M.Sc. (C. S.) PART II

SEMESTER - III (CBCS)

UBIQUITOUS COMPUTING

SUBJECT CODE: PSCS 301

Programme Co-ordinator : **Shri Mandar Bhanushe**
Head, Faculty of Science and Technology IDOL,
Univeristy of Mumbai - 400098

Course Co-ordinator : **Mr Sumedh Shejole**
Assistant Professor,
IDOL, University of Mumbai - 400098

Editor : **Ms. Prachita Sawant**
Assistant Professor,
IDOL, University of Mumbai - 400098

Course Writers : **Mr Santosh Patilshirke**
Smt. Janakibai Rama Salvi
Smt. Janakibai Rama Salvi College,
Kalwa (W), Thane.

: **Ms Arti Bansode**
SIES(Nerul) College of Arts,
Science and Commerce
Nerul, Navi Mumbai - 400706

August 2022, Print - 1

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

DTP composed and Printed by: Mumbai University Press

CONTENTS

M.Sc. (C. S.) Part II

SEMESTER - IIi (CBCS)

UBIQUITOUS COMPUTING

SYLLABUS

Course Code	Course Title	Credits
PSCS301	**Ubiquitous Computing**	**04**
Unit I: Basics of **Ubiquitous Computing** Examples of Ubiquitous Computing Applications, Holistic Framework for UbiCom: Smart DEI, Modeling the Key Ubiquitous Computing Properties, Ubiquitous System Environment Interaction, Architectural Design for UbiCom Systems: Smart DEI Model, Smart Devices and Services, Service Architecture Models, Service Provision Life Cycle.		
Unit II: Smart Mobiles, Cards and Device Networks Smart Mobile Devices, Users, Resources and Code, Operating Systems for Mobile Computers and Communicator Devices, Smart Card Devices, Device Networks. Human–Computer Interaction (HCI): Explicit HCI, Implicit HCI, User Interfaces and Interaction for Devices, Hidden UI Via Basic Smart Devices, Hidden UI Via Wearable and Implanted Devices, Human Centered Design (HCD).		
Unit III: Smart Environments Tagging, Sensing and Controlling, Tagging the Physical World, Sensors and Sensor Networks, Micro Actuation and Sensing: MEMS, Embedded Systems and Real Time Systems, Control Systems.		
Unit IV: Ubiquitous Communication Audio Networks, Data Networks, Wireless Data Networks, Universal and Transparent Audio, Video and Alphanumeric Data Network Access, Ubiquitous Networks, Network Design Issues.		

Text book: • Ubiquitous Computing Smart Devices, Environments and Interactions, Stefan Poslad, Wiley, 2009.
References: • Ubiquitous Computing Fundamentals. John Krumm, Chapman & Hall/CRC 2009. • Ambient intelligence, wireless networking and ubiquitous computing, Vasilakos, A., & Pedrycz, W. ArtechHouse, Boston, 2006. • http://www.eecs.qmul.ac.uk/~stefan/ubicom.

1

BASICS OF UBIQUITOUS COMPUTING

Unit Structure

1.1 BASICS OF UBIQUITOUS COMPUTING

Ubiquitous computing is a paradigm in which the processing of information is linked with each activity or object as encountered. It involves connecting electronic devices, including embedding microprocessors to communicate information. Devices that use ubiquitous computing have constant availability and are completely connected.

Ubiquitous computing focuses on learning by removing the complexity of computing and increases efficiency while using computing for different daily activities.

Ubiquitous computing is also known as pervasive computing, everyware and ambient intelligence.

IN OTHER WORDS Ubiquitous computing is a concept in software engineering and computer science where computing is made to appear anytime and everywhere. In contrast to desktop computing, ubiquitous computing can occur using any device, in any location, and in any format.

1.2 EXAMPLES OF UBIQUITOUS COMPUTING APPLICATIONS

1. Home Automation

In this era of computing, we all heard the concept of "Home Automation". In this process, we connect all possible devices to a LAN (Local Area Network) or directly to the internet. We use a parent device (Example: Amazon Alexa, Google Home, Apple Hub) to give the command to those devices. To imply this concept we give computing power to all connected devices. In this situation, we use ubiquitous computing.

2. IP camera

Nowadays consumer electronic devices are being networked. Some high-end cameras already have a wireless network capability and use it to send photos and videos to a nearby display or to upload those data to the internet. *"An Internet Protocol camera, or IP camera, is a type of digital video camera that receives control data and sends image data via the Internet"*.

3. RFID Devices

RFID stands for Radio Frequency Identification. RFID tags are passive of stamps which can communicate with RFID readers. We are using RFID chips massively in Credit or Debit cards to store our financial data in an encrypted format.

4. Smartphones

They are the most obvious and immediate example, since more often than not, one can just download an app and through it control their home automation systems and other ubiquitous applications.

5. Wearables

Wearables like smart watches , smart glasses etc are examples of ubiquitous computing Be it Apple watch or Fitbit, both come under ubiquitous computing.

6. Self-driving vehicles

self-driving cars Also good examples of ubiquitous computing experience. In fact, an autonomous vehicle that can identify its authorized passenger through smartphone proximity, dock and charge itself when needed, and handle toll, emergency response, and fast-food payments itself by communicating with infrastructure would make for an even better instance of ubiquitous computing.

1.3 HOLISTIC FRAMEWORK FOR UBICOM: SMART DEI

Three approaches to analyse and design UbiCom Systems to form a holistic framework for ubiquitous computing are proposed called the smart DEI framework based upon:

1. Design architectures to apply UbiCom systems: Three main types of design for UbiCom systems are proposed: smart device, smart environment and smart interaction.

2. An internal model of the UbiCom system properties based upon five fundamental properties: distributed, iHCI, context-awareness, autonomy, and artificial intelligence. There are many possible sub-types of ubiquitous system design depending on the degree to which these five properties are supported and interlinked.

3. A model of UbiCom system's interaction with its external environments. In addition to a conventional distributed ICT system device interaction within a virtual8 environment (C2C), two other types of interaction are highlighted: (a) between computer systems and humans as systems (HCI); (b) between computers and the physical world (CPI).

Smart DEI stands for the Smart Devices, Environments and Interactions model. It is pronounced 'Smart Day' in order to allude to the fact that the model focuses on the use of systems support for daily activities

The Smart DEI model represents a holistic framework to build diverse UbiCom systems based on smart devices, smart environments and smart interaction. These three types of design can also be combined to support different types of smart spaces, e.g., smart mobile devices may combine an awareness of their changing physical environment location in order to optimise the routing of physical assets or the computer environment from a different location. Each smart device is networked and can exchange data and access information services as a core property.

1.4 MODELING THE KEY UBIQUITOUS COMPUTING PROPERTIES

Ubiquitous computing represents a powerful shift in computation, where people live, work, and play in a seamless computer-enabled environment, interleaved into the world. Ubiquitous computing postulates a world where people are surrounded by computing devices and a computing infrastructure that supports us in everything we do.

Conventional networked computer9 systems10 or Information Communication Technology (ICT) systems consider themselves to be situated in a virtual world or environment of other ICT systems, forming a system of ICT systems. Computer systems behave as distributed computer systems that are interlinked using a communications network. In conventional ICT systems, the role of the physical environment is restricted, for example, the physical environment acts as a conduit for electronic communication and power and provides the physical resources to store data and to execute electronic instructions, supporting a virtual ICT environment. Because of the complexity of distributed computing, systems often project various degrees of transparency for their users and providers in order to hide the complexity of the distributed computing model from users, e.g., anywhere, anytime communication transparency and mobility transparency, so that senders can specify who to send to, what to send rather than where to send it to.

Human–computer interaction (HCI) with ICT systems has conventionally been structured using a few relatively expensive access points. This primarily uses input from keyboard and pointing devices which are fairly obtrusive to interact with.

1.5 UBIQUITOUS SYSTEM ENVIRONMENT INTERACTION

1. At a high level of abstraction, we can distinguish three types of system environment for each particular UbiCom system: (1) other UbiCom systems which form the ICT infrastructure, supporting services and act as middleware for that particular ICT system applications (virtual worlds).
2. human individuals and human organisations.
3. physical world systems including ecological and biologic systems. Together, the virtual (computer) environment humans and the physical world can be considered as forming an external environment for UbiCom systems. Note that each of these three main environments appear to have quite different design models and processes. Physical world phenomena are governed by well-known laws of physical forces such as electromagnetism and gravity. Living entities in the physical world are governed by ecological models of their habitat. Human living entities are often governed by economic and social models.

   A UbiCom system is often organised conventionally as a layered information system stack with a bottom layer of information resources, a middle layer of processing and a top layer of user information abstractions to view and interact with the information. A common communications pipe allows these to be distributed in different ICT systems.

There are three basic types of environment for UbiCom systems:

(1) the infrastructure of other ICT systems.

(2) the physical world environment

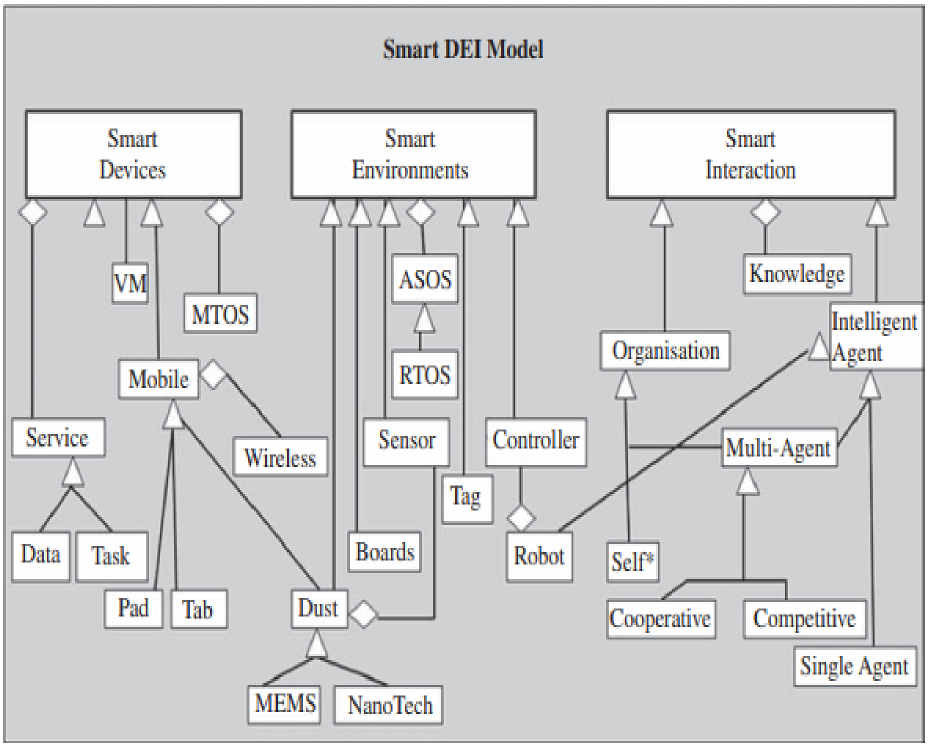
(3) the human environment.

Several basic types of system environment interaction occur: between humans and ICT systems, HCI between ICT systems and the physical world, CPI between ICT systems, C2C or CCI. In addition, interactions can occur between the non-ICT systems such as between different physical world entities and between humans (H2H or HHI), also called social interaction. These types of interaction all coexist.

1.5.1 Architectural Design for UbiCom Systems: Smart DEI Model, Smart Devices and Services

Three basic architectural design patterns for ubiquitous ICT system: smart devices, smart environment and smart interaction are proposed. Here the concept smart simply means that the entity is active, digital, networked, can operate to some extent autonomously, is reconfigurable and has local control of the resources it needs such as energy, data storage, etc. It follows that these three main types of system design may themselves contain sub-systems and components at a lower level of granularity that may also be considered smart, e.g., a smart environment device may consist of smart sensors and a smart controller, etc.

Three basic architectural design patterns for ubiquitous ICT system: smart devices, smart environment and smart interaction are proposed. Here the concept smart simply means that the entity is active, digital, networked, can operate to some extent autonomously, is reconfigurable and has local control of the resources it needs such as energy, data storage, etc. It follows that these three main types of system design may themselves contain sub-systems and components at a lower level of granularity that may also be considered smart, e.g., a smart environment device may consist of smart sensors and a smart controller, etc.



Smart devices

Smart devices, e.g., personal computer, mobile phone, tend to be multi-purpose ICT devices, operating as a single portal to access sets of popular multiple application services that may reside locally on the device or remotely on servers. There is a range of forms for smart devices. Smart devices tend to be personal devices, having a specified owner or user. In the smart device model, the locus of control and user interface reside in the smart device. The main characteristics of smart devices are as follows: mobility, dynamic service discovery and intermittent resource access (concurrency, upgrading, etc.). Devices are often designed to be multi-functional because these ease access to, and simplify the interoperability of, multi-functions at run-time. However, the trade-off is in a decreased openness of the system to maintain (upgrade) hardware components and to support more dynamic flexible run-time interoperability.

Weiser's ICT Device Forms: Tabs, Pads and Boards

Weiser's ICT Device Forms: Tabs, Pads and Boards We often think of computers primarily in terms of the multi-application personal or server computers, as devices with some type of screen display for data output and a keyboard and some sort of pointing devices for data input. As humans, we routinely interact with many more devices that have single embedded computers in them, such as household appliances, and with complex machines that have multiple embedded computers in them. Weiser noted that there was a trend away from many people per computer to one computer per person, through to many computers per person. Computer-based devices tend to become smaller and lighter in weight, cheaper to produce. Thus devices can become prevalent, made more portable and can appear less obtrusive. Weiser considered a range of device sizes in his early work from wearable centimetre-sized devices (tabs), to hand-held decimetre-sized devices (pads) to metre-sized (boards) displays. ICT Pads to enable people to access mobile services and ICT tabs to track goods are in widespread use. Wall displays are useful for viewing by multiple people, for collaborative working and for viewing large complex structures such as maps. Board devices may also be used horizontally as surface computers as well used in a vertical position.

Extended Forms for ICT Devices: Dust, Skin and Clay

The three forms proposed by Weiser (1991) for devices, tabs, pads and boards, are characterised by: being macro-sized, having a planar form and by incorporating visual output displays. If we relax each of these three characteristics, we can expand this range into a much more diverse and potential more useful range of ubiquitous computing devices. First, ICT devices can be miniaturised without visual output displays, e.g., Micro ElectroMechanical Systems (MEMS), ranging from nanometres through micrometers to millimetres. This form is called Smart Dust. Some of these can combine multiple tiny mechanical and electronic components, enabling an increasing set of functions to be embedded into ICT devices, the physical environment and humans.

Today MEMS, such as accelerometers, are incorporated into many devices such as laptops to sense falling and to park moving components such as disk arms, are being increasingly embedded into widely accessed systems. They are also used in many devices to support gesture-based interaction. Miniaturisation accompanied by cheap manufacturing is a core enabler for the vision of ubiquitous computing (Section 6.4). Second, fabrics based upon light-emitting and conductive polymers, organic computer devices, can be formed into more flexible non-planar display surfaces and products such as clothes and curtains . MEMS devices can also be painted onto various surfaces so that a variety of physical world structures can act as networked surfaces of MEMS. This form is called Smart Skins. Third, ensembles of MEMS can be formed into arbitrary three-dimensional shapes as artefacts resembling many different kinds of physical object. This form is called Smart Clay.

Mobility

Mobile devices usually refer to communicators, multimedia entertainment and business processing devices designed to be transported by their human owners, e.g., mobile phone, games consoles, etc. There is a range of different types of mobiles as follows:

- **Accompanied:** these are devices that are not worn or implanted. They can either be portable or hand-held, separate from, but carried in clothes or fashion accessories. • Portable: such as laptop computers which are oriented to two-handed operation while seated. These are generally the highest resource devices.
- **Hand-held:** devices are usually operated one handed and on occasion hands-free, combining multiple applications such as communication, audio-video recording and playback and mobile office. These are low resource devices.
- **Wearable:** devices such as accessories and jewellery are usually operated hands-free and operate autonomously, e.g., watches that act as personal information managers, earpieces that act as audio transceivers, glasses that act as visual transceivers and contact lenses. These are low resource devices.
- **Implanted or embedded:** these are often used for medical reasons to augment human functions, e.g., a heart pacemaker. They may also be used to enhance the abilities of physically and mentally able humans. Implants may be silicon-based macro- or micro-sized integrated circuits or they may be carbon-based, e.g., nanotechnology .

1.6 SERVICE ARCHITECTURE MODELS

Partitioning and Distribution of Service Components

There is a range of designs for dividing and distributing services that depend on:

(1) the application

(2) the type of communication service

(3) the type of access device used.

In low resource devices, a chess application executes remotely because there is not enough CPU power to execute the application locally. Relatively high resource access devices can act self-sufficiently, operating in an offline mode as monolithic or standalone devices the appliance model.

Relatively resource poor access devices, such as lightweight mobile devices, are often designed so that service execution largely occurs over the network, the utility model. However, the need to cope with unreliable and low performance networks, as well as the need to adapt power consumption to the power reserves available in mobile applications supports the case for some degree of self-reliance and local processing support elements of an appliance model. There is a need to balance the degree of local information storage and local processing against the degree of remote processing and communication bandwidth required. The balance may need to shift dynamically depending on the type of ICT infrastructure available and on the type of applications.

Multi-tier Client Service Models

In a single tier, monolithic system, the whole application service resides locally, when it is operating. The system may be networked so that under special conditions it can go online to seek help when its operation is interrupted or because of local failures. In a two tier, thin client server, the access device (or client device or terminal) supports data access or presentation, service processes execute remotely and the information associated with these services is stored remotely. In a two tier, fat client server model, the access device can support some local processing and some local use of services but can also invoke remote services.

Distributed Data Storage

Some of the components such as (information) resources, processing and access can be further split to support different types of application. Types of systems in which information resources are divided and distributed include: transaction monitors where data transactions are created by distributed data sources such as point of sale terminals where data warehouses; centralised analysis of centrally stored data sub sets that are periodically extracted from distributed data resources is supported; distributed databases where queries are distributed to multiple heterogeneous databases, each individual database is mapped into a common form using a database wrapper.

Distributed Processing

Although a single CPU client server type architecture is the dominant processing model used in distributed systems, sometimes more processing power is needed for a short time. One way to achieve this is by dividing the processing, distributing it among multiple remote processors, each executing part of the processing in parallel and then reassembling the results from the individual pieces to form the whole. For this to be worthwhile, the time gained in increasing the processing must be significantly more than the time taken to partition and distribute the tasks, collect the individual results and reassemble them. There are several different computing architectures to achieve parallel processing.

Client–Server Design

The client server model has the advantage of more centralised control of distribution but has the disadvantage that the distribution and configuration of servers are fixed. The client server model is an asymmetric distributed computing model with respect to the resources and the direction of the interaction. Servers are usually resource rich, e.g., have a higher storage capacity, more powerful CPUs to support intensive processing tasks, a high bandwidth, always on network connection in order to service multiple service requests, and act as a shared data repository. In contrast to servers, clients are relatively resource poor. Client server interaction is also asymmetric: client processes on access devices initiate the interaction, making requests to application service processes on servers that wait for client requests. This asymmetry simplifies the synchronisation between clients which start requesting while servers which start waiting for client requests.

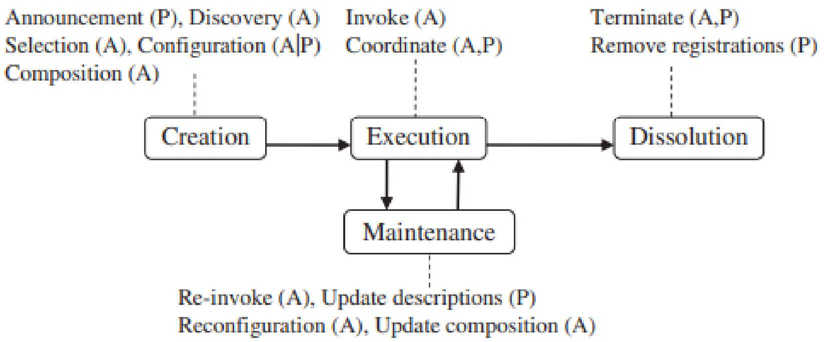
Proxy-based Service Access

Some applications, in order to mask the complexity of communication from being supported in client access process, use a client proxy. Proxy based service access can:

- offload presentation processing and network processing from the low resource client access device to the third proxy node, e.g., the client proxy could speak XML to a Web server but use a simpler message protocol when communicating with the mobile device, thus avoiding having to parse complex XML data structures in the access terminal
- hide the heterogeneity of different terminal types and different types of networks from the access applications
- simplify and compose access to multiple service providers. A proxy may request a default connec tion, reducing the processing and communication with wireless end devices. It can support a mobile portal model by aggregating content from multiple service providers.
- reduce the complexity of communication used in access devices, e.g., often complex hierarchical data structures need to be encoded and decoded into more efficient serialised data structure for transmission e.g. by compressing data or transcoding data at the server side thus reducing the amount of air wireless bandwidth consumed.
- enable devices to operate intermittently in a disconnected state. Devices may power off, move out of range of the wireless network, or simply choose to operate in a disconnected mode • shield network-based applications from the mobility of the access devices.

