery-powered nodes that are furnished with one or more sensors, embedded CPUs, and low-power radios. As a multihop wireless network, these nodes cooperate with one another. To successfully complete the task for which they are deployed, wireless sensor nodes depend on the design of effective MAC-layer protocols for WSNs.

The primary determinant of WSN performance is the selection of the medium access control protocol. An effective MAC layer protocol for WSN must take into account a number of factors. Battery-powered sensor network nodes are typical, and it is frequently challenging, if not impossible, to replace or recharge them. To increase the network's lifespan, an efficient MAC-layer protocol design for a WSN must also be energy efficient. In order to accommodate changes in network size, node density, and topology, the MAClayer protocol must also be scalable. Finally, when designing MAC layer protocols for WSNs, access equity, low latency, high throughput, and bandwidth utilization are also crucial considerations.

As more WSNs continue to appear, interest in the development of a MAC-layer protocol for sensor networks is expected to remain high. Also, recent advancements in cognitive radio are probably going to change how the

MAC-layer protocol for WSNs is designed. A wireless network device fitted with such a radio will be better able to adapt to and interact with its environment while carefully controlling its energy consumption thanks to the direct environment interaction capabilities of cognitive radios.

3.6 LIST OF REFERENCES

1) Protocols and Architectures for Wireless Sensor Network, Holger Kerl, Andreas Willig, John Wiley and Sons, 2005

2) Wireless Sensor Networks Technology, Protocols, and Applications,Kazem Sohraby, Daniel Minoli and TaiebZnati, John Wiley & Sons, 2007

3) Mobile communications, Jochen Schiller,2nd Edition, Addison wisely, Pearson Education,2012

4) Fundamentals of Wireless Sensor Networks, Theory and Practice, WaltenegusDargie, Christian Poellabauer, Wiley Series on wireless Communication and Mobile Computing, 2011

5) Networking Wireless Sensors, Bhaskar Krishnamachari, Cambridge University Press, 2005

3.7 UNIT END EXERCISES

1) Illustrate the performance requirements of MAC protocols.

2) Describe the MAC Protocols for WSNs.

3) Write a note on Schedule-based protocols.

4) Explain the Random access-based protocols.

5) Describe the Sensor-MAC Case Study.

6) What do you mean by Periodic listen and sleep operations?

7) Explain Schedule selection and coordination.

8) What do you mean by Schedule synchronization?

9) What is Adaptive listening?

10) Explain Access control and data exchange.

11) Write a note on Message passing.

4

ROUTING PROTOCOLS

Unit Structure

4.0 OBJECTIVES

- To examine fundamental routing difficulties in WSNs and offer various development methods for routing protocols in these networks
- To draw focus on the particular characteristics of the traffic that is often generated in WSNs
- To understand basic routing strategies used to strike a balance between responsiveness and energy efficiency

4.1 INTRODUCTION

Whether they are made up of fixed or mobile sensor nodes, WSNs can be deployed to support a wide range of applications in a number of contexts. Depending on the application, these sensors are placed in different ways. For example, sensor nodes are often installed ad hoc in environmental monitoring and surveillance applications to cover the precise area to be watched (e.g., C1WSNs). Smart wearable wireless devices and biologically compatible sensors can be strategically affixed to or implanted within the human body for medical purposes to monitor the

patient's vital signs. As soon as they are deployed, sensor nodes form an autonomous wireless ad hoc network that needs little to no maintenance.After that, sensor nodes work together to complete the duties required by the application for which they were installed.

The primary duty of wireless sensor nodes is to detect and gather data from a target domain, process the data, and communicate the information back to specified sites where the underlying application lives, notwithstanding the diversity in the goals of sensor applications. The creation of an energy-efficient routing protocol is necessary to create pathways between sensor nodes and the data sink in order to do this operation effectively. The network lifetime must be maximized by the path selection process. The routing challenge is extremely difficult due to the features of the environment that sensor nodes normally operate in as well as significant resource and energy limitations.

4.2 DATA DISSEMINATION AND GATHERING

An essential component of WSNs is how data and queries are transmitted from the base station to the location where the target phenomena are being observed. Direct data interchange between each sensor node and the base station is an easy way to complete this task. Nevertheless, a single-hop solution is expensive since nodes that are distant from the base station risk fast running out of energy, substantially reducing the network's lifetime.This is especially true if the wireless sensors are set up to cover a big area of land or if they are movable and could wander away from the base station.

Data sharing between the sensors and base stations is typically done utilizing multi-hop packet transmission over short communication distances to solve the drawbacks of the single-hop strategy. In particular in very dense WSNs, such an approach results in significant energy savings and lowers communication interference amongst sensor nodes competing for the channel. Figure 1 shows data forwarding between the sensors that collect data and the sinks that make it available. Data gathered by the sensors is transferred to the base station utilizing multi-hop pathways in response to requests made by the sinks or when particular events take place in the region being monitored.It is important to note that, depending on the application, sensor nodes may gather data that has been linked while travelling to the base station.

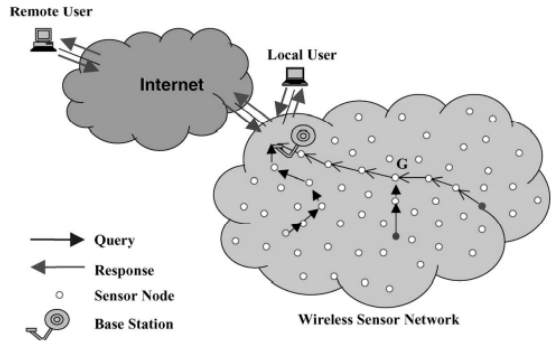


Figure 1: Multi-hop data and query forwarding

Data packets between the source and the destination must be forwarded by intermediate nodes in a multi-hop WSN. The main purpose of the routing algorithm is to choose which group of intermediary nodes should be used to create a data-forwarding path between the source and the destination. Routing in large-scale networks is, in general, a demanding problem whose solution must take into account a number of complex design requirements, such as correctness, stability, and optimality with regard to several performance metrics. In order to meet the traffic demands of the supported application while extending the life of the network, it is necessary to solve additional issues brought about by the fundamental features of WSNs in combination with severe energy and bandwidth limits.

4.3 ROUTING CHALLENGES AND DESIGN ISSUES IN WIRELESS SENSOR NETWORKS

Although WSNs and wired and ad hoc networks have a lot in common, they also have a few distinctive qualities that make them stand out from other networks. These distinctive qualities bring novel routing design requirements that go above and beyond those commonly found in wired and wireless ad hoc networks into sharp focus. Achieving these design specifications poses a different and particular set of difficulties. These difficulties can be attributed to a number of things, including severe energy constraints, constrained computing and communication capabilities, the dynamically changing environment in which sensors are deployed, special data traffic models, and application-level requirements for quality of service.

4.3.1 Network Scale and Time-Varying Characteristics

With severe energy constraints, sensor nodes can only operate with limited processing, storage, and communication capabilities. The densities of the WSNs may differ greatly, ranging from very sparse to very dense, due to the numerous possible sensor-based applications. However, in many applications, the sensor nodeswhich can sometimes number in the hundreds or even thousandsare set up haphazardly and frequently without supervision over large coverage regions. As a result of the requirement to self-organize and conserve energy, sensor nodes in these networks behave in a dynamic and highly adaptive manner, continually adjusting to their level of activity or lack thereof.In order to avoid the severe performance deterioration of the supported application, sensor nodes may also be necessary to modify their behavior in response to the irregular and unpredictable behavior of wireless connections induced by excessive noise levels and radio-frequency interference.

4.3.2 Resource Constraints

In order to deploy sensor nodes widely and cheaply, complexity is kept to a minimum. WSNs must operate on limited battery reserves while achieving a long lifetime, hence energy is a major challenge. A significant source of power consumption in wireless networks is multihop packet

transmission. The duty cycle of the wireless sensors can be dynamically controlled to lower energy consumption. Yet, many mission-critical sensor applications make the energy management challenge particularly difficult. Due to the demands of these applications, it is necessary to concurrently maintain a set level of sensing and communication performance limits. Hence, the topic of how to create scalable routing algorithms that can function well under a variety of performance limitations and design requirements arises.The development of these protocols is fundamental to the future of WSNs.

4.3.3 Sensor Applications Data Models

The information flow between the sensor nodes and the data sink is described by the data model. How data are sought and used depends a lot on the type of application used in these models. To fulfil the data-gathering demands and interface requirements of various sensor applications, a number of data models have been proposed. Data collection models for a class of sensor applications must be based on periodic sampling or be triggered by the occurrence of particular events. Before being sent to the data sink, data can be collected, stored, and possibly processed by a sensor node in other applications.A third category of sensor applications, however, necessitates bidirectional data models since they call for two-way communication between sensors and data sinks.The complexity of the route design challenge is increased by the requirement to support numerous data models. It becomes a massive design and engineering challenge to tailor the routing protocol to the particular data needs of an application while also supporting a wide range of data models and providing the best possible performance in terms of scalability, reliability, responsiveness, and power efficiency.

4.4 ROUTING STRATEGIES IN WIRELESS SENSOR NETWORKS

It is possible to argue that the WSN routing problem represents a conventional trade-off between responsiveness and efficiency. The overhead necessary to accommodate the constrained processing and communication capabilities of sensor nodes must be weighed against the requirement to do so. In a WSN, overhead is typically assessed in terms of mobile node processing demands, bandwidth use, and power consumption. The key to solving the routing problem is devising a plan to effectively balance these conflicting demands. Additionally, given the inherent properties of wireless networks, it is crucial to consider whether or not the existing routing protocols created for ad hoc networks are enough to handle this difficulty.

Ad hoc network routing methods can be divided into different categories based on how information is collected, retained, and used to compute paths based on the information that has been collected. Proactive, reactive, and hybrid techniques can all be distinguished from one another. The proactive approach, also known as table-driven, focuses on redistributing routing data on a regular basis to keep routing tables accurate and

consistent across all network nodes. The network's structure can either be flat or hierarchical. Flat proactive routing strategies may be able to determine the best routes. In a situation where things are changing quickly, the overhead needed to compute these pathways can be too high.Large ad hoc networks' routing needs are better met by hierarchical routing.Reactive routing techniques create routes to a small number of destinations as needed. Typically, these methods don't keep track of global data on all network nodes. In order to find paths between a source and a destination, they must consequently rely on a dynamic route search. Often, this entails flooding a route discovery query with responses that are sent back via the opposite direction. The ways in which the flooding process is managed by the reactive routing strategies to cut down on communication overhead and the ways in which routes are calculated and reestablished when failure occurs are different.

In order to achieve stability and scalability in massive networks, hybrid techniques rely on the existence of network structure. These techniques divide the network into clusters that are mutually nearby and are dynamically maintained as nodes enter and exit the clusters to which they are assigned. Clustering offers a framework that can be used to condense the routing algorithm's response to alterations in the network environment. The use of proactive routing within a cluster and reactive routing between clusters can be used to create a hybrid routing method. The primary difficulty is lowering the overhead needed to keep the clusters running.

4.4.1 WSN Routing Techniques

The routing protocols for WSNs must take into account the network nodes' power and resource constraints, the wireless channel's time-varying quality, and the potential for packet loss and delay. There have been various proposed routing strategies for WSNs to address these design constraints. All nodes are treated as peers in a flat network topology used by one class of routing techniques. A flat network architecture provides a number of benefits, including low infrastructure maintenance costs and the opportunity for various paths to be discovered between communication nodes for fault tolerance.

A second category of routing protocols puts a structure on the network in order to increase its stability, scalability, and energy efficiency. In this class of protocols, network nodes are grouped together into clusters, with the cluster leader, for instance, being the node with the highest residual energy. The cluster leader is in charge of directing information between clusters and organizing activity inside the cluster. Clustering has the ability to save energy use and increase network longevity.

A third category of routing protocols uses a data-centric strategy to spread information throughout the network. The method uses attribute-based naming, where a source node instead of a specific sensor node queries an attribute for the phenomena. By giving duties to sensor nodes and expressing inquiries about particular properties, the interest diffusion is accomplished. Various tactics, such as broadcasting, attribute-based

multicasting, geo-casting, and any casting, can be used to convey interests to the sensor nodes.

A sensor node is addressed by a fourth class of routing protocols based on its location. Applications where the node's location inside the network's geographic coverage is important to the query sent by the source node can benefit from location-based routing. Such a query could indicate the location of a certain place in the network environment or a specific area where a phenomenon of interest might occur.

4.4.2 Flooding and Its Variants

In both wired and wireless ad hoc networks, flooding is a typical approach widely employed for path discovery and information dissemination. The routing approach is straightforward and doesn't rely on pricey network topology upkeep or difficult route discovery algorithms. Each node that receives a data or control packet delivers the packet to all of its neighbors as part of the reactive strategy known as flooding.An information package follows every route after transmission. The packet will ultimately get there unless the network is shut down. Moreover, the transmitted packet takes the new routes when the network architecture changes. The idea of flooding in a data communications network is depicted in Figure 2. As seen in the image, flooding in its most basic form can lead to network nodes endlessly replicating packets.

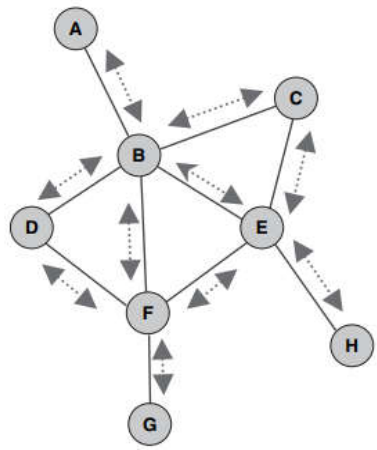


Figure 2:Flooding in data communications networks

A hop count field is typically present in a packet to prevent it from recirculating endlessly in the network. The hop count is initially chosen to roughly equal the network's diameter. The hop count decreases by one for each hop the packet makes as it moves through the network. The packet is only discarded when the number of hops reaches zero. A time-to-live field, which keeps track of how many time units a packet is permitted to live in the network, can be used to achieve a similar result. The packet is no longer forwarded once this period has passed.By giving each data packet a unique identification, flooding can be made even worse by making each network node discard every packet it has already

transmitted.A recent history of the traffic must be kept as part of this method in order to track which data packets have already been forwarded.

Flooding has a number of drawbacks when employed in WSNs, despite how straightforward its forwarding rule is and how little maintenance it needs. The first problem with flooding is that it might cause a traffic jam. Duplicate control or data packets transmitted repeatedly to the same node result in this unwanted outcome. Flooding's second negative aspect is the overlap issue it creates. Resource blindness is the third and most severe disadvantage of flooding. Flooding utilizes a straightforward forwarding rule that ignores the sensor nodes' energy limitations while routing packets. As a result, the node's energy may quickly run out, drastically decreasing the network's lifetime.

Despite its straightforward forwarding rule and reasonably inexpensive cost A derivative strategy called gossiping has been suggested as a remedy for flooding's drawbacks. Similar to flooding, gossiping also relies on a straightforward forwarding rule and does not demand expensive topology maintenance or sophisticated route discovery techniques. Instead of broadcasting a data packet to every neighbor as is the case with flooding, gossiping calls for each node to deliver the incoming packet to a randomly chosen neighbor. The packet is forwarded to the neighbor picked by the randomly chosen neighbor after it has been received by that neighbor.Until the packet arrives at its designated location or the maximum hop count is reached, this procedure iteratively continues. By restricting the number of packets each node delivers to its neighbor to one copy, gossip avoids the implosion issue. A packet may have high latency while travelling to its destination, especially in a large network. This is primarily due to the protocol's randomness, which essentially investigates one path at a time.

4.4.3 Sensor Protocols for Information via Negotiation

A family of data-centric negotiation-based information dissemination protocols for WSNs is called Sensor Protocols for Information through Negotiation (SPIN). These protocols' principal goal is to effectively distribute observations made by individual sensor nodes to all of the network's sensor nodes. In WSNs, straightforward protocols like flooding and gossiping are frequently proposed to achieve information dissemination. Flooding calls for sending copies of the data packet to each node's neighbors until it reaches every node in the network.In contrast, gossip just needs that a node receiving a data packet transmit it to a randomly chosen neighbor, using randomization to decrease the number of duplicate messages.

Both flooding and gossiping are easy and appealing because they don't require topology upkeep and use basic forwarding principles. Nevertheless, as the network's size and traffic load increase, these algorithms' performance quickly deteriorates in terms of packet latency and resource use. This performance flaw is often brought on by geographic overlap and traffic implosion.Several copies of the same data

are sent to the same sensor node as a result of traffic implosion.Geographic overlap, on the other hand, results in nodes covering the same region disseminating identical data items to network sensor nodes unnecessarily. Basic protocols like floods and chitchat do not modify their behavior to adjust communication and computation to the present status of their energy supply. This lack of resource awareness and flexibility could significantly shorten the network's lifespan as highly active nodes risk quickly running out of energy.

The fundamental goal of SPIN and its associated family members is to fix the flaws and improve the performance of traditional information distribution protocols. This family of protocols' fundamental principles are resource adaptability and data negotiation. Before any data are sent across network nodes, semantic-based data negotiation demands that nodes running SPIN "learn" about the content of the data. By having nodes associate metadata with the data they produce, SPIN takes advantage of data naming to undertake negotiations before delivering the actual data.A receiver may send a request to access the advertised data if they show interest in its content. By ensuring that data are sent only to interested nodes, this type of negotiation prevents traffic implosion and greatly lowers the amount of redundant data that is transmitted over the network. Furthermore, by allowing nodes to limit their requests to mention only the data they are interested in receiving, the usage of meta data descriptors avoids the chance of overlap.

Resource adaption enables SPIN-powered sensor nodes to adjust their operations to the state of their available energy sources. Before sending or processing data, each node in the network can probe the corresponding resource management to keep track of its resource usage. When the energy level drops, the node may scale back or stop performing specific tasks, such forwarding third-party information and data packets. The SPIN resource adaptability function enables nodes to prolong their lives and, as a result, the network's lifespan.

Three different sorts of messages are used by SPIN-running nodes for negotiation and data delivery. New data is advertised among nodes using the first message type, ADV. A network node can advertise its data to the other nodes in the network by first sending an ADV message that contains the information describing the data. Requesting an advertised piece of valuable data is done using the second message type, REQ.A network node interested in obtaining specific data sends a REQ message to the metadata advertising node after receiving an ADV containing metadata, and the node subsequently sends the requested data. The actual data gathered by a sensor and a metadata header are both included in the third message type, DATA.Generally speaking, the data message is bigger than the ADV and REQ messages. The latter messages are typically much smaller than the corresponding data packet and just carry metadata. Energy use can be significantly reduced by limiting the duplicate transmission of data messages utilizing semantic-based negotiation.

The fundamental operation of SPIN is shown in Figure 3, where sensor node A, the data source, sends an ADV message containing the information characterizing its data to sensor node B, its close neighbor. Node B sends a REQ message to request the data after expressing interest in it. Node B transmits an ADV message after receiving the data to inform its close neighbors of the new information. Only nodes C, E, and G, three of these neighbors, show interest in the information. These nodes send node B a REQ message, and node B responds by sending the requested data to each of the requesting nodes.

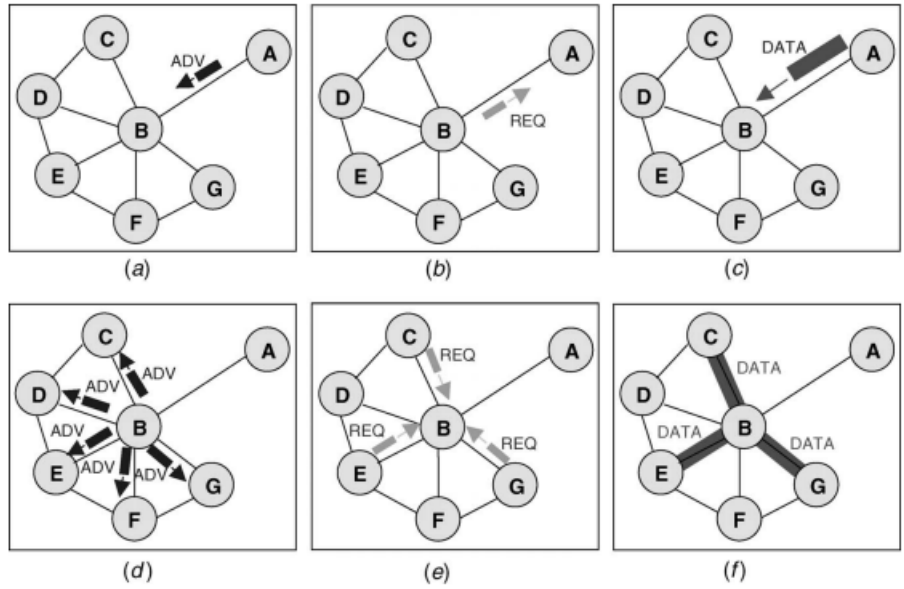


Figure 3: SPIN basic protocol operations

4.4.4 Low-Energy Adaptive Clustering Hierarchy

A routing technique called Low-energy Adaptive Clustering Hierarchy (LEACH) is created to gather and send data to a base station or other data sink. The primary goals of LEACH are:

- Prolongation of the network's life
- Each sensor node in the network using less energy
- Using data aggregation to minimize communication messages

LEACH uses a hierarchical technique to divide the network into a number of clusters in order to accomplish these goals. A chosen cluster leader is in charge of overseeing each cluster. The cluster leader takes on the obligation to complete several duties.Data from the cluster's members are periodically collected as the first task.The cluster head aggregates the data after collecting it in an effort to eliminate duplication among associated values. The direct transmission of the aggregated data to the base station is the cluster head's second primary responsibility. The sent data is combined and sent over a single hop. Figure 4 shows the network model utilized by LEACH.The creation of a TDMA-based schedule, in which each node of the cluster is given a time slot that it can use for transmission, is the cluster head's third major responsibility. The cluster head broadcasts the schedule

to the other cluster members. LEACH nodes employ a code-division multiple access-based communication protocol to lessen the possibility of sensor collisions both inside and outside the cluster.

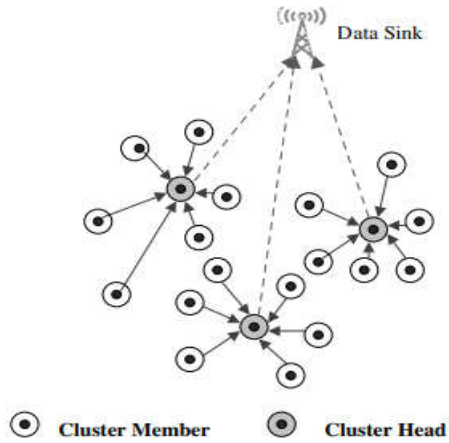


Figure 4: LEACH network model

LEACH's fundamental processes are divided into two separate phases. Figure 5 presents an illustration of these stages. Cluster creation and cluster-head selection are the two steps that make up the setup phase's initial phase. Data gathering, aggregation, and delivery to the base station are the main objectives of the second phase, often known as the steady-state phase. To reduce the protocol overhead, it is believed that the setup will last considerably less time than the steady-state phase.

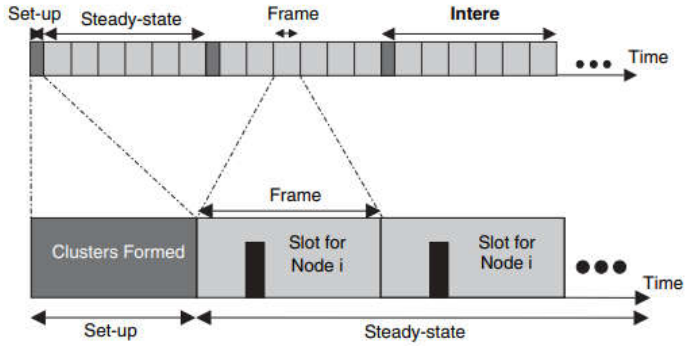


Figure 5: LEACH phases

LEACH possesses a number of characteristics that allow the technique to use less energy. All sensor nodes in LEACH are required to use a certain amount of energy since they round-robin take on the cluster head function according to their remaining energy. As LEACH is a fully distributed algorithm, the base station's control information is not needed. Local cluster management eliminates the requirement for knowledge of the entire world's networks. Additionally, since nodes are no longer required to send their data straight to the sink, data aggregation by the cluster also makes a significant contribution to energy savings. LEACH outperforms traditional routing protocols, such as direct transmission and multihop routing, minimum-transmission-energy routing, and static clustering-based routing algorithms, as demonstrated by simulation.

4.4.5 Power-Efficient Gathering in Sensor Information Systems

A family of routing and information-gathering protocols for WSNs includes hierarchical PEGASIS and the power-efficient collecting in sensor information systems (PEGASIS) extension. PEGASIS has two dual-purpose goals. Secondly, by establishing a high level of energy efficiency and uniform energy usage across all network nodes, the protocol attempts to increase the lifespan of a network. Second, the protocol aims to shorten the time that data takes to reach the sink.