1.7 SERVICE PROVISION LIFE CYCLE

The provision of application services for smart devices entails the management of distributed services throughout the whole of their life cycle and not just in specific phases such as service discovery. For example, smart devices that operate in dynamic environments, such as smart mobile devices, cannot assume that services will remain static during service operation. The design of service provision must contend with intermittent service access and handovers between different service instances, e.g., wireless communication handovers, There are two separate aspects to this, first, defining a generic life cycle model for service provision and, second, to manage this life cycle. In a simple service provision lifecycle model, only two of the five servicemodel components are active, the processing services or service provision and service access or clients, the other three components, communication, stored information and information sources, are treated as passive components.



In the service creation phase, service processes register themselves in service directories. Service requesters in access nodes search for services (information processes and repositories). Services get selected, configured, and multiple services need to be composed, e.g., multiple services need to be composed to capture, annotate, transmit, edit, store, configure and print images. In the operational or service execution phase, services are invoked and multiple interlinked services may need to be coordinated. In the service maintenance phase, service processes, access configurations and service compositions can be updated. In the service dissolution phase, services may be put off line or terminated temporarily by the processes themselves or by requesters. Services may also be terminated permanently and removed.

The design for the service lifecycle depends on application requirements such as the type of mobility needed . For example, a static device such as a set top audio video receiver can support both dynamic service initiation and execution. This enables the device to be preconfigured using default factory settings and then shipped to be used in different regions in which it must detect and tune itself to the variable regional RF broadcast signal sources. This can also enable a static smart device to switch to an alternative service provider when a fault occurs, providing the user has permission to access it, possibly via another service contract.

❄❄❄❄❄❄❄

2

SOCIAL NETWORK ANALYSIS

Unit Structure

2.1 SMART MOBILE DEVICES, USERS, RESOURCES AND CODE

Users are naturally mobile, e.g., users can move in between Internet nodes, to log on and to access Web based content and email, anywhere, anytime. Users can carry personalised mobile networked devices with them to access services filtered according to their personal prefer ences and to be aware of their location context and adapt to it . Other types of inanimate hosts such as transport vehicles can also act as mobile hosts.

Each of the main components of a UbiComm system can be mobile:

1. virtual processing and services at the operating system and application level, code, access devices
2. hardware and data resources

Mobile Service Design

Mobility service design builds upon the basic design of smart devices and services but is a more specialised variant of it. In order to simplify access by applications and users, mobility, e.g., locating and addressing mobile users and routing data to mobile receivers, should be designed to be more transparent to applications and users. There are three kinds of transpar ency for mobile services:

(1) user virtual environments (UVE)

(2) mobile virtual terminals (MVT)

(3) virtual resource management (VRM). UVE provides users with a uniform view of their working environments independent of the location and terminal type

2.2 MOBILE CODE

Code is usually designed to be downloaded from a remote service point. Installation requires configuring the code installation onto each platform. This is automated using a Makefile, e.g., installers can instruct the

Makefile in the file system to install the code. Once it is installed, it often remains at one point in the ICT infrastructure, although it can be maintained and updated. Service access devices can also download new operational capabilities at run time without requiring the capacity to store all possibly needed service support in advance this reduces the need for resource rich service access devices in dynamic environments. This paradigm enables providers to maintain, e.g., upgrade and fix, code in consumer devices with a network connection without the provider having to ship physical media to customers

Mobile Devices and Mobile Users

Device mobility can be viewed from several dimensions:

(1) in terms of physical dimensions)

(2) in terms of whether or not the device is mobile or some kind of host to which it is attached to is mobile

(3) in terms of what kind of host, mobile devices can be bound to

(4) in terms of how devices are attached to a host

(5) in terms of when the mobility occurs. Each of these is discussed in turn.

2.3 OPERATING SYSTEMS FOR MOBILE COMPUTERS

Desktop computers and high ICT resourced mobile devices such as smart phones (fat client devices) often require more complex multi process control and the use of a Multi-Tasking Operating Systems (MTOS) to support the execution of multiple users and applications.

Microkernel Designs

In a macro kernel or Monolithic Kernel Operating System all the system utilities such as hardware related drivers, memory management, process support, process scheduling, network pro tocol stack and file system are in one, single, large, kernel system .

The main benefit of the monolithic kernel system is that it is more efficient for a single processor system because fewer context switches are needed. Context switches are only needed between processes and the kernel utilities and not generally between processes.

The main drawback of a monolithic kernel system for low resource systems such as mobile systems is that use of the kernel is quite large and requires many system resources. In addition, the monolithic kernel is potentially more complex and has more points of failure and can require more updates, e.g., in order to support adding new hardware. In contrast, in a microkernel, only the fundamental parts of the operating system such as basic memory management, (limited) process management and inter process control are supported. The potential benefits of a small kernel are that it is more manageable in a low resource environment and more robust.

It can still function even when system utilities, not in the kernel, fail. The drawbacks are that there is potentially more context switching between application, non kernel utility and the kernel utility process execution contexts, thus potentially lowering the performance when run on single processor systems.

Mobility Support

Some original designs for OS for low resource device such as mobile devices were based on creating a cut down version of PC type MTOSs, e.g., Windows Mobile, while other designs for an OS were specifically oriented to a lower resource mobile device and other specialist characteristics from the ground up. Mobile device design is still evolving. Some aspects of mobile OS support are even being proposed to produce low cost, low resource, low power, PC design. In addition, to the normal OS ICT support, mobile devices have several more specialised requirements including HCI for small mobile user interfaces, heterogeneous communication support ,intermittent connections , data management ,mobility support and power management .

The core OS kernel should be small for a low resource device. Basic support is needed for memory management, to prevent memory leaks and to release system resources as soon as they are no longer needed. Good strategies for resource reuse are vital as resources are limited. There are different options for multi-tasking support. Multi-tasking is useful to support communications capable real time performance in order to talk, to count down alarms that were set and to run and access data and applications on the phone all at the same time, e.g., Symbian OS.

Alternatively, a system can schedule one task at a time, wait for it to finish and then switch to another one, i.e., non-pre-emptive task scheduling, e.g., Palm OS. Small mobile devices tend not to use magnetic disk based secondary storage for persistence, because of the relatively slow access speeds, damage to moving parts and higher power consumption. Permanent storage is used in the form of a Flash ROM to retain files and data, however, flash memory is slower than RAM. Mobile devices tend to boot from ROM. However, flash memory uses less power, so the battery life of devices can be longer. Note in the past, certain types of ROM had a shorter lifetime in terms of number of read writes they supported before they failed.

Smart Card Devices

A smart card is a plastic card embedded with digital memory and (usually) a microprocessor chip, as opposed to cards which store data on magnetic strips. It is reprogrammable, stores and processes data in the card and transacts data between card users and applications. Data can be stored and processed within the card's memory or microprocessor, which is accessed using a card reader (Shelfer and Procaccino, 2002). Smart cards are small and easy to carry around. They provide a secure data container, can be used for authentication purposes, e.g., as a hardware based digital signature, and can be used for metered services .

Many things found in a person's wallet have the potential to be replaced by a smart card, including driver's licence, insurance information, chip and pin credit or debit bank card, travel card and ticket, etc. Smart cards potentially represent a virulent form of Privacy Invasive Technology or PIT.

Smart Card OS

The primary tasks of a smart card operating system on behalf of applications are transferring data to and from the smart card, controlling the execution of commands, managing files, managing and executing cryptographic algorithms to protect access to stored data, and mana ging and executing program code. The components of a typical card OS are combined into a single IC chip to reduce size. To standardise the communication protocol between the on card application, and off card programs accessed via the reader, the ISO/IEC 7816 standard has been defined.

Device Networks

The goal of a device network is to enable a wide variety of devices to interoperate. These activities include home automation such as light and climate control, person aware systems, home security care in the community and pervasive AV content access . The enablers for this are home network and service infrastructures that are easy to install, to configure and to maintain; low cost (capital and operational cost) devices and infrastructures, and useful applications. In part, this is due to the ease of distribution throughout the home without the main tenance cost of re wiring and because the added insecurity of wireless does not yet appear to be a major concern for users. Currently, the home device infrastructure is highly fragmented and far from being a seamless infrastructure. First, many electronic devices are monolithic, may not be digital and are not network enabled.

❄❄❄❄❄❄❄

3

CLOUD COMPUTING –II

Unit Structure

3.1 HUMAN–COMPUTER INTERACTION (HCI):

The term Human Computer Interaction (HCI) has been in widespread use since the advent of the IBM computer for personal use in the mid 1980s. However, the groundwork for the field of HCI certainly started earlier, at the onset of the Industrial Revolution. Tasks became automated and power assisted, primarily to save labour, but also motivated by the need to overcome some limitations in human abilities and to perform tasks at a reduced cost. This triggered an interest in studying the interaction between humans and machines in order to make the interaction between them more effective. To enable humans to effectively interact with devices to perform tasks and to support activities, systems need to be designed to be useful and to be usable.

3.2 EXPLICIT HCI:

Motivation and Characteristics The basic concepts of HCI are:

- **humans:** single or multiple users, with diverse physical and mental abilities, interacting coopera tively or competitively
- **computers:** not just PCs but also a range of embedded computing devices and a range of device sizes such as dust, tabs, pads and boards
- **interaction:** may be directed via a command or by manipulating virtual objects (windows, desktop) but it can also involve more natural interaction such as speech interaction, gestures, etc.

HCI refers to the processes and the models for design and the operational interface for some products. For many users, the User Interface (UI) part of the system is the product. Explicit HCI puts the user at the centre of the interactive systems, so that the control of the system, responds to and is driven externally by the user, rather than the system being driven internally.

Inefficiently designed UIs can lead to both higher training costs and higher usage costs and of course leads to lower product sales. The reasons for higher training costs include users spending time working out what is happening, trying out inappropriate computer services and impaired task quality. Users may feel that a particular machine force them to do tasks in ways they prefer not to (no control). Users may have to re learn how to perform tasks: starting work again lowers productivity. Poor UIs can lead to higher error rates not acceptable in safety critical systems that require some human interaction.

The motivation for HCI is clear; to support more effective use in three ways to be:

- **useful:** accomplish a user task that the user requires to be done;
- **usable:** do the task easily, naturally, safely (without danger of error), etc.
- **used:** enrich the user experience by making it attractive, engaging, fun, etc.

Complexity of Ubiquitous Explicit HCI

Explicit HCI is complex for UbiCom scenarios even if it is well designed for individual devices because we need to use tasks as part of activities that require access to services across multiple devices, because devices can be used by different types of people, because users are engaged in multiple concurrent activities, because users are engaged in activities which may occur across multiple physical environments, because activities may be shared between participants, and because activities on occasion need to be suspended and resumed

3.3 IMPLICIT HCI

The concept of implicit HCI (iHCI), proposed by Schmidt (2000) was introduced in Section 1.2.3. It is defined as ‘an action, performed by the user that is not primarily aimed to interact with a computerised system but which such a system understands as input’. To support iHCI interaction is more about C2H (Computer to Human) Interaction. Implicit interaction is based on the assumption that the computer has a certain understanding of users’ behaviour in a given situation: it is a user aware type of context aware system.

This knowledge of users’ behaviours in given situations is then considered an additional input to the computer while doing a task (Schmidt, 2000). Implicit interaction can allow some of the interaction to be hidden from users and hence for the device to become invisible. Implicit interaction is also introduced as we seek to design systems with which we can interact with in a more natural1 way. For example, if we use hand gestures, such as a clap, to control a device to switch it on and off, we may also clap our hands for other reasons such as to express an emotion. Unless the context of the gesture is also defined and shared, clapping cannot be unambiguously used to imply that it is a command to switch a device on or off. There are some obvious challenges in supporting implicit interaction. It can sometimes be complex to accurately and reliably determine the user context because of: the non-determinism of the subject, the user has not clearly decided what to do, or because of non-determinism in the subject's environment.

Implicit interaction, in contrast to explicit interaction, reduces or even removes explicit user control. The user context determination may however invade and distract users' attention in order to directly interact with them or the determination may be inaccurate because it can only indirectly model users, e.g., via observing user interaction and inferring user behaviour. Systems may also require time in order to learn and build an accurate model of a user. Partial iHCI systems can be very useful in practice if designed appropriately. For example, they can be used to anticipate actions, to prioritise choices and to lessen the overload on, and the interaction by, users

3.4 USER INTERFACES AND INTERACTION FOR DEVICES

Four of the most commonly used networked ICT devices are the personal computer in its various forms such as desktop and laptop, hand held mobile devices used for communication, games consoles and remote controlled AV displays, players and recorders. Each of these can be designed to perform a common set of functions such as AV player, record, output, data and AV communication and to act as a hub or portal for multiple service access and interoperability.

Diversity of ICT Device Interaction

The term computer in the acronym HCI has a much more diverse meaning within the field of UbiCom. It refers to any device with a programmable IC chip inside, including a range of multi task operating system (MTOS) devices. Common multi task devices include desktop and laptops PCs, mobile phones, games consoles, AV recorders and players such as televisions, radios and cameras. Each of these supports keypad or keyboard haptic inputs and audio outputs and output display interfaces.

The user interfaces for these devices are primarily visual. Note the design for a universal visual interface interaction model will not work equally effectively across the wide variety of display and input device types and sizes.

Even more numerous than MTOS computers are computers embedded in devices that perform specialised tasks such as various household appliances, vehicle control systems, travel ticket machines, cash dispensers, building controls, etc. Their interfaces are far more diverse and in many cases have less input controls, typically consisting of a set of control buttons and knobs, and outputs such as one or more LEDs (Light Emitting Diodes) and LCDs (Liquid Crystal Displays).

six different form factors for ICT devices were considered. Of these, decimetre sized, from about 5 to 20 centimetres are the most common form for ICT devices today (Figure 5.1). There are several dimensions devices could be characterised according to:

- **size:** hand sized, centimetre sized, decimetre sized versus micro sized versus body sized or larger
- **haptic input:** two handed versus one handed versus hands free operation
- **interaction modalities:** single versus multiple
- **single user versus shared interaction:** in personal space, friends' space or public space
- **posture for human operator:** lying, sitting, standing, walking, running, etc.
- **distance of output display to input control:** centimetres to metres
- **position during operation:** fixed versus mobile;
- **connectivity:** standalone versus networked, wired versus wireless
- **tasking:** single task devices versus multi task devices
- **multimedia content access:** voice and text communication oriented, alpha numeric data or text oriented, AV content access
- **integrated:** embedded integrated devices versus dynamically interoperable devices.

3.5 HIDDEN UI VIA BASIC SMART DEVICES

The WIMPS style interface which dominates PCs is considered by many computer scientists to be obtrusive in the sense that it requires users to consciously think about how to operate a mouse pointer interface and which keys to press to use the computer. The computer itself is localised and users must go to its location to use it. In contrast, systems which can be situated where our activities are based and which directly make use of natural human sensory input offer a less obtrusive computer interface.

Multi-Modal Visual Interfaces

The modality of interaction refers to a mode of human interaction using one of the human senses and to the type of computer input. The categories of human senses are sight, touch, hearing, smell, and taste. ICT systems have modalities that are equivalent to some of the human senses such as cameras (sight), input devices such as touchscreens, keypads and pointer devices (touch, haptic), microphones (hearing) and the use of various chemical sensors and analysers (smell and taste).

The majority of interactive systems predominantly use a single visual mode of output interaction between a system and a human user but this can overload humans as the world becomes more digitally interactive and as more objects can seek to interact with the user at any one time. The use of multiple sensory channels can alleviate this bottleneck by increasing the bandwidth available. Human interaction is naturally multi modal in the sense that users can use multi modes of input and output to an extent concurrently. Systems typically use multiple instances of haptic modes, for input mouse and keyboard. Human users can receive multiple inputs from other humans and systems, i.e., listening to a voice and looking at someone. Users can also transmit multiple outputs, i.e., gesturing and talking, at the same time

Gesture Interfaces

Gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body, with the intent of conveying meaningful information about interacting with the environment. There are three main types of body gestures: hand and arm gestures, head and face gesture and full body movement

Gesture Interfaces Gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body, with the intent of conveying meaningful information about interacting with the environment. There are three main types of body gestures: hand and arm gestures, head and face gesture and full body movement

Gestures can also be classified into contactful gestures, e.g., handshake, touchscreen gesture input or contactless gestures, e.g., waving at a person or camera to attract attention. This type of classification also reflects 2D gestures on a touch sensitive planar surface as opposed to 3D gestures using a hand held device incorporating gyroscopic or accelerometer sensors

3.6 HIDDEN UI VIA WEARABLE AND IMPLANTED DEVICES

Posthuman Technology Model

In the Posthuman model (Hayles, 1999), technology can be used to extend a person's normal conscious experience and sense of presence, across space and time. There are three types of post human technology:

accompanied, wearable and implants. Accompanied technology is technology external to the body which accompanies it but is not directly attached to it, e.g., personalised mobile devices, smart cards, smart keys, etc.

Wearable technology is technology external to and directly attached to humans, e.g., hearing aids and wireless earpieces attached to mobile phones to support hands free use of phones. Implants are technology internal to the body. The obvious applications for implants are medical, to use various prosthetics and bio implants to overcome paralysis in limbs and muscles and to help regulate and treat irregular biological phenomena such as heart activity.

Virtual Reality and Augmented Reality

Most computers currently present **visual** information in two dimensions, although simple three dimensional or 3D effects can be created by using shadows, object occlusion and perspective. These are an important element of games consoles which heightens user satisfaction in the main group of users. More realistic 3D effects can be created by mimicking the stereoscopic vision of eyes

where each eye sees a slightly different perspective of the same scene. For example, a 3D headset or goggles that contains two miniature screens, each one showing the same scene from a slightly different perspective. Alternative techniques to simulate 3D are to either use polarised light or to blank out each eye synchronised to the computer frame rate so that each eye sees alternate images. As the head moves, sensors detect the change in angle to view the scene and the changing scene perspective is calculated and presented.

Wearable Computer Interaction

The essence of wearable computing is to embed computers into anything that we normally use to cover or accessorise our bodies. This includes clothes, jewellery, watches, eye wear, teeth wear, ear wear, headwear, footwear, and any other device that we can comfortably attach to our bodies and allow to behave as hidden computers. In a broader sense, devices can also be embedded in the environment that accompany us, in our transport vehicles are extensions of wearable computing. Wearable computers are especially useful when computer access is needed while a user's hands, voice, eyes or attention are actively engaged within a physical environment.

specified three criteria to define wearable computing.

- **Eudaemonic criterion (in the user's personal space):** the ICT device appears to be part of the user as considered by the user and observers of the user.

- **Existential criterion (iHCI Control by user):** ICT devices are controllable by the user. This control need not require conscious thought or effort, but the locus of control must be such that it is within the user's domain.
- **Ephemeral criterion (responsiveness):** interactional and operational delays are non existent or very small.
    - **Operational constancy:** It is always active while worn.
    - **Interactional constancy:** One or more output channels are accessible (e.g. visible) to the user at all times, not just during explicit HCI.

Head(s)-Up Display

(HUD) Head(s) Up Display or HUD, is any type of display that presents data without blocking the user's view (Sutherland, 1968). This technique was pioneered for military aviation and is now being used in commercial aviation and cars. There are two types of HUD. In a fixed HUD, the user looks through a display element attached to the airframe or vehicle chassis, the system projects the image with semi transparency onto a clear optical element and the user views the world through it (augmented reality). In a head mounted display, the system precisely monitors a user's direction of gaze and determines the appropriate image to be presented. The user wears a helmet or other headgear which is securely fixed to the user's head so that the display element does not move with respect to the user's eye.

Eyetap EyeTap

is a device that is worn in front of the eye that acts as a camera to record the scene available to the eye, and acts as a display to superimpose a computer generated imagery on the original scene available to the eye (Mann and Fung, 2002). An EyeTap is similar to a HUD but differs in that the scene available to the eye is also available to the computer that projects the HUD. This enables the EyeTap to modify the computer generated scene in response to the natural scene. The EyeTap uses a beam splitter to send the same scene (with reduced intensity) to both the eye and a camera. The camera then digitises the reflected image of the scene and sends it to a computer. The computer processes the image and then sends it to a projector. The projector sends the image to the other side of the beam splitter so that this computer generated image is reflected into the eye to be super imposed on the original scene. One use, for instance, would be a Sports EyeTap that enables the wearer to follow a particular player in a field

Virtual Retinal Display (VRD)

Virtual Retinal Display (VRD), also known as a retinal scan display (RSD), draws a raster display (like a television) directly onto the retina of the eye (Johnston and Willey, 1995). The user sees what appears to be a conventional display floating in space in front of them. This is in contrast to past systems that have been made by projecting a defocused image directly in front of the user's eye on a small 'screen', normally in the form of large sunglasses. The user focused their eyes on the back ground, where the screen appeared to be floating. The disadvantages of these systems were: the limited area covered by the 'screen'; the heavy weight of the small televisions used to project the display, and the fact that the image would appear focused only if the user was focusing at a particular 'depth'. Limited brightness made them useful only in indoor settings.

Clothes as Computers

Unlike HUD, EyeTap and VRD that focus on single sensors, clothes as computers use a network of sensors that can be worn and the data from them fused to allow other types of non visual context awareness. Van Laerhoven et al. (2002) have reported their experiences with a body distributed sensor system that integrated tens of accelerometers spread over the body into a garment with the majority on the legs and the rest divided over the arms and upper body. The accelerometers for the legs were integrated into a harness to enable testing and capturing of data from multiple users of different figure heights, while the others were attached on regular clothing using Velcro. The experiments indicated that it is feasible to distinguish certain activities of a wearer whose clothing has an embedded distributed sensor network. These activities could also include gestures made by the user, and basic events related to garments, such as putting on a coat or taking off a coat. These can be recognised with a reasonably high precision

3.7 HUMAN CENTERED DESIGN (HCD).

Conventional system design focuses on designing the core of the system to support sets of service actions or functions in order to handle service requests. In contrast, human centred design, also called user centred design, focuses on the design of the part of the system human users interact with, the user interface. For many users, the user interface represents the product. Whereas conventional system design mostly involves users only at the start and the end of system development life cycle, human centred design involves users throughout the whole of the design life cycle.

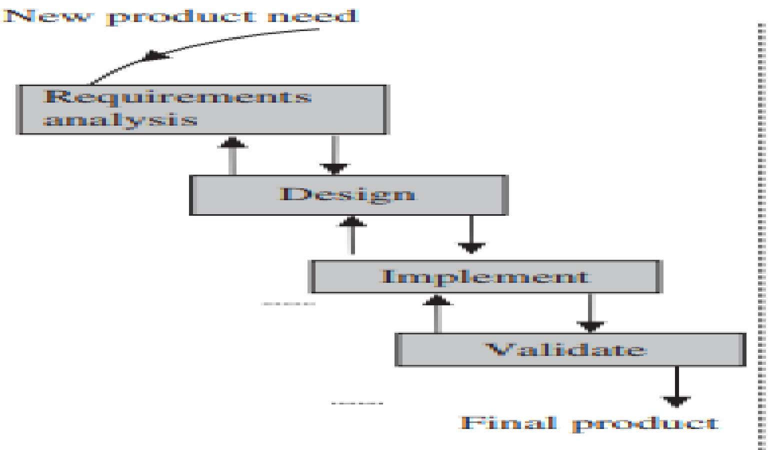
Human-Centred Design Life-Cycle The basic phases of development of user centred interactive system design are similar to those used in conventional functional system design (Figure 5.5). This development cycle comprises four phases of development:

(1) requirement gathering and analysis

(2) modelling and design

(3) implementation

(4) testing.



Functional system design

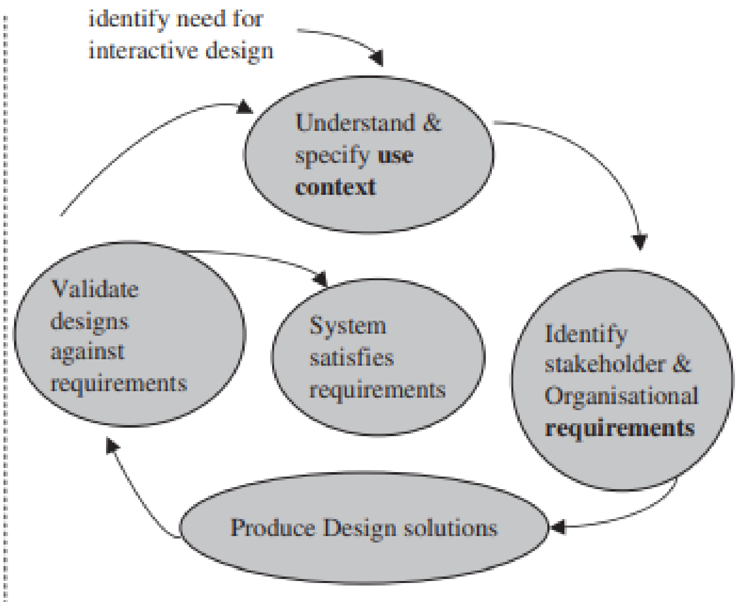
The ISO 1340719 human centred design process for interactive systems specifies four principles of design:

(1) the active involvement of users and a clear understanding of user and task requirements

(2) an appropriate allocation of function between users and technology based upon the relative competence of the technology and humans

(3) iteration is inevitable because designers hardly ever get it right the first time

(4) a multi-disciplinary approach to the design



Human-centered design

Methods to Acquire User Input and to Build Used Models

The basic techniques used to acquire user input for the formative evaluation of designs can also be used throughout the human centred design life cycle from requirements gathering, to model user behaviour. Selection of appropriate user input techniques will depend on a variety of factors, such as the different sub types of environment requirements, cost and time constraints, and availability of different types of users.

Inspection or heuristic evaluation: experts and experts as users inspect the user interface and are asked to consider and document how a set of usability heuristics apply when carrying out prespecified user tasks. There are different types of usability heuristics that apply when evaluating different types of interactive products such as online communication, web sites, desktop computers and mobile devices.

Cognitive walk through: this is an alternative approach to inspection in which designers and expert evaluators as individuals or groups walk through the sequences of a task and document whether or not a user will know how to achieve a task, what actions are available, whether or not users can interpret the system response to an action.

Observing users is the main technique used to gather indirect input about users. The same tools that are used to observe users in controlled environments (in the lab) can be used in uncontrolled environments (in the field) but the way in which they are used differs.

Predictive models provide ways of soliciting user feedback about products or designs, without necessarily directly involving users. Quantitative models of user interaction sequences such as clicking keys, moving a mouse pointer, thinking and moving between mouse and keyboard, can be used to calculate task efficiency such as how long it takes to complete a task and to quantitatively compare user interactions.

❄❄❄❄❄❄❄

4

TAGGING, SENSING AND CONTROLLING

Unit Structure

4.0 OBJECTIVES

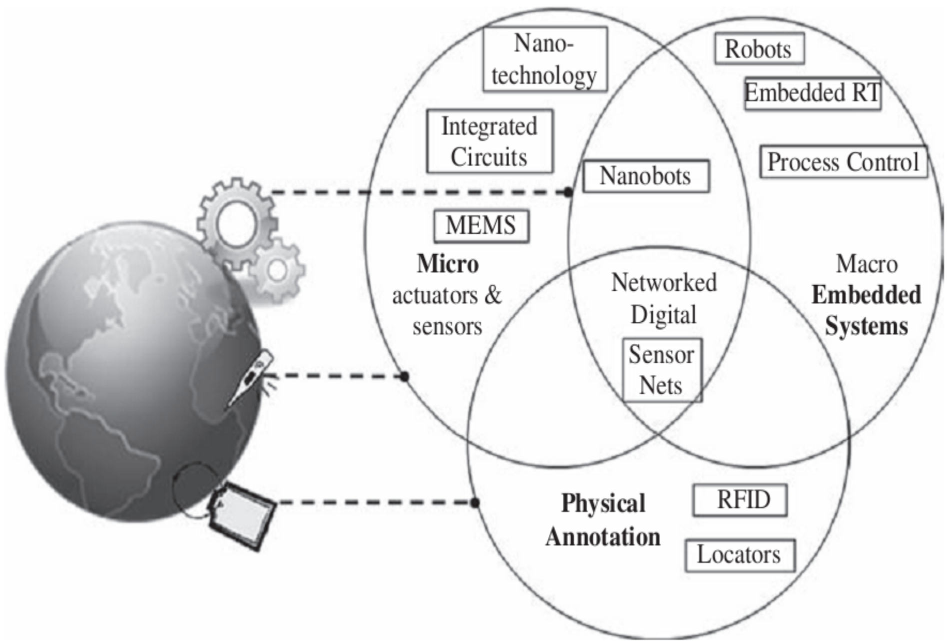
To understand sensors and its applications, tagging and controlling physical world objects.

4.1 INTRODUCTION

Technology is changing by the minute be it electrical ,mechanical or information technology. It has already become a part of our life. The demand for newer technologies had put the focus on inventing low cost, mass produced, faster to produce and market and miniature size devices. The demand is for low cost and low maintenance devices. Micro Electro Mechanical Systems (MEMS) enable sensing and actuation down to a scale of a nanometer. These devices are small in size and can be carried around on embedded in the environment. They are connected to each other by the Internet thus enabling truly ubiquitous computing.

Embedded systems allow making an animate as well as inanimate ubiquitous by embedding MEMS and providing control over them. Human being can interact with devices at a global scale. These interactions are not always explicit and come naturally to human beings like walking or talking. We can annotate things so that they can be interacted with. For example: locating your lost mobile phone or keys.

The three driving factors for truly ubiquitous and pervasive computing to materialize are MEMS, Embedded Systems and Physical annotation.



Ubiquitous Computing: Smart Devices, Environment and Interactions

4.1.1 Chapter Overview

This unit can be divided into three sections:

- Tagging - It refers to tagging physical objects with identifiers
- MEMS - Smaller devices that can be easily blended in the physical environment.
- Interaction – Connecting devices and interacting with them via embedded controllers.

4.2 TAGGING THE PHYSICAL WORLD

Physical world objects can be identified by tagging those using digital tags like RFID. Digital tags are networked electronic devices with an identity. Tagging a physical object allows us to interact with them or track them in a virtual space. Tagging can also be used with augmented reality. AR maps physical world objects into computer artefacts. It can be used to enhance the physical world objects with computer assisted information. AR environments can be used for information or for interaction.

Physical world objects can be identified by means of tagging and their virtual views can be captured using image, audio or video recording. These can further be used for location is space and time as well as for extracting features of interest. Devices such as phones and cameras capture location along with time while annotating images/videos captured.

4.2.1 Life-Cycle for Tagging Physical Objects

The tagging of physical objects using digital tags consists several processes like:

- capturing a physical view or recording of physical objects or some object feature such as:
  - moving or placing reader in range of an object with a tag; (Barcode)
  - moving objects with physical tags in the range of readers;(Toll booth)
  - capturing a physical view of an object and its surroundings.
- Identifying physical objects:
  - using the captured ID from an object to identify the object.
  - assigning an ID to a physical view of the object and then identifying the object within that view perhaps by relative position, e.g., the object is the bottom right corner.
- Anchoring or relating objects:
  - defining the relationship of objects with respect to a physical view, e.g., marking a location on a physical map;
  - defining the relationship of objects with respect to a virtual view, e.g., marking a location on a virtual map

- Organization or structuring: Displaying objects in relation with different types of objects within the same view, e.g., a location on the map with satellite, road and terrain view.
- Presentation: superimposing graphic annotation on physical world views such as maps (pin on a map to mark location) or by different degrees of detaching annotations in different forms from the physical views.
- Management: managing the annotation processes and data including, creating editing, removing, recycling, storing, querying and access control to the annotation data.

4.2.2 Tags: Types and Characteristics

Tagging can be performed in various ways. Mackay describes three basic approaches to augment physical world objects for use in virtual (computer) environments:

(1) augmenting the user – The user perception of the physical world is augmented by generating a digitized view of the physical world. E.g. use AR.

(2) augmenting the physical object – The physical objects can be digitized by attaching a tag to a physical object. e.g. use RFIDs;

(3) augmenting the environment – The environment around the user and object can be digitized and presented in a virtual environment e.g. Map with user's current location.

Hansen's analysis focuses on the presentation of the annotation

- Co-located or Local - Whether user of the annotation is onsite with the physical object
- Not Co-located or Remote - Whether user of the annotation is offsite with the physical object
- Linked or anchored – Virtual tag is attached to the object
- Not augmented – Not connected to the object

AR is an example of Co-Located reality whereas VR is not co-located. In AR, the user is in a real physical environment and so are the annotations. In VR, the user is interacting with the virtual view of the world and the annotations are attached to the virtual view. VR can not necessarily be true. For ex: playing a game. The tags used in web-based information system are an example of tags not connected to any physical objects.

Physical tags are used to identify physical things. They are used to identify, describe and represent the physical things as virtual tags. These tags can be used to manipulate the virtual environment. Virtual tags can be considered as metadata or descriptions or annotations about the physical object which can be stored onsite i.e. on the object itself or offsite in a database.

A physical tag can be linked to multiple virtual tags but a virtual tag can be not always be associated with a physical tag. Since, a physical tag is associated with an object it needs to be updated as soon as the object state changes.

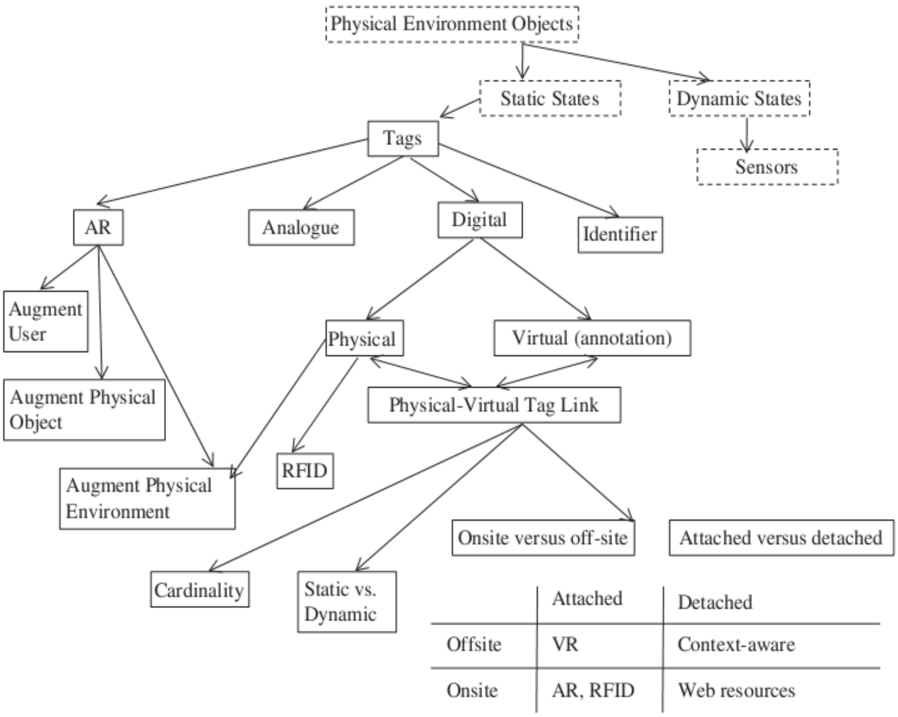


Figure 6.2 Taxonomy for types and characteristics of tags

Ubiquitous Computing: Smart Devices, Environment and Interactions

4.2.3 Physical and Virtual Tag Management

Managing tags can be challenging with respect to the varied characteristics of physical as well as virtual tags.

Tags can be attached to an object while its creation or attached later on. Ex: Printing a barcode on the device while manufacturing or sticking a barcode on the device post production. Tags can be permanent or removable. Ex: IMEI is permanent whereas mobile number can be changed. Tags can be removed after the end of the lifecycle or never removed. Ex: the tags attached to apparels in a shop are removed after purchase. Tags can disposed of when the object is disposed. We should be able to read tags outdoors in noisy, wet, dark or very bright environments. Once a tag is read, its annotation data should be stored, distributed and integrated immediately. An area can have multiple tags and readers overlapping with each other leading to redundant capture of data. In such case, applications and businesses should apply filtering and define the level of aggregation, reporting and analysis, which can be dynamically reconfigured.

Some design challenges include interoperability between different annotation systems and if the user is aware that he is being tagged or tracked (privacy issues).

4.2.4 RFID Tags

Radio Frequency Identifier (RFID) tags are attached to physical objects. They are used to identify the physical objects in a physical as well as virtual world. They might also contain some additional information along with the object identifier. RFID is truly wireless because it does not need to be line-of-sight to be read. RFID tags are used to tag and track high value goods as well as objects that are on the move. Barcode is more widely used than RFID for low cost consumer goods.

RFID tags can be used for various application like tracking rental car, location of delivery items, hospital patients, prisoners, payment in toll booth, etc. RFID tags are now used for payment of goods by hovering the card over a POS.

RFID operate in a promiscuous mode i.e. they reply to a generic scan without authentication or some form of activation code. Although , retail operations form the major applications of RFID, it was also used in the Second World War for identifying soldiers on field.

RFID tags are classified based on whether they have their own energy supply as active or passive. Active tags have their own energy supply whereas passive tags get their energy supply from the readers. Active tags are expensive, have longer range and require more maintenance as compared to passive tags.

RFID system consists two main components: the tag itself and a reader that scans the tag and reads the information it contains.

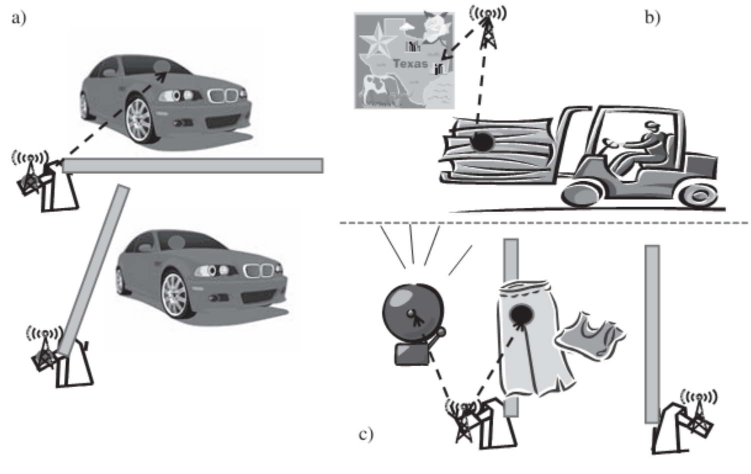


Figure 6.3 RFID tag application: (a) transponders in cars cause toll barriers to automatically lift as cars approach; (b) tags on pallet of goods tell distributers where goods are located; and (c) tags on clothes in retail outlets can signal alarms if they are removed without permission

4.2.4.1 Active RFID Tags

Active RFID tags have a long range and hence are used on large assets which need to be tracked over distances like cargo containers, rail cars, etc. They usually operate at 455 MHz, 2.45 GHz, or 5.8 GHz frequencies and they typically have a read range of 20-100 m, costing from 10 to 50 US dollars. The cost of the tags depends on the composition of the memory size, the battery life, on-board sensors and the ruggedness of the tag. There are two types of active tags: transponders and beacons.

Transponders are tags that wake up on receiving a signal from a reader. The RFID based toll payment collection is an example of Active Transponders. A car has an active transponder pasted on its windshield and the booth has a reader. As the car approaches a tollbooth, a reader at the booth sends out a signal waking up the transponder which replies by sending its unique ID to the reader. Transponders are active only when it is within range of a reader thus conserving battery life.

Beacons are used in applications where the precise location of an object is sought. The beacon emits a signal with its unique identifier at pre-set intervals. At least three reader antennas positioned around the perimeter where assets are tracked, pick up this signal. Beacons can be used along with GPS for long range tracking.

Active tags can incorporate sensors to collect additional information like the temperature around the asset and check if the asset has been dropped.

4.2.4.2 Passive RFID Tags

The tags used in ID cards for access are passive RFID tags. They get their energy supply from the reader (10 μW - 1 mW). They are cheaper, costing about 20 40 US cents, than the active tags. They are low maintenance and support shorter access range (few cm to 10 m). It consists a microchip attached to an antenna. Passive RFID comes in many forms. They can be printed and glued to an ID Card, embedded in a plastic card or any packing that is resistant to heat, cold or harsh cleaning chemicals. They operate either at the frequencies of LF:124 kHz, 125 kHz or 135 kHz or at HF:13.56 MHz or UHF: 860 MHz to 960 MHz. Low frequency can easily penetrate water and materials hence it an ideal choice for applications where the tag needs to be read through certain soft materials and water at a close range.

The power transfer from the reader to the passive tags can be done in two ways: near field and far field.