The network model that PEGASIS is considering assumes a uniform distribution of nodes over a region. It is assumed that nodes are aware of the placements of all other sensors globally. Also, they can adjust their power to cover virtually any area. The nodes might additionally have radio transceivers that support CDMA. Data collection and delivery to a sink, often a wireless base station, is the responsibility of the nodes. The objective is to create a routing structure and an aggregation system to balance energy consumption among the sensor nodes, reduce energy consumption, and transmit the aggregated data to the base station with the least amount of delay possible.PEGASIS employs a chain structure as opposed to other protocols, which rely on a tree structure or a cluster-based hierarchical organization of the network for data collection and dissemination.

Nodes communicate with their nearest neighbors using this structure. The farthest node from the sink is where the chain is assembled from. Network nodes are gradually added to the chain, starting from the end node's nearest neighbor. The nearest neighbor to the top node in the existing chain is added to the chain first when adding nodes that are currently outside of it, and this process continues until all nodes have been added. A node utilizessignal strength to calculate the distance to each of its neighbors, then uses that distance to determine which neighbor is the closest. This data is used by the node to modify the signal strength such that only the closest node is audible.

The chain leader is chosen from among the nodes in the chain. Its duty is to deliver the compiled data to the base station. After each round, the chain's position is changed by the chain leader. Rounds can be regulated by the data sink, and a strong beacon that it issues can trigger the change from one round to the next. The chain's nodes alternate taking the leading position to ensure that the overall energy consumption of the network is balanced.However, it should be noted that nodes acting as chain leaders may be arbitrarily far from the data sink.

In PEGASIS, data aggregation is accomplished along the chain. The aggregation procedure can be carried out successively as follows in its most basic form. The last node on the right end of the chain receives a token first from the chain leader. The end node sends its data to its downstream neighbor in the chain towards the leader after getting the token. The downstream neighboring node receives the data from the neighboring node that aggregates them. The cycle repeats itself until the

leader receives the compiled data.The same aggregation method is used until the data reach the leader after the leader provides a token to the left end of the chain after getting the data from the right side of the chain. The leader gathers the data and sends it to the data sink after receiving it from both ends of the chain. Although straightforward, the sequential aggregation approach may cause significant delays in the delivery of the aggregated data to the base station. However, if arbitrarily close simultaneous transmission cannot be done without causing signal interference, then such a sequential system might be required.

Using parallel data aggregation along the chain is one possible method to lessen the time needed to send aggregated data to the sink. If the sensor nodes have CDMA-capable transceivers, a high level of parallelism can be attained. A hierarchical structure can be "overlaid" over the chain and utilized to do data aggregation using the additional capability to carry out arbitrarily near transmissions without interference. Nodes at a specific level of the hierarchy broadcast to a close neighbor at a higher level of the hierarchy once each round. The leader at the top of the hierarchy receives the aggregated data at the end of this process. The latter sends the base station the final data aggregate.

Consider the scenario shown in Figure 6 to provide an example of the chain-based approach. In this illustration, it is assumed that every node uses a greedy algorithm to build the chain and has global knowledge of the network. Moreover, it is presumable that nodes transmit to the base station in rounds, with node i mod N where N is the total number of nodes transmitting the aggregate data to the base station in round i. Node 3, in the chain's third position, is the round 3 leader according to this assignment.The neighbor to the right must get data from every node in an even position. Node 3 stays in an odd position at the following level. Because of this, all nodes in an even position combine their data and send it to the appropriate neighbors. Node 3 is no longer at an unusual position at the third level. The only node other than node 3 to reach this level is node 7, which gathers its data and delivers it to node 3. Node 3 then combines the information received with its own information before sending it to the base station.

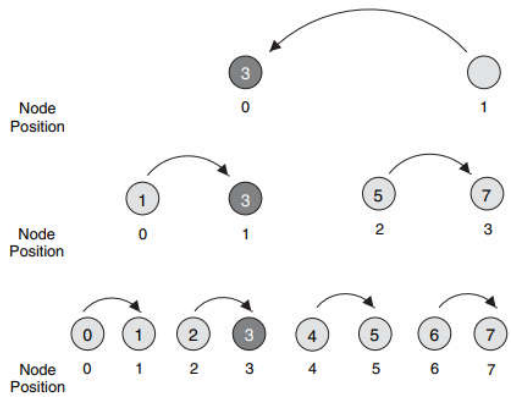


Figure 6: Chain-based data gathering and aggregation scheme

The chain-based binary technique uses significantly less energy because nodes work in close proximity to one another. In addition, the technique ensures that the leader will receive the aggregated data after log2N steps because the hierarchical, treelike structure is balanced. In PEGASIS, a chain-based binary aggregation approach has been utilized as an alternative to high parallelism. It has been demonstrated that the strategy works best with CDMA-capable sensor nodes in terms of the energy-delay product needed for each round of data collection, a parameter that balances the energy and delay costs.

4.4.6 Directed Diffusion

A data-centric routing protocol for information collection and sharing in WSNs is called directed diffusion. The protocol's primary goal is to generate significant energy savings in order to increase the network's lifespan. This goal is accomplished through directed diffusion, which maintains message-exchange interactions between nodes contained inside a certain network area. Direct diffusion can nevertheless achieve resilient multipath delivery and adapt to a small portion of network pathways by using localized contact. Significant energy savings are achieved by the protocol's special feature and the nodes' capacity to aggregate responses to queries.

Direct diffusion's primary components are interests, data messages, gradients, and reinforcements. Directed diffusion uses a publish-and-subscribeinformation model in which an inquirer expresses an interest using attribute–valuepairs. An interest can be thought of as a question or an interrogation that expresses the object of the inquiry. An example of how an interest in hummingbirds can be communicated using a set of attribute-value pairs is shown in Table 1. Sensor nodes that can serve the request return the relevant data.

Table 1:Interest Description Using Value and Attribute Pairs

Attribute–Value Pair	Description
Type = Hummingbirds	Detect hummingbird location
Interval = 20 ms	Report events every 20 ms
Duration = 10 s	Report for the next 10 s
Field = $[(x_1, y_1), (x_2, y_2)]$	Report from sensors in this area

The data sink periodically broadcasts an interest message to each neighbor for each current sensing task. As a named data interest, the message spreads throughout the sensor network. This exploratory interest message's main goal is to see if there are any sensor nodes that can accommodate the desired interest.An interest cache is maintained by each sensor node. The interest cache contains entries for each different interest. The cache item has a timestamp field, several gradient fields for each neighbor, and a duration field, among other fields.The timestamp of the most recent matching interest was contained in the timestamp column. Both the data rate and the direction in which data are to be delivered are specified in each gradient field. The interval attribute of the interest is used to

determine the data rate's value.The duration field provides an estimate of the interest's lifespan. The attribute's timestamp is used to calculate the duration's value. Interest spread in a WSN is shown in Figure 7.

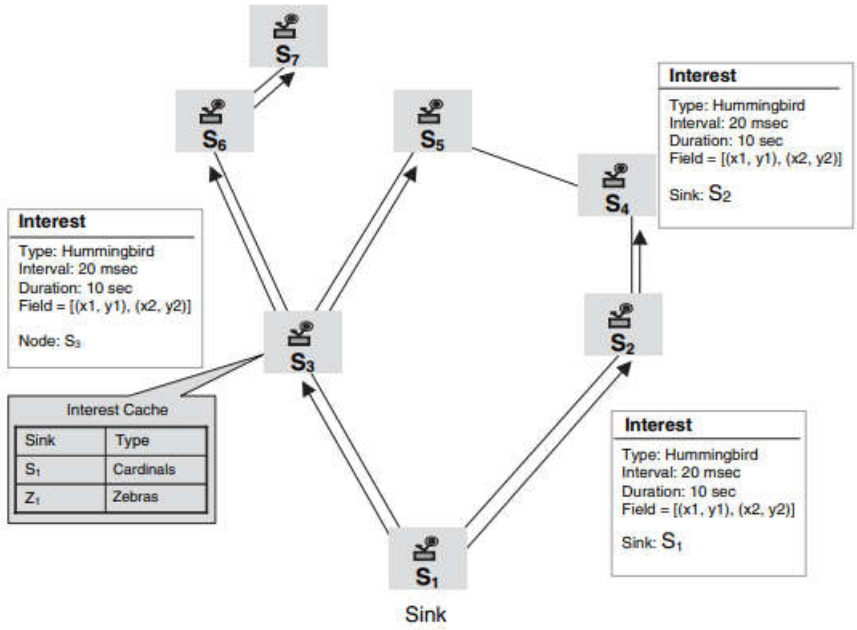


Figure 7: Interest propagation

A gradient can be compared to a reply link that directs traffic to the node next to the source of the interest. The discovery and development of pathways between the data sinks interested in the identified data and the nodes that can serve the data are made possible by the diffusion of interests over the entire network in conjunction with the establishment of gradients at the network nodes. An event-detection sensor node looks for a corresponding record in its interest cache. In the event that a match is found, the node computes the greatest event rate among all of its outgoing gradients first.It then configures its sensing component to sample events at this maximum pace. The node then broadcasts an event summary to each neighbor it has a gradient for. When a neighboring node receives data, it looks through its cache for an entry with the same interest. In the absence of a match, the node discards the data message without taking any further action. If such a match occurs and the received data message does not already have a corresponding data cache entry, the node adds the message to the data cache and broadcasts the data message to the surrounding nodes.

A node checks its interest cache after receiving an interest to see if it already has an entry for that interest. The receiving node makes a new cache entry if such an entry does not already present. The node then instantiates the parameters of the freshly generated interest field using the data from the interest. Moreover, the entry is configured to have a single gradient field pointing at the neighboring node from where the interest is received, with the event rate provided. The node updates the timestamp and duration fields of the matching entry if there is a match between the interest received and a cache entry.The node adds a gradient with the

value specified in the interest message if the entry does not already have one for the sender of the interest. The node merely updates the timestamp and duration fields if the matching interest item has a gradient for the interest sender. When a gradient expires, it is taken out of the interest entry. The first gradient configuration is shown in Figure 8.

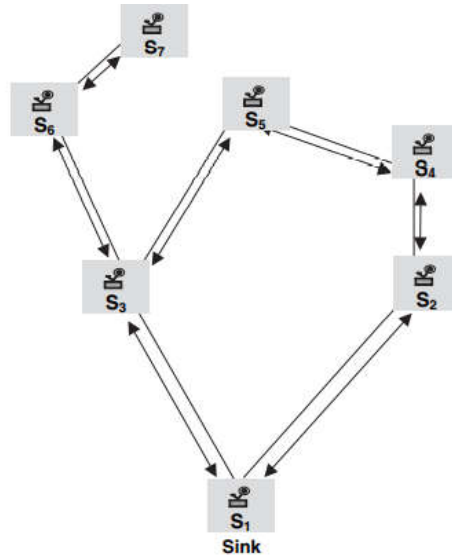


Figure 8: Initial gradient setup

A sink creates numerous pathways during the gradient building stage. By raising its data flow, the sink can employ these routes to events of greater quality. Via a process called path reinforcement, this is accomplished. The sink may decide to support one or more specific neighbors. In order to accomplish this, the sink resends the initial interest message across the chosen paths at a greater data rate, reinforcing the source nodes' incentives to submit data more frequently. The most effective path can then be kept while the others are negatively reinforced.By timing out any high-data-rate gradients in the network besides those that are explicitly reinforced, negative reinforcement can be produced. Data delivery along a reinforced channel is depicted in Figure 9.

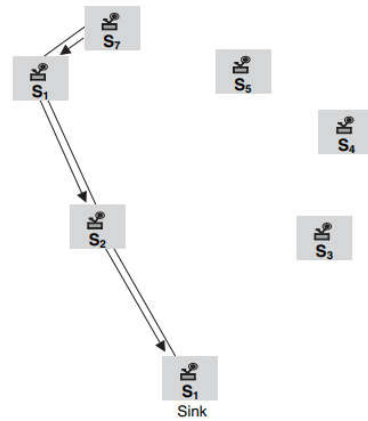


Figure 9: Data delivery along a reinforced path

Directed diffusion can be used to fix link failures brought on by external elements that affect the communications channel as well as node failures or performance degradation brought on by node energy loss or total depletion. Data loss or a reduction in rate are often indicators of these failures. An alternative path that is sending at slower rates can be found and strengthened when a path connecting a sensing node and the data sink fails. Lossy links can also be negatively reinforced by allowing the neighbor's cache expire over time or by delivering interests at the exploratory data rate.

Diffusion that is directed has the ability to save a lot of energy. It can attain comparatively high performance over non-optimized paths thanks to its localized interactions.The resulting diffusion processes are also resilient to a variety of network dynamics. Node addressing is unnecessary due to its data-centric methodology.The query-on-demand data model, however, is closely related with the directed diffusion paradigm. This might restrict its applicability to applications that make sense as a data model, where the process of interest matching can be carried out quickly and clearly.

4.4.7 Geographical Routing

Geographical routing's major goal is to create an effective route search towards the target using location data. Geographical routing is ideal for sensor networks because it removes redundant packets from diverse sources, allowing for a reduction in the number of broadcasts to the base station. In sensor networks, the need for data aggregation results in a shift from the traditional address-centric paradigm, where communication takes place between two addressable endpoints, to a data-centric paradigm, where the content of the data is more significant than the identity of the node that gathers the data.In this new paradigm, an application may send a query to learn more about a phenomenon that is occurring in or close to a certain physical location or prominent landmark. For instance, scientists interested in traffic flow patterns may want to know the typical quantity, size, and speed of cars using a certain stretch of a roadway. The data substance is more significant than the name of the sensors that gather and distribute data about traffic flow on a particular stretch of route. Also, a number of nodes that are situated near the specified stretch of the roadway might take part in gathering and aggregating the information required to respond to the query.Conventional routing techniques are not well adapted to handle multidimensional queries that are spatially specific because they are typically designed to find a path between two addressable endpoints. Geographical routing, on the other hand, uses location data to get from one place to another, using the location of each node as the address.

Geographical routing has a low computational and communication overhead and is compatible with data-centric applications. With conventional routing methods, such as those used in distributed shortest-path routing protocols for wired networks, a router may need to be aware of the full network topology or a summary of it in order to determine the shortest way to each destination.Also, routers are required to update the

state representing the current topology on a regular basis and whenever a link fails in order to maintain correct pathways to all destinations. According to the product of the number of routers and the rate at which the topology changes in the network, the requirement to update the topology state continuously may result in significant overhead.

Geographical routing, on the other hand, does not necessitate keeping a "heavy" state at the routers to monitor the topology's present state. Making the appropriate forwarding decisions simply requires the propagation of single-hop topological information, such as the location of the "best" neighbor. Maintaining internal data structures like routing tables is unnecessary due to the self-describing nature of geographical routing and its localized decision-making process.As a result, the control overhead is significantly decreased, improving its scalability on big networks. These characteristics make geographical routing a workable option for routing in sensor networks with limited resources.

4.5 SUMMARY

The routing challenge is extremely difficult due to the features of WSNs and the environment in which sensor nodes are frequently installed.In this chapter, we concentrated on the fundamental problems with routing in WSNs and discussed various methods for creating routing protocols for these networks. We provide a quick taxonomy of the fundamental routing techniques utilized to balance responsiveness and energy efficiency. We reviewed several protocols that deal with the issue of routing in contemporary WSNs. The routing issue has several workable solutions that have come to light. As the use of WSNs in many industries grows, improvements in network hardware and battery technology will open the door to realistic, economically viable implementations of these routing protocols.

4.6 LIST OF REFERENCES

1) Protocols and Architectures for Wireless Sensor Network, Holger Kerl, Andreas Willig, John Wiley and Sons, 2005

2) Wireless Sensor Networks Technology, Protocols, and Applications,Kazem Sohraby, Daniel Minoli and TaiebZnati, John Wiley & Sons, 2007

3) Mobile communications, Jochen Schiller,2nd Edition, Addison wisely, Pearson Education,2012

4) Fundamentals of Wireless Sensor Networks, Theory and Practice, WaltenegusDargie, Christian Poellabauer, Wiley Series on wireless Communication and Mobile Computing, 2011

5) Networking Wireless Sensors, Bhaskar Krishnamachari, Cambridge University Press, 2005

4.7 UNIT END EXERCISES

1) What do you mean by Data Dissemination and Gathering?
2) Illustrate Routing Challenges and Design Issues in Wireless Sensor Networks.
3) Explain Network Scale and Time-Varying Characteristics.
4) What is Resource Constraints?
5) Write a note on Sensor Applications Data Models.
6) Explain Routing Strategies in Wireless Sensor Networks.
7) Describe WSN Routing Techniques.
8) Write a note on Flooding and Its Variants.
9) What is Sensor Protocols for Information via Negotiation?
10) Explain Low-Energy Adaptive Clustering Hierarchy.
11) Describe Power-Efficient Gathering in Sensor Information Systems.
12) Write a note on Directed Diffusion.
13) What is Geographical Routing?

5

TRANSPORT CONTROL PROTOCOLS

Unit Structure

5.0 OBJECTIVES

- To understand the different traditional transport controlprotocol
- To get familiar with the design issues, examples and performance of Transport Control Protocols

5.1 INTRODUCTION

Physical, data link, network (or internetworking), transport, and other higher levels like session, presentation, and application make up the architecture of computer and communication networks. For its immediate upper layer, a service consumer, each lower layer serves as a service

provider. Via service access points, neighboring layers communicate with one another (SAPs). For instance, the network layer, which sits directly above the link layer, receives link services from the data link layer. The transport layer, which is the layer above it, receives addressing and routing services from the network layer, while the layers above it receive message transportation services from the transport layer. In this architecture, practically all nodes only include the lower three layers.Yet, the transport and the layers above it only exist at end points or hosts and carry out end-to-end protocol operations.

End-to-end segment transportation is offered by the transport layer, in which messages are split up into a series of segments at the source and then put back together again at the destination nodes. The procedures utilized to deliver the segments to the target nodes and/or the underlying delivery protocol structures are not concerns of the transport layer. Examples of transport protocols include the user datagram protocol (UDP), the sequenced packet exchange protocol (SPX), the transport control protocol (TCP), and NWLink (Microsoft's implementation of IPX/SPX). The Internet frequently uses TCP and UDP.TCP can be classified as either connection-oriented and connectionless.

5.2 TRADITIONAL TRANSPORT CONTROL PROTOCOLS

5.2.1 TCP

On the Internet, TCP is the most widely used connection-oriented transport control protocol. TCP is where some applications, such HTTP and FTP, are located. In order to provide dependable, orderly, controllable, and elastic transmission, TCP makes advantage of network services offered by the IP layer. TCP operation is divided into three stages:

1) **Connection establishment:** During this stage, a logical connection for TCP is formed. A logical connection is an association between a TCP sender and recipient that can be uniquely identified by their IP addresses and TCP port numbers. There could be many connections active between endpoints at once. These connections will have separate TCP port numbers despite sharing the same IP address. A three-way handshake is used by TCP to establish a connection.The TCP sender and receiver will negotiate parameters like the beginning sequence number, window size, and others during the handshake and will let each other know when data transmission can start.

2) **Data transmission:** TCP enables dependable and well-organized information transfer between the sender and the recipient. When a segment is lost, TCP utilizes (accumulative) ACK to find it. The segment header's sequence number allows for an ordered transmission. TCP also supports flow control and congestion control with sender-adjustable transmission rates. This task is carried out via TCP using a window-based approach, in which the sender manages a variable

called cwnd (congestion window). The maximum number of segments that the TCP sender can send is cwnd. After receiving an ACK from the receiver or following a timeout, cwnd is updated. Both flow control and delivery notification are performed using ACK, therefore the two tasks are somewhat intertwined.

3) **Disconnect:** The connection will be cut off and the relevant resource released once the data transmission is finished.

Via cwnd, TCP controls its flow and congestion. The procedure is divided into three stages:

1) **Slow start:** All transmissions begin slowly by default. For each ACK that is received during this phase for a segment that was transmitted, the cwnd rises by one.If ACK is not received because of segment loss, cwnd consequently rises.

2) **Congestion avoidance:** The system enters the congestion avoidance state when cwnd reaches a maximum value (threshold). After each ACK is received in this condition, cwnd is only increased by 1/cwnd. The sender keeps track of time for each segment sent. The system enters the slow start phase once more, the threshold is set to half of the current cwnd, the segment timer is doubled, and the cwnd is reset if the timer ends before an ACK corresponding to the segment is received. Round-trip time (RTT), which is determined by the ACK, is used to update the timer. A segment has been lost during transmission if the sequence number acknowledged in two consecutive ACKs is in sequence.In this scenario, cwnd will be cut in half at the same time that the system status switches to fast recovery and fast retransmission (FRFT).

3) **FRFT:** The same technique that updates cwnd in the congestion avoidance state is also employed in the rapid recovery and fast retransmission state. There is no need to reset cwnd because, usually speaking, random segment loss does not necessarily indicate that there is high congestion.A timeout, however, typically denotes significant traffic volume and/or a broken link.

TCP mechanisms provide adaptable flow and congestion control as a result. As can be seen from the foregoing, (1) high throughput results from cwnd increasing quickly and oscillating around a big value when there is little to no congestion and few segment losses. On the other hand, if cwnd has a low value, the TCP throughput will be low. (2) When RTT is low, ACKs are received rapidly, and cwnd rises similarly quickly. The sender will consequently experience significant throughput. (3) It goes without saying that high throughput also results from big segment sizes. Moreover, the theoretical analysis of TCP confirms these.

5.2.2 UDP

UDP is a transport protocol without connections. This protocol does not have any techniques for recovering lost information because it exchanges

datagrams without a sequence number between the transmitter and the receiver. It cannot guarantee ordered transmission because the datagrams do not include a sequence number. Moreover, it lacks features for flow control or congestion. As UDP does not perform congestion or flow control, it can end up outperforming TCP in situations where both protocols are active. A TCP-friendly rate control (TFRC) for UDP has been suggested in recent years to implement a certain amount of control in this protocol.When TCP and UDP are available on a connection, the fundamental principle underlying TFRC is to deliver almost comparable throughput to both protocols.

5.2.3 Mobile IP

In order to offer terminal mobility in an all-IP network, mobile IP is presented as a global mobility management solution at the network layer. TCP/early IP's design did not take mobility into account. Currently, the IP address serves as both a terminal identity and a network location identifier for terminals. Addresses are also utilized throughout the routing procedure. To separate the two, however, there must be some sort of process. Two new entities and one new IP address are introduced by mobile IP, which is intended to alleviate this issue.The two new entities are 1] the home agent (HA), which is situated in the home network of the mobile terminal and is in charge of managing its IP addresses and packet forwarding, and 2] the foreign agent (FA), which is situated in the network that the mobile terminal visits. HA and FA can be addressed worldwide and have static IP addresses.Care of address (COA), the IP address acquired from FA after the mobile device enters a new network, is the new IP address introduced for mobility.

When a terminal enters a new network, it registers with the FA of the new network and then gets a COA. This is how mobile IP works. The COA is now communicated to the terminal's HA by either the terminal or the FA. The HA will then pass the packets to the mobile terminal's COA when a related terminal sends data to the mobile terminal. Direct packet transmission occurs from the mobile terminal to the matching terminal.The triangular routing method, which results in a longer path from the corresponding terminal to the mobile and, thus, low efficiency, is hence an asymmetrical routing procedure between the two terminals. Even though the physical link may have enough bandwidth, the TCP sender is obliged to drop its rate during the mobility process since handoff comes as a result of movement and may result in packet loss and TCP timeout.

5.3 TRANSPORT PROTOCOL DESIGN ISSUES

WSNs should be developed with consideration for energy conservation, traffic control, data dissemination reliability, security, and management. These challenges can be studied independently in each layer or cooperatively across layers and frequently include one or more tiers of the hierarchical protocol. Congestion control, for instance, might just affect the transport layer, whereas energy conservation might affect the physical, data connection, network, and even all other upper layers. Congestion

control and loss recovery are the two primary objectives of the design of transport control protocols. Finding the beginning of congestion and pinpointing its location and timing is necessary for congestion control.For instance, monitoring node buffer occupancy or link load can be used to find congestion (such as wireless channel). Selective packet dropping at a congestion point, as in active queue management (AQM) schemes, rate adjustment at the source node, as in the technique of additive increase multiplicative decrease (AIMD) in TCP, and the use of routing techniques are some of the methods used to control congestion in the traditional Internet. Because sensors have limited resources, it is important to carefully evaluate how to identify congestion and how to avoid it in WSNs.These protocols must take into account energy conservation, simplicity, scalability, and strategies for extending the lifespan of sensor batteries.One may, for instance, use an end-to-end mechanism, like the one found in TCP, or hop-by-hop backpressure, like the kind used in frame relay networks or the asynchronous transfer mode (ATM).End-to-end strategies are highly straightforward and reliable;however, they could increase network traffic. Yet, hop-by-hop methods typically identify congestion rapidly and add less extra network traffic as a result. While developing congestion control algorithms for WSNs, significant consideration should be given to the trade-off between end-to-end and hop-by-hop processes due to energy constraints at the sensors.