Far field as the name suggest can signal information over greater distances. Near field is based upon electromagnetic induction. An RFID reader passes a large alternating current through its electromagnetic coil (antenna) which results in an alternating magnetic field within its locality. When a tag with a smaller coil is placed in this field an alternating voltage appears across the tag. This voltage is rectified and coupled to a capacitor that accumulates sufficient charge to power the tag chip. The tag reader uses the energy to vary the magnetic field through its antenna to send a signal containing the tag ID to the reader antenna. Far field passive RFID interaction is based upon capturing EM waves propagating from a dipole antenna attached to the reader. A smaller dipole antenna in the tag receives this energy as an alternating voltage difference and again can use this to charge itself with energy. Near field magnetic induction use back scattering to transmit a signal from the tag to the reader.

4.2.5 Personalised and Social Tags

Every individual has a personal view of the world, which can be annotated and recorded. Memex was a device in which an individual could store all his books, records, and communications. It could be mechanized for consultation with exceeding speed and flexibility.

MyLifeBits is a project used to record all the personal experiences of an individual. Another initiatives such as Semacode proposes a scheme to define labels that can be automatically processed from captured images and linked to a Web based spatial information encyclopedia.

Semacode encodes URLs as part of 100 character string encoded into 2D barcodes using a software. Mobile devices can install a Semacode scanner that reads the codes, parse them and loads the associated resource onto a user device.

Physical artefacts are annotated, using simple alphanumeric codes, to reduce costs by detecting and recuperating misplaced assets. Annotation for personal use is subjective, multi-modal and represented using multimedia. Subjective annotations are used in multiple contexts, multiple applications and multiple activities by users.

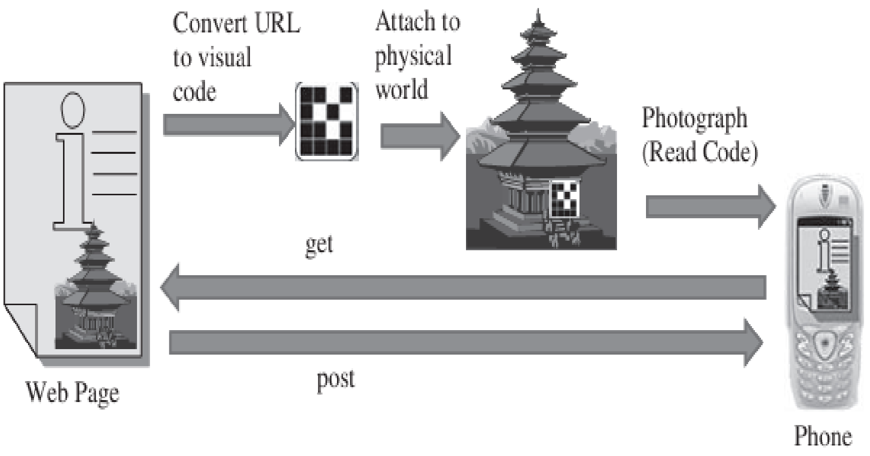


Figure 6.4 The processes of augmented reality tagging

Ubiquitous Computing: Smart Devices, Environment and Interactions

4.2.6 Micro Versus Macro Tags

RFID tags are attached to various types of objects. Hence, they can be micro or macro sized. MEMS and nanotechnology enable the manipulation of the size of the RFID tags from macro to nano. MEMS are cheaper to mass produce and can be invisible to the unaided human eye. The size of a wireless antennae transceiver is dependent on the wavelength of the wireless signal transceiver and is typically of the order of about 5 cm for a 2.45 GHz signal.

4.3 SENSORS AND SENSOR NETWORK S

Sensors are transducers that convert a physical phenomenon (heat, light, sound) into electrical signals. They act as inputs to a system so that it can adapt to perform specialized rather than general purpose functions, e.g., a temperature sensor can automatically switch off if the heat is too high.

Sensors can be used to monitor environments, track assets ,detect changes in the environment, control a system with respect to the environment within a defined range of changes; adapt services to improve their utility.

Sensors are networked. Sensor architecture consist of sensors which are data generators, as well as intermediate or services nodes that receive, post process and store data. Sensor data keeps changing with respect to the changes in phenomena they monitor. The data processing for sensor is more complex than tags.

4.3.1 Overview of Sensor Net Components and Processes

The main components of a typical sensor network system are sensors connected in a network. A sensor access node services these sensors.

There are three types of sensor nodes:

(1) common nodes - responsible for collecting sensor data;

(2) sink nodes - responsible for receiving, storing, processing, aggregating data from common nodes;

(3) gateway nodes - connect sink nodes to external entities.

Access nodes are a combination of sink and gateway nodes. Some nodes can act as a sink node as well as an access node.

In a sensor network, energy needed for communication rather than processing is the primary reason for energy utilization. Sensors have a low power, short range wireless interface to communicate with other sensors and readers within their range. To reduce power consumption and redundancy, only a single sensor source and sensors along a single path forward the data.

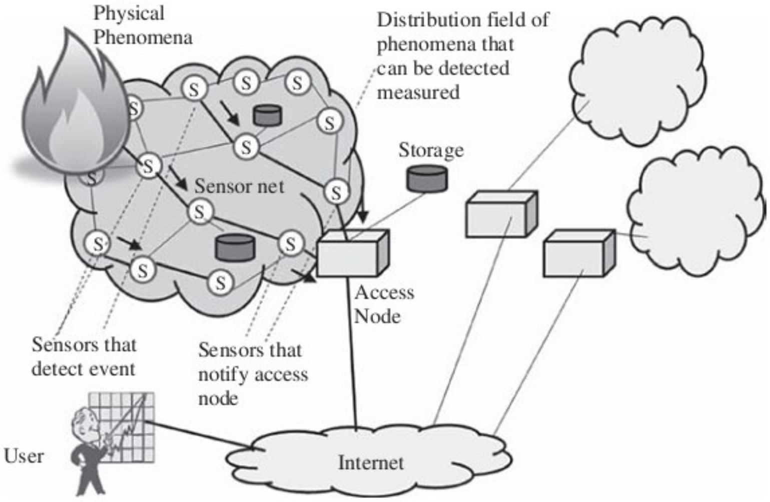


Figure 6.5 A sensor network used to detect increases in heat and report these to a user

Ubiquitous Computing: Smart Devices, Environment and Interactions

Sensor access nodes combine data from multiple sensors. The data can be stored locally or forwarded to another storage location. The access node acts as a base station that routes queries to other sensors. Sensors can be micro or macro sized.

In the Figure 6.5, two sensors are damaged by fire, three sensors are in the range of the event and two sensors are in the range of the access node. The three sensors are not in the range of the access node so they have to route their data via other sensors. Sensor networks consists large numbers of heterogeneous sensors and nodes. Hence, it is challenging to manage all of these constraints while creating a system that functions properly for the application domain.

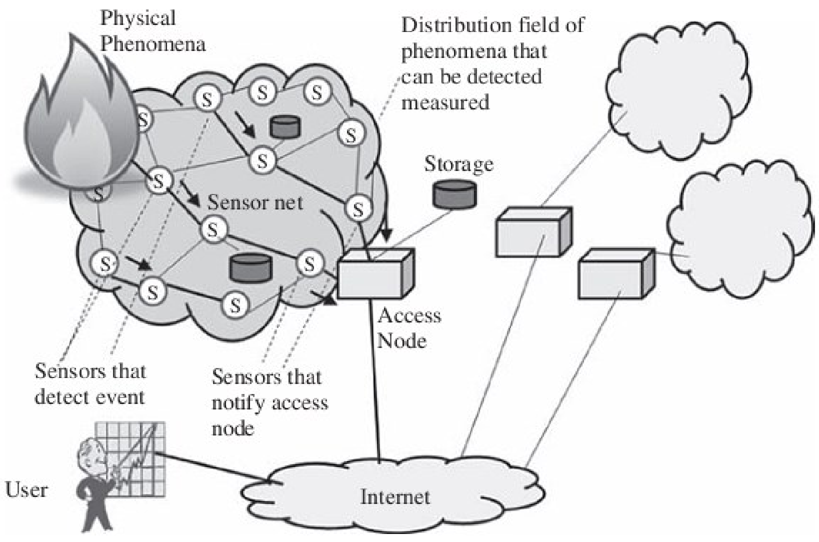


Figure 6.5 A sensor network used to detect increases in heat and report these to a user

Sensors can be characterised in terms of:

- **Characteristics of the phenomena being sensed :** Sensors can be used to sense a variety of physical phenomenon like location, temperature, etc. It can be used to monitor the usage of a resource.
- **Variability:** The type of environment being sensed; application tasks; the spatial distribution, i.e., objects of interest in the environment which may move.
- **Sensor physical characteristics:** Sensors are differentiated based on their power capacity, size and mobility.
- **Functional complexity:** Sensors can be basic that act as input or more complicated like they can be reconfigurable, self-configurable and self-optimising. They can work alone or with multiple sensors or can be part of an embedded control system.

The challenges of designing and deploying sensors and solutions are summarized in Table 6.1.

Table 6.1 Challenges in designing and deploying sensors and some corresponding solutions

Challenges of a sensor net system	Design solutions
Sensor energy is a scarce resource for data transmission	Use a sensor net that deploys, low power, short range transmissions Network sensors into mesh networks and use multi hop transmissions Filter data in situ and transmit only filtered data
Limited memory and computation power in sensors	Harvest renewable energy from the environment and store
Dynamic and non deterministic spatial temporal distribution of events. May not be able to pre determine how to optimally deploy individual sensors	Use a sensor net to increase the sensor density around estimated signal source positions when deterministic; Design sensor distributions to be reconfigurable, self organising, to be mobile Support variable sampling and support bursty data collection
Sensor failure is common due to a lack of power, physical damage, active (jamming) or passive environmental interference of the transceiver, access node or non optimal positioning	Use dense networks of low power sensors with redundant paths to route data through the network Use of counter measures and frequency shifts Locators and trackers are needed to locate (moving) sensors and can be used to position them
Multi hop sensor networks may have a dynamic topology. No global knowledge about structure of network may be known	Use specialised routing protocols to work over dynamic mesh topologies
Sensors can be too costly to update once deployed	Design sensors and sensor access nodes to be low maintenance. Support sensor redundancy
Sensors can generate huge quantities of data	Use in situ data processing both in the sensor and the sensor access node

Ubiquitous Computing: Smart Devices, Environment and Interactions

Sensor functions can be layered in a protocol stack based on the physical network, data network, data processing and sensor choreography.

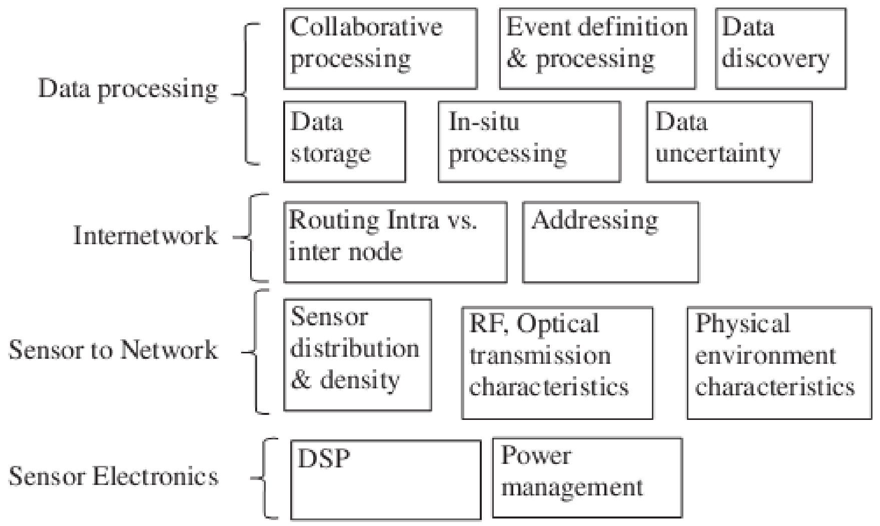


Figure 6.6 The main functional characteristics for sensor net deployment

Ubiquitous Computing: Smart Devices, Environment and Interactions

4.3.2 Sensor Electronics

Sensor processing can be split into four major functions: sensing, processing, transceiving and power. The signals received by the sensors are filtered and amplified, the ADC converts them into digital signals, which is further subjected to some Digital Signal Processing before being modulated for transmission.

MEMS can be used to reduce the size of the circuit and aggregation of multiple components into a single chip. Sensors should be designed such that they can operate unattended, have a long life and low duty cycle.

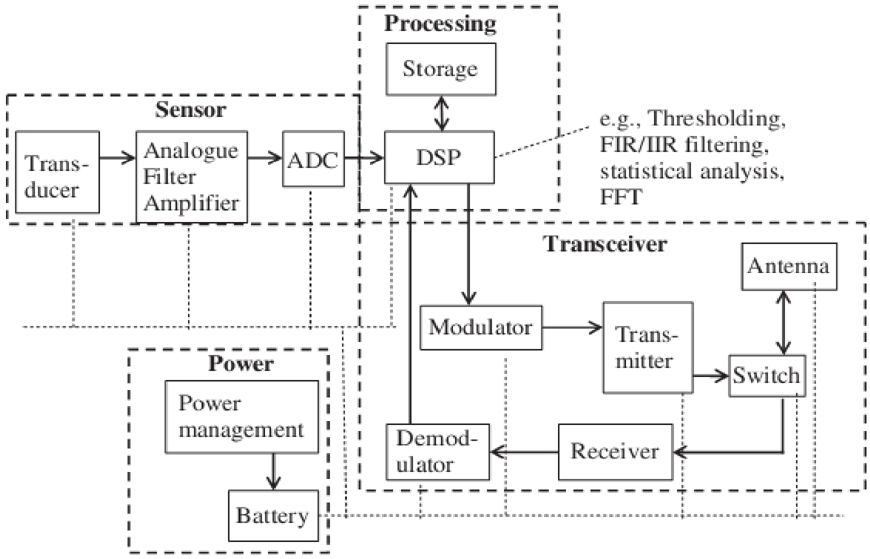


Figure 6.7 Block diagram for a sensor electronics circuit

Ubiquitous Computing: Smart Devices, Environment and Interactions

4.3.3 Physical Network: Environment, Density and Transmission

(#ref: Stefan Poslad - Ubiquitous Computing: Smart Devices,Environment and Interactions, pg 213-214)

Sensors may be deployed in three phases:

(1) pre-deployment phase – placement of sensors in individual or multiple locations

(2) deployment phase – sensors may move because of signal phenomena changes, energy optimisation or task changes

(3) redeployment phase - additional sensors are placed.

Multiple low-cost, low-power, short- range sensors are preferable over few long-range, high-power, high cost sensors because:

Dense networks improve SNR. The distance between the source of interest and sensor is a very important factor for getting accurate signals.

Consider acoustic sensing in a two dimensional plane. The acoustic power received, Pr, at distance d from a power source, Ps, varies inversely to the square of the distance between them (Equation 4.1):

$Pr \, \alpha \, Ps/d^2$ (4.1)

The SNR of the power received signal to the noise level signal power level, Pn is expressed in a logarithm decibel scale:

SNR = 10 log(Pr/Pn)=10logPs 10logPn 20log d (4.2)

Increasing the sensor density by a factor of k reduces the average distance to target inversely by the square root of r. Thus, SNR advantage of denser network is given in Equation 4.3.

SNRdif=$SNR(\frac{d}{\sqrt{k}}) \;\; SNR(d) = 20\log(\mathrm{d}/\left(\frac{d}{\sqrt{k}}\right)) = 10\log k$ (4.3)

Thus, an increase in the sensor density by a factor of k improves the SNR at a sensor by 10 logdB for acoustic type signals. Sensors need to be distributed in such a way that they can maximise the coverage areas and be local enough to detect strong enough signals from the physical phenomenon of interest. They must be distributed so that there is some overlap between adjacent sensors network coverage so that they find a data transmission path to a sensor node. The distribution and coverage of the sensors should be arranged to match the distribution and coverage of the physical phenomena of interest in order to optimise detection. However, using a uniform sensor distribution may not always be optimal, for example, because signal attenuation and obstacles are non-uniform. Increasing the sensor density may have no effect when the density of obstacles is similar to the density of sensors. It may not always be possible to pre determine an optimal sensor deployment distribution. Hence, there is a degree of sensing uncertainty. Random deployment could also be used, e.g., in inaccessible terrains or disaster relief operations. In addition, sensors that are mobile may be able to self organise themselves into an optimum configuration. Second, the energy efficiency for communication can be increased through the use of a multi hop topology for the sensor network. In an N hop network, overall distance for transmission is Nd where d is the (average) one hop distance. The minimum receiving power is Pr and the power at the transmission node is Ps; a is the RF attenuation coefficient which is typically in the range 2 5 because of multipath and other interference effects:

$$Pr \ \alpha \ Ps/d^a \qquad (4.4)$$

$$Ps \ \alpha \ d^a Pr \qquad (4.5)$$

Therefore the power advantage of an N hop transmission versus a single hop transmission over the same distance Nr, Pdif, is given in Equation 4.6:

$$Pdif=Ps(Nd)/(N.Ps(Nd))=(Nd)^a\ Pr/(N.d^a\ Pr)=N^{(a-1)} \qquad (4.6)$$

However, the power increase must be balanced against the disadvantages of using more relay nodes: the increased power use by all the components, the increased cost in using more sensors and the increased latency in forwarding messages over multi hops.

In addition, power management can also be supported by optimising routing or processing management to consider powering down transceivers in redundant sensors and sensor routes. Relevant information can be aggregated during multi hop data exchange and if sensors support data exchange via multi paths, the system has some resilience against individual sensor node failures.

4.3.4 Data Network: Addressing and Routing

A distributed system is made up of multiple nodes and each node needs to be identified uniquely. Usually the nodes are named based on the location of the node in the network. In fixed topology physical networks, the nodes are assigned an IP address and usually hierarchically distributed. New nodes are discovered using node discovery which is based on attribute based addressing scheme, independent of the network topology. Service level networks use multiple levels of indirection.

Directed diffusion is an addressing scheme for sensor networks that supports in network processing to leverage CPU communications trade offs for sensor networks. It reduces the number of indirections and operates directly over low level (hop by bop) communication protocols.

Directed diffusion is data centric routing. It names the data (not nodes) with externally relevant attributes such as data type, time, location of node and SNR. It supports network aggregation and processing of data sources.

A node in a sensor network plays multiple roles like being a source a data along with aggregating, combining and processing data from other nodes. A node may act as a server that publishes data when a condition arises or as a client for triggering event data (detecting a fire).

Routing algorithms for sensor networks can be classified according to type of network structure (flat or hierarchical) or according to interaction protocol (multipath, query based or negotiation based).

A number of routing protocols are used in sensor networks like:

- **SMEC -** Creates a subgraph of the sensor network that contains the minimum energy path.
- **Flooding -** Sends data to all neighbour nodes. regardless if they receive it before or not.
- **Gossiping -** Sends data to one randomly selected neighbour.
- **SPIN -** Sends data to only interested sensor nodes.
- **SAR-** Creates multiple trees where the root of each tree is one hop neighbour from the access node.
- **LEACH -** Forms clusters to minimise energy dissipation.

4.3.4.1 Sensor Net works Versus AdHoc Net works

Sensor networks are dense and are composed of larger number of node of the same type. Hence, each node may not be uniquely addressable.

They have a dynamic topology. In order to optimize a low SNR and power usage, nodes need to adapt to locally detected sensors, to unknown signal characteristics, to the (self) reconfiguration of in situ nomadic sensors and to dynamic routing.

Sensor nodes broadcast data whereas ad hoc networks transmit to a specific destination. Sensor networks use a dynamic, mobile, and unreliable mesh topology. They do not have a universal routing protocols and central registry of sensor locations and routes. Each sensor can be a router, a data source and/or a data sink and gateway.

4.3.5 Data Processing: Distributed Data Storage and Data Queries

Sensor networks are made up of many sensors distributed over a network. The processing is also distributed as it involves collecting events from all the sensors and aggregating it. Althogh, each sensor performing independently would be easier, it might also require a large amount of storage.

Sensor data can be stored in a RDBMS server which is further aggregated and dumped in a data warehouse. SQL can be used to query sensor data. However, many events may contain little information of value because readings are constant. Another approach is to filter data at each node and store only filtered events in the database. Sensor database shouled support distributed query processing over sensor network.

Sensor nodes can be clustered using attribute based sensor addressing or applying recursive data aggregation process. e.g., the SINA (Sensor Information Networking Architecture).

In SINA, SQTL (Sensor Query and Tasking Language) acts as an API between between sensor applications and the SINA middleware. It supports three types of events:

(1) Event generated when a message is received by a sensor nod e;

(2) Events triggered periodically by a timer; and

(3) Events caused by the expiration of a timer.

Node collaboration is required to filter out duplicate events and to switch off active sensors that produce or forward redundant or non-significant information leading to wastage of power and bandwidth resources. An isolated event may not be significant or may be too easily recognised as a false positive or false negative. This is information based sensor tasking.

4.4 MICRO ACTUATION AND SENSING: MEMS

MEMS (Micro electro mechanical systems) are small sized electronic devices (micron to millimeter sized) which operates independently or are part of an array. They can be used as sensors or actuators.

MEMS actuators convert an electrical signal into physical phenomena that involves movinf or controlling motors, pneumatic actuators, hydraulic pistons and relays. MEMS sensors convert some environmental phenomena such as temperature, humidity and pressure into an electrical signal.

MEMS are small components with high operating frequencies.

MEMS can be considered as a form of programmable matter because these can self-assemble in arbitrary three dimensional shapes. MEMS can be used to achieve a synthetic reality which allows the physical realisation of all computer generated objects, naturally without any sensory augmentation(head mounted displays).

4.4.1 Fabrication

MEMS are based upon silicon based Integrated Circuit (IC), also called chip, design. e.g., A traditional thermometer is based upon a liquid, such as mercury, expanding along a tube referenced to a calibrated scale whereas an electronic thermometer can be built out of a thermocouple and IC amplifier.

ICs consist of several layers of p-type and n-type doped silicon added to a substrate. Photolithography is used to fabricate the circuit. The ICs are first covered with a photoresistant chemical. The pattern to be fabricated is drawn on a photomask. The photolithography systems shines the UV light through the photomask, projecting a shadow onto the layer. The chemical layer that comes in contact with the UV light reacts and hardens. The remaining chemical is loose and can be removed. MEMS have the advantage of power reduction and can be batch produced.

Components in MEMS are closer together and hence they can be switched quickly leading to lower power consumption and better performance.

MEMS can also be used to produce micromachines. They can be fabricated in different ways using: bulk micro machining (etching into the substrate); surface micro machining (building up layers above the substrate and etching); and by machining LIGA deep structures.

4.4.2 Micro-Actuators

Micro-actuators are different from macro actuators because they function differently. They are engineered in a different way using IC design and nanotechnology.

MEMS actuator applications include:

- Micro mirror array based projectors (micro projectors) -Used to project content from smaller devices on a large screen display. They can be used in navigations systems (Heads Up Displays or HUDs), medical diagnosis and treatment and manufacturing to produce reference points for drilling.
- Inkjet printers heads : MEMS can be used to control ink deposits onto paper. MEMS devices can also be printed on paper using an ink jet printer.
- Optical switches: Optical cross connect switches (OXC) are used to switch high speed optical signals in a fibre optic network.
- Micro fluid pumps: They consists a fluid actuator, a fluidic control device, and micro plumbing. They should be made from materials that biocompatible, e.g.,Parylene.
- Miniature RF transceivers: can replace passive low Q by putting them on a single high Q MEMS RF transceiver chip which greater miniaturisation of communicators. Q indicates the rate of energy dissipation relative to the oscillation frequency.
- Miniature storage devices: A singe IC chip can be used to store gigabytes of non-volatile data which has low power, and low data latency.

4.4.3 Micro-Sensors

Micro-Sensors are small, inexpensive and easy to deploy giving suitable performance with respect to sensor functionality and communications capability. These millimetre scale nodes are called ‘Smart Dust’.

Smart Dust motes can be circulated by air currents, remain suspended in air and can sense and communicate for hours or even days. They contain micro sensors, an optical receiver, passive and active optical transmitters, signal processing and control circuitry, and a thick film battery power source. Efficient power management is an important factor in design of smart dust. Stored energy is about 1J targeted to be consumed at 10 mW throughout the day.

Power consumption by a power optimised CPUs is 1 nJ per 32 bit instruction. RF data transmissions consume a lot of power, e.g., Bluetooth radio frequency (RF) communication chips use about 100 nJ per bit transmitted. Micros sensors have a very limited space for antennas hence they use need high frequency data transmission.

MEMS sensor applications:

- Accelerometers detect the rapid negative acceleration of a car to determine when a collision is occurring and the severity of the collision. Hence, it is used to control the safety airbag release in cars.
- Angular rate sensors and gyroscopes can be used to measure the rotational velocity or angular rate of an object.

4.4.4 Smart Surfaces, Skin, Paint, Matter and Dust

MEMS can either be attached permanently to a surface making smart surfaces or can re-organise forming smart structures. A smart surface is used to create a smart environment. A surface that can sense vibrations is an example of a smart surface. A fine powder of a piezoelectric material, called lead zirconate titanate (PZT), is used to paint the surface. PZT crystals produce an electrical signal that is proportional to the force when they are stretched or squeezed.

MEMS can be mixed with a lot of materials such as paints or gels to be spread on surfaces or they can be embedded into surfaces as well as carried on air or water.

MEMS can be painted on bridges and buildings to sense and report traffic, wind loads and monitor structural integrity. It can sense vibrations, monitor the premises for intruders, and cancel noise. MEMS can be woven in organic polymers that have light emitting and conductive properties to form smart skin and smart clothes. MEMS can be networked in MEMS nets.

The Claytronics Project proposed using masses of thousands to millions of sensor, actuator and locomotion MEMS devices that behave as malleable programmable matter and can recreate artefacts for a wide range of physical shapes and objects. This is smart matter.

Smart matter should be self assembled in any arbitrary 3D shape, to achieve a synthetic reality. It should be based on nanotechnology so that to matter can be engineered on the molecular level. Smart dust and smart matter are also referred to as amorphous computing and spray computing.

4.4.5 Downsizing to Nanotechnology and Quantum Devices

Moore's Law states that the number of transistors on an IC chip doubles about every two years. This does not mean that the communication and software performance increases in the same proportion. The reason why software performance does not increase is that the time taken for changes in the computation architecture may be 5-10 years. Also, the communications capability also takes time to increase. Hardware capability can be scaled easily by increasing the density of IC chips. IC chip sizes are decreasing three times every six to seven years fue to advancements in photolithography. The size of the chips can be reduced by using smaller marks which are based on shorter wavelength of the light used.

Extreme ultra violet photolithography, to generate shorter wavelength light of 13.5 nm, may lead to delays in increasing the transistor density. Thin layer of insulation for the transistor gate is breaking down due to size reduction. Hence, nanotechnology can be used for new materials that can be modelled and designed at the molecular level. Nanotechnology can be defined as the manipulation, precision placement, measurement, modelling, and manufacture to create systems with less than 100 nm.

Nanotechnology is based upon a broader range of materials and mechanisms and sizes down to the molecular level whereas MEMS focuses on semiconductor based single IC chip technology and micro machining.

Nanotechnology uses a Bottom Up approach whereas MEMS use a top down approach for designing.

Nanotechnology is used to make smaller electronic components. However, when electronic components approach the nanometer size, electrons begins to reveal their quantum nature, e.g., electrons have the potential of crossing a transistor even if it is switched off. Secondly, thermal noise causes local molecules' movement because of heat. This phenomenon limits what can be done mechanically at the molecular scale.

Nanotechnology uses a bottom up approach i.e.to assemble custom made molecular structures for specific applications. The biggest challenge is to understand and model materials at this level. A lot of study related to combinations of materials, in particular compounds and their physical and functional properties needs to be conducted.

Nanotechnology requires two main types of engineering support:

- molecular positional assembly – It deals with manipulation and positioning of individual atoms

- massive parallelism – so that multiple molecules can be worked upon simultaneously for faster build time instead of working on one molecule at a time.

4.5 EMBEDDED SYSTEMS AND REAL-TIME SYSTEMS

An embedded system is a component in a larger system that performs a single dedicated task. They are designed to perform a single task as opposed to general purpose computers. They differ from general purpose (MTOS) systems in three main ways:

(1) They focus on single task enactment.

(2) They are Safety critical

(3) They work on real time

Embedded systems work on only one process so they need a simple and cheaper operating system and hardware. An embedded system may or may not be visible to a user. It may or may not have a visible control interface for human users. It can be local and fixed or may be mobile and distributed. It is programmable and contains one or more microprocessors or microcontrollers.

A microprocessor is an integrated circuit, which is the CPU for a computer or embedded controller. A microcontroller is a microprocessor with additional types of processor that supports other devices and is integrated into a single package. Other types of device support devices may include serial (COM), ports, parallel ports, USB ports, Ethernet ports, A/D and D/A, interval timers, watchdog timers, event counter/timers, real time clocks (RTC), Digital Signal Processing (DSP) and numeric coprocessors.

Embedded system may contain programmable logic elements such as FPGA (Field Programmable Gate Arrays) or application specific integrated circuit (ASIC).

They use a rich variety of microprocessors, hundreds of types that are dedicated to a specific task or tasks, e.g., peripherals, networking, etc. Every embedded design (hardware and software) is unique and may be designed for its own rigidly defined operational bounds, e.g., heating system.

They are designed for the highest possible performance at the lowest cost and often operate under moderate to severe real time constraints.

Embedded systems are safety critical hence they are designed to be fault tolerant. They have constraints with respect to power consumption. They are designed to operate in wider range of physical environmental conditions like damp, hot, cold, wet and dark conditions. They might lack storage capacities and might not have a common operating system. They might need specialist development tools.

Modern cars network multiple embedded systems for antilock brake systems (ABS), cruise control, climate control, wing mirrors, locomotion and drive sensor data monitoring, etc.

Different embedded sub systems can be interconnected into a holistic control and monitoring system for example Controller Area Network (CAN). A vehicle can contain two or three separate CANs operating at different transmission rates. A low speed CAN, running at less than 125 Kbps, is not safety critical and works on energy saving sleep mode. It is responsible for managing a car's comfort electronics for seat and window movement controls. A higher speed CAN runs real time critical functions such as engine management, antilock brakes, and cruise control.

Some embedded systems can be hard-linked to specific networked services. For example, electric appliances used at home can be designed to alert the manufacturer's service center when it malfunctions or needs a service. ATMs, POS terminals, information and entertainment devices and GPS are examples of networked embedded devices.

4.5.1 Application- Specific Opera ting Systems (ASOS)

Embedded systems usually perform a single task hence they require an application specific operating system (ASOS), which is customized and reconfigured as per the needs of the application. ASOS give a higher performance at a low cost by eliminating unnecessary processes that are usually part of a general purpose OS. Data compression techniques can be applied to reduce the data storage requirement. A key design issue is whether the ASOS can be updated post deployment to accommodate new requirements.

4.5.2 Real-Time Operating Systems for Embedded Systems

Real time embedded systems are a subset of embedded system applications which perform extremely safety critical tasks. Failure of a safety critical systems may cause the system to become unsafe or dangerous and can cause a big financial loss or worst a loss of life. Cars, ships, airplanes, medical instrumentation monitoring and control, multimedia streaming, factory automation, financial transaction processing and video games machines are examples of safety critical systems. They work in extreme constraints and cannot fail.

Real Time Operating Systems or RTOS is a system constrained in time. It might lock specific process to prevent process swapping overhead. It reacts to external events that interrupt it based on priority. Its focuses on scheduling so that processes can meet real time constraints and optimize memory allocation, process context switching time and interrupt latency. The two key factors that affect the response time are process context switching and interrupt latency.

The timing of a result is just as important as the result itself. A correct answer produced too late is just as bad as an incorrect answer or no answer at all. If the timing constraints are not met, it is considered as a system failure.

There are three categories of real time systems: soft, hard or firm.

Soft RTS- Response times are not very strict in soft RTS. The timing requirements are defined by using an average response time. A single delayed response does not make much of a difference whereas multiple delayed responses can affect the system. An airline reservation system is an example of a soft RTS.

Hard RTS – Timing requirements are extremely vital. A late response is considered as a system

Failure. Activities must be completed within a specified deadline, that can be a specific time, a time interval, or the arrival of an event; to avoid failure. Prediction of computation times can help in avoiding failures. Pacemaker is an example of hard RTS system where late response can inflict biological damage.

Firm RTS – The timing requirements are a combination of both hard (longer) and soft (shorter) computation times. Ventilator is an example of firm RTS. The system is allowed a few seconds delay in the initiation of a patient's breath however it must ventilate a patient given number of times within a given time period.

4.6 CONTROL SYSTEMS (FOR PHYSICAL WORLD TASKS)

Ubiquitous systems are smart when they are self-controlling systems. The most basic form is control is activated when defined thresholds are crossed, e.g., a geyser switches off once it reached an upper threshold. A thermostat is another example where the control switches the heating on when the temperature falls below the lower threshold or switches the cooling on when it rises above the upper threshold.

The outputs are regulated by Feedback control which continuously monitors the output using sensors. There are two basic kinds of feedback: Positive and Negative. Negative feedback reduces the system state or output whereas positive feedback amplifies the system state or output.

Negative feedback can combine a derivative of the output with the input to regulate the output. In a simple proportional control system, a signal is negatively fed back to a system in proportion to the degree the system output diverges from the reference value.

4.6.1 Programmable Controllers

Programmable controllers are used to support much more configurable and flexible control, e.g., microcontrollers. The hardware architecture of microcontrollers is much simpler than that of the more general purpose processor mother boards in PCs.

Microcontrollers architecture is very simple. It integrates all the RAM and non-volatile memory on the same chip(4-bit) as the CPU. Microcontrollers do not have display hence their I/O control is simple. The complexity in microcontrollers come from additional components like digital signal processors (DSPs).

Microcontrollers were programmed in assembly language. However, they can now be programmed in high level languages like C. The control programs are developed on an emulator in a PC environment, which is written on the target device for execution, maintenance, debugging and validation. More recent forms of microcontrollers allow the program to debug the software directly on the chip using chip debug circuitry accessed by an in-circuit emulator with a JTAG interface. JTAG is special four or five pin interface added to a chip, designed so that multiple chips can have their JTAG interfaces pipelined together. It can also be used as an FPGA (Field Programmable Gate Array) programming interface.

Threshold control systems are event driven i.e. they are event condition action or ECA systems. The JTAG (Joint Test Action Group) refers to the IEEE 1149.1 standard entitled Standard Test Access Port and Boundary Scan Architecture.

4.6.2 Simple PID-Type Controllers

A PID controller combines Proportional, Integral and Derivative type control. A proportional signal is the error signal multiplied by a constant. It is fed to a hardware drive. An integral controller works on the past history of the control. It is used to add long term precision to the control loop by adding all the preceding inputs. It is is always used in conjunction with the proportional controller. A derivative controller is used to stabilize the control by predicting plant behavior using feedback loop.

PID controllers are used in situations where P-type controller output is not regulated correctly, e.g., There is a lot of delay in the response time of the plant when the input is changed. Algorithms like P, PI, PD or PID can be easily hard coded in the controller with minor adjustment controls.

4.6.3 More Complex Controllers

PID controllers can be used for coarse control of robots' arms and locomotion. They are based on the assumption that some prior-knowledge of the plant to be controlled exits based on which approximating models can be formed. These models can further be tuned using feedback to cope with the uncertainties of the plant parameters.

These algorithms are not well suited to tasks that need finer or dynamic control, dealing with uncertainties in the control and the need to support adaptive control.

Sources of uncertainty in a control are:

- **environment dynamics :** stochastic form of physical objects that naturally varies
- **random action effects :** actions can have different effects because of variations in physical characteristics of things
- **inaccurate models :** some characteristics of the physical environment are inherently non deterministic;
- **approximate computation :** e.g. nonlinear behavior is often approximated by a linear solution in a part of its range;

- **sensor limitations:** improper sensor placement, partial views of world, and poor signal reception, see Table 6.1.

Adaptive control and robust control are two of the main techniques for controlling uncertain systems. Adaptive control uses online identification. It identifies the plant parameters of interest using error predictions (indirect adaptive control), or by monitoring the parameters ,adjusting them and tracking errors (direct adaptive control). Adaptive control is a nonlinear feedback technique that performs identification and control simultaneously. It is used to deal with systems with large uncertainties but is sensitive to unstructured uncertainties. A robust controller is designed to make the system insensitive to all uncertainties. The final controller has a fixed structure. A robust controller is suitable for dealing with small uncertainties.

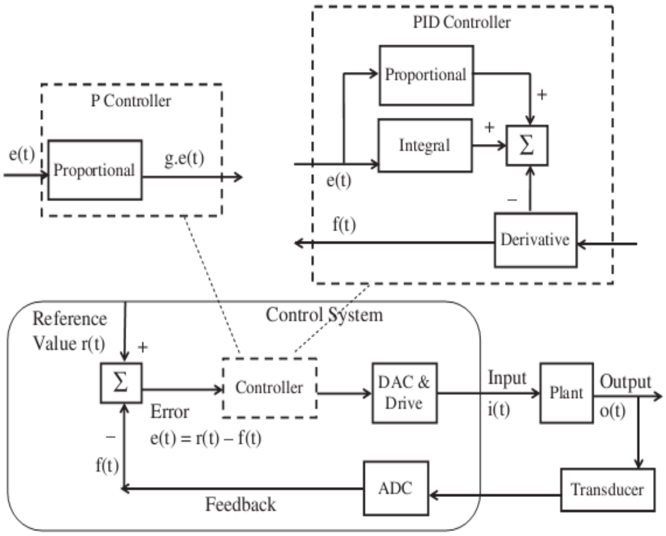


Figure 6.9 Two simple control systems: a proportional type controller (top) and a PID type controller (bottom)

#imgref: Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions

4.7 EXERCISES

1. Explain the life cycle of tagging.
2. Describe and discuss Active and Passive tags
3. What is a PID Controller?
4. Explain a Sensor net Architecture with an example.

4.8 REFERENCES:

1. Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions- Chapter 6

❄❄❄❄❄❄❄

5

UBIQUITOUS COMMUNICATION

Unit Structure

5.0 OBJECTIVES

To understand the different types of networks that can be employed for creating a ubiquitous network.

5.1 INTRODUCTION

Ubiquitous computing means accessing information anywhere and at any time. This is possible only by means of communication networks that are accessible everywhere. Although, ubiquitous networks are designed with the communication access as its part , it is dependent on the external network to be accessible everywhere.

Different Ubiquitous applications have different network and service requirements. Some of the challenges are whether the external networks can be configured or controlled, can they be exposed via some interfaces and whether we need to be connected to the network all the time or intermittently.

Ubiquitous communication is made up of various types of networks that deal with different forms of communication. Data communication focuses on the communication of alphanumeric data. Telecommunication focuses on voice communication. Broadcast audio video networks use separate radio and TV networks or wireless networks.

5.2 AUDIO NETWORKS

Audio networks were the first types of networks that became part of our lifestyle. There are two basic types of audio networks: audio unicast networks (PSTN) and an audio broadcast (radio) networks.

5.2.1 PSTN Voice Networks

PSTN stands for Public Switched Telephone Network(PSTN). It is designed to support voice communication. In the beginning, the PSTN used analogue transmissions for communication. However, it was later on replaced with digital transmission, as it is more cost effective.

However, the edge of the network

that connects to homes and business, the local loop, still remains analogue in most regions. Many workplaces today still use separate networks for voice and data although there is a progression towards combined voice, data and audio video networks.

PSTN consists of Telephones as access devices. In a commercial setup,these are connected to a private circuit switched network or Private Branch Exchange(PBX). The PBX is used to control access to a smaller number of external connections (extensions in an office).

For residential use, the telephone is connected to a PSTN using a single line local loop that is further connected to a local switching station.

A call once placed goes through the mobile switching office before being connected to the destination telephone number. With the changing technology, PSTN can also be accessed using fixed(landline) or mobile phones. It can also be used to connect to the Internet as well as to connect to a LAN. The LAN can be further connected to an external network via a router managed by an Internet Service Provider (ISP).

PSTNs use circuit switching. Circuit switching is having a dedicated end-to-end link between devices for the entire duration of transmission. The link is released once the communication is terminated.

PSTNs are extremely resilient. A typical call is handled by a hierarchy of five levels of switching offices where each switching office handled a range of telephone number ranges. A typical call in circuit switching goes through three phases: Setup, Transmission and termination. In the setup phase, a dedicated link is established between the source and the destination number. If the number is from a different range, it is passed through a hierarchy of switching office until a higher level office can connect to the range of the destination telephone number. In some cases, some switching offices on the same level may also be cross linked to prevent the need to pass calls up to a higher level switching and to provide greater resilience through alternate paths. Once the destination number is located, a ringtone can be heard which confirms the successful establishment of connection. A dial tone is basically the network trying to connect to the destination device. This link is used for voice communication between the parties until the duration of call. The data sent during the duration of call is the transmission phase. The advantage of having a dedicated link is that it is naturally easier to maintain a higher QoS. However, it leads to the wastage of resources when there is a pause in communication.

Digital transmission allowed the optimum utilization of the bandwidth by multiplexing multiple communications together. Multiplexing is combining data from multiple devices into one channel. Hence, data could be sent along with voice communication, which was not possible with analog transmission.

Telecoms networks support global interoperability and standards that allow users in one world region to (voice) call users in other regions.