In wireless sensor networks, packet loss is typically brought about by poor wireless channel quality, sensor failure, and/or congestion. In order to convey accurate information, WSNs need to ensure a particular level of reliability at the application or packet level through loss recovery. Packet-level dependability is necessary for certain essential applications since they depend on the reliable transmission of every packet. Application reliability is more significant than packet-level reliability since certain applications only require a proportionately reliable transfer of packets. Wireless sensor networks can detect packet loss using the same conventional techniques as packet-switched networks.Each packet, for instance, may carry a sequence number, and a receiver may use sequence numbers to detect packet loss. Using an end-to-end or hop-by-hop control, ACK and/or NACK can be used to recover lost packets after packet loss has been detected. Efficiency is maintained in terms of energy if there aren't many packets in transit and few retransmissions are needed. Less in-transit packets may be produced as a result of effective congestion control. Fewer retransmissions occur as a result of an efficient loss recovery strategy. In conclusion, the issue of transport control protocols for sensor networks essentially comes down to energy efficiency.The following elements must to be taken into account while designing transport protocols for WSNs:

1) **Regulate traffic flow and ensure data delivery reliability:** Since the majority of the data come from the sensor nodes to the sink, there may be congestion around the sink.Although the MAC protocol can recover packets that have been lost due to bit errors, it is unable to handle packets that have been lost due to buffer overflow. A packet loss recovery mechanism, similar to the ACK and selective ACK used in

TCP, is required for WSNs.Furthermore, whereas in traditional networks every packet's accurate transmission is ensured, trustworthy delivery with WSNs may mean something else.For some sensor applications, WSNs simply need to reliably receive packets from a portion of the local sensors, not from every sensor node. The design of WSN transport protocols may benefit greatly from this discovery. Also, as it may reduce packet loss and hence increase energy conservation, using a hop-by-hop strategy for congestion control and loss recovery may be more successful. The intermediary nodes' need for buffers can be reduced by the hop-by-hop technique.

2) In order to hasten connection establishment, increase throughput, and decrease transmission latency, transport protocols for wireless sensor networks should be made simpler or use a connectionless protocol. The majority of WSN applications are reactive, which means they observe passively and hold data until an event occurs before delivering it to the sink. Due to an occurrence, these programs might only need to send a few packets.

3) As packet loss results in energy loss, WSN transport techniques should minimize packet loss. The transport protocol should employ active congestion control (ACC) to prevent packet loss at the cost of slightly lower connection usage. Congestion avoidance is started by ACC before congestion actually happens.As an illustration of ACC, the sender (or intermediary nodes) may lower its sending (or forwarding) rate when the downstream neighbors' buffer size rises above a particular threshold.

4) Fairness for a range of sensor nodes should be ensured by the transport control protocols.

5) A transport protocol should, if at all possible, be created with cross-layer optimization in mind. For instance, if a routing algorithm alerts the transport protocol of a route failure, the protocol can infer that packet loss is not due to congestion but rather to the failure of the route rather than congestion. In this scenario, the sender is free to stick with its current rate.

5.4 EXAMPLES OF EXISTING TRANSPORT CONTROL PROTOCOLS

Examples of several transport protocols designed for WSNs are shown in Table 1. Most examples can be grouped in one of the four groups: upstream congestion control, downstream congestion control, upstream reliability guarantee, and downstream reliability guarantee.

TABLE 1 Several Transport Protocols for WSNs

Attributes	CODA	ESRT	RMST	PSFQ	GARUDA
Direction	Upstream	Upstream	Upstream	Downstream	Downstream
Congestion					
Support	Yes	Passive	No	No	No
Congestion detection	Buffer occupancy channel condition	Buffer occupancy	—	—	—
Open- or closed-loop congestion control	Both	No	—	—	—
Reliability					
Support	No	Yes	Yes	Yes	Yes
Packet or application reliability	—	Application	Packet	Packet	Packet
Loss detection	—	No	Yes	Yes	Yes
End-to-end (E2E) or hop-by-hop (H&H)	—	E2E	HbH	HbH	HbH
Cache	—	No	Option	Yes	Yes
In- or out-of-sequence NACK	—	N/A	In-sequence	Out-of-sequence	Out-of-sequence
ACK or NACK	—	ACK	NACK	NACK	NACK
Energy conservation	Good	Fair	—	—	Yes

5.4.1 CODA (Congestion Detection and Avoidance)

The three components of CODA, an upstream congestion control method, are closed-loop end-to-end multisource regulation, open-loop hop-by-hop backpressure, and congestion detection. By keeping track of wireless channel load and current buffer occupancy, CODA makes an effort to identify congestion. When buffer occupancy or wireless channel load surpasses a certain level, congestion is assumed to have taken place.Using an open-loop hop-by-hop backpressure, the node that has detected congestion will then alert its upstream neighbor to lower its rate. The upstream neighbor nodes use techniques like AIM to reduce their output rate.Finally, CODA regulates a multisource rate through a closed-loop end-to-end approach, as follows: (1) When a sensor node exceeds its theoretical rate, it sets a ‘‘regulation’’ bit in the ‘‘event’’ packet; (2) If the event packet received by the sink has a ‘‘regulation’’ bit set, the sink sends an ACK message to the sensor nodes and informs them to reduce their rate; and (3) if the congestion is cleared, the sink will send an immediate ACK control message to the sensor nodes, informing them that they can increase their rate. CODA’s disadvantages are its unidirectional control, only from the sensors to the sink; there is no reliability consideration; and the response time of its closed-loop multisource control increases under heavy congestion since the ACK issued from the sink will probably be lost.

5.4.2 ESRT (Event-to-Sink Reliable Transport)

ESRT which provides reliability and congestion control, belongs to the upstream reliability guarantee group. It periodically computes a reliability figure (r), representing the rate of packets received successfully in a given time interval. ESRT then deduces the required sensor reporting frequency (f) from the reliability figure (r) using an expression such as f = G(r). Finally, ESRT informs all sensors of the values of (f) through an assumed channel with high power. ESRT uses an end-to-end approach to guarantee a desired reliabilityfigure through adjusting the sensors’ reporting frequency. It provides overall reliability for the application. The additional benefit of ESRT is energy conservation through control of reporting frequency. Disadvantages of ESRT are that it advertises the same reporting frequency to all sensors (since different nodes may have

contributed differently to congestion, applying different frequencies would be more appropriate) and considers mainly reliability and energy conservation as performance measures.

5.4.3 RMST (Reliable Multisegment Transport)

The upstream direction of packet transmission is guaranteed by RMST. Either intermediate nodes operate in noncache mode, where only end hosts cache the sent packets for end-to-end recovery, or they cache each packet to enable hop-by-hop recovery. Both cache and noncache modes are supported by RMST. Moreover, for loss detection and alerting, RMST employs timer-driven and selective NACK techniques. Lost packets are tracked down in the cache mode hop by hop using the intermediary sensor nodes. The NACK will be forwarded upstream towards the source node if an intermediate node is unable to find the lost packet or if it is operating in noncache mode. In order to assure application reliability, RMTS is created to run above the routing protocol directed diffusion.Problems with RMST are lack of congestion control, energy efficiency, and application-level reliability.

5.4.4 PSFQ (Pump Slowly, Fetch Quickly)

By pacing data at a reasonably moderate rate and allowing sensor nodes that suffer from data loss to recover any missing segments from close neighbors, PSFQ distributes data from sink to sensors. This strategy is a part of the downstream reliability guarantee category. The goal is to localize data recovery among close neighbors in order to minimize loss recovery and achieve loose delay bounds. Pump, fetch, and report are the three processes that make up PSFQ. PSFQ functions as follows: Until all of the data fragments have been sent, Sink broadcasts a packet to its neighbors every T time units. The sensor node enters fetch mode when a sequence number gap is found and sends a NACK in the reverse path to retrieve the lost fragment.Unless the number of times the NACK is sent exceeds a set limit, the neighbor nodes do not relay the NACK. Lastly, using a straightforward and scalable hop-by-hop report method, the sink can request information from sensors regarding the status of data delivery. The following drawbacks of PSFQ: Its slow pump causes a significant delay, it cannot detect packet loss for single packet transmission, and its hop-by-hop recovery with cache requires bigger buffer sizes.

5.4.5 GARUDA

The downstream reliability category includes GARUDA. It is built on a two-tier node design, and core sensor nodes are chosen from nodes that are 3i hops away from the sink (i is an integer). Second-tier nodes are the noncore nodes that are still present. A nearby core node is selected by each noncore sensor node to serve as its core node. Core nodes are used by noncore nodes to recover lost packets. GARUDA detects and notifies losses via a NACK message. Loss recovery is done in two different ways: between core sensor nodes and between noncore sensor nodes and their core node.Retransmission to recover lost packets appears to be a combination of pure hop by hop and end to end, for this reason. In order to

ensure the success of single or first packet delivery, GARUDA designs a repeating wait for first packet (WFP) pulse transmission. In order to create a two-tier node architecture, the hop number and core sensor nodes are also computed and chosen via pulse transmission. Inconsistency in the upstream direction and a lack of congestion control are two drawbacks of GARUDA. At the time of this writing, GARUDA's published results were devoid of any reliability findings or performance evaluations against alternative algorithms like PSFQ.

5.4.6 ATP (Ad Hoc Transport Protocol)

A receiver and network-assisted end-to-end feedback control algorithm underlies ATP's operation. It makes use of selective ACKs (SACKs) to recover from packet loss. The sum of exponentially distributed packet queuing and transmission delay, or D, is computed by intermediate network nodes in ATP. The inverse of D is chosen as the needed end-to-end rate. The values of D are calculated over all packets that pass through a certain sensor node, and if they are greater than the value piggybacked in each outgoing packet, the field is updated before the packet is forwarded.Inverse of D is calculated by the receiver and fed back to the sender to determine the necessary end-to-end rate. As a result, the transmitter is able to intelligently modify its transmission rate based on the value obtained from the receiver. Selective ACKs (SACKs) are used by ATP as an end-to-end technique for loss detection to ensure reliability. Because ATP separates congestion control from reliability, it outperforms TCP in terms of fairness and throughput. The question of whether ATP is best for an end-to-end control method is raised by the fact that energy concerns are not taken into account for this design.

5.5 PERFORMANCE OF TRANSPORT CONTROL PROTOCOLS

This section compares the performance of WSN congestion and loss quantitatively. Energy consumption, which is calculated for end-to-end and hop-by-hop situations, is the parameter used to compare congestion. Loss performance is a different metric that is dependent on cache and noncache methods.

5.5.1 Congestion

End to end and hop by hop are two common methods for reducing congestion. The source node must identify congestion in either the receiver-assisted (ACK-based loss detection) mode or the network-assisted mode in an end-to-end protocol like standard TCP (using explicit congestion notification). Rate modifications therefore only take place at the source node. In hop-by-hop congestion control, intermediate nodes alert the originating connection node when there is congestion. Hop-by-hop control may be able to clear congestion more quickly than the end-to-end method while also lowering packet loss and energy usage in sensor nodes.

Here, a straightforward model is presented to assist in understanding how congestion control affects energy efficiency. We make the following presumptions:

- Between sources and sink nodes, there are h > 1 hops, and each hop causes a delay d. C is the link capacity.
- The network experiences uniform congestion. The frequency of congestion is f, and it depends on the buffer size, traffic patterns, and network topology.
- Congestion will be noticed when the total rate of source transmission reaches C(1+a).
- The average amount of energy needed to send or receive a packet over each link is e.

With the end-to-end strategy, 1.5hd is often needed to alert the source of the beginning of congestion. All nodes can send up to C(1+a)(1.5hd) packets during this window (between the time that congestion occurs and the source is alerted), with the exception of the congested link, when traffic is limited to C(1.5hd). As a result, n_e=a.C (1.5hd) can be used to estimate the number of packets lost in this situation as a result of congestion.

The time needed to start congestion control corresponds to merely a single hop's worth of delay (d) in the hop-by-hop method. As a result, before congestion is reduced, packet loss is roughly equal to n_b = aCd.

Let Nd(T) represent the number of packets dropped owing to congestion during the time interval T, and let Ns(T) represent the number of packets successfully transmitted via the congested network. Each dropped packet has made 0.5H hops on average. The energy effectiveness of a congestion control device is defined as

$$E_c = \frac{N_s(T)He}{N_s(T)He + N_d(T)(0.5H)e} = \frac{N_s(T)}{N_s(T) + 0.5N_d(T)}$$

where Ec is the mean energy ratio required to send one packet successfully. In ideal situations, when there is no congestion, Ec would be 1. Therefore, for end-to-end congestion control,

$$E_c = \frac{N_s(T)}{N_s(T) + 0.5N_d(T)} = \frac{TC}{TC + 0.5f\,Tn_e} = \frac{4}{4 + 3fahd}$$

and for the hop-by-hop control,

$$E_c = \frac{N_s(T)}{N_s(T) + 0.5N_d(T)} = \frac{TC}{TC + 0.5f\,Tn_h} = \frac{2}{2 + fad}$$

The two equations above show that an end-to-end mechanism's energy efficiency depends on the path length (H), but hop-by-hop control is independent of the path length and yields a better efficiency ratio.

The ratio of all packets discarded in the sensor network to all packets received at the sink for hop-by-hop congestion control is the energy tax, according to CODA. The lower ratio therefore denotes more energy efficiency. Figure 1 shows CODA's energy efficiency.

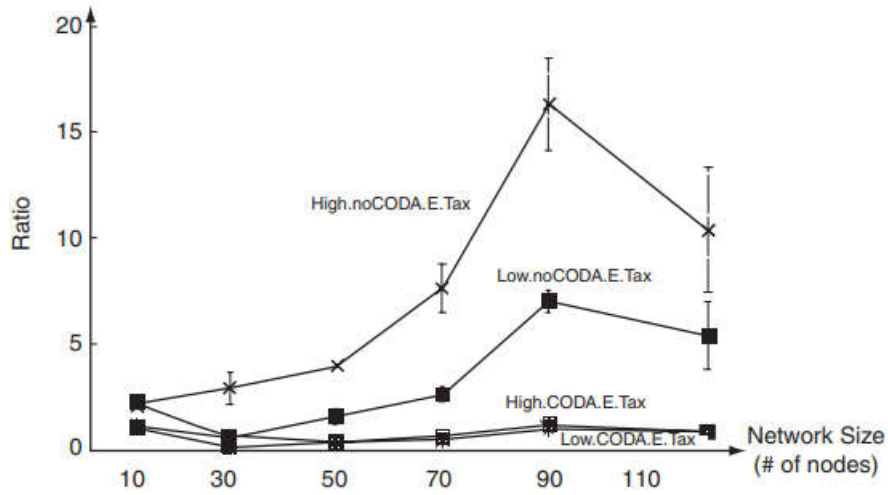


Figure 1: Energy tax in CODA as a function of network size for high- and low-data-rate traffic. The difference between the data points with and without CODA indicates the energy saving achieved by CODA

5.5.2 Packet loss recovery

How to recover lost packets is the issue we address in this section.Cache and noncache recovery are typically the two approaches that are available for this purpose. A similar end-to-end ARQ (automatic repeat request) to the conventional TCP is noncache recovery. Using a hop-by-hop methodology, cache-based recovery relies on retransmissions between two nearby nodes and caching at the intermediate nodes. Therefore, retransmissions may happen in h hops in the noncache situation, necessitating greater overall energy.The node that replicates transmitted packets locally for a predetermined amount of time is referred to as the cache point, while the node that packets are dropped due to congestion is referred to as the loss point. The length of the retransmission path, or l_p, will be defined as the quantity of hops between the caching node and the node where the loss occurs. $l_p = h1$, where h1 is the number of hops from the loss point to the source node, in the noncache case.If lost packets are located on nearby nodes in the cache example, l_p can be 1. Packet copies can only be stored for a finite amount of time because sensor nodes have a finite amount of buffer space. Because of this, l_p in the cache scenario may be greater than 1 but lower than h1($1<l_p<h1$). Different techniques in cache-based recovery may introduce varied energy efficiency and have different retransmission path lengths (l_p).

By using cache-based recovery, each packet is kept at each intermediate node it passes through until it is successfully received by its neighboring node or until a timeout occurs (whichever is sooner). In this instance, $l_p$is probably extremely near to 1. Distributing caching is a different technique that would distribute packet copies among intermediate nodes. Only one

or a few intermediate nodes store each packet.In addition to using less buffer space than traditional caching, distributed caching may have a longer l_p than traditional caching (but still be smaller than in the noncache situation).

The performance of several loss recovery strategies that might offer dependability via the connection, transport, and application levels were examined by RMST.The effectiveness of end-to-end loss recovery and hop-by-hop loss recovery in the transport layer is compared in Figure 2 from. The comparison is based on how many transmissions are necessary to send 10 packets over a network in 10 hops. This graph demonstrates how the number of end-to-end retransmissions doubles when the success rate falls below 0.95, which reduces energy efficiency.

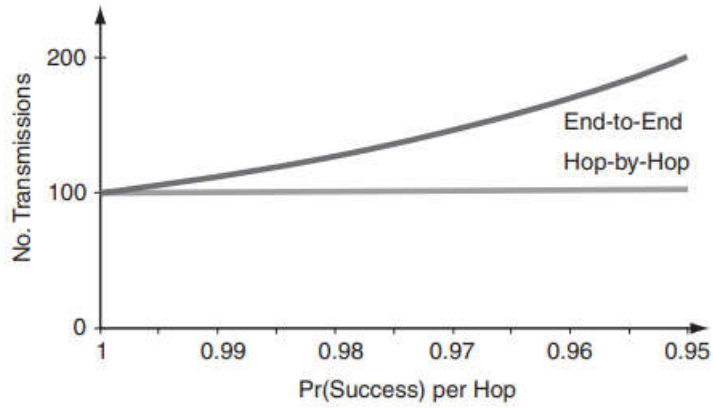


Figure 2: Hop-by-hop versus end-to-end: number of transmissions required to send 10 packets in 10 hops

5.6 SUMMARY

We gave a general overview of the wireless sensor network transport control protocol in this chapter. The drawbacks of the TCP and UDP protocols were examined, along with the reasons why they weren't appropriate for wireless sensor networks. Also, a study of various sensor transport control protocols that are now in use was given, along with a list of issues with those protocols. Designing transport control methods for wireless sensor networks requires careful consideration of the following points:

1. The effectiveness of protocols and the effectiveness of congestion-control techniques.Efficient techniques provide high throughput while minimizing packet loss.

2. The reliability of the transport layer, the need for loss recovery at the transport layer, and the most efficient and effective mechanism.Any such techniques should ideally have minimal buffering requirements.

3. Fairness between sensor nodes located at various ranges from the sink.

4. Using a cross-layer optimization technique to boost performance.

5.7 LIST OF REFERENCES

1) Protocols and Architectures for Wireless Sensor Network, Holger Kerl, Andreas Willig, John Wiley and Sons, 2005

2) Wireless Sensor Networks Technology, Protocols, and Applications,Kazem Sohraby, Daniel Minoli and TaiebZnati, John Wiley & Sons, 2007

3) Mobile communications, Jochen Schiller,2nd Edition, Addison wisely, Pearson Education,2012

4) Fundamentals of Wireless Sensor Networks, Theory and Practice, WaltenegusDargie, Christian Poellabauer, Wiley Series on wireless Communication and Mobile Computing, 2011

5) Networking Wireless Sensors, Bhaskar Krishnamachari, Cambridge University Press, 2005

5.8 UNIT END EXERCISES

1) Explain: i] TCP ii] UDP.

2) Describe Mobile IP.

3) Illustrate and explain Transport Protocol Design Issues.

4) Explain the examples of Existing Transport Control Protocols

5) Write a note on CODA (Congestion Detection and Avoidance).

6) Explain ESRT (Event-to-Sink Reliable Transport).

7) Discuss the RMST (Reliable Multi segment Transport).

8) What is PSFQ (Pump Slowly, Fetch Quickly)?

9) Explain GARUDA.

10) What is ATP (Ad Hoc Transport Protocol)?

11) Describe the performance of Transport Control Protocols.

12) What do you mean by Congestion?

13) Explain the concept of Packet loss recovery.

6

INTRODUCTION, WIRELESS TRANSMISSION AND MEDIUM ACCESS CONTROL

Unit Structure

6.0 OBJECTIVES

- To get familiar with wireless transmission and medium access control
- To get acquaint with the signaling and propagation involved and associated with the wireless transmission

6.1 INTRODUCTION

In ten years, what will computers look like? No one can forecast the future with absolute certainty, but most computers will undoubtedly be portable. How will consumers utilize computers or other communication tools to access networks? a growing number wirelessly, that is, without any wires. How will people spend the most of their time while on vacation at work? Numerous people will be mobile, which is currently one of the main features of contemporary civilization. Consider an aircraft with 800 seats, for instance. Passengers on modern aircraft currently have minimal

network connection, and future aircraft will provide simple Internet access.

In this case, the only method of transferring data to and from passengers will be a mobile network travelling at a high rate of speed above ground and connected by a wireless link. Consider vehicles with Internet connection and countless embedded processors that must interact with devices like cameras, cell phones, CD players, headsets, keyboards, intelligent traffic signs, and sensors. This wide range of tools and programmes demonstrates the current importance of mobile communications.

The definitions of the terms "mobile" and "wireless" as they are used should be given before showing more applications. User mobility and device portability are two different types of mobility. The term "user mobility" describes a user who has access to the same or equivalent telecommunication services at various locations; in other words, the user is mobile and the services follow them. Simple call-forwarding solutions from the telephone or computer desktops that enable roaming (i.e., have the same appearance no matter which computer a user logs into the network with) are examples of systems that support user mobility.

When a communication device is portable, it can be moved (with or without a user). To ensure that communication is still feasible when the device is moving, numerous procedures both inside the device and in the network must be in place. The mobile phone system is a common illustration of a system that supports device portability, as the system automatically switches the device from one radio transmitter (also known as a base station) to another if the signal deteriorates. Most of the scenarios include simultaneous user mobility and gadget portability.

The word "wireless" is applied to gadgets. This only explains how to connect to a network or other communication partners without using a wire. Transmission of electromagnetic waves through "the air" takes the role of the cable (although wireless transmission does not need any medium).

6.2 APPLICATIONS

Although wireless networks and mobile communications can be advantageous for many applications, some application settings appear to be tailor-made for their use. Some of them are included in the sections below:

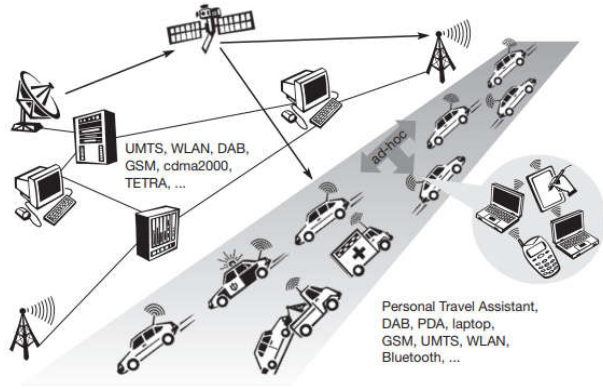
1] Vehicles

While some already exist in today's cars, there will be many more wireless communication systems and mobility-aware applications in cars of the future. Digital audio broadcasting (DAB) with 1.5 Mbit/s allows for the reception of music, news, traffic updates, weather forecasts, and other broadcast information. A universal mobile telecommunications system (UMTS) phone with 384 kbit/s voice and data connectivity might be

available for personal communication. Satellite communication can be employed in remote places, and the global positioning system is used to identify the car's current location. (GPS). To transmit information quickly in an emergency or to preserve a safe distance from one another, nearby cars create a small ad hoc network.In the event of an accident, not only will the airbag deploy, but a provider will also receive an emergency call alerting the police and ambulance service. This technology is already in some cars. In the future, vehicles will communicate with one another via an ad-hoc network in order to alert them about accidents and help them slowdown in time, even before a driver is aware of one. Already, trains, trucks, and buses send maintenance and logistical data to their base of operations, improving fleet management and saving time and money.

A typical setup for mobile communications including numerous wireless devices is shown in Figure 1. Mobile phone networks (GSM, UMTS) and trunked radio systems (TETRA) will connect to networks with a fixed infrastructure to form wireless LANs. (WLAN). Additionally, satellite communication lines may be utilized. It's more likely that the networks inside each automobile and those between cars would operate haphazardly.Personal digital assistants (PDA), computers, and mobile phones, such as those connected via Bluetooth, can all be a part of wireless pico networks within a car.