5.2.2 Intelligent Networks and IP Multimedia Subsystems

The initial voice based communication networks were not easy to be customized for new services. The Intelligent Networks(IN) network service model, standardised by the ITU-T enabled telecommunication networks to offer new value added and customised voice services to users such as toll free calls('0800' numbers). The advantage of IN is that it supports component based services in the existing general-purpose computer nodes eliminating the need for special switching nodes. Thus, the service providers can easily introduce new services using existing infrastructure networks.

Active development in new IN services has declined in recent years as the emphasis now-a-days is more on the development of telecom services and APIs rather than on developing new telecom network protocols. Hybrid IN and Internet service architectures for mobile users are being proposed such as IP Multimedia Subsystems (IMS). It was originally developed for 3G wireless networks but now it also supports WLAN and fixed networks. Users can access IMS using IP.

A key challenge for the application layer is the ability to handle voice/video session, multimedia conference, messaging and Presence over IP. The IETF SIP (Session Initiation Protocol) supports these services.

A typical SIP system consists of :

- A series of mediators –where the users sign in and are authenticated,
- Location servers - used to track user locations,
- Presence servers- to detect if users are active,
- Proxy servers and redirect servers – to assist in call forwarding
- Multi-point Control Unit(MCU) – to mix multimedia streams.

SIP can use three different types of MCU: full mesh, mixer and multicast. In a full mesh, every participant builds a signalling path with every other participant. It sends an individual copy of the media stream to the others. This is used only in very small groups. A mixer or bridge combines several media streams. It uses 2Mbit/s demodulation and supports Multimedia Access using the DECT Multimedia Access Profile (DMAP). corDECT is a low cost digital system based upon DECT used to provide wireless voice and data services in rural areas.

DECT and Bluetooth are similar in terms of their applications and market share. Majority of LAN cordless household telephones use DECT while the majority of WAN mobile phones, games consoles, etc. use Bluetooth in UK. Bluetooth is preferred over DECT in a wide rane of devices.

5.2.5 Audio Broadcast (Radio Entertainment) Networks

Broadcasting means sending it to all. Audio broadcast consists of one sender and many receivers. The FM radio that we listen to is the best example of an Audio Broadcast Network. Using Radio network for transmission has several benefits.

Audio broadcast supports one sender to many receivers. It supports a multi modal interface i.e. a human being can perform their activities while listening to an audio broadcast. Eg: You can drive, sweep or cook while listening to an audio broadcast. Radio receivers can be tuned to receive a wide spectrum of RF broadcasts on multiple channels. Analogue radio has now-a-days been replaced by digital radio. The Eureka 147 Digital Audio Broadcast (DAB) standard is the most commonly used standard, which is coordinated by the World DMB Forum for a digital radio. It uses the MPEG-1 Audio Layer 2 audio (MP2) codec for audio broadcasting while personal players use the MP3 codec.

The main objectives of DAB were:

(1) to provide radio at CD quality;

(2) to provide better in-car reception quality as compared to a FM analogue radio;

(3) efficient use of the spectrum;

(4) to allow tuning by the name of the station rather than by frequency;

(5) to allow data to be transmitted.

5.3 DATA NETWORKS

Internet has brought about a lot a change in the communication technology. Internet differs from the traditional communication model in various ways.

First- The shift from (single tasking) batch computers to (multi tasking) time shared computers. Second- The direct connection between computer nodes in the form of a peer-to-peer (P2P) mesh network. This setup was however changed due to scalability issue. Now, the computers were connected via intermediate nodes, dedicated network computers, known as Interface Message Processors . The IMPs did not deal with permanent storage of data but simply stored and forwarded data.

Third-The shift from analogue to digital communication. It was done to avoid signal degradation across multinode networks.

Fourth- a network which supports high capacity and resilient network paths.

Fifth – Data was split into fixed size manageable data packets.

Sixth- The shift from a circuit switched Telecoms network model to a packet switched data model. Packet switching is dividing the data into manageable chunks called packets. The packets take different routes to reach the destination. There is no dedicated link like circuit switched network.

5.3.1 Network Protocol Suites

Protocol is a set of rules that govern data communication. Protocols are useful for standardizing the different aspects of communication like the size of the packet, the address type, the window size, etc. Network protocols usually define the size in which the entire data is to be divided called as the packet size. Packets make it easier to handle data while transmission. The packet size can differ with respect to the underlying network. If the size of the packet is bigger than the network capacity it is split into smaller packets. This is called Segmentation. There should be synchronization between the sender and receiver while transmitting the packet. Hence, a specialized set of packets called as synchronization (SYN) and acknowledgement (ACK/NAK) packets are used. Every packet has a sender and destination address which enables interleaving of packets from different devices together. This is called as multiplexing. The packets can travel through a network having a different IP version or protocol. In such cases, the packets are encapsulated with a new header specific to the network. This is called as tunneling. It hides the complexities related to the transmission.

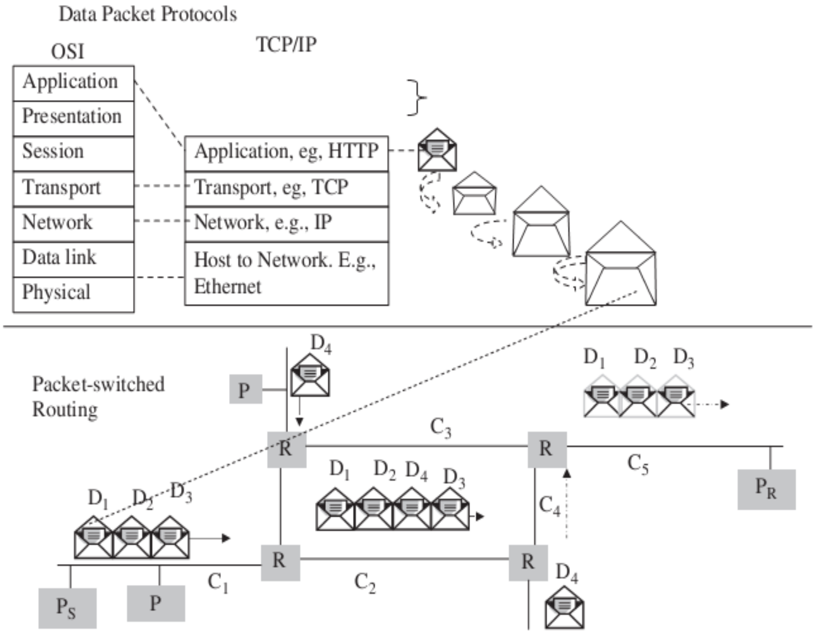


Figure 11.1 Data messages for an application are fragmented into packets D_1 to D_3 for delivery across distinct communication networks C_1 to C_5. Data protocols can be combined to encapsulate data corresponding to a higher level more complex protocol and to map data into a simpler lower level one

#imgref: Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions

5.3.2 Addressing

Every device on the network is assigned a unique address that will identify it on the network. The addresses can be allotted in a flat manner or in a hierarchy called as subnets or supernets. A router is a device which reads the address on the packet and forwards it to the destination. The address given to a device on the internet is called as in IP Address. An IPv4 address is 32 bit in size. It can be represented as four decimal numbers separated by dots (ex: 10.1.1.20).The numbers can be in the range of 0 to 255. An IP address consists of two parts: network address and host address. The host address identifies the device whereas the network address helps in identifying the network a particular device belongs to. IPv4 supports 2^{32} combinations of IP address. However, with the increase in the number of devices on the internet, IPv4 falls short of the number of addresses needed. IPv6 is 128 bits and support 2^{128} addresses. IPv4 is still the most widely used address owing to the different ways for address reuse. However, we might run out of it and then we will be forced to use IPv6.

5.3.3 Routing and Internetworking

The internet has enabled communication on a wider scale. We can easily communicate with devices outside our domain or devices that are geographically far off. The data to be transmitted is divided into packets that take different routes to reach the destination. The routers are responsible for forwarding packets. Routers read the address on the packets and depending on the current network status forward the packet on a particular route. They refer to the routing tables which hold the port through which the data is to be forwarded. This is called routing. Routing is done at the network level so the users are not aware of it. Routers keep updating each other about the network status by means of various routing protocols like RIP, OSPF, etc. In the Open Shortest Path First(OSPF) protocol, each OSPF router broadcasts its routes and in turn requests information about routes that it can connect to. This helps in dynamic routing which is aimed at fixed wired infrastructure networks. Since, the packets in a particular data transmission can take different routes to reach its destination, packets may arrive out of order or may be lost in transmission. The lost packets need to be retransmitted and once all the packets have arrived they need to reassembled. Data networks use a type of simple unreliable Packet Switched Data Network (PSDN) protocol called Internet Protocol (IP).The IP protocol is unreliable hence it is used in conjunction with another reliable protocol named the TCP protocol which works at the transport layer. The combination protocol is called

TCP/IP which provides reliable and sequenced data delivery. TCP/IP uses retransmission, acknowledgement packets, timeouts and windowing protocols to ensure reliable data delivery. TCP/IP is connection oriented yet connectionless. It establishes a connection between the source and destination. Once the connection is established, data transmission starts in the form of packets. This is called as s 3 way handshake. Each packet travels using a different path. The windowing protocol is used to keep a track of packets sent at the source and packets received at the destination. Signaling packets like ACK and NAK are used to acknowledging the delivery of packet or to request a particular missing packet. A packet which is not received within a particular timeframe is retransmitted as it is assumed to be lost. TCP used a variety of mechanisms for efficient data transmission. It uses flow control for adjusting the transmission rate to avoid overwhelming the receiver and congestion control to avoid congestion in the network. TCP is used only for unicast communication. UDP is used for multicasting. TCP needs connection setup between devices. UDP does not perform connection setup. UDP is simply about sending the data to the destination. It does not have retransmission or reliable data delivery. Multimedia streaming uses UDP. The Real time Streaming Protocol (RTSP), Real time Transport Protocol (RTP) and the Real time Transport Control Protocol (RTCP) were specifically designed to stream media over networks. The latter two are built on top of UDP. Reliable protocols, such as the Transmission Control Protocol (TCP) can also be used for more reliable media streaming. However, the media streams may freeze while the protocol detect packet loss and try retransmission. The freezing effect can be minimized using buffering of data for display.

5.4 WIRELESS DATA NETWORKS

Wireless networks have a lot of advantages over wired networks. Some of them are:

Anywhere: Wireless networks are independent of the location unlike wired networks. You can access wireless network anywhere. The only catch is that there is a network access device and the user is in the range or the wireless transmitter.

Mobility: Wireless communication networks can be accessed while in motion. A user can move from one transmitter to another seamlessly. The cost of installing multiple wireless transmitters is much less than laying a wired network.

Less disruptive: The laying down if wireless network does not demand much change in the underlying physical infrastructure. Hence, it is usually used in areas where wired networks would be considered too inconvenient, disruptive or expensive to install, e.g., in old historical buildings and in emergency situations.

Adaptivity: Wireless networks can be scaled up or down depending on the demand on the network. The network coverage can easily be expanded or shrinked with respect to the density of coverage, installation of transmitters and capacity.

5.4.1 Types of Wireless Network

A wide variety of wireless networks are available with respect to the type of infrastructure, the network range, frequencies used, the type of signal modulation to increase channel efficiency and channel sharing, the bandwidth available and power consumption. Networks that are global and cover wide area are infrastructure dependent and use fixed transmitters like the satellite, mobile phone, WLAN or WiFi, WiMax,etc . Adhoc wireless network use the transmitters and routers which are dynamic, e.g., packet radios and sensor nets.

Mobile wireless networks can be

- Global e.g., satellite
- Wide Area Networks covering hundreds to thousands of km, e.g., mobile phone networks such as GSM, TDMA,CDMA
- Local Area Networks(LAN) and Metropolitan Area Networks(MAN) covering 100 m to 5 km, e.g. WLAN
- Personal Area Networks(PAN) covering 1 to 10 m, e.g., Bluetooth, ZigBee and InfraRed.

The range of potential access depends on many factors like the frequency of transmission, the strength of the transmitter and attenuation of the signal by moisture, water and different kinds of solid objects. Use of higher frequency gives greater data transmission rate but suffers from attenuation.

Table 11.1 A comparison of the characteristics of wireless networks used for different kinds of services

Device / Service	Frequency (Hz)	Transmitter ange (M)	Bit rate (bps)	Energy (W) & other factors, e.g., attenuation
Dust:				
Smart Dust (Berkeley) Sensor		1 20KM	1 M (burst)	0.1 nJ/bit
Radio:				
AM	0.5 1.6 M		Analogue	20 KW transmitter
Short wave:	5.9 26.1 M		Analogue	
Citizens Band (CB):	26.9 27.4 M		Analogue	
FM	88 108 M	100 k	Analogue	20 KW transmitter
DAB	174 240 M	80 160 k	128 K	
Television				
Analogue TV	75, 200 M	100 K	Analogue	
Cable TV / channel	6 M	6M / channel	0.5 10 M	
Satellite TV	10.9 14.5 G	40000 k		
Telecoms				
Mobile phones	0.8,0.9,1.8 G	0.1 5 k	50 400 k	
DECT	40 50,900 M	100	56 K	
PSTN	0.3 3.3 k	N/A	15 64 k	
ADSL	0.003 8 M	5 K	0.5 10 M	
Data networks				
Ethernet	20,100,250,600 M	100	10,100,1000 M	
WiFi	2.4 G	100	2 54 M	500 mW
WiMax	2 10 G	8 K	70 M	
Consumer electronics:				
Garage doors, alarms	40 M	10	Analogue	
Baby monitors	49 M	10	Analogue	
Radio control cars	72 M	10	Analogue	
Microwave ovens	2.4 G	0.2	Analogue	Shielded
TV remote control IR	35 K	0.2	Analogue	Line of sight
Other short range:				
RFID	12,13,900 M	0 5	0.1	
Bluetooth	125 135 K	30	1 M	60 mW (active mode)
ZigBee	0.9, 2.4 G	100	20 250 k	1 100 mW (active mode)
UWB	2.45 G	10	0.1 5 G	250 mW
Other long range:				
Wildlife tracking	215 220 M			
Air traffic control radar	960 1215 M	50 100 K	Analogue	
GPS	1.2 1.6 G	40 M		Line of sight
MIR space station	145 437 M	40 M		
Deep space radio	2.29 2.3 G	>400000 K		20 KW transmitter

#imgref: Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions

Spatial Efficiency (SE) in Bits per Second per unit area is a useful metric to describe the data rates within a local area. As power consumption is a constraint for mobile wireless transceivers, a power efficiency metric, Bits per Second per Watt, is used.

With the proliferation of new wireless services offered over multiple networks, allocation of radio frequency spectrum and support for multiprotocol wireless networks is a design challenge. The efficient use of RF spectrum is possible by using various techniques like smart antennas, smart modulation and digital signal processsing, multi user detection, ad hoc networking and software radio.

Software radios move the radio functionality from hardware into software. It moves analogue/digital conversion as close to the receiving antenna as possible. It performs software processing instead of hardware processing. It facilitates a transition from dedicated to general-purpose hardware.

5.4.2 WLAN and WiMAX

Wireless LANs (WLANs) are local area wireless networks that adhere to the IEEE 802.1l set of standards. There are various versions of 802.11 that support different data rates. The 802.11g standard ratified in June 2003 works in the 2.4 GHz band and supports a maximum raw data rate of 54 Mbps.

A typical WLAN network is made up of computers with WLAN cards that connect to the WLAN access points. These access points have a wired connection to the internet.

Wi-Fi is a registered trademark of the Wi-Fi Alliance. The Wi-Fi Alliance is a trade organisation that tests and certifies equipment compliance with the 802.11x standards.

The Worldwide Interoperability for Microwave Access(WiMAX) belongs to the WiMAX Forum. It is a wireless wide area broadband access technology offering 10 Mb/s over 10 KM although speeds up to 70 Mb/s are achievable over 10 KM. It is based upon the IEEE 802.16 standard.

The IEEE 802.11 MAC uses CSMA ('listen before talk') and is connectionless. The IEEE 802.16 uses MAC and supports full QoS, bandwidth on demand.It is connection oriented and supports centralised control and scheduling and offers multimedia support.

5.4.3 Bluetooth

Bluetooth is the most widely used technology for short range wireless communication. It supports a distance of 1 to 100 m depending on the class of device and power. Bluetooth is used for applications that are used for local communication as well as local control of devices. Bluetooth does not require a line of sight between the transmitter and receiver. Bluetooth is widely used for hands-free mobile-phone –headsets, car-kits and wireless controllers of game consoles.

Bluetooth uses the same radio frequencies as WLAN. It has a higher power consumption that results in a stronger connection. Bluetooth does not require a lot for configuration for setup. It is easy to connect devices using Bluetooth. Bluetooth device advertise themselves along with the services provided by the device. Hence, it becomes easier to discover and connect to new devices and utilize the services provided by them.

WLAN requires a lot of configuration during set up. It uses access points/nodes and is better suited for operating full-scale networks. It enables a faster connection, a better range from the base station, and better security than Bluetooth.

Bluetooth devices use an auto discovery mode to enable devices to discover each other within range. Bluetooth devices form a small adhoc networks called as a Piconet, which supports upto 8 devices. A piconet consists of one master and the devices connected to it act as slaves. A collection of piconet is called as a Scatternet. The scatternets are connected using a slave node as a gateway between 2 piconets. Scatternets use the TDM-MAC scheme to support shared access.

5.4.4 ZigBee

ZigBee is a suite of communication protocols from the ZigBee alliance formed in 2002. It uses small, low power digital radios based on the IEEE 802.15.4 standard for Wireless Personal Area Networks (WPAN). It uses a beaconing technique for node discovery where a node continuously transmits small packets to advertise its presence to other devices. The devices then connect to each other and start networks allowing other devices to join in.

ZigBee operates in the industrial, scientific and medical (ISM) radio bands; 868 MHz in Europe, 915 MHz in the USA and 2.4 GHz in most jurisdictions worldwide.

ZigBee protocols are usually used for embedded applications that require low data rates, low power consumption and low latency. ZigBee's current focus is to define a general purpose, inexpensive, self organising, mesh network that consumes very less amount of power. This will be useful in devices be used for industrial control, embedded sensing, medical data collection, smoke and intruder warning, building automation and home automation which can run for a year or two using the original battery. This technology needs less coding than Bluetooth.

Bluetooth needs a lot of power hence its devices go to sleep mode to reduce power consumption. A Bluetooth device takes around 3 seconds to wake up and respond. ZigBee can wake up and tranmsit a packet across a network connection in around 15 milliseconds.

There are three different types of ZigBee device.

- ZigBee coordinator (ZC): It is the root of the network tree and can bridge to other networks.
- ZigBee Router (ZR): It acts as an intermediate router, passing data from other devices.
- ZigBee End Device (ZED): The only function it supports is communication with its parent node that is either the coordinator or a router. It is incapable of relay data from other devices. It requires the least amount of memory of the 3 devices.

ZigBee protocols use an adhoc, on-demand, distance vector to automatically construct a low speed adhoc network of nodes. In most large network instances, the network will be a cluster of clusters. It can also form a mesh or a single cluster.

5.4.5 Infrared

Infrared (IR) is a short-range, low-bandwidth data communication technology used to communicate between devices. It is commonly used for remote controls and for communication between the computers and its peripherals (mouse).

Remote controls and IrDA devices use Infrared Light-Emitting-Diodes (LEDs) to emit infrared radiation. A plastic lens focuses this radiation into a narrow beam. The beam is switched on and off i.e. modulated to encode the data. At the receiver's end, a silicon photodiode is used to convert the infrared radiation to an electric current. It responds only to the rapidly pulsing signal created by the transmitter, and filters out slowly changing infrared radiation from ambient light.

Infrared communication is efficient indoors as it cannot be used outside due to interference from sunlight. It requires a line of sight communication between the transmitter and the receiver. It is high frequency so it does not penetrate walls and hence does not interfere with other devices in adjoining rooms. The most common application of Infrared is in remote controls to command appliances.

5.4.6 UWB

Ultra Wideband (UWB) is a technology that supports a data rate of more than 100 m bps using a large bandwidth that is greater than 500 MHz. It works on a low power range and over short distances.

Unlicensed use of UWB in the 3.1 to 10.6 GHz frequency range has been authorised by the FCC. This means that the UWB band can be efficiently used for short range, high data rate communication like BANs, PANs and within buildings as well for longer range, low data rate applications like radar, collision obstacle avoidance, precision altimetry and imaging systems.

UWB transmits signals at baseband frequencies. In pulse based UWB, the transmitter only needs to operate during the pulse transmission, producing a strong duty cycle on the radio and minimising baseline power consumption. As most of the complexity of UWB communication is in the receiver, simple, low power transmitters can be supported.

5.4.7 Satellite and Microwave Communication

Satellites cover a wide geographic area and hence it has the potential for truly ubiquitous global communication. Certain commercial satellites use parts of the microwave range frequencies for transmission.

The first communication satellites, TELSTAR 1 in the early 1960s, did not provide global coverage but acted as a mobile relay to store and forward transmissions between ground stations in different locations. The ground stations could only communicate for part of the day.