Figure 1.
A typical application of mobile communications: road traffic



Consider instances where there is train or air travel. Here, speed can lead to a variety of issues. While trains and contemporary aeroplanes may move at speeds of up to 900 km/h and 350 km/h, respectively, many technologies cannot function if a mobile device's relative speed is greater than, for example, 250 km/h for GSM or 100 km/h for AMPS. Only a few technologies, such as DAB, are capable of speeds of up to 900 km/h(unidirectional only).

2] Emergencies

Just consider the advantages of an ambulance having a reliable wifi connection to a medical facility. From the accident scene, critical information regarding injured people can be conveyed to the hospital. For this specific accident type, the essential preparations can be made, and professionals can be consulted for an early diagnosis. In the event of a natural disaster, such as a hurricane or earthquake, wireless networks are the only available form of communication. Only wireless, decentralized ad

hoc networks survive in the worst scenarios. In addition to the regular cable telephone system failing, all mobile phone systems needing base stations also fail if all cabling fails.

3] Business

Today's travelling salesperson requires immediate access to the company's database to make sure that the files on his or her laptop represent the current situation, to allow the business to monitor all of its travelling employees' activities, to maintain consistent databases, etc. The laptop can become a truly mobile office with wireless connection, but effective and strong synchronization techniques are required to guarantee data consistency. Figure 2 depicts what could occur when staff members attempt to communicate inappropriately.The laptop at home connects to the Internet using DSL and a WLAN or LAN. When the WLAN coverage runs out, it is necessary to switch to a different technology, like an improved version of GSM, before leaving the house.Data rates decrease when travelling at a higher pace due to interference and other causes. In addition to gas, some gas stations provide WLAN hot spots.There is already wifi connectivity available aboard trains. Before getting to the office, it could be required to switch to a few more different technologies. Mobile communications should always provide the best access to the internet, the company's intranet, or the telephone network, regardless of the time and location.

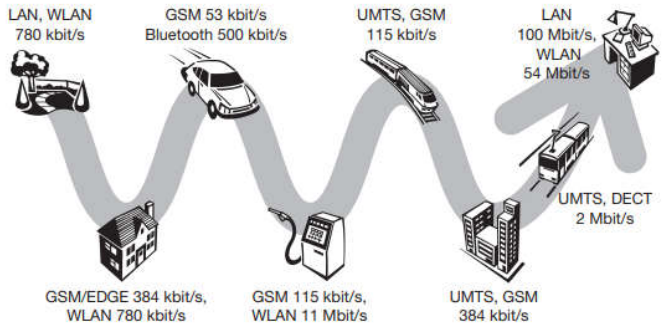


Figure .2 Mobile and wireless services – always best connected

4] Replacement of wired networks

In some circumstances, such as with remote sensors, at trade exhibitions, or in old buildings, wireless networks can also be used in place of wired networks. It is frequently impractical to link remote sensors for weather forecasts, earthquake detection, or to give environmental data due to financial considerations. In this case, wireless connections, like those provided by satellite, can be useful. Tradeshows require a highly flexible infrastructure, but installing cable takes a lot of time and often proves to be excessively rigid. WLANs frequently take the place of cabling at computer shows.Computers, sensors, or information displays in historic structures are additional uses for wireless networks since excessive cabling could damage priceless walls or flooring. The use of wireless access points in a room corner may be the answer.

5] Infotainment and more

Internet everywhere? Not if wireless networks are absent! Think about a city's travel guide. Static data can be downloaded from a CD-ROM, DVD, or even the Internet at home. However, wireless networks can deliver current information at any suitable location. By determining your location via GPS, a local base station, or triangulation, the tour guide may provide you with information on a building's past while simultaneously downloading details about a concert that will be taking place there that same night over a local wireless network. You can select a seat, pay using electronic money, and email these details to a service provider. In order to enable, for instance, ad-hoc gaming networks as soon as players meet to play together, entertainment and games are a rising area of wireless network applications.

6] Location dependent services

Many research projects in mobile computing and wireless networks attempt to conceal the fact that a wireless link is more error prone than a connected one or that network access has changed (for example, from mobile phone to WLAN or between various access points). Many protocols attempt to improve link quality via encoding methods or retransmission to ensure that applications intended for fixed networks continue to function. Mobile IP seeks to mask the fact that shifting access points by diverting packets while preserving the same IP address.However, it is frequently necessary for an application to "know" anything about the location or the user may require location data for additional activities. Different services that may depend on the actual location can be identified include follow-up services, location aware services, privacy services, information services, and support services.

7] Mobile and wireless devices

Even if there are already a lot of wireless and mobile gadgets on the market, there will be a lot more in the future. Such gadgets are not specifically categorized in terms of size, shape, weight, or computer capability. Currently, laptops are regarded as being at the top of the mobile device spectrum.Future mobile and wireless gadgets will be more potent, lighter, and equipped with brand-new user and network interfaces. The energy source is one significant issue that has not yet been resolved. A device requires more power the more features are integrated into it. Assuming the same technology, the device's battery life decreases with increasing performance. Furthermore, energy is used up quickly during wireless data transfer.

6.3 A SHORT HISTORY OF WIRELESS COMMUNICATION

The history of Wireless Communications started with the understanding or magnetic and electric properties observed during the early days by the Chinese, Greek and Roman cultures and experiments carried out in the

17th and 18th centuries. Here are some selected events in the development of Wireless Communications.

1807 – French mathematician Jean Baptiste Joseph Fourier discovered Fourier's theorem

1820 – Danish physicist Hans Christian Orsted discovered the electromagnetic field caused by electric current. The French physicist Dominique Francois Jean Arago showed that a wire became a magnet when current flowed through it. French mathematician and physicist Andre-Marie Ampere discovered electrodynamics and proposed an Electromagnetic Telegraph.

1831 – British scientist Michael Faraday discovered electromagnetic induction and predicted existence of electromagnetic waves.

1834 – American inventor Samuel Finley Breese Morse invented the code for telegraphy named after him.

1847 – German physiologist and physicist Hermann Ludwig Ferdinand von Helmholtz suggested electrical oscillation

1853 – William Thomson (Lord Kelvin) calculated the period, damping and intensity as a function of the capacity, self-inductance and resistance of an oscillatory circuit.

1857 – Feddersen verified experimentally the resonant frequency of a tuned circuit as suggested by Helmholtz in 1847.

1864 – Scottish mathematician and physicist James Clerk Maxwell formulated the electromagnetic theory of light and developed the general equations of the electromagnetic field. He formulated 20 equations that were later simplified into the 4 basic equations we use today.

1866 - American dentist Dr. Mahlon Loomis described and demonstrated a wireless transmission system which he patented in 1866. Loomis demonstrated the transmission of signals between two mountains, a distance of 22 km.

1882 – American physicist, Amos Emerson Dolbear, was granted a patent for a wireless transmission system using an induction coil, microphone and telephone receiver and battery. Nathan Stubblefield transmitted audio signals without wires.

1883 – Irish physicist and chemist George Francis FitzGerald published a formula for the power radiated by a small loop antenna.

1884 – German physicist Heinrich Rudolf Hertz wrote Maxwell's equations in scalar form by discarding the concept of aether reducing it from 20 to 12 equations.

1885 – Thomas Edison patented a system of wireless communication by electrostatic induction.

1886 – Heaviside introduced impedance as the ratio of voltage over current. Hertz started his work to demonstrate the existence of radio waves and published his results in 1888.

1887 – English physicist Oliver Joseph Lodge discovered Sympathetic Resonance (standing waves) in wires.

1888 – Hertz produced, transmitted, and received electromagnetic waves (5 m to 50 cm) using reflectors to concentrate the beam. Hertz also discovered the principle for Radar. Heaviside wrote Maxwell's equations in vector form – the four equations we use today. Italian Galileo Farrari and Croatian-American Nilola Tesla independently produced rotating fields using 2-phase currents. Austrian engineer Ernst Lecher established the relation between frequency, wire length, velocity of propagation and the electrical constants of the wire.

1890 – Lecher used standing waves produced in parallel wires to measure frequency. Tesla introduced high frequency currents in therapeutics as he observed that current of high frequency could raise the temperature of living tissue. Tesla also patented his Tesla Coil which was used later in every spark gap generator to produce high frequency signals. Heinrich Rubens and R. Titter made a sensitive bolometer which measured the intensity of electromagnetic waves by means of the heat generated in a thin wire.

1893 – English physicist Joseph John Thomson published the first theoretical analysis of electric oscillations within a conducting cylindrical cavity of finite length suggesting the possibility of wave propagation in hollow pipes (waveguides). Hertz conducted experiments of EM shielding and for coaxial configuration.

1895 – Marconi transmitted and received a coded message at a distance of 1.75 miles near his home in Bologna, Italy. Indian physicist, Sir Jagadis Chunder Bose generated and detected wireless signals and produced many devices such as waveguides, horn antennas, microwave reflectors and more.

1897 – Marconi demonstrated a radio transmission to a tugboat over an 18 mile path over the English Channel. The first wireless company, Wireless Telegraph and Signal Company was founded – they bought most of Marconi's patents. Lord Rayleigh suggests EM wave propagation in waveguides and analysis of propagation through dielectrically filled waveguides. Lodge patented various types of antennas.

1899 – Marconi sent the first international wireless message from Dover, England to Wimereux, France.

1900 – Tesla obtained patents on System of Transmission of Electrical Energy which the US recognized as the first patents on Radio. Tesla is the first person to describe a system of determining the location of an object using radio waves – Radar.

1902 – Fessenden patented the Heterodyne receiver. American Cornelius D. Ehret filed patents covering the transmission and reception of coded signals or speech (Frequency Modulation – FM). Poulsen was the first to develop the CW transmitter.

1903 – Marconi established a transmission station in South Wellfleet, MA – the dedication included exchanges of greetings between American President Theodore Roosevelt and British King Edward VII. G.

1904 – Frank J. Sprague developed the idea of the printed circuit. W. Pickard filed a patent application for a crystal detector where a thin wire was in contact with silicon. It was the central component in early radio receivers called crystal radios. J. C. Bose was granted a patent on point contact diodes that were used for many years as detectors in the industry. Fleming suggested the rectifying action of the vacuum-tube diode for detecting high frequency oscillation – the first practical radio tube.

1905 – Fessenden invented the superheterodyne circuit.

1906 – Lee de Forest patented the general principle of omni-range using a rotating radio beam keyed to identify the sector forming 360 degree sweep of the beam. He also invented the three-electrode valve or vacuum tube triode that was instrumental in the development of transcontinental telephony in 1913. Poulsen transmitted music by wireless using an arc transmitter with 1 kW of input power and a 200 feet high antenna that was heard 300 miles away.

1909 – Marconi and Braun shared the Nobel Prize for Physics for their contributions to the physics of electric oscillations and radiotelegraphy.

1911 – Von Lieben and Eugen Riesz developed a cascade amplifier. Hugo Germsback, an American novelist, envisaged the concept of pulse radar in one of his works where he proposed the use of a pulsating polarized wave, the reflection of which was detected by an actinoscope.

1911 – Engineers start to realize that the triode can also be used for transmitter and oscillator – the three-electrode vacuum tube was included in designs for telephone repeaters in several countries.

1912 – G. A. Campbell developed guided wave filters. Sinding and Larsen transmitted TV by wireless using 3 channels. The Institute of Radio Engineers was formed in the US.

1914 – The German physicist Walter Schottky discovered the effect of electric field on the rate of electron emission from thermionic-emitters named after him. Fleming discovered the atmospheric refraction and its importance in the transmission of EM waves around the Earth. Carl R. Englund was the first to develop the equation of a modulated wave (AM) and also discovered the frequencies related to sidebands. Frequency modulation of a carrier was proposed to accommodate more channels within the available bandwidths.

1915 – Schottky stated work on the space-charge-grid tube and a screen grid tube or Tetrode that achieved good amplification by placing a screen grid between the grid and the anode.

1916 – Leon m Brillouin and Georges A. Beauvais patented the R-C coupled amplifier. F. Adcock used open vertically spaced aerials for direction finding in aircraft and granted British patent.

1918 – Armstrong invented the Superheterodyne Radio Receiver using 8 valves – most receivers still use this design today. Langmuir patented the feedback amplifier. E. H. O Shaughnessy development of direction finding was one of the key weapons in England during WWI – Bellini-Tosi aerials were installed around the coast to locate transmission from ships and aircrafts. Louis Alan Hazeltime invented the neutrodyne circuit with tuned RF amplifier with neutralization.

1919 – Marconi-Osram company developed the U-5 twin-anode full-wave rectifier. Joseph Slepian filed a patent application for a vacuum tube electron multiplier. Sir Robert Alexander Watson-Watt patented a device for radiolocation by means of short-wave radio waves – the forerunner of the Radar system.

1921 - E. S. Purington made the all-electric frequency modulator. A.W. Hull invented the Magnetron oscillator operating at 30 kHz and output power of 8 kW and 69 percent efficiency. E. H. Colpitt and O. B. Blackwell developed modulation of an audio frequency carrier by signals of lower audio frequency for carrying telephony over wires. S. Butterworth published a classic paper on HF resistance of single coil considering skin and proximity effect.

1922 – Walter Guiton Cady invented the piezoelectric (Quartz) crystal oscillator. The BBC broadcasts is first news program.

1923 – The decibel ($1/10^{th}$ of a bel, after A. G. Bell, inventor of the telephone) was used to express the loss in a telephone cable. H. W. Nichols developed point-to-point communication using single side-band communication. D.C Prince analyzed Class A and Class C amplifiers. Scottish engineer Antoine Logie Barid built and patented the first practical TV. Watson-Watt perfected the radiolocation device by displaying radio information on a cathode ray oscilloscope telling the radar operator the direction, distance and velocity of the target. Ralph Vinton Lyon Hartley showed that the amount of information that can be transmitted at a given time is proportional to the bandwidth of the communication channel. H. Flurschein filed a patent on radio warning system for use on vehicles.

1924 – J.R. Carson showed that energy absorbed by a receiver is directly proportional to its bandwidth and extended Lorentz's reciprocity theory to EM fields to antenna terminals. Lloyd Espenschied invented the first radio altimeter. The mobile telephone was invented by Bell Telephone Company and introduced to NYC police cars.

1925 – First conference on frequency allocation was held in Geneva. Joseph Tykocinski-Tykociner demonstrated that the characteristics of a full size antenna can be replaced with sufficient accuracy from measurements made on a small short wave in the rage of 3 to 6 m.

1926 – L.E. Lilienfield patented the theory of the Field-Effect Transistor. Japanese engineers Hidetsugu Yagi and Shintaro Uda developed the Yagi antenna, a row of aerials consisting of one active antenna and twenty undriven members as a wave canal. Hulsenback and Company patented identification of buried objects using CW radar.

1927 – R. V. Hartley developed the mathematical theory of communications. Harold Stephen Black of Bell Laboratories conceived the negative feedback amplifier. A. de Hass studied fading and independently developed diversity reception system.

1928 – Baird conducted the first transatlantic TV broadcast and built the first color TV. Nyquist published a classic paper on the theory of signal transmission in telegraphy. He developed the criteria for the correct reception of telegraph signals transmitted over dispersive channels in the absence of noise. C.S. Franklin patented the coaxial cable in England to be used as an antenna feeder.

1929 – L. Cohen proposed circuit tuning by wave resonance (resonant transmission line) and its application to radio reception. H.A. Affel and L. Espenscheid of AT&T/Bell Labs created the concept of coaxial cable for a FDMA multi-channel telephony system. K. Okabe made a breakthrough in cm-waves when operating his slotted-anode magnetron (5.35 GHz). Hans Erich Hollmann patented the idea of a reflex klystron with his double-grid retarding-field tube. W.H. Martin proposed the Decibel as a transmission unit.

1931 – H. diamond and F. W. Dunmore conceived a radio beacon and receiving system for blind landing of aircraft. H. E. Hollmann built and operated the first decimeter transmitter and receiver at the Heinrich Hertz Institute. He called the device the magnetron.

1932 – The word Telecommunication was coined and the International Telecommunications Union (ITU) was formed. George C. Southworth and J. F. Hargreaves developed the circular waveguide. Karl Jansky accidentally discovered radio noise coming from outer space giving birth to radio astronomy. R. Darbord developed the UHF Antenna with parabolic reflector.

1933 – Armstrong demonstrated Frequency Modulation (FM) and proposed FM radio in 1936. C.E. Cleeton and N. H. Williams made a 30 GHZ CW oscillator using a split-anode magnetron.

1934 – The Federal Communications Commission (FTC) was created in the US. W.L. Everitt obtained the optimum operating conditions for Class C amplifiers. F. E. Terman demonstrated a transmission line as a resonant circuit. German physicist Oskar Ernst Heil applied for a patent on

technology relating electrical amplifiers and other control arrangements that was the theoretical invention of capacitive current control in FETs.

1935 – C. J. Frank of Boonton Radio Corp demonstrated Q-meter at the fall meeting of IRE – the ratio of reactance to resistance of a coil as its "Quality Factor" was first suggested about 1926. A French TV transmitter was installed on top of the Eiffel Tower. Watson-Watt developed and patented the first practical radar for use in the detection of airplanes in England. H. E. Hollmann filed a patent for the multi-cavity magnetron (granted in 1938).

1936 – H. W. Doherty developed a new high efficiency power amplifier for modulated waves, Doherty amplifier, at Bell Labs. English engineer Paul Eisler devised the Printed Circuit. N. H. Jack patented the semi-rigid coaxial cable using thin soft copper tube as the outer conductor. Harold Wheeler used two flat copper strips side by side to make a low loss transmission line that could be rolled to save space. H. T. Friis and A. C. Beck invented the horn reflector antenna with dual polarization.

1937 – Grote Rober constructed the first radio telescope. W. R. Blair patented the first anti-aircraft fire control radar. Russell H. Varian and his brother Sigurd Varian along with William Hansen developed the reflex Klystron. Alex H. Reeves invented pulse-code modulation for digital encoding of speech signals.

1938 – E. L. Chaffee determined the optimum load for Class B amplifiers. IRE published standards on transmitters, receivers and antennas. Claude Elwood Shannon recognized the parallels between Boolean algebra and the functioning of electrical switching systems. W. R. Hewlett developed the Wien-bridge (RC) oscillator. P. H Smith at RCA developed the well known Smith Chart. N. E. Lindenblad of RCA developed a coaxial horn antenna. John Turton Randall and Albert Boot developed the cavity magnetron that becomes the central components to radar systems.

1941 – W. C. Godwin developed the direct-coupled push-pull amplifier with inverse feedback. Siemens & Halske made the Ge diode – R. S. Ohl made the Si junction diode. Sidney Warner realized a two-way police FM radio.

1943 – H. J. Finden developed the frequency synthesizer. Austrian engineer Rudolf Kompfner developed the traveling wave tube. C. K. Chang developed frequency modulation of RC oscillators. C. F. Edwards developed microwave mixers. H. T. Friis developed noise figures of radio receivers.

1944 – Harold Goldberg suggested pulse frequency position modulation. E. C Quackenbush of Amphenol developed the VHF coaxial connectors. Paul Neil of Bell Labs developed Type N connectors. Maurice Deloraine, P. R. Adams and D. H. Ranson applied for patents covering switching by pulse displacement a principle later defined as time-slot interchange – Thus, Time-Division Multiplexing (TDMA) was invented. Radio

Research Lab developed radar countermeasures (jamming) in the 25 MHz to 6 GHz range.

1946 – S. L. Ackerman and G. Rappaport developed a radio control systems for guided missiles. E. M. Williams developed the radio frequency spectrum analyzer.

1947 – G. E. Mueller and W. A. Tyrrel developed the dielectric rod antenna. John D. Kraus invented the helical antenna. W. Tyrell proposed hybrid circuits for microwaves, H. E. Kallaman constructed the VSWR indictor meter.

1948 – W. H. Branttain, J. Bardeen and W. Shockley of Bell Labs built the junction transistor. E. L. Ginzton and others developed distributed wideband amplifier using pentodes in parallel. Shannon laid out the theoretical foundations of digital communications in a paper entitled "A Mathematical Theory of Communication." Paine described the BALUN.

1949 – E. J. Barlow published the principle of operation of Doppler Radar.

1950- J. M. Janssen developed the sampling oscilloscope.

1951- Charles Hard Townes published the principle of the MASER (Microwave Amplification by Stimulated Emission of Radiation). The Laboratoire Central des Telecommunications in Paris developed the first model of a time-division multiplex system connecting subscriber line by electronic gates handling amplitude modulated pulses.

1952 – C. L. Hogan demonstrated a microwave circulator.

1955 – R. H. DuHamel and D. E. IsBelll develop the log periodic antenna. John R. Pierce proposed using satellites for communications. Sony marketed the first transistor radio.

1957 – Soviet Union launched Sputnik I that transmitted telemetry signals for about 5 months. German physicist Herbert Kroemer originated the concept of the heterostructure bipolar transistor (HBT).

1958 – Robert Noyce (Intel) and Jack Kilby (TI) produced the first Si integrated circuit (IC).

1962 – G. Robert-Pierre Marie patented the wide band slot antenna. S. R. Hofstein and F. P. Heiman developed MOS IC.

1963 – W. S. Mortley and J. H. Rowen developed surface acoustic wave (SAW) devices. John B. Gunn of IBM demonstrated microwave oscillations in GaAs and InP diodes. The Institute of Electrical and Electronic Engineers (IEEE) was formed by merging the IRE and AIEE.

1964 – R. L. Johnson, B. C. De Loach and B. G. Cohen developed the IMPATT diode oscillator. COMSAT and INTELSAT started launching a series of communications satellites that were the building blocks in the global network of international communications satellites.

1969 – The first digital radio-relay system went into operation in Japan using 2 GHz operating frequency. ARPANET was launched (precursor to Internet).

1971 – Statek began manufacturing and marketing quartz oscillators that were made using their patented photolithographic process.

1978 – AT&T Bell Labs started testing a mobile telephone system based on cells.

1980 – CW performance of GaAs MESFET reached 10 W at 10 GHz. ATLAS I EM pulse simulator was built for testing large aircraft – it was the largest wooden structure in the world (400 x 105 x 75 m).

1989 – F. Laleari invented the broadband notch antenna

1990 – WWW was developed

6.4 WIRELESS TRANSMISSION: FREQUENCY FOR RADIO TRANSMISSION

Numerous frequency bands are available for radio transmission. There are pros and downsides to each frequency band. The frequency range that can be employed for data transmission is depicted in rough detail in Figure 3.In the graph, frequencies from 300 Hz to over 300 THz are depicted. The wavelength λ is directly connected to the frequency by the following equation:

$\lambda = c/f$

where $c \cong 3 \cdot 108$ m/s (the speed of light in vacuum) and f the frequency. For traditional wired networks, frequencies of up to several hundred kHz are used for distances up to some km with twisted pair copper wires, while frequencies of several hundred MHz are used with coaxial cable (new coding schemes work with several hundred MHz even with twisted pair copper wires over distances of some 100 m). Fiber optics are used for frequency ranges of several hundred THz, but here one typically refers to the wavelength which is, e.g., 1500 nm, 1350 nm etc. (infra-red).

At several kHz, or the very low frequency (VLF) band, radio transmission begins. These waves are incredibly long. Submarines use low-frequency (LF) waves because they can travel through water and track the surface of the earth. These frequencies are still used by some radio stations, for instance in Germany between 148.5 kHz and 283.5 kHz. The transmission of hundreds of radio stations often takes place in the medium frequency (MF) and high frequency (HF) bands using either amplitude modulation (AM) between 520 kHz and 1605.5 kHz, short wave (SW) between 5.9 MHz and 26.1 MHz, or frequency modulation (FM) between 87.5 MHz and 108 MHz.The frequencies setting the boundaries of these bands are normally established by national law and vary from one nation to another. Because of ionosphere reflection, short waves are frequently employed for (amateur) radio transmission throughout the world. Up to 500 kW of

transmit power is available, which is significantly more than the 1 W of a cell phone.

The TV stations follow when we go to higher frequencies. The 174-230 MHz and 470-790 MHz very high frequency (VHF) and ultra-high frequency (UHF) bands are used to transmit traditional analogue TV. This frequency range is also used for digital audio broadcasting (DAB) (223–230 MHz and 1452–1472 MHz), as well as for planned or installed digital TV (470–862 MHz), which reuses parts of the previous analogue TV channels. In addition, UHF is utilized for analogue mobile phones (450-465 MHz), digital GSM (890-960 MHz, 1710-1880 MHz), digital cordless phones (1880-1900 MHz), 3G cellular networks (1900-1980 MHz, 2020-2025 MHz, 2110-2190 MHz), and many other applications.

Super high frequencies (SHF) are primarily utilized for fixed satellite services in the C-band (4 and 6 GHz), Ku-band (11 and 14 GHz), or Ka-band (between 2 and 40 GHz). (19 and 29 GHz). The extremely high frequency (EHF) band, which is near to infrared, is where certain devices are planned. To prevent interference, all radio frequencies are regulated; for example, German law regulates frequencies between 9 kHz and 275 GHz.

Optical transmission, which is utilized for both wireless communications and fibre optical networks, is the next step into higher frequencies.For directed links, such as using laser links to connect several buildings, infrared (IR) transmission is employed. IrDA, the most used IR technology, connects laptops, PDAs, and other devices using wavelengths between 850 and 900 nm. Finally, wireless transmission has been possible for thousands of years using visible light. Even while interference makes light less reliable, it is still helpful since it has human receivers built in.

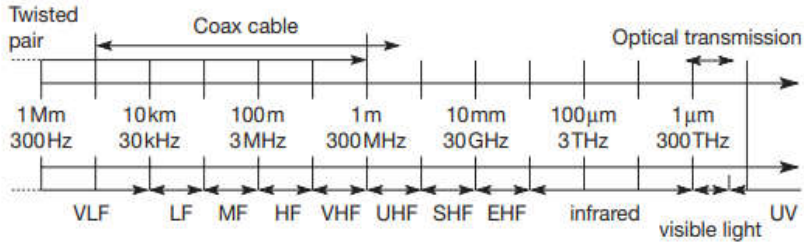


Figure 3: Frequency spectrum

6.5 SIGNALS

Data is physically represented by signals. The only way for users of a communication system to exchange data is through the signaling of signals. Data, or bits, are converted into signals and back again via Layer 1 of the ISO/OSI basic reference model. Signals are products of space and time. The data values are represented by signal parameters. Periodic signals, particularly those using sine waves as carriers, are the most intriguing sorts of signals for radio transmission. A sine wave's function as a whole is:

$$g(t) = A_t \sin(2\pi f_t t + \varphi_t)$$

The signal's amplitude (A), frequency (f), and phase shift (φ) are its parameters.As a result, A_t, the amplitude as a component of the function g may likewise fluctuate with time. The periodicity of the signal is expressed by the frequency f, where T = 1/f is the period. (In equations ω, is often substituted for 2f.) Additionally, the frequency f may alter with time, so f_t. The signal's shift in relation to the same signal without a shift is finally determined by the phase shift. Figure 4 illustrates an illustration of shifting a function. This compares a sine function with and without a phase shift φ, with the same amplitude and frequency, respectively.

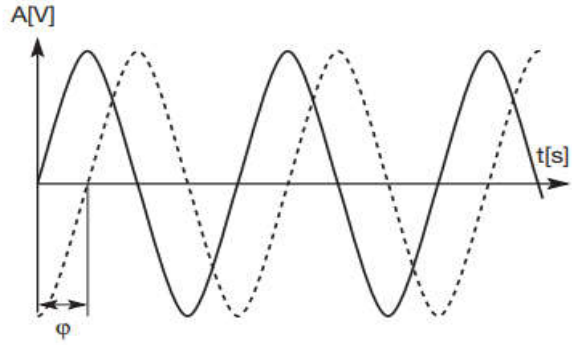


Figure 4: Time domain representation of a signal

Sine waves are of special interest, as it is possible to construct every periodic signal g by using only sine and cosine functions according to a fundamental equation of Fourier:

$$g(t) = \frac{1}{2}c + \sum_{n=1}^{\infty} a_n \sin(2\pi nft) + \sum_{n=1}^{\infty} b_n \cos(2\pi nft)$$

The coefficients a and bn in this equation represent the amplitudes of the nth sine and cosine functions, while the parameter c specifies the Direct Current (DC) component of the signal. The equation demonstrates that arbitrary periodic functions can only be built using an endless number of sine and cosine functions.However, the frequencies of these functions (the so-called harmonics) are a multiple of the fundamental frequency f and rise with a growing parameter n.Any media has a finite bandwidth, including the air, cable, transmitters, etc. There is also a frequency upper bound.Since all practical transmitting systems have bandwidth limitations and can never transmit arbitrary periodic functions, it is sufficient to take into account a small number of sine and cosine functions to generate periodic functions. We only need to be aware that transmitted signals can be conceptualized as being made up of one or more sine functions.The examples that follow always show the situation where there is just one frequency, or one sine function.

The time domain is a common representation format for signals (Figure 4). Here, a signal's amplitude A is plotted against time (time is often

expressed in seconds; amplitudes can, for example, be expressed in volts). This is also the common oscilloscope representation that is well-known. This illustration can also demonstrate a phase change.

If a signal contains a large number of distinct frequencies, representations in the time domain can be troublesome (as the Fourier equation indicates). In this instance, the frequency domain provides a more accurate representation of a signal(Figure 5). Here, the signal's amplitude at a certain frequency is plotted against frequency. Figure 5 only displays one peak, and the signal is a single sine function with only one frequency component. Arbitrary periodic functions would have a wide range of peaks, or the signal's frequency spectrum. A spectrum analyzer is a device for displaying frequencies.Using the inverse Fourier transformation, Fourier transformations are a mathematical tool for converting from the time domain to the frequency domain.

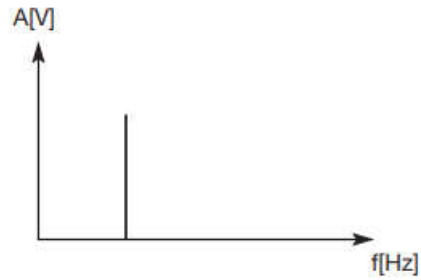


Figure 5: Frequency domain representation of a signal

Figure 6's representation of the phase domain offers a third way to represent signals. The amplitude M and phase of a signal are shown in polar coordinates in this diagram, which is also known as a phase state or signal constellation diagram. (The vector's length denotes amplitude and angle, phase shift.) The x-axis, which is also known as In-Phase (I), shows a phase of 0. Quadrature (Q) would be a point on the y-axis with a phase shift of 90° or /2.

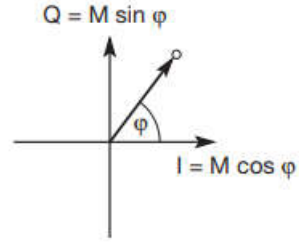


Figure 6: Phase domain representation of a signal

6.6 ANTENNAS

Antennas transmit and receive electromagnetic radiation from space through a wire or coaxial cable(or any other appropriate conductor).The isotropic radiator, a point in space that radiates equally in all directions, serves as a theoretical reference antenna. All places with equal power are situated on a sphere with the antenna at its centre. Figure 7, which shows a

two-dimensional cross-section of the actual three-dimensional pattern, shows that the radiation pattern is symmetric in all directions.

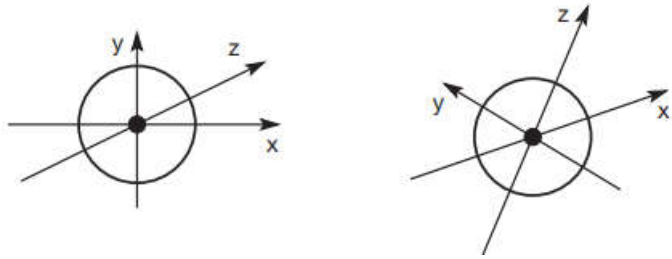


Figure 7: Radiation pattern of an isotropic radiator

But in reality, such an antenna does not exist. Real antennas all have directional effects, which means that the radiation intensity varies depending on the direction the antenna is facing. A thin, center-fed dipole, also known as a Hertzian dipole, is the most basic actual antenna and is depicted in Figure 8 (right-hand side). The dipole comprises of two equal-length collinear conductors separated by a tiny feeding gap. The dipole's length is not arbitrary, but, for instance, cutting the wavelength λ in half of the signal's transmission yields extremely effective energy radiation. The length of λ/4 is effective when put on a car's roof. Additionally called the Marconi antenna

Figure 8: Simple antennas

As seen in Figure 9, a λ/2 dipole radiates uniformly or omnidirectionally in one plane and in a figure-eight pattern in the other two. The only way this kind of antenna can overcome environmental obstacles is by increasing the signal's power. Challenges may include hills, valleys, structures, etc.

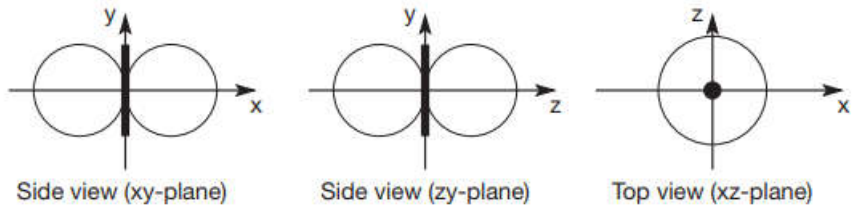


Figure 9: Radiation pattern of a simple dipole

An omnidirectional radiation pattern is not very useful if an antenna is installed, for instance, in a valley or between two structures. Directional antennas with predetermined set favored transmission and reception directions can be employed in this situation. The radiation pattern of a

directional antenna with its main lobe facing the x-axis is depicted in Figure 10. Satellite dishes are an exceptional type of directional antenna.

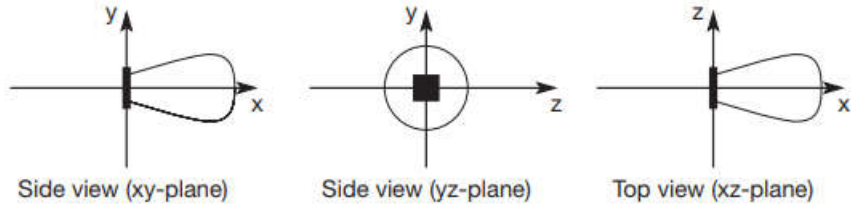


Figure 10:Radiation pattern of a directed antenna

Cellular systems frequently employ directed antennas. A sectorized antenna can be built by combining several directed antennas on a single pole. It is possible to sectorize a cell into, say, three or six sectors, allowing for frequency reuse. The radiation patterns of these sectorized antennas are depicted in Figure 11.

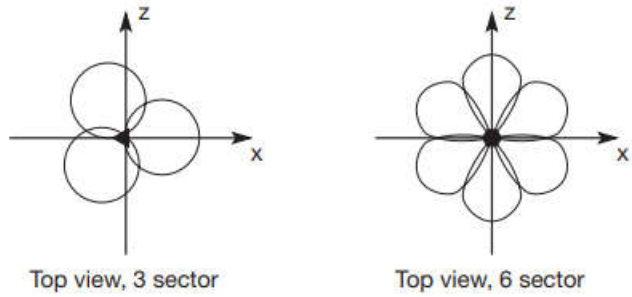


Figure 11: Radiation patterns of sectorized antennas

It is also possible to combine two or more antennas to enhance reception by reducing the impacts of multi-path propagation.Different diversity methods are possible with these antennas, commonly known as multi-element antenna arrays. Switched diversity or selection diversity is one such strategy where the receiver always employs the antenna element with the highest output. The power of all signals is combined to achieve gain by diversity combining. To prevent cancellation, the phase is first rectified (cophasing). As seen in Figure 12, various plans are viable. On top of a ground plane, two λ/4 antennas are combined on the left with a λ/2 gap between them.On the right, three standard λ/2 dipoles are combined with a distance of λ/2 between them. Spacing could also be in multiples of λ/2.

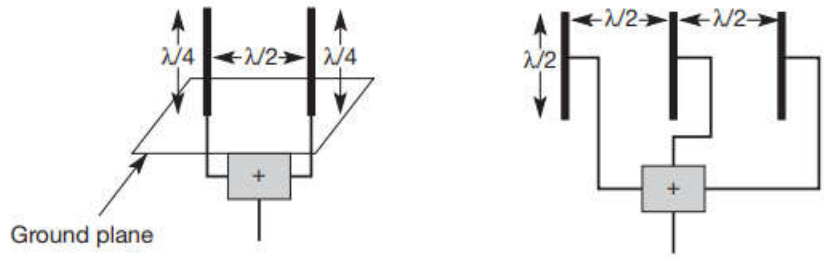


Figure 12:Diversity antenna systems

Smart antennas, which combine several antenna elements (also known as antenna array) with signal processing to optimize the radiation/reception pattern in response to the signal environment, offer a more sophisticated solution. These antennas can adjust to variations in transmission conditions, reception power, and a variety of signal propagation effects. Beam formation can also involve antenna arrays. This would be an extreme example of a directed antenna that may employ space division multiplexing to track a single user. Base stations wouldn't be the only devices that could follow users with a specific beam. Wireless devices may also aim their electromagnetic radiation at a base station rather than at a person, for example. This would aid in lowering the radiation absorbed.

6.7 SIGNAL PROPAGATION

Wireless communication networks have signal senders and receivers much like wired networks do. However, these two networks show significant disparities in terms of signal propagation. In contrast to wired networks, where a signal may only travel via a wire, wireless networks do not have a wire for the signal to use to establish the direction of propagation(which can be twisted pair copper wires, a coax cable, but also a fiber etc.). The wire usually displays the same traits at each place as long as it is not cut or damaged. Depending on the length, one can exactly predict how a signal will behave while passing over this wire, for example, received power. This predictable behavior for wireless transmission only applies when there is nothing between the sender and the receiver, or in a vacuum. This is illustrated in figure 13.

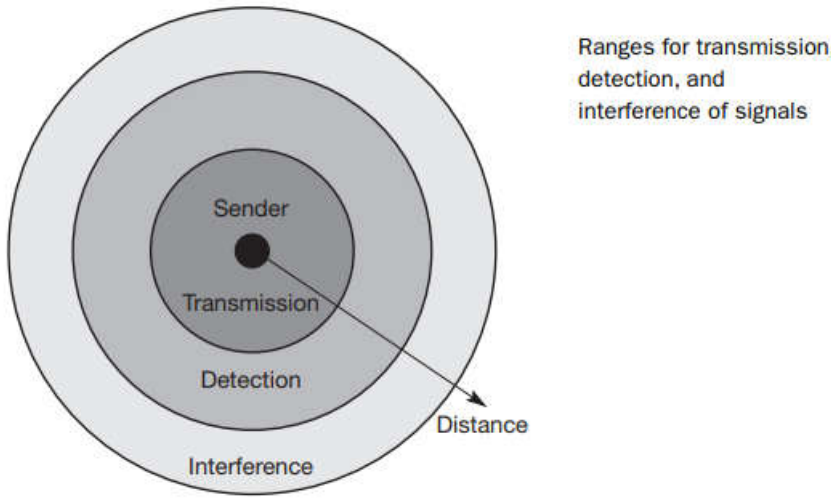


Figure 13

- Transmission range: Within a certain radius of the sender transmission is possible, i.e., a receiver receives the signals with an error rate low enough to be able to communicate and can also act as sender.
- Detection range: Within a second radius, detection of the transmission is possible, i.e., the transmitted power is large enough

to differ from background noise. However, the error rate is too high to establish communication.

- Interference range: Within a third even larger radius, the sender may interfere with other transmission by adding to the background noise. A receiver will not be able to detect the signals, but the signals may disturb other signals.

Cells arranged around a transmitter are the result of this straightforward and excellent plan. Real life, on the other hand, does not take place in a vacuum; radio transmission must deal with obstacles like our atmosphere, mountains, structures, moving senders and receivers, etc. The three circles mentioned above will actually be oddly shaped polygons whose shapes vary with time and frequency.

6.8 MULTIPLEXING

Multiplexing is not only a fundamental mechanism in communication systems but also in everyday life. Multiplexing describes how several users can share a medium with minimum or no interference. One example, is highways with several lanes. Many users (car drivers) use the same medium (the highways) with hopefully no interference (i.e., accidents). This is possible due to the provision of several lanes (space division multiplexing) separating the traffic. In addition, different cars use the same medium (i.e., the same lane) at different points in time (time division multiplexing).

Multiple access schemes are used to allow multiple mobile users to share a finite quantity of radio spectrum at the same time.

6.8.1 Multiple Access Techniques

It is common in wireless communication systems for subscribers to be able to send information from the mobile station to the base station while simultaneously receiving information from the base station to the mobile station.

A cellular system splits an area into cells, with each cell's mobile unit communicating with a base station. The primary goal of cellular system design is to maximise channel capacity, or the ability to handle as many calls as feasible in a given bandwidth while maintaining a sufficient level of service.

There are numerous options for granting access to the channel. These primarily comprise the following:

- Frequency Division Multiple Access (FDMA)
- Time Division Multiple Access (TDMA)
- Code Division Multiple Access (CDMA)
- Space Division Multiple Access (SDMA)
- Spread Spectrum Multiple Access

These strategies can be classed as narrowband or wideband systems depending on how the available bandwidth is allocated to the users.

- Narrowband systems

Narrow band systems operate with channels that are significantly narrower than the coherence bandwidth. Narrow band TDMA allows users to share the same channel but assigns each user a distinct time slot, allowing a small number of users to be separated in time on a single channel.

- Wideband systems

In wideband systems, a single channel's transmission bandwidth is substantially wider than the channel's coherence bandwidth. As a result, multipath fading has a minor impact on the received signal in a wideband channel, and frequency selective fades only occur in a limited portion of the signal bandwidth.

6.8.2 Frequency Division Multiple Access (FDMA)

The FDMA standard is the foundation for enhanced mobile phone services. The following are the characteristics of FDMA:

- For each user to access the network, FDMA assigns a different sub-band of frequency.
- When FDMA is not in use, the channel is left idle rather than being assigned to other users.
- Narrowband systems use FDMA, which is less complicated than TDMA.
- To reduce adjacent channel interference, tight filtering is used.
- In FDMA, the base station BS and the mobile station MS transmit and receive data at the same time.

6.8.3 Time Division Multiple Access (TDMA)

TDMA is utilised instead of FDMA in situations where continuous transmission is not required. The following are some of the characteristics of TDMA:

- TDMA allows several users to share a single carrier frequency by using non-overlapping time intervals.
- In TDMA, data is transmitted in bursts rather than continuously. As a result, the process is simplified.
- Duplexers are not required because TDMA employs different time slots for transmission and reception.
- TDMA offers the advantage of allowing different users to be assigned varied numbers of time slots every frame.
- By concatenating or reassigning time slots based on priority, bandwidth can be provided on demand to multiple users.

6.8.4 Code Division Multiple Access (CDMA)

Multiple access techniques such as code division multiple access employ a single channel to broadcast data simultaneously from multiple transmitters. The following are its characteristics:

- Instead than being assigned a specific frequency, each CDMA customer uses the entire available spectrum.
- For voice and data connections, CDMA is highly recommended.
- While many codes share the same CDMA channel, users with the same code can connect with one another.
- CDMA has a greater capacity for airspace than TDMA.
- CDMA handles the hand-off between base stations quite well.

6.8.5 Space Division Multiple Access (SDMA)

Space division multiple access, also known as spatial division multiple access, is a MIMO (multiple-input multiple-output) architecture technique that is commonly used in wireless and satellite communication. It has the following characteristics:

- Using the same channel, all users can communicate at the same time.
- SDMA is fully interference-free.
- A single satellite can communicate with many satellites using the same frequency receiver.
- The base station in SDMA can monitor a moving user thanks to the use of directional spot-beam antennas.
- Controls the amount of energy radiated by each user in space.

6.8.6 Spread Spectrum Multiple Access

Signals with a transmission bandwidth greater than the minimum needed RF bandwidth are used in spread spectrum multiple access (SSMA).

Spread spectrum multiple access approaches are divided into two categories:

- Frequency hopped spread spectrum (FHSS)
- Direct sequence spread spectrum (DSSS)

- Frequency hopped spread spectrum (FHSS)

This is a digital multiple access system in which the carrier frequencies of individual users within a wideband channel are altered in a pseudo random manner. The digital data is split down into uniformly sized bursts that are then sent over various carrier frequencies.

- Direct sequence spread spectrum (DSSS)

This is the most widely utilised CDMA technology. A Pseudo Random Noise Code multiplies the message signal in DS-SS. Each user is assigned

a code word that is orthogonal to the codes of other users, and the receiver must know the transmitter's code word in order to detect the user.

Another sort of spread spectrum is hybrid, which is made up of combinational sequences. Another type that is rarely mentioned is time hopping.

Spread spectrum systems become bandwidth efficient in a multiple user scenario because numerous users can share the same spread spectrum bandwidth without interfering with one another.

6.9 MODULATION

Modulation is a process of changing the characteristics of the wave to be transmitted by superimposing the message signal on the high-frequency signal. In this process video, voice and other data signals modify high-frequency signals – also known as the carrier wave. This carrier wave can be DC or AC or pulse chain depending on the application used. Usually, a high-frequency sine wave is used as a carrier wave signal.

These modulation techniques are classified into two major types: analog and digital or pulse modulation.

Different Types of Modulation

The two types of modulation: analog and digital modulation techniques have already been discussed. In both the techniques, the baseband information is converted to Radio Frequency signals, but in analog modulation, these RF communication signals are a continuous range of values, whereas in digital modulation these are prearranged discrete states.

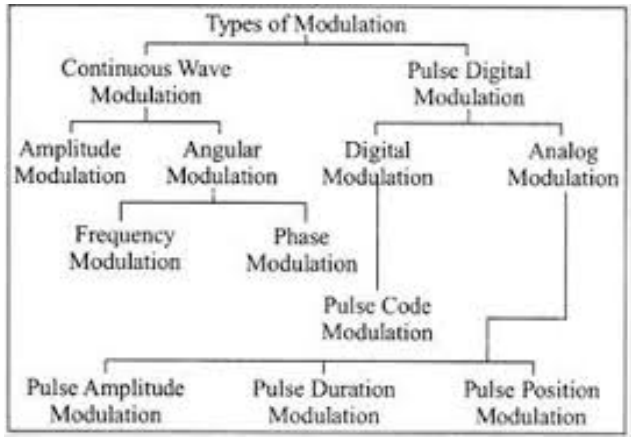


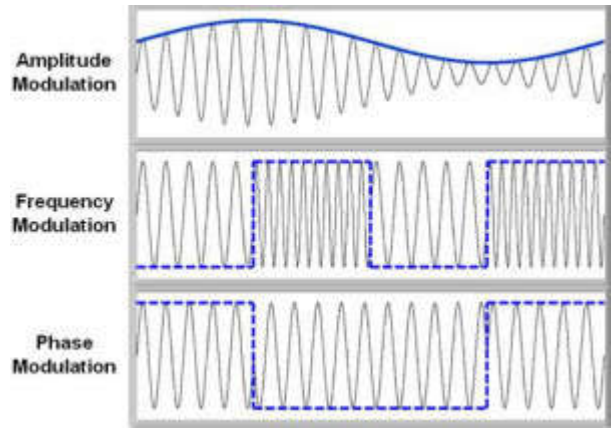
Figure: Types of modulation

Analog Modulation

In this modulation, a continuously varying sine wave is used as a carrier wave that modulates the message signal or data signal. The Sinusoidal wave's general function is shown in the figure below, in which, three parameters can be altered to get modulation – they are mainly amplitude, frequency, and phase, so the types of analog modulation are:

- Amplitude modulation (AM)
- Frequency modulation (FM)
- Phase modulation (PM)

In amplitude modulation, the amplitude of the carrier wave is varied in proportion to the message signal, and the other factors like frequency and phase remain constant. The modulated signal is shown in the below figure, and its spectrum consists of a lower frequency band, upper-frequency band, and carrier frequency components. This type of modulation requires greater bandwidth, more power. Filtering is very difficult in this modulation.

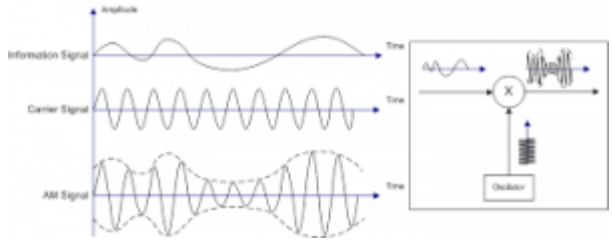


Types of Analog Modulation

Frequency modulation (FM) varies the frequency of the carrier in proportion to the message or data signal while maintaining other parameters constant. The advantage of FM over AM is the greater suppression of noise at the expense of bandwidth in FM. It is used in applications like radio, radar, telemetry seismic prospecting, and so on. The efficiency and bandwidths depend on the modulation index and maximum modulating frequency.

In phase modulation, the carrier phase is varied in accordance with the data signal. In this type of modulation, when the phase is changed it also affects the frequency, so this modulation also comes under frequency modulation.

Analog modulation (AM, FM, and PM) is more sensitive to noise. If noise enters into a system, it persists and gets carried till the end receiver. Therefore, this drawback can be overcome by the digital modulation technique.