Geostationary satellites, launched in the mid of 1960s, remained in a fixed location and operated in the 4 and 6 GHz range(C Band). Geostationary satellites have simpler antennae design and configuration. These satellites can be interlinked to provide global coverage.

Satellite design must contend with station keeping design, e.g., the influence of the sun and moon in causing orbit perturbations, communication payload design, handling the large round trip delay for transmitting and receiving signals and efficient multiple channel utilisation.

Satellites were considered superior to undersea RF cables in terms of both bandwidth and cost for long distance communication. However, with the advent of fibre optic cables and the installation of terrestrial VHF transmitters for mobile phones, satellite communication is no longer considered superior.

Satellite communication was majorly used for wide area TV video broadcasts in the 2000s. However, cable networks have proved to be an equal for satellite communication. Satellite communication is used to serve very large numbers of thin routes, which cannot economically be served by cable.

New satellite systems, operating in the Ka band spectrum at 30 and 20 GHz, were developed in mid 1990s to provide two-way interactive services. These systems offered higher data rates 120 Gbps per satellite, a smaller equipment size and high gain spot beams to sequentially interconnect specific uplinks to specific downlinks using FDM or TDMA to access multiple channels. However, the Ka band frequencies are more susceptible to propagation impairments and are increasingly affected by the Earth's atmosphere.

5.4.8 Roaming between Local Wireless LANs

The internet that is accessed on wireless devices support roaming over several short range communications. It has an advantage of selection of an optimum network based on security, bandwidth, etc when multiple networks are available and it can be used to connect short-range services to long-range services. E.g. to notify someone remotely that something of interest remote is happening locally.

Let’s take an example of a logistics company which uses WLAN access points(short-range) in warehouses. This access points with interlinked scanners audit tagged goods (packages) that are being loaded, unloaded and are being stored. Global positioning system (long-range) capabilities built into the device can help dispatchers know the exact location of any truck. It can help in warning a driver about to deliver a package to the wrong location. Customers get up-to-date information about the location of their package as and when its location is updated.

Phones can use WLAN to access the Internet or use WLAN and VoIP to connect to an access node when in range of a WLAN transmitter instead of using WWAN network(data pack in cellphones) which is costly and slower to transfer data. A operating system support is required to support pipelines or transitions across heterogeneous networks.

Generic Access Network (GAN), also known as Unlicensed Mobile Access (UMA), is a telecommunication system that allows seamless roaming and handover between LAN and WAN using dual mode mobile phones. The idea behind GAN is to enable GSM mobile operators to offer fully converged connectivity using their existing network so that subscribers can move seamlessly from one cellular network to another without any disconnection during their calls. The variety of services and applications offered to customers has increased with the increase in the number of cellular operators. Hence, networks coverage has become a significant issue. Base stations act as a gateway. GAN technology is not used with 3G telecom networks. Multimode handsets can be used by operators to enable seamless access to services on different networks.

Femtocells are small cellular access points. They provide enhanced coverage in residential environments. They are used to provide fixed mobile converged voice, data and video services to consumers (IPTV).

Femtocells work with all UMTS terminals and are activated based on presence based activation. They supports standards beyond High Speed Packet Access(HSPA).

Femtocells does not really affect the dual mode handset market like GAN but both technologies can survive with new handover techniques based on SIP, e.g., VCC (Voice Call Continuity)

5.5 UNIVERSAL AND TRANSPARENT AUDIO, VIDEO AND ALPHANUMERIC DATA NETWORK ACCESS

Users access different types of content on different types of network. Each network has its own characteristic and properties. Each of them was developed by different technical groups. The users selection of network is purely based on costs and availability. Once a network is selected then the user decides what services to subscribe to over them. The high cost of mixing and matching of services from different providers is a prime deterrent for users to access multiple redundant networks. However, now-a- days services are being delivered over a common network such as IP or cable as well as network service providers bundle multiple service together as a single package.

There is a lot of difference between Audio-Video broadcast Content based networks and Telecom Networks.

First, the AVCBN streams multimedia i.e. audio and video content. These content is time-sensitive and any delay might lead to jitter. The data contains frames or individual data elements that needs to be accessed at the same rate it was created.

Second, Broadcast networks are simplex whereas Telecoms networks are duplex. Video content is usually region specific whereas telecom is global. Video content is richer and cutomised to a specific region in terms of language and culture. Hence, there are many national competitive and incompatible video broadcast systems. Telephone calls that are usually unicast whereas AVCBN is broadcast. AV CBN allows users control over the content. They can interact with the content by performing actions like view, pause, record and review or skip parts of audio video content. Streaming video data is more complex than streaming only audio data.

Internet was designed for transmitting alphanumeric data. It supports audio-video content but it is still maturing. The audio –video content broadcast entertainment network is quite matured. The adoption of compatible standards for the triple play (audio, video and alphanumeric data) will facilitate their integration. Common codecs could be reused across multiple audio video services, e.g. across digital radio, TV, voice broadcasts, audioplayers and the Internet to facilitate better Audio Video Content broadcasting on the Internet.

5.5.1 Combined Voice and Data Networks

The residential buildings usually have a single external communication line that is accessed via a modem. There are different types of modems like Cable, Digital Subscriber Lines (DSL) and its many variants, such as Asynchronous Digital Subscriber Line (ASDL) and air DSL. Users access multiple services, such as voice and text applications, and video over this line. It is cost effective. The voice and text data is transmitted over a single external telephone line connection (broadband). It has replaced ISDN and the traditional dial-up modem. Analog phone lines have a frequency of 3KHZ which is used to transmit voice and text data over telephone lines. The lower frequency signals are modulated to higher frequency to avoid interference with each other.

Cable TV modems utilize the unused bandwidth from analogue TV channels for data transmission. A splitter is used to split the TV signal from the data signal and numbers of individual cable users are multiplexed together at the cable TV provider.

DSL and cable modem use CDMA or FDMA to allow text and voice data to be used concurrently on the same access network. (Figure 11.2).

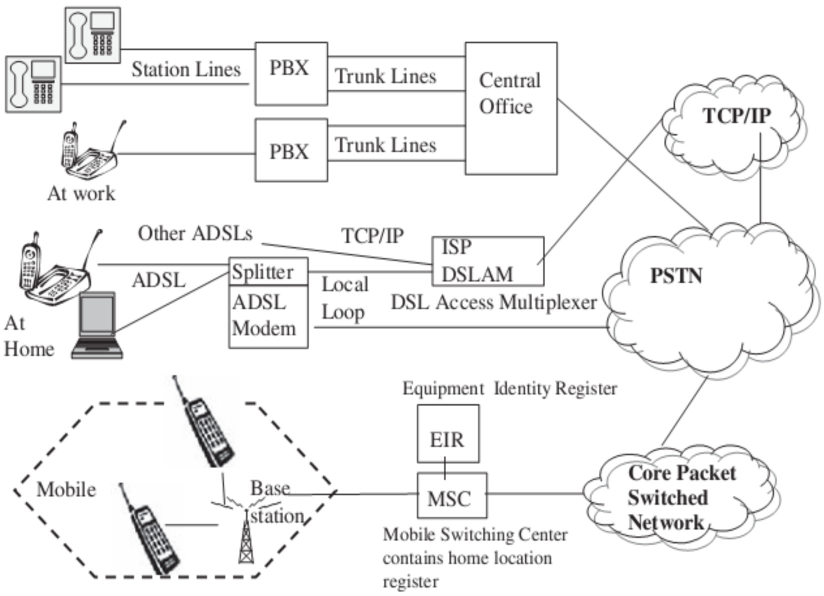


Figure 11.2 A typical telecoms network that can support voice and data over fixed and wireless links

#imgref: Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions

Voice over IP (VoIP) uses an IP or packet switched data network, to combine text data and voice to be transmitted over the same network. Three important communication requirements to be looked at while transmitting voice as data packets over IP are: to minimize delays, jitter and packet loss.

Delays can be caused by IP networks storing and buffering data during transfer. It may mislead the sender or receiver into thinking that the other party has paused the communication, when they have not. The communication is full duplex. Jitter refers to the variability of time during the transmission of data through the network. Packets can be get lost if there is a congestion in the network or if the receiver is overwhelmed with packets.

Delays, jitter and packet loss can be minimized by adding more network capacity, or by network routers being designed to prioritise data packets. The packets are classified as gold, silver, bronze, etc as per their priority level. The routers first process the packets with higher priority. So, voice packets can be given a higher priority over data packets for faster processing.

PSTNs use a codec of 64 kbps to digitise voice. VoIP digitisation can vary from 4 to 64 Kbps depending on the compression scheme and the sound quality. PSTN provide users with a range of voice services like signaling the caller ID before call is connected and also to support multi party conference calls.

VoIP uses two basic protocols namely IETF's SIP and ITU's H.323. Both of these use the Realtime Transport Protocol (RTP) for transmission of time sensitive messages. SIP is the more widely used protocol.

VoIP can be accessed via a software application installed on a phone or on a computer or via a dedicated VoIP hardware phone linked using a serial line such as USB or via a wireless link.

5.5.2 Combined Audio-Video and Data Content Distribution Networks

As mentioned before, the delivery of audio-video streams in more complex because of the synchronization between the audio and video of the content. Also, users need to interact with video streams where they can play, pause, rewind or record the content.

Video networks have a different view with respect to various stakehilders like viewers, content creators, network providers, content distributors, TV or display device manufactures and users. Users select services based upon cost , choice of content and availability within a specific geographic region.

There are three types of conventional audio video content delivery networks viz VHF TV, satellite TV and cable TV. Analog television can include digital data in the Vertical Blanking Interval (VBI) of the television signal. VBI has been used for Teletext data transmission in Europe as well as USA. Teletext is used for broadcasting a range of text based information like news, sports results, weather and displaying the TV schedules. It is also used for displaying the subtitles during video play. Teletext organizes information in numbered pages. The user can select a page to access particular information.

The Video content is usually broadcast in digital encoded form rather than analogue form. There are multiple standards for digital video broadcasting like the European DVB system, the US ATSC system, the Japanese ISDB system as well as some proprietary versions of these. The DVB system is most widely used because of its similarity to TCP/IP at an abstract level. DVB transmits content over multiple types of physical link such as satellite (DVB S, DVB S2 and DVB SH) cable (DVB C); terrestrial television (DVB T) and terrestrial television for hand helds (DVB H). Data is transmitted as MPEG 2 transport streams. The core link standards of ISDB are similar to those of DVB and are all based on the MPEG 2 video and audio coding and transport streams.

An illustrative video broadcast network that uses cable, based upon DVB C is shown in Figure 11.3. The cable provider's central office is called as a Head End. It multiplexes and encodes the

individual audio, video, and data components into an MPEG 2 stream. Streams may also originate from content providers and other broadcasters. Multiple program streams are multiplexed into an MPEG 2 Transport Stream at the Head End and are then passed over fibre optical physical links to the Cable Modem Termination Systems (CMTSs). Each CMTS can connect about 1000 or so customer premises. Splitters can be used at Customers' premises to multiplex video with text and voice data.

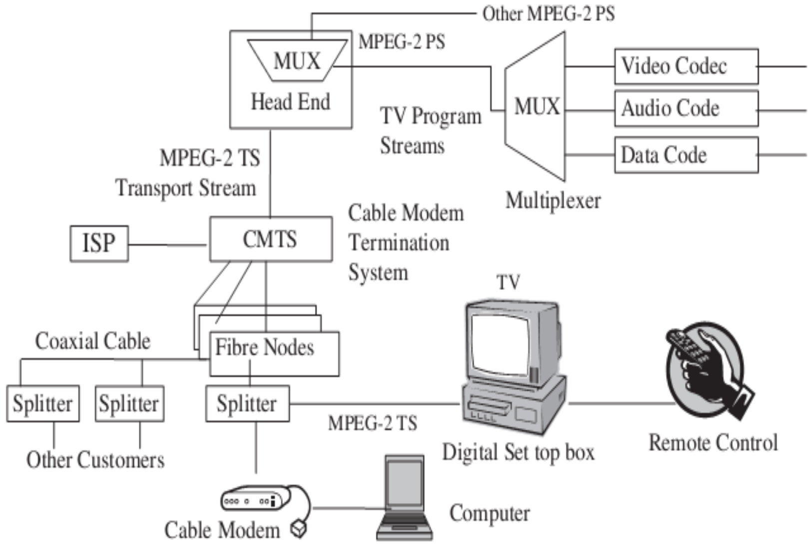


Figure 11.3 A video broadcasting network over cable that also supports the cable provider operating as an ISP

#imgref: Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions

Freeview is a digital video broadcast system based upon open standards used for UK digital terrestrial video broadcasts. It represents the first horizontal market in digital television in the world, with several multiplex operators and an open market in receivers. It uses an open standard API as its operating system. It uses the ISO MHEG 5 standard based on the work of DAVIC.

Digital TV broadcast both standard definition television (SDTV) and High Definition Television (HDTV) content. SDTV standards are analogue in nature. They derive much of their structure from the need to be compatible with analogue television. One of the challenges is to use the same network for various multimedia applications simultaneously. The concept of transmitting audio video and alphanumeric data content over fixed and mobile networks is called quad play networks. Streaming multimedia content which have different sensitivities to time delays and jitter can be a challenge. It can be achieved by adding protocols such as MPLS, Diffserv and RSVP to an existing IPv4 networks. MultiProtocol Label Switching(MPLS) was designed to provide a unified data carrying service for both circuit based and packet switching networks. It can be used to carry many different kinds of traffic over different types of underlying physical and logical networks.

MPLS supports processes packets faster by routing packets as part of a stream, Differentiated Services (Diffserv) tags priority label for each packet individually. The Resource Reservation Protocol (RSVP) reserves paths for different priority data. However, it cannot handle variable delays to individual packets hence it is used less.

5.5.3 On-demand, Interactive and Distributed Content

Conventional Audio Video content was very expensive to create. It was created to cater to a large diverse audience and was delivered over a small number of broadcast channels. The content was delivered with limited audience interaction. The interaction that the audience could perform was to decide which programs to view.

Video On demand (VoD) dedicates a single channel to each user. It enables the user to interact with the video to pause, rewind, fast forward, etc. TV services over IP based network (IPTV) uses MPEG 2 as the content format. DVB technologies acts and interface between an IP network and a digital TV receiver to enable live media broadcast and user interaction to pause and continue content on demand.

5.6 UBIQUITOUS NETWORKS

5.6.1 Wireless Networks

The advent of wireless technology may make us believe that we can have complete ubiquitous access. However, it is not really possible to have complete access but try to provide maximum ubiquitous access. Wireless networks suffer from various issues like attenuation, signals being obstruction due to solid objects or unable to penetrate walls. For example, GPS cannot be used indoors because it suffers attenuation by buildings. Analog signals tend to fade due to multipath effects however they can be overcome in digital transmission by error correction techniques.

Transmitter used in wireless networks have a limited range. When multiple transmitters are located in the same space they may use different frequency ,code or time division multiplexing to avoid interference with each other. When a user is moving between adjacent transmitters; receivers that switch between frequencies easily will allow seamless transmission. Frequency allocation may be region specific so same frequency cannot be used universally. Hence, it can be difficult to switch between wireless networks.

Wireless terminals need to be plugged to a power outlet. A wireless device consumes more power and can drain quickly. A device that runs out of power cannot be used for transmission. A lot of power saving mechanisms can be implemented like going to sleep, transmitters polling for receivers in range or vice versa.

Security is another cause of concern for wireless devices. It can easily be eavesdropped on or can be access if no security measures are implemented.

Wireless networks, particularly WLAN, use the unlicensed frequency band 802.1 which can be used by other devices like a microwave oven and it might interfere in transmision.

If wireless networks use a single RF channel in each node, that channel is used both for peer to peer access and for 'backbone' access to a wireless access point or gateway to a wired network. To overcome these problems, MESH networks have been proposed, these use multiple RF transceivers and channels, together with collision detection and avoidance.

5.6.2 Power Line Communication (PLC)

Power Line Communication (PLC) uses the electricity power line for transmission of data. It is a wonderful alternative to ubiquitously access data and audio video content. The electricity line network can modulate electricity as a signal and can be used as a channel to communicate data and audio video content. PLC describes a range of systems for using electricity distribution wires for transmission of data. The carrier superimposes a modulating signal over the standard 50 or 60 Hz alternating current (AC). PLC can be used in home automation for remote control of lighting and appliances without the installation of additional control wiring, e.g., X10 and the HomePlug powerline alliance. HomePlug defines a number of standards for PLC, for transmitting HDTV and VoIP around the home. PLC can also be used to provide Internet content in electric vehicles that are permanently plugged to an electricity supply like electric trains, buses and trams as well as in underground city metro where mobile communication suffers because signals cannot penetrate ground.

5.6.3 Personal Area Networks

Wireless Personal Area Networks (WPAN) have been specified by the IEEE P802.15 working group. A Personal Area Network (PAN) extends up to 10 meters in all directions and envelops two or more objects or persons whether stationary or in motion.

Initially WLAN standard was considered for PAN. However, it was dropped because of its high power requirements and higher network management overhead.

PAN are usually implemented using Bluetooth, Zigbee and InfraRed. PAN applications can be as simple as the connection between your mobile and laptop, it can be a multi network phone like intercom, cordless phone and cell phone; Internet access on the mobile phone while in motion;

interactive local conferencing to instantly exchange information; hands free head set to communicate and (voice) control and automatic synchronisation of information for mobile users.

5.6.4 Body Area Networks

Body Area Networks are networks made up of sensors that are wearable or implanted on a human body. These sensors communicate with each other and other devices. They are therefore used for monitoring the vital body parameters and movements. The data collected by these sensors is either stored on the device on the body or uploaded on an external storage device.

A BAN can be used for the simultaneous acquisition of EEG, (electrical brain activity), ECG (electrical cardiac activity) and EMG (electrical muscle activity) signals and transmission of these signals to a base station.

Bluetooth and Zigbee cannot fulfill the power and bandwisth requirements of a BAN. BAN radio chipsets consume in the order of 10 to 100 mW for supporting data rates of 100 to 1000kbps, leading to a power efficiency of roughly 100 to 1000 m W/Mbps or nJ/bit. Since, wireless transmitters need to be one or more order power efficient , UWB is used because it supports a data rate of 10 kbps with 5mW power consumption.

A BAN can be considered as a subset of a PAN. It is more personal and it uses the bosy as a conduit between devices.

Electrical conduction through the body can cause inductive and capacitive fields to protrude outside the body. Zimmerman (1996) distinguishes nearfield (capacitive and inductive electric field effects) and far field (radio) communication in PANs. Far field communication is more susceptible to eavesdropping and interference. Radio transmission loses signal strength with square of distance whereas nearfield loses strength distance cubed. Near field operates at lower frequencies (0.1 to 1MHz) than farfield.

Zimmerman has proposed the use of electronic devices connected as part of near field BANs to exchange digital information by capacitively coupling picoamp currents through the body. He has demonstrated a near field BAN system to exchange business cards via a handshake.

5.6.5 Mobile Users Networks

Mobile user networks deals with networks that are wired or wireless, which can be accessed by a user in motion. Mobility refers to mobile applications where users are free to continuously move anywhere within range of a wireless network. Portability refers to a mobile device that is stopped or suspended during the roaming itself but can be attached at discrete access points (Laptop).

The challenges that networks face while supporting mobile users include how to locate and address mobile users, how to route data to mobile receivers, channel allocation for multiple users whether or not there is a fixed network topology or infrastructure or an ad hoc one.

Mobile network support can be classified on the basis of whether or not the network infrastructure is fixed or dynamic; whether or not the user is assigned a new ('care of') address every time the user changes location or keeps the same address independent of location. Mobile user support at the network level allows transparency to mobile applications.

5.6.5.1 Mobile Addresses

The transmission of data for a mobile device is possible only if the user's location is known or if there is an address for communication. There are two basic approaches to communicating with a mobile user: keeping the address the same in different locations or giving a new address for every change of location. If the same address is kept in different locations, then the network must keep track of users' nearest base station or access point address. DHCP (Dynamic Host Configuration Protocol) is used for acquiring new address.

If the user changes address depending on location then routing becomes more complex. The user's address is updated by a base station paging mobile users periodically or the mobile user can notify services of updates whenever there is a significant move.