AM

Digital Modulation

For better quality and efficient communication, the digital modulation technique is employed. The main advantages of digital modulation over analog modulation include permissible power, available bandwidth, and high noise immunity. In digital modulation, a message signal is converted from analog to digital message and then modulated by using a carrier wave.

The carrier wave is keyed or switched on and off to create pulses such that the signal is modulated. Similar to the analog, here the parameters like amplitude, frequency, and phase variation of the carrier wave decides the type of digital modulation.

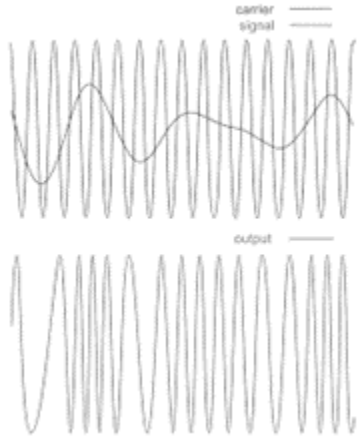
The types of digital modulation are based on the type of signal and application used such as Amplitude Shift Keying, Frequency Shift Keying, Phase Shift Keying, Differential Phase Shift Keying, Quadrature Phase Shift Keying, Minimum Shift Keying, Gaussian Minimum Shift Keying, Orthogonal Frequency Division Multiplexing, etc., as shown in the figure.

Amplitude shift keying changes the amplitude of the carrier wave based on the baseband signal or message signal, which is in digital format. It is used for low-band requirements and is sensitive to noise.

In frequency-shift keying, the frequency of the carrier wave is varied for each symbol in the digital data. It needs larger bandwidths as shown in the figure. Similarly, the phase shift keying changes the phase of the carrier for each symbol and it is less sensitive to noise.

Frequency Modulation

In order to create a frequency modulated wave, the frequency of the radio wave is varied in accordance with the amplitude of the input signal.



Frequency Modulation

When the audio wave is modulated with that of the radio frequency carrier signal, then the generated frequency signal will change its frequency level.

The variation by which the wave moves upward and downward is to be noted. This is termed as deviation and is generally represented as kHz deviation.

As an instance, when the signal has a deviation of either + or – 3kHz, then it is represented as ±3kHz. This means that the carrier signal has up and downward deviation of 3kHz.

Broadcasting stations that need very high-frequency range in the frequency spectrum (in the range of 88.5 – 108 MHz), they need certainly a large amount of deviation which is nearly ±75 kHz. This is called wide-band frequency modulation. The signals in this range hold the ability to assist the high quality of transmissions, whereas they require higher bandwidth too. In general, 200kHz is permitted for every WBFM. And for narrowband FM, a deviation of ±3 kHz is enough.

While implementing an FM wave, it is more beneficial to know the effectivity range of the modulation. This stands as the parameter in stating factors such as knowing the type of signal whether wide band or narrow band FM signal. It also helps in making sure that the whole receivers or transmitters that are in the system are programmed to adapt to the standardized range of modulation as this shows an impact on the factors such as the channel spacing, bandwidth of the receiver, and others.

So, to signify the modulation level, modulation index and deviation ratio parameters are to be determined.

The different types of frequency modulation include the following.

Narrow band FM

- This is termed as the type of frequency modulation where the modulation index value is too minimal.
- When the modulation index value is < 0.3, then there will be an only carrier and corresponding sidebands having bandwidth as twice the modulating signal. So, $\beta \leq 0.3$ is called narrow band frequency modulation.
- The maximum range of modulating frequency is of 3 kHz
- The maximum frequency deviation value is 75 kHz

Wide band FM

- This is termed as the type of frequency modulation where the modulation index value is large.
- When the modulation index value is > 0.3, then there will be more than two sidebands having bandwidth as twice the modulating signal. When the modulation index value increases, then the number of sidebands gets increased. So, $\beta > 0.3$ is called narrow band frequency modulation.

- The maximum range of modulating frequencies is in between 30 Hz – 15 kHz
- The maximum frequency deviation value is 75 kHz
- This frequency modulation needs a higher bandwidth range which is almost 15 times ahead of the narrow band frequency modulation.

The other types of modulation techniques used in the communication system are:

- Binary phase shift keying
- Differential phase-shift keying
- Differential quadrature phase shift keying
- Offset quadrature phase shift keying
- Audio FSK
- Multi FSK
- Dual-tone FSK
- Minimum shift keying
- Gaussian minimum shift keying
- Trellis coded type of modulation

6.10 CELLULAR SYSTEMS

SDM is used in cellular networks for mobile communications. Each transmitter, commonly referred to as a base station, serves a specific cell. Cell radii can be as small as a few metres in a building, hundreds of metres in a city, or even tens of kilometres across the nation. As seen in Figure 14, cells are never perfectly round or hexagonal; instead, the shape of a cell depends on its surroundings (buildings, mountains, valleys, etc.), the weather, and occasionally even the strain on the system. This strategy is frequently used in mobile telecommunication systems, where a mobile station inside a base station's cell can connect with that base station and vice versa.

Figure
Cellular system with three and seven cell clusters

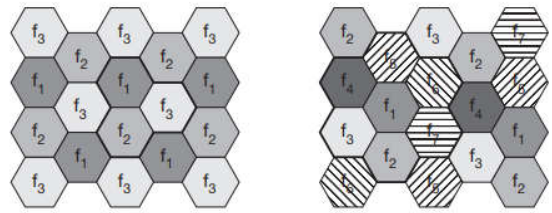


Figure 14

In this context, the question of why mobile network operators do not employ strong transmitters with large cells like, for example, radio stations do, instead of installing thousands of base stations across a nation (which is rather expensive), emerges.Small cell cellular systems provide the following benefits:

- **Greater capacity:** Frequency reuse is possible with SDM. One transmitter can use the same frequencies if it is far from another, or outside the interference range. This frequency is prohibited for other users since most mobile phone systems assign frequencies to certain users (or specific hopping patterns). However, frequencies are a finite resource, and each cell can only support a very small number of concurrent users. Large cells cannot accommodate more users. Instead, they are constrained to fewer potential consumers per km^2. In cities where a significant number of people use mobile phones, there is another justification for employing very small cells.
- **Less transmission power:** While power issues may not pose a significant threat to base stations, they do pose a threat to mobile stations. The few Watts of transmit power now available would not be sufficient for a receiver located far from a base station. However, energy is a significant issue for portable, mobile electronics.
- **Local interference only:** Interference issues are exacerbated by large separations between the sender and receiver. Mobile and base stations only have to contend with "local" interference when using tiny cells.
- **Robustness:** Because they are decentralized, cellular systems are better able to withstand the failure of a single component. If one antenna malfunctions, communications are only impacted locally.

Additionally, small cells have a few drawbacks:

- Cellular systems require a sophisticated infrastructure to link all base stations. This costs quite a bit because it requires numerous antennas, switches for call forwarding, position registers to locate a mobile station, etc.
- Handover needed:When switching from one cell to another, the mobile station must complete a handover. This can occur frequently, depending on the size of the cell and the rate of movement.
- Frequency planning: To avoid interference between transmitters using the same frequencies, frequencies have to be distributed carefully. On the one hand, interference should be avoided, on the other, only a limited number of frequencies is available.

Different transmitters that are inside each other's interference range employ FDM to prevent interference. The hopping pattern must be coordinated if FDM and TDM are mixed. Within the interference range, the basic rule is to avoid using the same frequency at the same time(if CDM is not applied).Figure 14 depicts two potential models for producing cell patterns with less interference. Clusters of cells are formed; on the left side, three cells create a cluster, while on the right, seven cells do the same. A cluster's cells all employ different sets of frequencies. Sets f1, f2, and f3 are used by the cluster's first, second, and third cells, respectively,

on the left.The pattern will resemble something slightly different in actual transmission. To illustrate the model simply, the hexagonal layout was chosen. The repeating of the same frequency sets is also demonstrated by this pattern. A sender's transmission strength must be constrained to prevent interference with the subsequent cell using the same frequencies.

Sectorized antennas can be utilised to further reduce interference (particularly under specific traffic conditions, such as the number of users per km^2). In a cluster of three cells, as shown in Figure 15, three sectors are utilised per cell. Generally, with larger cell radii, sectorized antennas make more sense than omni-directional antennas.

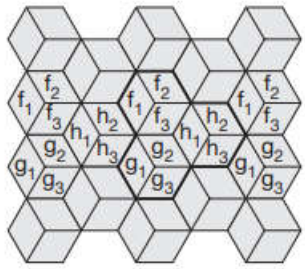


Figure
Cellular system with three cell clusters and three sectors per cell

Figure 15

If traffic demand varies, the fixed frequency assignment to cell clusters and cells is not very effective. It might make sense to "borrow" frequencies, for example, if one cell has a strong load while the neighbouring cell has a light burden. More frequencies are dynamically assigned to cells with more traffic.While the first fixed method is known as fixed channel allocation (FCA), this one is known as borrowing channel allocation (BCA). The GSM system uses FCA because it is easier to use, but rigorous traffic analysis is needed before installation.

DECT has a dynamic channel allocation (DCA) system in place. In this system, frequencies can be freely assigned to cells, but they can also only be borrowed. The risk of interference with cells using the same frequency exists with dynamic frequency assignment to cells. The adjacent cells can block the "borrowed" frequency.Such intricate frequency planning and elaborate channel allocation algorithms are not necessary in cellular systems that use CDM rather than FDM. Users are distinguished in this case based on the code they employ rather than frequency. Another issue with cell planning is that cell size is dependent on the level of load. As a result, CDM cells are frequently referred to as "breathing." A cell can cover a wider area when under a light load, but as the load increases, it contracts. If there are more users in a cell, the noise level will increase, which is the cause of this.more noise levels result in more route loss and transmission mistakes. Last but not least, mobile stations farther from the base station disconnect from the cell. (This is like attempting to talk to someone who is far away at a busy party.) This behaviour is shown in figure 16 with a user transmitting a high bit rate stream inside a CDM cell. Two users leave the cell as a result of this additional user, causing the cell to decrease. In a real-world scenario, an additional user might ask for a

video broadcast (at a high bit rate), while the rest speak normally (low bit rate).

Figure
Cell breathing depending on the current load

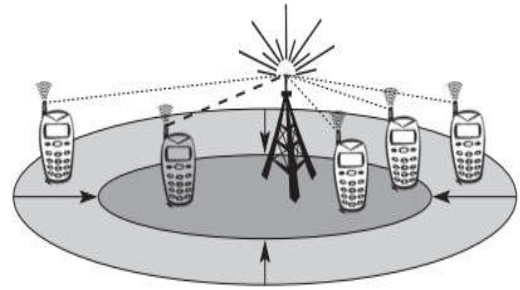


Figure 16

6.11 SUMMARY

Antennas are required for the transmission and receiving of electromagnetic waves, which are the foundation of wireless communication. Omni-directional antennas are preferred for mobile devices while directed antennas are frequently used in mobile phone system base stations. Electromagnetic waves can experience a variety of side effects when travelling from originator to receiver. The common effectsinclude shadowing, fading, reflection, diffraction, and scattering. Multi-path propagation is one of the main issues in wireless communication as a result of all these phenomena. As a result of intersymbol interference, or when one symbol is "smeared" into another symbol as a result of delay spread, multi-path propagation reduces the channel bandwidth.

Since wireless transmission uses just one "medium," multiplexing techniques can be used to increase overall capacity. SDM, FDM, TDM, and CDM are considered the standard schemes. Data must be "translated" into a signal with a specific carrier frequency in order to achieve FDM. Consequently, two modulation steps are possible. Analog modulation pushes the signal's centre frequency up to the radio carrier whereas digital modulation encrypts data into a baseband signal. Many bits can be encoded into a single-phase shift using some cutting-edge techniques, which increases efficiency.

Spread spectrum technology can be used to implement a number of functionalities. One is (at least some) security because the signal appears to be noise to someone who doesn't know the spreading code. The code space is the foundation for spread spectrum special medium access techniques.Since the signal is dispersed over a wider bandwidth, narrowband interference only affects a tiny portion of the signal, making a transmission more resistant to it thanks to spread spectrum.

Finally, we discussed about Cellular systems. SDM is used by cellular systems to increase the total capacity of mobile phone systems.Although these systems necessitate meticulous planning (i.e., matching the cell size

to the anticipated traffic), they offer one of the fundamental approaches to effectively using the limited frequency resources.

6.12 LIST OF REFERENCES

1) Protocols and Architectures for Wireless Sensor Network, Holger Kerl, Andreas Willig, John Wiley and Sons, 2005

2) Wireless Sensor Networks Technology, Protocols, and Applications,Kazem Sohraby, Daniel Minoli and TaiebZnati, John Wiley & Sons, 2007

3) Mobile communications, Jochen Schiller,2nd Edition, Addison wisely, Pearson Education,2012

4) Fundamentals of Wireless Sensor Networks, Theory and Practice, Waltenegus Dargie, Christian Poellabauer, Wiley Series on wireless Communication and Mobile Computing, 2011

5) Networking Wireless Sensors, Bhaskar Krishnamachari, Cambridge University Press, 2005

6.13 UNIT END EXERCISES

1. State and explain the applications of wireless transmission.
2. Discuss about the history of wireless communication.
3. Explain Frequency for radio transmission.
4. Write a note on Signals.
5. Describe about Antennas and its types.
6. What is Signal propagation?
7. Explain Multiplexing.
8. Explain Modulation.
9. Write a note on Cellular systems.

7

TELECOMMUNICATION, SATELLITE AND BROADCAST SYSTEMS: GSM

Unit Structure

7.0 OBJECTIVES

- To understand the concept of GSM and how it is utilised for voice traffic
- To understand DECT architecture
- To gain basic insights about: GEO, LEO, MEO

7.1 INTRODUCTION

Digital cellular networks are the fastest-growing part of the market for mobile and wireless devices. They are wireless extensions of standard PSTN or ISDN networks, allowing for smooth roaming within the same country or even globally. These systems are mostly utilised for voice traffic nowadays. However, because data traffic is always increasing, this chapter discusses many wireless data transfer strategies using cellular systems.

The following are the global market figures for cellular networks (GSM Association, 2002). GSM is the most widely used digital system, accounting for over 70% of the market. The analogue AMPS system still has 3% of the market, while the Japanese PDC has 5%. (60 million users). The rest is split between CDMA (12%) and TDMA (10%) systems, as well as other technologies. Nearly everyone in Europe (about 370 million people) uses the digital GSM system, with almost no analogue systems remaining.In the United States and several other countries that have absorbed US technology, the situation is different (e.g., South Korea, Canada). With 107 million TDMA, 135 million CDMA, and just 16 million GSM users (North America only), the digital market is divided into TDMA, CDMA, and GSM systems. While Europe has only one digital system, the US market is fragmented into numerous. This causes serious coverage and service availability issues, and is an example of how market forces failed to deliver better services (compared to the common standard in Europe).

The figure below depicts the global number of subscribers to various mobile phone technology (GSM Association, 2002). The illustration mixes various versions of the same technology (e.g., GSM working on 900, 1,800, and 1,900 MHz). The graph's two upper lines depict the total number of users and 1998 predictions. It's interesting to note that no one predicted mobile communication technology's enormous success. The graph also demonstrates that analogue systems are no longer in use, with GSM dominating the present market.Second generation systems include GSM, TDMA, CDMA, and PDC. It's worth noting that mobile phones are now used by more people than landlines!In March 2002, the graphs of mobile and fixed users crossed.

Figure
Worldwide subscribers of different mobile phone technologies

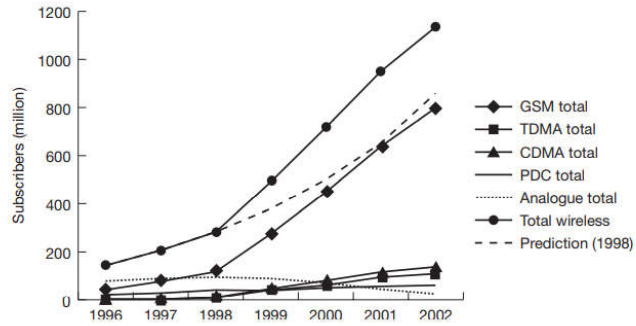


Figure: Worldwide subscribers of different mobile phone technologies

7.2 MOBILE SERVICES

GSM makes it possible to combine various voice and data services and communicate with current networks. Customers find a network appealing because of its services. Bearer, tele, and supplemental services are the three different categories of services that GSM has established. The subsections that follow provide descriptions of them.A reference model for GSM services is shown in Figure below.

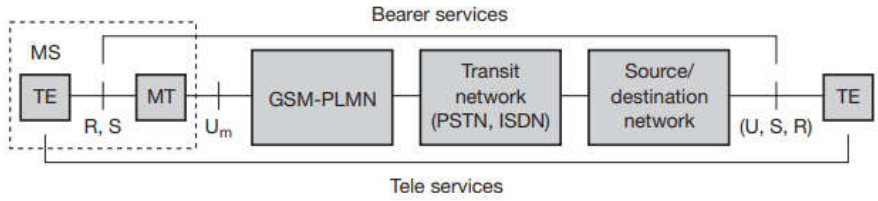


Figure: Bearer and tele services reference model

A mobile station MS is connected to the GSM public land mobile network (PLMN) via the Um interface. (GSM-PLMN is the infrastructure needed for the GSM network.) This network is connected to transit networks, e.g., integrated services digital network (ISDN) or traditional public switched telephone network (PSTN). There might be an additional network, the source/destination network, before another terminal TE is connected. Bearer services now comprise all services that enable the transparent transmission of data between the interfaces to the network, i.e., S in case of the mobile station, and a similar interface for the other terminal (e.g., S0 for ISDN terminals). Interfaces like U, S, and R in case of ISDN have not been defined for all networks, so it depends on the specific network which interface is used as a reference for the transparent transmission of data. In the classical GSM model, bearer services are connection-oriented and circuit- or packet-switched. These services only need the lower three layers of the ISO/OSI reference model.

The mobile termination (MT) in the mobile station MS handles all network-specific duties (TDMA, FDMA, coding, etc.) and provides the terminal (TE), which can then be network-independent, with an interface for data transmission (S). According to the ISDN reference model, additional interfaces, such R, may be required depending on TE's capabilities. Since tele services are application-specific, all seven layers of the ISO/OSI reference model can be required. These services are described from one terminal TE to another, or end-to-end.

1] Bearer services

Different data transmission protocols are specified by GSM, with the original GSM allowing for non-voice service data speeds of up to 9600 bit/s. Transparent and opaque, synchronous and asynchronous data transmission is possible with bearer services. Only the physical layer's (layer 1) capabilities are used by transparent bearer services for data transmission. If there are no transmission defects, data transmission has a

constant delay and throughput. Forward error correction (FEC), which incorporates redundancy into the data stream and aids in the reconstruction of the original data in the event of transmission faults, is the only method for improving transmission quality. Data rates of 2.4, 4.8, or 9.6 kbit/s are feasible, depending on the FEC.

Error correction and flow control are implemented via protocols at layers two and three in non-transparent bearer services. These services incorporate a radio link protocol while utilising transparent bearer services. (RLP). This protocol includes high-level data link control (HDLC) mechanisms and unique selective-reject techniques to force the retransmission of inaccurate data. Although less than 10^{-7} bit errors were obtained, throughput and delay may now differ depending on the quality of the transmission.

2] Tele services

GSM primarily focuses on teleservices that are voice-oriented. These include message services, basic data connection with terminals that are familiar from the PSTN or ISDN, and encrypted voice transfer. (e.g., fax). However, as telephony is the principal service, the fundamental objective of GSM was to provide high-quality digital voice transmission, at least providing the normal bandwidth of 3.1 kHz of analogue phone networks. While ordinary codecs are used to transmit analogue data for use with conventional computer modems found in, for example, fax machines, special codecs (coder/decoder) are utilised for voice communication.Another service offered by GSM is the emergency number.A useful service for very simple message transfer is the short message service(SMS), which offers transmission of messages of up to 160 characters.Another non-voice tele service is group 3 fax, which is available worldwide. In this service, fax data is transmitted as digital data over the analog telephone network according to the ITU-T standards T.4 and T.30 using modems.

3] Supplementary services

GSM service providers may provide additional services in addition to tele and bearer services. These services, which can vary from provider to provider, provide numerous upgrades for the basic telephony service, much like ISDN networks.User identification, call forwarding, and phone redirection are examples of typical services. There may be access to standard ISDN capabilities like multiparty communication and locked user groups. Companies are particularly interested in closed user groups because they enable features like a company-specific GSM sub-network that is only accessible to group members.

7.3 SYSTEM ARCHITECTURE

GSM, like many telecommunications systems, has a hierarchical, complicated system architecture with numerous entities, interfaces, and acronyms. The GSM system, as defined by ETSI, is depicted in the diagram. The radio subsystem (RSS), the network and switching

subsystem (NSS), and the operation subsystem (OS) make up a GSM system (OSS). The mobile stations (MS) and some antenna masts of the base transceiver stations are usually the only parts of the network that a GSM customer notices (BTS).

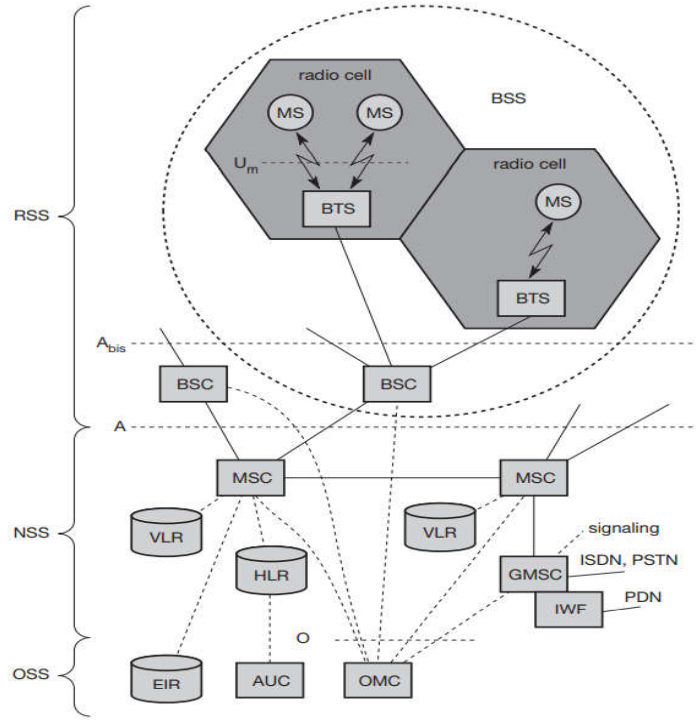


Figure: GSM system functional architecture

7.3.1 Radio subsystem

All radio-specific elements, such as mobile stations (MS) and base station subsystems, are included in the radio subsystem (RSS) (BSS). The RSS and NSS are connected via the A interface (solid lines) while the OSS is connected via the O interface (dashed lines) in the diagram above. The A interface employs circuit-switched PCM-30 systems (2.048 Mbit/s) to handle up to 30 64 kbit/s connections, whilst the O interface uses the Signalling System No. 7 (SS7) based on X.25 to carry management data to and from the RSS.

- **Base station subsystem (BSS):** A GSM network has numerous base station subsystems, each controlled by a base station controller (BSC). The BSS is responsible for maintaining radio connections to an MS, voice coding and decoding, and rate adaption to and from the wireless network. The BSS contains many BTSs in addition to a BSC.

- **Base transceiver station (BTS):** It contains all radio equipment required for radio transmission, such as antennas, signal processing, and amplifiers. A BTS connects to the MS via the U_m interface (ISDN U interface for mobile use) and to the BSC via the A_{bis} interface, and can build a radio cell or many cells using sectorized antennas. The Um interface contains all of the wireless transmission techniques (TDMA,

FDMA etc.) 16 or 64 kbit/s connections make up the A_{bis} interface. Depending on the context (buildings, open space, mountains, etc.) as well as projected traffic, a GSM cell can measure between 100 m and 35 km.

- **Base station controller (BSC):** The BSC is in charge of the BTSs. It manages the handover of radio frequencies from one BTS to another within the BSS and performs MS paging. At the A interface, the BSC multiplexes the radio channels onto the fixed network connections.

7.3.2 Network and Switching subsystem

The network and switching subsystems are the "heart" of the GSM system (NSS). The NSS connects the wireless network to conventional public networks, handles handovers between multiple BSSs, includes capabilities for global user localisation, and facilitates charging, accounting, and roaming of users across different providers and countries. The NSS is made up of the switches and databases listed below:

- **Mobile Services Switching Centres:** MSCs are high-speed digital ISDN switches. They use the A interface to link to other MSCs and BSCs, forming the GSM system's fixed backbone network.Typically, an MSC is in charge of numerous BSCs in a certain geographic area. Other fixed networks, such as PSTN and ISDN, are connected to a gateway MSC (GMSC). An MSC can additionally connect to public data networks (PDNs) like X.25 using additional interworking functions (IWF). An MSC is in charge of all signals for connection setup, connection release, and connection handover to other MSCs. For this, the standard signalling system No. 7 (SS7) is employed.SS7 includes all aspects of digital network control signalling (reliable routing and delivery of control messages, call establishment and monitoring).