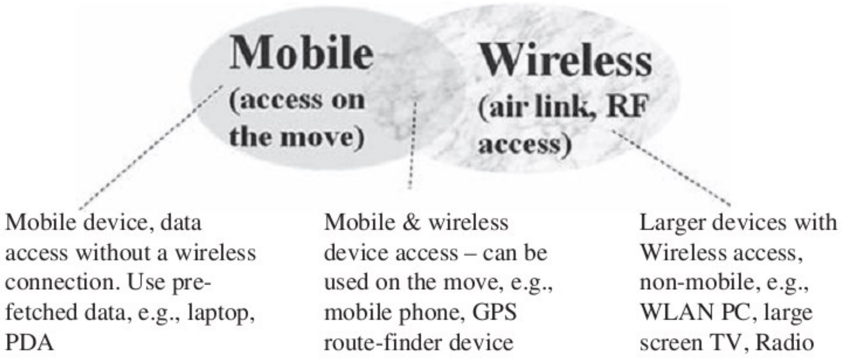


Figure 11.4 The difference between mobile and wireless

#imgref: Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions

5.6.5.2 Single -Path Routing

There are two types of routing for mobile users: single path routes and wireless ad hoc, multi path routes. Mobile IP consisting of Mobile Node, Home agent and foreign agent is an example of a single path route.

A mobile node is a host computer or router that moves from one network to another without changing its home IP address.

A Home Agent is a router on a mobile node's home network that is responsible for delivering data to the mobile node. It maintains the current location information for each mobile node.

A foreign agent is a router on a mobile node's visited network that works along with the home agent to complete the delivery of data to the mobile node while it is away from home.

Mobile IP performs three main functions:

(1) Discovery: mobility agents advertise their availability on each link for which they can provide a service. It is also used by mobile nodes for DHCP to receive a care of address;

(2) Registration: When the mobile node is away from home, it registers its care of address with its home agent;

(3) Tunneling: is used for data to be delivered to the mobile node when it is away from home, the home agent has to forward (tunnel) the data to the care of address.

5.6.5.3 Multi- Path Routing in Mobile Adhoc Networks (MANETs)

Transmission in fixed networks consists of computers having fixed address connected to a network. If devices fall in the same network domain they do not need routing. Devices on different network domains need routing.

Ad hoc networks are networks that are established on the go. They are completely volatile and are never fixed. These devices use connections that are established for the duration of one session. They do not require any base station or fixed router. Adhoc networks that support mobile nodes are called Mobile Ad hoc Networks(MANETs). In MANET each node is acting as a forwarding device. The determination of which nodes would forward data is made dynamically. Hence, they are called as Adhoc Networks. Adhoc networks are formed when devices discover each other and connect to them. If a device cannot find the destination address for data transmission, the network is flooded by broadcasts that are forwarded to each node. Connection can be direct or over multiple nodes (multihop ad hoc network). Routing protocols provide stable connections even if nodes are mobile. Agents are node that communicates over a network. MANETs can be used in situations establishing a network infrastructure is difficult like in natural disasters and in armed conflict situations. It is currently used in playstations.

Ad hoc routing is often more complex than single path routing. It can be classified based upon whether

- the routing is predetermined or on demand,
- uses periodic updates versus event driven updates to routes,
- uses flat versus hierarchical structured of network nodes,
- uses source routing versus hop by hop routing and
- whether single paths or multiple paths are used.

5.7 FURTHER NETWORK DESIGN ISSUES

5.7.1 Network Access Control

A network is made up of multiple devices sharing the link/channel simultaneously, which can lead to collision. Hence, every network has an access control mechanism for users to access network. There are various multiple access schemes available which can simply detect collision or detect and avoid collision like CSMA,TDMA,CDMA, etc.

TDMA stands for Time Division multiple access. In this scheme, every device is allotted a time slot during which the device is allowed to access the channel. GSM used TDMA.

Another popular scheme is CDMA, which is a type of spread spectrum. In CDMA, all devices can transmit simultaneously. The sender and receiver use its own chip sequence to encode/decode the data.

New generation of mobile technologies, use WCDMA which supports higher bandwidths. WLANs tend to suffer from a lot of collisions because the transmission by certain devices, which are out of range of the transmitting device, might not be visible. This is called as hidden station problem. Hence, WLANs use Multiple Access with Collision Avoidance (MACA).

MACA uses two control packets viz. Clear To Send(CTS) and Request To Send(RTS) for signaling before starting transmission. Whenever a wireless device needs to send data it sends a RTS alongwith the length of the data frame to notify all the devices before starting transmission. All devices on the wireless network receive a RTS. The device waits for the duration of packet i.e. it keeps silent and does not transmit. The receiving device on receiving RTS replies with a CTS notifying the sender its availability for transmission. All devices wait for the duration mentioned in the CTS and a random period of time before they start transmission.

Carrier Sense Multiple Access with Collision Detection (CSMA/CD) is a protocol used in Ethernet LAN that offers collision detection and collision avoidance. In this scheme, if the sender detects another transmission, it stops its current transmission and sends a jamming signal notifying other devices of probable collision. It then waits for a random amount of time before attempting retransmission.

Token based access control system use a signal called as token to implement access control. Any device which needs to transmit should capture the token. This token is circulated continuously throughout the system. Since, only one device can retain the token at a time, simultaneous multiple access is not allowed, avoiding collision. The token status is set to a state which indicate if the network is busy or free. If device wants to send data it captures the token and changes the state of the token to busy. Once the transmission is complete, the token state is set to free so others can transmit. There is a limit to the duration of transmission. There are designated nodes which manage the creation of tokens and deal with lost tokens.

5.7.2 Ubiquitous Versus Localised Access

Not all networks are designed to be ubiquitously accessed, some services support only local access. There are several scenarios at play here:

Services which are tailored to cater locally or a particular set of users and application like hospitals in the neighbourhood, hotels delivering food nearby, etc. Such services only advertise to local users or users with a particular radius.

Services can be local if they are available within a domain on a wired network like the internal network of an organization. All data on the networks is available only to on-premise users on the network.

Services can be limited to access for users operating on a Virtual Private Networks or VPN.

Wide Area Networks are used to support ubiquitous access. A WAN is made up of multiple distinct networks that are further internetworked. These networks should be designed to offer seamless handover or transparent routing between different networks.

There are three types of WAN:

1. Internet (Wired and Wireless)
2. Satellite Networks (Wireless)
3. Telephone Networks (Wired and Wireless)

Each use their own protocol for managing networks.

5.7.3 Controlling Network Access: Firewalls, NATs and VPNs

The access to the network may be limited to a specific user group. Usually, access to ICT devices can be overloaded or overused by freeloaders leading to genuine users being refused connection. The access to such resources can be restricted by limited access based on IP Address, Authentication based on Digital certificates or by the use of firewalls, NATs and VPNs.

A firewall monitors all the incoming and outgoing packet to a network. It acts as a gatekeeper to the network. The access rules are based on various parameters like IP Address, port number, application or service type, etc. There are two levels of firewalls implementation: packet level and application level firewalls.

Packet level filter as the name suggests examine the various fields of incoming as well as outgoing packets. They apply the rules and perform action accordingly. Packets filters are usually used as first line of defense. They are easy to implement and are great for basic filtering of traffic. However, they are susceptible to packet spoofing. Packet spoofing is replacing the original IP address of a packet (attacker) with another IP Address (genuine or unsuspecting user).

Application level firewalls make it mandatory to login into the firewall computer for access to the network.

Network Address Translation (NAT) is used to use a limited set of IP Address in a network with several computers. NAT has two set of IP Address: Private addresses and Licensed IP Addresses. The devices on the network use private IP addresses. These cannot be used for external communication. Whenever a packet is sent across the network, NAT replaces the internal private address with the external IP address. NAT can be used to solve the problem of limited IP address as well as for security purposes.

Firewalls support multihomed host i.e. they have multiple interface to a network or every interface might cater to a different network. Organizations can use different combinations of Firewalls and NATs to protect vital network resources. With the increase in the use of Internet on devices at home, using firewalls at home has also become a necessity.

Access to networks can be limited to the type of user or role specific. Some users can be given limited remote access to the network by means of a Virtual Private Network or VPN.

VPN use tunneling for accessing network resources. Tunelling is establishing lower layer virtual channel for data transmission which is hidden from higher layers. Since, multiple heterogenous applications share the same tunnel, the packets are encrypted. Users accessing the VPN are authenticated at a VPN client or access device. VPN tunnels can be differentiated based upon the type of encryption or access control used.

5.7.4 Group Communication: Transmissions for Multiple Receivers

Group communication is also called as multicast communication. It is used for sending messages from a single source to multiple destinations like passing information within a group or broadcasting information to a limited devices/users.

Multicast can be done easily with hardware support. However, in the absence of hardware support, messages are sent sequentially by the routers.

Multicast is used for sending messages in bulk. Hence, to avoid the overhead of managing large groups, the group members are split into hierarchies, scaling the management of users to smaller groups.

There are different types of multicast:

- Unreliable – Messages are transmitted without confirming the receipt of message. There is no effort taken to ensure delivery of message.
- Reliable- The transmission of message is monitored to provide best effort delivery. It keeps track of the transmission of message. Some members of the group may not receive a message.
- Atomic multicast : The message is received by all or none of the users of the group.

5.7.5 Internetworking Heterogeneous Networks

Universal access means any type of data can be accessed anywhere over any kind of network simultaneously. There are various types of communication networks, which might not be compatible with each other. They might use different technology or protocols or media for communication. For example, internet can be accesed using wired , wi-fi or Bluetooth network.

Physical media have different capacities and hence different requirements with respect to repeaters and amplifiers to overcome attenuation. Every network has its own protocol for handling and forwarding of data. Some networks are diverse with respect to the type of data they support.

Networks can also be differentiated with respect to their topology, architecture or how networks are interlinked and managed.

Hence, it is absolutely essential to standardize the techniques to interlink heterogenous networks such that the interconnection complexities remain hidden. Standards for universal networks are based upon the TCP/IP and Universal Mobile Transmission Service (UMTS).

5.7.6 Global Use: Low-Cost Access Networks for Rural Use

Wireless networks are the backbone of truly ubiquitous computing. However, in practice, wireless networks are not available everywhere. Wireless network communication and content delivery networks are built and operated as commercial businesses that have to generate sufficient revenue to exist.

The total worldwide Internet usage penetration is only about 18% and it is even lesser in certain regions. For example, it is 4% in Africa. This is contrast to the global population using GSM type mobile phone technology (29%) and more if other types of mobile phone are also included.

The cost of accessing wireless networks can be a deterrent for people in rural areas. Hence, there is a need to provide low cost networks and access terminals.

CorDECT has been proposed in rural india to support local access. It is based on DECT, which was designed for cordless telephones. It uses MCTDMA for multiple access. It uses both time and frequency division to accommodate multiple channels. It supports data rates of 35 and 70kbps over a distances of up to 10 km.

Another way of providing low cost access is to use TV broadcast network. The PrintCast system can be used to transmit information that can be printed on paper. It augments the TV VBI signal with print related information to allow users to print Data. PrintCast combines print data with AV signal using a device called an inserter. The inserter encodes the signal and transmits it via traditional broadcast equipment. Viewers use a PrintCast decoder to decode the data from an analog television signal and relay it to a printer. PrintCast can be used for voter information guides for elections, advertisements coupons, homework exercises for distance learning, and printing recipes for cooking shows.

Very Small Aperture Terminal (VSAT) is a product range based on principles for designing small ground terminals. It enables cheaper satellite access, and Spread Spectrum, a communications technique for the transmission of signal over some band of radio spectrum.

Mesh network can also be used for providing low cost networks and access terminals.

5.7.7 Separating Management and Control from Usage

A plethora of options is available to separate concerns about application use from concerns of control and management of networks.

Management is concerned with the global, system wide, regulation function, FCAPS functions, etc. Control is usually limited to local functions such as synchronisation, error correction, network congestion, and forwarding and routing.

In an application centric model, the user data flow, the control and management are application driven take place over a single communication channel.

The control signaling is carried out by reserving a portion of the channel or adding a separate channel. Adding a separate channel is better because it is not affected by the user error or signaling error. PSTN and TCP/IP-FTP use separate channel for sending control information.

Certain application use their own dedicated network. Hence, the management needs to be factored out of the applications and networks.

5.7.8 Service-Oriented Networks

Application services differ with respect to their requirements for transmission such as security, scalability, performance, latency, etc. Hence, application services were coupled to specific networks.

It is not possible for a single network to support different levels of service. Hence, services are designed to be functional across heterogenous networks.

Service architectures were network oriented. A user had to subscribe to a particular network to use a service. Ex: subscribe to telecom for audio calls and content networks for audio-video content.

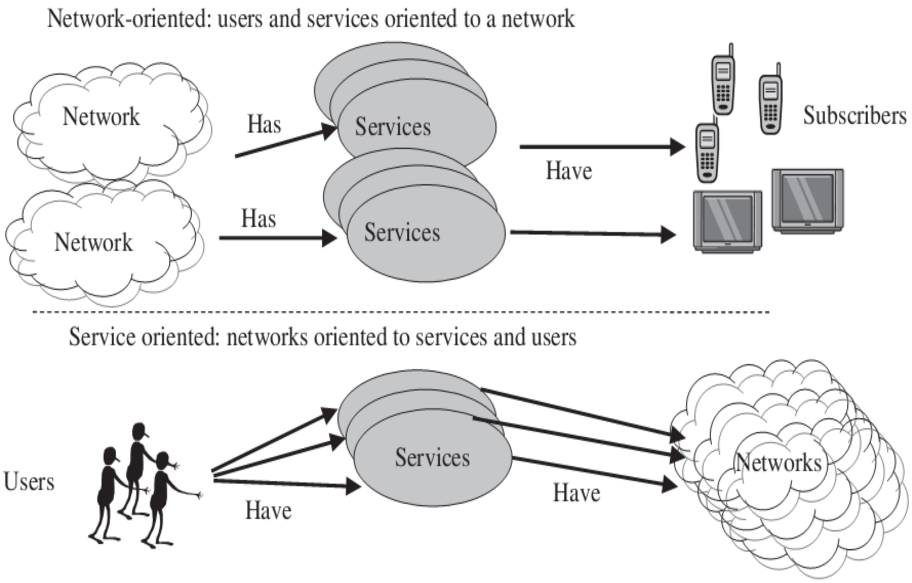


Figure 11.7 From network oriented service models to service oriented network models

#imgref: Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions

5.7.8.1 Service-Orientation at the Network Edge

Universal communication is supported by partitioning the communication network into an access network (edge network) and a core network. Although, it might appear as a single universal network. The core network comprises of the interlinked network. The Access or Edge networks are used by users and applications to access the core network.

A major point to consider is whether the complexity or intelligence related to heterogeneous service is to be part of the core network. Although there are many theories related to this design issue, it is preferable to keep the core network simple and try to use a simple and neutral network. The applications in the Edge network should handle all the complexities and intelligence. Internet is the best example of universal interoperability network which supports a variety of services across heterogeneous devices. However, some service providers are against the concept of universal access for fear of losing their market share.

Network homogeneity can lead to reduction in diversity and by creating a bias in the market against certain types of applications. This can further lead to monopoly power by dampening incentives to invest in alternative network neutrality.

Hence, Yoo suggests that it is better to have a heterogeneous network. The core network should support multiple core network functions. Nikolaidis states that users expect a diverse portfolio of universal services from the network.

5.7.8.2 Content-based Networks

Traditional network communication was based on a sender transmitting some data to a receiver. It was destination specific. Content-based network on the other hand is based on the content of message to be transmitted. Senders publish content with no specific receiver in mind. The receivers subscribe to the type of content that they are interested in with no regards to the location or source of the content.

Conventional CBN allocate bands of addresses to each type of content then multicast the content to this band of addresses. Another approach is to use a publish/ subscribe middleware service. However, both of these fail to support diverse kinds of content and interests along with high performance.

In CBN, each node advertises a r-predicate or receiver predicate. A r predicate indicates the data of interest that the node intends to receive. Nodes send datagrams that are forwarded to nodes with r-predicates.

The router is responsible for forwarding and routing of content. The router uses an internal forwarding table that is updated by gathering, combining, and exchanging predicates and other routing information with adjacent nodes.

5.7.8.3 Programmable Networks

The router controls environment algorithms and updates the router states in a network. They cannot be updated by the service providers. Hence, it is difficult to deploy new network services. Programmable networks are network where some of the network elements can be reprogrammed dynamically. This is achieved by injecting executable codes, supporting application specific services, into network elements such as routers and switches. Programmable networks might lead to increased complexity, instability and security issues due to malicious code being programmed into the core network elements.

DARAPA's Active Networks (AN) program and the Open Signalling (Opensig) community are two of the main initiatives for establishing programmable networks identified by Campbell et al. AN is injecting executable code in the network whereas Opensign seeks to establish open programmable network interfaces.An example of open network APIs is the Parlay/OSA initiative.

5.7.8.4 Overlay Networks

An overlay network is a virtual network built on top of the physical network. It provides infrastructure to one or more applications by handling data similar to a physical network. It is operated by third parties which may be a collection of end users. Overlay networks allow the introduction of new services without making changes to the existing network. It also support different service levels for applications by deploying multiple Application-specific overlay networks.

Overlay networks were used by Telcos to offer non-geographical phone number services such as free phone (0800) numbers, local rate phone and premium phone rate numbers which charge the same rate irrespective of the location.

5.7.8.5 Mesh Networks

Mesh topology consists of a dedicated point-to-point link between every node of the network. It needs a lot of wiring and is costly to implement. Hence, it is usually used only for smaller networks. Mesh topology is used along with other topologies to connect networks.

Wireless Mesh Networks (WMNs) are partial mesh, ad hoc, networks. They perform better at a lower cost and at a lower power output compared to other types of WLAN. WMN consume less power because it uses a set of lower power multihop transmissions.

WMN is a probable solution in rural areas. WMN receiver are complex and costly because they act as a relay. WMN does not have a separate base station. Each wireless receiver acts as a relay point or node for other receivers within range. The data signal passes through several nodes as it traverses from the core network to its destination.

WMNs support interoperability between heterogeneous wireless networks. Each node operates, as a host as well as a router, forwarding packets on behalf of other nodes that may not be within direct wireless transmission.

DHT is a hash table (HT) which stores values that can be looked up if you have the key. A simple (hash) function converts the key (name) to a storage address. A DHT distributes and stores the HT over multiple computer nodes. DHT are usually designed as a circular, double linked list.

Multihop routing is used by sensors to reduce power transmission costs. Dedicated mesh routers are equipped with additional routing capabilities and bridging and gateway function to other networks.

WMN support Adhoc multihop routing by dynamically self-organising and self-configuring mesh connectivity. The main characteristics and benefits of WMN as:

- Multi hop wireless network: extends coverage range without sacrificing the channel capacity, without a direct line of sight (LOS) link.
- Ad hoc networking: supports self forming, self healing, and self organisation for multipoint to multipoint communications enabling the network to grow gradually as needed.
- Mobility dependence on the type of mesh nodes: mesh routers usually have minimal mobility,while mesh clients can be stationary or mobile nodes.
- Multiple types of network access: both backhaul access to the Internet and peer to peer (P2P) communications are supported.
- Mesh routers usually do not have strict power consumption constraints and may not be appropriate for some types of mesh clients, those that act as sensors in a sensor network where power consumption is the primary concern.
- WMNs are compatible and can be integrated with multiple IEEE 802.11 (Wifi) type networks.
- WMNs aim to diversify the capabilities of ad hoc networks.

Ad hoc networks can be considered as a subset of WMNs. WMNs consists of a wireless backbone with mesh routers to provide large coverage, connectivity, and robustness in the wireless domain. They reduce the load and energy consumption on end user devices.

Mesh routers can perform peer-to-peer routing and backbone access functionalities which separates the types of traffic in the wireless domain and increases performance. Ad hoc networks provide routing using the end user devices, the network topology and connectivity which further depends on the movement of users. Adhoc network routing protocols, network configuration and deployment is complex.

5.7.8.6 Cooperative Net works

Network access devices are programmed to access only one type of network. They cannot support heterogeneous network access. However, some network access devices have inbuilt support for heterogeneous network access, e.g., to Bluetooth, Infrared, WiFi and GSM networks. Each of these networks must be used one at a time. Multiple types of the same type of physical and network layer may exist because multiple independent users and providers may offer overlapping wireless networks within the same vicinity. However, they do not interoperate.

Cooperative communication is designed to enable single antenna mobile access devices to being Multiple Input Multiple Outputs (MIMO) systems. It can solve the problem of signal fading due to thermal noise, shadowing due to fixed obstacles and signal attenuation. It transmits independent copies of the signal that will face independent fading. This generates diversity and can effectively combat the deleterious effects of fading.

Cooperative communication treats communication nodes as a part of the WMN. It multicasts the same message over multiple routes with different fading effects to the same receiver. It can also be used to reduce power consumption for transmission.

5.8 EXERCISES

1. What is the advantage of Bluetooth over InfraRed?
2. Describe the types and characteristics of mobile networks
3. Discuss overlay networks
4. How is WMN better than WLAN?

5.9 REFERENCES:

1. Stefan Poslad - Ubiquitous Computing: Smart Devices, Environment and Interactions- Chapter 11

❄❄❄❄❄❄❄