- **Home Location Register (HLR):** The HLR is the most significant database in a GSM system since it stores all user-relevant data. This includes static data like the mobile subscriber ISDN number (MSISDN), subscribed services (including call forwarding, roaming limitations, and GPRS), and the international mobile subscriber identity (IMSI). The current location area (LA) of the MS, the mobile subscriber roaming number (MSRN), the current VLR, and MSC are all examples of dynamic information. The information in the HLR is updated as soon as an MS departs its current LA.This data is required to pinpoint a user's location within the global GSM network. In a single HLR, which also allows billing and accounting, all of these user-specific information pieces exist just once for each user. HLRs can manage data for millions of clients and contain highly specialised data bases that must meet particular real-time standards in order to respond to queries within certain time frames.

- **Visitor Location register (VLR):** The VLR associated with each MSC is a dynamic database that stores all necessary information for

MS users currently in the LA associated with the MSC (e.g., IMSI, MSISDN, HLR address). When a new MS joins a LA for which the VLR is responsible, the HLR copies all necessary information for this user. This VLR and HLR structure prevents frequent HLR updates and long-distance user information communication.

7.3.3 Operation subsystem

The operating subsystem (OSS), the third component of a GSM system, contains the functions required for network operation and maintenance. The OSS has its own network entities and communicates with others via SS7 signalling (see Figure above). The entities listed below have been defined.

- **The operation and maintenance centre (OMC):** It uses the O interface to monitor and control all other network elements (SS7 with X.25). Traffic monitoring, network entity status reports, subscriber and security management, and accounting and billing are all common OMC administration responsibilities.Telecommunication management networks (TMNs), as defined by the ITU-T, are used by OMCs.

- **Authentication Centre (AuC):** Because the radio interface and mobile stations are particularly vulnerable, a separate authentication centre (AuC) has been established to protect user identity and data transmission. The AuC contains the authentication methods as well as the encryption keys, and it generates the values required for user authentication in the HLR. The AuC might be located in a designated protected area within the HLR.

- **The Equipment Identity Register (EIR):** EIR is a database that contains all IMEIs, or all device identifications registered for this network. MSs are easily stolen because they are mobile. Anyone with a valid SIM might use the stolen MS. The EIR maintains a stolen (or locked) device blacklist. An MS is theoretically worthless once the owner has reported a theft. Unfortunately, different providers' blacklists are not always synchronised, making it feasible to use a device in another operator's network without permission. A list of valid IMEIs (white list) and a list of malfunctioning devices are also included in the EIR (gray list).

7.4 RADIO INTERFACE

The radio interface is the most significant interface in a GSM system since it contains several techniques for multiplexing and media access. GSM uses cells with BTS to implement SDMA and assigns an MS to each BTS. As demonstrated in Figure, FDD is also utilised to segregate the downlink and uplink. TDMA and FDMA are combined in media access. For FDMA, GSM 900 employs 124 channels, each 200 kHz wide, whereas GSM 1800 requires 374 channels. Channels 1 and 124 are not used for transmission in GSM 900 due to technical reasons.In most cases, 32 channels are dedicated to corporate data, while the remaining 90 are dedicated to

customers. After that, each BTS maintains a single channel for organisational data and up to ten channels for user data, for example.The sample below uses the GSM 900 system, while GSM operates similarly at 1800 and 1900 MHz.

The TDM used is also shown in the figure. A GSM TDMA frame separates each of the 248 channels in time, i.e., each 200 kHz carrier is subdivided into frames that are repeated constantly. A frame lasts 4.615 milliseconds. A frame is split into 8 GSM time slots, each of which represents a physical TDM channel and lasts 577 seconds. Every 4.615 ms, each TDM channel occupies the 200 kHz carrier for 577 seconds.

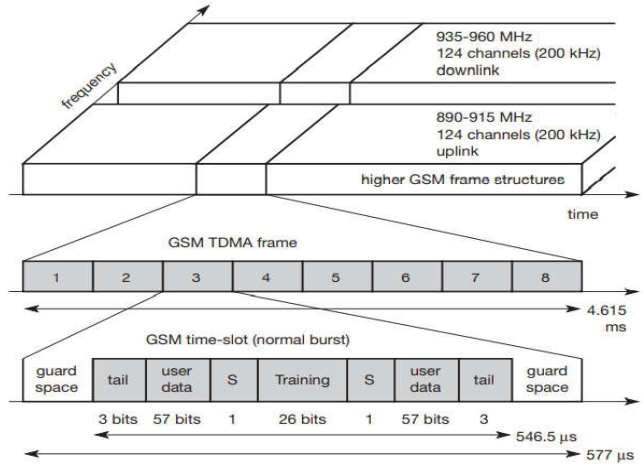


Figure: GSM TDMA frame, slots and bursts

Bursts of data are sent out in short increments. Figure depicts a typical burst utilised for data transmission (user and signalling data) inside a time slot. The burst in the diagram is only 546.5 seconds long and comprises 148 bits. The remaining 30.5 seconds are used as guard space to prevent bursts from overlapping due to varying path delays and to allow the transmitter to turn on and off. Filling the entire slot with data enables for 156.25bit transmission in 577 seconds. Each radio carrier sends roughly 270 kbit/s across the Um interface, and each physical TDM channel has a raw data throughput of about 33.8 kbit/s.

A regular burst's first and last three bits (tail) are all set to 0 and can be exploited to improve receiver performance. The training sequence in the middle of a slot is used to adapt the receiver's parameters to the current path propagation characteristics and, in the case of multi-path propagation, to select the strongest signal. If the data field contains user or network control data, the flag S is set.A frequency correction burst allows the MS to correct the local oscillator to avoid interference with neighbouring channels, a synchronisation burst with an extended training sequence synchronises the MS with the BTS in time, an access burst is used for the initial connection setup between MS and BTS, and finally a dummy burst is used if no data is available for a slot, according to ETSI (1993a).

Simple transmitter hardware is possible due to two factors: first, the uplink and downlink slots of a physical TDM channel are separated in frequency

(45 MHz for GSM 900, 95 MHz for GSM 1800 using FDD). The TDMA frames, on the other hand, are time shifted for three slots, thus if the BTS provides data in slot one on the downlink at time t_0, the MS accesses slot one on the uplink at time t_0+3.577 microsecond. A full-duplex transmitter is unnecessary for an MS; a basic half-duplex transmitter that switches between receiving and sending is sufficient.

GSM specifies an optional slow frequency hopping strategy to mitigate frequency selective fading. Based on a shared hopping sequence, MS and BTS may change the carrier frequency after each frame. Between uplink and downlink slots, an MS adjusts its frequency.

7.5 PROTOCOLS

GSM protocol architecture including signalling protocols and interfaces is depicted in the diagram below. The U_m interface is of particular significance because the other interfaces are between entities in a fixed network. All radio-specific functions are handled by Layer 1, the physical layer. This involves creating bursts in one of five possible formats, multiplexing bursts into a TDMA frame, synchronisation with the BTS, detection of idle channels, and downlink channel quality measurement. The physical layer at Um uses GMSK for digital modulation and provides data encryption and decryption, but only between MS and BSS over the air interface, rather than end-to-end.

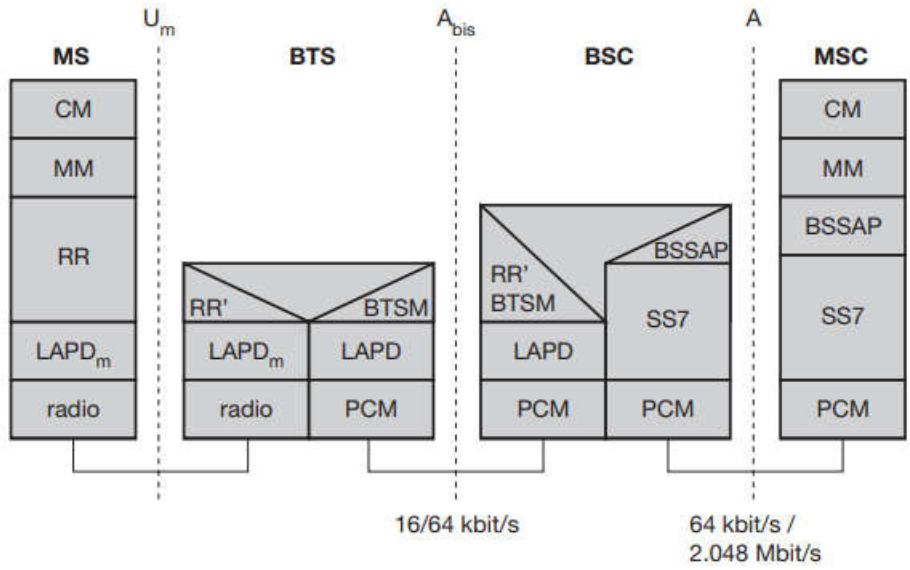


Figure: Protocol architecture for signalling

Individual route delays between an MS and the BTS are also corrected during synchronisation. Because all MSs in a cell use the same BTS, they must all be synced to it. The BTS creates the temporal structure of frames, slots, and other elements. The differing round trip times is a problem in this situation (RTT). The RTT of an MS close to the BTS is quite short, whereas an MS 35 kilometres away already has an RTT of roughly 0.23 ms. Large guard spaces would be necessary if the MS further away

employed the slot structure without correction, because 0.23 ms is already 40% of the 0.577 ms provided for each slot.As a result, the BTS sends the current RTT to the MS, which adjusts its access time to ensure that all bursts arrive to the BTS within their time limitations. This technology cuts the guard time in half, from 30.5 seconds to 5%. The variable timing advance can be used to control access, allowing a burst to be pushed up to 63 bits earlier, with each bit lasting 3.69 microseconds (which results in 0.23ms needed). A burst cannot be shifted sooner than 63-bit times because the variable timing advance cannot be extended.As a result, the maximum distance between an MS and a BTS is 35 kilometres. It may be possible to receive the signals at larger distances; however, access to the BTS cannot be authorised to avoid collisions.

The physical layer's key responsibilities include channel coding and error detection/correction, which are closely related to the coding techniques.Different forward error correction (FEC) techniques are used extensively in channel coding. FEC adds redundancy to user data, making it possible to discover and remedy certain problems. The degree of redundancy, coding technique, and further interleaving of data to mitigate the impacts of burst mistakes determine the strength of an FEC scheme. The FEC is also the reason why, contrary to the ISO/OSI reference model, error detection and correction occurs in layer one rather than layer two. The GSM physical layer attempts to fix mistakes but does not send incorrect data to the higher layers.

GSM uses several coding systems with different correction capacities for different logical channels. To obtain a data rate of 22.8 kbit/s (using the 13 kbit/s from the voice codec plus redundancy, CRC bits, and interleaving), speech channels require further coding of voice data following analogue to digital conversion. Because speech was anticipated to be the primary service in GSM, the physical layer includes features like voice activity detection (VAD), which transmits voice data only when a voice signal is present.This approach helps to reduce interference because a channel may be silent around 60% of the time (assuming only one person speaks at a time and some extra time is required to move between speakers). The physical layer provides a comfort noise to simulate a connection (total silence would likely mislead a user), but no actual transmission takes occur during times of silence (e.g., if a user needs time to ponder before talking). The noise is even customised to the communication partner's existing background noise.

All of this data interleaving for a channel to reduce interference due to burst mistakes and the logical channel's recurrence pattern results in a transmission delay. A TCH/FS has a 60 ms delay and a TCH/F9.6 has a 100 ms delay (within 100 ms signals in fixed networks easily travel around the globe). If communicating with an MS instead of a regular fixed station (telephone, computer, etc.), these periods must be added to the transmission delay, and they may affect the performance of any upper layer protocols, such as computer data transmission.

7.6 LOCALIZATION AND CALLING

The automatic, worldwide localization of users is a key aspect of the GSM system. The system always knows where a user is, and the same phone number can be used anywhere in the world. Even if a user does not use the mobile station, GSM performs periodic location updates to provide this service (assuming the MS is still logged into the GSM network and is not totally switched off). The current location (just the location area, not the specific geographical location) is always stored in the HLR, and the VLR in charge of the MS tells the HLR when the location changes.The HLR transfers all user data to the new VLR as soon as an MS enters the range of a new VLR (a new location region). Roaming is the process of switching VLRs while maintaining continuous availability of all services. Roaming can take place within a single provider's network, between two providers in the same country (national roaming is frequently not supported owing to operator rivalry), or between different carriers in other countries (international roaming). People usually identify the term roaming with international roaming because it is this form of roaming that makes GSM so appealing: one device, 190 countries!

Several numbers are required to locate and address an MS:

- **Mobile Station International ISDN Number (MSISDN):** For a GSM customer, the phone number is the most important number. Remember that the phone number is not linked to a specific device, but rather to the SIM, which is unique to each user. For addresses, the MSISDN uses the ITU-T standard E.164, which is also used in fixed ISDN networks. The country code (CC) (e.g., +49 179 1234567 with 49 for Germany), the national destination code (NDC) (i.e., the network provider's address, e.g., 179), and the subscriber number make up this number (SN).

- **International mobile subscriber identity (IMSI):** GSM employs the IMSI to identify subscribers internally. A mobile country code (MCC) (e.g., 240 for Sweden, 208 for France), a mobile network code (MNC) (i.e., the network provider's code), and lastly a mobile subscriber identity number (MSIN) make up an IMSI (MSIN).

- **Temporary mobile subscriber identity (TMSI):** GSM employs the 4 byte TMSI for local subscriber identification to disguise the IMSI, which would reveal the actual identity of the user signalling over the air interface. TMSI is chosen by the current VLR and is only valid for a limited time and within the VLR's location region (for ongoing communication, TMSI and LAI are sufficient; the IMSI is not required). A VLR may also modify the TMSI on a regular basis

- **Mobile station roaming number (MSRN):** MSRN is another temporary address that hides a subscriber's identity and location. This address is generated by the VLR in response to a request from the MSC, and it is also kept in the HLR. The current visitor country code

(VCC), the visitor national destination code (VNDC), the current MSC's identifier, and the subscriber number are all contained in the MSRN. The MSRN assists the HLR in locating an incoming call subscriber.

All of these numbers are required to locate a subscriber and maintain a mobile station connection. The mobile terminated call (MTC) is an intriguing circumstance in which a station calls another mobile station (which could be outside the GSM network or another mobile station). The essential processes for connecting the calling station to the mobile user are shown in the diagram. In step one, a user phones a GSM subscriber's phone number. The fixed network (PSTN) recognises that the number belongs to a GSM network user (based on the destination code) and forwards the call setup to the Gateway MSC.The GMSC locates the subscriber's HLR (which is encoded in the phone number) and informs it of the call setup. The HLR now verifies that the number is valid and that the user has subscribed to the requested services before requesting an MSRN from the current VLR. The HLR can determine the MSC responsible for the MS after receiving the MSRN and sends this information to the GMSC. The call setup request can now be forwarded to the MSC defined by the GMSC.

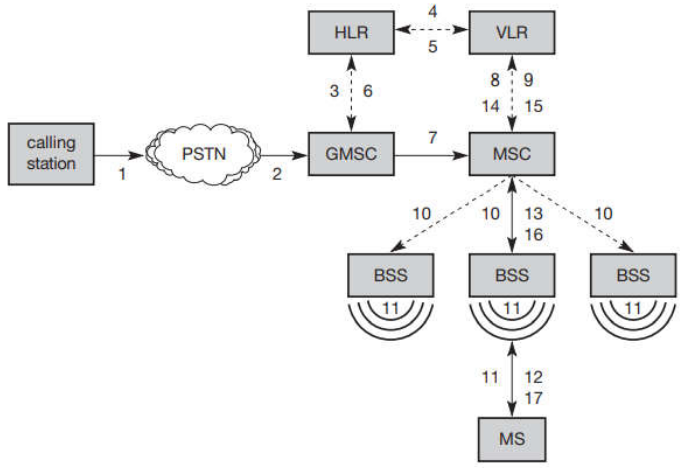


Figure: Mobile Terminated Call (MTC)

The MSC is in charge of all subsequent steps from this point forward. It starts by asking the VLR for the MS's current state. If the MS is available, the MSC commences paging in all cells it is responsible for (i.e., the location area, LA), because searching for the correct cell would take too long (although this strategy puts some strain on the signalling channels, therefore optimizations exist). This paging signal is sent to the MS by all BSS BTSs. If the MS responds, the VLR must do security checks (set up encryption etc.). The VLR then instructs the MSC to establish a link with the MS (steps 15 to 17).

When compared to an MTC, making a mobile originating call (MOC) is more easier (Referfigure below). The MS sends a request for a new connection to the BSS, which passes it on to the MSC. The MSC then

determines whether or not this user is authorised to make a call using the desired service, as well as the availability of resources across the GSM network and into the PSTN. The MSC establishes a connection between the MS and the fixed network if all resources are available.

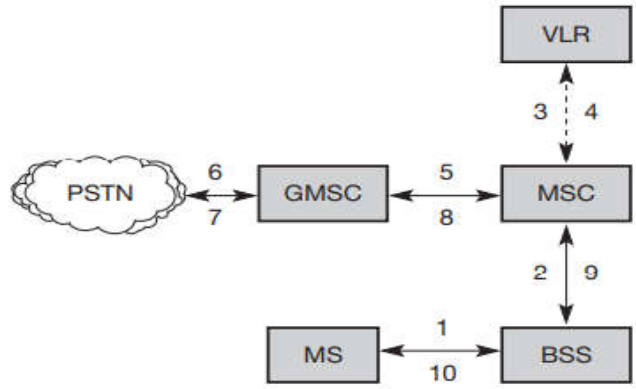


Figure: Mobile originated call (MOC)

Other messages are exchanged between an MS and a BTS during connection establishment in addition to those stated above (in either direction). Before the phone rings, these messages can be heard as crackling noise on radios or poorly insulated loudspeakers. The messages for an MTC and MOC are shown in the diagram below. Paging is only required for an MTC, after which comparable message exchanges take place. The channel access via the randomaccess channel (RACH) with sequential channel assignment is the initial stage in this context; the channel assigned could be a traffic channel (TCH) or a slower signalling channel SDCCH.

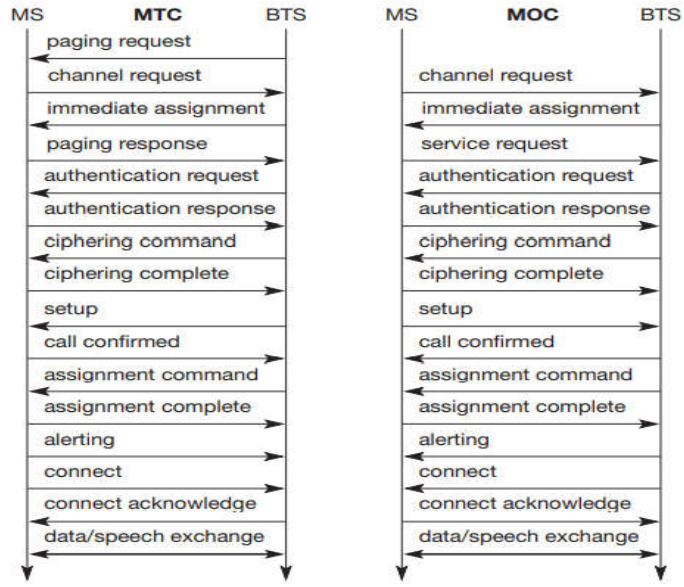


Figure: Message flow for MTC and MOC

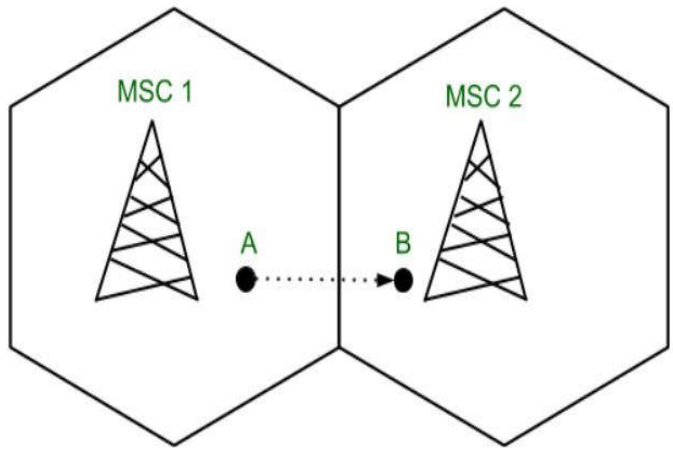
The authentication of the MS and the switch to encrypted connection are the following stages required for communication security.TCH is now assigned by the system (if this has not been done). This has the advantage of requiring an SDCCH only during the initial setup phases. No TCH has been blocked if the setup fails. Using a TCH from the start, on the other hand, has a speed advantage.

The actions that follow are dependent on whether you're using MTC or MOC. If someone calls the MS, it now responds with 'alerting,' indicating that the MS is ringing, and 'connect,' indicating that the user has clicked the connect button. If the MS has initiated the call, the identical procedures occur in reverse. Both parties can share data after the connection is acknowledged.

The connection is closed by sending a user-initiated disconnect message (on both sides), then relinquishing the connection and the radio channel.

7.7 HANDOVER

Handover or handoff refers to the procedure of transferring ongoing call or data connectivity from one Base Station to another in cellular telecommunications. When a phone goes to a different cell while a call is in progress, the MSC (Mobile Switching Center) transfers the call to a new channel associated with the new Base Station.



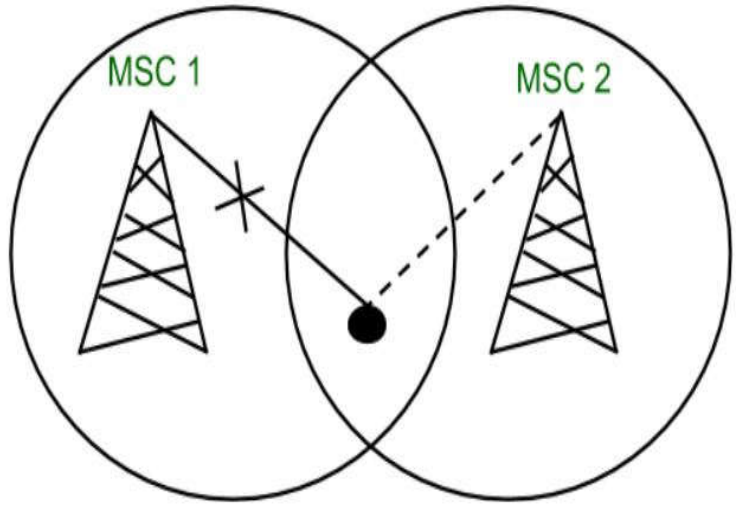
When a mobile user A moves from one cell to another, the signal strength of BSC 1 decreases while the signal strength of BSC 2 improves, allowing the mobile user to continue making calls or accessing data without interruption.

7.7.1 Types of Handoff

1] Hard handoff

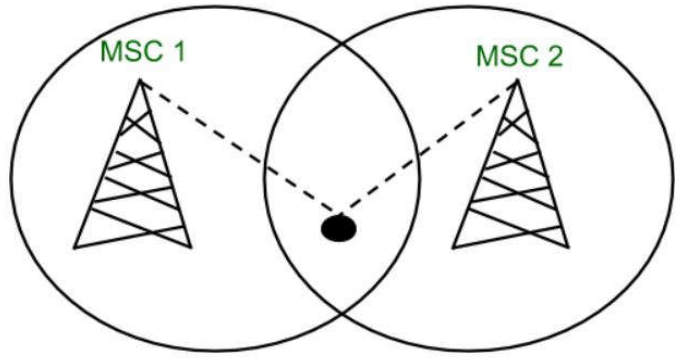
When transitioning from one Base Station to another Base Station, there is an actual interruption in connectivity. The Base Station and MSC are not burdened since the changeover occurs so swiftly that the users are barely

aware of it. The quality of the connection is poor. The 'break before make' policy was implemented by Hard Handoff.



2] Soft handoff

When radio signals are added or removed from the Base Station, at least one of the links is retained in Soft Handoff. The'make before break' principle was implemented by Soft Handoff. Hard Handoff is more expensive than Soft Handoff.



7.7.2 Situations for triggering handoff

Handoffs occur in any of the following circumstances:

- When a subscriber in a call or data session travels out of one cell's coverage region and into another cell's coverage area, a handoff is triggered to ensure service continuity. The duties that the first cell was performing are now being delegated to the second cell.
- Each cell has a specified capacity, which means it can only serve a certain number of subscribers. A handoff occurs when the number of users using a particular cell hits its maximum capacity. If the subscriber is within the overlapping service area of both cells, some calls are moved to nearby cells.
- Microcells are often subdivided from larger cells. When responsibilities are transferred from the large cell to the smaller cell

and vice versa, a handoff may occur. A travelling user, for example, is moving inside the jurisdiction of a huge cell. If the traveller comes to a halt, the jurisdiction is moved to a microcell to relieve the large cell's load.

- Handoffs can also happen when multiple calls using the same frequency for communication collide.

7.7.3 Mobile assisted handoff

Mobile assisted handoff(MAHO) is a mechanism in which mobile devices assist the Base Station Controller (BSC) in transferring a call to a different BSC. GSM cellular networks employ it. A handoff in other systems, such as AMPS, is completely the responsibility of the BSC and the Mobile Switching Centre (MSC), with no involvement from the mobile device. When a mobile station in GSM is not using its time slots for communication, it measures signal quality and communicates that information to the BSC. This information is used by the BSC to complete handoff.

7.8 SECURITY

GSM provides a variety of security services based on information saved in the AuC and individual SIMs (which is plugged into an arbitrary MS). The SIM card stores personal and confidential information and is secured with a PIN to prevent unwanted access.(The secret key Ki, for example, is saved in the SIM and is used for authentication andencryption procedures.) GSM's security services are described in detail below:

- **Access control and authentication:** The first step is to verify that the SIM user is legitimate. To utilise the SIM, the user must enter a secret PIN.The subscriber authentication is the next stage. A challenge-response method is used in this step.

- **Confidentiality:** All user-related data is encrypted for privacy. Following authentication, the BTS and MS encrypt speech, data, and signalling. This level of confidentiality occurs just between MS and BTS, not from end to end or throughout the entire fixed GSM/telephone network.

- **Anonymity:** All data is encrypted before transmission to ensure user anonymity, and user identifiers (which might indicate an identity) are not used over the air. Instead, GSM sends out a temporary identification (TMSI) that the VLR assigns after each position update. The TMSI can also be changed by the VLR at any moment.

In order to provide security services in GSM, three algorithms have been specified.A3 is used for authentication, A5 is used for encryption, and A8 is used to generate a cypher key. Only algorithm A5 was publicly disclosed in the GSM standard, but algorithms A3 and A8 were kept secret but standardised with open interfaces. Both A3 and A8 are no longer classified, having been made public in 1998 via the internet. This proves

that security by obscurity is ineffective.The algorithms, it turned out, aren't very good. Users can employ greater end-to-end encryption or network providers can utilise stronger authentication techniques. Algorithms A3 and A8 (or their equivalents) are proprietary and can be found on the SIM and in the AuC.Only A5 must be identical across all providers in terms of device implementation.

4.8.1 Authentication

A subscriber must be authenticated before he or she may utilise any GSM network service. The SIM, which stores the individual authentication key Ki, the user identity IMSI, and the authentication algorithm A3, is used for authentication. The challenge-response approach is used for authentication: the access control AC generates a random number RAND as a challenge, and the SIM within the MS responds with SRES (signed response) (Refer following Figure). For each IMSI, the AuC generates the basic random values RAND, signed answers SRES, and cypher keys Kc, then sends this information to the HLR. The present VLR asks the HLR for the necessary RAND, SRES, and Kc values.

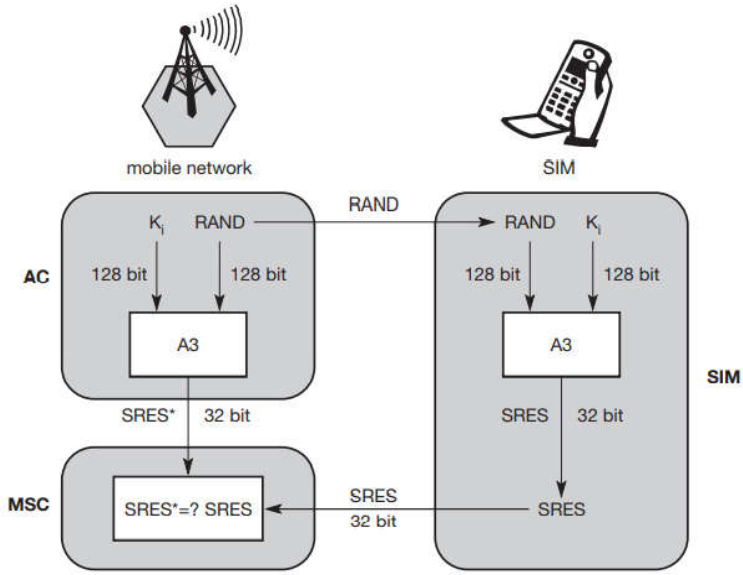


Figure: Subscriber authentication

The VLR delivers the SIM the random value RAND for authentication.With RAND and the key Ki, dubbed A3, both the network and subscriber modules conduct the identical function. The MS returns the SIM's SRES, allowing the VLR to compare the two values. The VLR approves the subscriber if they are the same; otherwise, the subscriber is rejected.

4.8.2 Encryption

All transmissions including user-related information are encrypted in GSM over the air interface to preserve privacy. MS and BSS can start employing encryption after authentication by applying the cypher key K_c (the actual position of security functions for encryption, BTS and/or BSC,

is vendor dependant). The algorithm A8 is used to produce K_c from the individual key K_i and a random value. It's worth noting that the SIM in the MS and the network both use the same random number RAND to calculate K_c. Over the air interface, the key K_c is not communicated.

The algorithm A5 and the cypher key K_c can now be used by MS and BTS to encrypt and decode data. As seen in the diagram below, K_c should be a 64-bit key, which isn't particularly strong but provides adequate security against simple eavesdropping.However, the internet publishing of A3 and A8 revealed that in some implementations, 10 of the 64 bits are always set to 0, resulting in a key length of only 54 bits. As a result, the encryption is significantly less secure.

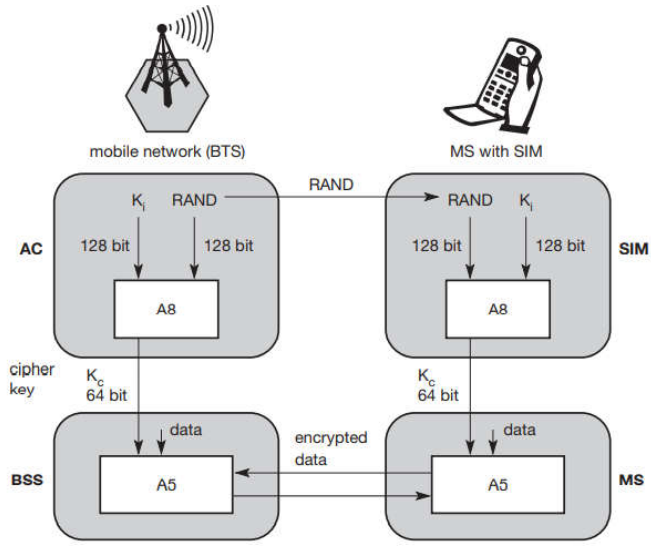


Figure: Data Encryption

7.9 DECT: SYSTEM ARCHITECTURE

Depending on its intended function, a DECT system may have a variety of distinct physical implementations. Different DECT entities can be spread, replicated, etc., and merged into a single physical unit. On the same logical reference model of the system architecture, as depicted in followingfigure, all implementations are based.

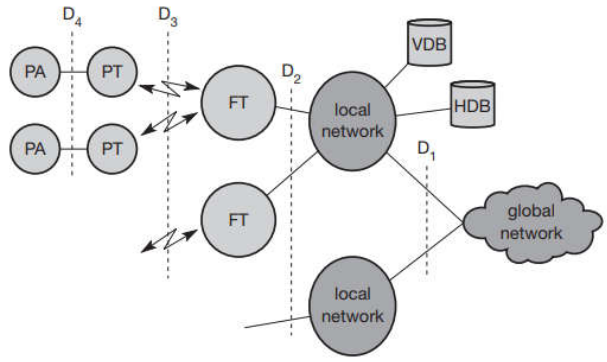


Figure: DECT system architecture reference model

The local communication system is linked to the outside world via a global network, which provides its services over the D1 interface. Public switched telephone networks (PSTN), public land mobile networks (PLMN), such as GSM, or packet switched public data networks are examples of global networks. (PSPDN). These networks provide a variety of services, such as data transportation, address translation, and data routing between local networks.

In the DECT environment, local networks provide local telecommunication services that can range from straightforward switching to clever call forwarding, address translation, etc. Such networks include analogue or digital private branch exchanges (PBXs) or LANs, such as those that adhere to the IEEE 802.x family of LAN standards.All normal network tasks must be integrated in the local or global network, where the databases home data base (HDB) and visitor data base (VDB) are also located, despite the DECT system's relatively simple core. With features that are comparable to those in the HLR and VLR in GSM systems, both databases facilitate mobility. Incoming calls are automatically routed to the DECT user's current subsystem, and the current VDB notifies the HDB of any location changes.

The fixed radio termination (FT) and portable radio termination (PT) make up the DECT core network, which essentially just offers a multiplexing service. At the fixed network side and mobile network side, respectively, FT and PT cover layers one through three. Additionally, a device may support a number of portable applications (PA).

7.10 PROTOCOL ARCHITECTURE

The OSI reference model is followed by the DECT protocol reference architecture. The physical layer, medium access control, and data link control8 for both the control plane (C-Plane) and the user plane are represented in the following diagram as the layers covered by the standard. (U-Plane). User data from layer two is transferred directly to the U-Plane thanks to the specification of an additional network layer for the C-Plane.All lower layers of a DECT system are vertically covered by a management plane.

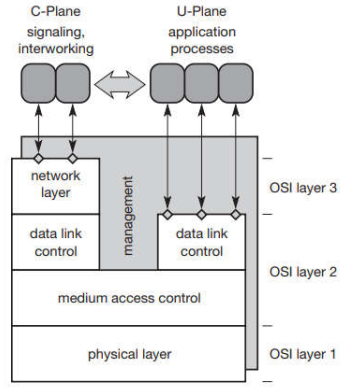


Figure: DECT protocol layers

7.10.1 Physical layer

The physical layer of every wireless network includes all operations for modulation and demodulation, incoming signal detection, sender/receiver synchronization, and gathering status data for the management plane. The physical channel structure is generated by this layer with a predetermined, guaranteed throughput. The physical layer assigns a channel for data transmission in response to a request from the MAC layer.

7.10.2 Medium access control layer

By activating and deactivating physical channels, the media access control (MAC) layer creates, maintains, and releases channels for higher layers. Multiple logical channels are multiplexed onto physical channels using MAC. There are logical channels for broadcast messages, user data transmission, paging, and signalling network control. The segmentation and reassembly of packets as well as error control and error correction are additional services provided.

7.10.3 Data link control layer

The data link control (DLC) layer creates and maintains reliable connections between the mobile terminal and the base station. Two services have been defined for the C-Plane: a connectionless broadcast service for paging (called Lb) and a point-to-point protocol similar to LAPD in ISDN, but adapted to the underlying MAC (called LAPC+Lc). Several services exist for the U-Plane, e.g., a transparent unprotected service (basically a null service), a forward error correction service, rate adaptation services, and services for future enhancements. If services are used, e.g., to transfer ISDN data at 64 kbit/s, then DECT also tries to transfer 64 kbit/s. However, in case of errors, DECT raises the transfer rate to 72 kbit/s, and includes FEC and a buffer for up to eight blocks to perform ARQ. This buffer then introduces an additional delay of up to 80 ms.

7.10.4 Network layer

DECT's network layer, which only exists for the C-Plane, is comparable to those in ISDN and GSM. This layer offers services for resource requests, checks, reservations, controls, and releases at fixed stations (wireless connections to fixed networks) and mobile terminals (wireless connection). Identity management, authentication, and location database management are all handled by the mobility management (MM) component of the network layer. Setup, release, and negotiation of connections are handled by call control (CC). The interworking unit that links the DECT system with the outside world is connected by two message services, the connectionless message service (CLMS) and the connectionless message service (COMS).

7.11 TETRA

Another means of transmitting wireless data are truncated radio systems.These systems employ a wide range of radio carriers, but they only ever temporarily pair a user with a particular carrier based on demand. While traditional systems require organisations like taxi services, transportation providers with fleet management systems, and rescue teams to each have their own distinct carrier frequency, trunked radio systems allow these organisations to share a large number of frequencies for improved frequency reuse using FDM and TDM approaches. Although they are not accessible to the general public, these radio systems frequently provide interfaces to the fixed telephone network, including voice and data services.These networks are not only more straightforward than the majority of other networks, but also more dependable and reasonably priced to set up and run since they just need to service the local users' operating areas, such as a city taxi service.

ETSI standardised the TETRA system (terrestrial trunked radio)9 in 1991 to enable a uniform system across Europe (ETSI, 2002; TETRA MoU, 2002). This system should take the place of regional systems that commonly connect to an X.25 packet network, such as MODACOM, MOBITEX, and COGNITO in Europe.TETRA offers two standards: the Voice+Data (V+D) service (ETSI, 1998l) and the packet data optimized (PDO) service (ETSI, 1998m). While V+D offers circuit-switched voice and data transmission, PDO only offers packet data transmission, either connection-oriented to connect to X.25 or connectionless for the ISO CLNS (connectionless network service). The latter service can be point-to-point or point-to-multipoint, the typical delay for a short message (128 byte) being less than 100 ms. V+D connection modes comprise unicast and broadcast connections, group communication within a certain protected group, and a direct ad hoc mode without a base station. However, delays for short messages can be up to 500 ms or higher depending on the priority.

TETRA also offers bearer services of up to 28.8 kbit/s for unprotected data transmission and 9.6 kbit/s for protected transmission. Examples for end-to-end services are call forwarding, call barring, identification, call hold, call priorities, emergency calls and group joins. The system architecture of TETRA is very similar to GSM. Via the radio interface Um, the mobile station (MS) connects to the switching and management infrastructure (SwMI), which contains the user data bases (HDB, VDB), the base station, and interfaces to PSTN, ISDN, or PDN. The system itself, however, is much simpler in real implementation compared to GSM, as typically no handover is needed. Taxis usually remain within a certain area which can be covered by one TETRA cell. Several frequencies have been specified for TETRA which uses FDD (e.g., 380–390 MHz uplink/390–400 MHz downlink, 410–420 MHz uplink/420–430 MHz downlink). Each channel has a bandwidth of 25 kHz and can carry 36 kbit/s. Modulation is DQPSK. While V+D uses up to four TDMA voice or data channels per carrier, PDO performs statistical multiplexing. For accessing a channel, slotted Aloha is used.

7.12 UMTS AND IMT- 2000

Figure below depicts a simplified UMTS reference design that applies to both UTRA and non-UTRA solutions (3GPP, 2000). The UTRA network (UTRAN) is a radio network subsystem that manages cell level mobility (RNS). Radio channel ciphering and deciphering, handover control, radio resource management, and other services are all performed by the RNS. The radio interface U_u (which is similar to the Um interface in GSM) connects the UTRAN to the user equipment (UE). UTRAN interfaces with the core network(CN) via the I_u interface (which is comparable to the A interface in GSM). If there is no dedicated connection between the UE and the UTRAN, the CN contains capabilities for inter-system handover, gateways to other networks (fixed or wireless), and location management.

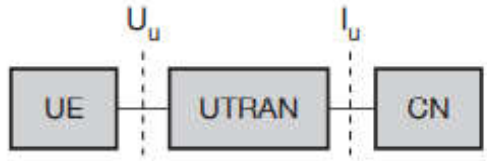


Figure: Main components of UMTS system architecture

The following basic architecture is further subdivided into domains by UMTS (Refer figure below). A single user is assigned to the user equipment domain, which contains all of the functionalities required to access UMTS services. The USIM domain and the mobile equipment domain are both contained under this domain. The USIM domain contains the UMTS SIM, which handles encryption and authentication operations for users and maintains all user-related data for UMTS. This USIM is usually associated with a service provider and includes a microprocessor for a better programme execution environment (USAT, UMTS SIM application toolkit). The end device is classified as mobile equipment. All radio transmission functions, as well as user interfaces, are housed here.

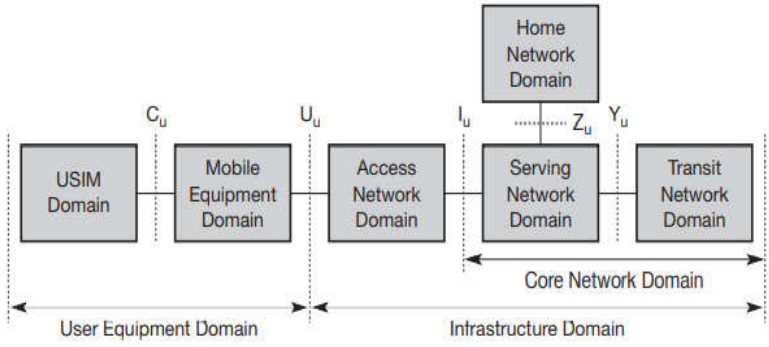


Figure: UMTS domain and interfaces

All users share the infrastructure domain, which provides UMTS services to all accepted users. The access network domain, which houses radio access networks (RANs), and the core network domain, which houses access network independent operations, make up this domain. The core

network domain can be divided into three distinct domains, each with its own set of tasks. All functions currently employed by a user to access UMTS services are included in the serving network domain. The home network domain encompasses all functions connected to a user's home network, such as user data look-up. Finally, the transit network domain may be required if the serving network is unable to communicate directly with the home network.The core network's three domains could all be the same physical network. These domains are merely functional descriptions.

7.13 SATELLITE SYSTEMS: HISTORY, APPLICATIONS

- **HISTORY**

After World War II, satellite communications were first developed. The ability to create rockets that could launch radio transmitters into orbit was known to scientists. Arthur C. Clarke's essay on "Extra Terrestrial Relays" was published in 1945. However, it wasn't until 1957, in the midst of the Cold War, when the Soviet Union abruptly launched the first satellite SPUTNIK, shocking the Western world.SPUTNIK was essentially a small sender that transmitted a sporadic "beep," which is in no way analogous to a satellite today. However, this was sufficient for the US to devote all of its resources to creating its first satellite. The first reflecting communication satellite ECHO was launched into space in about three years, in 1960.ECHO was in space. ECHO was essentially a mirror in the sky that reflected signals to allow for communication. The first geostationary (or geosynchronous) satellite, SYNCOM, was launched three years later. Geostationary satellites still serve as the foundation for aerial news broadcasts today. Their permanent position in the sky is their main advantage. They appear to be anchored to a specific point because their spin is synchronised with the earth's rotation.

The first commercial geostationary communication satellite INTELSAT 1 (also known as ‘Early Bird’) went into operation in 1965. It was in service for one-and-a-half years, weighed 68 kg and offered 240 duplex telephone channels or, alternatively, a single TV channel. INTELSAT 2 followed in 1967, INTELSAT 3 in 1969 already offered 1,200 telephone channels. While communication on land always provides the alternative of using wires, this is not the case for ships at sea. Three MARISAT satellites went into operation in 1976 which offered worldwide maritime communication. However, Sender and receiver still had to be installed on the ships with large antennas (1.2 m antenna, 40 W transmit power). The first mobile satellite telephone system, INMARSAT-A, was introduced in 1982. Six years later, INMARSAT-C became the first satellite system to offer mobile phone and data services. (Data rates of about 600 bit/s, interfaces to the X.25 packet data network exist.) In 1993, satellite telephone systems finally became fully digital with INMARSAT-M. The actual mobility, however, was relative from a user’s point of view, as the devices needed for communication via geostationary satellites were heavy (several kilograms) and needed a lot of transmit power to achieve decent data rates.

Nineteen ninety-eight marked the beginning of a new age of satellite data communication with the introduction of global satellite systems for small mobile phones, such as, e.g., Iridium and Globalstar. There are currently almost 200 geostationary satellites in commercial use which shows the impressive growth of satellite communication over the last 30 years. However, satellite networks are currently facing heavy competition from terrestrial networks with nationwide coverage or at least enough coverage to support most applications and users.

- APPLICATIONS

  Satellites have historically been applied in the following fields:

- Weather prediction: Several satellites send images of the world using infrared or visible light, for example. It would be impossible to forecast hurricanes without the assistance of satellites.
- Radio and TV broadcast satellites: Satellites used for radio and TV broadcasts make thousands of radio and television programmes accessible. Since it costs less to install and typically requires no additional fees, this technology competes with cable in many locations. In central Europe, modern satellite dishes have diameters of 30 to 40 cm. (the diameters in northern countries are slightly larger).
- Military satellites: One of the earliest applications of satellites was their use for carrying out espionage. Many communication links are managed via satellite because they are much safer from attack by enemies.
- Satellites for navigation: The global positioning system (GPS), which was once primarily utilised for military purposes, is today well-known and accessible to everybody. Worldwide exact localization is possible with the technology, and with some additional techniques, the precision can reach a few metres. The majority of ships and aircraft use GPS in addition to more conventional navigation systems. GPS receivers are commonly seen in trucks and autos. This technology is also employed, for instance, to manage the truck fleet or to locate stolen vehicles.

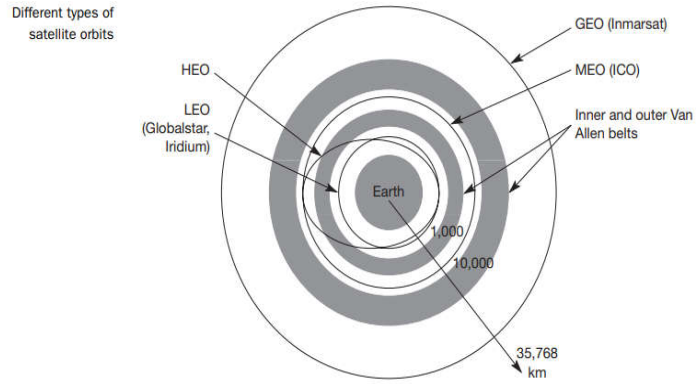
7.14 BASICS: GEO, LEO, MEO

- **Geostationary (or geosynchronous) earth orbit (GEO):** GEO satellites have a distance of almost 36,000 km to the earth. Examples are almost all TV and radio broadcast satellites, many weather satellites and satellites operating as backbones for the telephone network

- **Medium earth orbit (MEO):** MEOs operate at a distance of about 5,000–12,000 km. Up to now there have not been many satellites in this class, but some upcoming systems (e.g., ICO) use this class for various reasons

- **Low earth orbit (LEO):** While some time ago LEO satellites were mainly used for espionage, several of the new satellite systems now rely on this class using altitudes of 500–1,500 km

- **Highly elliptical orbit (HEO):** This class comprises all satellites with noncircular orbits. Currently, only a few commercial communication systems using satellites with elliptical orbits are planned. These systems have their perigee over large cities to improve communication quality.



7.15 SUMMARY

GSM has been shown as the most successful second generation digital cellular network for the most part in this chapter. Although GSM was originally developed for voice communication, the chapter demonstrated how allows for more data-oriented transmission. This evolution comprises the move from a circuit-switched network to a packet-switched system that is more similar to the internet architecture.Other systems presented include DECT, thedigital standard for cordless phones, and TETRA, a trunked radio system. DECT can be used for wireless data transmission on a campus or indoors, but also for wireless local loops (WLL). For special scenarios, e.g., emergencies, trunked radio systems such as TETRA can be the best choice. They offer a fast connection setup (even within communication groups) and can work in an ad hoc network, i.e., without a base station. This chapter also presented an overview of current and future third generation systems. UMTS, a proposal of operators and companies involved in the GSM business, was discussed in more detail.

7.16 LIST OF REFERENCES

1) Protocols and Architectures for Wireless Sensor Network, Holger Kerl, Andreas Willig, John Wiley and Sons, 2005

2) Wireless Sensor Networks Technology, Protocols, and Applications,Kazem Sohraby, Daniel Minoli and TaiebZnati, John Wiley & Sons, 2007

3) Mobile communications, Jochen Schiller,2nd Edition, Addison wisely, Pearson Education,2012

4) Fundamentals of Wireless Sensor Networks, Theory and Practice, Waltenegus Dargie, Christian Poellabauer, Wiley Series on wireless Communication and Mobile Computing, 2011

5) Networking Wireless Sensors, Bhaskar Krishnamachari, Cambridge University Press, 2005

7.17 UNIT END EXERCISES

1) State Mobile services.
2) Explain GSM System architecture.
3) Write a note on Radio interface.
4) Illustrate different Protocols.
5) Explain Localization and Calling.
6) Write a note on Handover.
7) Describe different Security.
8) Explain DECT: System architecture.
9) Explain the protocol architecture.
10) Write a note on TETRA.
11) Explain UMTS and IMT- 2000.
12) Describe: Satellite Systems: History, Applications
13) Write a basic note on: GEO, LEO, MEO